Universität Rostock

Traditio et Innovatio

Coarse-Grained Modeling of Interactions between Pollutants and Soil Organic Matter

Kumulative Dissertation zur Erlangung
des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.) der
Mathematisch–Naturwissenschaftlichen Fakultät
der Universität Rostock

Vorgelegt von:

Lorenz Friedrich Dettmann

Matrikel–Nr.: 215204751

Eingereicht am:

14.11.2025

Betreuer:

Priv.-Doz. Dr. Ashour Ahmed

Prof. Dr. Oliver Kühn

https://doi.org/10.18453/rosdok_id00005889

Institut für Physik

Mathematisch Naturwissenschaftliche Fakultät | Universität Rostock

Albert-Einstein-Straße 23–24, D–18059, Rostock



Dieses Werk ist lizenziert unter einer
Creative Commons Namensnennung 4.0 International Lizenz.

Abstract

Soil organic matter (SOM) is a highly complex and heterogeneous component of soils, whose molecular structure and interactions strongly influence soil properties and functions. Despite its ecological importance, many of the microscopic mechanisms that govern SOM organization and its interactions with other compounds remain poorly understood, largely due to experimental constraints. This is particularly true for interactions with environmental pollutants and emerging contaminants such as nanoplastics (NPs), which are difficult to probe directly. Molecular simulations offer a complementary approach to addressing this gap, provided that relevant time and length scales of the underlying processes can be accessed. This work presents a coarse-grained modeling framework for SOM, derived from the Vienna Soil Organic Matter Modeler 2 (VSOMM2) and implemented in an end-to-end tool named SOM2CG. The tool converts atomistic SOM models generated by the VSOMM2 into coarse-grained representations, enabling microsecond-scale simulations of SOM–pollutant interactions. For the coarse-grained model development, VSOMM2 fragment molecules were parametrized using the Swarm-CG program and by matching transfer free energies before assembling into larger SOM molecules for benchmarking against atomistic references. To capture chemical diversity, coarse-grained SOM models of varying compositions were generated and automatically parametrized. The systems were subsequently applied to study pollutant mobility, focusing on hexachlorobenzene (HCB) and sulfanilamide (SAA), and to investigate SOM interactions with NPs, including both single polymer chains and a polyethylene (PE) nanoparticle under different pH conditions. The parametrization of the fragment database yielded coarse-grained molecules that reproduced size and shape with satisfactory accuracy, as reflected in the radius of gyration and solvent-accessible surface area. Thermodynamic consistency was supported by transfer-free-energy calculations, which showed good agreement with hexadecane–water and chloroform–water references, while larger deviations were observed for octanol–water values. Simulation of coarse-grained Leonardite humic acid revealed that the models reproduced system densities close to the atomistic references, particularly at higher hydration levels. Extending the analysis to diverse SOM compositions demonstrated that electrostatic interactions dictated structural packing, with acidity influencing both densities and potential energy profiles. Pollutant simulations showed that SAA largely followed water dynamics and exhibited higher diffusivity, whereas HCB partitioned more strongly into the SOM phase and displayed reduced mobility. Microsecond simulation times further revealed the temporary trapping of both pollutants within the SOM matrix. Finally, nanoparticle simulations demonstrated that low pH conditions destabilized SOM aggregates, promoting their dilution and enhancing PE surface coverage. Overall, this work presents a coarse-grained SOM modeling framework that captures key structural and interaction mechanisms across diverse compositions. It provides molecular-level insights into pollutant mobility as well as NP–SOM interactions, and lays the foundation for future studies on SOM–pollutant processes at extended spatial and temporal simulation scales.

Zusammenfassung

Organische Bodensubstanz (SOM) ist ein hochkomplexer und heterogener Bestandteil von Böden, dessen molekulare Struktur und Wechselwirkungen die Eigenschaften und Funktionen des Bodens stark beeinflussen. Trotz ihrer ökologischen Bedeutung sind viele der mikroskopischen Mechanismen, die die Organisation der SOM und ihre Wechselwirkungen mit anderen Verbindungen steuern, nach wie vor kaum verstanden, was vor allem auf experimentelle Einschränkungen zurückzuführen ist. Dies gilt insbesondere für Wechselwirkungen mit toxischen Substanzen und neu auftretenden Schadstoffen wie Nanoplastik (NP), die nur schwer direkt untersucht werden können. Molekulare Simulationen bieten einen ergänzenden Ansatz, um diese Lücke zu schließen, vorausgesetzt, dass die relevanten Zeit- und Längenskalen der zugrunde liegenden Prozesse zugänglich sind. Diese Arbeit präsentiert einen grobkörnigen Modellierungsrahmen für SOM, der aus dem Vienna Soil Organic Matter Modeler 2 (VSOMM2) abgeleitet und in einem End-to-End-Tool namens SOM2CG implementiert wurde. Das Tool wandelt atomistische SOM-Modelle, die mit dem VSOMM2 generiert wurden, in grobkörnige Darstellungen um und ermöglicht so Simulationen der Wechselwirkungen zwischen SOM und Schadstoffen im Mikrosekundenbereich. Für die Entwicklung des grobkörnigen Modells wurden VSOMM2-Fragmentmoleküle mit dem Programm Swarm-CG parametrisiert und mit freien Transferenergien abgeglichen, bevor sie zu größeren SOM-Molekülen für den Vergleich mit atomistischen Referenzen zusammengebaut wurden. Um die chemische Vielfalt zu erfassen, wurden grobkörnige SOM-Modelle unterschiedlicher Zusammensetzung generiert und automatisch parametrisiert. Die Modelle wurden anschließend zur Untersuchung der Schadstoffmobilität unter Verwendung von Hexachlorbenzol (HCB) und Sulfanilamid (SAA) sowie zur Untersuchung der Wechselwirkungen von SOM mit NP, darunter sowohl einzelnen Polymerketten als auch einem Polyethylen (PE)-Nanopartikel unter verschiedenen pH-Bedingungen, eingesetzt. Die Parametrisierung der Fragmentdatenbank ergab grobkörnige Moleküle, die Größe und Form mit zufriedenstellender Genauigkeit reproduzierten, was sich im Gyrationsradius und der lösungsmittelzugänglichen Oberfläche widerspiegelte. Thermodynamische Konsistenz wurde durch Berechnungen der freien Transferenergie bestätigt, die eine gute Übereinstimmung mit den Referenzwerten für Hexadecan-Wasser und Chloroform-Wasser zeigten, während bei den Oktanol-Wasser-Werten größere Abweichungen beobachtet wurden. Simulationen von grobkörniger Leonardit-Huminsäure ergaben, dass die Modelle Systemdichten reproduzierten, die nahe an den atomistischen Referenzwerten lagen, insbesondere bei höheren Wassergehalten. Die Ausweitung der Analyse auf verschiedene SOM-Zusammensetzungen zeigte, dass elektrostatische Wechselwirkungen die strukturelle Aggregation förderten, wobei der Säuregehalt sowohl die Dichten als auch die potenziellen Energieprofile bestimmte. Weitere Simulationen mit Schadstoffen ergaben, dass SAA weitgehend der Wasserdynamik folgte und eine höhere Diffusionsfähigkeit aufwies, während HCB stärker in die SOM-Phase überging und eine verringerte Mobilität zeigte. Simulationszeiten im Mikrosekundenbereich offenbarten darüber hinaus eine vorübergehende Fixierung beider Schadstoffe

innerhalb der SOM-Matrix. Schließlich demonstrierten Nanopartikel-Simulationen, dass niedrige pH-Werte die SOM-Aggregate destabilisierten, ihre Auflösung förderten und die PE-Oberflächenbedeckung erhöhten. Insgesamt schafft diese Arbeit einen grobkörnigen SOM-Modellierungsrahmen, der wichtige Struktur- und Wechselwirkungsmechanismen über verschiedene Zusammensetzungen hinweg erfasst. Sie liefert Einblicke auf molekularer Ebene in die Mobilität von Schadstoffen sowie in NP-SOM-Wechselwirkungen und ermöglicht zukünftige Studien über Prozesse zwischen Bodenmolekülen und Schadstoffen auf erweiterten räumlichen und zeitlichen Simulationsskalen.

Gutachter: Priv.-Doz. Dr. Ashour Ahmed
Leibniz-Institut für Katalyse e. V. (LIKAT Rostock)
und Institut für Physik, Universität Rostock

Prof. Dr. Oliver Kühn
Institut für Physik, Universität Rostock

Prof. Dr. Chris Oostenbrink
Institut für Molekulare Modellierung und Simulation,
Universität für Bodenkultur Wien

Associate Prof. Dr. Mark Miller
Department of Chemistry, Durham University

Tag der Verteidigung: 24.04.2026

Eidesstattliche Versicherung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, alle Ausführungen, die anderen Schriften wörtlich oder sinngemäß entnommen wurden, kenntlich gemacht sind und die Arbeit in gleicher oder ähnlicher Fassung noch nicht Bestandteil einer Studien- oder Prüfungsleistung war.

Rostock, 14.11.2025

Lorenz Friedrich Dettmann

Contents

List of Abbreviations **VIII**

1 Introduction **1**

2 Theoretical Background **7**
2.1 Thermodynamic Ensemble 7
2.2 Molecular Dynamics 8
2.3 Multiscale Modeling and Force Fields 11
2.4 Coarse-Graining Approach 15
2.5 Martini Force Field 19
2.6 Free-Energy Calculations 21
2.7 COSMO-RS 22

3 Molecular Modeling and Analysis **24**
3.1 VSOMM Functionality 24
3.2 Modeling and Parametrization Approach 26
3.3 Analysis 34
3.4 Computational Details 35

4 Results and Discussion **36**
4.1 Benchmarking the Coarse-grained Fragment Database 36
4.2 Coarse-Grained Modeling of Leonardite Humic Acid 38
4.3 Reparametrization and SOM Diversity 40
4.4 Pollutant Mobility in SOM Systems 43
4.5 SOM Heteroaggregation with NP 46

5 Summary and Conclusions **48**

Appendices **67**

List of Abbreviations

B Bidentate

BAR Bennett Acceptance Ratio

COSMO-RS Conductor-like Screening Model for Real Solvents

DBI Direct Boltzmann Inversion

DFT Density Functional Theory

EMD Earth Mover's Distance

HCB Hexachlorobenzene

HS Humic Substance

IHSS International Humic Substances Society

LHA Leonardite Humic Acid

M Monodentate

MD Molecular Dynamics

MSD Mean-Squared Displacement

NMR Nuclear Magnetic Resonance

NP Nanoplastic

OSC Outer-Sphere Complex

PE Polyethylene

PEO Poly(ethylene oxide)

PMF Potential of Mean Force

POP Persistent Organic Pollutant

Py-FIMS Pyrolysis-Field Ionization Mass Spectrometry

SAA Sulfanilamide

SASA Solvent Accessible Surface Area

SOM Soil Organic Matter

TMS Total Mismatch Score

VSOMM Vienna Soil Organic Matter Modeler

XANES X-Ray Absorption Near Edge Structure

1 Introduction

Soil is a highly dynamic and heterogeneous system, containing a spectrum of chemical components such as minerals, organic matter, water, and air. It holds the largest biogeochemically active carbon pool and is connected with long-term climate regulation (Masoom et al., 2016; Jackson et al., 2017; Scharlemann et al., 2014). Understanding the geochemical processes that operate within soil is therefore crucial to using its potential in mitigating climate change (Paustian et al., 2016; Kibblewhite et al., 2007). Beyond carbon, soil also sequesters and regulates other indispensable resources, such as nitrogen and phosphorus. As the fundamental component of terrestrial ecosystems, it plays a vital role in providing nutrients, supporting plant growth, and fulfilling functions important for the survival of animals and humans (Blume et al., 2016b). However, these functions are increasingly threatened by human-made compounds that are released into the environment (Xu et al., 2022; Hanif et al., 2024), including a wide range of contaminants such as persistent organic pollutants (POPs) (Jones and de Voogt, 1999), micro- and nanoplastics (NPs) (Guo et al., 2020; Pérez-Reverón et al., 2023), heavy metals (Zwolak et al., 2019), and antibiotics (Cycoń et al., 2019). POPs, in particular, are toxic compounds with long lifetimes due to their resistance to natural degradation processes (Ritter et al., 1995), allowing them to spread across large distances via air and water and remain in the environment (Jones and de Voogt, 1999; Luarte et al., 2023). In addition, polar organic substances such as pharmaceutical residues and personal care products are highly mobile and can disrupt terrestrial ecosystems, a characteristic which is connected to their significant biological activity (Wilkinson et al., 2022; Gworek et al., 2021; Zheng and Guo, 2021). Addressing the diverse and complex threats to soil systems requires an interdisciplinary research approach to fully understand pollutant behavior and develop effective mediation strategies (Hou et al., 2020). Figure 1.1 schematically illustrates the pathways followed by anthropogenic pollutants in terrestrial environments.

To systematically investigate how pollutants interact within the complex soil environment, it is essential to disentangle its underlying physiochemical processes. A particularly reactive and biochemically active component of the soil is its organic fraction, known as soil organic matter (SOM). Although SOM typically accounts for only 1-5 % of topsoils, it plays an important role in global carbon cycling and the soil's functionality (Blume et al., 2016a). SOM itself is a heterogeneous and dynamic mixture, comprising decomposed plant and animal residues, microbial biomass, and various organic compounds. Its final composition and behavior are shaped by inputs from plants and microorganisms, the composition of the parent material, and climatic factors (Paul, 2016; Chen et al., 2019). By improving the structure of the soil, SOM enhances water infiltration and

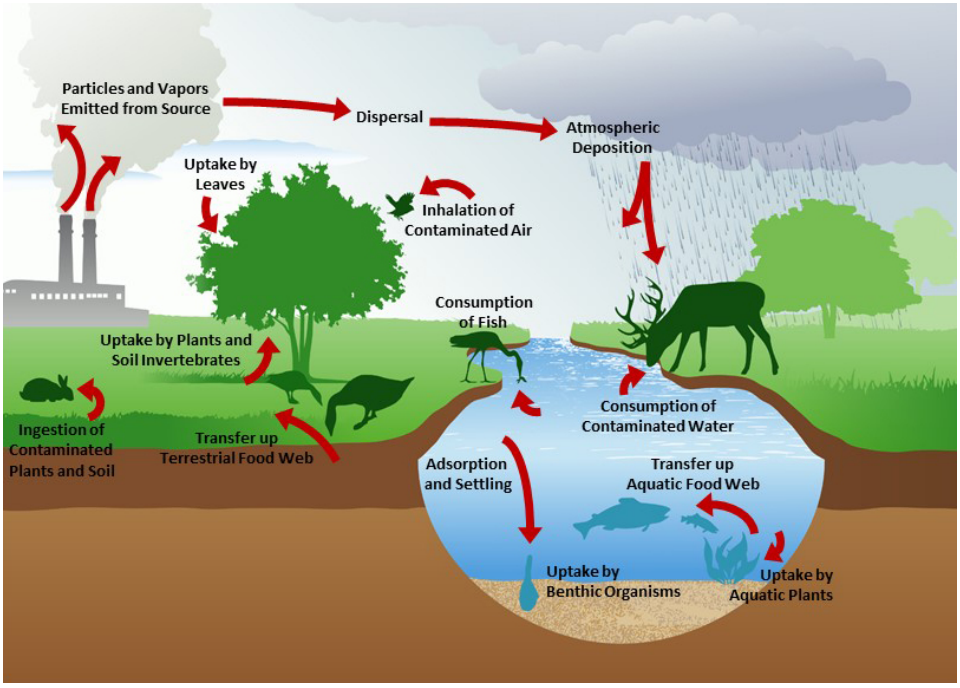


Figure 1.1: Generic pathways followed by anthropogenic pollutants into terrestrial ecosystems. The emissions from industrial facilities illustrated here are representative of other pollutant entry sources, such as historical residues of pesticide use, or the application of sewage sludge or animal manure in the agricultural sector. Adapted from U.S. Environmental Protection Agency (2025).

retention (Hudson, 1994; Rawls et al., 2003), and increases its nutrient-holding capacity, thereby supporting the soil's fertility (Roy et al., 2006). Because of its large surface area and chemically diverse functional groups, SOM plays a decisive role in the sorption of chemical contaminants, making it a key focus of research in the study of pollutant–soil interactions. In the early stages of degradation, organic inputs are referred to as raw organic matter, while the more stable and decomposed fraction, often termed humus, represents the end product of this process (Blume et al., 2016a; Galicia-Andrés et al., 2021a). A major component of this humified material are compounds known as humic substances (HSs). To systematically investigate the structure, function, and interaction of SOM, researchers have traditionally relied on alkaline extraction methods to isolate HSs. Although there is ongoing debate about the representativeness of these extracts for SOM (Gerke, 2018; Kleber and Lehmann, 2019), HSs continue to serve as valuable modeling systems (Olk et al., 2019a,b). In particular, the reference materials provided by the International Humic Substances Society (IHSS) (IHSS, 2025), which were obtained through standardized protocols from diverse global environments, offer a consistent basis for comparative and mechanistic studies of SOM chemistry. As a complementary approach, Leinweber and colleagues demonstrated that combining pyrolysis-field ionization

mass spectrometry (Py-FIMS) with X-ray absorption near edge structure (XANES) spectroscopy reveals the molecular composition and binding environment of SOM, including its organo-mineral associations (Leinweber et al., 2007; Ahmed et al., 2012; Chen et al., 2018). This approach made large parts of the previously unresolved nitrogen and carbon composition of SOM identifiable and linked them to distinct molecular structures and binding forms, thereby enhancing the general understanding of SOM's chemical complexity and heterogeneity.

Due to the rapid increase in computational power over the last few decades, molecular simulations have become an important tool in soil science, complementing advances in experimental techniques (Kubicki, 2016). Molecular dynamics (MD) simulations, in particular, offer powerful insights into the nanoscale dynamics and interactions within complex systems such as SOM. These simulations help to elucidate sorption processes central to the environmental fate of pollutants. However, accurately modeling SOM remains a major challenge due to its structural and chemical heterogeneity (Lehmann and Kleber, 2015). Various molecular SOM models have been proposed in previous research. An early influential approach was introduced by Schulten and colleagues (Schulten et al., 1991; Schulten and Schnitzer, 1995), which conceptualized SOM as a polymeric network composed of long-chain alkyl structures and aromatic rings, and was subsequently refined to include carbohydrate and protein units (Schulten and Schnitzer, 1993, 1997). This model was characterized by numerous interaction sites, enabling studies of the sorption of xenobiotic compounds to be conducted (Schulten and Leinweber, 2000; Thiele-Bruhn et al., 2004; Schwarz et al., 2012). Although the assumption that SOM is composed of macromolecular and high-weight molecular structures has since been reconsidered (Schulten, 1999b,a; Sutton et al., 2005), these models provided an important conceptual foundation for understanding SOM structure. As an alternative, researchers have also modeled individual SOM functional groups to reduce the systems' complexity and allow for more physically accurate simulations, e.g., those based on quantum mechanical methods (Aquino et al., 2007; Tunega et al., 2020). However, capturing the diverse chemistry of SOM requires a large and representative set of molecules. Addressing this issue, Ahmed and colleagues developed a set of over 30 small organic molecular fragments designed to represent major SOM functional groups, which were identified through Py-FIMS and XANES spectroscopy (Ahmed et al., 2012, 2015; Gros et al., 2017). This approach enabled simulations to be carried out across three levels of complexity: individual functional groups, compound classes, and SOM as a whole. Through the use of quantum-chemical calculations and quantitative structure–activity relationship models, key molecular descriptors governing sorption mechanisms were identified. This comprehensive strategy closely aligned with experimental data and provided mechanistic insights into SOM–pollutant interactions that remain experimentally inaccessible.

Ongoing research into the nature of HSs has led to the development of increasingly refined models and computational strategies for representing SOM. Based on analytical evidence, Piccolo proposed that HSs do not consist of large macromolecules but rather of supramolecular assemblies (Piccolo, 2001, 2002). These aggregates are held together

through non-covalent interactions of the molecules such as electrostatic and dispersive forces (Aquino et al., 2011; Schaumann et al., 2013; Gerzabek et al., 2022). In order to implement and explore this supramolecular concept in a computational framework, the Vienna Soil Organic Matter Modeler (VSOMM) was introduced (Sündermann et al., 2015). This approach represents a major advance in the molecular modeling of SOM by enabling the generation of condensed-phase HS models that reflect experimental elemental compositions and ^{13}C NMR assignments. Using a fragment database, the VSOMM facilitates the creation of structurally and chemically diverse HS molecular systems that more closely align with the natural heterogeneity of SOM. Agreement with experimental data has been proven in relation to various physicochemical properties of the Leonardite humic acid (LHA) (Petrov et al., 2017), and its models have been applied to study interactions with small organic molecules (Petrov et al., 2019), proteins (Gotsmy et al., 2021), and mineral surfaces (Galicia-Andrés et al., 2021b). To extend the fragment database, a second version of the modeler, VSOMM2, was subsequently released, optimizing the match between organic compositions and input parameters (Escalona et al., 2021). These tools allow for the exploration of HS properties using MD simulations, facilitating direct comparison with experimental data. They have expanded the potential for systematic and versatile SOM modeling at the atomistic level.

Despite the continuing growth in computational power, atomistic MD simulations, where all atoms are explicitly represented, are still constrained to relatively small spatial and temporal scales, rarely extending beyond hundreds of nanometers or several microseconds. This restricts the modeling of complex environmental systems such as SOM. For example, simulating supramolecular SOM structures and their interactions with large biomolecules requires the inclusion of a vast number of atoms, which significantly increases the computational costs. This limitation can become further apparent when studying dynamic adsorption or diffusion processes, which demand extensive sampling and long simulation times. In such cases, the efficiency and feasibility of atomistic MD can be compromised. To overcome these challenges, coarse-grained simulation techniques have been developed (Noid, 2013; Noid et al., 2024). Coarse-graining involves grouping several atoms into larger pseudo-particles known as beads, thereby reducing the number of particles and interactions within the molecular system. Consequently, this approach accelerates the simulations and enables the exploration of longer timescales and larger spatial domains. Sacrificing atomistic resolution for improved performance is an inherent trade-off, as atomistic details cannot be retained or only described implicitly. Nevertheless, this constraint is not a substantial disadvantage in the context of SOM, for which characteristic substructures and functional groups are known, whereas no exact overall molecular structure can be defined. This justifies an abstracted and averaged representation of the molecules in certain instances. The process of coarse-graining includes a mapping scheme from the atomistic representation to the coarse-grained representation and involves parametrizing the interactions between the beads. Since this process is not trivial and can vary depending on the desired accuracy and the respective area of application, various coarse-graining methods have been developed (Kmiecik et al.,

2016; Jin et al., 2022). In so-called bottom-up approaches, the interaction potentials are systematically derived from atomistic reference simulations, while top-down methods adjust these to reproduce macroscopic or experimental data.

In response to the growing interest in producing efficient simulations of large and complex molecular systems, coarse-grained force fields have been developed in recent years (Orsi and Essex, 2011; Machado et al., 2019). These frameworks provide pre-defined parameter sets that govern the interactions between coarse-grained beads. Of these, the Martini force field has become one of the most widely adopted coarse-grained force fields used for biomolecular and soft matter simulations (Marrink et al., 2007; Souza et al., 2021). It offers a broad spectrum of pre-parametrized bead types representing a wide range of chemical functionalities, which facilitates the construction of diverse molecular models. The available bead-bead interaction parameters allow for the straightforward definition of nonbonded interactions between different molecular compounds, thus reducing the need for an extensive system-specific parametrization. This is especially advantageous given the complexity and variability inherent in SOM systems. Despite this potential, no coarse-grained SOM model has yet been developed that allows different compositions to be simulated, as is possible with the VSOMM2 in atomistic resolution. Consequently, a transferable coarse-grained model of SOM that accounts for compositional variability is currently lacking, which restricts the investigation of large-scale molecular dynamics and pollutant–SOM heteroaggregation processes. This absence constitutes a key limitation for modeling soil systems and addressing environmental problems, such as those caused by anthropogenic contaminants that threaten ecological functions. Understanding how such pollutants interact with SOM, a key component governing sorption processes and chemical fate in soils, is crucial for assessing environmental risks and improving remediation strategies.

The aim of this thesis is therefore to introduce a coarse-grained model of SOM, derived from atomistic structures generated by the VSOMM2 and designed to be compatible with the Martini 3 force field (Souza et al., 2021). The specific objectives are to:

- develop a coarse-grained SOM modeling approach based on the VSOMM2 that accounts for compositional variability and enables the description of different SOM samples,
- parametrize and benchmark the model to investigate structural and dynamic properties of distinct SOM systems at coarse-grained resolution, and
- explore the mobility and molecular-level interactions of pollutants in SOM matrices, as well as heteroaggregation processes between NPs and SOM, through accelerated coarse-grained simulations.

Building on the VSOMM2's ability to generate structurally diverse SOM models, an initial parametrization of the VSOMM fragments is performed and validated against corresponding atomistic systems. An automated Python-based tool is then developed

to enable the conversion of arbitrary VSOMM2 structures into coarse-grained Martini representations, thereby streamlining the model generation process. These coarse-grained SOM models are subsequently employed to investigate the interactions with representative compounds from two major pollutant classes: POPs and polar organic contaminants, in particular hexachlorobenzene (HCB) and sulfanilamide (SAA). Facilitated by the coarse-grained representation, these simulations explore the mobility and binding behavior of pollutants across chemically diverse SOM systems, expanding on insights from prior experimental and theoretical studies (Thiele-Bruhn et al., 2004; Kubicki, 2006; Ahmed et al., 2015). In addition, a study of the interactions between SOM and NPs is conducted, focusing on both polymer chain incorporation into SOM matrices and SOM adsorption onto a plastic nanoparticle. The rest of this thesis is structured as follows:
Chapter 2 presents the theoretical background, including an overview of molecular simulations, force fields, coarse-graining principles, and parametrization strategies.
Chapter 3 outlines the modeling approach used for SOM parametrization and provides details of the study design, examining the pollutant interactions with SOM.
Chapter 4 presents and discusses the key results of this work.
Chapter 5 concludes with a general summary.
The appendix lists the peer-reviewed publications that form the basis of this cumulative thesis.

2 Theoretical Background

This chapter introduces the fundamental concepts and computational methods used throughout this thesis. The primary technique is the MD simulation, which enables the complex, heterogeneous interactions between pollutants and SOM to be examined at the molecular scale. The coarse-graining strategy, which constitutes a central aspect of this research, is introduced to improve the computational efficiency, which enables the investigation of larger length- and time-scales of the molecular systems. Finally, an overview of the Martini coarse-grained force field and the methods used to calculate free-energy quantities using molecular simulations is provided.

2.1 Thermodynamic Ensemble

Computational-chemistry methods enable the calculation of a system's total energy as a function of atomic coordinates and provide instantaneous positions and velocities for every particle; information that conventional experiments cannot directly access. However, in order to compare such microscopic data with experiments or to predict macroscopic observables, they usually have to be connected to thermodynamic properties.

Consider a system in a given thermodynamic *macrostate*, specified by any three of the following variables: N (number of particles), V (volume), T (temperature), P (pressure), E (total energy), or μ (chemical potential). Each macrostate corresponds to many possible *microstates*, distinguished by their particle positions and velocities. For example, there are several ways for atoms to generate the same temperature or pressure in a molecular system. The set of all microstates compatible with a particular macrostate is called a *thermodynamic ensemble* (Sherman, 2016).

The probability of finding the system in microstate i with total energy E_i is

$$p_i = \frac{1}{Z} \exp(-\beta E_i)\,, \tag{2.1}$$

where $\beta = 1/(k_{\mathrm{B}}T)$ and the partition function

$$Z = \sum_i \exp(-\beta E_i) \tag{2.2}$$

normalizes all probabilities. The ensemble average, or expectation value, of an observable O is as follows

$$\langle O \rangle = \sum_i p_i\, O_i = \frac{1}{Z} \sum_i O_i \exp(-\beta E_i)\,, \tag{2.3}$$

providing a bridge between micro- and macrostates.

Several thermodynamic relations can be derived from Equation (2.3), for instance, the internal energy

$$U = \langle E \rangle = \frac{1}{Z} \sum_i E_i \exp(-\beta E_i) = -\frac{\partial}{\partial \beta} \ln Z. \tag{2.4}$$

A naïve strategy would involve evaluating the energies of *all* the microstates and inserting them into Equation (2.3). In practice, this is impossible because the number of microstates is immense. The entropy,

$$S = k_\mathrm{B} \ln \Omega, \tag{2.5}$$

where Ω denotes the number of accessible microstates, illustrates the scale of this task: the standard molar entropy of water is $69.95\,\mathrm{J\,mol^{-1}\,K^{-1}}$ (Chase, 1998), giving $\ln \Omega \approx 5 \times 10^{24}$ for one mole of water, which is far beyond the reach of computational evaluation.

Fortunately, in many cases, only a small subset of microstates contributes significantly to the expectation value because some are far more probable than others. If these dominant microstates can be sampled, generating accurate macroscopic averages becomes feasible. Furthermore, many systems exhibit periodic characteristics or structural patterns, allowing the simulation of much smaller subsystems rather than all the molecules in one mole. These ideas form the basis of the MD simulations and related methods employed throughout this work.

2.2 Molecular Dynamics

The MD simulation is a powerful and increasingly widely used technique for exploring molecular behavior. In a typical setup, a finite simulation box contains N atoms, each with mass m_j and position $\vec{r}_j$. Their kinetic and potential energies arise, respectively, from the velocities $\vec{v}_j$ and the inter-atomic interactions. In the classical limit, the quantized nature of the motions of the atomic nuclei is neglected, and the equations of motion are reduced to Newton's second law:

$$\vec{F}_j = m_j \vec{a}_j = m_j \frac{d^2 \vec{r}_j}{dt^2} = -\vec{\nabla}_j V_N\left(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N\right), \tag{2.6}$$

where $\vec{a}_j$ is the acceleration of atom j and V_N is the potential energy function of the entire system.

With a given V_N, the equations of motion can be integrated by discretizing the problem with respect to the time t:

$$t_n = t_0 + n\Delta t, \tag{2.7}$$
$$(\vec{r}_j)_n = \vec{r}_j(t_n), \tag{2.8}$$
$$(\vec{v}_j)_n = \vec{v}_j(t_n), \tag{2.9}$$

where t_0 is the start time and Δt is the time interval between two steps n and $n+1$.

The GROMACS software package (Abraham et al., 2015) employs the leap-frog integrator (Hockney and Eastwood, 1988), which evaluates velocities at half-integer steps and positions at integer steps:

$$(\vec{v}_j)_{n+\frac{1}{2}} = (\vec{v}_j)_{n-\frac{1}{2}} + (\vec{a}_j)_n \Delta t, \tag{2.10}$$

$$(\vec{r}_j)_{n+1} = (\vec{r}_j)_n + (\vec{v}_j)_{n+\frac{1}{2}} \Delta t. \tag{2.11}$$

Other common methods, such as the velocity-Verlet algorithm (Verlet, 1967), follow the same principles.

By iteratively calculating positions and velocities, a trajectory is generated that samples the system's phase space and yields the instantaneous value $O(t_n)$ of an observable O. Estimating the ensemble average $\langle O \rangle$ from the system's time evolution is based on the ergodic principle (Birkhoff, 1931):

$$\langle O \rangle = \frac{1}{Z} \sum_i O_i \exp(-\beta E_i) \overset{!}{=} \lim_{N_s \to \infty} \frac{1}{N_s} \sum_{n=0}^{N_s - 1} O(n \Delta t), \tag{2.12}$$

where N_s is the total number of steps. This states that, with infinite time, the system that is bound to the phase trajectory will come arbitrarily close to every possible point in the phase space. Although t_s is finite in practice, a sufficiently long run samples the most probable regions of the phase space, allowing the time average to approach the true ensemble average. Therefore, MD simulations provide a way to estimate thermodynamic properties without the need to calculate the partition function explicitly (Sherman, 2016).

To reproduce experimental conditions, an MD simulation must actively control macroscopic variables such as temperature and pressure. The controlled quantities depend on the chosen thermodynamic ensemble. At the beginning of a simulation, when no initial velocities are given, they are randomly set following the Maxwell–Boltzmann probability distribution (Abraham et al., 2025):

$$p(\vec{v}_j) = \sqrt{\frac{m_j}{2\pi k_\mathrm{B} T}} \exp\left(-\frac{m_j \vec{v}_j^2}{2 k_\mathrm{B} T} \right), \tag{2.13}$$

which yields the correct kinetic-energy distribution at temperature T.

In this work, the canonical (NVT) and isothermal–isobaric (NPT) ensembles are utilized, where temperature and, for NPT, pressure are kept close to the prescribed values. The temperature control is achieved with a thermostat, while the pressure control relies on a barostat.

The canonical velocity rescaling thermostat (Bussi et al., 2007) rescales every velocity by the same factor α, $\vec{v}_j \leftarrow \alpha \vec{v}_j$, such that the instantaneous kinetic energy K approaches the desired average $\hat{K} = N_f/(2\beta)$ at the target temperature $\hat{T}$:

$$\alpha = \sqrt{\frac{\hat{K}}{K}} = \sqrt{\frac{\hat{T}}{T}}, \tag{2.14}$$

with N_f being the number of degrees of freedom. Applying a single, abrupt rescaling would disturb the dynamics of the molecules considerably, which can be prevented by distributing the adjustment over several steps. Furthermore, to enforce the canonical distribution, a target kinetic energy K_t is stochastically selected at each step, which leads to the following propagation rule:

$$dK = (K_t - K)\frac{dt}{\tau_T} + 2\sqrt{\frac{KK_t}{N_f}}\,\frac{dW}{\sqrt{\tau_T}}, \tag{2.15}$$

where τ_T is the temperature coupling time and dW a Wiener noise term. Far from equilibrium, the deterministic first term dominates, modifying the system dynamics toward $\hat{T}$. Near equilibrium, the noise term maintains the correct canonical sampling. If the noise term is suppressed ($dW = 0$), Equation (2.15) reduces to the Berendsen thermostat (Berendsen et al., 1984), which controls temperature efficiently but does not generate an exact canonical ensemble.

Similarly, the pressure can be regulated by coupling the simulation box to an external bath. A barostat rescales the box vectors, which are collected in the matrix $\bar{b}$, and the particle coordinates such that the instantaneous pressure tensor $\bar{P}$ relaxes toward a target value $\bar{P}_t$.

One of the simplest schemes is the Berendsen barostat (Berendsen et al., 1984), which imposes first-order relaxation,

$$\frac{d\bar{P}}{dt} = \frac{\bar{P}_t - \bar{P}}{\tau_p}, \tag{2.16}$$

with the pressure coupling time τ_p. For isotropic coupling, the pressure reduces to the scalar

$$P = \frac{1}{3}\,\mathrm{Tr}(\bar{P}), \tag{2.17}$$

and a uniform scaling matrix adjusts both $\bar{b}$ and all the particle positions accordingly. While this algorithm maintains the correct average pressure, it does not produce the true NPT ensemble, which can be important for properties sensitive to pressure fluctuations. Consequently, it is best used for equilibration phases, where rapid relaxation is desired before a production run begins.

For a correct sampling of the NPT ensemble as well as for systems where the box shape should respond to anisotropic deformations, the Parrinello–Rahman (Parrinello and Rahman, 1981) method is preferred. Introducing scaled coordinates $\vec{s}_i$ via $\vec{r}_i = \bar{b}\,\vec{s}_i$, the system's Lagrangian L becomes

$$L = \frac{1}{2}\sum_{i=1}^{N} m_i \frac{d\vec{s}_i^{\,T}}{dt}\bar{G}\,\frac{d\vec{s}_i}{dt} - V_N(\{\vec{s}_i\}) + \frac{1}{2}W\,\mathrm{Tr}\!\left(\frac{d\bar{b}^T}{dt}\,\frac{d\bar{b}}{dt}\right) - P_t V, \tag{2.18}$$

where V is the box volume, $\bar{G} = \bar{b}^T\bar{b}$ is the metric tensor, and

$$W = \frac{3\tau_p^2 l}{4\pi\beta_T} \tag{2.19}$$

is a mass parameter that governs the strength of the coupling. Here, l is the largest box-matrix element and β_T is the isothermal compressibility.

Applying the Euler–Lagrange equations yields the following equations of motion:

$$\frac{d^2\vec{s}_i}{dt^2} = \bar{b}^{-1}\frac{\vec{F}_i}{m_i} - \bar{G}^{-1}\frac{d\bar{G}}{dt}\frac{d\vec{s}_i}{dt}, \tag{2.20}$$

$$\frac{d^2\bar{b}}{dt^2} = \frac{V}{W}\left(\bar{P} - \mathbb{I}P_t\right)\left(\bar{b}^{-1}\right)^T. \tag{2.21}$$

The first equation describes the particle motion in the deforming box. The second equation governs the box's acceleration in response to pressure differences. Because these are second-order equations, large oscillations can arise if the system is far from equilibrium, so the Parrinello–Rahman scheme is most suitable for production runs using an already equilibrated system. A more recent alternative approach introduced a stochastic term, analogous to the canonical velocity-rescale thermostat, to combine first-order relaxation with correct ensemble sampling (Bernetti and Bussi, 2020).

When a particle approaches the edge of the simulation cell, boundary conditions must be defined. The most common choice is to impose *periodic boundary conditions* (Allen and Tildesley, 2017): a particle that leaves the box on one side re-enters from the opposite side. Formally, this is equivalent to surrounding the central box with an infinite lattice of translated replicas that contain identical particle configurations, so the simulated system is effectively without boundaries.

To prevent a particle from interacting with its own periodic images, the *minimum-image convention* is applied. For pairwise nonbonded forces, a cutoff radius R_c is chosen such that it does not exceed half the length of the shortest box vector,

$$R_c < \frac{1}{2}\min\left(\vec{b}_1, \vec{b}_2, \vec{b}_3\right), \tag{2.22}$$

ensuring that, for any reference particle, only the nearest image of every other particle is considered. For long-range interactions, most notably Coulomb forces, this truncation can introduce substantial errors. Such interactions therefore require specialized treatments, which are discussed in Section 2.3.

2.3 Multiscale Modeling and Force Fields

The potential energy function in Equation (2.6), which is used to describe all occurring interactions, will ultimately determine the dynamics of a molecular system. The representation of these interactions is dependent on the level of theory with which the models will be described. The molecular systems can be simulated at different resolutions, which further determines the accuracy of the models and, at the same time, the attainable time and length scales. Figure 2.1 illustrates different approaches which are commonly

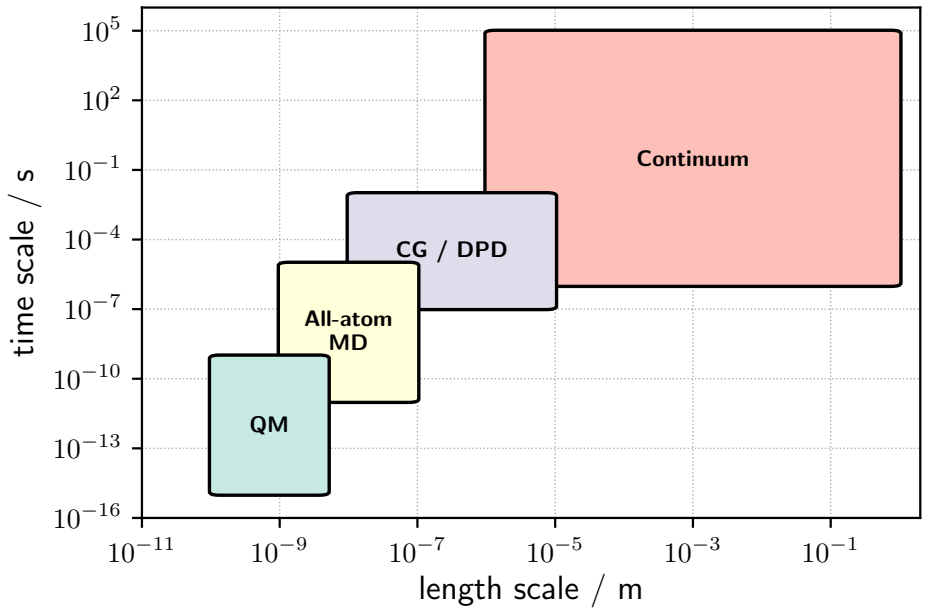


Figure 2.1: Estimates of the typical length and time scales of various methods for simulating molecular and material systems based on Ozboyaci et al. (2016). These include quantum mechanical (QM), classical all-atom, coarse-grained (CG), and dissipative particle dynamics (DPD) approaches, as well as continuum methods.

used in simulations to represent the molecules or bulk materials. Quantum mechanical methods are situated at the bottom left. In these approaches, the dynamics of electrons are explicitly treated together with those of the nuclei. A commonly used approach is density-functional-theory MD (DFT-MD) (Hohenberg and Kohn, 1964; Car and Parrinello, 1985), in which the electronic energy is determined as a functional of the electron density, while the nuclei are treated classically using Newton's equations of motion (Equation 2.6). Based on the Born–Oppenheimer approximation (Born and Oppenheimer, 1927), which decouples the electronic and nuclear dynamics, this method allows the description of a wide range of chemical reactions. DFT-MD has been successfully applied to investigate molecular-level interactions relevant in the context of soil science (Tunega et al., 2004; Kerisit et al., 2016; Ahmed et al., 2018). In general, DFT-MD provides a highly accurate description of the molecular dynamics and interactions. However, this accuracy comes at a high computational cost, as the Schrödinger equation must be solved at each MD step due to the explicit treatment of electrons.

A cheaper and often more intuitive description of the molecular interactions can be achieved by treating the electronic degrees of freedom implicitly and representing each atom as a point charge. In this approach, the interactions are described by empirical parameters and by solving the classical equations of motion. As shown in Figure 2.1,

depending on the available computational resources, all-atom MD simulations can reach length scales of up to hundreds of nanometers and, in the best cases, simulation times in the microsecond regime. In such cases, the potential energy function depends solely on the positions of the atoms. Therefore, fully classical simulations provide adequate accuracy primarily when quantum effects play only a minor role in the processes of interest (van Gunsteren and Oostenbrink, 2024).

The implicit treatment of the electronic degrees of freedom requires the definition of functional forms for the potential energy function, together with empirical parameters. A collection of such parameters, designed to describe both bonded and nonbonded interactions of molecules, is referred to as a force field. These parameters are determined in order to reproduce specific target properties, which vary depending on the force field. For example, the GROMOS force field employs a top-down approach, where parameters are mainly fitted to match experimental and thermodynamic properties such as liquid–liquid partitioning data (Schmid et al., 2011; Reif et al., 2012). In contrast, the CHARMM force field is based on a bottom-up approach that involves building parameters starting from quantum mechanical data (Brooks et al., 2009). So-called hybrid schemes also exist, in which part of the system is described at the quantum-mechanical level while the remainder is modeled using a classical force field. These so-called QM/MM approaches have been successfully applied in previous work on the investigation of the binding of phosphorus compounds to mineral surfaces (Ganta, 2023).

In classical MD simulations, the potential energy function is generally composed of bonded and nonbonded terms. While the exact functional forms can vary depending on the applied force field, a common separation ansatz is employed. The bonded interactions typically include bond stretching, angle bending, and dihedral torsions and can be expressed in the form:

$$V_{\mathrm{B}} = \frac{1}{2}k^{r}_{ij}\left(r_{ij} - r^{0}_{ij}\right)^2 + \frac{1}{2}k^{\theta}_{ijk}\left(\theta_{ijk} - \theta^{0}_{ijk}\right)^2 + k^{\phi}_{ijkl}\left[1 + \cos\left(n\phi_{ijkl} - \phi^{0}_{ijkl}\right)\right], \tag{2.23}$$

where r_{ij} is the distance between particles i and j, θ_{ijk} is the angle formed by particles i, j, and k, and ϕ_{ijkl} denotes the dihedral angle, defined by the particle planes (i, j, k) and (j, k, l). The parameters r^{0}_{ij}, θ^{0}_{ijk}, and ϕ^{0}_{ijkl} are the corresponding equilibrium values, respectively. The *proper* dihedral angle, as shown in Equation 2.23, has a periodic form, which is particularly useful for describing conformational transitions such as between *cis* and *trans* configurations. The *improper* dihedral angle, by contrast, is often modeled with a harmonic potential to maintain planarity in certain groups of particles.

A modified approach to describing angular motion is introduced with the restricted bending potential (Bulacu et al., 2013):

$$V_{\mathrm{ReB}} = \frac{1}{2}k^{\theta}_{ijk}\frac{\left(\cos\theta_{ijk} - \cos\theta^{0}_{ijk}\right)^2}{\sin^2\theta_{ijk}}, \tag{2.24}$$

which prevents the angle θ_{ijk} from approaching 180°. This restriction can be important because if the particles, that are involved in constructing a dihedral angle become

collinear, it becomes undefined. By ensuring well-defined dihedral angles, restricted bending improves the numerical stability of the simulations. This functional form is particularly common in coarse-grained molecular models, which are introduced in Section 2.4.

Although classical mechanics generally provides a good approximation for atomic dynamics, light bonded atoms such as hydrogen can exhibit motions that deviate from purely classical behavior described by a harmonic oscillator (Allen and Tildesley, 2017). To obtain a realistic yet computationally efficient description, bonds are often treated as constraints in MD (Hess et al., 1997). This description reasonably approximates the quantum-mechanical bond behavior in the ground state (Abraham et al., 2025) while removing high-frequency vibrations, thereby allowing for larger integration time steps and faster simulations. Consequently, constraints are widely applied not only to bonds involving hydrogen but also to those between heavier atoms and in coarse-grained models to enhance numerical efficiency.

In addition to bonded terms, the potential energy function in classical MD simulations contains nonbonded interactions, which are usually described as a sum of Lennard–Jones and Coulomb terms:

$$V_{\mathrm{NB}} = V_{\mathrm{LJ}} + V_{\mathrm{C}} = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + f\frac{q_i\, q_j}{\epsilon_r\, r_{ij}}, \tag{2.25}$$

where $f = \frac{1}{4\pi\epsilon_0}$, ϵ_0 is the vacuum permittivity, and ϵ_r is the relative permittivity. Here, ϵ_{ij} effectively represents the depth of the Lennard–Jones potential, and σ_{ij} is the distance at which the potential becomes zero and can be interpreted as the minimum possible contact distance between two nonbonded particles. These values are central parameters in force fields, and specific combination rules are used to determine ϵ_{ij} and σ_{ij} for interactions between different particle types. The Lennard–Jones term proportional to r^{-6} describes attractive dispersion (dipole–dipole) interactions, which can be viewed as an effective potential resulting from the polarization of electron clouds. The repulsive r^{-12} term models the Pauli repulsion. Although this exponent is chosen for computational efficiency, more realistic short-range repulsion can be represented by the Buckingham potential (Buckingham, 1938), which employs an exponential form at increased computational cost. These functional forms illustrate how classical force fields implicitly account for the effects of electrons and their interactions.

As previously stated, Coulomb interactions act over long ranges, which makes it essential to properly treat cutoff effects, otherwise the abrupt truncation at the cutoff distance R_c can lead to discontinuities in the potential function. One common approach that can be used to mitigate this problem involves modifying the Coulomb interaction under the assumption of a constant dielectric environment beyond R_c, characterized by a dielectric constant ϵ_{RF}. In this reaction-field approach (Tironi et al., 1995), the potential takes the form

$$V_{\mathrm{C,RF}} = f\frac{q_i\, q_j}{\epsilon_r}\left[\frac{1}{r_{ij}} + \frac{\epsilon_{\mathrm{RF}} - \epsilon_r}{2\epsilon_{\mathrm{RF}} + \epsilon_r}\frac{r_{ij}^2}{R_c^3} - \frac{3\epsilon_{\mathrm{RF}}}{2\epsilon_{\mathrm{RF}} + \epsilon_r}\frac{1}{R_c}\right] \tag{2.26}$$

where the first term represents the standard Coulomb interaction, the second term accounts for the interaction with the medium outside the cutoff sphere, and the third term ensures that the potential smoothly reaches zero at R_c. Inside the cutoff, Coulomb interactions are treated explicitly, whereas outside the cutoff, the surrounding medium is approximated as having a uniform dielectric constant. Physically, this treatment implicitly models the process in which particles within the cutoff sphere induce a polarization in the medium, which in turn generates a reaction field that has an effect on the particles. A similar consideration applies to Lennard–Jones interactions, where the potential can be shifted such that the potential energy is zero at the cutoff distance (de Jong et al., 2016). This ensures a transition of the potential energy to zero and avoids artifacts from abrupt truncation.

2.4 Coarse-Graining Approach

While classical atomistic models can provide highly detailed insights into molecular systems, their computational cost increases rapidly with system size and simulation time. Coarse-graining constitutes a commonly used strategy for overcoming these limitations (see the middle region of Figure 2.1). It aims to reduce the complexity of a system by lowering its dimensionality, i.e., by decreasing the number of particles, while retaining the system's most important structural and dynamical characteristics (Noid, 2013). Figure 2.2 illustrates this translation, where an atomistic representation of a molecule is systematically converted into a coarse-grained one. The decisive advantage is immediately apparent: in the coarse-grained representation, the same system is described by fewer particles, which significantly reduces the number of interactions to be evaluated and thereby accelerates the calculations.

When designing coarse-grained models using atomistic data, an explicit connection between the atomistic (reference) system and its coarse-grained representation is required. This connection can be formalized via a mapping matrix $\bar{M}$, which transforms the coordinates of a system composed of n atoms into those of N coarse-grained beads:

$$\vec{R}^N = \bar{M}\,\vec{r}^n, \tag{2.27}$$

where $\vec{r}^n$ and $\vec{R}^N$ are the atomistic and coarse-grained coordinates, respectively. The mapping matrix defines which atoms are grouped together and how each atomistic coordinate contributes, via specified weights, to the position of the corresponding coarse-grained bead. Common choices for these weights are center-of-mass and center-of-geometry mappings.

The choice of mapping, i.e., how atoms are grouped into beads, is largely determined by the desired level of coarse-graining. While in principle this is arbitrary, mappings typically follow chemical intuition or established schemes (see, for example, the guidelines for the Martini model in Section 3.2). More recently, algorithmic approaches have been

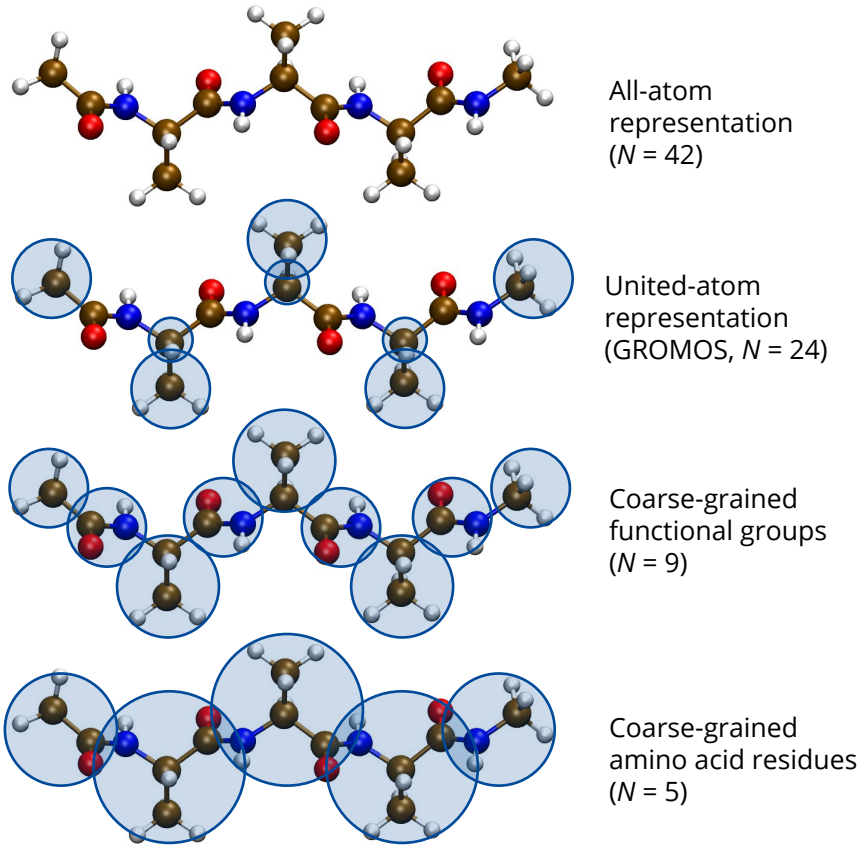


Figure 2.2: Hierarchical representation of a peptide illustrating the principle of coarse-graining. The molecule is depicted at different levels of resolution, from an all-atom representation to a united-atom representation (as used in certain force fields such as GROMOS) and coarse-grained models based on grouped functional units or entire amino acid residues. The choice of the representation will be reflected in the definition of the mapping matrix $\bar{M}$. With increasing coarse-graining level, the number of particles N decreases accordingly. Adapted from Shell (2016).

developed to automate and optimize the mapping process (Webb et al., 2019; Szczuka et al., 2025).

In principle, a coarse-grained model should reproduce the physical behavior of the atomistic system as closely as possible. To this end, the interactions between the coarse-grained beads are described by an effective potential energy function V_{CG} for the coarse-grained molecules. Determining the parameters of this potential can be addressed by considering the configurational probability, which assigns probabilities to

states according to their potential energy:

$$\rho_{\mathrm{AT}}(\vec{r}^{\,n}) = \frac{\exp[-\beta\, V_{\mathrm{AT}}(\vec{r}^{\,n})]}{Z_{\mathrm{AT}}}, \tag{2.28}$$

where V_{AT} is the atomistic potential energy function and Z_{AT} is the atomistic configurational partition function. Here, $\rho_{\mathrm{AT}}(\vec{r}^{\,n})$ denotes the configurational probability density obtained by considering only the position-dependent part of the canonical probability distribution. The momentum contributions can be treated separately, as outlined in Noid et al. (2008).

When mapping the atomistic configurations to the coarse-grained ones, the atomistic configurational probability is expressed in terms of the coarse-grained coordinates $\vec{R}^{\,N}$ using the Dirac delta distribution:

$$\rho_{\mathrm{AT}}(\vec{R}^{\,N}) = \int \rho_{\mathrm{AT}}(\vec{r}^{\,n})\, \delta\Big(\bar{M}\vec{r}^{\,n} - \vec{R}^{\,N}\Big)\, d\vec{r}^{\,n}. \tag{2.29}$$

Assuming that the coarse-grained model is also sampled in the canonical ensemble, an exact coarse-grained model potential would reproduce the mapped distribution:

$$\rho_{\mathrm{CG}}(\vec{R}^{\,N}) = \rho_{\mathrm{AT}}(\vec{R}^{\,N}). \tag{2.30}$$

Using Equations 2.29 and 2.30, and defining the coarse-grained configurational probability analogously, the following equation is obtained:

$$\frac{\exp\Big[-\beta\, V_{\mathrm{CG}}(\vec{R}^{\,N})\Big]}{Z_{\mathrm{CG}}} = \int \frac{\exp[-\beta\, V_{\mathrm{AT}}(\vec{r}^{\,n})]}{Z_{\mathrm{AT}}}\, \delta\Big(\bar{M}\vec{r}^{\,n} - \vec{R}^{\,N}\Big)\, d\vec{r}^{\,n}. \tag{2.31}$$

Taking the logarithm of Equation 2.31 and incorporating the (state-independent) partition function into an additive constant yields an expression for the optimal coarse-grained potential in terms of the underlying atomistic system:

$$V_{\mathrm{CG}}(\vec{R}^{\,N}) = \underbrace{-k_{\mathrm{B}}T\, \ln \int \exp[-\beta\, V_{\mathrm{AT}}(\vec{r}^{\,n})]\, \delta\Big(\bar{M}\vec{r}^{\,n} - \vec{R}^{\,N}\Big)\, d\vec{r}^{\,n}}_{A_{\mathrm{AT}}(\vec{R}^{\,N})} + \text{const.} \tag{2.32}$$

Here, the first term $A_{\mathrm{AT}}(\vec{R}^{\,N})$ is the atomistic free-energy surface along the coarse-grained coordinates, i.e., the (generally many-body) potential of mean force (PMF) (Shell, 2016). Hence, if V_{CG} reproduces the underlying free-energy surface of the atomistic system, the thermodynamic averages computed in the coarse-grained model will match those of the atomistic reference. Other mapping-dependent constants only contribute to the additive constant and therefore do not affect forces or canonical averages.

However, the PMF is usually a many-body quantity because integrating out the eliminated degrees of freedom induces complex couplings between the remaining coarse-grained coordinates. In molecular simulations, the functional form of such many-body

potentials is not known a priori, so they are typically approximated by choosing a restricted functional form and calibrating its parameters to reproduce selected target quantities of the reference model. Consequently, commonly used pairwise additive potentials are inherently unable to fully reproduce the exact PMF. On the other hand, pair potentials are computationally efficient, and various *bottom-up* parametrization strategies approximate the PMF by targeting quantities obtained from atomistic simulations, for example by matching radial distribution functions, reproducing mean forces, or minimizing the relative entropy (Noid, 2013). Due to the central role played by free-energy quantities in coarse-grained models, *top-down* approaches instead use experimental thermodynamic data, particularly free-energy differences such as solvation or transfer free energies, as target quantities that the coarse-grained model should reproduce (see also Section 2.5).

An example of a bottom-up parametrization strategy for bonded interactions that assumes the validity of the separation ansatz (i.e., that the potential energy can be decomposed into bonded and nonbonded contributions) is the *Boltzmann inversion*. This approach follows directly from the relationship between the probability distribution and the potential energy in a canonical ensemble (please compare to Equation 2.28). For an independent degree of freedom q, which can represent a bond length r, an angle θ, or a dihedral angle ϕ, the bonded potential $V_q(q)$ can be obtained by inverting the corresponding probability distribution $p_q(q)$ as

$$V_q(q) = -k_\mathrm{B}T \ln[p_q(q)]\,. \tag{2.33}$$

In practice, the normalization factor (partition function) inside the logarithm is omitted, as it only contributes an additive constant to the potential and is therefore irrelevant for the forces.

By applying the mapping to the coordinates from atomistic reference simulations, histograms of the bonds, angles, and dihedral angles can be generated. These histograms, H_q, are then converted into probability distributions by undertaking a rescaling (Junghans et al., 2021):

$$p_r(r) = \frac{H_r(r)}{4\pi r^2}, \quad p_\theta(\theta) = \frac{H_\theta(\theta)}{\sin(\theta)}, \quad p_\phi(\phi) = H_\phi(\phi). \tag{2.34}$$

The resulting probability distributions can then be transformed into potentials either via tabulated forms or by fitting them to analytical potential functions. This approach, also referred to as *Direct Boltzmann Inversion* (DBI) (Tschöp et al., 1998), assumes that the bonded degrees of freedom are statistically independent and therefore neglects cross-correlations between them. To account for such couplings, the potentials can be refined in an iterative procedure, as implemented, for example, in the Swarm-CG method (Empereur-Mot et al., 2020) (see Section 3.2).

2.5 Martini Force Field

As outlined above, coarse-grained models require reference information to assign parameters to the beads, which represent groups of atoms. Under comparable simulation conditions and within a consistent mapping scheme, identical functional groups across different molecules can be expected to yield similar interaction parameters after parametrization of the cross-interactions. Rather than repeating this parametrization process for every new system, it is therefore advantageous to establish a transferable parameter set that sufficiently approximates common molecular building blocks. This concept underlies coarse-grained force fields such as Martini, which provide a set of predefined bead types. In particular, the Martini force field has proven to be highly effective in biomolecular simulations, where certain types of functional groups frequently occur.

The Martini coarse-grained force field was first introduced by Marrink et al. (2004) for the simulation of lipid membranes. Subsequently, its applicability was extended to a wide variety of biomolecular and material systems, including peptides, proteins, carbohydrates, nucleic acids, carbon-based materials, ionic liquids, polymers, and nanoparticles (Marrink et al., 2007, 2023). The Martini force field version 3 included a comprehensive reparametrization of the model (Souza et al., 2021). Martini 3 introduced a rebalanced interaction matrix alongside an expanded set of bead types, also incorporating implicit representations of specific functional group properties, such as hydrogen bonding abilities or altered electronic density.

Figure 2.3 provides an overview of the available bead types. The mapping is generally performed according to a four-to-one scheme, in which on average four heavy atoms together with their associated hydrogens are represented by a single (regular) coarse-grained bead. This representation can be refined by using small and tiny beads (denoted S and T, respectively), which represent groups of three and two heavy atoms, to more accurately capture ring-like structures or compact functional groups. The force field contains the four principal bead classes, C (nonpolar), N (intermediate polarity), P (polar), and Q (charged), with gradual polarity differences defined within each class. In addition, the bead class X was introduced in Martini 3 to represent halogen compounds. To improve the miscibility and packing behavior across systems represented at different resolutions, the relative sizes of the beads were rebalanced. Within the charged bead class (Q), five subtypes are available to reproduce ion-specific effects consistent with the Hofmeister series.

The nonbonded interactions are described by the Lennard-Jones potentials, which differentiate the polarity levels, while charged beads additionally interact through Coulombic forces. Martini follows a top-down parametrization strategy, where the Lennard-Jones parameters are calibrated against experimental thermodynamic data such as free energies of transfer, heats of vaporization, densities, and miscibility trends (Souza et al., 2021). Bonded interactions, in contrast, are typically derived from atomistic simulations, where probability distributions of mapped molecular configurations are reproduced, which can be achieved by the Boltzmann inversion (Equation 2.33).

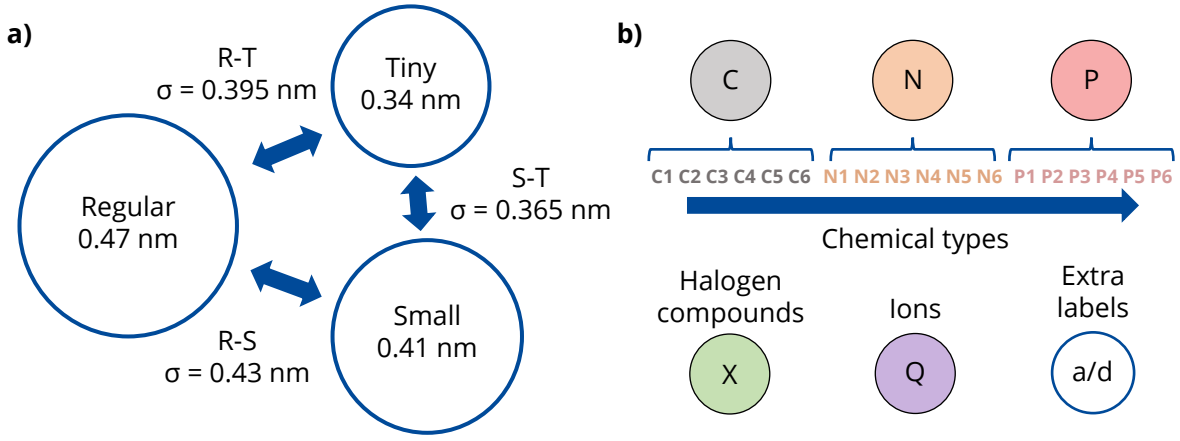


Figure 2.3: Panel a) shows the three Martini bead sizes and the Lennard–Jones σ values for each size combination. Panel b) shows the three main bead classes, each with six polarity levels, along with special bead classes and extra labels. Note that X and Q classes include additional subdivisions, and further labels exist beyond those shown for hydrogen bonding (Souza et al., 2021).

Despite its broad range of applications, it is important to recognize the intrinsic limitations of Martini models (Alessandri et al., 2019). In Hamiltonian-based coarse-grained models such as Martini, the dynamics of the molecules are artificially accelerated because coarse-graining smooths the underlying free-energy landscape. Consequently, the whole spectrum of frictional forces that are present in atomistic systems is not reproduced, leading to systematically faster diffusion. While this feature can be advantageous for sampling long-timescale conformational changes, it complicates a direct interpretation of the simulated time scale. Another limitation arises from the enthalpy–entropy imbalance inherent in the potentials (Jarin et al., 2021). Because the interaction parameters are independent of the thermodynamic state applied, Martini models do not reproduce the correct temperature dependence of entropic contributions. The model accuracy at temperatures that differ from those used during parametrization can therefore deviate from the atomistic behavior. It is thus advisable to apply Martini models only under thermodynamic conditions close to those used in their parametrization and validation.

The SOM systems considered in this work comprise a diverse set of chemical functional groups. Its systematic building-block parametrization allows the Martini force field to represent these organic groups through predefined bead types with well-defined cross-interaction parameters. This modular strategy enables the construction of new molecules by assembling existing beads. Since nonbonded interactions are specified through a global interaction matrix, cross-interactions between soil components and pollutant molecules are automatically defined once each component is mapped to its coarse-grained representation. This facilitates the efficient and transferable modeling of chemically complex systems such as pollutants interacting with SOM.

2.6 Free-Energy Calculations

MD simulations have traditionally been employed to explore the equilibrium and dynamical properties of a system in a single state. However, many processes of chemical and biological relevance involve a transition between two states of a system, such as conformational changes of a molecule or the binding of a ligand to a binding site. In such cases, free-energy calculations provide a rigorous framework with which to quantify how spontaneously a process takes place and how much work is released or required during the transformation. Importantly, free-energy differences can also be directly compared to experimental data, which makes them a valuable bridge between simulation and experiment (Mey et al., 2020; York, 2023).

Since free energy is a state function, its difference between two states depends solely on the initial and final states, not on the path that connects them. In molecular simulations, this property enables the use of arbitrary, even unphysical, pathways to calculate free-energy differences. Methods that apply this principle are referred to as *alchemical free-energy methods* (Mey et al., 2020). They employ artificial coupling pathways that gradually turn on or off interactions of a molecule with its environment, which can be computationally far more efficient than simulating the actual physical transformation. A typical example is the determination of the solvation free energy, where the interactions of a molecule with the solvent are gradually switched off to obtain the free-energy difference between solvated and vacuum states. Such approaches are also essential for calculating transfer free energies and partitioning, quantities that are highly relevant in environmental chemistry and drug discovery (York, 2023). Moreover, free-energy methods can be applied to study rare events, where potential energy barriers or slow dynamics prevent adequate sampling using plain MD simulations.

A major challenge in free-energy calculations is that free energies are not observables that can be computed by simple averaging. Instead, they must be obtained by constructing a thermodynamic pathway that connects the states of interest. In practice, this is achieved by introducing a coupling parameter λ that interpolates between the Hamiltonians of the two states. The system is then simulated in a series of n intermediate states, $0 \leq \lambda \leq 1$, such that sufficient phase-space overlap exists between neighboring states (Shirts and Pande, 2005). The free-energy difference can then be computed as

$$\Delta G = G(\lambda_n = 1) - G(\lambda_1 = 0) = \sum_{i=1}^{n-1} \Delta G(\lambda_i \to \lambda_{i+1}). \tag{2.35}$$

One of the most widely used estimators for computing $\Delta G(\lambda_i \to \lambda_j)$ is the Bennett Acceptance Ratio (BAR) method (Bennett, 1976). BAR is a statistically optimized perturbation method that makes use of information from both neighboring ensembles and utilizes their overlap to minimize variance and bias (Bruckner and Boresch, 2011; de Ruiter et al., 2013). The free-energy difference between two states i and j is given by

$$\Delta G(\lambda_i \to \lambda_j) = k_\mathrm{B} T \ln\left(\frac{\langle f(U(\lambda_i) - U(\lambda_j) + C)\rangle_{\lambda_j}}{\langle f(U(\lambda_j) - U(\lambda_i) + C)\rangle_{\lambda_i}}\right) + C, \tag{2.36}$$

where $U(\lambda)$ is the potential energy in state λ, $f(x) = 1/(1 + e^x)$ is the Fermi–Dirac function, which satisfies the detailed balance condition and is the statistically optimal variance-minimizing weighting. C is a constant determined iteratively such that

$$\langle f(U(\lambda_i) - U(\lambda_j) + C)\rangle_{\lambda_j} = \langle f(U(\lambda_j) - U(\lambda_i) + C)\rangle_{\lambda_i}. \tag{2.37}$$

In practice, the ensemble averages are approximated by trajectory averages. When C is determined in this way, the following equation is obtained:

$$\Delta G(\lambda_i \to \lambda_j) = C - k_\mathrm{B}T \ln \frac{N_j}{N_i}, \tag{2.38}$$

with N_i and N_j denoting the number of frames sampled in states i and j, respectively.

The term "acceptance ratio" originates from the method's derivation in the context of the Metropolis Monte Carlo algorithm, where the free-energy difference can be interpreted as a ratio of acceptance functions with an optimal weighting. In practice, BAR is applied by performing alchemical simulations along a set of λ-states, thus determining the energy differences between states in a post-processing step, and calculating the individual free-energy increments that add up to the total ΔG.

2.7 COSMO-RS

As discussed in the previous section, the Martini coarse-grained model relies on partitioning data as a reference for parametrizing nonbonded interactions. However, for many molecules, especially when considering partitioning beyond the commonly considered octanol–water case, such as hexadecane-water or chloroform-water, experimental reference data such as transfer free energies are often unavailable. Although atomistic simulations utilizing the BAR method can, in principle, estimate such properties, these simulations become prohibitively costly when the molecule exhibits numerous degrees of freedom that must be sampled extensively.

A computationally efficient alternative is provided by the Conductor-like Screening Model for Real Solvents (COSMO-RS), originally introduced by Klamt and colleagues (Klamt, 1995; Klamt et al., 2010). COSMO-RS predicts thermodynamic properties based on stationary quantum-chemical calculations, circumventing the need for experimental data or exhaustive MD simulations. At its core, it relies on the approximation that interactions between molecules in the liquid phase can be represented as contacts between molecular surface segments, each characterized by its screening charge density σ. This surface-segment approach models the solvent as a densely packed ensemble of interacting fragments. Quantum-chemical calculations, typically using DFT, are able to produce the ground-state electron density and, hence, the screening charge density of a molecule embedded in a perfect conductor. From this, the molecule's σ-profile, $p^X(\sigma)$, is constructed, which encodes the distribution of its surface area as a function of σ. This descriptor captures the molecule's polarity landscape, which is crucial for representing

its intermolecular interactions. For two interacting surface segments with densities σ and σ', the interaction energy, $E_{\text{int}}(\sigma, \sigma')$, is described by:

$$E_{\text{int}}(\sigma, \sigma') = a_{\text{eff}} \frac{\alpha_{\text{misfit}}}{2} (\sigma + \sigma')^2 + a_{\text{eff}}\, c_{hb}(T)\, \min\{0,\ \sigma + \sigma_{hb}\}\, \max\{0,\ \sigma' - \sigma_{hb}\} \quad (2.39)$$

The first term accounts for the electrostatic "misfit" energy due to non-complementary charge densities, while the second term represents hydrogen-bonding interactions, which are activated only when donor and acceptor characteristics exceed the threshold σ_{hb}. The coefficients a_{eff}, α_{misfit}, $c_{hb}(T)$, and σ_{hb} are global parameters and are typically fitted or parametrized against experimental data. Building on this, the model defines the so-called σ-potential of a solvent S, denoted as $\mu_S(\sigma)$, which quantifies the chemical potential of inserting a segment with charge density σ into the solvent. It is given by:

$$\mu_S(\sigma) = -k_{\text{B}}T \ln \left[\int p_S(\sigma') \exp\left(-\frac{1}{k_{\text{B}}T} \left(E_{\text{int}}(\sigma, \sigma') - \mu_S(\sigma') \right) \right) d\sigma' \right] \quad (2.40)$$

Here, $\mu_S(\sigma)$ must be determined self-consistently via iteration. The convolution with the solvent σ-profile, $p_S(\sigma')$, represents the statistical integration over all the surface contacts. Once $\mu_S(\sigma)$ is known, the chemical potential of a solute X in a solvent S is obtained by integrating over the solute's σ-profile:

$$\mu_S^X = \int p^X(\sigma)\, \mu_S(\sigma)\, d\sigma + \mu_{\text{comb}}^X \quad (2.41)$$

The first term captures the direct segment–segment interactions, while the second term accounts for the combinatorial contribution arising from size and shape mismatch between solute and solvent. Depending on the COSMO-RS implementation, the dispersion energy is either added here as a separate contribution to the chemical potential or included within the interaction energy term (Equation 2.39) (Gerlach et al., 2022). Finally, the transfer free energy of solute X from, for example, water to octanol, $\Delta G_{W \to OC}$, is estimated as the difference in its chemical potentials in these solvents (Warnau et al., 2021):

$$\Delta G_{W \to OC} = \mu_{OC}^X - \mu_W^X \quad (2.42)$$

This quantity reflects the thermodynamic preference of the solute for one solvent over the other and is the fundamental parametrization reference within the Martini model development.

In summary, COSMO-RS offers an efficient and broadly applicable framework for estimating chemical potentials, activity coefficients, solvation free energies, and partitioning properties, based on first-principles quantum chemistry. Therefore, the approach can serve as a valuable tool for generating reference data for Martini parametrizations, especially when experimental values are lacking.

3 Molecular Modeling and Analysis

The following sections present the molecular modeling approach, including the parametrization of the coarse-grained SOM models and the setups of the molecular systems investigated. An important objective of this work is to establish a comprehensive parametrization workflow for the coarse-grained representation of VSOMM2-generated SOM models. As will become clearer below, the principal challenge lies in developing a transferable parametrization scheme capable of converting not only specific SOM compositions but the entire spectrum of VSOMM2-generated molecular structures into their coarse-grained counterparts. To this end, the framework of the Martini 3 coarse-grained force field was employed as the methodological basis. The coarse-grained interaction parameters determined within this framework were subsequently used to construct the parametrization workflow developed in this work. For the various investigations presented, multiple coarse-grained SOM systems were generated, all based on the same underlying parametrization strategy.

Section 3.1 describes the functionality of the VSOMM, which forms the basis of the parametrization approach explained in Section 3.2. This includes the creation of the coarse-grained fragment database, the generation of coarse-grained SOM models by combining these fragments, and the automation of this process. Furthermore, the model systems used to investigate the interactions between pollutants or NPs and SOM are introduced. An overview of the most important analysis variables and how they were calculated is provided in Section 3.3. Finally, Section 3.4 summarizes the computational details of the simulations.

3.1 VSOMM Functionality

The goal of the VSOMM is to generate condensed-phase models of SOM in order to investigate the structure and dynamics of these systems with varying chemical compositions. As SOM consists of organic compounds with complex and heterogeneous properties, a systematic and flexible modeling approach is required. To this end, the modeler constructs SOM models based on a fragment database and user-defined input parameters. The resulting HS molecules are composed of the diverse fragments, also referred to as building blocks, and are heterogeneously distributed within the simulation cell. The organic fragments were initially derived from theoretical descriptions of HS molecules reported in the literature (Sein et al., 1999; Alvarez-Puebla et al., 2006; Albers et al., 2008). Subsequently, the database was extended by including fragments consisting

of methoxyl and carboxyl groups bound to aromatic rings, whose presence in HS has been confirmed by previous experiments (Saiz-Jimenez and De Leeuw, 1986; Chefetz et al., 2002; González-Pérez et al., 2004). A further reason for doing so was to increase the chemical diversity of the models and to better reflect the organic composition of HS, specifically the distribution of organic functional groups, rather than the elemental composition, as the former is considered more meaningful for reflecting the SOM chemistry. Such organic composition of HS samples is typically determined via experiments using nuclear magnetic resonance (NMR) spectroscopy, particularly ^{13}C NMR. An overview of the final fragment database used by the VSOMM2 can be found in the Supplementary Data of Escalona et al. (2021) or the Supporting Information of Dettmann et al. (2024b).

The modeler generates SOM molecules by assembling building blocks such that the concentration of functional groups matches that of the user-defined input values. These values are based on HS samples as defined by the IHSS, which provides information about fractions of functional groups such as carbonyl, carboxyl, aromatic, acetal, heteroaliphatic, and aliphatic groups, as well as elemental fractions for carbon (C) and nitrogen (N). A genetic algorithm with a fitness function is employed by the VSOMM2 to iteratively select and combine building blocks until the target composition is achieved. The pH of the HS systems is modeled by adjusting the ratio of deprotonated to protonated carboxylate groups using a fixed pK_a value. Additional parameters such as the number of fragments per molecule, which affects the molecular weight of the HS molecules, and the number of water molecules, which determines the system's dilution, can also be specified. The model construction workflow is illustrated in Figure 3.1.

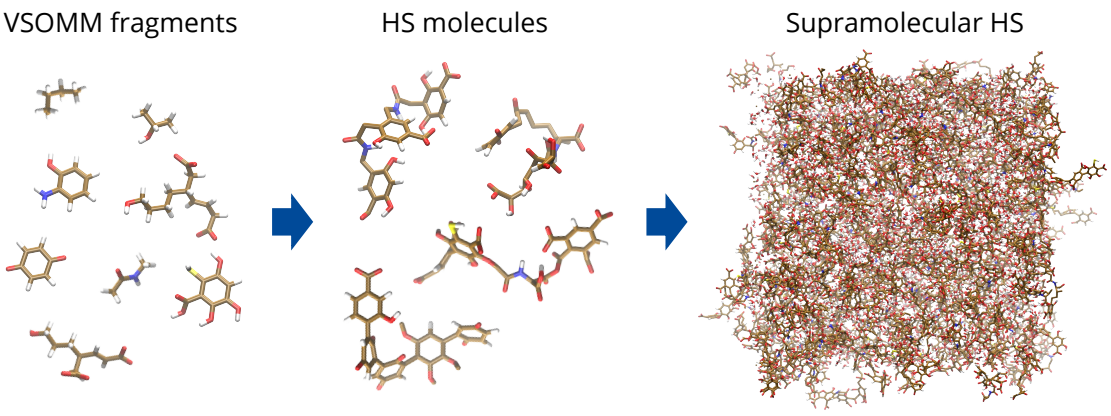


Figure 3.1: Schematic representation of the VSOMM2 model construction principle. Simple organic fragments are combined to form larger HS molecules and supramolecular HS assemblies (left to right). The complete fragment database comprises additional entities not shown in this illustration. Adapted from Gerzabek et al. (2022).

After determining the sequence of fragments to assemble into larger molecules based

on the given input, the process of model generation includes an energy minimization procedure, followed by equilibration simulations of the individual molecules. These molecules are then distributed within a periodic simulation box, solvated using the SPC water model (Berendsen et al., 1981), and neutralized with sodium or calcium counterions. As the final structures are typically of lower density due to the build-up process, they require further equilibration via MD simulations. To describe the interactions, the atomistic version of the modeler employs parameters from the GROMOS force field (Schmid et al., 2011). Following this approach, the modeler is capable of constructing chemically diverse condensed-phase SOM systems, using, for example, IHSS samples as references.

3.2 Modeling and Parametrization Approach

Coarse-Grained Fragment Database

In order to apply the VSOMM2 modeling approach within the coarse-grained framework, the first step in this work involved constructing a fragment database in the coarse-grained representation. This was followed by the determination of bonded interaction parameters between different fragments, enabling their assembly into larger coarse-grained SOM molecules. The Martini framework provides a set of guidelines for coarse-graining molecules to ensure force field compatibility and transferability. These guidelines include preserving functional groups and molecular symmetries as far as possible and using specific bead sizes for different chemical groups, for instance, tiny beads for aromatic moieties to maintain planarity and small beads for aliphatic ring structures. Additional rules are detailed in the Martini 3 publication (Souza et al., 2021) and in the introduction to parametrizing small molecules (Alessandri et al., 2022).

According to the Martini building block approach, pre-parametrized beads are, in principle, interchangeable between different molecules. For example, aromatic moieties are commonly represented using the TC5 bead. These standard beads serve as the starting point for the bead assignment in the coarse-graining process. Further refinement is undertaken based on the structural characteristics of individual molecules. For instance, molecules such as caffeine, with conjugated carbonyl groups, may require modified bead assignments to reflect their altered polarity (Alessandri et al., 2023). Additionally, matching thermodynamic data, particularly transfer free energies, is used to validate or assign bead types, especially when dealing with unknown or novel functional groups that are not covered by standard Martini bead definitions. Based on this approach, the VSOMM2 fragments were coarse-grained accordingly. The resulting fragment database, along with the assigned bead types, is presented in Figure S1 in the Supporting Information of Dettmann et al. (2024b).

To verify the developed bead assignments that would determine the nonbonded interactions, transfer free energies of the coarse-grained fragments were calculated and

compared to reference data, including octanol–water, hexadecane–water, and chloroform–water transfer free energies. While experimental values are typically used for such comparisons, most of the fragments examined do not belong to chemical substances that have been intensively studied before, and therefore no experimental data is generally available. Thus, the COSMO-RS approach was employed in this work to calculate these free energies, which served as a reference. This approach is introduced in section 2.7.

The bonded interactions of a molecule represent the effective connections between the underlying molecular parts, which are represented by coarse-grained beads. Determining how the beads are connected to each other within a molecule and which interaction potentials should be applied, follows directly from the mapping of the molecule. The interaction types used are described by bond, angle, and dihedral angle potentials. Furthermore, virtual sites are applied, which are interacting particles whose positions are determined as a linear combination of coordinates of other "constructing" beads, allowing certain ring molecules to be represented in the coarse-grained Martini model. Since, in contrast to nonbonded interactions, no data bank for bonded interaction parameters is available, these have to be determined using reference data. To generate these, each atomistic fragment molecule is first solvated in a water box and simulated for a certain time to sample its conformational space, resulting in a trajectory of the molecule. The atoms of the fragment are then grouped according to the defined mapping, yielding the so-called "mapped trajectory". An ideally parametrized coarse-grained fragment would exactly reproduce the motion observed in this trajectory. To derive optimal bonded parameters from the mapped trajectory, the Swarm-CG program was used in this work. It matches the probability distributions of bonded interactions in the coarse-grained representation to those of the mapped molecule by performing iterative coarse-grained simulations of the same fragment in water, under conditions identical to the atomistic simulations (e.g., same box size and solvent content, with one water bead representing four atomistic water molecules). These distributions describe, for example, the occurrence of distances between two bonded beads observed over the trajectory. The workflow for generating the coarse-grained fragment database is depicted in Figure 3.2.

As bonded interactions are often cross-correlated, meaning that adjusting one bond or angle can influence neighboring ones, Swarm-CG applies a particle swarm optimization algorithm (Kennedy and Eberhart, 1995; Nobile et al., 2018) to efficiently explore the multidimensional parameter space and to identify an optimal solution. The level of agreement between the atomistic and coarse-grained distributions is quantified using the Earth Mover's Distance (EMD) (Rubner et al., 1998), which is incorporated into a total mismatch score (TMS) calculated from all bonded distributions (further details of the TMS are described in the Supporting Information of Dettmann et al. (2024a)). The optimization process in Swarm-CG is designed to minimize this TMS as much as possible. In addition to examining the match between bonded distributions, the radius of gyration (R_g) and the solvent-accessible surface area (SASA) of the atomistic fragments were used as reference parameters for the coarse-grained fragments to evaluate the quality of the coarse-grained model. The calculation of both is described in Section 3.3.

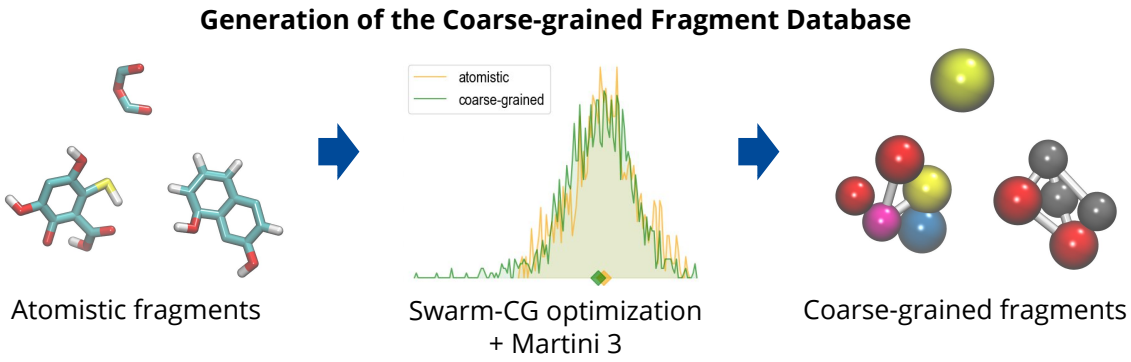


Figure 3.2: Workflow for generating the coarse-grained fragment database. Atomistic fragments are simulated in water to sample representative configurations. Corresponding Martini 3 mappings are assigned, and parameters are iteratively optimized using Swarm-CG to match the bonded probability distributions of the coarse-grained (green) and mapped atomistic (yellow) models. The resulting set of optimized coarse-grained fragments constitutes the fragment database (Dettmann et al., 2024b).

Assembly of Coarse-Grained SOM

Once the determined bonded and nonbonded interaction parameters have been stored in the coarse-grained fragment database, coarse-grained SOM models can be simulated by connecting different fragments to larger HS molecules according to the atomistic reference structures and following the construction philosophy of the VSOMM2. At this stage, the interaction parameters within the fragment database are fixed.

LHA is a widely studied standard used in HS research due to its well-characterized chemical composition (Thorn et al., 1989). Its reproducibility and ease of experimental use make it a common subject of investigation in soil science (Aeschbacher et al., 2009; Petrov et al., 2017). For explorative purposes, coarse-grained models of LHA were generated in this work. To introduce compositional variability, three distinct LHA systems were created using different random seeds in the VSOMM2 model generation process. The corresponding atomistic structures were simulated via MD to obtain reference trajectories for the coarse-grained models.

While the bonded interaction parameters within individual fragments were already defined, as detailed in the previous section, connecting these fragments to form larger SOM molecules requires additional bonded parameters. However, deriving these parameters using the same iterative optimization strategy as for the fragment database, based on Swarm-CG, is computationally demanding for such large molecules. Therefore, the DBI method was employed instead. This approach enables the estimation of interaction parameters directly from the probability distributions of the bonded interactions in the mapped atomistic trajectories, without iterative optimization. The required distributions

were extracted from the atomistic simulations of the LHA systems.

This approach allows coarse-grained SOM models to be constructed and simulated based on atomistic reference data. In principle, any SOM composition can be modeled in this way, provided that atomistic reference simulations of the corresponding system have been carried out.

Automated Model Construction and Parameter Rebalancing

While the combination of interaction parameters from the fragment database and the additional bonded parameters determined via DBI enables the simulation of any coarse-grained SOM system, conducting atomistic reference simulations for full SOM systems remains time-consuming, particularly when targeting larger molecular structures, which is an important motivation for coarse-graining. This challenge becomes even greater when exploring various SOM compositions. In addition to these limitations, it was observed that the existing set of parameters only supported a time step of 10 fs without causing numerical instabilities, which are connected to a divergence in the integration of the equations of motion. This time step represents the lower limit typically used in Martini simulations (Grünewald et al., 2020) and indicates that the forces arising from the interactions in the system are too large for stable integration at larger time steps.

To address these issues, the model parametrization was repeated and refined using Swarm-CG with an imposed upper limit on the force constants of the angle potentials. These angle interactions were identified as a major source of large instantaneous forces, often resulting in simulation crashes at higher time steps. Since constraining force constants might compromise the accuracy of the model, the impact on distribution matching was quantified using the TMS by comparing values before and after applying the restriction. Furthermore, to account for the missing bonded interaction parameters between different fragment molecules, which are essential for assembling larger SOM structures, a conformer search approach was incorporated into the parametrization workflow. The idea behind this method is to generate atomistic conformers of large SOM molecules, providing a computationally cheaper alternative to full MD simulations for sampling the conformational space. From these static molecular configurations, equilibrium bond lengths and angles for the coarse-grained models can be extracted by averaging over a diverse set of conformers after applying the corresponding mapping scheme. Specifically, the atomistic conformers were generated using the ETKDG method (Riniker and Landrum, 2015; Wang et al., 2020), a stochastic conformer generation algorithm based on distance geometry (Blaney and Dixon, 1994).

The refined parametrization workflow (illustrated in Figure 3.3), incorporating the modified parameters for the coarse-grained fragment database and the ETKDG approach, was implemented in the SOM2CG tool (Dettmann, 2025). This enables the automated conversion of atomistic VSOMM2 topology files into a coarse-grained representation, eliminating the need for additional atomistic MD simulations as reference input. This significantly improves the accessibility of the coarse-grained models, reducing both the

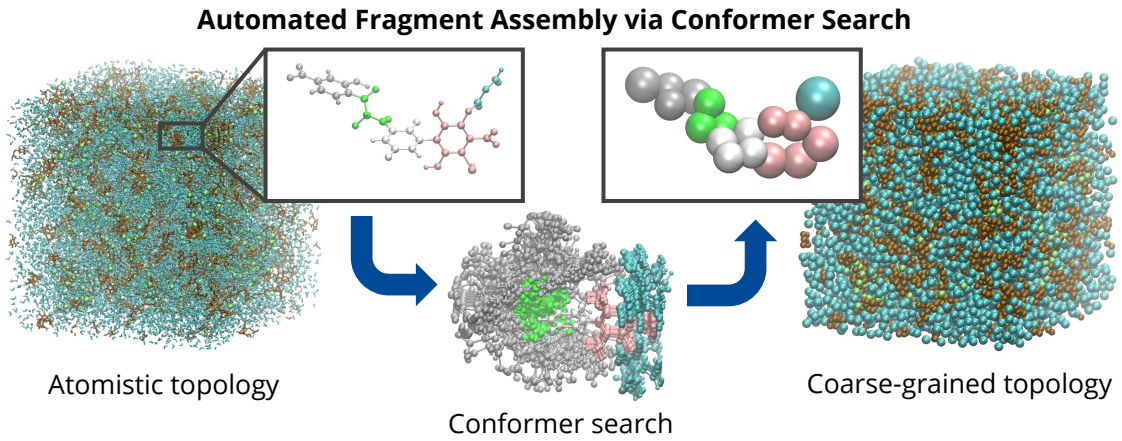


Figure 3.3: Automated parametrization of atomistic VSOMM2 systems into coarse-grained representations using the conformer search approach. Molecular information from atomistic HS molecules is processed, and conformer searches provide equilibrium data for fragment assembly, which, together with the coarse-grained fragment database, enables the generation of coarse-grained topologies. The entire workflow is fully automated (Dettmann et al., 2024a).

time required for parametrization and the complexity of the setup. Because the approach relies on the fragment database, any SOM system constructed with the VSOMM2 can, in principle, be converted into a coarse-grained representation. This allows for the efficient and flexible simulation of chemically diverse SOM compositions at coarse-grained resolution.

To explore how the composition of SOM influences its properties at extended time and length scales, a selection of HS samples, documented by the IHSS, was generated using the VSOMM2 and subsequently converted into coarse-grained systems using the SOM2CG tool. The selection of specific IHSS samples was guided by a principal component analysis conducted by Escalona et al. (Escalona et al., 2023), who classified HS samples based on their acidic and aromatic content. Their two-dimensional compositional map enabled the identification of representative systems with high and low acidity, high and low aromaticity, and intermediate characteristics, thereby capturing a broader spectrum of different SOM characteristics. Five HS samples, labeled SOM I to V, were chosen to reflect the intrinsic heterogeneity of natural SOM and are described in Table 1 of the corresponding study (Dettmann et al., 2024a). Furthermore, the water content of each system was also systematically varied to assess hydration-dependent effects. The properties investigated included system densities, potential energies, and other structural quantities.

Pollutant Parametrization and Investigation

Pollutants in terrestrial ecosystems interact with SOM, but their behavior is difficult to study experimentally at the molecular level due to the system's complexity and heterogeneity. The availability of coarse-grained SOM systems developed in this work facilitates the investigation of such interactions through molecular simulations. Specifically, two environmentally relevant compounds were chosen for this study: HCB, a POP, and SAA, a polar antibiotic. Both were investigated in the context of their interaction with SOM using the coarse-grained modeling approach. As no Martini models were available for these compounds, their coarse-grained representations were created. The parametrization strategy followed a similar protocol to that used for the SOM fragment molecules. Bonded interaction parameters were derived by conducting atomistic simulations of the individual pollutant molecules in water to sample their conformational space, which served as reference data for the Swarm-CG optimization. Mapping schemes were defined by preserving key functional groups and respecting molecular symmetry, in line with the principles of the Martini framework. In the following section, the mapping choices are briefly outlined to make it easier to understand the parametrization procedure. Please refer to Figure 3.4 for a comparison between the atomistic and coarse-grained representation of SAA and HCB.

For HCB, the mapping strategy followed the general approach used for aromatic compounds in Martini, where the six-fold symmetry of the atomistic ring is reduced to a three-fold symmetry by combining two carbon atoms of the aromatic ring together with their attached substituents into a single bead. In contrast to benzene, where hydrogen atoms are attached to the carbons, HCB has chlorine substituents, meaning that each bead represents four non-hydrogen atoms. In this case, a regular-sized bead was used to represent each aromatic moiety of HCB, resulting in a total of three regular-sized beads for the entire molecule. Since such beads have not previously been applied to this type of aromatic system in Martini, the bead type was determined by comparing the transfer free energies of the HCB model to experimental and COSMO-RS reference values. To this end, free-energy calculations were conducted for the coarse-grained HCB model.

In contrast, the structure of SAA requires a different mapping strategy for its aromatic parts and attached functional groups. The aromatic ring is represented by standard beads for aromatic moieties (TC5 bead), with the attached amino group mapped to a TN6d bead. The sulfonamide group forms a fully branched five-atom unit, including one carbon atom from the aromatic ring, and is represented by a regular-sized bead. While the bead types for the aromatic and amino groups are predefined within the Martini framework, the bead type for the sulfonamide group was unknown and thus required parametrization. This was achieved by varying the bead type until the calculated transfer free energy of the complete SAA model sufficiently matched the corresponding reference values. The transfer free energies were determined using the same free-energy calculation approach as for HCB.

Following the parametrization of HCB and SAA, the resulting coarse-grained models

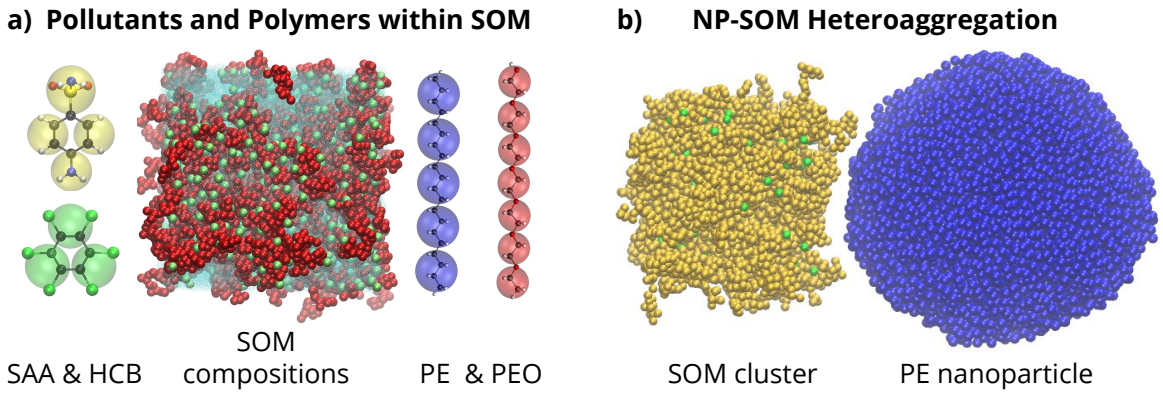


Figure 3.4: Overview of the model systems used to investigate pollutant interactions with SOM and NP–SOM heteroaggregation. a) Atomistic and coarse-grained representations of the pollutants SAA (yellow) and HCB (green), as well as polymer chains of PE (blue) and PEO (red). The actual PE chains used in the simulations consisted of seven coarse-grained beads. For both pollutants and polymers, 20 distinct starting positions were defined within the SOM systems to generate independent trajectories, and the SOM composition was varied. The HS molecules are depicted in dark red, water in cyan and calcium ions in lime. b) Initial setup of the NP–SOM heteroaggregation simulations, representative of a series of systems at different pH conditions in water, which is not shown for clarity (Dettmann et al., 2025a,b).

were employed to study their interactions and mobility within representative SOM systems. A characteristic behavior of pollutants in SOM matrices is their entrapment within voids, as suggested by experimental observations and previous simulations (Ahmed et al., 2015). Thus, five representative SOM systems were selected, following the same rationale as in the previous section to cover a range of chemical compositions. Each of these systems was simulated separately with HCB and SAA, respectively. Figure 3.4 a) provides an overview of the atomistic and coarse-grained representations of SAA and HCB, as well as an illustration of an SOM system under investigation. The water content was not varied; instead, the systems with the highest hydration levels from the investigation, described in the previous section, were used. A high water content promotes greater diffusivity, thereby enhancing the sampling of heterogeneous structures, an essential requirement given the complexity of the SOM matrix. To further ensure sufficient exploration of different interaction sites, each pollutant–SOM combination was simulated in 20 independent replicas, differing in the initial spatial placement of the pollutant molecule within the SOM system. The coadsorption of multiple pollutant molecules was not examined, as the focus was solely on single-molecule pollutant–SOM interactions.

For further details of the pollutant parametrization, free-energy calculations, and the SOM model systems, please refer to the Materials and Methods section of Dettmann et al. (2025a).

Modeling NP–SOM Interactions

Motivated by the considerable uncertainty surrounding the fate and effects of NPs in terrestrial ecosystems (Baho et al., 2021), the final study focused on modeling polymer plastics in the presence of SOM molecules. Plastic particles in the environment occur in a wide range of sizes, polymer types, and degrees of environmental aging, such as oxidation (Ivleva, 2021). To capture different aspects of polymer–SOM interactions, the model systems were designed from two complementary perspectives.

In the first approach, polymer chains were embedded in SOM matrices in a setup analogous to the pollutant–SOM study described in the previous section. Polyethylene (PE), one of the most widely produced plastics (Geyer et al., 2017) and a major constituent of global plastic waste, was chosen as the representative polymer type. In addition, its oxidized counterpart poly(ethylene oxide) (PEO) was included to represent a variation in polarity. Both polymers in their atomistic and coarse-grained representations are shown in Figure 3.4 a), together with an SOM system that formed part of the investigation. Here, a PE bead represents two monomers, i.e., four carbon atoms and the associated hydrogen atoms, while a PEO bead describes one monomer of the polymer. The coarse-grained models were constructed using the polyply tool (Grünewald et al., 2022). For the simulations, the previous SOM model set was adapted and reduced to two HS systems: one with a high concentration and one with a low concentration of acidic functional groups. This simplification focused the investigation on a single compositional variable, as the pollutant study had shown that carboxylate concentration strongly influences the pollutant behavior in SOM. The rest of the simulation design followed the protocol established for the pollutant systems, including the use of multiple initial placements of the polymer chains within the SOM matrices and the application of a relatively high water content.

In the second approach, an NP particle was modeled, consisting of relatively long PE chains (each comprising 500 beads) aggregated in water, where the total number of beads was estimated using Flory theory and parameters from previous work (Dettmann et al., 2021). The resulting particle adopted a roughly spherical shape with a diameter of approximately 10 nm. An SOM cluster, comprising HS molecules and stabilized by calcium ions, was then positioned near the NP particle to study potential heteroaggregation and interaction motifs (see Figure 3.4 b) for the initial position of a representative system). Furthermore, the pH conditions were systematically varied by adjusting the fraction of protonated to unprotonated carboxylate groups, enabling the assessment of pH effects on NP–SOM interactions. This setup provided a basis for examining how nanoscale plastic particles interact with complex SOM matrices under different chemical conditions.

3.3 Analysis

Radius of Gyration

The R_g is a measure of the compactness of a molecule or structure and is defined as the mass-weighted root-mean-square distance of the considered N particles to their center of mass $\vec{r}_c$:

$$R_g = \sqrt{\left(\sum_{i=1}^{N} m_i \left\|\vec{r}_i - \vec{r}_c\right\|^2\right)\left(\sum_{i=1}^{N} m_i\right)^{-1}}. \tag{3.1}$$

It is a straightforward metric that can be directly calculated from molecular simulation trajectories based on particle coordinates. Within the Martini 3 framework, the R_g is particularly useful for parametrization purposes, as it enables the comparison of the spatial extent and flexibility (via the R_g distribution) of coarse-grained molecules with those of their all-atom counterparts. The R_g can also be related to dilute-solution thermodynamics through the Flory theory (Rubinstein and Colby, 2003), offering a way to evaluate whether the balance between self-interactions and solvent interactions in a coarse-grained model is accurately represented (Grünewald et al., 2022). It is likewise a useful quantity for polymer systems, where it can be employed to quantify collapse or compaction effects in confined molecular environments such as SOM, which is characterized by the presence of molecular voids.

Solvent-Accessible Surface Area

The SASA is the area traced by a spherical probe particle, representing the solvent, as it rolls over a molecule's van der Waals surface, a concept formalized by Lee and Richards (1971) and refined by the Shrake–Rupley algorithm (Shrake and Rupley, 1973). In GROMACS, the SASA is computed using the Double Cubic Lattice Method, a fast numerical variant of the Shrake–Rupley approach that preserves the same geometric definition while accelerating the evaluation for larger systems (Eisenhaber et al., 1995). Within the Martini framework, the SASA serves as a structural target linking molecular shapes and volumes between atomistic and coarse-grained models, thereby contributing to achieving accurate molecular packing. Recent Martini 3 parametrizations explicitly compare the coarse-grained SASA against the atomistic ones, for example in carbohydrate models (Grünewald et al., 2022). Likewise, the SASA benchmarking against atomistic references was applied to the SOM fragments and pollutant molecules in this work.

Diffusion Coefficient

Pollutants interact with soil components through various mechanisms, which can influence their mobility and, in some cases, lead to aggregation. A key quantity for describing how fast molecules move within a molecular system is the diffusion coefficient. This

parameter is of particular interest as it can provide insight into the accumulation behavior of molecules, which is directly related to their fate in ecosystems.

In molecular simulations, diffusion is typically determined from the mean squared displacement (MSD) $\langle r^2(t)\rangle$, which quantifies the average squared distance traveled by the molecules over time:

$$\langle r^2(t)\rangle = \frac{1}{N}\sum_{i=1}^{N}\left\|\vec{r}_i(t) - \vec{r}_i(0)\right\|^2. \tag{3.2}$$

From the MSD, the diffusion coefficient D can be obtained using the Einstein relation (Einstein, 1905):

$$D = \lim_{t\to\infty}\frac{\langle r^2(t)\rangle}{6\,t}. \tag{3.3}$$

In practice, this is achieved by performing a linear fit to the MSD curve in the region where it exhibits approximately linear behavior. In order to determine the diffusion coefficient in this way, the MSD must show a distinct linear regime over time. Before the MSD calculation, molecular trajectories need to be unwrapped to remove discontinuities caused by the periodic boundary conditions.

3.4 Computational Details

The following software packages, algorithms, and simulation parameters were used throughout the investigations. All MD simulations in this work were performed using the GROMACS software package, version 2019.4 (Abraham et al., 2015). Analyses were carried out with GROMACS tools and custom Python scripts, employing the packages NumPy (Harris et al., 2020), MDAnalysis (Michaud-Agrawal et al., 2011; Gowers et al., 2016), Matplotlib (Hunter, 2007), IPython (Pérez and Granger, 2007), and RDKit (Landrum et al., 2023). Molecular structures and trajectories were visualized using VMD (Humphrey et al., 1996). For the coarse-grained systems, all simulations employed the leap-frog integrator with a time step of either 10 fs or, after reparametrization, 20 fs. The systems were simulated at $T = 298.15$ K using the velocity-rescale thermostat (Bussi et al., 2007) to maintain temperature. During initial equilibration of the box dimensions, the Berendsen barostat (Berendsen et al., 1984) was used for the pressure coupling. The production runs were carried out with the Parrinello–Rahman barostat (Parrinello and Rahman, 1981). In both cases, the reference pressure was set to 1 bar. Nonbonded Lennard-Jones interactions used a cutoff of 1.1 nm, the standard for the Martini force field, together with the Verlet cutoff scheme. Coulomb interactions were evaluated with the same cutoff using the reaction-field method. Bond constraints were applied using the LINCS algorithm (Hess et al., 1997), and periodic boundary conditions were imposed in all three dimensions for every simulation box. The COSMO-RS calculations were carried out using the Amsterdam Modeling Suite version 2023.1 (Pye and Ziegler, 1999; Louwen et al., 2023). Further details of specific parameter choices and calculation setups can be found in the respective computational details sections of the individual publications.

4 Results and Discussion

This chapter summarizes the main results of the parametrizations and simulations of the coarse-grained SOM systems. A major motivation for these investigations is the limited experimental access to molecular-level interactions in SOM and the opportunity to extend existing VSOMM simulation approaches to larger spatial and temporal scales through coarse-graining. Two main objectives were pursued in this work: first, the development and parametrization of coarse-grained SOM models based on structures generated with the VSOMM2; and second, the application of these models to study the mobility of pollutants within SOM systems as well as the heteroaggregation of SOM with NPs.

4.1 Benchmarking the Coarse-grained Fragment Database

The first step in the generation of coarse-grained SOM models involved determining the bonded and nonbonded interaction parameters for the fragment database. The bonded interactions were optimized using Swarm-CG, with the main objective of maximizing the overlap between the mapped and coarse-grained bonded distributions, which was quantitatively evaluated via the TMS. Further measures relating to the parametrization accuracy, such as the R_g (see Figure 2 in Dettmann et al. (2024b)), supported this approach, suggesting that correctly reproducing bonded distributions also yields realistic molecular sizes and shapes. A good agreement with the atomistic SASA was achieved (Figure 4.1 a)), as indicated by the high correlation coefficient of $R^2 = 0.985$. Although the SASA was not a direct target in the Swarm-CG optimization, the molecular shapes were adequately captured. Minor deviations were nevertheless observed, which can be attributed to the influence of bead sizes on the coarse-grained representation; these could, in principle, be reduced by manual fine-tuning. Since the molecular shape directly affects the packing of molecules in condensed-phase systems, and the SASA is known to be an important determinant for reproducing bulk properties such as the system density in Martini 3 (Alessandri et al., 2022), its accurate representation gives confidence in the overall parametrization.

The nonbonded interactions were assigned using the standard Martini 3 bead definitions and validated against transfer free energies calculated with COSMO-RS (Figure 4.1 b)). For this purpose, the transfer free energies between octanol–water, hexadecane–water, and chloroform–water phases were determined to assess the partitioning behavior of the coarse-grained fragments in different chemical environments. As the bead assignment

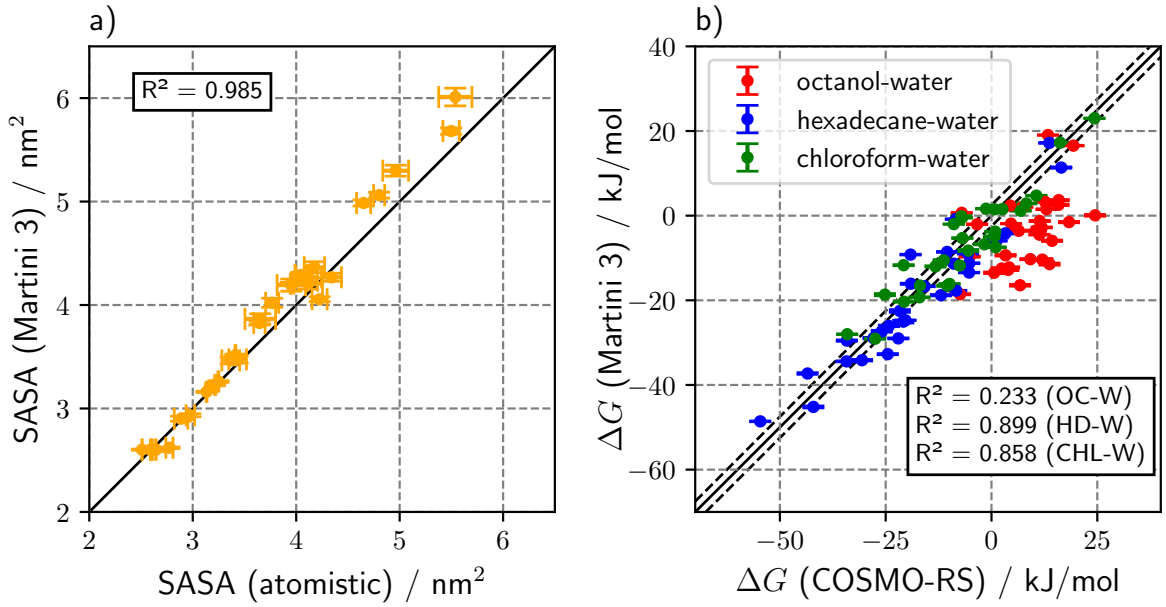


Figure 4.1: Parametrization results of the fragment database. Panel a) shows a comparison of the SASA values of coarse-grained fragments with their atomistic counterparts, with error bars indicating standard deviations. Panel b) depicts the transfer free energies (ΔG) calculated with the coarse-grained fragment models, compared to reference data obtained using the COSMO-RS method. The ΔG values for octanol–water (OC–W), hexadecane–water (HD–W), and chloroform–water (CHL–W) are shown in red, blue, and green, respectively (Dettmann et al., 2024b).

directly defines the polarity and partitioning tendencies of a fragment, it is essential to correctly reproduce these reference values. The results showed good correlations for the hexadecane–water and chloroform–water targets, indicating a reliable reproduction of the thermodynamic properties of the fragment molecules. In contrast, the octanol–water values exhibited larger deviations, which can be partly explained by the lower hydrophobicity of octanol compared to hexadecane or chloroform and the resulting narrower spread of the values. A potential improvement could be achieved by slightly adjusting the bead assignments, although this would be a time-consuming task given the large number of molecules involved. Recently developed automated parametrization tools for Martini 3 may facilitate such refinements in future studies (Szczuka et al., 2025). It should also be noted that the optimal benchmark would involve using experimental free-energy data, which would provide a more accurate reference and thereby improve the assessment of the model performance. Overall, however, the coarse-grained fragment models sufficiently reproduced the key target properties, and the resulting fragment database provided a robust foundation for the construction of coarse-grained SOM systems in this work.

4.2 Coarse-Grained Modeling of Leonardite Humic Acid

LHA was chosen as an application case for constructing coarse-grained SOM models since, as previously stated, it represents a well-defined sample and has been frequently investigated in previous studies (Petrov et al., 2017, 2019). The major challenge in this context was to assemble the coarse-grained fragments into larger HS molecules with appropriate bonded parameters. To address this objective, the DBI approach was applied and benchmarked by comparing the average bonded distribution values of coarse-grained and mapped reference systems, as shown in Figure 6 in Dettmann et al. (2024b). Using this procedure, distinct LHA systems were constructed and subsequently converted to the coarse-grained representation through the assembly of coarse-grained fragments.

To evaluate the influence of the determined interaction parameters on the macroscopic properties of SOM systems, the densities of the LHA models were calculated and comparisons were made between atomistic and coarse-grained representations. Figure 4.2 a) shows the system densities as a function of the water content, where the latter is expressed by the heavy-atom fraction of water, x_{water}, with respect to the entire system. In both representations, the densities decrease with increasing water content, reflecting more compact LHA structures under dry conditions. A systematic underestimation of the atomistic densities by the coarse-grained systems is apparent, although the discrepancy becomes smaller at higher water contents. The largest deviation between the densities for the driest cases is approximately 11 %, a value that remains acceptable given the complexity of the systems. However, this finding contrasts with the good agreement between the atomistic and coarse-grained SASA values. One explanation for this is that, at lower water contents, the pore and void sizes in the LHA structures are reduced, making it more difficult for coarse-grained water beads to accurately fill these cavities. This interpretation is supported by a comparison with simulations employing tiny water beads (2:1 mapping), which led to reduced deviations in density, particularly at lower water contents (Figure 5 in Dettmann et al. (2024b)), although some discrepancies remained.

Another decisive factor affecting the density deviations is the interaction between calcium ions and carboxylate groups. Figure 4.2 b) shows the radial distribution functions between these groups, comparing the mapped and coarse-grained representations for the system with the lowest water content. The mapped atomistic system serves as a reference, as it allows a direct comparison to be made with the coarse-grained model. The radial distribution function indicates the most probable distances between carboxylate groups and calcium ions. While the mapped system exhibits multiple peaks, the coarse-grained system produces only a single peak at a larger distance, demonstrating that average calcium–carboxylate interactions are shifted outward in the coarse-grained representation. The importance of such interactions for the aggregation of HS systems has been demonstrated previously by Galicia-Andrés et al. (2021a). The systematic underestimation of calcium–carboxylate interactions in the coarse-grained model most likely contributes to less compact packing and thus to lower densities. At higher water

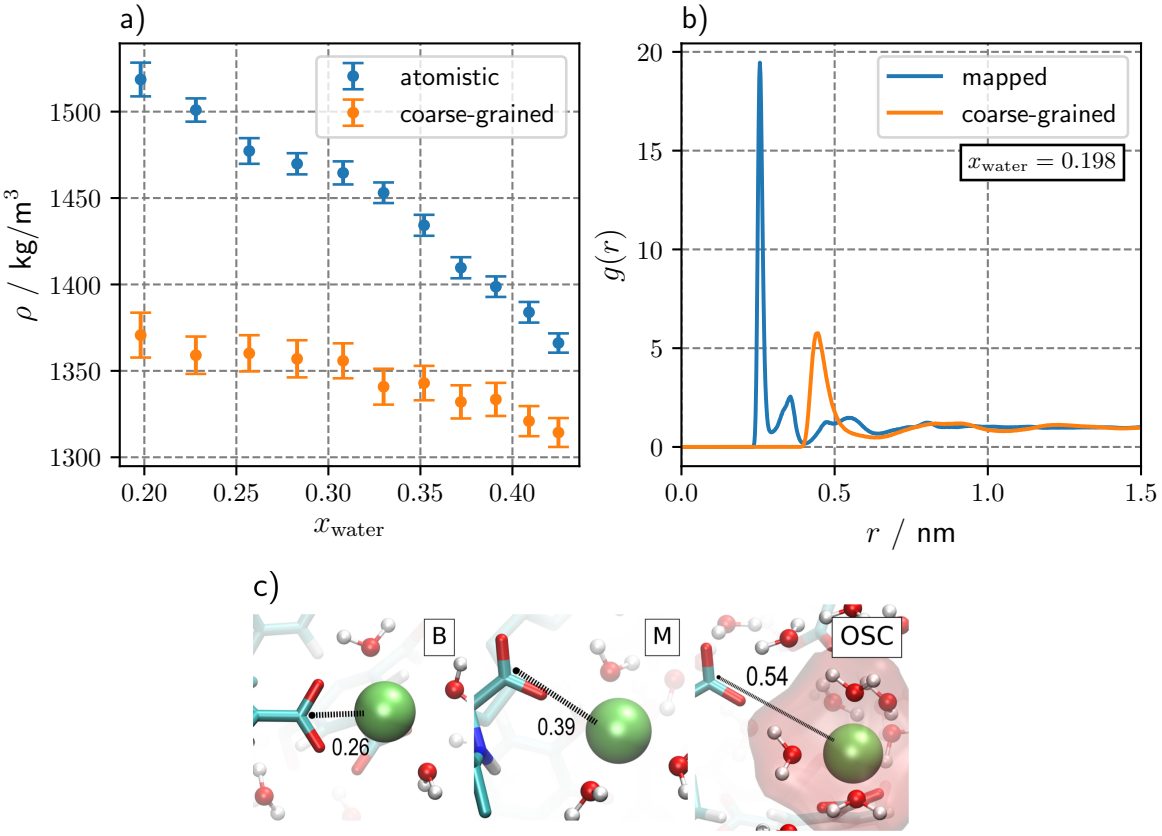


Figure 4.2: Panel a) shows the system densities of LHA under different hydration levels, comparing atomistic and coarse-grained LHA systems; error bars denote standard deviations. Panel b) depicts the radial distribution functions between carboxylate groups and calcium ions for the lowest water content, comparing the mapped atomistic system (blue) with the coarse-grained model (orange). Panel c) illustrates distinct calcium–carboxylate complexes identified in the atomistic systems, namely bidentate (B), monodentate (M), and outer-sphere complexes (OSCs). LHA molecules are shown as sticks, water molecules as balls and sticks, and calcium ions as green spheres. Oxygen, carbon, nitrogen, and hydrogen atoms are represented in red, cyan, blue, and white, respectively. Center-of-geometry distances between carboxylate groups and calcium ions (black) are given in nanometers (Dettmann et al., 2024b).

contents, the discrepancy in the radial distribution functions becomes smaller (see Figure 7 in Dettmann et al. (2024b)), which is consistent with the reduced deviations in density observed under these conditions.

The multiple peaks observed in the mapped radial distribution functions (Figure 4.2 b)) indicate the presence of distinct configurations of calcium–carboxylate complexes. A

categorization of these complexes from the atomistic analysis is illustrated in Figure 4.2 c), which shows why such behavior cannot be captured by two coarse-grained point particles interacting through simple spherical potentials. In particular, bidentate (B) complexes play a central role in promoting tighter packing of the structures. The first peak in Figure 4.2 b) can be attributed to these complexes. As the coarse-grained model can only reproduce an average interaction distance, the structural contribution of B complexes is lost, which helps to explain the systematic underestimation of densities. A more detailed discussion of these configurations and additional analyses of hydrogen bonding are provided in Dettmann et al. (2024b).

It is inherent in the coarse-graining approach that not all configurational features of the atomistic systems can be reproduced, which in this case affects the system densities, particularly under dry conditions. Such deviations at low water contents may also reflect a general limitation of the Martini framework, suggesting that it would be preferable to use more hydrated systems when aiming for reliable quantitative results. Nevertheless, the deviations at the higher densities do not compromise the overall applicability of the model. Future improvements could involve the use of coarse-grained calcium ions represented by tiny beads, which would allow closer calcium–carboxylate interactions and thereby mimic the effect of B complexes more accurately.

4.3 Reparametrization and SOM Diversity

The initial coarse-grained LHA models revealed two main challenges: limited numerical stability, which required a reduced integration time step and thus decreased simulation performance, and the reliance on atomistic MD simulations for constructing the coarse-grained models using the DBI approach. Therefore, it is to be expected that the bonded interaction parameters connecting different fragments will be sensitive to the sampling quality of the reference systems in the MD simulations, which most likely varies with system composition and water content, and for which no standardized procedure exists. To address these issues, a modified parametrization protocol was developed, enabling the automated construction of coarse-grained SOM models without the need for atomistic reference simulations. In the following section, the main outcomes of this reparametrization are summarized, together with the results obtained for SOM systems of diverse compositions.

In the initial coarse-grained LHA models, the integration time step had to be limited to 10 fs, which is acceptable within the Martini framework but significantly reduces simulation efficiency compared to the 20 fs time step typically used in Martini models. To improve numerical stability and thereby enable larger time steps, a reparametrization of the bonded interaction parameters was carried out using Swarm-CG with restricted force constants. Detailed analyses of the trajectory data indicated that high-frequency vibrational motions, in particular those arising from angle potentials, were a major source of the numerical instabilities. To evaluate the effect of the reparametrization,

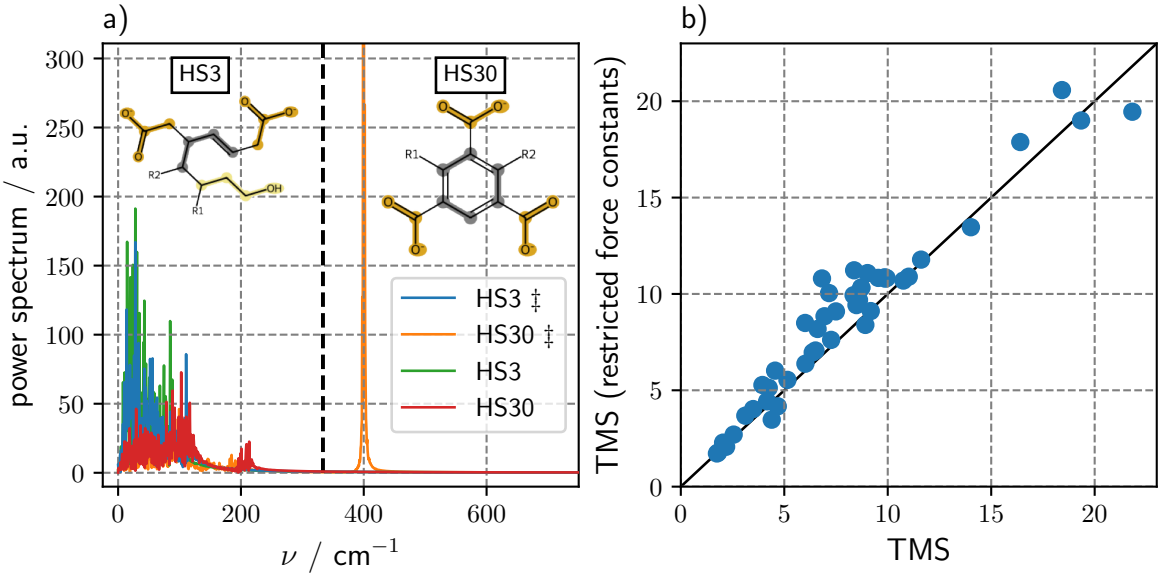


Figure 4.3: Panel a) shows the vibrational power spectra of the fragments HS3 and HS30, where ‡ denotes the spectra before reparametrization. The dashed line indicates the maximum frequency that can be resolved with a time step of 20 fs. Panel b) compares the TMSs of the individual fragments before (x-axis) and after (y-axis) reparametrization. Values close to the diagonal line indicate only minor changes, demonstrating that the restriction of force constants had little impact on the accuracy of the models (Dettmann et al., 2024a).

the vibrational power spectra of individual fragments solvated in water were calculated. Figure 4.3 a) shows spectra for the fragments HS3 and HS30 before (‡) and after reparametrization. While HS3 was only marginally affected, HS30 displayed a pronounced reduction of high-frequency components, with the peak around 400 cm^{-1} vanishing after reparametrization. Thus, the vibrational motions could be resolved with larger integration steps, which was confirmed by stable simulations of SOM systems using a time step of 20 fs.

Restricting the force constants reduces the accessible parameter space, which could in principle lead to a less optimal reproduction of bonded distributions and thus to a higher mismatch with the reference. To assess this effect, the TMS of each fragment was compared before and after reparametrization, as shown in Figure 4.3 b). Each data point belongs to one fragment, with the two TMS values plotted against each other. Since the TMS is an absolute measure of the mismatch that depends on the number of bonded terms within a fragment, only comparisons of the same fragment are meaningful, and deviations from the diagonal line indicate changes due to the reparametrization. With a few exceptions, most fragments showed very similar TMS values before and after

reparametrization. The overall changes were small, suggesting that restricting the force constants did not significantly compromise the accuracy of the models. Given the clear gain in numerical stability and performance, this outcome indicates that the approach used represents a beneficial compromise.

Beyond improving the numerical stability, the parametrization protocol was further extended to enable the generation of coarse-grained topologies for SOM systems of different compositions. For this purpose, the ETKDG method was incorporated into the fragment assembly, allowing automated parametrization without the need for atomistic reference trajectories. In this context, it was necessary to determine the optimal number of conformers to achieve convergence of the equilibrium distances and angles (see Figure 5 in Dettmann et al. (2024a)). The entire parametrization procedure was implemented in a Python tool (Dettmann, 2025).

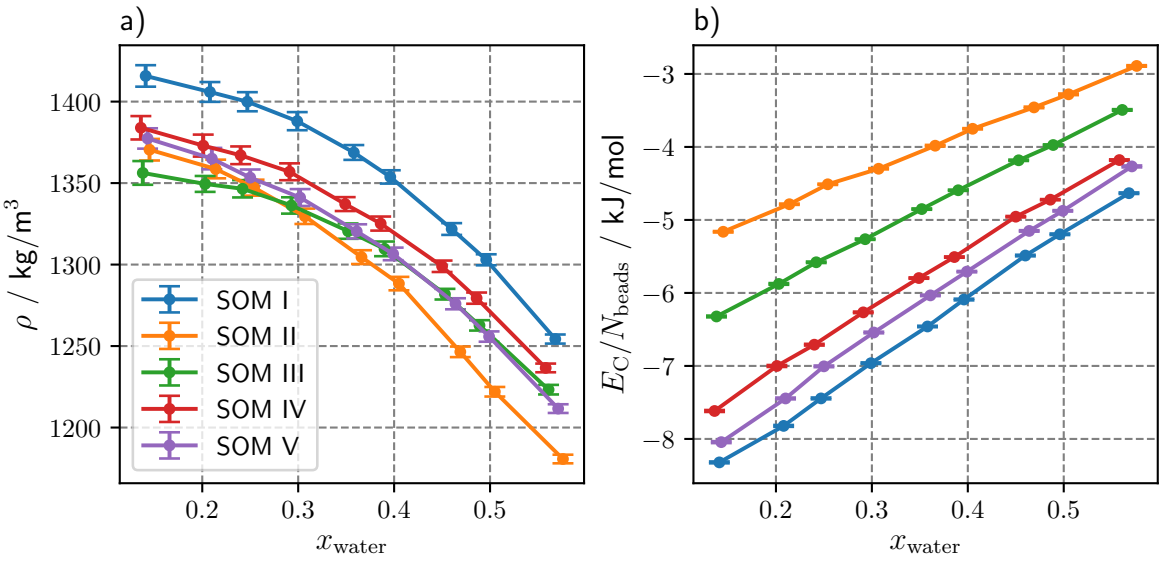


Figure 4.4: Panel a) shows the average system densities of the different SOM compositions as a function of water content. Panel b) depicts the Coulomb energies E_C, normalized by the number of HS beads N_{beads}, for the same systems and hydration levels. In both panels, error bars represent standard deviations (Dettmann et al., 2024a).

Using the modified parametrization approach, five different SOM compositions representing HS samples from the IHSS were investigated under varying hydration levels. Figure 4.4 a) shows the time-averaged system densities as a function of the water content x_{water}. In all cases, the densities decreased with increasing water content, reflecting the more compact packing of SOM structures under dry conditions. A comparison across the compositions revealed that systems with higher acidity, i.e., a larger abundance of carboxylate groups, reached higher densities at comparable hydration levels. Specifically,

the order from high to low acidity is SOM I (blue), SOM IV (red), and SOM II (orange). This trend is consistent with stronger calcium–carboxylate interactions, which are assumed to occur more prominently at higher acidity levels and represent a key driving force behind the compactification of the SOM structures.

This interpretation is further supported by the energetic contributions of the SOM systems. Figure 4.4 b) shows the Coulomb energies, which arise from interactions between carboxylate groups and calcium ions. Although the concentration of these groups remains constant within a given SOM type, the Coulomb energies decrease with lower water contents, indicating shorter average distances and thus a tighter packing of charged groups. This enhanced electrostatic stabilization promotes aggregation under dry conditions and explains the higher densities observed, which is consistent with atomistic studies of HS systems (Galicia-Andrés et al., 2021a; Borisover et al., 2025). At the same time, repulsive contributions counterbalance the attractive forces, leading to a saturation of the densities at lower hydration levels, which is also reflected in the other energetic components reported in the original study (Dettmann et al., 2024a).

Overall, these results demonstrate that the coarse-grained framework is capable of capturing the general trends regarding the compactification behavior of diverse SOM types, thereby providing a proof of principle for comparative studies across chemically distinct compositions. Beyond the density and energetic analyses, the simulations also revealed a demixing process at high water contents for the SOM II system, where the HS and water phases became increasingly separated over time. Therefore, the computational performance of the coarse-grained models offered a decisive advantage. Using benchmark calculations, the performance could be extended by two orders of magnitude compared to atomistic simulations, enabling the investigation of processes such as aggregation and phase behavior on time scales that would otherwise be inaccessible.

4.4 Pollutant Mobility in SOM Systems

The developed coarse-grained SOM modeling approach made it possible to extend the investigations to the interaction of chemical compounds with the SOM matrix. In the context of molecular-scale interactions of pollutants with SOM, HCB and SAA were selected as representative cases of a hydrophobic POP and a polar pharmaceutical pollutant, respectively. Their distinct chemical properties make them suitable model systems for studying how pollutant polarity and hydrophobicity influence partitioning and mobility within SOM. For this purpose, five SOM systems of different compositions were investigated under high hydration conditions, corresponding to a water fraction of $x_{\text{water}} \approx 0.567$. This choice was guided by the density results discussed previously (Figure 4.2 a)), where the coarse-grained models showed improved accuracy at higher hydration levels.

Analogous to the VSOMM fragments, the two pollutants were parametrized using Swarm-CG for the bonded interactions, while the nonbonded bead types were chosen

by matching experimental transfer free energies. In addition, the SASA was used as a further reference with which to validate the molecular shape. The outcomes are depicted in Figure 4.5, which shows the applied mapping and assigned Martini beads for SAA (a) and HCB (b). In the case of SAA, the sulfonamide group was represented by the strongly polar P5 bead, while the halogen groups of HCB were best described by the highly hydrophobic X1 bead, consistent with their chemical character. Panels c) and d) illustrate the Connolly surfaces of the atomistic (cyan) and coarse-grained (orange) models, demonstrating good agreement in shape and SASA. Further details of the parametrization procedure, including the numerical SASA and ΔG values, are provided in the original study (Dettmann et al., 2025a). In summary, a suitable Martini representation of both pollutants was determined, reproducing the chemical character and molecular shape with sufficient accuracy to enable the subsequent investigation of their interactions with coarse-grained SOM systems.

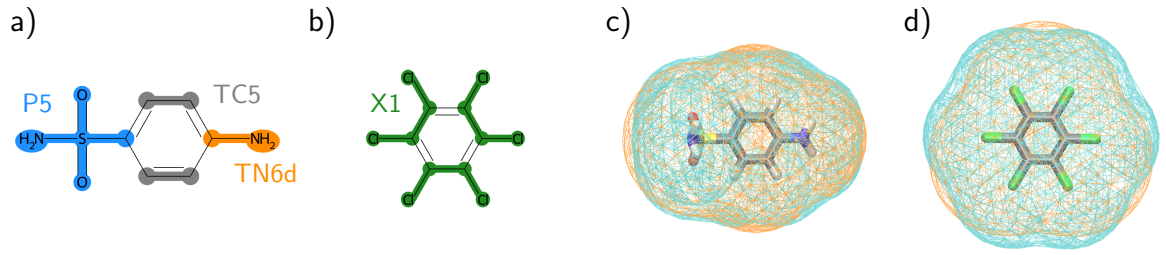


Figure 4.5: Panels a) and b) show the mapping and assigned Martini bead types for SAA and HCB after parametrization, respectively. Panels c) and d) depict the Connolly surfaces of the atomistic (cyan) and coarse-grained (orange) models, illustrating the agreement in molecular shape (Dettmann et al., 2025a).

The mobility of the pollutants was assessed by calculating diffusion coefficients from the MSDs according to Equation 3.3. The results are shown in Figure 4.6 a) for SAA (orange) and HCB (green) across the different SOM compositions, with the diffusion coefficients of the water beads (blue) provided as an additional reference. The diffusion behavior of SAA closely followed that of water in all systems, consistent with a coordination number analysis indicating that SAA is predominantly located in the aqueous phase (see Figure 3 in the publication). Accordingly, water diffusion has the strongest influence on SAA mobility. In contrast, HCB displayed markedly lower diffusion coefficients in almost every case. In SOM II and SOM III, which are characterized by low acidity and high aromaticity, respectively, the diffusion coefficient of HCB is particularly low compared to that of water in the corresponding systems. This reduced mobility reflects the tendency of HCB to accumulate in SOM, driven by its strong partitioning into the hydrophobic matrix and direct interactions with SOM functional groups (Ahmed

et al., 2014). Although Martini coarse-graining is known to accelerate the dynamics artificially, qualitative comparisons of diffusion coefficients can still provide useful insights, as discussed in Section 3.2 of Dettmann et al. (2025a).

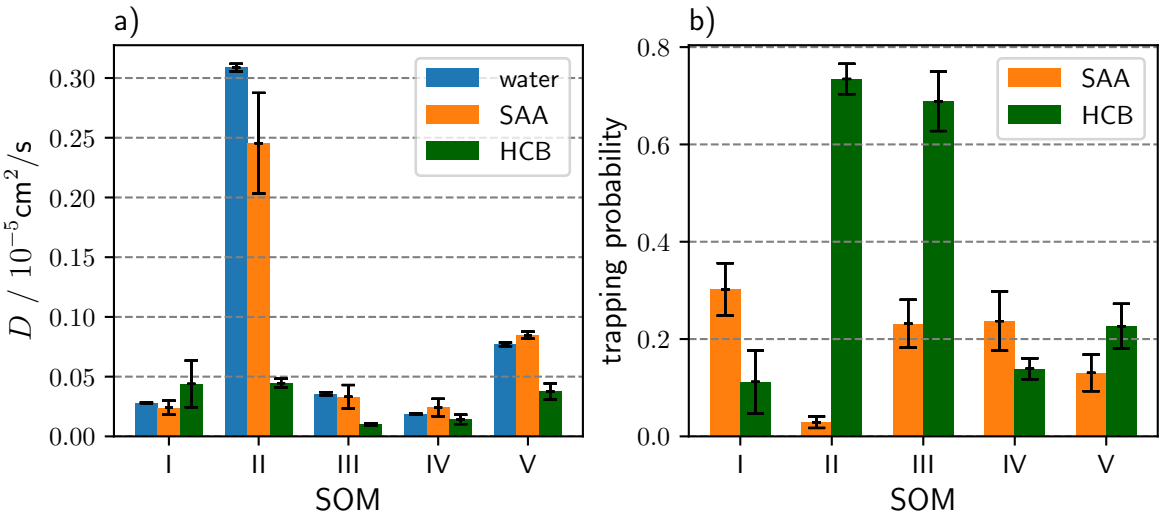


Figure 4.6: Panel a) shows the diffusion coefficients of water beads (blue), SAA (orange), and HCB (green) for the different SOM compositions. Panel b) depicts the trapping probabilities of SAA (orange) and HCB (green) in the same systems. Both quantities are average values from the 20 individual trajectories for each pollutant and SOM system (Dettmann et al., 2025a).

Another property used to characterize the influence of SOM composition on pollutant mobility was the analysis of trajectory intervals in which the pollutant molecules were trapped within the SOM matrix. These intervals were identified using a hidden Markov model (Rabiner, 1989) that distinguished distinct mobility states based on the spatial displacements of the molecules (for details, see the Supporting Information of Dettmann et al. (2025a)). The ratio of trapped intervals to the total simulation time, referred to as the trapping probability, is presented in Figure 4.6 b) for both pollutants across the five SOM systems investigated.

In the case of SAA, the analysis revealed a strong anticorrelation with water diffusion, as confirmed by a Pearson correlation coefficient of $r = -0.90$. This finding indicates that SAA mobility is largely governed by the dynamics of the aqueous phase: higher water diffusion reduces the likelihood of SAA being trapped within the SOM matrix. In contrast, the trapping probability of HCB correlates strongly with the coordination numbers of the HS molecules ($r = 0.97$), reflecting the tendency of HCB to accumulate in the HS phase. Accordingly, direct interactions with HS functional groups, particularly in SOM II and SOM III, exert the strongest influence on HCB trapping.

More detailed analyses highlighting the role of specific functional groups in the

SOM–pollutant interactions can be found in Dettmann et al. (2025a). Overall, the results provide molecular-level insights into the trajectory-resolved interactions of SAA and HCB, which are difficult to access experimentally. Beyond the systems studied here, the approach offers a transferable framework for investigating other pollutants, as well as their interactions with diverse soil constituents.

4.5 SOM Heteroaggregation with NP

NPs represent an emerging environmental risk with largely unknown behavior in soils. To investigate their molecular-scale interactions with SOM, two complementary approaches were employed. First, single polymer chains were embedded into SOM matrices of different acidity to examine their partitioning, dominant interaction sites, and mobility. One of the most important findings was that the polarity of the polymers had a decisive influence on their diffusion and partitioning within the SOM systems. PEO showed increased mobility, interacting strongly with water and polar SOM functional groups. In contrast, PE displayed acidity-dependent partitioning, including trapping in water-filled voids under acidic conditions and binding to hydrophobic moieties under less acidic conditions, with low diffusivity overall. For further details about this part of the investigation, please refer to Dettmann et al. (2025b).

Second, the association between a PE nanoparticle and an SOM cluster in aqueous solution was studied under different pH conditions. The latter setup enabled the direct analysis of NP–SOM heteroaggregation. Figure 4.7 a) – d) shows representative end configurations of the trajectories for the PE nanoparticle (blue) and the SOM cluster (orange) at pH values from 5 to 8. In all cases, adsorption of the SOM cluster onto the nanoparticle surface was observed, and the cluster remained attached throughout the entire simulation time. The extent of SOM dilution varied with the pH. At the lowest pH (Figure 4.7 a)), the cluster exhibited a pronounced degree of dispersion, whereas at higher pH levels the SOM structure remained more compact. This trend was quantified by a cluster size analysis, as presented in Figure 10 in Dettmann et al. (2025b).

In Figure 4.7 e), the coverage of the PE nanoparticle surface is shown as a function of pH, expressed as the fraction of surface beads in contact with HS molecules within a defined cutoff. The coverage is separated into contributions from the largest remaining cluster and the dissolved HS molecules, which persist after partial dilution of the original aggregate. At a low pH, both the largest cluster and the dissolved molecules contribute substantially to the coverage, leading to high overall values. With an increasing pH, the coverage decreases and eventually approaches a saturated level.

The observed coverage behavior can be rationalized by considering the structural response of the HS aggregates. At a low pH, the largest cluster undergoes deformation and spreads more extensively across the PE surface, as indicated by a sphericity analysis, which increases its contact area (please refer to the Supporting Information of the corresponding study). At the same time, the higher number of dissolved HS molecules

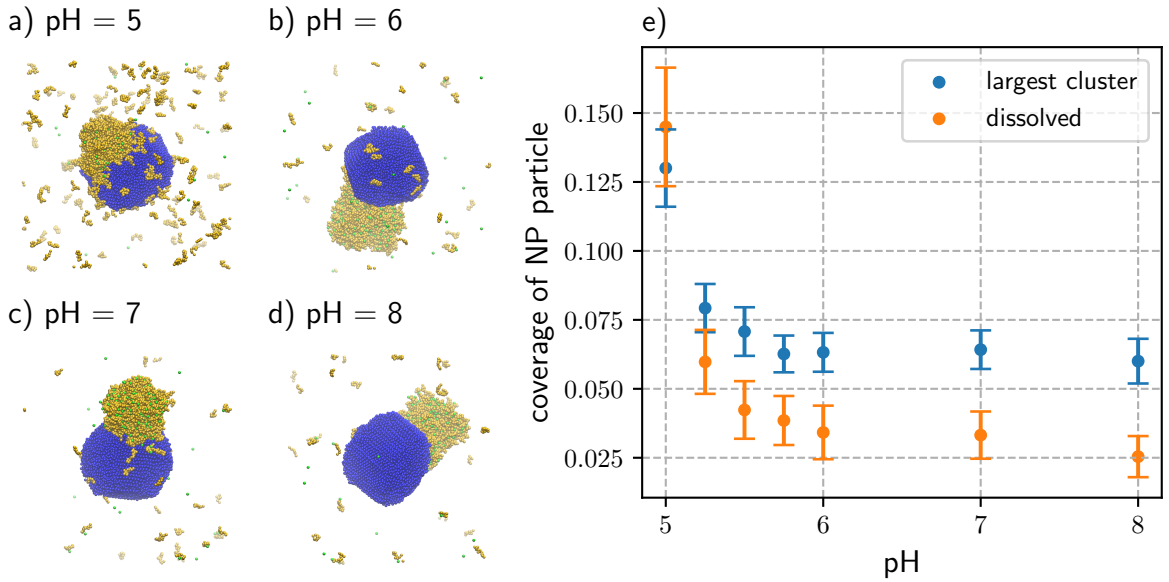


Figure 4.7: Panels a) – d) show snapshots of SOM–NP associations at different pH values (PE particle in blue, HS molecules in orange, calcium ions in green; water omitted for clarity), with the PE particle exhibiting a diameter of roughly 10 nm. Panel e) depicts the coverage of the NP surface by HS molecules as a function of pH, separated into contributions from the largest cluster (blue) and dissolved HS molecules (orange) (Dettmann et al., 2025b).

present in the solution, resulting from the greater dilution under low pH conditions, further contributes to the coverage. However, the concentration of dissolved HS molecules is inherently dependent on the system design and could itself be regarded as a separate environmental variable, which is why the coverage was differentiated into cluster and dissolved contributions for clearer interpretation.

Further interaction analyses presented in Dettmann et al. (2025b) suggested that the observed dilution is linked to the protonation state of carboxylate groups, which influences their stabilizing interactions with calcium ions (see Figure 12 of the study). Overall, the results indicate that a stronger NP–SOM association can be expected at a lower pH, although additional factors such as nanoparticle surface modification, ionic strength of the solution, and an alternative model representation of the pH would be necessary to refine this picture. However, the current work demonstrates the capability of coarse-grained simulations to capture NP–SOM interactions over extended timescales and larger system sizes, providing a basis for future studies of environmental processes that remain challenging to investigate experimentally.

5 Summary and Conclusions

SOM plays a central role in terrestrial ecosystems, yet its molecular-scale interactions remain difficult to probe experimentally due to its inherent complexity. At the same time, pollutants such as POPs, pharmaceuticals, and NPs pose increasing environmental risks, while their behavior in SOM is still poorly understood. Molecular simulations provide a complementary research approach that can be employed to address these challenges. By applying a coarse-graining strategy, the accessible time and length scales of simulations can be extended, allowing the study of larger systems and long-term dynamics, while also simplifying structural features of SOM. Developing a coarse-grained SOM framework based on VSOMM2 structures, with the aim of enhancing the understanding of pollutant mobility and fate in SOM therefore constituted the central objective of this work.

Starting from the generation of a coarse-grained VSOMM fragment database, interaction parameters were derived and methods developed to assemble the fragments into SOM molecules, following the philosophy of the VSOMM2. The parametrization was carried out within the flexible Martini 3 force field, and the resulting SOM models were benchmarked against atomistic simulations to evaluate their strengths and limitations. An automated algorithm was developed to generate coarse-grained SOM topologies directly from VSOMM2 structures, facilitating the construction of systems with diverse compositions. Finally, the models were applied to study the interactions of SOM with distinct anthropogenic pollutants.

The following key findings about the development of the coarse-grained models as well as the insights into pollutant behavior with SOM emerged from this work:

- The Swarm-CG tool was employed to determine the bonded interaction parameters of the coarse-grained fragments. Both the R_g and the SASA showed good agreement with the atomistic references, indicating that molecular shape and size were reproduced sufficiently accurately. As a measure of thermodynamic properties, the transfer free energies were also captured effectively, particularly in the hexadecane–water and chloroform–water cases.

- Coarse-grained models of LHA were constructed at varying water contents and compared to atomistic reference systems. The assembly of fragments was achieved using the DBI method. The resulting densities of the coarse-grained models showed an acceptable level of agreement with the atomistic systems, especially under higher hydration conditions.

- The RDFs between the carboxylate groups and calcium ions were calculated, re-

vealing that short-range interactions were not reproduced in the coarse-grained models. Analyses of the atomistic systems showed the presence of distinct calcium–carboxylate complexes, which explained the density discrepancies between atomistic and coarse-grained systems.

- Revising the force constants of bonded interactions reduced the occurrence of high vibrational frequencies in the fragment molecules, as shown by the power spectra. This improvement enhanced simulation stability and allowed for larger time steps, resulting in a performance increase of about two orders of magnitude compared to the atomistic systems. The TMS indicated that the reparametrization did not significantly influence the agreement between the bonded distributions of atomistic and coarse-grained fragments.

- The automated parametrization approach demonstrated its capability by enabling the generation and analysis of coarse-grained SOM models with varying compositions. Calculations of the potential energies of nonbonded interactions showed that Coulomb interactions acted as the main driving force behind SOM packing, with the strength of these interactions being connected to the acidity of the SOM systems.

- The developed models were applied to address environmentally relevant processes, beginning with an assessment of the mobility of the pollutants SAA and HCB in SOM systems. Determination of diffusion coefficients revealed that the SAA diffusion closely followed that of the water beads, whereas HCB exhibited markedly reduced mobility, except in systems with increased carboxylate content. Additional analysis using a hidden Markov model resolved distinct trapping states, differentiating between fixation within the SOM and time intervals of free diffusion.

- Using the same investigation strategy focusing on SOM compositions with high and low acidity, the behavior of PEO and PE polymers within SOM matrices was likewise examined. PEO exhibited elevated mobility, remaining predominantly in the water phase. In contrast, PE showed distinct partitioning behavior depending on the SOM acidity, with overall reduced mobility connected to either entrapment within water-filled SOM voids or enhanced interactions with hydrophobic regions of SOM.

- Furthermore, the association between SOM and a PE nanoparticle was examined under different pH levels. At a lower pH, reduced interactions between carboxylate groups and calcium ions destabilized SOM aggregates, promoting their dispersion and leading to increased coverage of the nanoparticle surface.

Regarding future research, several directions for further development of the modeling approach emerged. Refining the force constants for bonds and angles by employing all-atom representations of the reference molecules could improve the parameter transferability

between Swarm-CG and ETKDG. In the long term, the construction of a comprehensive database of equilibrium and force constants would allow the parametrization step to be bypassed entirely. Beyond methodological refinements, the framework offers broad potential for future applications. It enables studies of diverse SOM compositions and can be readily extended to incorporate biologically relevant macromolecules within the Martini 3 framework. Furthermore, the integration of additional pre-parametrized pollutants as well as the investigation of large-scale SOM structural organization would considerably expand the scope of applications. Environmental relevance could be further enhanced by including aged or surface-functionalized NPs with polar or charged groups, other polymer types, and scenarios involving co-adsorption with pollutants. Finally, the influence of ionic strength and ion identity represents another important factor that could be explored to capture the complexity of natural soil environments more realistically. Altogether, these perspectives highlight the potential of the developed framework to act as a bridge between molecular modeling, experimental studies and environmental research by providing mechanistic insights into SOM processes and their interactions with anthropogenic pollutants.

Bibliography

Abraham, M., Alekseenko, A., Andrews, B., Basov, V., Bauer, P., Bird, H., et al. (2025). GROMACS 2025.2 Manual. doi:10.5281/zenodo.15387070

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1-2, 19–25. doi:10.1016/j.softx.2015.06.001

Aeschbacher, M., Sander, M., and Schwarzenbach, R. P. (2009). Novel Electrochemical Approach to Assess the Redox Properties of Humic Substances. *Environmental Science & Technology* 44, 87–93. doi:10.1021/es902627p

Ahmed, A. A., Kühn, O., and Leinweber, P. (2012). Controlled experimental soil organic matter modification for study of organic pollutant interactions in soil. *Science of The Total Environment* 441, 151–158. doi:10.1016/j.scitotenv.2012.09.042

Ahmed, A. A., Kühn, O., Aziz, S. G., Hilal, R. H., and Leinweber, P. (2014). How soil organic matter composition controls hexachlorobenzene–soil-interactions: Adsorption isotherms and quantum chemical modeling. *Science of The Total Environment* 476-477, 98–106. doi:10.1016/j.scitotenv.2013.12.096

Ahmed, A. A., Leinweber, P., and Kühn, O. (2018). Unravelling the nature of glyphosate binding to goethite surfaces by *ab initio* molecular dynamics simulations. *Physical Chemistry Chemical Physics* 20, 1531–1539. doi:10.1039/c7cp06245a

Ahmed, A. A., Thiele-Bruhn, S., Aziz, S. G., Hilal, R. H., Elroby, S. A., Al-Youbi, A. O., et al. (2015). Interaction of polar and nonpolar organic pollutants with soil organic matter: Sorption experiments and molecular dynamics simulation. *Science of The Total Environment* 508, 276–287. doi:10.1016/j.scitotenv.2014.11.087

Albers, C. N., Banta, G. T., Jacobsen, O. S., and Hansen, P. E. (2008). Characterization and structural modelling of humic substances in field soil displaying significant differences from previously proposed structures. *European Journal of Soil Science* 59, 693–705. doi:10.1111/j.1365-2389.2008.01036.x

Alessandri, R., Barnoud, J., Gertsen, A. S., Patmanidis, I., de Vries, A. H., Souza, P. C. T., et al. (2022). Martini 3 Coarse-Grained Force Field: Small Molecules. *Advanced Theory and Simulations* 5, 2100391. doi:10.1002/adts.202100391

Alessandri, R., Souza, P. C. T., Thallmair, S., Melo, M. N., de Vries, A. H., and Marrink, S. J. (2019). Pitfalls of the Martini Model. *Journal of Chemical Theory and Computation* 15, 5448–5460. doi:10.1021/acs.jctc.9b00473

Alessandri, R., Thallmair, S., Herrero, C. G., Mera-Adasme, R., Marrink, S. J., and Souza, P. C. T. (2023). A Practical Introduction to Martini 3 and its Application to Protein-Ligand Binding Simulations. In *A Practical Guide to Recent Advances in Multiscale Modeling and Simulation of Biomolecules* (AIP Publishing LLC), chap. 1. 1–34. doi:10.1063/9780735425279_001

Allen, M. P. and Tildesley, D. J. (2017). *Computer Simulation of Liquids* (Oxford University Press), second edn. doi:10.1093/oso/9780198803195.001.0001

Alvarez-Puebla, R., Valenzuela-Calahorro, C., and Garrido, J. (2006). Theoretical study on fulvic acid structure, conformation and aggregation. *Science of The Total Environment* 358, 243–254. doi:10.1016/j.scitotenv.2004.11.026

Aquino, A. J., Tunega, D., Pašalić, H., Schaumann, G. E., Haberhauer, G., Gerzabek, M. H., et al. (2011). Study of solvent effect on the stability of water bridge-linked carboxyl groups in humic acid models. *Geoderma* 169, 20–26. doi:10.1016/j.geoderma.2010.12.006

Aquino, A. J. A., Tunega, D., Haberhauer, G., Gerzabek, M. H., and Lischka, H. (2007). Interaction of the 2, 4-dichlorophenoxyacetic acid herbicide with soil organic matter moieties: a theoretical study. *European Journal of Soil Science* 58, 889–899. doi:10.1111/j.1365-2389.2007.00928.x

Baho, D. L., Bundschuh, M., and Futter, M. N. (2021). Microplastics in terrestrial ecosystems: Moving beyond the state of the art to minimize the risk of ecological surprise. *Global Change Biology* 27, 3969–3986. doi:10.1111/gcb.15724

Bennett, C. H. (1976). Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics* 22, 245–268. doi:10.1016/0021-9991(76)90078-4

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* 81, 3684–3690. doi:10.1063/1.448118

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., and Hermans, J. (1981). *Interaction Models for Water in Relation to Protein Hydration* (Springer Netherlands), chap. 21. 331–342. doi:10.1007/978-94-015-7658-1_21

Bernetti, M. and Bussi, G. (2020). Pressure control using stochastic cell rescaling. *The Journal of Chemical Physics* 153, 114107. doi:10.1063/5.0020514

Birkhoff, G. D. (1931). Proof of the Ergodic Theorem. *Proceedings of the National Academy of Sciences* 17, 656–660. doi:10.1073/pnas.17.2.656

Blaney, J. M. and Dixon, J. S. (1994). *Distance Geometry in Molecular Modeling* (John Wiley & Sons, Ltd), chap. 6. 299–335. doi:10.1002/9780470125823.ch6

Blume, H.-P., Brümmer, G. W., Fleige, H., Horn, R., Kandeler, E., Kögel-Knabner, I., et al. (2016a). *Soil Organic Matter* (Springer Berlin Heidelberg), chap. 3. 55–86. doi:10.1007/978-3-642-30942-7_3

Blume, H.-P., Brümmer, G. W., Fleige, H., Horn, R., Kandeler, E., Kögel-Knabner, I., et al. (2016b). *Threats to the Soil Functions* (Springer Berlin Heidelberg), chap. 10. 485–559. doi:10.1007/978-3-642-30942-7_10

Borisover, M., Petrov, D., Oostenbrink, C., and Galicia-Andrés, E. (2025). Diluting humic substances in water in molecular dynamics simulations: Are aggregates stable? *Colloids and Surfaces A: Physicochemical and Engineering Aspects* 704, 135507. doi:10.1016/j.colsurfa.2024.135507

Born, M. and Oppenheimer, R. (1927). Zur Quantentheorie der Molekeln. *Annalen der Physik* 389, 457–484. doi:10.1002/andp.19273892002

Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., et al. (2009). CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* 30, 1545–1614. doi:10.1002/jcc.21287

Bruckner, S. and Boresch, S. (2011). Efficiency of alchemical free energy simulations. I. A practical comparison of the exponential formula, thermodynamic integration, and Bennett's acceptance ratio method. *Journal of Computational Chemistry* 32, 1303–1319. doi:10.1002/jcc.21713

Buckingham, R. A. (1938). The classical equation of state of gaseous helium, neon and argon. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 168, 264–283. doi:10.1098/rspa.1938.0173

Bulacu, M., Goga, N., Zhao, W., Rossi, G., Monticelli, L., Periole, X., et al. (2013). Improved Angle Potentials for Coarse-Grained Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* 9, 3282–3292. doi:10.1021/ct400219n

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* 126. doi:10.1063/1.2408420

Car, R. and Parrinello, M. (1985). Unified Approach for Molecular Dynamics and Density-Functional Theory. *Physical Review Letters* 55, 2471–2474. doi:10.1103/physrevlett.55.2471

Chase, M. W. J. (ed.) (1998). *NIST-JANAF Thermochemical Tables*. Journal of Physical and Chemical Reference Data, Monograph 9 (American Institute of Physics), fourth edn.

Chefetz, B., Tarchitzky, J., Deshmukh, A. P., Hatcher, P. G., and Chen, Y. (2002). Structural Characterization of Soil Organic Matter and Humic Acids in Particle-Size Fractions of an Agricultural Soil. *Soil Science Society of America Journal* 66, 129–141. doi:10.2136/sssaj2002.1290

Chen, C., Leinweber, P., Eckhardt, K.-U., and Sparks, D. (2018). The Composition and Stability of Clay-Associated Organic Matter along a Soil Profile. *Soil Systems* 2, 16. doi:10.3390/soilsystems2010016

Chen, X. D., Dunfield, K. E., Fraser, T. D., Wakelin, S. A., Richardson, A. E., and Condron, L. M. (2019). Soil biodiversity and biogeochemical function in managed ecosystems. *Soil Research* 58, 1–20. doi:10.1071/sr19067

Cycoń, M., Mrozik, A., and Piotrowska-Seget, Z. (2019). Antibiotics in the Soil Environment—Degradation and Their Impact on Microbial Activity and Diversity. *Frontiers in Microbiology* 10. doi:10.3389/fmicb.2019.00338

de Jong, D. H., Baoukina, S., Ingólfsson, H. I., and Marrink, S. J. (2016). Martini straight: Boosting performance using a shorter cutoff and GPUs. *Computer Physics Communications* 199, 1–7. doi:10.1016/j.cpc.2015.09.014

de Ruiter, A., Boresch, S., and Oostenbrink, C. (2013). Comparison of thermodynamic integration and Bennett acceptance ratio for calculating relative protein-ligand binding free energies. *Journal of Computational Chemistry* 34, 1024–1034. doi:10.1002/jcc.23229

Dettmann, L. F. (2025). SOM2CG: Coarse-graining soil-organic-matter models. `https://github.com/LorenzDettmann/SOM2CG`. GitHub repository, accessed 8 Aug 2025

Dettmann, L. F., Kühn, O., and Ahmed, A. A. (2021). Coarse-grained molecular dynamics simulations of nanoplastics interacting with a hydrophobic environment in aqueous solution. *RSC Advances* 11, 27734–27744. doi:10.1039/d1ra04439g

Dettmann, L. F., Kühn, O., and Ahmed, A. A. (2024a). Automated Parametrization Approach for Coarse-Graining Soil Organic Matter Molecules. *Journal of Chemical Theory and Computation* 20, 10684–10696. doi:10.1021/acs.jctc.4c01334. Reprinted with permission. Copyright 2024 American Chemical Society

Dettmann, L. F., Kühn, O., and Ahmed, A. A. (2024b). Martini-Based Coarse-Grained Soil Organic Matter Model Derived from Atomistic Simulations. *Journal of Chemical Theory and Computation* 20, 5291–5305. doi:10.1021/acs.jctc.4c00332. Reprinted with permission. Copyright 2024 American Chemical Society

Dettmann, L. F., Kühn, O., and Ahmed, A. A. (2025a). Coarse-grained Simulations of Sulfanilamide and Hexachlorobenzene Mobility in Soil Organic Matter. *Environmental Science: Advances* doi:10.1039/d5va00237k. Reprinted with permission from the RSC

Dettmann, L. F., Kühn, O., and Ahmed, A. A. (2025b). Coarse-Grained Simulations of the Nanoplastic Interaction with Soil Organic Matter. *The Journal of Physical Chemistry B* 129, 11605–11615. doi:10.1021/acs.jpcb.5c06072. Reprinted with permission. Copyright 2025 American Chemical Society

Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik* 322, 549–560. doi:10.1002/andp.19053220806

Eisenhaber, F., Lijnzaad, P., Argos, P., Sander, C., and Scharf, M. (1995). The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *Journal of Computational Chemistry* 16, 273–284. doi:10.1002/jcc.540160303

Empereur-Mot, C., Pesce, L., Doni, G., Bochicchio, D., Capelli, R., Perego, C., et al. (2020). *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* 5, 32823–32843. doi:10.1021/acsomega.0c05469

Escalona, Y., Petrov, D., Galicia-Andrés, E., and Oostenbrink, C. (2023). Exploring the Macroscopic Properties of Humic Substances Using Modeling and Molecular Simulations. *Agronomy* 13, 1044. doi:10.3390/agronomy13041044

Escalona, Y., Petrov, D., and Oostenbrink, C. (2021). Vienna soil organic matter modeler 2 (VSOMM2). *Journal of Molecular Graphics and Modelling* 103, 107817. doi:10.1016/j.jmgm.2020.107817

Galicia-Andrés, E., Escalona, Y., Oostenbrink, C., Tunega, D., and Gerzabek, M. H. (2021a). Soil organic matter stabilization at molecular scale: The role of metal cations and hydrogen bonds. *Geoderma* 401, 115237. doi:10.1016/j.geoderma.2021.115237

Galicia-Andrés, E., Oostenbrink, C., Gerzabek, M. H., and Tunega, D. (2021b). On the Adsorption Mechanism of Humic Substances on Kaolinite and Their Microscopic Structure. *Minerals* 11, 1138. doi:10.3390/min11101138

Ganta, P. B. (2023). *Revealing the binding mechanism of phosphates at the mineral-water interface: insights from QM/MM molecular dynamics simulations*. Ph.D. thesis, University of Rostock, Rostock, Germany. doi:10.18453/rosdok_id00004870

Gerke, J. (2018). Concepts and Misconceptions of Humic Substances as the Stable Part of Soil Organic Matter: A Review. *Agronomy* 8, 76. doi:10.3390/agronomy8050076

Gerlach, T., Müller, S., de Castilla, A. G., and Smirnova, I. (2022). An open source COSMO-RS implementation and parameterization supporting the efficient implementation of multiple segment descriptors. *Fluid Phase Equilibria* 560, 113472. doi:10.1016/j.fluid.2022.113472

Gerzabek, M. H., Aquino, A. J. A., Balboa, Y. I. E., Galicia-Andrés, E., Grančič, P., Oostenbrink, C., et al. (2022). A contribution of molecular modeling to supramolecular structures in soil organic matter. *Journal of Plant Nutrition and Soil Science* 185, 44–59. doi:10.1002/jpln.202100360

Geyer, R., Jambeck, J. R., and Law, K. L. (2017). Production, use, and fate of all plastics ever made. *Science Advances* 3, e1700782. doi:10.1126/sciadv.1700782

González-Pérez, J. A., González-Vila, F. J., Almendros, G., and Knicker, H. (2004). The effect of fire on soil organic matter—a review. *Environment International* 30, 855–870. doi:10.1016/j.envint.2004.02.003

Gotsmy, M., Escalona, Y., Oostenbrink, C., and Petrov, D. (2021). Exploring the structure and dynamics of proteins in soil organic matter. *Proteins: Structure, Function, and Bioinformatics* 89, 925–936. doi:10.1002/prot.26070

Gowers, R., Linke, M., Barnoud, J., Reddy, T., Melo, M., Seyler, S., et al. (2016). MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In *Proceedings of the 15th Python in Science Conference* (SciPy), 98–105. doi:10.25080/majora-629e541a-00e

Gros, P., Ahmed, A., Kühn, O., and Leinweber, P. (2017). Glyphosate binding in soil as revealed by sorption experiments and quantum-chemical modeling. *Science of The Total Environment* 586, 527–535. doi:10.1016/j.scitotenv.2017.02.007

Grünewald, F., Alessandri, R., Kroon, P. C., Monticelli, L., Souza, P. C. T., and Marrink, S. J. (2022). Polyply; a python suite for facilitating simulations of macromolecules and nanomaterials. *Nature Communications* 13. doi:10.1038/s41467-021-27627-4

Grünewald, F., Punt, M. H., Jefferys, E. E., Vainikka, P. A., König, M., Virtanen, V., et al. (2022). Martini 3 Coarse-Grained Force Field for Carbohydrates. *Journal of Chemical Theory and Computation* 18, 7555–7569. doi:10.1021/acs.jctc.2c00757

Grünewald, F., Souza, P. C. T., Abdizadeh, H., Barnoud, J., de Vries, A. H., and Marrink, S. J. (2020). Titratable Martini model for constant pH simulations. *The Journal of Chemical Physics* 153, 024118. doi:10.1063/5.0014258

Guo, J.-J., Huang, X.-P., Xiang, L., Wang, Y.-Z., Li, Y.-W., Li, H., et al. (2020). Source, migration and toxicology of microplastics in soil. *Environment International* 137, 105263. doi:10.1016/j.envint.2019.105263

Gworek, B., Kijeńska, M., Wrzosek, J., and Graniewska, M. (2021). Pharmaceuticals in the Soil and Plant Environment: a Review. *Water, Air, & Soil Pollution* 232. doi:10.1007/s11270-020-04954-8

Hanif, M. N., Aijaz, N., Azam, K., Akhtar, M., Laftah, W. A., Babur, M., et al. (2024). Impact of microplastics on soil (physical and chemical) properties, soil biological properties/soil biota, and response of plants to it: a review. *International Journal of Environmental Science and Technology* 21, 10277–10318. doi:10.1007/s13762-024-05656-y

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., et al. (2020). Array programming with NumPy. *Nature* 585, 357–362. doi:10.1038/s41586-020-2649-2

Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997). LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* 18, 1463–1472. doi:10.1002/(sici)1096-987x(199709)18:12⟨1463::aid-jcc4⟩3.0.co;2-h

Hockney, R. and Eastwood, J. (1988). *Computer Simulation Using Particles* (CRC Press), first edn. doi:10.1201/9780367806934

Hohenberg, P. and Kohn, W. (1964). Inhomogeneous Electron Gas. *Physical Review* 136, B864–B871. doi:10.1103/physrev.136.b864

Hou, D., Bolan, N. S., Tsang, D. C., Kirkham, M. B., and O'Connor, D. (2020). Sustainable soil use and management: An interdisciplinary and systematic approach. *Science of The Total Environment* 729, 138961. doi:10.1016/j.scitotenv.2020.138961

Hudson, B. D. (1994). Soil organic matter and available water capacity. *Journal of Soil and Water Conservation* 49, 189–194. doi:10.1080/00224561.1994.12456850

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 14, 33–38. doi:10.1016/0263-7855(96)00018-5

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* 9, 90–95. doi:10.1109/mcse.2007.55

IHSS (2025). International Humic Substances Society – Official Website. `https://humic-substances.org`. Accessed: July 22, 2025

Ivleva, N. P. (2021). Chemical Analysis of Microplastics and Nanoplastics: Challenges, Advanced Methods, and Perspectives. *Chemical Reviews* 121, 11886–11936. doi:10.1021/acs.chemrev.1c00178

Jackson, R. B., Lajtha, K., Crow, S. E., Hugelius, G., Kramer, M. G., and Piñeiro, G. (2017). The Ecology of Soil Carbon: Pools, Vulnerabilities, and Biotic and Abiotic Controls. *Annual Review of Ecology, Evolution, and Systematics* 48, 419–445. doi:10.1146/annurev-ecolsys-112414-054234

Jarin, Z., Newhouse, J., and Voth, G. A. (2021). Coarse-Grained Force Fields from the Perspective of Statistical Mechanics: Better Understanding of the Origins of a MARTINI Hangover. *Journal of Chemical Theory and Computation* 17, 1170–1180. doi:10.1021/acs.jctc.0c00638

Jin, J., Pak, A. J., Durumeric, A. E. P., Loose, T. D., and Voth, G. A. (2022). Bottom-up Coarse-Graining: Principles and Perspectives. *Journal of Chemical Theory and Computation* 18, 5759–5791. doi:10.1021/acs.jctc.2c00643

Jones, K. and de Voogt, P. (1999). Persistent organic pollutants (POPs): state of the science. *Environmental Pollution* 100, 209–221. doi:10.1016/s0269-7491(99)00098-6

Junghans, C., Andrienko, D., Bot, V., Rosenberger, D., Krajniak, J., marvinbernhardt, et al. (2021). *votca/csg-manual: v1.6.4*. doi:10.5281/zenodo.4435216

Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*. vol. 4, 1942–1948. doi:10.1109/icnn.1995.488968

Kerisit, S., Bylaska, E. J., Massey, M. S., McBriarty, M. E., and Ilton, E. S. (2016). Ab Initio Molecular Dynamics of Uranium Incorporated in Goethite (α-FeOOH): Interpretation of X-ray Absorption Spectroscopy of Trace Polyvalent Metals. *Inorganic Chemistry* 55, 11736–11746. doi:10.1021/acs.inorgchem.6b01773

Kibblewhite, M., Ritz, K., and Swift, M. (2007). Soil health in agricultural systems. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363, 685–701. doi:10.1098/rstb.2007.2178

Klamt, A. (1995). Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *The Journal of Physical Chemistry* 99, 2224–2235. doi:10.1021/j100007a062

Klamt, A., Eckert, F., and Arlt, W. (2010). COSMO-RS: An Alternative to Simulation for Calculating Thermodynamic Properties of Liquid Mixtures. *Annual Review of Chemical and Biomolecular Engineering* 1, 101–122. doi:10.1146/annurev-chembioeng-073009-100903

Kleber, M. and Lehmann, J. (2019). Humic Substances Extracted by Alkali Are Invalid Proxies for the Dynamics and Functions of Organic Matter in Terrestrial and Aquatic Ecosystems. *Journal of Environmental Quality* 48, 207–216. doi:10.2134/jeq2019.01.0036

Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E., and Kolinski, A. (2016). Coarse-Grained Protein Models and Their Applications. *Chemical Reviews* 116, 7898–7936. doi:10.1021/acs.chemrev.6b00163

Kubicki, J. D. (2006). Molecular Simulations of Benzene and PAH Interactions with Soot. *Environmental Science & Technology* 40, 2298–2303. doi:10.1021/es051083s

Kubicki, J. D. (ed.) (2016). *Molecular Modeling of Geochemical Reactions: An Introduction* (John Wiley & Sons, Ltd). doi:10.1002/9781118845226

Landrum, G., Tosco, P., Kelley, B., Ric, Cosgrove, D., sriniker, et al. (2023). rdkit/rdkit: 2023_03_3 (Q1 2023) Release. doi:10.5281/zenodo.8254217

Lee, B. and Richards, F. (1971). The interpretation of protein structures: Estimation of static accessibility. *Journal of Molecular Biology* 55, 379–400. doi:10.1016/0022-2836(71)90324-x

Lehmann, J. and Kleber, M. (2015). The contentious nature of soil organic matter. *Nature* 528, 60–68. doi:10.1038/nature16069

Leinweber, P., Kruse, J., Walley, F. L., Gillespie, A., Eckhardt, K.-U., Blyth, R. I. R., et al. (2007). Nitrogen K-edge XANES – an overview of reference compounds used to identify 'unknown' organic nitrogen in environmental samples. *Journal of Synchrotron Radiation* 14, 500–511. doi:10.1107/s0909049507042513

Louwen, J., Pye, C., van Lenthe, E., Austin, N., McGarrity, E., Xiong, R., et al. (2023). AMS 2023.1 COSMO-RS. SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands. `https://www.scm.com`

Luarte, T., Gómez-Aburto, V. A., Poblete-Castro, I., Castro-Nallar, E., Huneeus, N., Molina-Montenegro, M., et al. (2023). Levels of persistent organic pollutants (POPs) in the Antarctic atmosphere over time (1980 to 2021) and estimation of their atmospheric half-lives. *Atmospheric Chemistry and Physics* 23, 8103–8118. doi:10.5194/acp-23-8103-2023

Machado, M. R., Barrera, E. E., Klein, F., Sóñora, M., Silva, S., and Pantano, S. (2019). The SIRAH 2.0 Force Field: Altius, Fortius, Citius. *Journal of Chemical Theory and Computation* 15, 2719–2733. doi:10.1021/acs.jctc.9b00006

Marrink, S. J., de Vries, A. H., and Mark, A. E. (2004). Coarse Grained Model for Semiquantitative Lipid Simulations. *The Journal of Physical Chemistry B* 108, 750–760. doi:10.1021/jp036508g

Marrink, S. J., Monticelli, L., Melo, M. N., Alessandri, R., Tieleman, D. P., and Souza, P. C. T. (2023). Two decades of Martini: Better beads, broader scope. *WIREs Computational Molecular Science* 13, e1620. doi:10.1002/wcms.1620

Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., and de Vries, A. H. (2007). The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *The Journal of Physical Chemistry B* 111, 7812–7824. doi:10.1021/jp071097f

Masoom, H., Courtier-Murias, D., Farooq, H., Soong, R., Kelleher, B. P., Zhang, C., et al. (2016). Soil Organic Matter in Its Native State: Unravelling the Most Complex Biomaterial on Earth. *Environmental Science & Technology* 50, 1670–1680. doi: 10.1021/acs.est.5b03410

Mey, A. S., Allen, B. K., Bruce Macdonald, H. E., Chodera, J. D., Hahn, D. F., Kuhn, M., et al. (2020). Best Practices for Alchemical Free Energy Calculations [Article v1.0]. *Living Journal of Computational Molecular Science* 2, 18378. doi: 10.33011/livecoms.2.1.18378

Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., and Beckstein, O. (2011). MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* 32, 2319–2327. doi:10.1002/jcc.21787

Nobile, M. S., Cazzaniga, P., Besozzi, D., Colombo, R., Mauri, G., and Pasi, G. (2018). Fuzzy Self-Tuning PSO: A settings-free algorithm for global optimization. *Swarm and Evolutionary Computation* 39, 70–85. doi:10.1016/j.swevo.2017.09.001

Noid, W., Szukalo, R. J., Kidder, K. M., and Lesniewski, M. C. (2024). Rigorous Progress in Coarse-Graining. *Annual Review of Physical Chemistry* 75, 21–45. doi: 10.1146/annurev-physchem-062123-010821

Noid, W. G. (2013). Perspective: Coarse-grained models for biomolecular systems. *The Journal of Chemical Physics* 139, 090901. doi:10.1063/1.4818908

Noid, W. G., Chu, J.-W., Ayton, G. S., Krishna, V., Izvekov, S., Voth, G. A., et al. (2008). The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *The Journal of Chemical Physics* 128, 244114. doi: 10.1063/1.2938860

Olk, D. C., Bloom, P. R., De Nobili, M., Chen, Y., McKnight, D. M., Wells, M. J. M., et al. (2019a). Using Humic Fractions to Understand Natural Organic Matter Processes in Soil and Water: Selected Studies and Applications. *Journal of Environmental Quality* 48, 1633–1643. doi:10.2134/jeq2019.03.0100

Olk, D. C., Bloom, P. R., Perdue, E. M., McKnight, D. M., Chen, Y., Farenhorst, A., et al. (2019b). Environmental and Agricultural Relevance of Humic Fractions Extracted by Alkali from Soils and Natural Waters. *Journal of Environmental Quality* 48, 217–232. doi:10.2134/jeq2019.02.0041

Orsi, M. and Essex, J. W. (2011). The ELBA Force Field for Coarse-Grain Modeling of Lipid Membranes. *PLoS ONE* 6, e28637. doi:10.1371/journal.pone.0028637

Ozboyaci, M., Kokh, D. B., Corni, S., and Wade, R. C. (2016). Modeling and simulation of protein–surface interactions: achievements and challenges. *Quarterly Reviews of Biophysics* 49, e4. doi:10.1017/s0033583515000256

Parrinello, M. and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* 52, 7182–7190. doi:10.1063/1.328693

Paul, E. A. (2016). The nature and dynamics of soil organic matter: Plant inputs, microbial transformations, and organic matter stabilization. *Soil Biology and Biochemistry* 98, 109–126. doi:10.1016/j.soilbio.2016.04.001

Paustian, K., Lehmann, J., Ogle, S., Reay, D., Robertson, G. P., and Smith, P. (2016). Climate-smart soils. *Nature* 532, 49–57. doi:10.1038/nature17174

Pérez, F. and Granger, B. E. (2007). IPython: A System for Interactive Scientific Computing. *Computing in Science and Engineering* 9, 21–29. doi:10.1109/mcse.2007.53

Pérez-Reverón, R., Álvarez-Méndez, S. J., González-Sálamo, J., Socas-Hernández, C., Díaz-Peña, F. J., Hernández-Sánchez, C., et al. (2023). Nanoplastics in the soil environment: Analytical methods, occurrence, fate and ecological implications. *Environmental Pollution* 317, 120788. doi:10.1016/j.envpol.2022.120788

Petrov, D., Tunega, D., Gerzabek, M. H., and Oostenbrink, C. (2017). Molecular Dynamics Simulations of the Standard Leonardite Humic Acid: Microscopic Analysis of the Structure and Dynamics. *Environmental Science & Technology* 51, 5414–5424. doi:10.1021/acs.est.7b00266

Petrov, D., Tunega, D., Gerzabek, M. H., and Oostenbrink, C. (2019). Molecular modelling of sorption processes of a range of diverse small organic molecules in Leonardite humic acid. *European Journal of Soil Science* 71, 831–844. doi:10.1111/ejss.12868

Piccolo, A. (2001). The supramolecular structure of humic substances. *Soil Science* 166, 810–832. doi:10.1097/00010694-200111000-00007

Piccolo, A. (2002). The supramolecular structure of humic substances: A novel understanding of humus chemistry and implications in soil science. In *Advances in Agronomy* (Elsevier). 57–134. doi:10.1016/s0065-2113(02)75003-7

Pye, C. C. and Ziegler, T. (1999). An implementation of the conductor-like screening model of solvation within the Amsterdam density functional package. *Theoretical Chemistry Accounts* 101, 396–408. doi:10.1007/s002140050457

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77, 257–286. doi:10.1109/5.18626

Rawls, W., Pachepsky, Y., Ritchie, J., Sobecki, T., and Bloodworth, H. (2003). Effect of soil organic carbon on soil water retention. *Geoderma* 116, 61–76. doi:10.1016/s0016-7061(03)00094-6

Reif, M. M., Hünenberger, P. H., and Oostenbrink, C. (2012). New Interaction Parameters for Charged Amino Acid Side Chains in the GROMOS Force Field. *Journal of Chemical Theory and Computation* 8, 3705–3723. doi:10.1021/ct300156h

Riniker, S. and Landrum, G. A. (2015). Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *Journal of Chemical Information and Modeling* 55, 2562–2574. doi:10.1021/acs.jcim.5b00654

Ritter, L., Solomon, K. R., Forget, J., Stemeroff, M., and O'Leary, C. (1995). *Persistent Organic Pollutants: An Assessment Report on DDT, Aldrin, Dieldrin, Endrin, Chlordane, Heptachlor, Hexachlorobenzene, Mirex, Toxaphene, Polychlorinated Biphenyls, Dioxins and Furans.* Tech. rep., International Programme on Chemical Safety (IPCS). `https://chm.pops.int/portals/0/docs/from_old_website/documents/background/assessreport/en/ritteren.pdf`

Roy, R. N., Finck, A., Blair, G. J., and Tandon, H. L. S. (2006). *Plant nutrition for food security: A guide for integrated nutrient management.* No. 16 in FAO Fertilizer and Plant Nutrition Bulletin (Food and Agriculture Organization of the United Nations). `https://www.fao.org/3/a0443e/a0443e.pdf`

Rubinstein, M. and Colby, R. H. (2003). *Polymer Physics* (Oxford University Press). doi:10.1093/oso/9780198520597.001.0001

Rubner, Y., Tomasi, C., and Guibas, L. (1998). A metric for distributions with applications to image databases. In *Sixth International Conference on Computer Vision* (Bombay, India: IEEE), 59–66. doi:10.1109/iccv.1998.710701

Saiz-Jimenez, C. and De Leeuw, J. (1986). Chemical characterization of soil organic matter fractions by analytical pyrolysis-gas chromatography-mass spectrometry. *Journal of Analytical and Applied Pyrolysis* 9, 99–119. doi:10.1016/0165-2370(86)85002-1

Scharlemann, J. P., Tanner, E. V., Hiederer, R., and Kapos, V. (2014). Global soil carbon: understanding and managing the largest terrestrial carbon pool. *Carbon Management* 5, 81–91. doi:10.4155/cmt.13.77

Schaumann, G. E., Gildemeister, D., Kunhi Mouvenchery, Y., Spielvogel, S., and Diehl, D. (2013). Interactions between cations and water molecule bridges in soil organic matter. *Journal of Soils and Sediments* 13, 1579–1588. doi:10.1007/s11368-013-0746-7

Schmid, N., Eichenberger, A. P., Choutko, A., Riniker, S., Winger, M., Mark, A. E., et al. (2011). Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European Biophysics Journal* 40, 843–856. doi:10.1007/s00249-011-0700-9

Schulten, H.-R. (1999a). Analytical pyrolysis and computational chemistry of aquatic humic substances and dissolved organic matter. *Journal of Analytical and Applied Pyrolysis* 49, 385–415. doi:10.1016/s0165-2370(98)00137-5

Schulten, H.-R. (1999b). Interactions of dissolved organic matter with xenobiotic compounds: Molecular modeling in water. *Environmental Toxicology and Chemistry* 18, 1643–1655. doi:10.1002/etc.5620180806

Schulten, H.-R. and Leinweber, P. (2000). New insights into organic-mineral particles: composition, properties and models of molecular structure. *Biology and Fertility of Soils* 30, 399–432. doi:10.1007/s003740050020

Schulten, H.-R., Plage, B., and Schnitzer, M. (1991). A chemical structure for humic substances. *The Science of Nature* 78, 311–312. doi:10.1007/bf01221416

Schulten, H.-R. and Schnitzer, M. (1993). A state of the art structural concept for humic substances. *Naturwissenschaften* 80, 29–30. doi:10.1007/bf01139754

Schulten, H.-R. and Schnitzer, M. (1995). Three-dimensional models for humic acids and soil organic matter. *Naturwissenschaften* 82, 487–498. doi:10.1007/bf01134484

Schulten, H.-R. and Schnitzer, M. (1997). Chemical Model Structures for Soil Organic Matter and Soils. *Soil Science* 162, 115–130. doi:10.1097/00010694-199702000-00005

Schwarz, J., Thiele-Bruhn, S., Eckhardt, K.-U., and Schulten, H.-R. (2012). Sorption of Sulfonamide Antibiotics to Soil Organic Sorbents: Batch Experiments with Model Compounds and Computational Chemistry. *International Scholarly Research Notices* 2012, 159189. doi:10.5402/2012/159189

Sein, L. T., Varnum, J. M., and Jansen, S. A. (1999). Conformational Modeling of a New Building Block of Humic Acid: Approaches to the Lowest Energy Conformer. *Environmental Science & Technology* 33, 546–552. doi:10.1021/es9805324

Shell, M. S. (2016). *Coarse-Graining with the Relative Entropy* (John Wiley & Sons, Ltd), chap. 5. 395–441. doi:10.1002/9781119290971.ch5

Sherman, D. M. (2016). *Introduction to the Theory and Methods of Computational Chemistry* (John Wiley & Sons, Ltd), chap. 1. 1–31. doi:10.1002/9781118845226.ch1

Shirts, M. R. and Pande, V. S. (2005). Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *The Journal of Chemical Physics* 122, 134508. doi:10.1063/1.1877132

Shrake, A. and Rupley, J. (1973). Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *Journal of Molecular Biology* 79, 351–371. doi:10.1016/0022-2836(73)90011-9

Souza, P. C. T., Alessandri, R., Barnoud, J., Thallmair, S., Faustino, I., Grünewald, F., et al. (2021). Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nature Methods* 18, 382–388. doi:10.1038/s41592-021-01098-3

Sündermann, A., Solc, R., Tunega, D., Haberhauer, G., Gerzabek, M. H., and Oostenbrink, C. (2015). Vienna Soil-Organic-Matter Modeler—Generating condensed-phase models of humic substances. *Journal of Molecular Graphics and Modelling* 62, 253–261. doi:10.1016/j.jmgm.2015.10.007

Sutton, R., Sposito, G., Diallo, M. S., and Schulten, H. (2005). Molecular simulation of a model of dissolved organic matter. *Environmental Toxicology and Chemistry* 24, 1902–1911. doi:10.1897/04-567r.1

Szczuka, M., Pereira, G. P., Walter, L. J., Gueroult, M., Poulain, P., Bereau, T., et al. (2025). Fast parameterization of Martini3 models for fragments and small molecules. *bioRxiv* doi:10.1101/2025.07.13.664596

Thiele-Bruhn, S., Seibicke, T., Schulten, H., and Leinweber, P. (2004). Sorption of Sulfonamide Pharmaceutical Antibiotics on Whole Soils and Particle-Size Fractions. *Journal of Environmental Quality* 33, 1331–1342. doi:10.2134/jeq2004.1331

Thorn, K. A., Folan, D. W., and MacCarthy, P. (1989). *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 89-4196, U.S. Geological Survey. doi:10.3133/wri894196

Tironi, I. G., Sperb, R., Smith, P. E., and van Gunsteren, W. F. (1995). A generalized reaction field method for molecular dynamics simulations. *The Journal of Chemical Physics* 102, 5451–5459. doi:10.1063/1.469273

Tschöp, W., Kremer, K., Batoulis, J., Bürger, T., and Hahn, O. (1998). Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polymerica* 49, 61–74. doi:10.1002/(sici)1521-4044(199802)49:2/3⟨61::aid-apol61⟩3.0.co;2-v

Tunega, D., Gerzabek, M. H., Haberhauer, G., Lischka, H., Solc, R., and Aquino, A. J. A. (2020). Adsorption process of polar and nonpolar compounds in a nanopore model of humic substances. *European Journal of Soil Science* 71, 845–855. doi:10.1111/ejss.12865

Tunega, D., Gerzabek, M. H., and Lischka, H. (2004). Ab Initio Molecular Dynamics Study of a Monomolecular Water Layer on Octahedral and Tetrahedral Kaolinite Surfaces. *The Journal of Physical Chemistry B* 108, 5930–5936. doi:10.1021/jp037121g

U.S. Environmental Protection Agency (2025). EPA EcoBox Tools by Exposure Pathways – Exposure Pathways in ERA. `https://www.epa.gov/ecobox/epa-ecobox-tools-exposure-pathways-exposure-pathways-era`. Last updated July 14, 2025

van Gunsteren, W. F. and Oostenbrink, C. (2024). Methods for Classical-Mechanical Molecular Simulation in Chemistry: Achievements, Limitations, Perspectives. *Journal of Chemical Information and Modeling* 64, 6281–6304. doi:10.1021/acs.jcim.4c00823

Verlet, L. (1967). Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* 159, 98–103. doi:10.1103/physrev.159.98

Wang, S., Witek, J., Landrum, G. A., and Riniker, S. (2020). Improving Conformer Generation for Small Rings and Macrocycles Based on Distance Geometry and Experimental Torsional-Angle Preferences. *Journal of Chemical Information and Modeling* 60, 2044–2058. doi:10.1021/acs.jcim.0c00025

Warnau, J., Wichmann, K., and Reinisch, J. (2021). COSMO-RS predictions of logP in the SAMPL7 blind challenge. *Journal of Computer-Aided Molecular Design* 35, 813–818. doi:10.1007/s10822-021-00395-5

Webb, M. A., Delannoy, J.-Y., and de Pablo, J. J. (2019). Graph-Based Approach to Systematic Molecular Coarse-Graining. *Journal of Chemical Theory and Computation* 15, 1199–1208. doi:10.1021/acs.jctc.8b00920

Wilkinson, J. L., Boxall, A. B. A., Kolpin, D. W., Leung, K. M. Y., Lai, R. W. S., Galbán-Malagón, C., et al. (2022). Pharmaceutical pollution of the world's rivers. *Proceedings of the National Academy of Sciences* 119, e2113947119. doi:10.1073/pnas.2113947119

Xu, B., Yang, G., Lehmann, A., Riedel, S., and Rillig, M. C. (2022). Effects of perfluoroalkyl and polyfluoroalkyl substances (PFAS) on soil structure and function. *Soil Ecology Letters* 5, 108–117. doi:10.1007/s42832-022-0143-5

York, D. M. (2023). Modern Alchemical Free Energy Methods for Drug Discovery Explained. *ACS Physical Chemistry Au* 3, 478–491. doi:10.1021/acsphyschemau.3c00033

Zheng, W. and Guo, M. (2021). Soil–Plant Transfer of Pharmaceuticals and Personal Care Products. *Current Pollution Reports* 7, 510–523. doi:10.1007/s40726-021-00207-2

Zwolak, A., Sarzyńska, M., Szpyrka, E., and Stawarczyk, K. (2019). Sources of Soil Pollution by Heavy Metals and Their Accumulation in Vegetables: a Review. *Water, Air, & Soil Pollution* 230, 164. doi:10.1007/s11270-019-4221-y

Appendices

Declaration of Digital and AI-Assisted Tools

The following digital tools were used during the preparation of this thesis. I remain fully responsible for all content, analyses, and conclusions.

- *ChatGPT* (OpenAI, chat.openai.com) was exclusively used to:
  - suggest alternative phrasings to improve clarity and tone;
  - brainstorm search terms and obtain general topic overviews to plan literature searches; and
  - troubleshoot LaTeX syntax for equations.

  All citations and technical content were verified against primary sources, and AI-generated text was not used as a direct source. *ChatGPT* was not employed to generate research data, results, or interpretations.
- *LanguageTool* (languagetool.org) was used for grammar and spell checking.

Peer-Reviewed Publications

For each of the publications below, my supervisors, Dr. Ashour Ahmed and Prof. Oliver Kühn, collaborated with me in designing the studies and developing the corresponding model systems. I performed the simulations and post-processing of the simulation data. I prepared the figures included in the publications, which were finalized by me or my supervisors after consultation. The analysis and interpretation of the results were undertaken jointly with my supervisors. I wrote the first draft of each publication and worked with my supervisors on the revision and finalization of the manuscripts.

Published Papers

1. Dettmann, L. F., Kühn, O., and Ahmed, A. A. (2024). Martini-Based Coarse-Grained Soil Organic Matter Model Derived from Atomistic Simulations. *Journal of Chemical Theory and Computation* 20, 5291–5305. DOI: 10.1021/acs.jctc.4c00332.
2. Dettmann, L. F., Kühn, O., and Ahmed, A. A. (2024). Automated Parametrization Approach for Coarse-Graining Soil Organic Matter Molecules. *Journal of Chemical Theory and Computation* 20, 10684–10696. DOI: 10.1021/acs.jctc.4c01334
3. Dettmann, L. F., Kühn, O., and Ahmed, A. A. (2025). Coarse-Grained Simulations of Sulfanilamide and Hexachlorobenzene Mobility in Soil Organic Matter. *Environmental Science: Advances* DOI: 10.1039/d5va00237k
4. Dettmann, L. F., Kühn, O., and Ahmed, A. A. (2025). Coarse-Grained Simulations of the Nanoplastic Interaction with Soil Organic Matter. *The Journal of Physical Chemistry B* 129, 11605–11615. DOI: 10.1021/acs.jpcb.5c06072

JCTC Journal of Chemical Theory and Computation

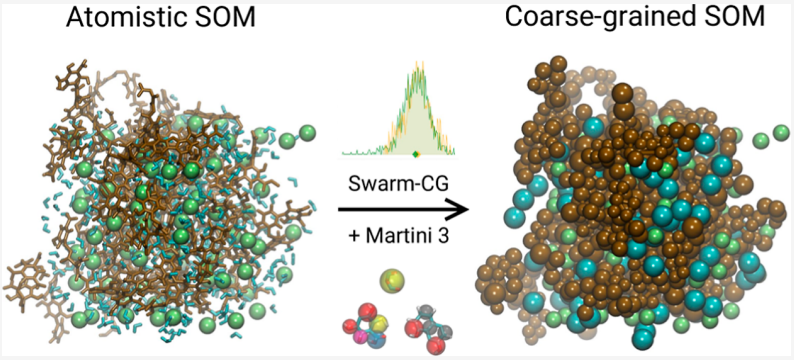
pubs.acs.org/JCTC Article

Martini-Based Coarse-Grained Soil Organic Matter Model Derived from Atomistic Simulations

Lorenz F. Dettmann,* Oliver Kühn, and Ashour A. Ahmed

Cite This: J. Chem. Theory Comput. 2024, 20, 5291−5305 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: The significance of soil organic matter (SOM) in environmental contexts, particularly its role in pollutant adsorption, has prompted an 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 models 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 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 showed partial agreement, however, reproducing certain peaks was challenging due to fixed bead sizes. Detailed analysis of atomistic systems revealed different configurations between the groups, 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.

Downloaded via UNIV ROSTOCK on July 17, 2025 at 12:37:50 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

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, as it provides a reservoir of essential 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 topsoils, 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 sulfur.

Received: March 14, 2024
Revised: May 22, 2024
Accepted: May 22, 2024
Published: June 4, 2024



ACS Publications
© 2024 The Authors. Published by American Chemical Society

https://doi.org/10.1021/acs.jctc.4c00332
J. Chem. Theory Comput. 2024, 20, 5291−5305

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.[1] Examples of contaminants include persistent organic pollutants,[3] micro- and nanoplastics,[4,5] heavy metals,[6] and antibiotics,[7] 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.[8−13]

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.[14] 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).[15,16] 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,[2,17−20] including interactions with proteins[21] and mineral surfaces,[22] 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.[23] Thereby, atoms are grouped into new particles, so-called beads, which effectively reduces 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.[24,25] Similar to atomistic methods, force fields for coarse-grained simulations exist.[26−28] Of these, the Martini force field[29] is the most commonly used.

Martini uses a building block approach with preparametrized beads that are applicable to a broad range of chemical compounds. Recently, a new version, Martini 3, has been published, which involved a complete reparametrization of the force field.[30] 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,[31] 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. Beyond the benefits inherent in coarse-grained methodologies, which encompass the simulation of larger molecular systems and extended durations compared to atomistic approaches, there are additional advantages to modeling SOM. Notably, given the typical uncertainty surrounding the precise structure of SOM molecules, coarse-grained models offer a simplified representation that abstracts from atomistic intricacies through averaging.

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[32] tool. As an example, a coarse-grained model of the Leonardite humic acid (LHA) is created, which has been previously studied using MD simulations.[2,17−22] 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 provided 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.[16] 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 nonbonded interactions are then set according to the guidelines for parametrizing new molecules for the Martini 3 force field.[30] The parameters for bonded interactions are determined via Swarm-CG[32] using trajectories of isolated fragments in water in their atomistic representation as input. Transfer free energies, calculated using the COSMO-RS method[33−35] 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

https://doi.org/10.1021/acs.jctc.4c00332
J. Chem. Theory Comput. 2024, 20, 5291−5305

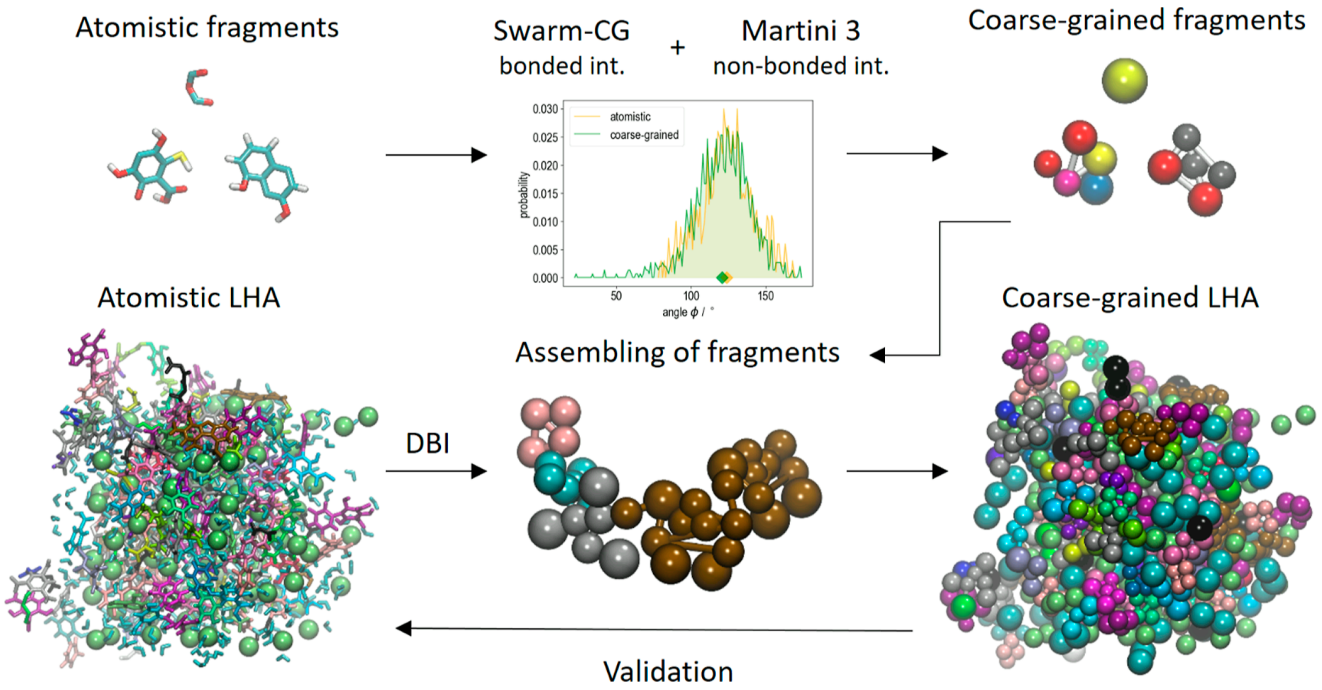


Figure 1. Overview of the procedure for creating the coarse-grained SOM models. Atomistic fragments are transformed into a coarse-grained representation using Swarm-CG and Martini 3. These fragments are assembled with parameters from DBI, determined from atomistic simulations, for bonded interactions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

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.[17,19] A graphical representation providing an overview of the molecular modeling and parametrization approach is illustrated in Figure 1.

2.2. Parametrization Approach. *2.2.1. Determination of Mapping and Nonbonded 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 reproduce 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 nonbonded interactions. In Martini, a building block approach is used with predefined 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 six 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,[30] 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

https://doi.org/10.1021/acs.jctc.4c00332
J. Chem. Theory Comput. 2024, 20, 5291−5305

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.[32] 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 nonbonded interactions, following a common practice in certain parametrizations in the frame of the Martini force field.[36−38]

2.2.3. Determination of Reference Properties. As mentioned earlier, the transfer free energy ΔG is 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[39] 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 first from the atomistic reference trajectories of isolated fragments and second 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 nonbonded 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, (noniterative) 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[40] version 2019.4. The analysis was conducted using GROMACS tools and the Python 3 package MDAnalysis.[41,42] Other packages that were utilized included NumPy,[43] SciPy,[44] IPython,[45] and Matplotlib.[46] Structures were visualized with VMD.[47] Unless otherwise stated, the solvation of molecules was performed using Packmol.[48]

2.3.1. Atomistic Isolated Fragments. For the atomistic simulations, the GROMOS force field version 54A7[49,50] was used, together with the SPC water model.[51] 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[52] at 298.15 K. Furthermore, equations of motion were integrated with the leapfrog 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[53] with a coupling constant of $\tau_p = 1.0$ ps and a reference pressure of 1 bar. In the case of short-range nonbonded 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.[16] The input parameters, which were based on an experimentally determined[54] 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 counterions because it is abundant in most soils and soil solutions. In total, 200 building blocks were used, with 5 building blocks per molecule.

https://doi.org/10.1021/acs.jctc.4c00332
J. Chem. Theory Comput. 2024, 20, 5291−5305

Table 1. Experimentally (exp) Determined Composition of Carbon and Nitrogen, and Functional Groups, Contained in LHA, and Corresponding Composition of the LHA Models[a]

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

[a]Experimentally determined values were given as input parameters to the modeler.

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 equilibration procedure included two *NPT* equilibrations instead of one *NVT* and one *NPT* equilibration. The Berendsen barostat was used for the first equilibration and the Parrinello−Rahman barostat for the second, which each lasted 1 ns. 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 or 25 ns, depending on the respective optimization cycle. For more details about the optimization procedure used by Swarm-CG, see the corresponding study.[32] The equations of motions were integrated using the leapfrog integrator, with a time step of 20 fs. For the equilibration, the Berendsen barostat[55] was used with a coupling constant of τ_p = 4.0 ps, whereas for the production run, the Parrinello−Rahman barostat was used with a coupling constant of τ_p = 8.0 ps. For each equilibration and production run, 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.[56] 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 ϵ_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.[30] 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 leapfrog 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 α = 0.5. Finally, the transfer free energies were calculated from the difference between two solvation free energies: $\Delta G_{S1\rightarrow S2} = \Delta G_{S1\rightarrow \emptyset} - \Delta G_{S2\rightarrow \emptyset}$. The solvation-free energies were calculated with the help of the Bennett acceptance ratio method[57] 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 representation of calcium ions utilized SD beads, consistent with the standard bead type for calcium. In contrast to the free energy calculations, charged fragments were also utilized in this study, following the atomistic representation. The mapped structure was solvated, first with regular-sized water (four water molecules represented by one bead), and second with tiny-sized water (two water molecules represented by one bead) 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 τ_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 AND 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 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 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 standard deviations for the values of the

https://doi.org/10.1021/acs.jctc.4c00332
J. Chem. Theory Comput. 2024, 20, 5291−5305

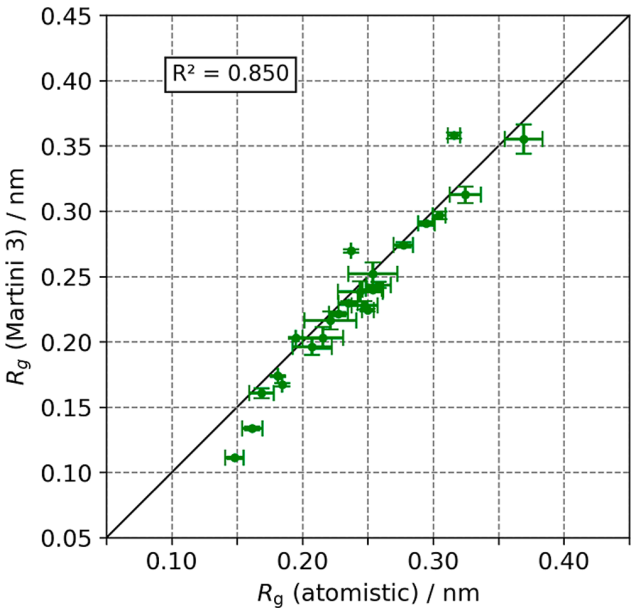


Figure 2. Comparison of time-averaged R_g of atomistic fragments with their coarse-grained counterparts (Martini 3); error bars according to standard deviations.

atomistic fragments are generally larger than those for the coarse-grained fragments. This implies a more rigid structure of the coarse-grained fragments, which exhibit less variability in R_g. 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 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.[58] 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.

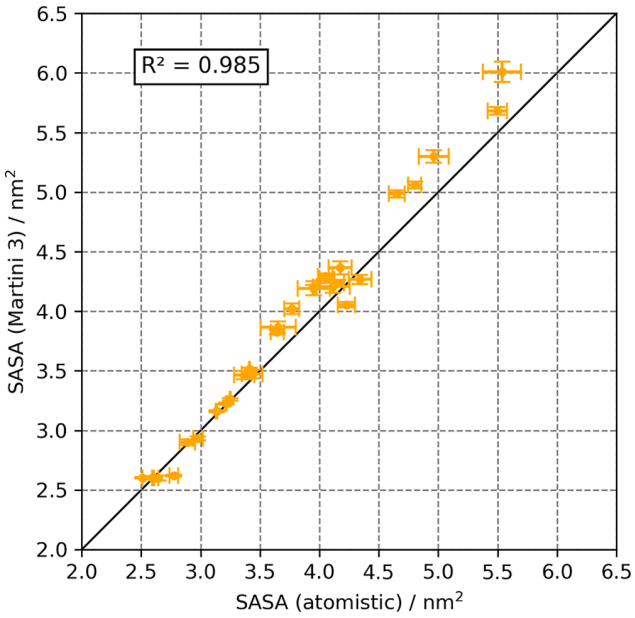


Figure 3. Comparison of time-averaged SASA of atomistic fragments with their coarse-grained counterparts (Martini 3); error bars according to the standard deviation.

In Figure 4, transfer free energies for each fragment are shown. The free energies from the coarse-grained fragments

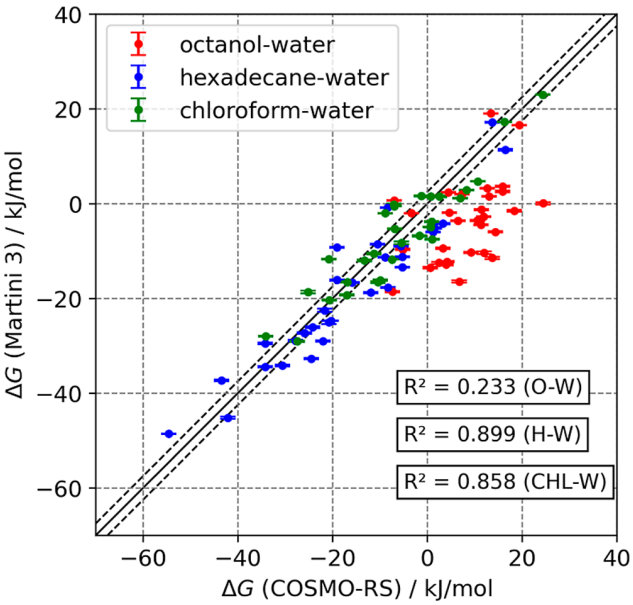


Figure 4. Comparison of octanol−water, hexadecane−water, and chloroform−water transfer free energies ΔG of coarse-grained fragments (Martini 3) with the reference. ΔG for coarse-grained fragments were determined from the difference between two solvation-free energies. Reference values were determined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

are plotted against reference values from COSMO-RS.[33−35] Here, the dotted lines represent a free energy deviation from the reference of $k_BT \approx 2.479$ kJ mol^{-1}. Points between these lines represent a very good agreement between the calculated and reference-free energy values.[59] Matching transfer free energies is one important criterion for accurately parametrizing molecules for the Martini force field.[30,58−60] In principle, this could have been done with automated tools[36,37] 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.[30] This improves the parametrization of a molecule with respect to its wide range of chemical properties. Following this approach and

https://doi.org/10.1021/acs.jctc.4c00332
J. Chem. Theory Comput. 2024, 20, 5291−5305

because experimental values were not available for each fragment, reference values were calculated by using COMSO-RS. Here, similar to the free energy calculations, hydrated octanol was employed to determine the octanol–water values.

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 lower correlation. To further investigate the octanol–water case, reference values were recalculated using ALOGPS,[61] 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 values, 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.

To validate COSMO-RS and ALOGPS as references for calculating transfer-free energies and comparing them against corresponding values from Martini 3 and experiments, COSMO-RS and ALOGPS calculations were performed for selected molecules previously studied in the main publication of Martini 3. Overall, good correlations are evident in the comparison with experimental data (see Figure S4 in the Supporting Information). This also holds true for the comparison with Martini 3, albeit with a higher degree of scatter observed than for the comparison of Martini 3 to experimental data (refer to the Martini 3 main publication). This finding supports the notion that COSMO-RS and ALOGPS are suitable for verifying general trends but may also contribute to a slightly lower overall correlation.

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[59]). 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 systems' densities were calculated with the help of the GROMACS tool *gmx energy*. As standard masses of Martini were employed for the coarse-grained model, the coarse-grained densities were adjusted by scaling them with the ratio of the mass of the associated atomistic system to the mass of the coarse-grained system, ensuring improved comparability. 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

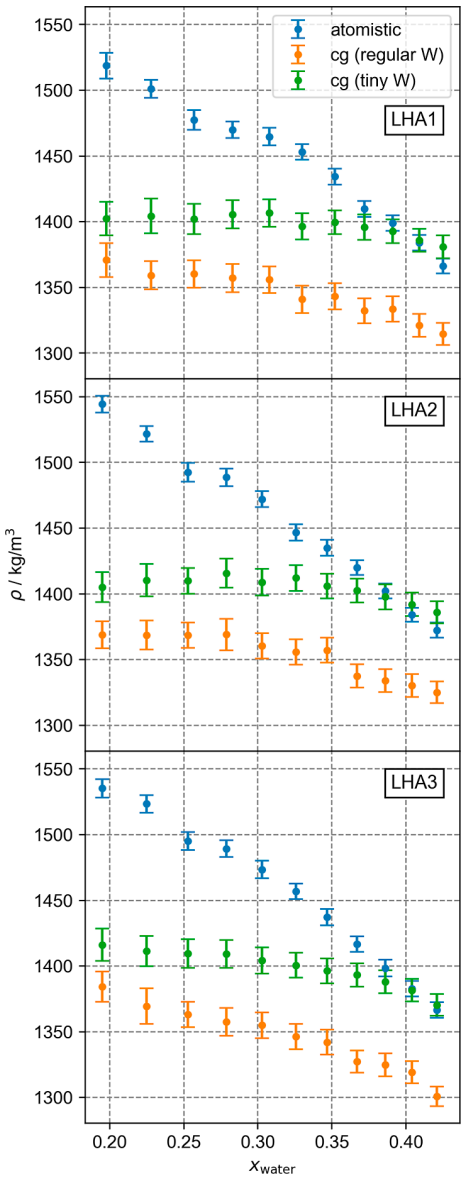


Figure 5. Time averaged system densities of the three LHA models, respectively, for different water contents. For each system, densities of the atomistic model (blue), coarse-grained model with regular water (orange), and coarse-grained model with tiny water (green) are plotted. Error bars indicate standard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

https://doi.org/10.1021/acs.jctc.4c00332
J. Chem. Theory Comput. 2024, 20, 5291–5305

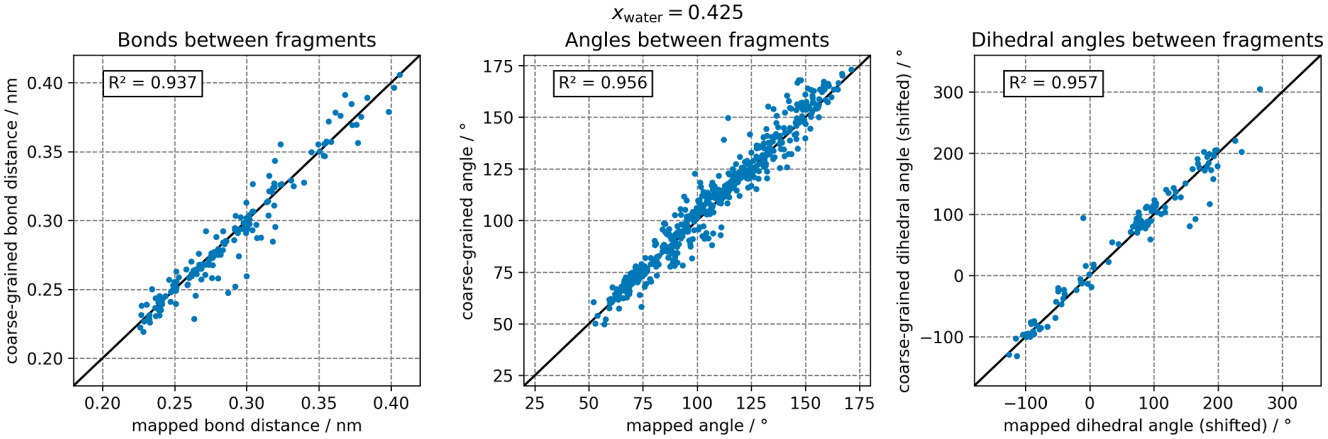


Figure 6. Comparison of average values of bonded interactions between a coarse-grained LHA structure and the mapped version of its atomistic counterpart. Here, the LHA1 model with x_{water} = 0.425 is considered. Bonded interactions are divided into bonds, angles, and dihedral angles between different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

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 the respective 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 assess the relaxation process of the structures, the time evolutions of the densities were calculated (refer to Figure S6 in the Supporting Information). The densities needed longer to equilibrate for systems with a lower water content. In the case of the LHA1 model with x_{water} = 0.228, the equilibration time had to be extended by 100 ns. This was done similarly for the LHA2 model with x_{water} = 0.195 and x_{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 nonbonded interactions could be adjusted, as they may have a significant influence on the overall density (see, for instance, Lutsyk et al.[62]). 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.[63] 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 significant 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_{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 S7. 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

https://doi.org/10.1021/acs.jctc.4c00332
J. Chem. Theory Comput. 2024, 20, 5291−5305

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 nonbonded interactions, as mentioned previously, the bead types for the ions and charged groups are mostly fixed for the Martini force field. 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.

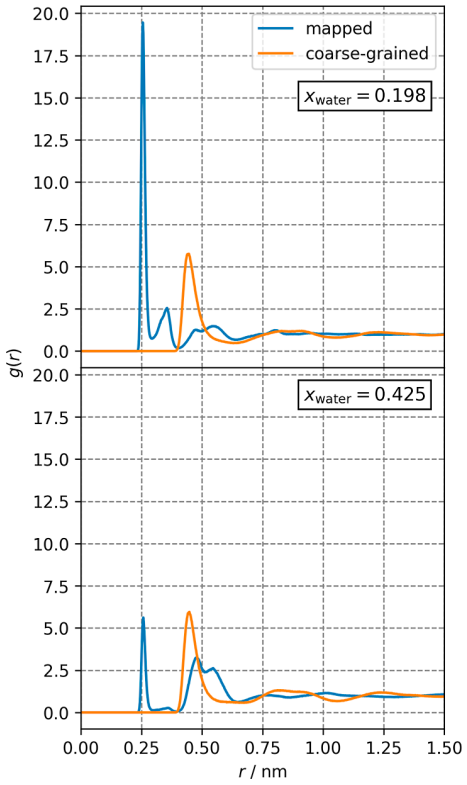


Figure 7. Comparison between mapped (blue) and coarse-grained (orange) RDFs, which were calculated between carboxylate groups and calcium ions. Here, the LHA1 model for x_{water} = 0.198 and for x_{water} = 0.425 is shown.

The 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 nonbonded interactions. One possible method, that involves conducting such matching, is the so-called iterative Boltzmann inversion.[64] 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 8a. Most notably, three peaks can be identified, which have been labeled as peaks I, II, and III, 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 8b), 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 ions are important for the SOM agglomeration process. Due to the presence of calcium ions, more dense SOM phases are created,

https://doi.org/10.1021/acs.jctc.4c00332
J. Chem. Theory Comput. 2024, 20, 5291−5305

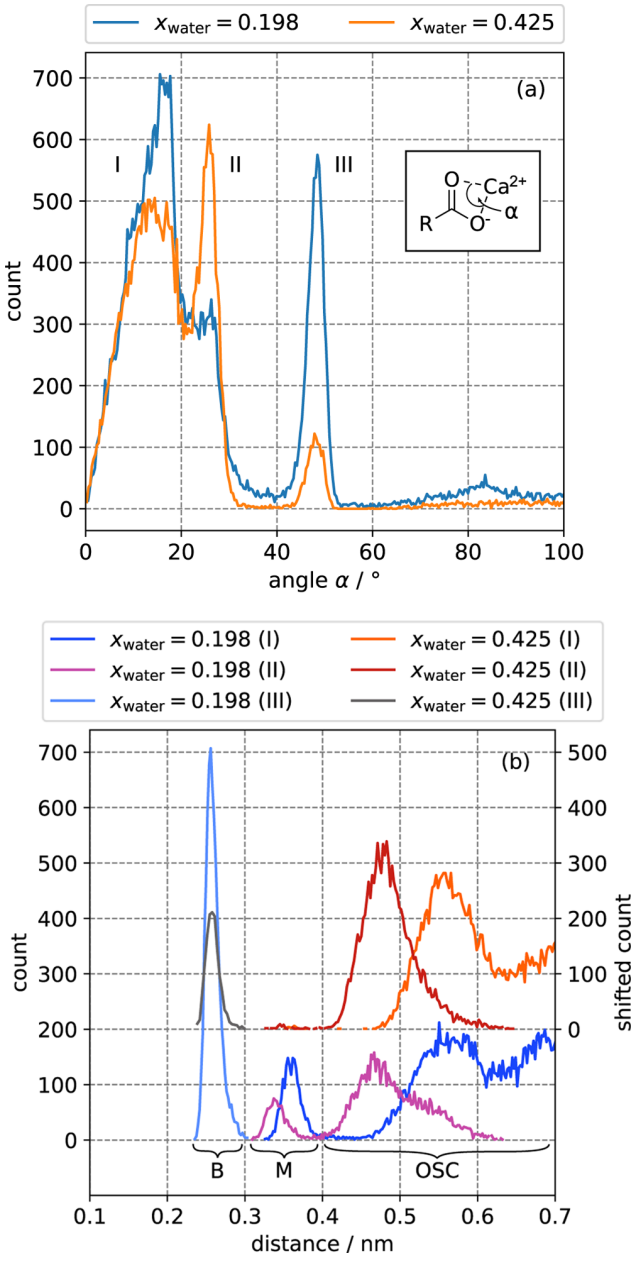


Figure 8. (a) Angle distributions between calcium ions and oxygen atoms of carboxylate groups, for lowest and highest water content considered. Visible peaks are divided into peaks I, II, and III. A cutoff radius of 0.7 nm between the calcium ions and the oxygen atoms was used for the angle calculations. (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.

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 8b. 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_{water} = 0.198. Configurations with a center of geometry distance larger than 0.40 nm can be assigned to OSCs. For the x_{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.

Additionally, Figure 7 adds another layer to this discussion. Atomistic simulations highlight the predominant role of direct calcium−carboxylate interactions (pertinent to the first peak from the left) at lower water content. This results in less pronounced calcium bridges (pertinent to the second peak) and a tighter packing of SOM molecules, consequently leading to a higher density. In contrast, at higher water content, the preference of calcium ions for solvation underscores the importance of calcium bridges (pertinent to the second peak) in influencing the packing of SOM molecules, thereby leading to a lower density. 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.

An alternative approach to mitigate density differences could involve representing a portion of calcium ions using smaller TD beads. This would accommodate various configurations (e.g., TD beads proportional to the number of B complexes, SD beads proportional to the number of OSCs). However, it is uncertain whether adjusting the TD to SD bead ratio based on the water content can be applied to diverse compositions. Consequently, we maintain the current representation of calcium ions with SD beads and keep the alternative approach open for future exploration.

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 large number 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 toward 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.

https://doi.org/10.1021/acs.jctc.4c00332
J. Chem. Theory Comput. 2024, 20, 5291−5305

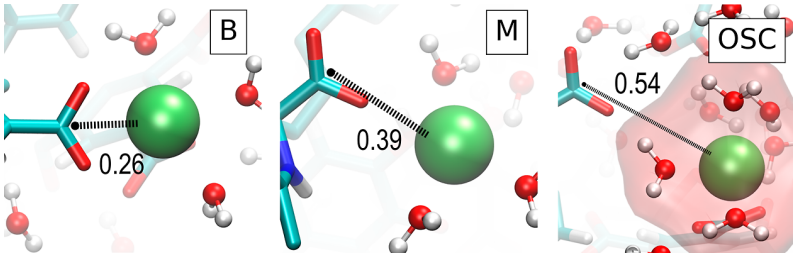


Figure 9. Trajectory snapshots for x_{water} = 0.198. Inner-sphere complexes [bidentate (B) and monodentate (M)], and an outer−sphere complex (OSC) are shown from left to right. Calcium ions are represented as green particles, LHA molecules as sticks, and water molecules as balls and sticks. Oxygen, 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.

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 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

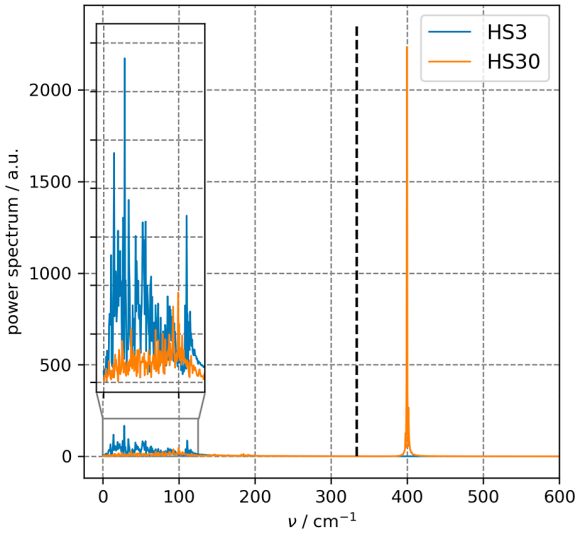


Figure 10. Power spectra of isolated fragments in water, calculated via the velocity autocorrelation function. As examples, the spectra for the fragments HS3 and HS30 are plotted. The dashed line denotes a frequency, with a time period of 5 times 20 fs. This frequency can be used as a guideline to verify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

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 leapfrog integrator, the time step should be at least 5 times shorter than the period of the fastest oscillation in the system.[65] 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. Furthermore, the investigation revealed that the instabilities are not attributed to the constraints, or VSs, as reported for cholesterol Martini models,[66] nor to the specific version of GROMACS employed. 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 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.[17,19] 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 11a, 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 S8 in the Supporting Information. Regarding the W−LHA hydrogen bonding in Figure 11a, 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[17] investigations.

https://doi.org/10.1021/acs.jctc.4c00332
J. Chem. Theory Comput. 2024, 20, 5291−5305

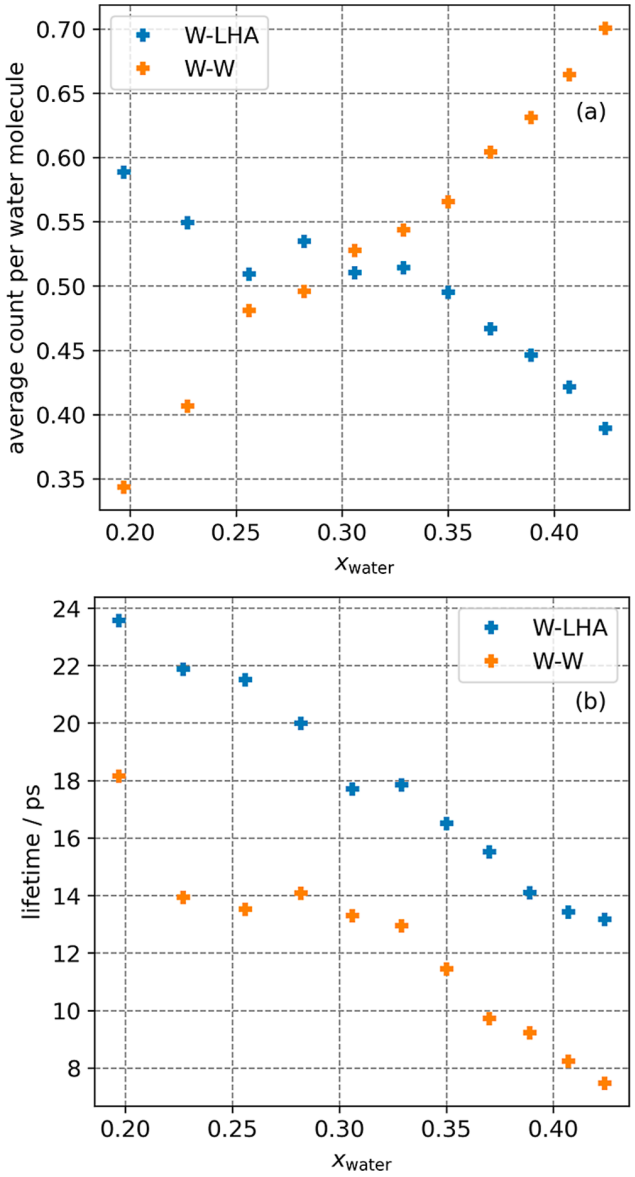


Figure 11. (a) Average number of hydrogen bonds per water molecule in the system depending on the water content. Distinctions were made between hydrogen bonds between the LHA molecules and water (W−LHA) and hydrogen bonds between the water molecules themselves (W−W). (b) Hydrogen bond lifetime, fitted 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.

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 11b, 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 comparable 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 nonbonded 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 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 counterions. 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

https://doi.org/10.1021/acs.jctc.4c00332
J. Chem. Theory Comput. 2024, 20, 5291−5305

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 facilitate easier access to these models, we are developing a tool that constructs coarse-grained SOM models without requiring trajectories. 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 properties such as density, RDF, and hydrogen bonds is lower compared to scenarios with higher water content. 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 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.

■ ASSOCIATED CONTENT

Data Availability Statement

All files required to reproduce the simulations and parametrization procedures in this study are available in Zenodo at https://doi.org/10.5281/zenodo.11191270.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.4c00332.

Information on the bead assignment, additional results from SASA and free energy calculations, and further explanation and results on the DBI approach (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Lorenz F. Dettmann − *Institute of Physics, University of Rostock, Rostock D-18059, Germany;* orcid.org/0000-0001-8407-4469; Email: lorenz.dettmann@uni-rostock.de

Authors

Oliver Kühn − *Institute of Physics, University of Rostock, Rostock D-18059, Germany; Department of Life, Light and Matter (LLM), University of Rostock, Rostock D-18059, Germany;* orcid.org/0000-0002-5132-2961

Ashour A. Ahmed − *Department of Life, Light and Matter (LLM), University of Rostock, Rostock D-18059, Germany;* orcid.org/0000-0002-5933-5249

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.4c00332

Notes

The authors declare no competing financial interest.

■ 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/Schachtschabel Soil Science*; Springer: Berlin Heidelberg, 2015.

(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) Jones, K.; de Voogt, P. Persistent organic pollutants (POPs): state of the science. *Environ. Pollut.* **1999**, *100*, 209−221.

(4) 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. *Environ. Int.* **2020**, *137*, 105263.

(5) 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. *Environ. Pollut.* **2023**, *317*, 120788.

(6) 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 Pollut.* **2019**, *230*, 164.

(7) Cycoń, M.; Mrozik, A.; Piotrowska-Seget, Z. Antibiotics in the Soil Environment—Degradation and Their Impact on Microbial Activity and Diversity. *Front. Microbiol.* **2019**, *10*, 338.

(8) Kubicki, J.; Apitz, S. Models of natural organic matter and interactions with organic contaminants. *Org. Geochem.* **1999**, *30*, 911−927.

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

(10) 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. *Eur. J. Soil Sci.* **2007**, *58*, 889−899.

(11) 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**, *508*, 276−287.

(12) 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.

(13) 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. *Environ. Sci. Pollut. Res.* **2023**, *30*, 36824−36837.

(14) Dettmann, L. F.; Kühn, O.; Ahmed, A. A. Coarse-grained molecular dynamics simulations of nanoplastics interacting with a hydrophobic environment in aqueous solution. *RSC Adv.* **2021**, *11*, 27734−27744.
(15) 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. *J. Mol. Graphics Modell.* **2015**, *62*, 253−261.
(16) Escalona, Y.; Petrov, D.; Oostenbrink, C. Vienna soil organic matter modeler 2 (VSOMM2). *J. Mol. Graphics Modell.* **2021**, *103*, 107817.
(17) 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. *Environ. Sci. Technol.* **2017**, *51*, 5414−5424.
(18) 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. *Eur. J. Soil Sci.* **2020**, *71*, 831−844.
(19) Escalona, Y.; Petrov, D.; Oostenbrink, C. Modeling soil organic matter: Changes in macroscopic properties due to microscopic changes. *Geochim. Cosmochim. Acta* **2021**, *307*, 228−241.
(20) 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.
(21) Gotsmy, M.; Escalona, Y.; Oostenbrink, C.; Petrov, D. Exploring the structure and dynamics of proteins in soil organic matter. *Proteins: Struct., Funct., Bioinf.* **2021**, *89*, 925−936.
(22) 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.
(23) 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.
(24) Joshi, S. Y.; Deshmukh, S. A. A review of advancements in coarse-grained molecular dynamics simulations. *Mol. Simul.* **2021**, *47*, 786−803.
(25) Noé, F.; Tkatchenko, A.; Müller, K. R.; Clementi, C. Machine Learning for Molecular Simulation. *Annu. Rev. Phys. Chem.* **2020**, *71*, 361−390.
(26) 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. *J. Chem. Theory Comput.* **2015**, *11*, 723−739.
(27) Machado, M. R.; Barrera, E. E.; Klein, F.; Sóñora, M.; Silva, S.; Pantano, S. The SIRAH 2.0 Force Field: Altius, Fortius, Citius. *J. Chem. Theory Comput.* **2019**, *15*, 2719−2733.
(28) Orsi, M.; Essex, J. W. The ELBA Force Field for Coarse-Grain Modeling of Lipid Membranes. *PLoS One* **2011**, *6*, No. e28637.
(29) 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. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(30) 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. *Nat. Methods* **2021**, *18*, 382−388.
(31) 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. *Environ. Sci. Technol.* **2018**, *52*, 14228−14234.
(32) 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.
(33) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99*, 2224−2235.
(34) 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.
(35) 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..
(36) Bereau, T.; Kremer, K. Automated Parametrization of the Coarse-Grained Martini Force Field for Small Organic Molecules. *J. Chem. Theory Comput.* **2015**, *11*, 2783−2791.
(37) 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.
(38) 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.
(39) Rowland, R. S.; Taylor, R. Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van der Waals Radii. *J. Phys. Chem.* **1996**, *100*, 7384−7391.
(40) 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.
(41) 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.
(42) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011**, *32*, 2319−2327.
(43) 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.
(44) 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.; et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261−272.
(45) Perez, F.; Granger, B. E. IPython: a System for Interactive Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 21−29.
(46) Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90−95.
(47) Humphrey, W.; Dalke, A.; Schulten, K. VMD − Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(48) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157−2164.
(49) Poger, D.; Van Gunsteren, W. F.; Mark, A. E. A new force field for simulating phosphatidylcholine bilayers. *J. Comput. Chem.* **2009**, *31*, 1117−1125.
(50) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and testing

https://doi.org/10.1021/acs.jctc.4c00332
J. Chem. Theory Comput. 2024, 20, 5291−5305

of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843−856.
(51) 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.
(52) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(53) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(54) 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.
(55) 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. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(56) de Jong, D. H.; Baoukina, S.; Ingólfsson, H. I.; Marrink, S. J. Martini straight: Boosting performance using a shorter cutoff and GPUs. *Comput. Phys. Commun.* **2016**, *199*, 1−7.
(57) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245−268.
(58) 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. *J. Chem. Theory Comput.* **2022**, *18*, 7555−7569.
(59) 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.
(60) 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. *Adv. Theory Simul.* **2021**, *5*, 2100391.
(61) Tetko, I. V.; Tanchuk, V. Y. Application of Associative Neural Networks for Prediction of Lipophilicity in ALOGPS 2.1 Program. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1136−1145.
(62) Lutsyk, V.; Wolski, P.; Plazinski, W. Extending the Martini 3 Coarse-Grained Force Field to Carbohydrates. *J. Chem. Theory Comput.* **2022**, *18*, 5089−5107.
(63) 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. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2022**, *13*, No. e1620.
(64) Reith, D.; Pütz, M.; Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* **2003**, *24*, 1624−1636.
(65) Apol, E.; Apostolov, R.; Berendsen, H. J.; Aldert Van Buuren; Par, B.; 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, M.; Murtola, T.; SzilArd, P. A.; 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.
(66) Borges-Araújo, L.; Borges-Araújo, A. C.; Ozturk, T. N.; Ramirez-Echemendia, D. P.; Fábián, B.; Carpenter, T. S.; Thallmair, S.; Barnoud, J.; Ingólfsson, H. I.; Hummer, G.; Tieleman, D. P.; Marrink, S. J.; Souza, P. C. T.; Melo, M. N. Martini 3 coarse-grained force field for cholesterol. *J. Chem. Theory Comput.* **2023**, *19*, 7387−7404.
(67) Cozzolino, A.; Conte, P.; Piccolo, A. Conformational changes of humic substances induced by some hydroxy-keto-and sulfonic acids. *Soil Biol. Biochem.* **2001**, *33*, 563−571.
(68) Schaumann, G. E. Soil organic matter beyond molecular structure Part I: Macromolecular and supramolecular characteristics. *J. Plant Nutr. Soil Sci.* **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. *Phys. Chem. Chem. Phys.* **2019**, *21*, 9845−9857.

https://doi.org/10.1021/acs.jctc.4c00332
J. Chem. Theory Comput. 2024, 20, 5291−5305

Supporting Information

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

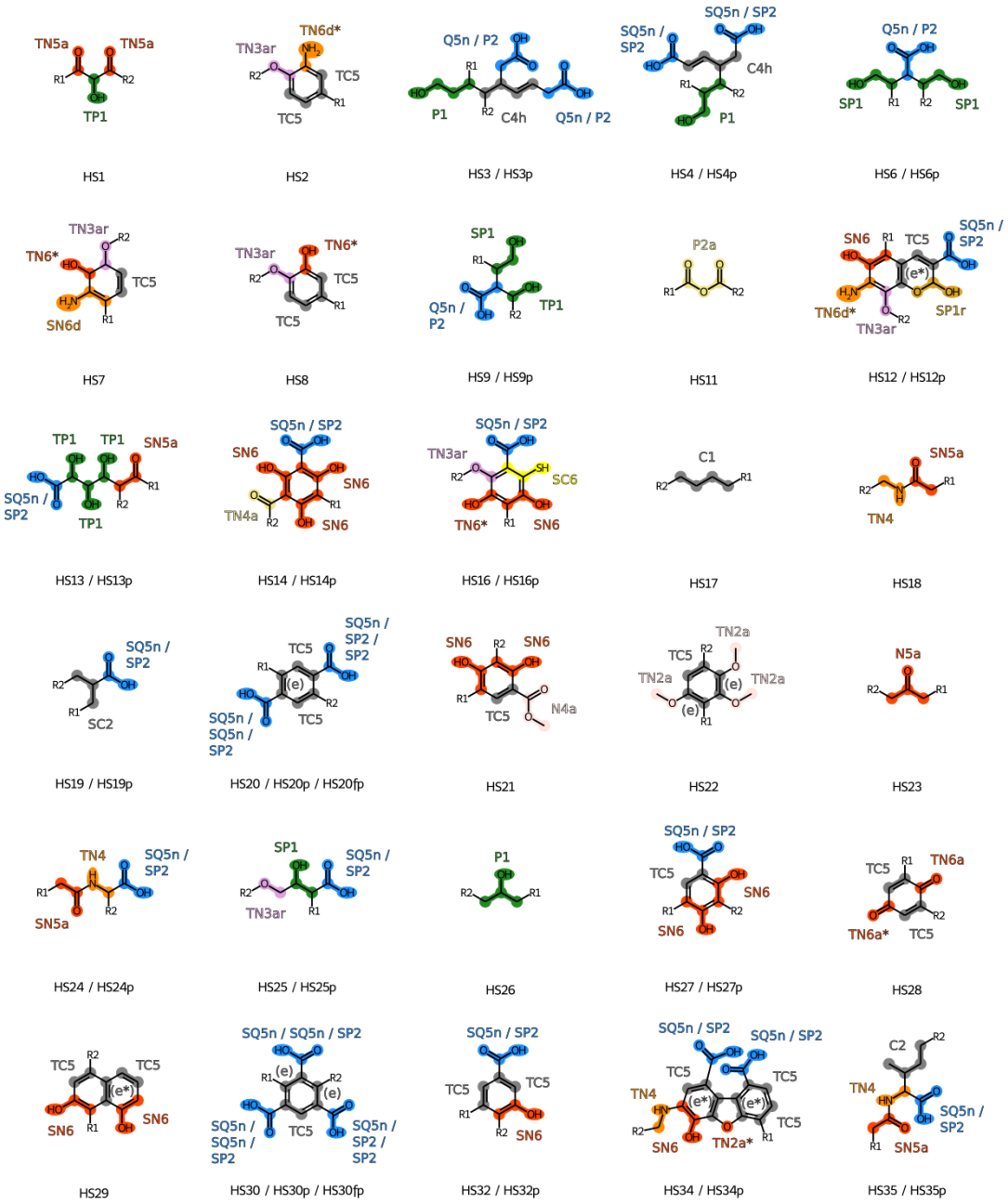


Figure S1: Fragments of the VSOMM2, shown in their atomistic representation (black) and with the applied Martini mapping (colored groups). TC5e beads were abbreviated with '(e)'. The '*' denotes the use of a virtual site (VS) for the corresponding bead. Certain fragments can be described by different protonation states, e.g. HS3 and HS3p. In the latter, the carboxylate groups are protonated. In this case, the beads are changed from Q5n to P2. For free energy calculations only, the fully protonated fragments HS20fp and HS30fp were introduced, which were not used for the SOM model construction. This was done to have a corresponding representative for the fragments HS20, HS20p, HS30, and HS30p for the free energy calculations, where only neutrally charged fragments were considered.

Table S1: List of beads applied to each fragment.

HS1
1: TN5a (1), 2: TP1 (2), 3: TN5a (1)
HS2
1: TC5 (3), 2: TC5 (3), 3: TN3ar (4), 4: TN6d (5)
HS3
1: P1 (2), 2: Q5n (6), 3: Q5n (6), 4: C4h (7)
HS3p
1: P1 (2), 2: P2 (8), 3: P2 (8), 4: C4h (7)
HS4
1: P1 (2), 2: C4h (7), 3: SQ5n (6), 4: SQ5n (6)
HS4p
1: P1 (2), 2: C4h (7), 3: SP2 (8), 4: SP2 (8)
HS6
1: SP1 (2), 2: Q5n (6), 3: SP1 (2)
HS6p
1: SP1 (2), 2: P2 (8), 3: SP1 (2)
HS7
1: SN6d (5), 2: TC5 (3), 3: TN3ar (4), 4: TN6 (9)
HS8
1: TC5 (3), 2: TC5 (3), 3: TN3ar (4), 4: TN6 (9)
HS9
1: SP1 (2), 2: Q5n (6), 3: TP1 (2)

HS9p

1: SP1 (2), 2: P2 (8), 3: TP1 (2)

HS11

1: P2a (10)

HS12

1: SN6 (9), 2: TC5 (3), 3: SQ5n (6), 4: SP1r (11), 5: TN3ar (4), 6: TC5e (12),

7: TN6d (5)

HS12p

1: SN6 (9), 2: TC5 (3), 3: SP2 (8), 4: SP1r (11), 5: TN3ar (4), 6: TC5e (12),

7: TN6d (5)

HS13

1: SN5a (1), 2: TP1 (2), 3: TP1 (2), 4: TP1 (2), 5: SQ5n (6)

HS13p

1: SN5a (1), 2: TP1 (2), 3: TP1 (2), 4: TP1 (2), 5: SP2 (8)

HS14

1: SN6 (9), 2: SN6 (9), 3: SQ5n (6), 4: SN6 (9), 5: TN4a (13)

HS14p

1: SN6 (9), 2: SN6 (9), 3: SP2 (8), 4: SN6 (9), 5: TN4a (13)

HS16

1: SN6 (9), 2: SC6 (14), 3: SQ5n (6), 4: TN3ar (4), 5: TN6 (9)

HS16p

1: SN6 (9), 2: SC6 (14), 3: SP2 (8), 4: TN3ar (4), 5: TN6 (9)

HS17

1: C1 (15)

HS18

1: SN5a (1), 2: TN4 (16)

HS19

1: SC2 (17), 2: SQ5n (6)

HS19p

1: SC2 (17), 2: SP2 (8)

HS20

1: TC5e (12), 2: TC5 (3), 3: SQ5n (6), 4: SQ5n (6), 5: TC5 (3)

HS20p

1: TC5e (12), 2: TC5 (3), 3: SP2 (8), 4: SQ5n (6), 5: TC5 (3)

HS20fp

1: TC5e (12), 2: TC5 (3), 3: SP2 (8), 4: SP2 (8), 5: TC5 (3)

HS21

1: SN6 (9), 2: TC5 (3), 3: N4a (18), 4: SN6 (9)

HS22

1: TC5e (12), 2: TN2a (19), 3: TC5e (12), 4: TN2a (19), 5: TN2a (19),

6: TC5 (3)

HS23

1: N5a (1)

HS24

1: SN5a (1), 2: SQ5n (6), 3: TN4 (16)

HS24p

1: SN5a (1), 2: SP2 (8), 3: TN4 (16)

HS25

1: SP1 (2), 2: SQ5n (6), 3: TN3ar (4)

HS25p

1: SP1 (2), 2: SP2 (8), 3: TN3ar (4)

HS26

1: P1 (2)

HS27

1: SN6 (9), 2: TC5 (3), 3: SQ5n (6), 4: SN6 (9)

HS27p

1: SN6 (9), 2: TC5 (3), 3: SP2 (8), 4: SN6 (9)

HS28

1: TC5 (3), 2: TN6a (20), 3: TC5 (3), 4: TN6a (20)

HS29

1: SN6 (9), 2: SN6 (9), 3: TC5 (3), 4: TC5 (3), 5: TC5e (12)

HS30

1: TC5e (12), 2: SQ5n (6), 3: SQ5n (6), 4: TC5 (3), 5: SQ5n (6), 6: TC5e (12)

HS30p

1: TC5e (12), 2: SQ5n (6), 3: SQ5n (6), 4: TC5 (3), 5: SP2 (8), 6: TC5e (12)

HS30fp

1: TC5e (12), 2: SP2 (8), 3: SP2 (8), 4: TC5 (3), 5: SP2 (8), 6: TC5e (12)

HS32

1: TC5 (3), 2: TC5 (3), 3: SQ5n (6), 4: SN6 (9)

HS32p

1: TC5 (3), 2: TC5 (3), 3: SP2 (8), 4: SN6 (9)

HS34

1: TC5 (3), 2: TC5 (3), 3: TC5 (3), 4: SN6 (9), 5: SQ5n (6), 6: SQ5n (6),
7: TN4 (16), 8: TC5e (12), 9: TC5e (12), 10: TN2a (21)
HS34p
1: TC5 (3), 2: TC5 (3), 3: TC5 (3), 4: SN6 (9), 5: SP2 (8), 6: SP2 (8),
7: TN4 (16), 8: TC5e (12), 9: TC5e (12), 10: TN2a (21)

HS35

1: SN5a (1), 2: TN4 (16), 3: SQ5n (6), 4: C2 (17)

HS35p

1: SN5a (1), 2: TN4 (16), 3: SP2 (8), 4: C2 (17)

1: Ketone (N5a)

2: Alcohol (P1)

3: Aromatics (C5)

4: Ether (end-group, TN3ar)

5: Primary amide (N6d)

6: Carboxylate (Q5n/SQ5n)

7: Alkene (C4), with higher self-interaction (h), like Butene[1]

8: Carboxylic acid (P2)

9: Phenol (N6)

10: HS11 (matched to transfer free energies, P2a)

11: Group containing Ether and Hydroxy group (inspired by carbohydrates[2], SP1r)

12: Aromatic (no hydrogens, C5e)

13: Ketone connected to aromatic compound

(inspired by Aldehyde group from Benzaldehyde[1], TN4a)

14: Thiol (C6)

15: Linear Alkane (C1)

16: Secondary Amide (N4)

17: Branched Alkane (C2)

18: Methyl-acetate group (inspired by Methyl-benzoate[1], N4a)

19: Methoxy group (inspired by Methoxybenzene[1], TN2a)

20: Ketone group within the aromatic compound (inspired by Benzoquinone[1], TN6a)

21: Ether group within the aromatic compound (inspired by Furan[1], TN2a)

Table S2: Molar fractions (x) and mass-weighted fractions (w) of LHA molecules, water, and calcium ions for different water contents, averaged over the three considered LHA systems. N_{water} is the number of water molecules in the systems.

	LHA		water		calcium	
N_{water}	x	w	x	w	x	w
500	0.775	0.722	0.196	0.202	0.029	0.076
600	0.745	0.695	0.226	0.232	0.028	0.073
700	0.718	0.671	0.254	0.259	0.027	0.070
800	0.693	0.648	0.280	0.284	0.026	0.067
900	0.670	0.628	0.305	0.308	0.025	0.065
1000	0.648	0.609	0.327	0.329	0.024	0.062
1100	0.627	0.591	0.349	0.349	0.024	0.060
1200	0.608	0.574	0.369	0.368	0.023	0.058
1300	0.590	0.559	0.388	0.385	0.022	0.056
1400	0.573	0.545	0.406	0.401	0.022	0.054
1500	0.557	0.531	0.422	0.416	0.021	0.052

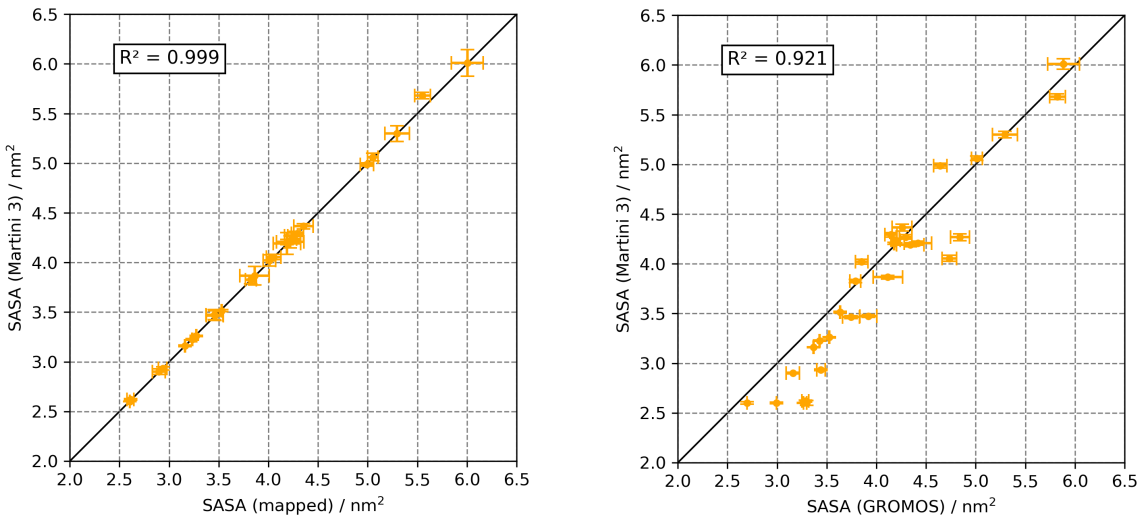


Figure S2: Comparison of time-averaged SASA of coarse-grained fragments with the mapped fragments (left) and with the atomistic fragments, using van der Waals radii from the GROMOS force field (right). An error bar denotes the standard deviation of an averaged value.

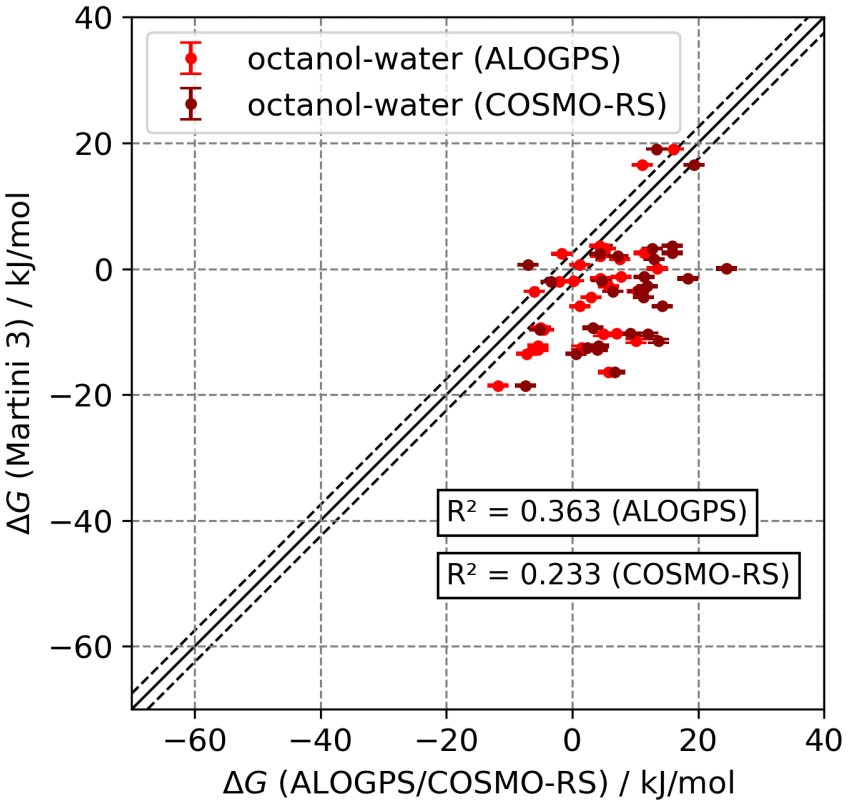


Figure S3: Comparison of octanol-water transfer free energies ΔG of coarse-grained fragments (Martini 3) with the reference. The reference values were calculated via partition coefficients from ALOGPS and from COSMO-RS.

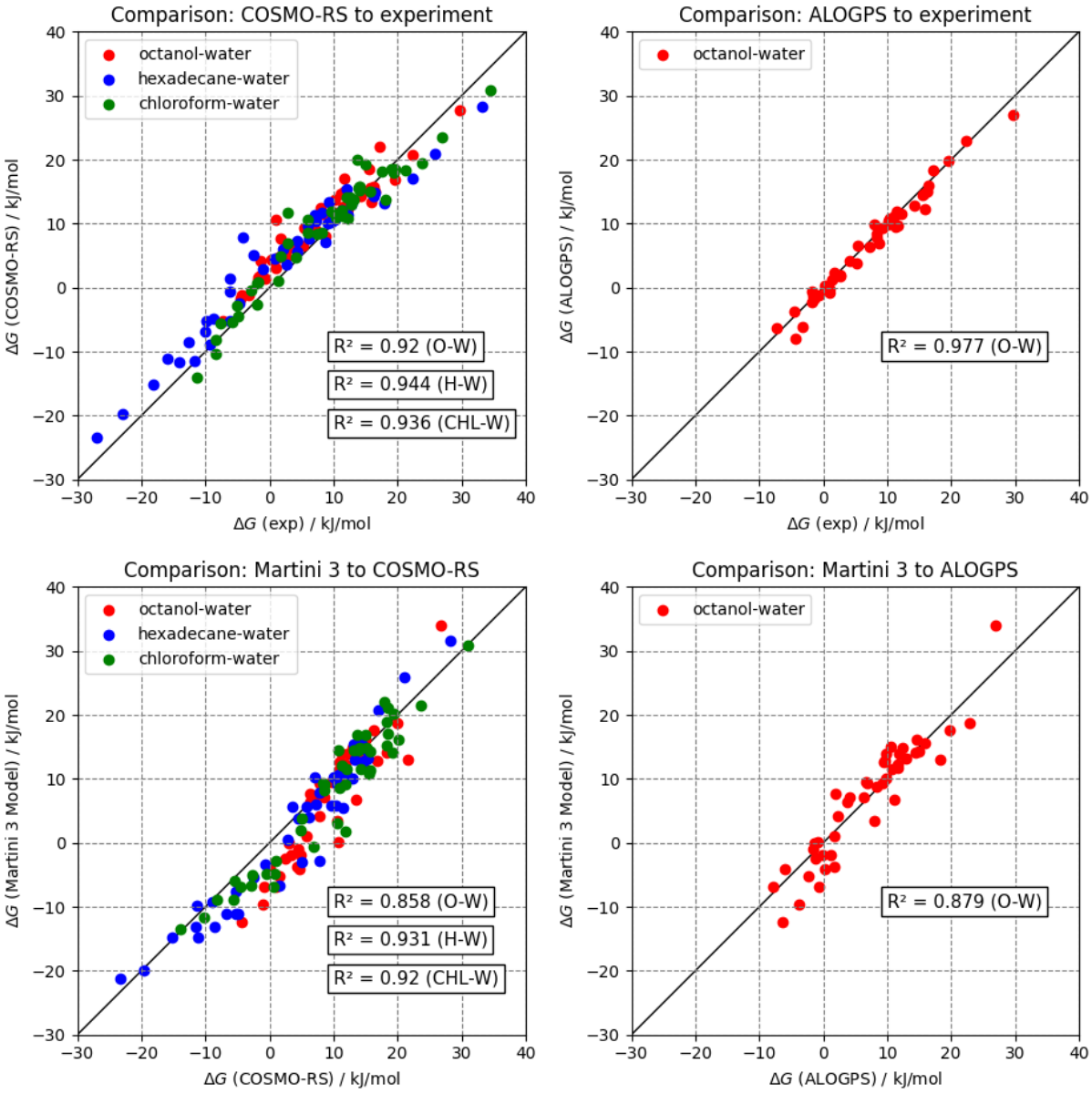


Figure S4: Comparison of transfer free energies ΔG for selected molecules from the Martini 3 main publication. Comparisons were made between COSMO-RS to experiment, ALOGPS to experiment, Martini 3 to COSMO-RS and Martini 3 to ALOGPS.

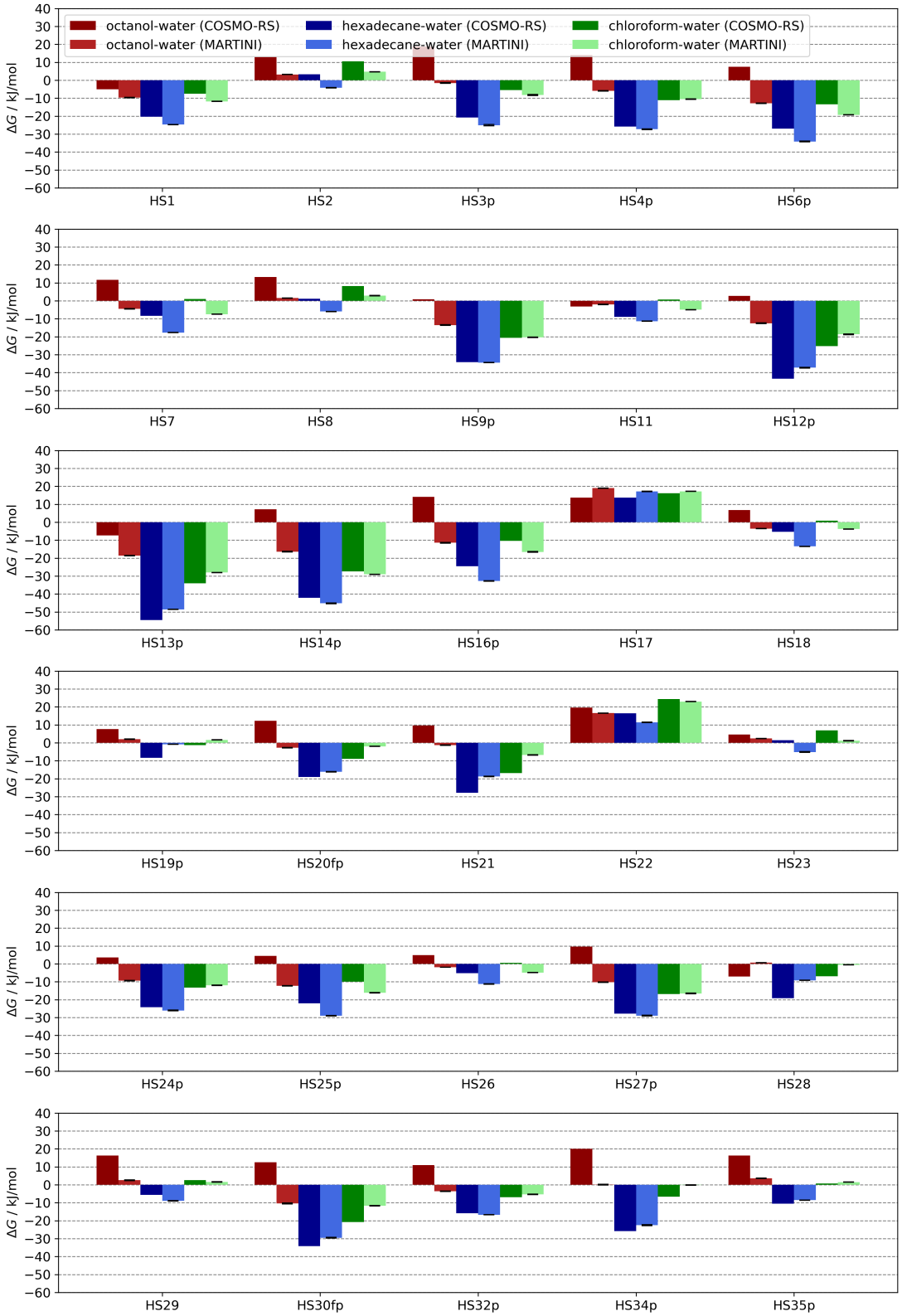


Figure S5: Transfer free energy values, shown separately for each fragment as a bar plot.

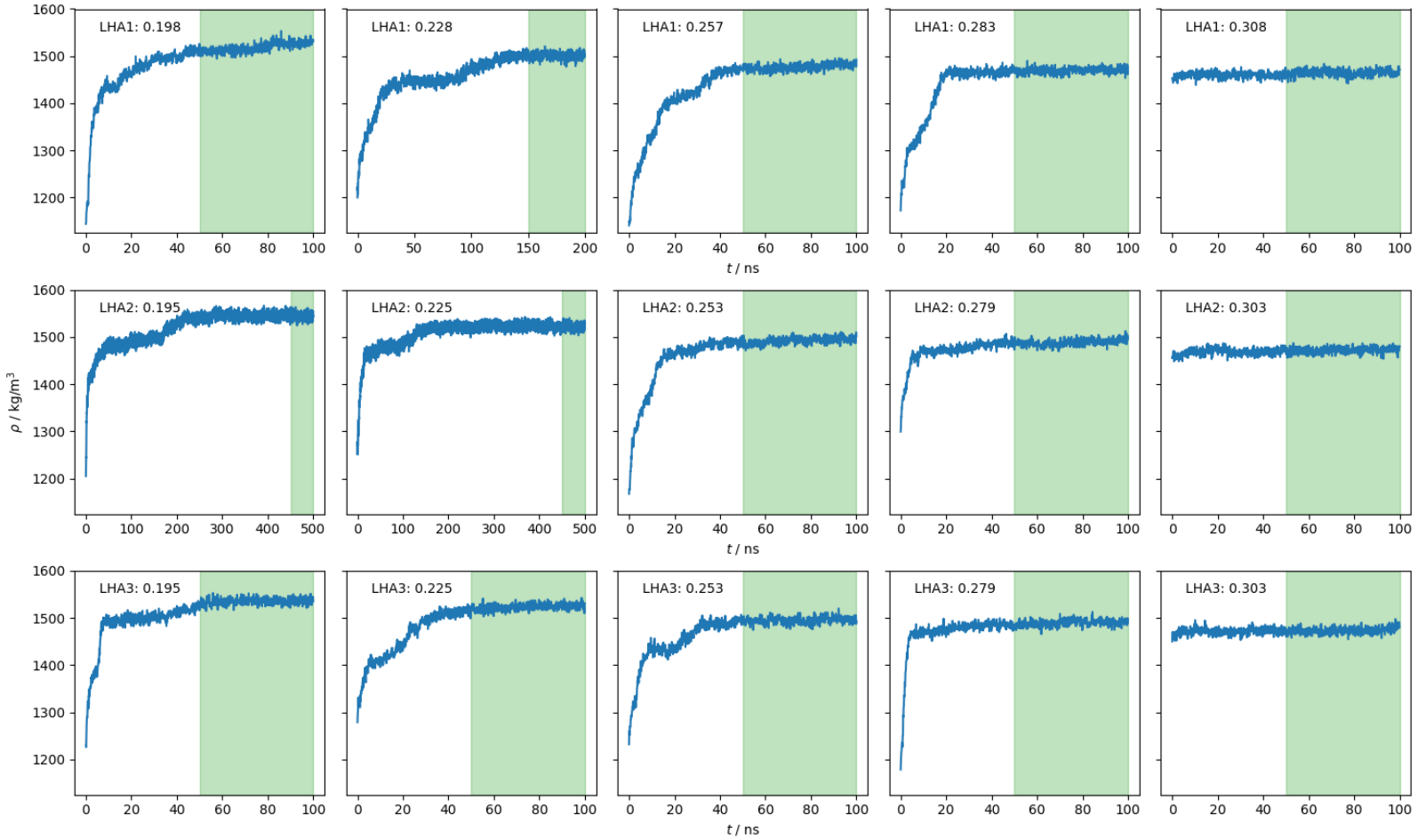


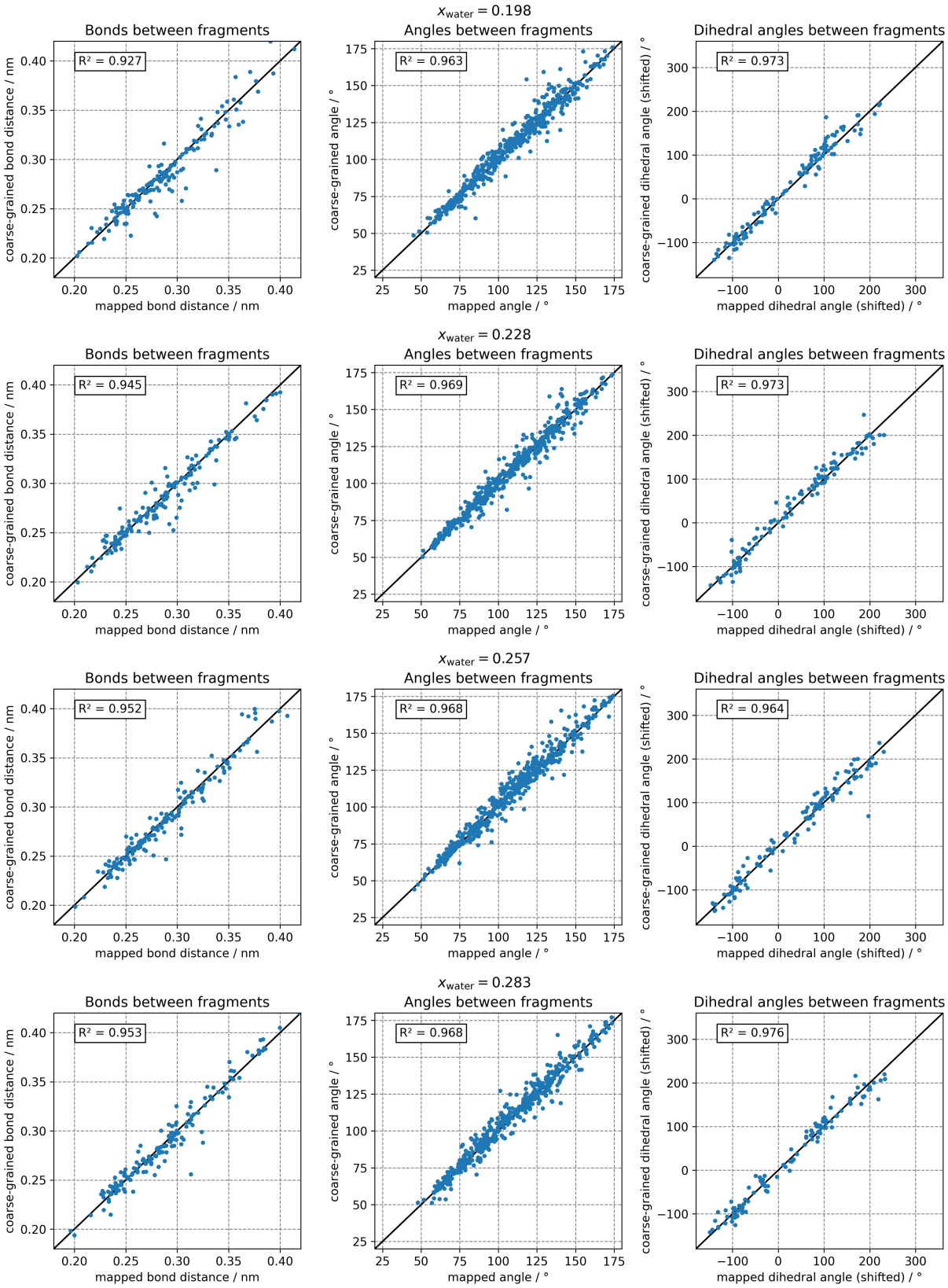
Figure S6: Time evolutions of the densities for the three LHA systems. Shown here are water contents from low to medium, to illustrate the longer relaxation times, especially for systems with lower water content. The green area marks the respective simulation section that was used to determine the mean densities.

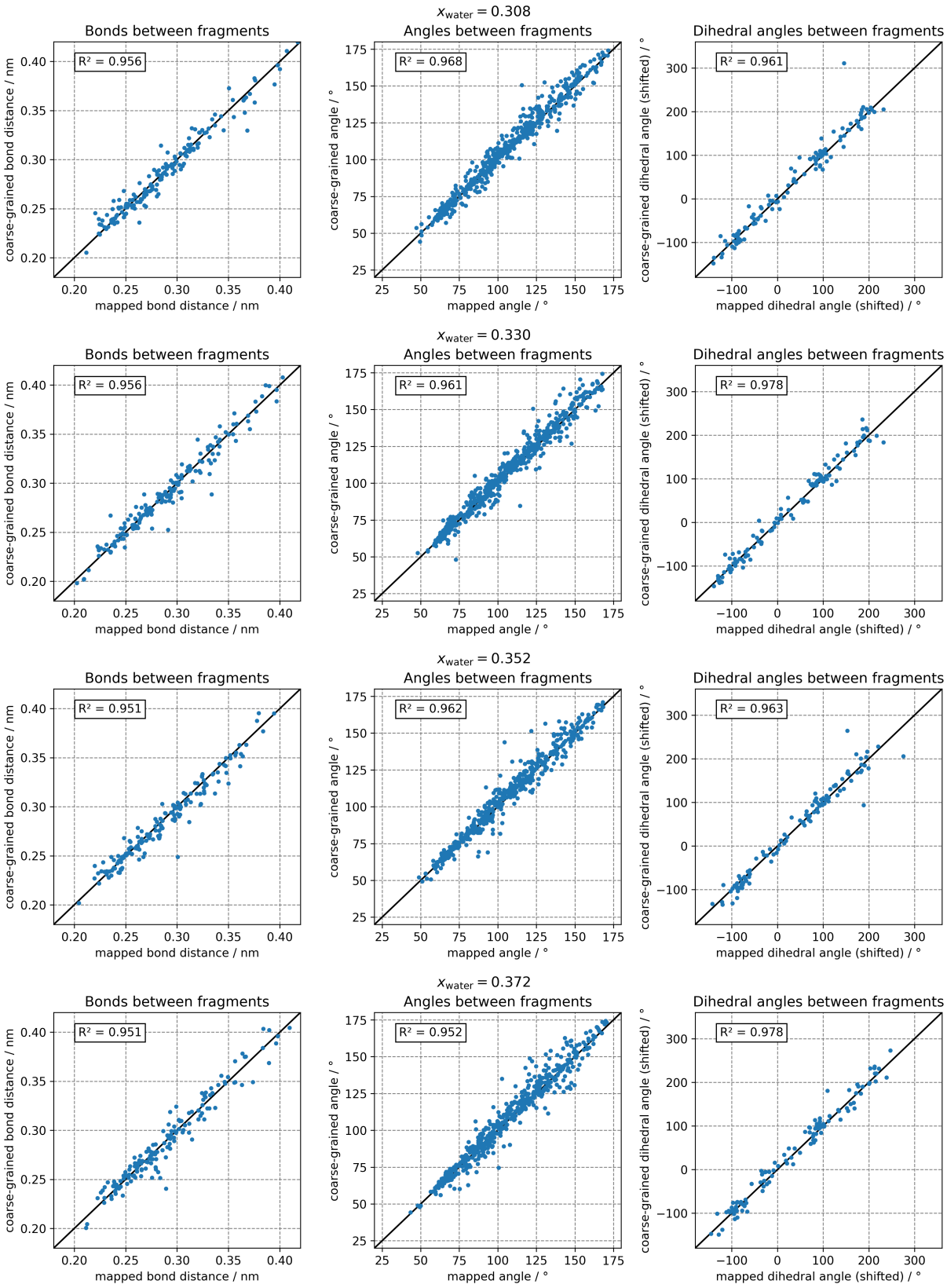
S1 Direct Boltzmann Inversion

(Direct) Boltzmann Inversion (DBI)[3] was applied to obtain bonded interaction potentials between different fragments from atomistic simulations. The potential $U(q)$ and the probability distribution $P(q)$ of a corresponding bonded interaction are linked in the following way: $U(q) = -k_B T \ln P(q)$, where q is either a bond distance r, an angle θ or a dihedral angle φ[4]. To obtain the probability distributions, histograms $H(q)$ had to be rescaled, depending on the considered interaction type: $P(r) = H(r)/4\pi r^2$, $P(\theta) = H(\theta)/sin(\theta)$, $P(\varphi) = H(\varphi)$. Using a Python script and MDAnalysis[5,6], the DBI procedure was implemented and conducted in the following way. The histograms were created with 50 bins for bonds and angles, and 100 bins for dihedral angles. After rescaling the histograms, the probability distributions could be determined. Exponential smoothing was applied to the distributions, with a smoothing parameter $\alpha = 0.55$. Afterward, the potential $U(q)$ was calculated with the aforementioned equation. To obtain GROMACS-compatible parameters, the potential was fitted using the SciPy package[7]. For the fitting procedure, the potentials that corresponded to the respective interaction type were used. To improve numerical stability, the following modifications were applied to the fitting parameters. The force constants of the bonds were capped, depending on the mass of the involved beads, such that the estimated period of the oscillation was greater than twenty times of the applied time step. Furthermore, the force constants of the restricted bending potentials and the proper dihedral angle potentials were limited to $1200\,\mathrm{kJ\,mol^{-1}}$ and $35\,\mathrm{kJ\,mol^{-1}}$ respectively. The restricted angle potentials were converted into harmonic angle potentials if the equilibrium angle obtained from the fit was less than 30° or greater than 150°. In this case, the dihedral angle potential was removed. In the case of bimodal distributions, the harmonic potential was fitted to the global minimum to prevent force constants that were too low. The overall approach was based on Swarm-CG's methods for obtaining initial interaction parameters for the optimization procedure. Further details on the implementation in Swarm-CG can be taken from the supporting information of the corresponding publication[8].

S2 Automated determination of bonded interactions between fragments

In general, the bonded interactions between fragments were determined based on their connections. The bonds between two beads, which connected two fragments, were set according to the underlying atomistic connection. For example, if two groups of atoms from different fragments were connected via a bond, the corresponding beads were also connected. This also defines which of the beads of a coarse-grained fragment have the first or the last index. For some fragments, a bond to the previous fragment will be connected to the same bead, as a bond to the next fragment (see e.g. HS3, or HS19). The angles were then determined between beads that connect different fragments and which are the nearest neighbors to these beads. In other words, an angle was defined between three consecutive beads where at least one of the bonds was a bond connecting different fragments. Finally, the dihedral angles were defined via pairs of the aforementioned angles. For example, if an angle potential was applied between the beads ijk and another one between jkl, a dihedral potential was applied between the beads $ijkl$.


$x_{water} = 0.198$
Bonds between fragments
R² = 0.927
coarse-grained bond distance / nm
mapped bond distance / nm
Angles between fragments
R² = 0.963
coarse-grained angle / °
mapped angle / °
Dihedral angles between fragments
R² = 0.973
coarse-grained dihedral angle (shifted) / °
mapped dihedral angle (shifted) / °
$x_{water} = 0.228$
Bonds between fragments
R² = 0.945
coarse-grained bond distance / nm
mapped bond distance / nm
Angles between fragments
R² = 0.969
coarse-grained angle / °
mapped angle / °
Dihedral angles between fragments
R² = 0.973
coarse-grained dihedral angle (shifted) / °
mapped dihedral angle (shifted) / °
$x_{water} = 0.257$
Bonds between fragments
R² = 0.952
coarse-grained bond distance / nm
mapped bond distance / nm
Angles between fragments
R² = 0.968
coarse-grained angle / °
mapped angle / °
Dihedral angles between fragments
R² = 0.964
coarse-grained dihedral angle (shifted) / °
mapped dihedral angle (shifted) / °
$x_{water} = 0.283$
Bonds between fragments
R² = 0.953
coarse-grained bond distance / nm
mapped bond distance / nm
Angles between fragments
R² = 0.968
coarse-grained angle / °
mapped angle / °
Dihedral angles between fragments
R² = 0.976
coarse-grained dihedral angle (shifted) / °
mapped dihedral angle (shifted) / °


$x_{water} = 0.308$
Bonds between fragments
R² = 0.956
coarse-grained bond distance / nm
mapped bond distance / nm
Angles between fragments
R² = 0.968
coarse-grained angle / °
mapped angle / °
Dihedral angles between fragments
R² = 0.961
coarse-grained dihedral angle (shifted) / °
mapped dihedral angle (shifted) / °
$x_{water} = 0.330$
Bonds between fragments
R² = 0.956
coarse-grained bond distance / nm
mapped bond distance / nm
Angles between fragments
R² = 0.961
coarse-grained angle / °
mapped angle / °
Dihedral angles between fragments
R² = 0.978
coarse-grained dihedral angle (shifted) / °
mapped dihedral angle (shifted) / °
$x_{water} = 0.352$
Bonds between fragments
R² = 0.951
coarse-grained bond distance / nm
mapped bond distance / nm
Angles between fragments
R² = 0.962
coarse-grained angle / °
mapped angle / °
Dihedral angles between fragments
R² = 0.963
coarse-grained dihedral angle (shifted) / °
mapped dihedral angle (shifted) / °
$x_{water} = 0.372$
Bonds between fragments
R² = 0.951
coarse-grained bond distance / nm
mapped bond distance / nm
Angles between fragments
R² = 0.952
coarse-grained angle / °
mapped angle / °
Dihedral angles between fragments
R² = 0.978
coarse-grained dihedral angle (shifted) / °
mapped dihedral angle (shifted) / °

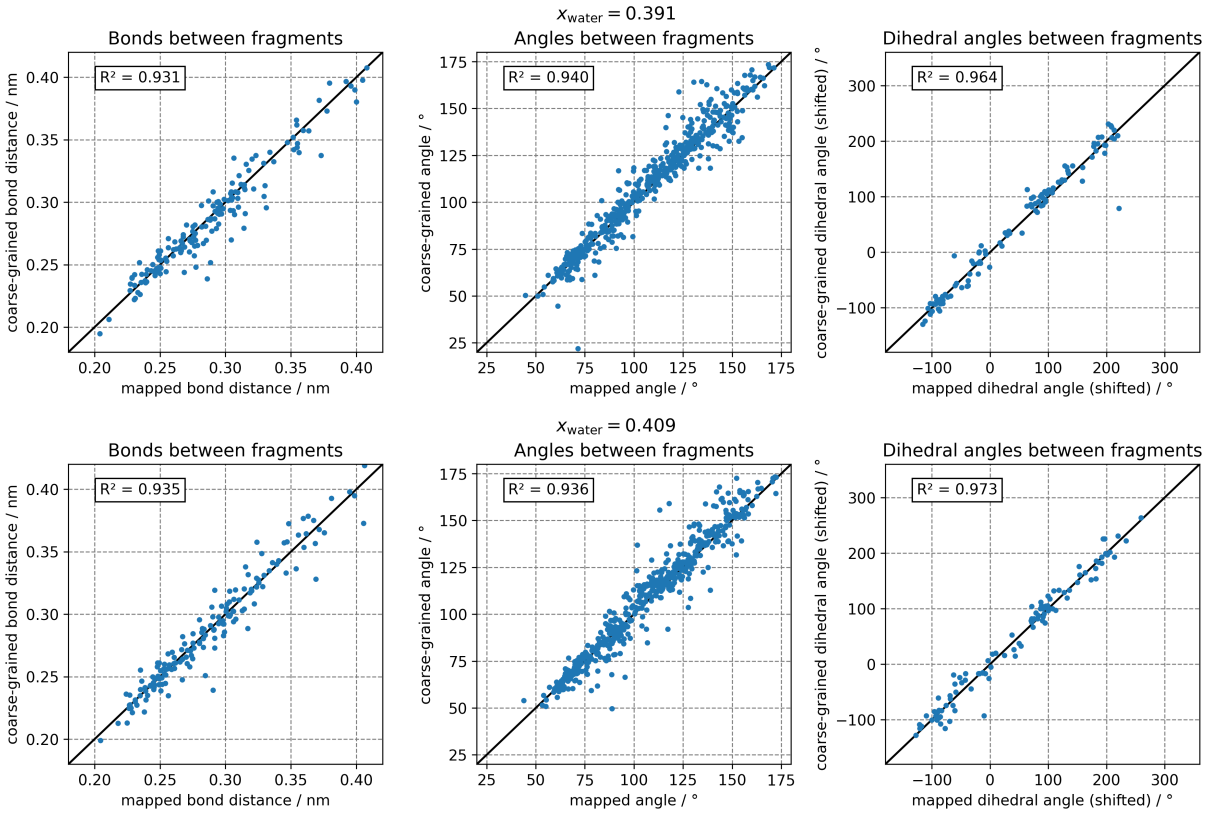


Figure S7: Comparison of average values of bonded interactions between a coarse-grained LHA structure and the mapped version of its atomistic counterpart. Different water contents x_{water} are considered. Bonded interactions are divided into bonds, angles, and dihedral angles between different fragments.

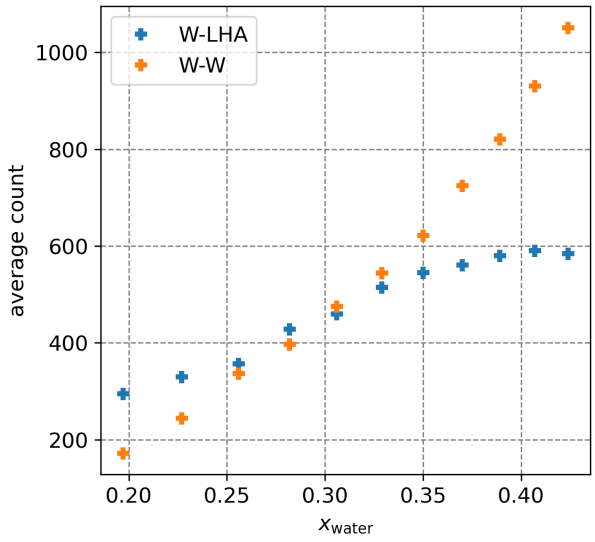


Figure S8: Time-averaged number of hydrogen bonds in the system, depending on the water content. A distinction was made between hydrogen bonds between the LHA molecules and water (W-LHA), and hydrogen bonds between the water molecules themselves (W-W).

References

(1) 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.

(2) 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.

(3) Tschöp, W.; Kremer, K.; Batoulis, J.; Bürger, T.; Hahn, O. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polymerica* **1998**, *49*, 61–74.

(4) Junghans, C.; Andrienko, D.; Votca Bot,; Rosenberger, D.; Krajniak, J.; Marvin-bernhardt,; Abrukhno,; Sikandar Y. Mashayak,; Bereau, T.; AlexanderWi,; , Jens,; Espinosadeoliveira, votca/csg-manual: v1.6.4. 2021; https://zenodo.org/record/4435216.

(5) 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.

(6) 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.

(7) 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.

(8) 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.

JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

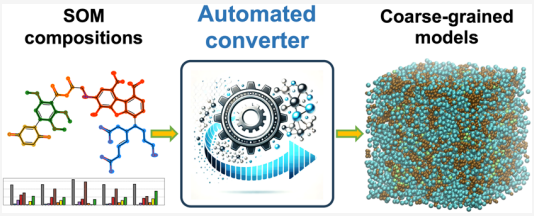
Automated Parametrization Approach for Coarse-Graining Soil Organic Matter Molecules

Lorenz F. Dettmann, Oliver Kühn, and Ashour A. Ahmed*

Cite This: J. Chem. Theory Comput. 2024, 20, 10684−10696 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

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 parametrization 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.



Downloaded via UNIV ROSTOCK on July 17, 2025 at 12:40:47 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

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.

Received: October 7, 2024
Revised: November 19, 2024
Accepted: November 20, 2024
Published: November 29, 2024



ACS Publications

© 2024 The Authors. Published by American Chemical Society

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, nonamide 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] 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 parametrization protocol. Instead of using atomistic simulations of SOM structures of VSOMM2 for the parametrization of the bonded interactions, we employ a conformer generation method known as experimental-torsion basic knowledge distance geometry (ETKDG).[55] Our new parametrization 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 model systems using a set of fragment molecules.[6,54] The parametrization approach utilized Swarm-CG[56] to optimize the bonded interaction

https://doi.org/10.1021/acs.jctc.4c01334
J. Chem. Theory Comput. 2024, 20, 10684−10696

parameters of the isolated fragments and to generate the corresponding coarse-grained structures. These fragments were then assembled into larger macromolecules[a], 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 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 parametrization 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 parametrization 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.

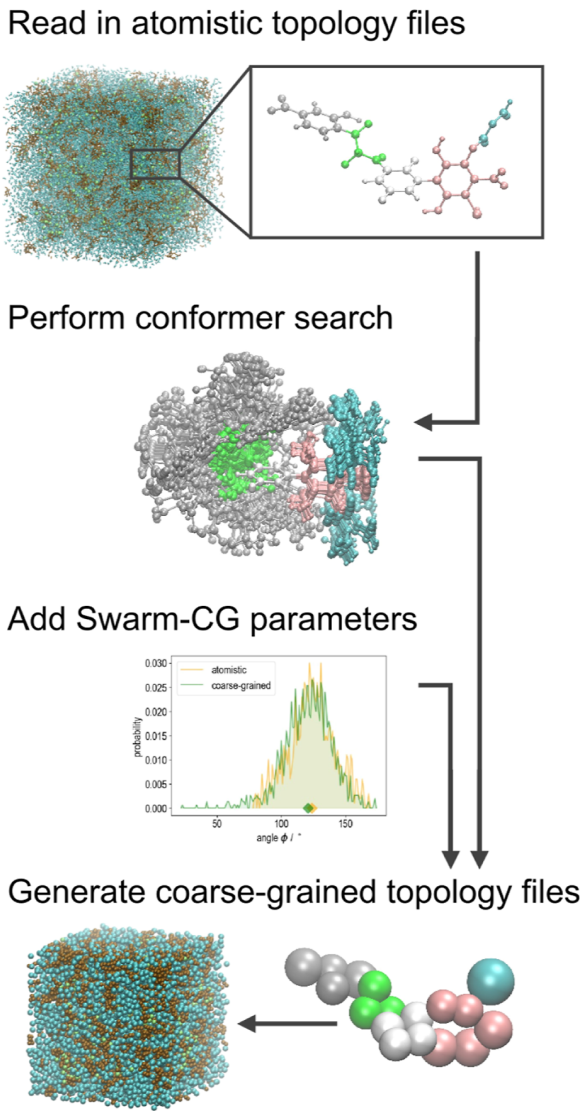


Figure 1. Graphical illustration of the modified parametrization approach. The tool that was developed reads the atomistic topology files. Subsequently, equilibrium values are determined using the conformer search (ETKDG) and augmented with values obtained from the Swarm-CG parametrization. Finally,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

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 IHSS via a principal component (PC) analysis, using aromaticity and acidity as PCs. 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.

https://doi.org/10.1021/acs.jctc.4c01334
J. Chem. Theory Comput. 2024, 20, 10684−10696

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 heteroaliphatic and aliphatic content

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 parametrization 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 parametrization 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.

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 leapfrog 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 τ_T = 1 ps. For the *NPT* equilibration, the Berendsen barostat[68] with a coupling constant of τ_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 cutoff of 1.1 nm with a dielectric constant of ε_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 parametrization 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 parametrization 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).

3.1. Benchmarking the Modified Parametrization Approach. *3.1.1. Revised Force Constants.* 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 cm^{-1} (orange) to approximately 200 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 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 leapfrog algorithm) and directly impacts the conservation of key quantities such as energy, temperature, and pressure.[72] A

https://doi.org/10.1021/acs.jctc.4c01334
J. Chem. Theory Comput. 2024, 20, 10684–10696

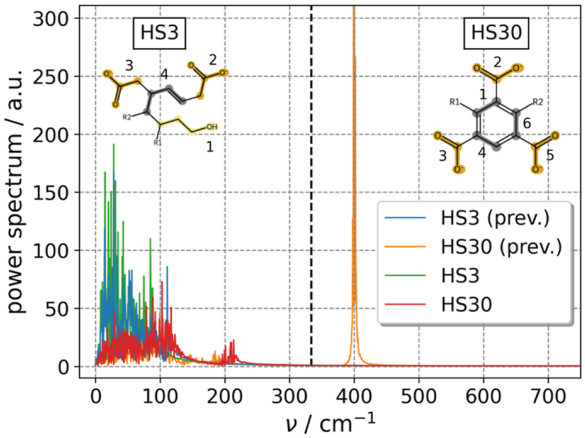


Figure 2. Power spectra of the velocity autocorrelation function for HS3 and HS30 fragments, obtained using the previous (prev.) parametrization approach (blue and orange) and the current restricted parametrization approach employed in this work (green and red). The colored groups of fragment molecules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

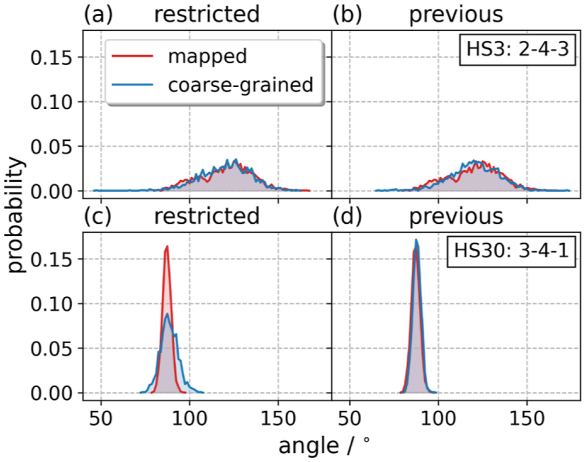


Figure 3. Comparison of the angle distributions of mapped (red) and coarse-grained (blue) models. The distributions of the angle between the beads 2, 4, and 3 (2−4−3) of the HS3 fragment are shown for the restricted parametrization (a) and previous parametrization (b). Similarly, the angle distributions 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.

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, 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 parametrization 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 3a,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 distribution to the reference for the restricted model, as observed in Figure 3c. 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 parametrization, 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

https://doi.org/10.1021/acs.jctc.4c01334
J. Chem. Theory Comput. 2024, 20, 10684−10696

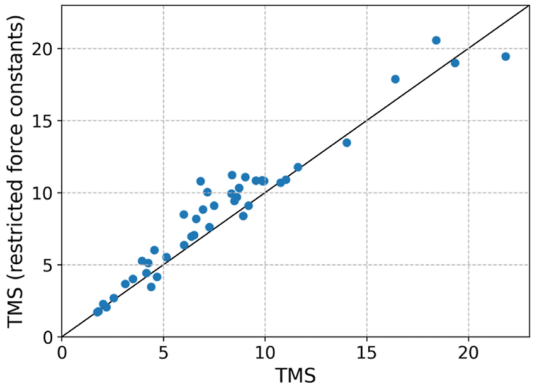


Figure 4. Comparison of the total mismatch scores (TMSs) of the fragments from the Swarm-CG parametrization. The values from the previous parametrization (*x*-axis) are plotted against those from the modified parametrization in this study (*y*-axis). Each point represents a comparison of the two TMS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

different points. A higher position in the upper half of the graph indicates a larger TMS value for the restricted parametrization 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. 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

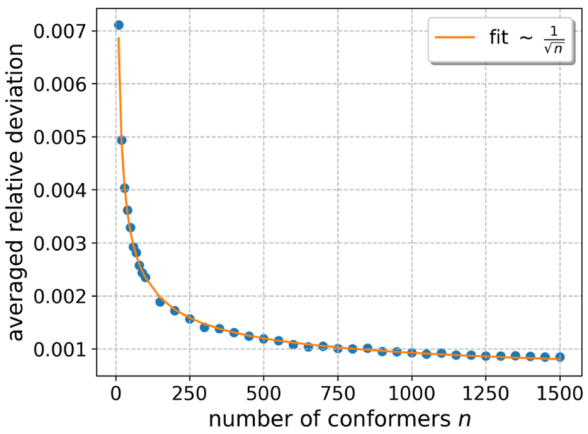


Figure 5. Convergence of the bond values for the SOM III model system as a function of the number of conformers. The plot depicts the relative deviation from the mean value, averaged across all bonded terms within the model system.

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 ‰ (0.001). 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 constructing 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 time frame. 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 μs 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) 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-

https://doi.org/10.1021/acs.jctc.4c01334
J. Chem. Theory Comput. 2024, 20, 10684−10696

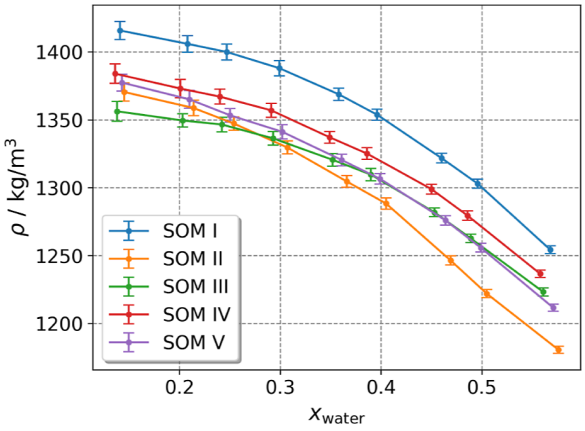


Figure 6. Average of system densities for the SOM types calculated for different water contents x_{water} in the systems; error bars denote standard deviations.

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 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 nonbonded 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

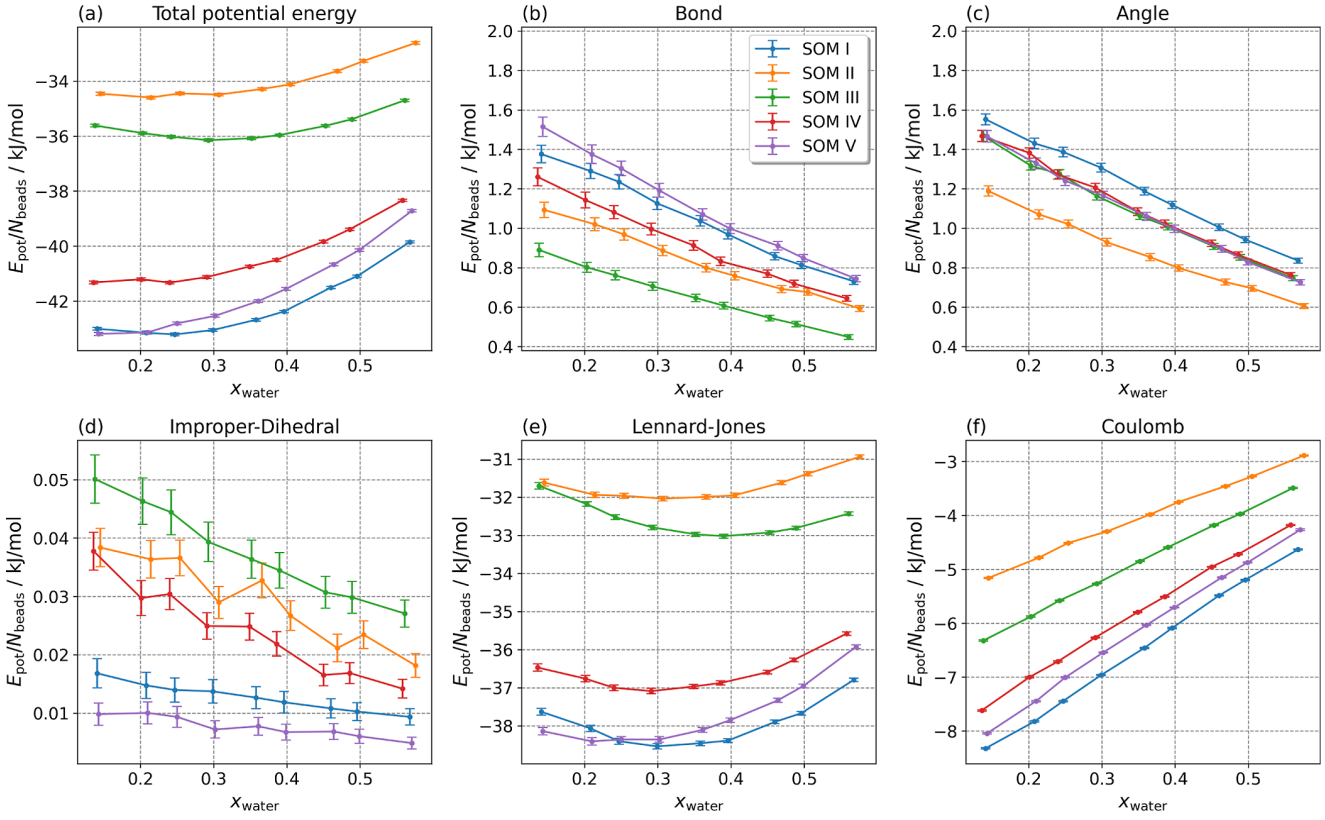


Figure 7. Time averaged total potential energies (a), and potential energies of bonded (b−d) and nonbonded interactions (e, f) for the different SOM models; error bars denote standard deviations. Please note the different scales of the *y*-axes.

https://doi.org/10.1021/acs.jctc.4c01334
J. Chem. Theory Comput. 2024, 20, 10684−10696

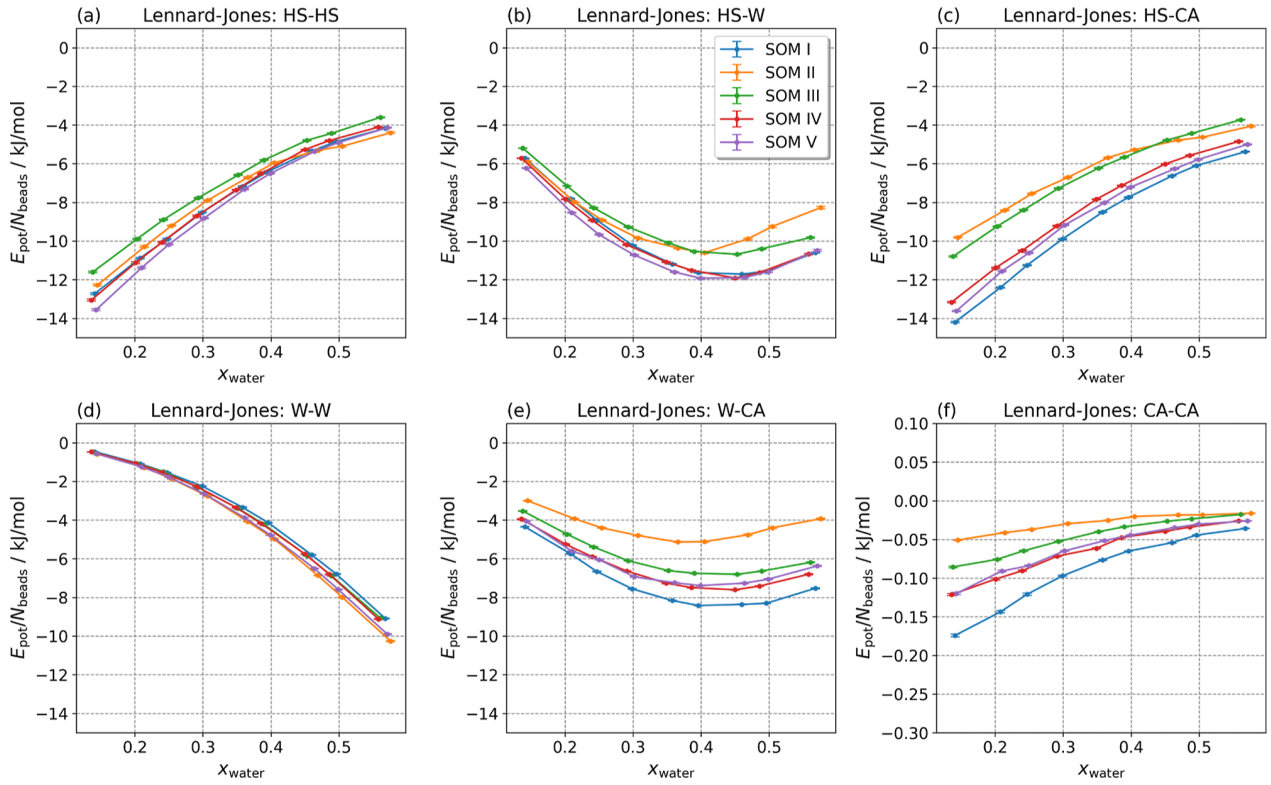


Figure 8. Time averaged Lennard-Jones potential energies for the different pairs of interactions between HS molecules, water (W) and calcium (CA); error bars denote standard deviations.

that the individual potential types can be regarded as a true division of the total potential energy.

Concerning the total potential energy in Figure 7a, 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 nonbonded 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 7b), angles (Figure 7c), and improper dihedral angles (Figure 7d) 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 nonbonded 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 7e 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_{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 7f), 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 possible 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

https://doi.org/10.1021/acs.jctc.4c01334
J. Chem. Theory Comput. 2024, 20, 10684−10696

different curves is most apparent in the HS-calcium (HS-CA) in (Figure 8c) and water-calcium (W-CA) Lennard-Jones interactions (Figure 8e). 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 8a) and HS−CA (Figure 8c) 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 8d). 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 8b) and W−CA Lennard-Jones interactions (Figure 8e) 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 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

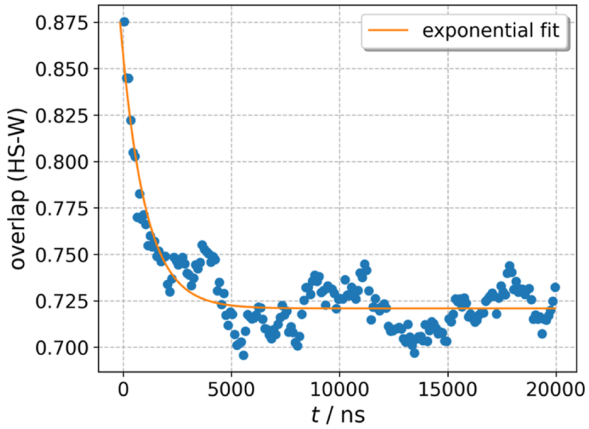


Figure 9. Overlap between normalized partial densities of HS and water over simulation time, in *y*-direction. SOM II with the highest water content is considered. Please note that the exponential fit was applied to obtain the average value, which the partial density oscillates around.

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 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.

3.3. Numerical Performance. To assess the improvement in numerical performance, measured as the length of trajectory produced within a given computational time, benchmarking

https://doi.org/10.1021/acs.jctc.4c01334
J. Chem. Theory Comput. 2024, 20, 10684−10696

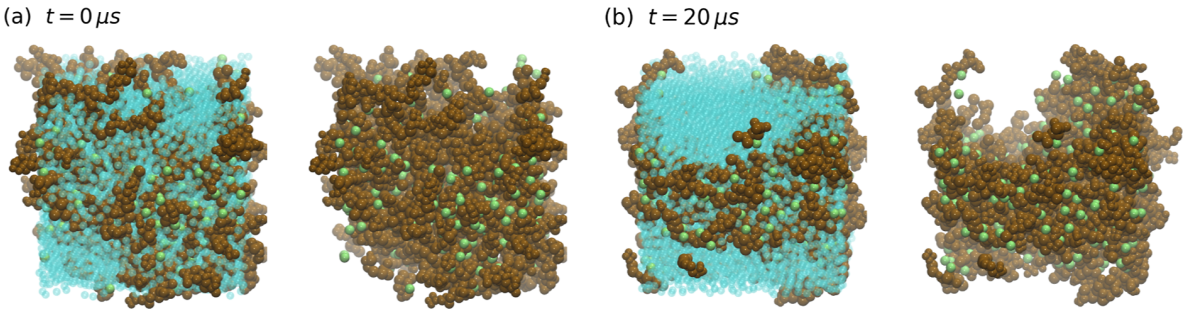


Figure 10. Snapshots from the trajectory of SOM II with the highest water content, at 0 μs (a) and after 20 μs (b). HS (brown), water (cyan) and calcium (green) are depicted. Each snapshot is shown with and without water, respectively.

simulations were performed using both atomistic and coarse-grained approaches. The performance ratio between coarse-grained and atomistic models, P_{cg}/P_{at} (Figure 11a), was therefore calculated, along with the speedup of the corresponding systems (Figure 11b). The speedup is defined as the ratio of serial to parallel runtime.

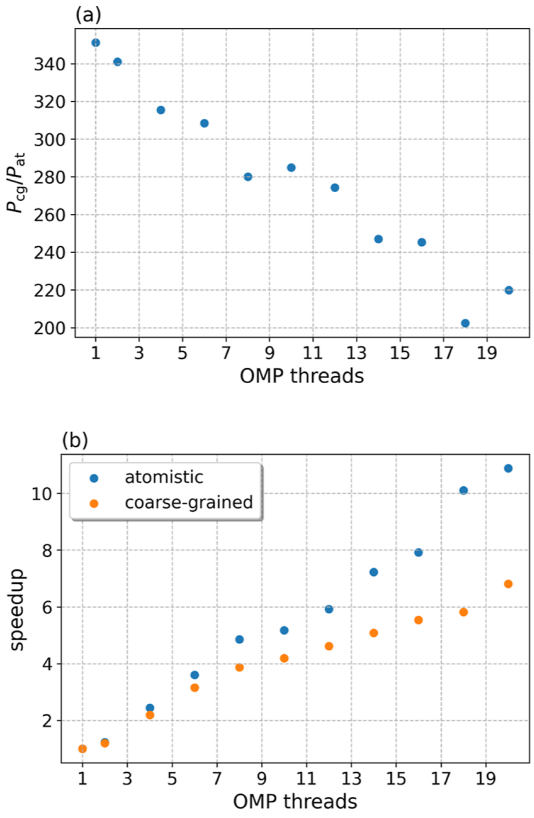


Figure 11. Performance calculations for SOM III with the highest water content, comparing atomistic and coarse-grained representation. The ratio of coarse-grained to atomistic performance is shown in (a). The speedup for different OMP Threads is depicted in (b).

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 11a, 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 11b, 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 2 orders of magnitude can be observed. This performance improvement can be primarily attributed to two factors. First, the time step 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 10-fold increase in time step, 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

https://doi.org/10.1021/acs.jctc.4c01334
J. Chem. Theory Comput. 2024, 20, 10684−10696

addition, it provides the coarse-grained topology files in a more efficient and user-friendly manner.

We first assessed how the modified parametrization 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 nonbonded 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 the Supporting Information of Escalona et al.[6] or 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 parametrization 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/doi/10.1021/acs.jctc.4c01334.

It includes details about the modified parametrization approach and conformer search, as well as additional information about the SOM compositions and further benchmark results (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Ashour A. Ahmed − *University of Rostock, Institute of Physics, D-18059 Rostock, Germany; Leibniz Institute for Catalysis (LIKAT), D-18059 Rostock, Germany;* orcid.org/0000-0002-5933-5249; Email: ashour.ahmed@uni-rostock.de

Authors

Lorenz F. Dettmann − *University of Rostock, Institute of Physics, D-18059 Rostock, Germany;* orcid.org/0000-0001-8407-4469

Oliver Kühn − *University of Rostock, Institute of Physics, D-18059 Rostock, Germany;* orcid.org/0000-0002-5132-2961

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.4c01334

Notes

The authors declare no competing financial interest.

■ ADDITIONAL NOTE

[a]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.

■ 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. *Environ. Sci. Technol.* **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. *Annu. Rev. Ecol. Evol. Syst.* **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 Manage.* **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. *Philos. Trans. R. Soc., B* **2008**, *363*, 685−701.

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

(7) Paul, E. A. The nature and dynamics of soil organic matter: Plant inputs, microbial transformations, and organic matter stabilization. *Soil Biol. Biochem.* **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 Res.* **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) *Molecular Modeling of Geochemical Reactions*; Kubicki, J. D., Ed.; 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. In *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 Sci.* **1997**, *162*, 115−130.
(15) Schulten, H.-R.; Leinweber, P. New insights into organic-mineral particles: composition, properties and models of molecular structure. *Biol. Fertil. Soils* **2000**, *30*, 399−432.
(16) Schulten, H.-R. Soil Mineral-Organic Matter-Microorganism Interactions and Ecosystem Health. In *Dynamics, Mobility and Transformation of Pollutants and Nutrients*; Elsevier, 2002; pp 351−381.
(17) Schulten, H. R.; Plage, B.; Schnitzer, M. A chemical structure for humic substances. *Sci. Nat.* **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. *J. Environ. Qual.* **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 Sci.* **2012**, *2012*, 1−10.
(21) Schulten, H.-R. Analytical pyrolysis and computational chemistry of aquatic humic substances and dissolved organic matter. *J. Anal. Appl. Pyrolysis* **1999**, *49*, 385−415.
(22) Schulten, H. Interactions of dissolved organic matter with xenobiotic compounds: Molecular modeling in water. *Environ. Toxicol. Chem.* **1999**, *18*, 1643−1655.
(23) Sutton, R.; Sposito, G.; Diallo, M. S.; Schulten, H. Molecular simulation of a model of dissolved organic matter. *Environ. Toxicol. Chem.* **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. *Eur. J. Soil Sci.* **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. *J. Phys. Chem. 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. *Sci. Total Environ.* **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. *Sci. Total Environ.* **2014**, *476−477*, 98−106.
(29) Ahmed, A. A.; Leinweber, P.; Kühn, O. A computational study of hexachlorobenzene-soil organic matter-interactions. *J. Theor. Comput. Chem.* **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. *Sci. Total Environ.* **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. *J. Mol. Graphics Modell.* **2015**, *62*, 253−261.
(35) Piccolo, A. The supramolecular structure of humic substances. *Soil Sci.* **2001**, *166*, 810−832.
(36) Piccolo, A. *Advances in Agronomy*; Elsevier, 2002; pp 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 water molecule bridges in soil organic matter. *J. Soils 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. *J. Plant Nutr. Soil Sci.* **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. *Environ. Sci. Technol.* **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. *Eur. J. Soil Sci.* **2020**, *71*, 831−844.
(42) Gotsmy, M.; Escalona, Y.; Oostenbrink, C.; Petrov, D. Exploring the structure and dynamics of proteins in soil organic matter. *Proteins: Struct., Funct., Bioinf.* **2021**, *89*, 925−936.
(43) Escalona, Y.; Petrov, D.; Oostenbrink, C. Modeling soil organic matter: Changes in macroscopic properties due to microscopic changes. *Geochim. Cosmochim. 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* **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. *J. Chem. Theory Comput.* **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. *J. Chem. Theory Comput.* **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*, No. 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. *J. Phys. Chem. 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. *Environ. Sci. Technol.* **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. *Environ. Sci. Technol.* **2024**, *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. *Nat. 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. *J. Chem. Theory Comput.* **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.:Processes Impacts* **2023**, *25*, 1082−1093.
(59) Dettmann, L. F.; Kühn, O.; Ahmed, A. Topology files and Python tool for "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. *Comput. Sci. Eng.* **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. In *Proceedings of the Python in Science Conference*; SciPy, 2016.
(64) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011**, *32*, 2319−2327.
(65) Pérez, F.; Granger, B. E. IPython: a System for Interactive Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 21−29.
(66) Humphrey, W.; Dalke, A.; Schulten, K. VMD − Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(67) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **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. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(70) Poger, D.; Van Gunsteren, W. F.; Mark, A. E. A new force field for simulating phosphatidylcholine bilayers. *J. Comput. Chem.* **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. *Eur. Biophys. J.* **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. *Phys. Chem. Chem. Phys.* **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. *J. Chem. Theory Comput.* **2024**, *20*, 5763−5773.
(74) Rubner, Y.; Tomasi, C.; Guibas, L. J. A metric for distributions with applications to image databases. In *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*; Narosa Publishing House, 2002.

Supporting Information

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

S1 Details of the modified parametrization approach

The modified parametrization approach enables coarse-grained SOM models from VSOMM2 topology files to be generated through an automated pipeline. The corresponding tool uses the mapping and non-bonded interactions from the previous parametrization[1]. In contrast to the previous parametrization procedure, the determination of bonded interactions is based on a conformer search using RDKit (further details are provided in Section S2). Further modifications that have been made include the following. Force constants were reparametrized using Swarm-CG calculations, with the angle force constants restricted to $350\,\mathrm{kJ\,mol^{-1}\,rad^{-2}}$ and the bond force constants restricted to $12\,000\,\mathrm{kJ\,nm^{-2}}$. Furthermore, the lowest angle force constant was set to $125\,\mathrm{kJ\,mol^{-1}\,rad^{-2}}$. To simplify the automatic parametrization procedure, proper dihedral angles were neglected in the case of linearly arranged beads (please refer to Section S2 for more information). Standard force constants for bonds and

angles between different fragments were applied, using a force constant of 5000 kJ nm^{-2} and 250 kJ mol^{-1} rad^{-2}, respectively. Furthermore, the virtual site of the HS28 fragment was replaced by a normal bead.

S2 Details of ETKDG

The ETKDG[2] approach applied to the coarse-grained models was adapted from the work by Potter et al.[3,4], who introduced this method for the parametrization of bonded terms of Martini models (https://github.com/cgkmw-durham/cg_param_m3). ETKDG is an efficient method for generating a conformational ensemble of a molecule that covers its accessible phase space. It combines the distance geometry approach, which uses purely geometric constraints to describe all possible conformations of a molecule, with empirical knowledge. The latter is utilized in the form of experimental torsional-angle preferences obtained from the Cambridge structural database. Additionally, so-called basic knowledge, such as flat aromatic rings, is incorporated to enhance the accuracy of the conformers generated.

The tool developed in this work reads the VSOMM2 atomistic topology files and recreates the SOM molecules within the RDKit framework. For each molecule, conformers are generated and minimized using the Universal force field[5]. The atom groups of the conformers are mapped according to the bead assignment defined in Dettmann et al.[1]. Finally, the bonds and angles between the beads are determined by averaging over the different mapped conformers generated for a molecule.

Dihedral angles are calculated from the minimized conformer with the lowest energy. This encompasses dihedral angles solely for the purpose of stabilizing certain ring molecules (improper dihedral angles), whereby the occurrence of an angle close to 180° between the constructing particles can be excluded. This avoids the utilization of restricted bending potentials[6]. The parameters of these potentials usually require manual fine-tuning for good stability, which is impractical for the heterogeneous systems and automated parametrization used in this work.

S3 Comparison of the approach to other Martini parametrization techniques

Various approaches have been developed for the automated parametrization of molecules within the Martini force field, including the approach by Potter et al.[3,4] and the Bartender tool developed by Pereira et al.[7]. For the approach of Potter et al., only the SMILES of the molecule is required. The algorithm automatically determines a mapping using a graph-based approach[8]. The bead assignment is conducted by matching octanol–water partitioning free energies of the individual atom groups, predominantly using those determined with ALOGPS[9]. Equilibrium distances and angles are determined by an ETKDG conformer search of the molecule. Purely generic force constants are used for bonds and angles. Overall, this approach offers a very user-friendly and automated parametrization in principle for any small organic molecule.

In the automated parametrization approach defined in the present work, mapping and bead types are defined and fixed for the whole fragment database and therefore do not have to be redefined for every new macromolecule. The definition of non-bonded interactions is based on already parametrized Martini 3 molecules and a comparison to COSMO-RS data[1]. For equilibrium distances and angles, the ETKDG method is utilized as well. However, the accuracy of the force constants is improved compared to using purely generic ones by applying Swarm-CG force constants for bonded interactions within the fragments.

As mentioned in the main manuscript, these force constants could also have been determined using the Bartender tool. This tool allows, with a given mapping, to determine the parameters for the bonded interactions of a coarse-grained molecule based on an automatically set up xTB[10] reference simulation (xTB: semiempirical Extended Tight-Binding method). The process is based on a non-iterative fitting of data from the reference simulation of the molecule. In principle, this allows the bonded interactions of any small molecule to be parametrized efficiently. In contrast, the Swarm-CG tool was used to parametrize the individual fragments. This tool performs sequential reference simulations to determine the

parameters of the bonded interactions. The classical GROMOS force field was used for the reference simulations. With the force constants restricted, as in this work, it can be assumed that both approaches provide similar results.

The approach from this work with the developed tool is limited to the coarse-graining of atomistic VSOMM2 SOM models due to the parametrization of the fragment molecules. However, the general approach, i.e. the more precise parametrization of repeating molecular fragments and their combination with an efficient method such as ETKDG as a reference, can be transferred to other suitable molecular systems. The prerequisite for this is a defined and limited set of fragment molecules. This could be used, for example, in the case of polymers, carbohydrates, proteins or other complex molecules composed of molecular units.

S4 Ab initio and experimental IR spectra of the HS30 fragment

The HS30 fragment, which corresponds to a trimesic acid molecule in its fully deprotonated state, has been studied using ab initio calculations and experimental techniques (please refer to Mahalakshmi and Balachandran[11]). This study identifies vibrational frequencies predominantly in the range of 500 to 3500 cm^{-1}, with specific modes including O–H stretching of the carboxylic acid group at 3420 cm^{-1}, benzene C–H stretching at 3100 and 3024 cm^{-1}, C=O stretching at 1723 cm^{-1}, benzene C–C stretching in the regions of 1460 to 1660 cm^{-1} and 1070 to 1150 cm^{-1}, and bending vibrations in the range of 966 to 1276 cm^{-1}. These values are significantly higher than the maximum frequencies observed in the coarse-grained model, which are around 400 cm^{-1}. This disparity arises because coarse-graining inherently simplifies the molecular system, eliminating many degrees of freedom necessary to capture high-frequency vibrational motions. Additionally, the atomistic reference simulations and parametrization techniques used to derive coarse-grained bonded interaction parameters are not explicitly designed to replicate detailed vibrational properties. Instead, coarse graining prioritizes computational efficiency and the representation of large-scale, slower dynamics over the reproduction of fine-grained vibrational features.

S5 Earth mover's distance and total mismatch score

The Earth mover's distance (EMD) was introduced by Rubner et al.[12] as a measure of distance, or dissimilarity, between two distributions. If such distributions are seen as masses of earth, the EMD reflects the minimal amount of work required to transform one distribution into the other by moving "distribution mass". The computation of the EMD is based on the transportation problem[13] and is defined in the following way:

$$\mathrm{EMD}(P, Q) = \frac{\min_{\{f_{ij}\}} \sum_{i,j} c_{ij} f_{ij}}{\sum_{i,j} f_{ij}}, \tag{1}$$

with two distributions P and Q, $\{f_{ij}\}$ representing the set of flows, and c_{ij} being the cost of moving a distribution unit from bin i to j, where $i \in P$ and $j \in Q$. Therefore, the EMD is calculated by minimizing the right term through the optimization of the set of flows.

Swarm-CG utilizes the EMD to evaluate the difference between coarse-grained and reference distributions of the bonded interactions. To quantify the match of all distributions for a given molecule, the total mismatch score (TMS) is used and defined via the following function[14]:

$$\mathrm{TMS} = \sqrt{\sum_i (C \cdot \mathrm{EMD}_{\mathrm{bond,i}})^2} + \sqrt{\sum_j \mathrm{EMD}^2_{\mathrm{angle,j}}} + \sqrt{\sum_k \mathrm{EMD}^2_{\mathrm{dihedral,k}}}. \tag{2}$$

C is a scaling factor that is used to convert between the different units of EMDs from bonds and (dihedral) angles. The standard value of C is equal to 50, which means that an EMD of 0.4 Å contributes in the same way as an EMD of 20°. If the units of the axes of the distributions are considered, the EMD would have the same unit as the x-axis. In this case, the following units would be applied: $[C] = {}^\circ\,\mathrm{Å}^{-1}$ and $[\mathrm{TMS}] = {}^\circ$.

S6 Comparison of ETKDG and Swarm-CG equilibrium values

Strictly speaking, the equilibrium values and force constants determined by Swarm-CG are interdependent. The equilibrium distances and angles are available for all interactions of a macromolecule only through ETKDG. Due to the supplementation of the Swarm-CG force constants to the coarse-grained models, we compared the equilibrium distances and angles, determined by Swarm-CG and ETKDG, in Figure S1. Panels (a) and (b) respectively show a comparison between the equilibrium distances and angles determined from Swarm-CG and from ETKDG, using a center of geometry mapping. Overall, good correlations are visible, with correlation coefficients of 0.902 and 0.928, respectively, for the distances and angles.

In general, different force fields can operate with varying equilibrium distances, contingent upon the corresponding parametrization procedure for determining these parameters[3]. Therefore, deviations in equilibrium distances of the coarse-grained model determined from the mapped molecules can arise when using different reference models. As a result, slight deviations can be expected. It should be noted that Swarm-CG iteratively optimizes the bonded interactions, which compensates for cross interactions and can further affect the generated parameters of the coarse-grained models. The role of cross interactions may be less significant within individual fragments, due to the application of a number of exclusions equal to 2.

It is presumed that the deviations are caused by the united-atom representation employed by the modeler and consequently applied to the reference for Swarm-CG. To test this hypothesis, Swarm-CG equilibrium distances and angles were recalculated using a center of mass mapping (see Figure S1 (c) and (d)). When using the center of mass mapping, the weight of the hydrogen atoms is reduced. This results in a lower deviation when comparing mapped structures created from a reference with an all-atom or united-atom representation. Indeed, in this particular instance, and especially regarding the equilibrium distances, the majority of values are situated on the diagonal. This suggests that the deviations produced when using the center of geometry mapping are mainly caused by the united-atom repre-

sentation of the reference, used for Swarm-CG. However, the center of geometry mapping is the standard way of mapping molecules for the Martini 3 force field, and was therefore applied, also because of the still overall good correlation. Reference simulations of atomistic fragments could be conducted again with an all-atom representation. However, this would require a high computational effort, which is beyond the scope of this study.

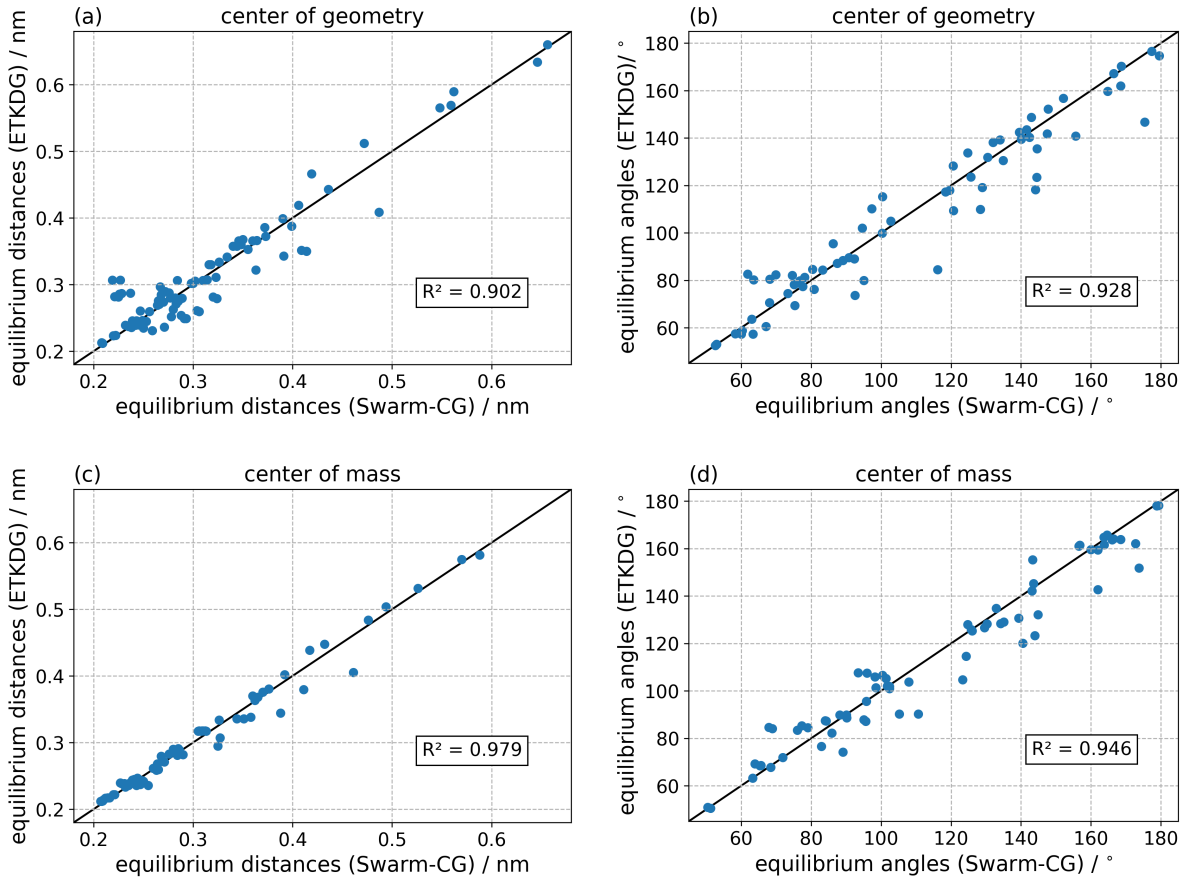


Figure S1: Comparison of the equilibrium values of bonds (a) and angles (b) determined by using Swarm-CG (x-axis) and ETKDG (y-axis) with a center of geometry mapping. Furthermore, the values of bonds (c) and angles (d) were calculated using a center of mass mapping. The bonds and angles considered belong to the fragment molecules of SOM III.

Table S1: Elemental and organic fractions of the models taken from the IHSS, namely: carbon (C), nitrogen (N), carbonyl, carboxylic, aromatic, acetal, heteroaliphatic and aliphatic fractions.

	C	N	carbonyl	carboxylic	aromatic	acetal	heteroaliphatic	aliphatic
SOM I	0.5057	0.0272	0.1212	0.2525	0.3031	0.0101	0.0909	0.2222
SOM II	0.5951	0.039	0.01	0.11	0.41	0.06	0.14	0.27
SOM III	0.6381	0.0123	0.08	0.15	0.58	0.04	0.01	0.14
SOM IV	0.5131	0.0234	0.0373	0.1936	0.4037	0.0621	0.1128	0.1905
SOM V	0.5234	0.0067	0.0495	0.1683	0.2178	0.0594	0.1584	0.3466

Table S2: x_{water} values for varying number of water molecules N_{water} in the respective atomistic systems.

N_{water}	2500	4000	5000	6500	8500	10000	13000	15000	20000
SOM I	0.141	0.208	0.247	0.299	0.358	0.396	0.460	0.496	0.568
SOM II	0.145	0.214	0.254	0.307	0.366	0.405	0.469	0.505	0.576
SOM III	0.138	0.203	0.242	0.293	0.352	0.390	0.453	0.489	0.561
SOM IV	0.136	0.201	0.240	0.291	0.349	0.386	0.450	0.486	0.558
SOM V	0.143	0.210	0.250	0.302	0.361	0.399	0.464	0.499	0.571

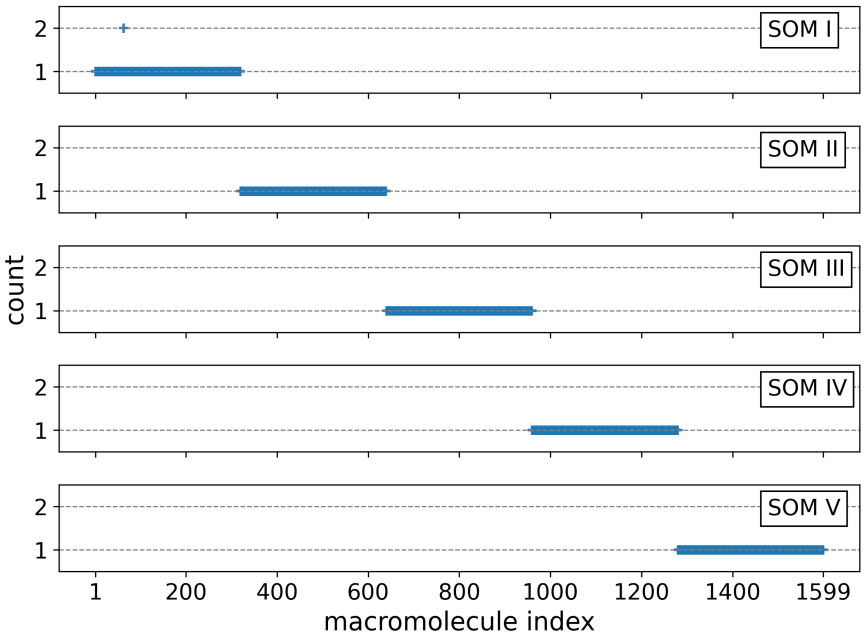


Figure S2: Illustration of the heterogeneity of the SOM compositions, based on the fragment sequences of the macromolecules. The same index is assigned to the macromolecules if they have the same sequence of fragments. Except SOM I, in which one of these molecules occurs twice, all macromolecules occurring are different, also when comparing different SOM compositions.

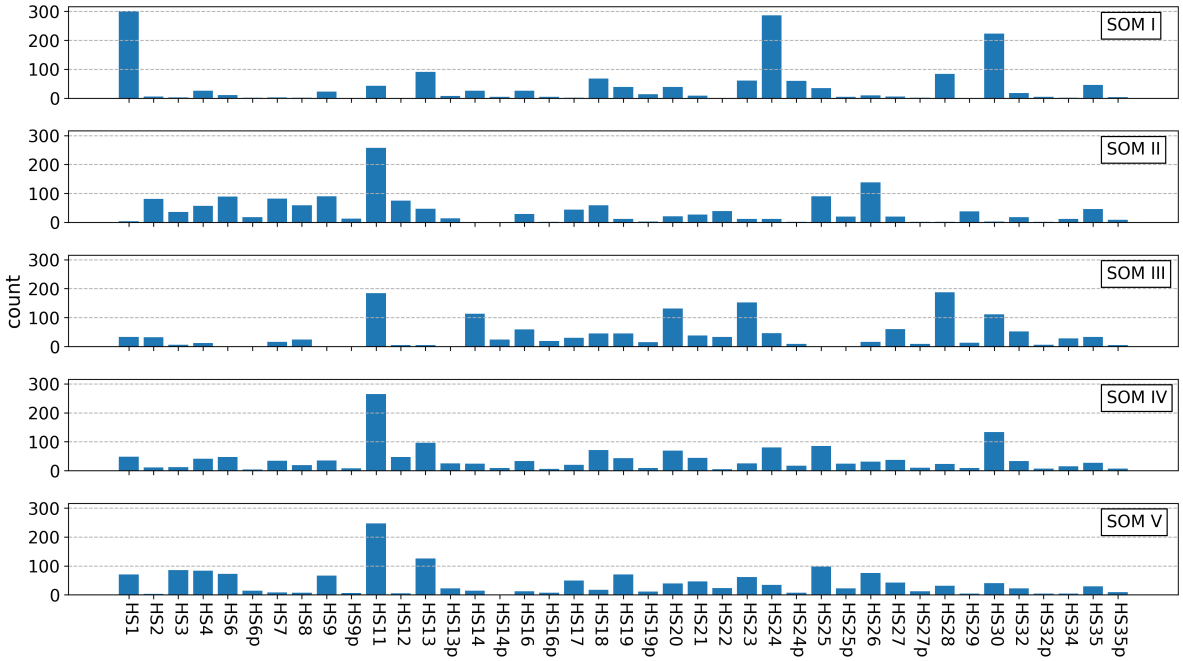


Figure S3: Distribution of fragments within the SOM compositions.

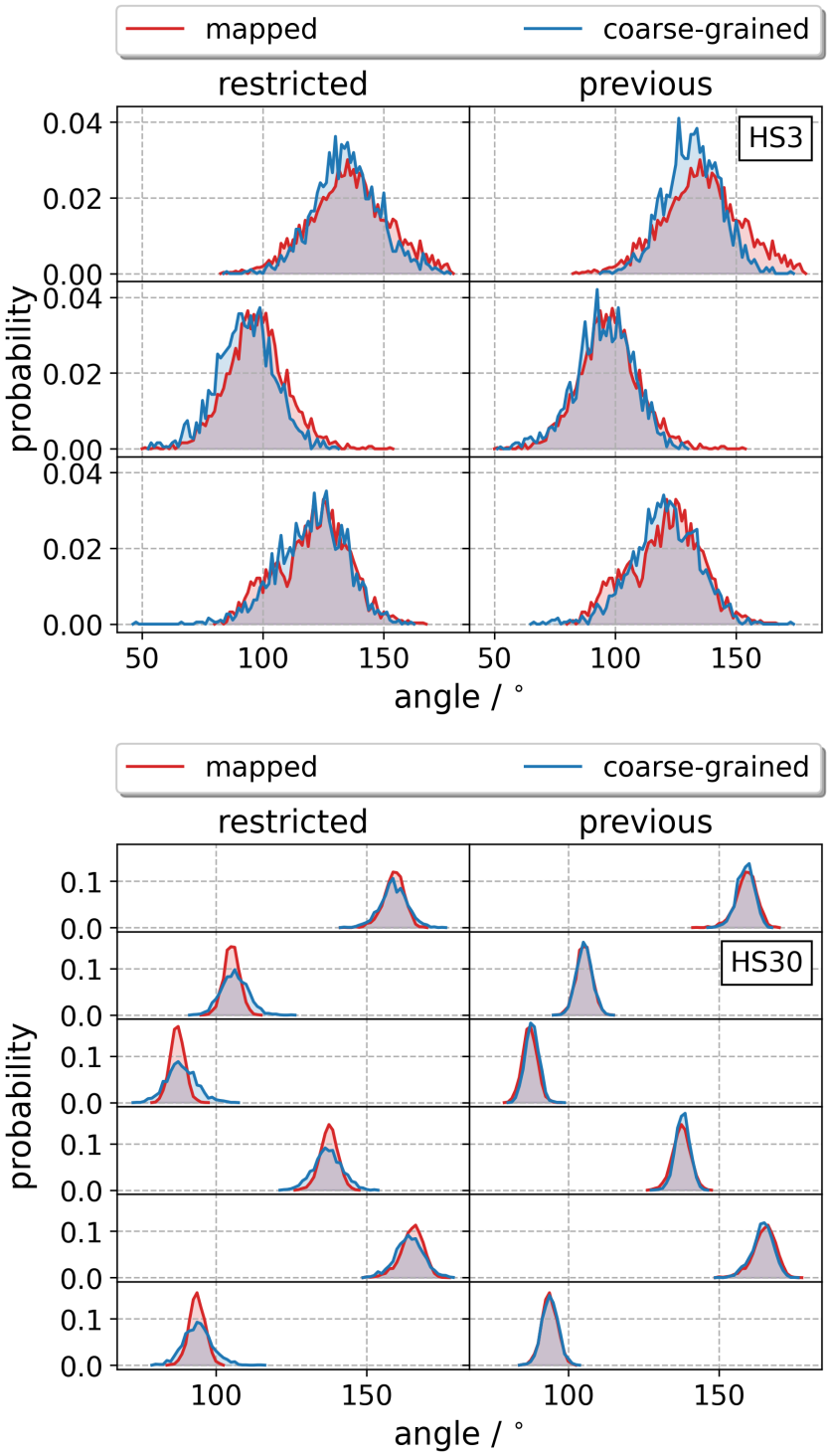


Figure S4: Comparison of the angle distributions of atomistic (red) and coarse-grained (blue) models. All angles for the HS3 and HS30 fragments are considered. From top to bottom, for HS3, the angles 1-4-2, 1-4-3, and 2-4-3, and for HS30, the angles 2-1-4, 2-1-6, 3-4-1, 3-4-6, 1-6-5, and 4-6-5 are shown. For each angle of both fragments, the distributions from the models of the restricted parametrization and the previous parametrization are depicted.

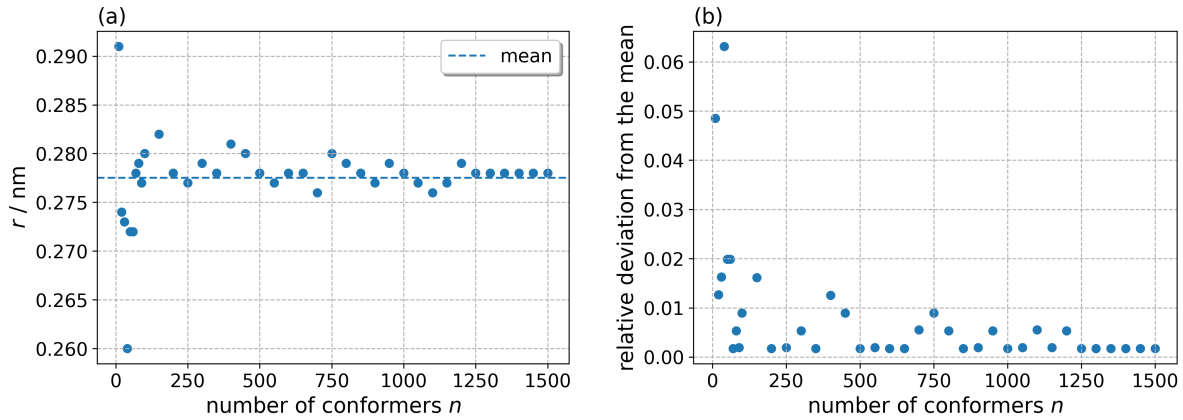


Figure S5: In (a), the value of one specific bond distance is considered, which changes over a varying number of conformers n considered in the calculation. As expected, the values converge to the mean as n increases. In (b), the change in these values relative to the mean is shown. To yield the curves in Figure 5 in the main manuscript, the change relative to the mean is calculated for each bond, and subsequently averaged over all bonds in the system. The entire procedure is performed analogously for the angles.

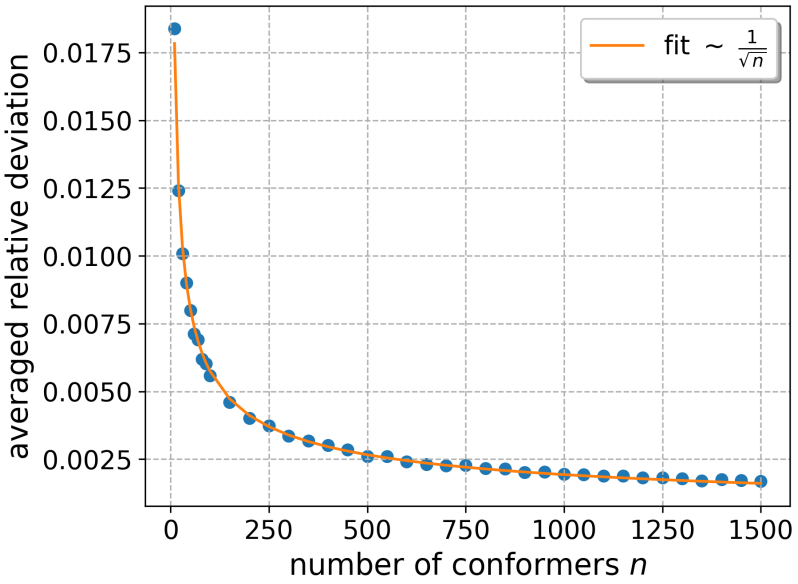


Figure S6: Convergence of the angle values for the SOM III model system as a function of the number of conformers. The relative deviation from the mean value, averaged over all bonded terms in the model system is depicted.

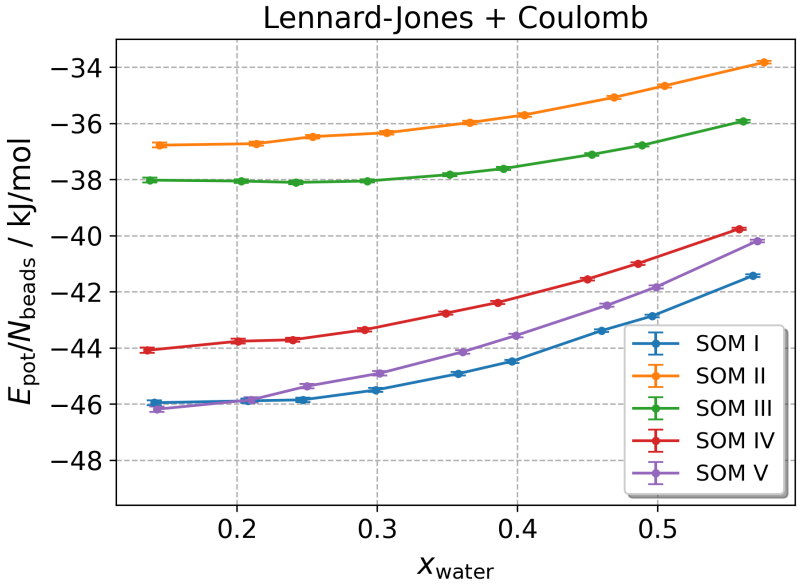


Figure S7: Time averaged sum of Lennard-Jones and Coulomb potential energies; error bars denote standard deviations. From the similarity to the total potential energy in Figure 7 (a) in the main manuscript, it can be inferred that the latter is mainly determined by the non-bonded interactions.

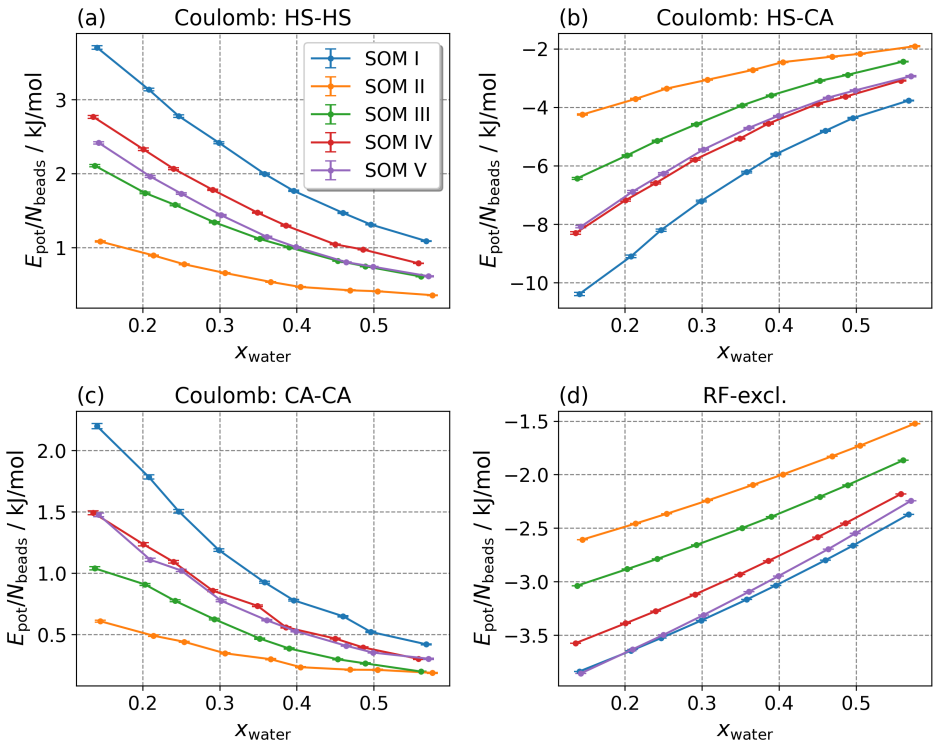


Figure S8: Time averaged Coulomb potential energies for the different pairs of interactions between HS molecules and calcium (CA), depicted in (a)-(c); error bars denote standard deviations. In (d), the reaction-field correction applied to the excluded pairs of beads (RF-excl.) is shown. For each SOM type respectively, the sum of the energy curves in (a)-(d) result in the Coulomb potential energies, depicted in Figure 7 (f) in the main manuscript.

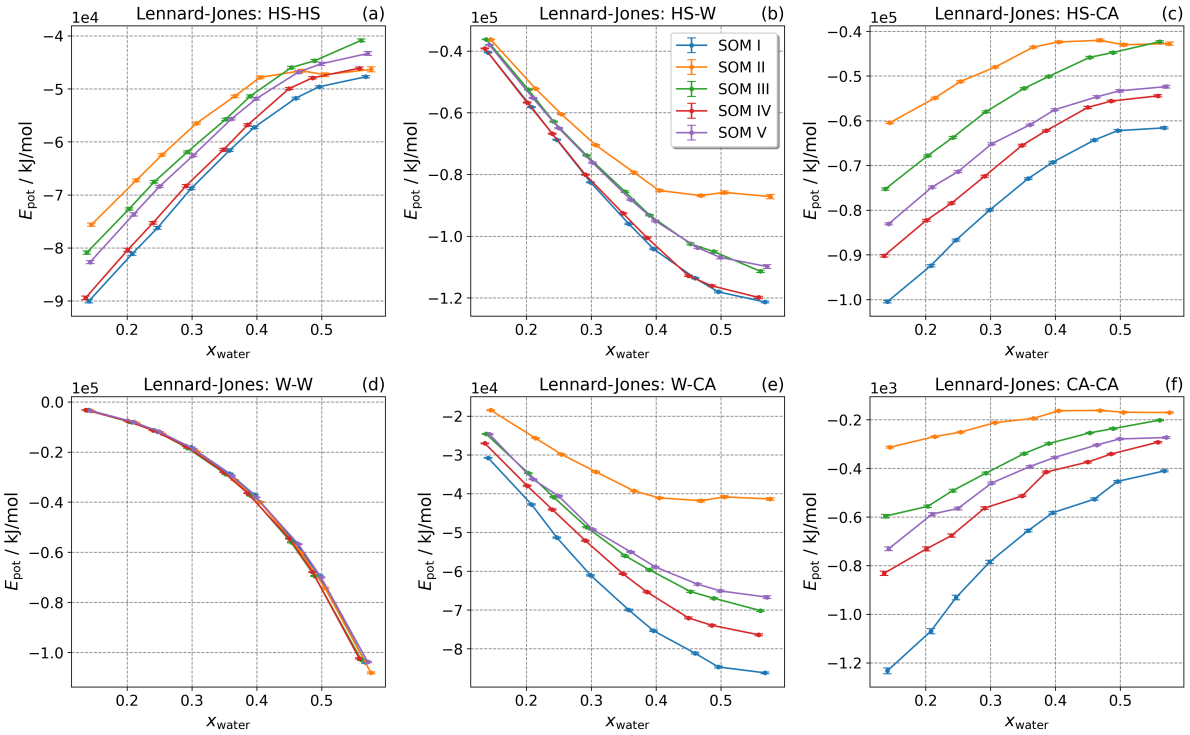


Figure S9: Time averaged Lennard-Jones potential energies for the different pairs of interactions between HS molecules, water (W) and calcium (CA); error bars denote standard deviations. These are the energies from Figure 8 of the main manuscript, which have not been normalized to the number of beads in the respective systems. As no normalization was used, these graphs are only used to illustrate the curves' progressions and not to compare the energies between the individual systems. For example, the saturation of the interaction sites with water can be seen from the curves in (b) and (d), which approach an almost constant value with increasing water content.

S7 Overlap calculation

The overlap of density distributions was used to quantify the demixing between HS molecules and water. It was calculated as follows. First, the partial densities ρ of HS and water along the dimensions x, y and z were determined using the *gmx density* tool, whereby the y-direction is used for the following formulas. These densities were normalized, resulting in a quantity that can be understood as a position density p:

$$p(y) = \frac{\rho(y)}{\int \rho(y) dy}. \tag{3}$$

Finally, the overlap O between the density of HS molecules p_{HS} and water p_{W} was calculated using the following formula:

$$O_y = \int \min(p_{\mathrm{HS}}, p_{\mathrm{W}}) dy. \tag{4}$$

O_y corresponds to the value of the green area, depicted in Figure S10. This quantity was determined for specified time intervals, yielding the individual points in Figure 9 of the main manuscript.

The following fit function with the parameters a, b and c was used for the fit in Figure 9:

$$f(t) = a \cdot \exp(b \cdot t) + c \tag{5}$$

Here, the parameter c specifies the value against which the overlap converges over time and quantifies the degree of demixing. Table S3 shows the values of c for the different SOM types with the highest water content. Only the parameter that has the lowest value of the fits in the x, y and z directions is shown for the respective systems, which corresponds to a maximum demixing in that direction.

Table S3: Lowest c value for the different SOM systems, which quantifies the degree of demixing between HS molecules and water. The corresponding direction, from which the lowest value was obrained, is shown in brackets after the respective SOM type.

	SOM I (z)	SOM II (y)	SOM III (x)	SOM IV (x)	SOM V (y)
c	0.858	0.721	0.875	0.783	0.822

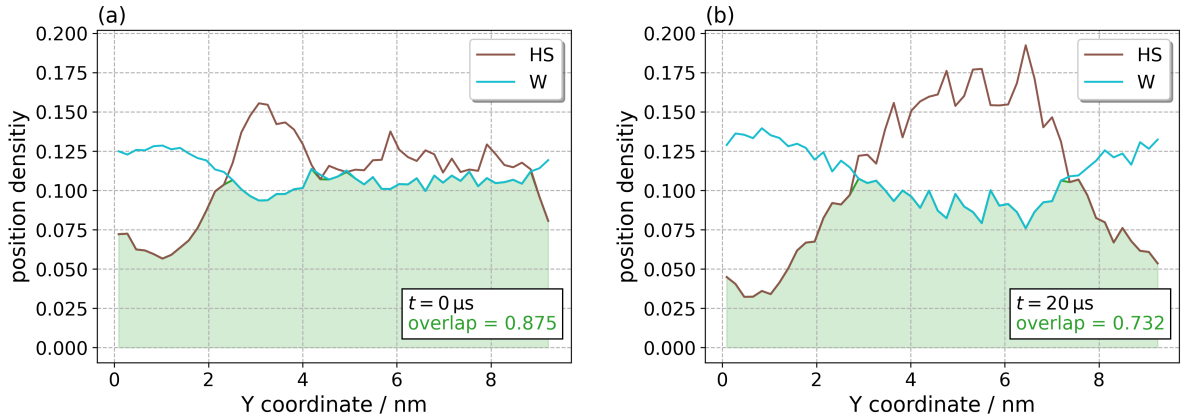


Figure S10: Partial densities represented as the position densities in the y-direction, at 0 µs (a) and at 20 µs (b). The density of the HS molecules (brown) and water (cyan) is depicted. The green area represents the overlap between the two density distributions. The two times selected here correspond to those in Figure 10 of the main manuscript, which shows the corresponding snapshots of the trajectory.

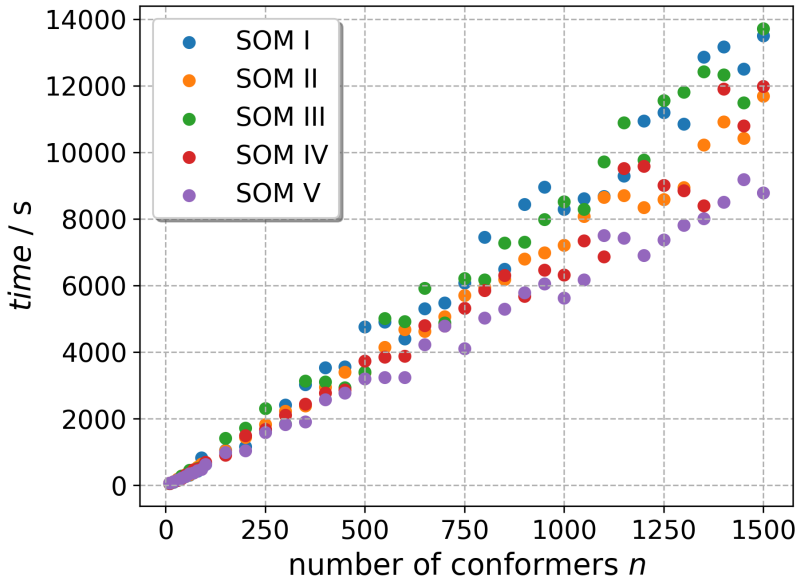


Figure S11: Parametrization time for generating the coarse-grained topology files for varying numbers of conformers. The Intel(R) Xeon(R) Platinum 8380H CPU @ 2.90GHz was used with 28 OMP threads.

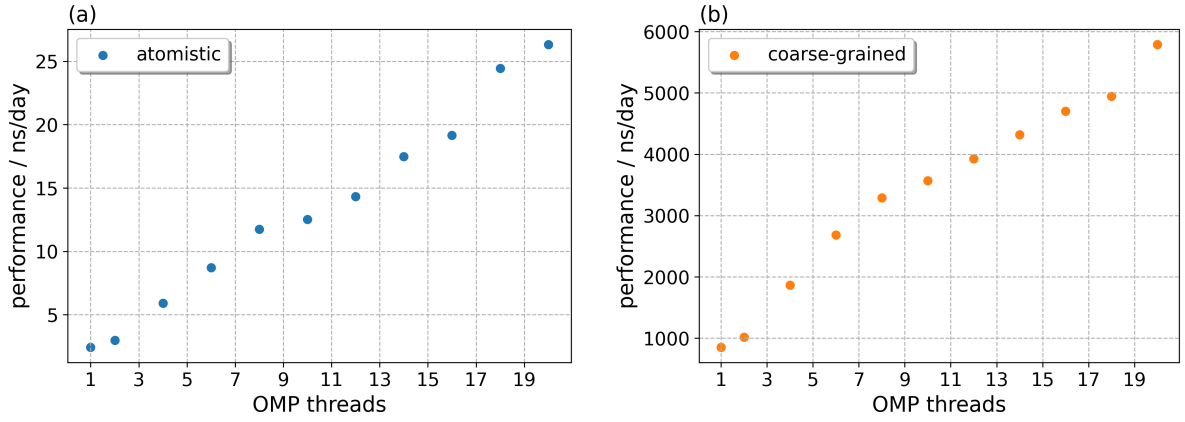


Figure S12: Performance of the atomistic (a) and coarse-grained (b) MD simulations of the SOM III system with the highest water content, using 10000 simulation steps. The Intel(R) Xeon(R) Platinum 8380H CPU @ 2.90GHz was used with OMP parallelization.

References

(1) 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.

(2) 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.

(3) 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.

(4) 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.

(5) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society* **1992**, *114*, 10024–10035.

(6) Bulacu, M.; Goga, N.; Zhao, W.; Rossi, G.; Monticelli, L.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved Angle Potentials for Coarse-Grained Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* **2013**, *9*, 3282–3292.

(7) 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.

(8) Webb, M. A.; Delannoy, J.-Y.; de Pablo, J. J. Graph-Based Approach to Systematic Molecular Coarse-Graining. *Journal of Chemical Theory and Computation* **2018**, *15*, 1199–1208.

(9) 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.

(10) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. *WIREs Computational Molecular Science* **2020**, *11*.

(11) Mahalakshmi, G.; Balachandran, V. FT-IR and FT-Raman spectra, normal coordinate analysis and ab initio computations of Trimesic acid. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2014**, *124*, 535–547.

(12) 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.

(13) Dantzig, G. B. In *Activity Analysis of Production and Allocation*; Koopmans, T. C., Ed.; John Wiley and Sons, 1951; pp 359–373.

(14) 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.

Environmental Science Advances

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Check for updates

Cite this: DOI: 10.1039/d5va00237k

Coarse-grained simulations of sulfanilamide and hexachlorobenzene mobility in soil organic matter

Lorenz F. Dettmann, [a] Oliver Kühn *[a] and Ashour A. Ahmed [b]

Soil organic matter (SOM) is a complex and heterogeneous molecular system, crucial for soil health and ecosystem functioning. Therefore, the release of pollutant molecules into the environment poses a significant environmental threat. Mechanisms that are governed by the interactions of these pollutants with SOM at the molecular level remain largely unexplored. In this study, coarse-grained molecular dynamics simulations were employed to investigate the behavior of hexachlorobenzene (HCB) and sulfanilamide (SAA) in humic substance (HS) systems with varying compositions. Diffusion coefficients indicated a strong influence of water on SAA, with SAA displaying higher mobility, whereas HCB exhibited greater accumulation in the HS phase. Calculations of spatial distributions supported these observations, showing that SAA is predominantly situated in the water phase, while HCB's interaction was influenced by the hydrophobicity of the SOM system. Simulations in the microsecond range, which were possible by the coarse-grained representation, revealed temporary trapping of SAA in the SOM matrix. These were anti-correlated with water diffusion, while HCB's behavior was dominated by direct pollutant–SOM interactions. In general, the coarse-graining approach provides novel insights into the trapping processes of pollutants in SOM and offers a realistic representation of molecular interactions at larger spatial and temporal scales. The proposed method enhances the understanding of pollutant mobility in soil systems, thus enabling future studies on the ecological impact of pollutant–SOM interactions.

Received 29th July 2025
Accepted 23rd September 2025

DOI: 10.1039/d5va00237k

rsc.li/esadvances

Open Access Article. Published on 26 September 2025. Downloaded on 10/24/2025 2:41:56 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Environmental significance

Pollutants released into terrestrial ecosystems, including polar and non-polar organic pollutants, threaten soil functions and human health. Their chemical stability enables them to evade natural degradation; however, their behavior in soil is not well understood due to its structural and chemical complexity. Soil organic matter (SOM), which is rich in diverse functional groups, is a principal sorbent that governs pollutant mobility. In this study, we employ coarse-grained molecular dynamics to investigate the diffusion of hexachlorobenzene and sulfanilamide within SOM of different compositions. Simulations in the microsecond range provide mechanistic insight into interactions between pollutants and SOM and offer a transferable framework for investigating pollutant dynamics in SOM.

1 Introduction

Soil organic matter (SOM) is a significant soil component, comprising decomposed plant and animal residues, microbial biomass, and various organic compounds.[1] It contributes to improved soil structure stability, increasing its water infiltration and retention capabilities.[2,3] By enhancing the soil's nutrient-holding capacity, SOM plays a crucial role in maintaining soil fertility and supporting plant growth.[4] Furthermore, SOM provides a habitat for soil microorganisms, which drive nutrient cycling and the functioning of the ecosystems.[5]

The soil's ecological functions are endangered by the release of chemical substances into the environment.[6,7] Persistent organic pollutants (POPs) are toxic chemical compounds with a long lifetime, resulting from their resistance to degradation through environmental processes.[8] These compounds are widespread in the environment and can be transported over long distances through air and water currents.[9,10] Moreover, polar organic pollutants, such as pharmaceutical antibiotics and personal care products, are of particular concern due to their potential to cause harm to terrestrial ecosystems.[11–13] Numerous studies have been carried out to investigate the sorption of POPs and polar organic pollutants in soil using experimental and theoretical methods.[14–35]

Understanding the interactions of pollutants within the soil matrix and the properties of SOM is crucial for optimizing remediation strategies, such as immobilization or bioremediation.[36] Traditionally, SOM has been closely studied by obtaining

[a]University of Rostock, Institute of Physics, Albert-Einstein-Str. 23-24, D-18059 Rostock, Germany. E-mail: oliver.kuehn@uni-rostock.de
[b]Leibniz Institute for Catalysis (LIKAT), Albert-Einstein-Str. 29a, D-18059 Rostock, Germany

© 2025 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 26 September 2025. Downloaded on 10/24/2025 2:41:56 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

so-called humic substances (HSs) *via* alkaline extraction, which benefits from standardized protocols and reference samples provided by the International Humic Substances Society (IHSS).[37] Despite ongoing discussions about how accurately HSs can reflect native SOM,[38–41] they remain the subject of research aiming to systematically understand more about SOM formation, composition, and activities in diverse environments.[42,43]

From a theoretical viewpoint, computational chemistry provides insight into processes at the molecular level and expands the understanding of the pollutant–SOM interaction. The modeling of SOM is challenging due to its heterogeneity and complexity, which has led to the ongoing development of various SOM models.[44–55] The Vienna Soil Organic Matter Modeler 2 (VSOMM2)[56,57] was recently released, which allows model systems of condensed phase HS to be generated, following a building block approach. Based on elemental and organic fractions as input parameters, heterogeneous systems of HS with different compositions can be investigated, thus approaching the properties of real SOM. These inputs (*e.g.*, bulk C to N ratios and fractions of aromatic, aliphatic, or acidic functional groups) are commonly used to describe SOM heterogeneity and serve as constraints on the generated structures, consistent with established practices in supramolecular SOM modeling (see Gerzabek *et al.*[58]).

While such systems consisting of larger, supramolecular structures can improve the accuracy of the models, the accessible time and length scales of simulations may be compromised. To reduce computational cost, coarse-graining techniques are often employed,[59–62] whereby multiple atoms are grouped into larger particles known as beads. This approach sacrifices atomistic detail for computational efficiency by reducing the number of interactions that must be evaluated. Coarse-graining strategies can follow either a top-down or bottom-up philosophy: in the top-down approach, interaction potentials are tuned to reproduce experimental observables or macroscopic properties, whereas in the bottom-up approach, they are derived systematically from atomistic simulations.[60] Furthermore, several coarse-grained force fields have emerged,[63–65] including the widely used Martini model,[66] which provides pre-parametrized beads to represent common chemical functionalities.

To apply the coarse-graining method to SOM systems, we have developed a tool[67,68] that converts systems generated by the VSOMM2 from an atomistic to a coarse-grained representation, using the Martini 3 framework.[69] On the one hand, this enables an abstraction of the SOM molecular structures, whose exact shape remains unknown. On the other hand, the simulation of diffusion processes in large SOM systems becomes possible, thus accounting for the heterogeneous nature of interaction sites of SOM. To the best of our knowledge, no molecular-level study has examined the interaction between pollutants and supramolecular SOM structures over extended time scales, with a focus on their diffusion and distinct interaction sites.

The present study fills this gap by investigating the mechanisms and key functional groups that drive the interaction and trapping of POPs and polar organic pollutants in SOM at a coarse-grained level of theory. Sulfanilamide (SAA) and hexachlorobenzene (HCB) were selected to represent distinct pollutant classes and interaction modes with SOM. SAA is a widely detected pharmaceutical with polar interaction sites, making it useful for probing site-specific interactions with carboxylic and phenolic moieties.[15] In contrast, HCB is a legacy POP with hydrophobic character that primarily partitions into aromatic or black-carbon-like domains.[50] Together, this pair of pollutants represents complementary sorption mechanisms, reflects environmental relevance, and allows comparison to previous studies. Their interactions with SOM model systems were analyzed, the latter being represented by HSs from different samples provided by the IHSS. Coarse-grained molecular dynamics (CGMD) simulations were employed to examine the mobility of pollutants in the SOM systems and their interaction with SOM functional groups. The analysis of the diffusion and trapping of the pollutant molecules within the SOM matrix provided insights into the underlying processes that govern their behavior in soil environments.

2 Materials and methods

2.1 Coarse-grained modeling approach

The parametrization of the coarse-grained pollutants and SOM models was conducted within the framework of the Martini 3 force field, which offers a wide range of available topologies for, *e.g.*, small molecules,[70] biomolecules,[71–74] or carbon materials.[75,76] In previous work,[67,68] coarse-grained SOM systems based on VSOMM2 structures were introduced, employing an automated parametrization approach for the generation of the models. Validation of the coarse-grained models involved matching the probability distributions of bonded interactions, using the Swarm-CG tool,[77] and reproducing thermodynamic observables, specifically transfer free energies.

The mapping of pollutant molecules was performed with careful consideration of chemical functionality and molecular symmetry. Martini 3 bead types were chosen to reflect the polarity and hydrogen-bonding capacity of functional groups, aiming to reproduce experimental octanol–water partition coefficients as a first-order validation of solute–solvent interactions. To this end, coarse-grained transfer free energies were calculated using the Bennett acceptance ratio method,[78] and compared against experimental data if available. While the Martini framework cannot fully capture directionality in hydrogen bonding, it allows to approximate the overall partitioning behavior relevant for environmental partitioning. Bonded interaction parameters were further refined using Swarm-CG, with atomistic MD trajectories of each pollutant in explicit water serving as a structural reference during iterative optimization. A detailed description of this protocol is included in Section S1 of the SI.

The primary focus of the investigation is the mobility and interactions of SAA and HCB in SOM systems, as captured by CGMD simulations. Using elemental and carbon distribution data on HS samples from the IHSS, five characteristically distinct HS systems were generated using VSOMM2 and subsequently converted into a coarse-grained representation. Guided by the principal component analysis of IHSS reference

 © 2025 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 26 September 2025. Downloaded on 10/24/2025 2:41:56 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Table 1 HS samples from the IHSS, which are modeled and analyzed in this work. The elemental and organic fractions of the HS models were used as input parameters for the generation with the VSOMM2. These include carbon (C), nitrogen (N), carbonyl, carboxylic (carbox.), aromatic (arom.), acetal, hetero-aliphatic (hetero-aliph.) and aliphatic (aliph.) fractions

			Elemental and organic fractions							
Identifier	HS sample	Characteristics	C	N	Carbonyl	Carbox.	Arom.	Acetal	Hetero-aliph.	Aliph.
SOM I	Elliot soil I (FA)	High carboxyl and carbonyl content	0.5057	0.0272	0.1212	0.2525	0.3031	0.0101	0.0909	0.2222
SOM II	Elliot soil IV (HA)	Low carboxyl and carbonyl content	0.5951	0.0390	0.0100	0.1100	0.4100	0.0600	0.1400	0.2700
SOM III	Leonardite (HA)	High aromatic content	0.6381	0.0123	0.0800	0.1500	0.5800	0.0400	0.0100	0.1400
SOM IV	Pahokee peat II (FA)	Intermediate carboxyl and aromatic content	0.5131	0.0234	0.0373	0.1936	0.4037	0.0621	0.1128	0.1905
SOM V	Suwannee river II (FA)	High hetero-aliphatic & aliphatic content	0.5234	0.0067	0.0495	0.1683	0.2178	0.0594	0.1584	0.3466

materials conducted by Escalona *et al.*,[79] we selected HS samples that span distinct regions along the acidity and aromaticity axes, thereby capturing the chemical diversity and inherent heterogeneity of SOM within a concise and manageable set of models. Please refer to Table 1, and Tables S1 and S2 (SI) for further details regarding the HS systems, which comprise information on the compositions and functional group fractions.

As initial configurations, equilibrated SOM structures derived from 20 μs long CGMD simulations from previous work were adopted, in which different water contents were considered.[68] To enhance the sampling of the interactions between pollutants and HSs, the SOM systems with the highest water content considered were utilized, corresponding to a heavy atom fraction of water of approximately 0.567. Here, one coarse-grained water bead represents four atomistic water molecules. Each of the five systems consists of 1600 VSOMM2 fragments, with five fragments per HS molecule and 5000 water beads.

The pollutants SAA and HCB were studied separately and positioned into the different SOM systems. Each simulation included a single pollutant molecule, representing a low environmental concentration of the pollutant. Additional simulations at higher pollutant concentrations are reported in Section S9 of the SI. To ensure sufficient sampling of the different interaction sites of the heterogeneous SOM structures, 20 simulations were conducted, in which the initial position of the pollutant was varied. These positions were further refined to ensure a minimum distance of 0.75 nm between the pollutant and HS molecules, thus preventing initial artificial trapping within a local minimum of the SOM matrix. Therefore, the pollutant molecules initially tended to interact with the HS molecules from the water phase.

2.2 Computational details

The CGMD simulations were performed using the GROMACS simulation package version 2019.4.[80] Each system underwent an energy minimization, followed by an NPT equilibration of 10 ns and the production run, which lasted 1.5 μs. The temperature and pressure in the systems were 298.15 K and 1 bar, respectively. Further computational details and information about the simulation parameters are provided in Section S2 (SI).

Python packages were utilized to carry out the analysis, including NumPy,[81] Matplotlib,[82] MDAnalysis[83,84] and IPython.[85] The molecular structures were visualized using VMD.[86]

3 Results and discussion

3.1 Pollutant parametrization

The assigned mappings for SAA and HCB are depicted in Fig. 1. First, the available standard Martini bead types were used for the atomic groups, including the aromatic moieties (TC5) and the amino group (TN6d) of SAA. The bead types for the sulfonamide group of SAA (Fig. 1(a) in blue) and for the HCB groups (Fig. 1(b) in green) were then determined by optimizing the octanol–water transfer free energies (ΔG), calculated with the models, against the experimental energies. In addition, the solvent accessible surface area (SASA) and Connolly surfaces (Fig. 1(c) and (d)) were calculated from the atomistic pollutant models to account for the reproduction of the molecular shape. A comparison of the values is shown in Table 2.

The octanol–water values of the models closely matched the experimental values in the range of $k_bT \approx 2.5$ kJ mol^{-1}, which is considered to be a good agreement for a Martini molecule parametrization.[87] The resulting bead types for the sulfonamide group and the HCB groups were P5 (strong polarity) and X1

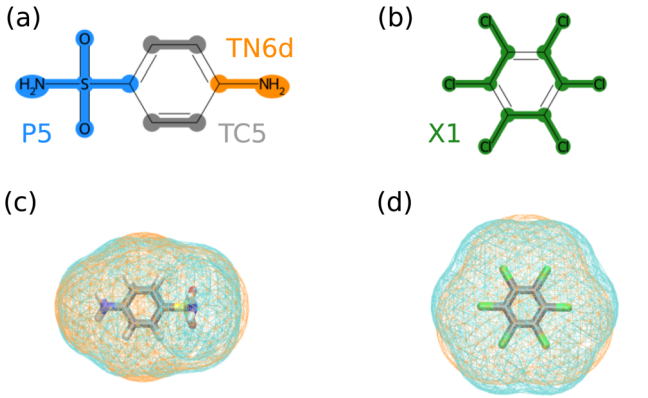


Fig. 1 Applied mapping to SAA (a) and HCB (b). The different colors and labels respectively represent the corresponding mapped groups and Martini bead types. Connolly surfaces of SAA (c) and HCB (d) are depicted for the atomistic (cyan) and coarse-grained models (orange).

Open Access Article. Published on 26 September 2025. Downloaded on 10/24/2025 2:41:56 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Table 2 Comparison of model predictions with experimental and reference values for SAA and HCB. Octanol–water transfer free energy values (ΔG) are compared to experimental values,[88] while the solvent-accessible surface area (SASA) values are compared to values from atomistic simulations. The energy values are in kJ mol^{-1} and surface areas in nm^2. The experimental transfer free energies were derived from partition coefficients

	SAA		HCB	
	Model	Reference	Model	Reference
ΔG	−5.27	−3.54	35.47	32.71
SASA	4.15	4.07	4.37	4.37

(halogen, lowest polarity), respectively, correctly reflecting the polarity of the corresponding groups. In accordance with the Martini 3 parametrization approach, no hydrogen bonding label was applied for the bead representing the sulfonamide group because it can act as both a donor and an acceptor.[69] In Martini, the hydrogen bonding labels can influence the interaction with other beads to mimic an overall reduced or increased attraction. However, it is important to note that the information about the directionality of such interactions cannot be reproduced by the coarse-grained model in any case.

The SASA values between the atomistic and coarse-grained models of the pollutants can be considered to be in sufficient agreement. The bond distance between the HCB groups obtained from the Swarm-CG optimization was increased by 12.7% to achieve a better agreement of the SASA values and to optimize the match between the Connolly surfaces of the atomistic and coarse-grained model. Further information can be found in Section S1 of the SI.

3.2 Diffusion coefficients

To assess the mobility of pollutants, diffusion coefficients were determined based on the mean squared displacements (MSDs) of SAA and HCB. Details regarding the calculation can be found in Section S3 (SI). Fig. 2 presents the diffusion coefficients of the water beads (blue), SAA (orange), and HCB (green) across different SOM types. Please note that the error bars in Fig. 2 and all following Figures containing bar plots are the standard errors arising from the averaging over the 20 individual trajectories for each case.

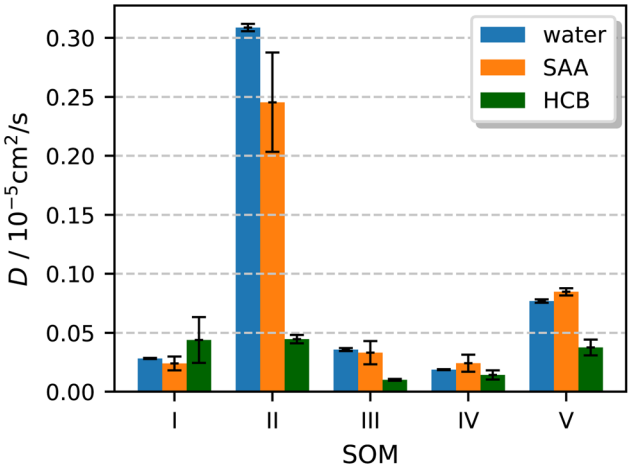


Fig. 2 Diffusion coefficients of the water beads (blue), SAA (orange), and HCB (green) for the different SOM types.

For each SOM system, the diffusion of SAA correlates with that of water, with the highest diffusion coefficient observed in the SOM II system. The strong influence of water diffusion on SAA suggests that SAA molecules either spend extended periods in the water phase or remain in continuous contact with water. The elevated water diffusion in the SOM II system can be attributed to the system's more hydrophobic nature (low acidity), which promotes local phase separation and results in a more interconnected water phase,[68] thereby facilitating higher water mobility.

The overall diffusion of HCB is slower, reflecting its high accumulation potential in SOM.[50,51] The lowest diffusion coefficient is observed in the SOM III system. This result aligns with a theoretical study on HCB, which reported relatively low diffusion in Leonardite humic acid (LHA),[54] corresponding to SOM III in this study. In contrast, in the SOM I system, HCB diffusion closely matches that of water, indicating that HCB is predominantly located in the water phase for this SOM type. Conversely, SOM II exhibits the greatest disparity between HCB and water diffusion, probably due to a strong accumulation of HCB in the HS phase.

It is important to note that Martini coarse-grained simulations are carried out without explicit frictional or stochastic forces, unlike methods based on the generalized Langevin equation.[89–91] This omission leads to artificially accelerated dynamics, as the loss of atomistic-level friction prevents proper dissipation of energy. While such acceleration can be advantageous for exploring long-timescale conformational changes, it compromises the accuracy of real-time dynamics compared to atomistic models. Furthermore, the degree of acceleration may vary depending on the coarse-graining resolution and mapping scheme. Several approaches have been developed to quantify or correct for this acceleration, including relative entropy methods,[92,93] excess entropy scaling relationships,[94–97] biasing potentials[98,99] and assessments based on molecular roughness.[100,101]

Therefore, the absolute diffusion coefficients shown in Fig. 2 are not directly comparable to experimental values if such data were available. Because coarse-graining in Martini models typically accelerates dynamics by roughly a factor of 3–8,[102] relative speed-ups of SAA and HCB are not expected to differ by orders of magnitude. Nevertheless, the precise magnitude of this artificial acceleration may vary between the two models. Moreover, the environmental heterogeneity in SOM (*e.g.*, highly charged *vs.* non-polar domains) may alter the effective friction and thus the magnitude of the coarse-grained speed-up. Consequently, direct comparison of their diffusion constants should be limited to qualitative trends rather than precise quantitative differences.

3.3 Coordination numbers

To assess the overall partitioning of the pollutants within the SOM systems and its connection to diffusion behavior, the

© 2025 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 26 September 2025. Downloaded on 10/24/2025 2:41:56 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

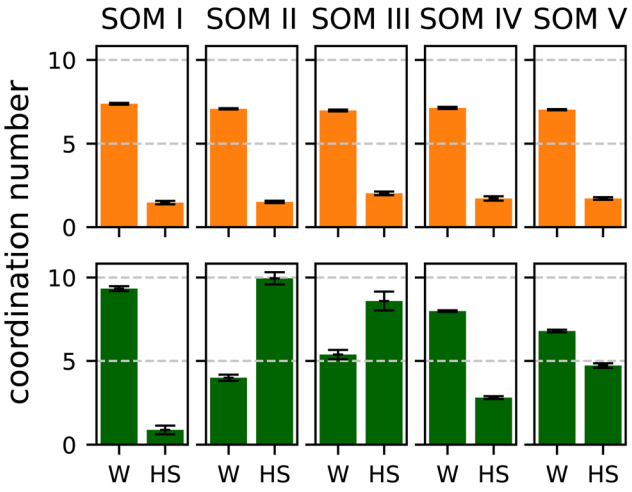


Fig. 3 Coordination numbers of water (W) and HS beads around SAA (orange) and HCB (green), averaged over the 20 trajectories for each pollutant and each SOM type.

average coordination numbers of water and HS beads around SAA and HCB were calculated and illustrated in Fig. 3 (see Section S4 (SI) for details of the calculation). The coordination number is a measure of how many beads of water or HS the pollutant molecules are surrounded by. Note that one water bead is representative of four atomistic water molecules.

SAA exhibits similar behavior across different SOM types, with a slightly higher value for the HS beads in the SOM III system. Compared to HS, the coordination number of water is approximately four to five times higher, indicating a relatively low but non-negligible SAA concentration within the HS phase. This explains the observed correlation between the calculated diffusion coefficients of SAA and water in Fig. 2. Furthermore, the results suggest that at high water contents, SAA diffusion is primarily influenced not by direct interactions with HS molecules but rather by the structural and spatial arrangement of HS molecules, which modulates the mobility and flow of the surrounding water beads.

In contrast, the coordination number of HS beads around HCB varies significantly depending on the SOM type. The highest value is observed for the SOM II (low acidity) and SOM III (high aromaticity) systems. A notable contribution is also evident in the highly aliphatic SOM (SOM V), while the lowest coordination number is found in SOM I, which exhibits high acidity. Overall, HCB has a higher concentration in the HS phase compared to SAA, suggesting that HCB has a stronger binding affinity to SOM. This distinct behavior of SAA and HCB in the SOM II system is illustrated in Fig. 4. Unlike SAA, HCB diffusion is strongly influenced by direct HCB–SOM interactions.

Interestingly, due to the low coordination number of HS beads, SOM I constitutes an exception, displaying a relatively low probability of HCB residing in the HS phase. This explains why, in SOM I, HCB exhibits diffusion comparable to that of water (see Fig. 2). Across all SOM systems, HS and water molecules are not perfectly mixed; instead, larger interconnected

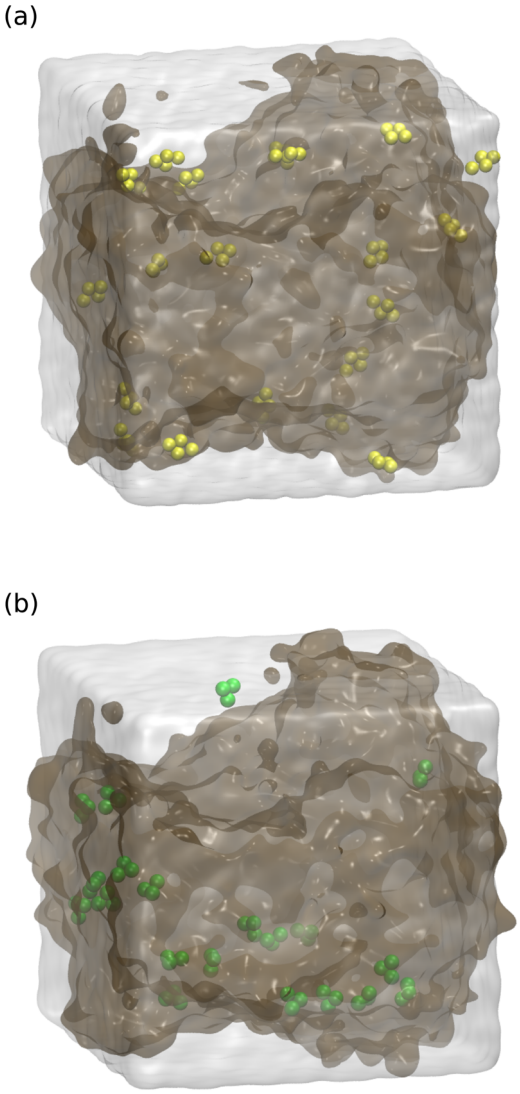


Fig. 4 Merged final frames from 20 independent replicas of SOM II. Colored spheres mark the pollutant bead positions in each replica's final frame; SAA: yellow in (a), HCB: green in (b). The brown (HS) and translucent white (water) density isosurfaces were obtained by averaging bead positions across the 20 final frames, then contouring at a fixed threshold to indicate regions frequently occupied by HS and water. The surfaces are replica-endpoint densities (not time averages); pollutant spheres are not averaged. HCB endpoints predominantly fall within HS-rich regions, whereas SAA endpoints are often in water.

water regions exist, as previously noted for SOM II. The size of these regions is likely to be determined by the specific composition and arrangement of HS molecules. Particularly in SOM I, HCB diffuses within this water phase rather than integrating into the matrix formed by HS molecules, as identified from simulation trajectories.

At first glance, this behavior appears contradictory given HCB's strong hydrophobicity. However, it is presumably caused by the high polarity of the abundant carboxylate groups in SOM I, which prevent HCB from being absorbed into the HS phase. Moreover, the variations in coordination number across the HS phase can be linked to the acidity of the SOM system. High-

Open Access Article. Published on 26 September 2025. Downloaded on 10/24/2025 2:41:56 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

acidity SOM (*e.g.*, SOM I) is associated with low hydrophobicity of HS molecules, while low-acidity SOM (*e.g.*, SOM II) exhibits higher hydrophobicity. This emphasizes the crucial role of HCB's hydrophobicity in its accumulation within SOM.

3.4 Interaction energies with functional groups

In order to characterize the interaction of the pollutants with SOM in more detail, interaction energies between the pollutant molecules and HS functional groups were calculated and plotted for the groups with the highest contributions to the interaction (see Fig. 5), considering the whole 1.5 μs of each of the production runs.

The negatively charged carboxylate groups exhibit the strongest interaction with SAA in each system. Furthermore, ketone, hydroxyl, phenol, acid anhydride, and aromatic functional groups play an important role. Overall, the preference for these functional groups can be explained by the hydrogen

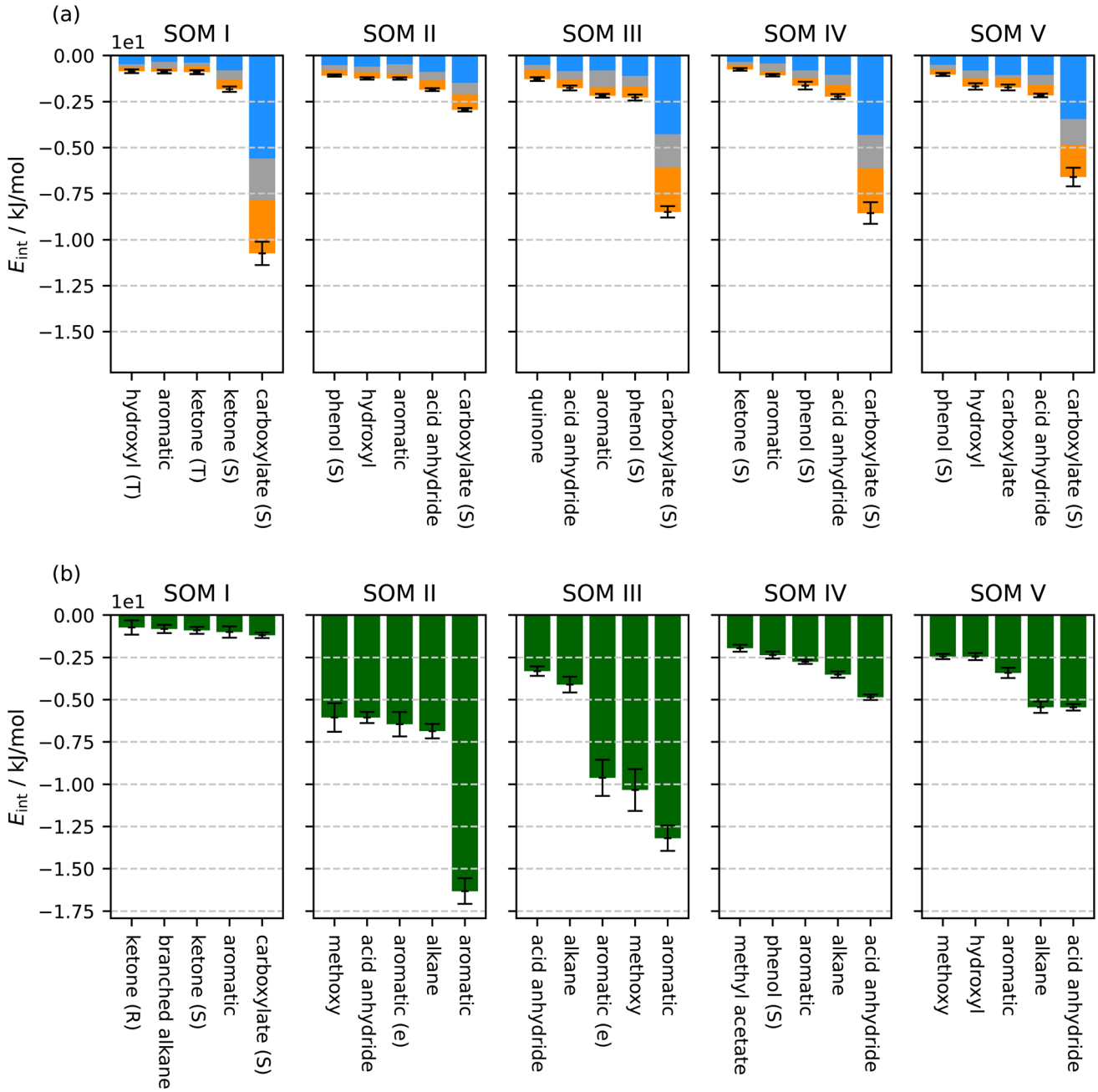


Fig. 5 Interaction energies (time-averaged) between SAA (a) or HCB (b) and beads representing different functional groups (*x*-axis). Values are further averaged across the 20 independent trajectories for every pollutant and SOM type. Proportions of the sulfonamide group (blue), aromatic ring (gray), and amino group (orange) are depicted for SAA. Green represents the HCB interaction energies. If available for a group, "T" or "S" stands for the Martini size of the bead, namely tiny or small, respectively. In the case of aromatic beads, a distinction is also made between those representing electron-rich regions of a ring, which are denoted by the letter "e".

 © 2025 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 26 September 2025. Downloaded on 10/24/2025 2:41:56 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

bonding ability of SAA and the hydrophilic character of the sulfonamide group. Conversely, the aromatic nature of SAA is reflected in the attraction to aromatic beads of the HSs, which is the most hydrophobic functional group in Fig. 5 that appears for SAA. The overall order for these groups is mostly determined by their abundance within each SOM system.

Another notable feature of SAA is its site-specific adsorption to the HSs, which is most apparent in the SOM III case when comparing the contribution of the sulfonamide group (blue) across different functional groups. The site-specific adsorption is characterized by selective interactions of the aromatic ring or sulfonamide group with specific functional groups or sites within SOM. The ratio $\xi = E_{\text{int,sulf.}}/E_{\text{int,arom.}}$ of interaction energies of the sulfonamide group compared to the aromatic part with the amino group (gray + orange) in SOM III is as follows for the five depicted functional groups, quinone, acid anhydride, aromatic, phenol, and carboxylate: 1.89, 1.94, 0.90, 1.88, and 2.36. This indicates, for example, a stronger attraction of the sulfonamide group to the negatively charged carboxylate groups ($\xi = 2.36$), and a higher attraction of the aromatic ring to the aromatic parts of the HS molecules ($\xi = 0.90$). This feature of SAA in the SOM interaction was reported by theoretical and experimental investigations[51,52] and is reproduced by the coarse-grained model, thus further validating the present parameters.

In comparison, for HCB, significant differences in the distribution of functional groups are visible. The hydrophobic aromatic and alkane groups show high contributions in every SOM type. The SOM I case shows HCB exhibiting exceptional behavior, whereby higher proportions of hydrophilic groups, *e.g.*, carboxylate and ketone groups, are also recognizable, but where the overall values are also low. The higher interaction energies occurring in SOM II and III underline the importance of aromatic and alkane groups for the interaction of HCB with the HS structures. Overall, the interaction energies reflect the strong hydrophobicity of HCB, which alters the distribution of preferred functional groups compared to SAA.

3.5 Pollutant trapping

To explore the potential of HS structures to fixate pollutant molecules, an analysis of the trapping of the pollutants was conducted. This could be realized using quantities such as mean-first-passage times[103] or residence time distributions.[104] However, these require systematically defined positions or volumes within the SOM systems, which is impractical in light of the large total number of trajectories. Instead, the time intervals along the CGMD trajectories wherein the pollutant molecules were fixated were determined using the pollutants' spatial displacements and a hidden Markov model[105] (please refer to Section S7 (SI) for details about the calculations containing values of the displacements and identified trapping times). This type of trapping interval, where SAA remains fixed in the SOM matrix for a certain time, is illustrated in the trajectory snapshots in Fig. 6. The ratio of trapped to the total simulation time (trapping probability) for the different SOM types and pollutants is depicted in Fig. 7.

In the case of SAA, the trapping probabilities are relatively high for SOM I, III, and IV. These systems exhibit a relatively low water diffusion (please compare to Fig. 2), which correlates with a higher probability of the molecules being trapped. In contrast, a low trapping probability is visible for SOM II, which is most likely connected to the high diffusion of SAA in this system, caused by the higher water diffusion. It is therefore assumed that the strong influence of the diffusion of water on SAA, indicated by the diffusion coefficients, is directly reflected in the probability of SAA being trapped in the SOM matrix, where water diffusion and trapping probability are anti-correlated. This is further supported by a strong negative Pearson correlation ($r = -0.90$) between water diffusion coefficients and SAA

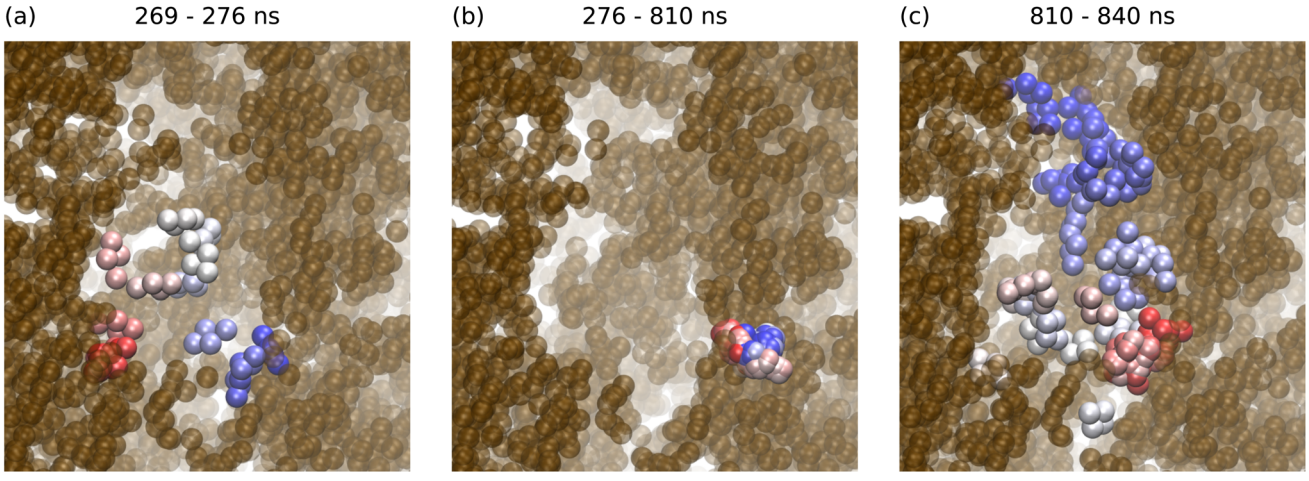


Fig. 6 Visualization of trapping and release of an SAA molecule in the SOM I system. The SAA trajectory is color-coded from red to blue, representing the movement from start to end point in the respective time interval. The HS molecules are depicted in brown, whereas the water beads are omitted for clarity. (a) SAA migrates from the bulk aqueous phase (red) into the HS matrix (blue). (b) The molecule then remains virtually stationary, illustrating its trapped state within the matrix. (c) Finally, SAA escapes, moving from the trapped location (red) back into the water phase (blue).

Open Access Article. Published on 26 September 2025. Downloaded on 10/24/2025 2:41:56 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

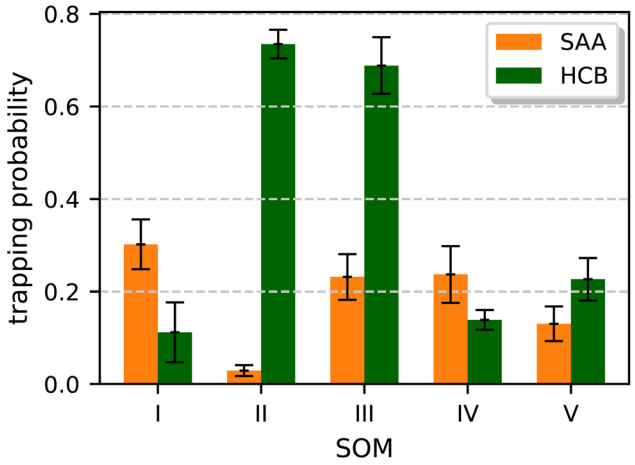


Fig. 7 Trapping probabilities of SAA (orange) and HCB (green) for the different SOM types.

trapping probabilities. The results further indicate that the molecules do not follow a continuous diffusion, but rather go through sequential intervals of trapping and free diffusion. This behavior can only be observed using domain sampling methods such as CGMD simulations, which are facilitated by the coarse-grained representation of the molecules.

In contrast, the trapping probability of HCB correlates more with the coordination numbers of HS beads, *i.e.*, with the probability of HCB being found in the HS phase (Fig. 3), emphasizing the influence of the direct interaction of HCB with SOM on diffusion. The Pearson correlation coefficient between coordination numbers and HCB trapping probabilities is $r = 0.97$, indicating a strong positive relationship.

However, the factors that influence the diffusion of the pollutant molecules in SOM appear to be complex. For example, when comparing SOM III and IV, an interesting behavior emerges: SOM III induces a much higher trapping probability than SOM IV, although the diffusion of HCB is similar in both cases. This suggests that the low diffusion of HCB can be caused by different mechanisms, which would be either the direct trapping of the molecule (SOM III) or the slowdown of the molecules *via* the interactions with the HS molecules without trapping (SOM IV). The difference between HCB and water diffusion for the respective cases, which is higher for the SOM III case, should probably also be taken into account here. Furthermore, it is assumed that the high trapping probabilities (above 0.5) are connected to the strong interaction between HCB and the HS aromatic compounds. This is supported by a complementary analysis of ring–ring contacts, which showed that trapping events coincide with an increased frequency of π–π stacking configurations (see Section S8 in the SI).

It is important to note that the SOM models used in this work represent highly hydrated, diluted systems. As such, the results concerning trapping and diffusion should primarily be interpreted in the context of such hydration levels. Since water facilitates molecular mobility, lower-moisture systems would be expected to exhibit reduced diffusion rates and increased trapping probabilities, particularly for polar pollutants such as SAA. Therefore, the observed interactions and mobility patterns may differ under drier or more heterogeneous soil conditions.

Using the presented methodology, molecular-level insights into how pollutants interact with SOM functional groups and how SOM composition influences pollutant mobility can be revealed, which has further implications for remediation strategies. Identifying preferential binding sites of pollutants can guide the design of amendments that either immobilize contaminants or enhance their degradation.[106] Future studies should aim to systematically characterize these binding preferences and define the conditions under which different pollutants are stabilized or mobilized, thereby allowing predictions of how amendments may alter these processes.

3.6 Implications of the Martini representation

The coarse-grained simplification that underpins the Martini models accelerates simulations, but inevitably sacrifices chemical resolution and can lead to artifacts in the simulation of the molecules. Many of the shortcomings of the Martini 2 force field have been alleviated in Martini 3. For instance, in Martini 2, the models' interactions caused an overstabilization of biomolecular interactions for soluble[107] and membrane proteins,[108,109] and also for carbohydrates.[110–112] The consequence was a too stable self-association in solution and an exaggerated tendency of membrane proteins to aggregate. Martini 3 alleviated this "stickiness" through a global reparametrization of bead interactions,[69,112] yet overcompaction for intrinsically disordered or multidomain proteins was noted.[113]

A second artifact arose when very short bond distances or mixed bead sizes were employed: the resulting high density of interaction sites generated artificial desolvation barriers and overstabilized dimers.[111] Contrary to previous versions, Martini 3 introduces a "size–shaped" mapping together with additional new bead sizes, aimed to restore the correct molecular volume and to improve the oil–water partitioning of small molecules.[69,112] This type of mapping was applied to all molecules investigated in this study. However, although Martini 3 remedies many of the deficiencies identified in the earlier version, one should continue to critically examine whether the remaining approximations are acceptable for the physicochemical properties under investigation.

Coarse-graining eliminates certain configurational degrees of freedom and thus reduces the intrinsic molecular entropy. In the Martini framework, this entropy loss is empirically compensated by enhancing non-bonded interaction strengths so that resulting free energies match reference experimental or atomistic data. While this ensures accurate reproduction of free energy differences, it distorts the enthalpy–entropy balance, making coarse-grained interaction energies model-specific and not directly interpretable as physical enthalpies.[114–118] Therefore, the absolute values in Fig. 5 should be interpreted with caution and used in combination with non-energetic measures such as coordination numbers (Fig. S6 in the SI). A more rigorous treatment would involve calculating sorption free energies, similar to the approach of Petrov *et al.*,[119] who investigated small organic compounds in LHA systems using

© 2025 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 26 September 2025. Downloaded on 10/24/2025 2:41:56 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

atomistic simulations. This approach could also be applied to drier SOM systems, which exhibit slower solute diffusion as a result of reduced water activity.[120,121] While this was beyond the computational scope of the present study, such approaches could be explored in future work to quantify specific pollutant–SOM interactions in a statistically rigorous framework.

3.7 Comparison with literature

Experimental and computational studies have shown that HCB and SAA become trapped in SOM, including reports of irreversible sorption into SOM voids.[24,122,123] While the coarse-grained HS model used here cannot represent the full structural and chemical complexity of natural SOM, it reproduces key qualitative patterns of pollutant immobilization. This consistency provides a mesoscale confirmation of known behaviors and highlights the utility of coarse-grained approaches in bridging molecular insight with environmental timescales.

Ahmed *et al.*[51] combined sorption experiments with all-atom MD simulations for simplified SOM models to show that HCB binds more strongly to SOM than SAA, attributed to HCB's hydrophobicity and aromaticity, which agrees with the calculated coordination numbers (Fig. 3) and interaction energies (Fig. 5). In a related density functional theory (DFT) study, Ahmed *et al.*[52] demonstrated that water destabilizes SAA–SOM complexes and that SAA shows a strong affinity for ionic sites, consistent with the observations of the present study.

Furthermore, previous work indicates that SAA sorption onto SOM is strongly composition-dependent.[15,51] This behavior was not observed in the present study, probably because the high water content favored SAA partitioning into the aqueous phase. Future work on varying water content or pH with coarse-grained simulations may further clarify these differences. Regarding HCB, DFT calculations on HCB–SOM interactions revealed that it strongly binds to functional groups such as alkylated aromatics, phenol, and lignin.[50] In the present study, similar contributions from aromatic, alkane, and phenol groups were evident in the binding of HCB.

4 Conclusions

While pollutants continue to pose a persistent threat to terrestrial ecosystems, their molecular–scale interactions with SOM remain only partially understood, a gap this study helps to address. Using CGMD simulations, we investigated the mobility of SAA and HCB in SOM systems of varied composition.

Diffusion analysis revealed contrasting behavior: SAA followed closely the diffusion of water beads, whereas HCB displayed reduced diffusion except in systems with high carboxylate content. Based on the calculation of coordination numbers, the partitioning behavior of the pollutants indicated that SAA's distribution between water and HS phases was nearly composition-independent, while HCB accumulated in HS matrices containing low carboxyl and low carbonyl content, or high aromatic content.

Interaction-energy calculations revealed a preference of SAA to carboxylate groups and other hydrophilic functionalities. Aromatic moieties also contributed, displaying a site-specific interaction of SAA to HS. For HCB, interactions were dominated by contacts with aromatic moieties, alkane chains, as well as methoxy and acid-anhydride functional groups.

Analysis using a hidden Markov model resolved distinct trapping states for both pollutants, encompassing fixation within the SOM matrix and periods of free diffusion. The extent of SAA trapping was anticorrelated with water diffusion, underscoring the influence of solvent mobility on its behavior. In contrast, HCB trapping correlated with encounters with HS molecules, indicating direct interactions with the HS matrix as an immobilizing factor.

The coarse-grained approach employed here enables the investigation of dynamic processes over extended time and length scales, thereby complementing previous studies and integrating into the current research framework of pollutants in soil. The method offers broad potential for future applications, including studies of varied SOM compositions, the integration of other pre-parametrized pollutants from the Martini 3 framework, and the investigation of large-scale SOM structural organization.

Conflicts of interest

There are no conflicts to declare.

Data availability

All topology files required to reproduce the simulations from this study are available in Zenodo at DOI: **https://doi.org/10.5281/zenodo.14749817**

Data supporting this article have been included as part of the Supplementary Information (SI). See DOI: **https://doi.org/10.1039/d5va00237k**. The SI includes details on the pollutant parametrization and HS samples, additional results on coordination numbers and interaction energies, and analyses of π–π stacking configurations between pollutants and SOM.

References

1 R. R. Weil and N. C. Brady, *The Nature and Properties of Soils*, Pearson, Upper Saddle River, NJ, 15th edn, 2016.
2 B. D. Hudson, *J. Soil Water Conserv.*, 1994, **49**, 189–194.
3 W. Rawls, Y. A. Pachepsky, J. Ritchie, T. Sobecki and H. Bloodworth, *Geoderma*, 2003, **116**, 61–76.
4 Food and Agriculture Organization of the United Nations, *Plant Nutrition for Food Security, a Guide for Integrated Nutrient Management*, Food & Agriculture Organization of the United Nations (FAO), Rome, Italy, 2006.
5 P. K. R. Nair, B. M. Kumar and V. D. Nair, in *Soil Organic Matter (SOM) and Nutrient Cycling*, Springer International Publishing, 2021, pp. 383–411.
6 B. Xu, G. Yang, A. Lehmann, S. Riedel and M. C. Rillig, *Soil Ecol. Lett.*, 2022, **5**, 108–117.
7 M. N. Hanif, N. Aijaz, K. Azam, M. Akhtar, W. A. Laftah, M. Babur, N. K. Abbood and I. B. Benitez, *Int. J. Environ. Sci. Technol.*, 2024, **21**, 10277–10318.

Open Access Article. Published on 26 September 2025. Downloaded on 10/24/2025 2:41:56 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

8 L. Ritter, K. R. Solomon, J. Forget, M. Stemeroff and C. O'Leary, *Persistent Organic Pollutants: an Assessment Report on DDT, Aldrin, Dieldrin, Endrin, Chlordane, Heptachlor, Hexachlorobenzene, Mirex, Toxaphene, Polychlorinated Biphenyls, Dioxins and Furans*, International Programme on Chemical Safety (IPCS), United Nations Environment Programme Technical Report PCS/95.39, 1995.
9 K. C. Jones and P. De Voogt, *Environ. Pollut.*, 1999, **100**, 209–221.
10 T. Luarte, V. A. Gómez-Aburto, I. Poblete-Castro, E. Castro-Nallar, N. Huneeus, M. Molina-Montenegro, C. Egas, G. Azcune, A. Pérez-Parada, R. Lohmann, P. Bohlin-Nizzetto, J. Dachs, S. Bengtson-Nash, G. Chiang, K. Pozo and C. J. Galbán-Malagón, *Atmos. Chem. Phys.*, 2023, **23**, 8103–8118.
11 J. L. Wilkinson, A. B. A. Boxall, D. W. Kolpin, K. M. Y. Leung, R. W. S. Lai, C. Galbán-Malagón, A. D. Adell, J. Mondon, M. Metian, R. A. Marchant, A. Bouzas-Monroy, A. Cuni-Sanchez, A. Coors, P. Carriquiriborde, M. Rojo, C. Gordon, M. Cara, M. Moermond, T. Luarte, V. Petrosyan, Y. Perikhanyan, C. S. Mahon, C. J. McGurk, T. Hofmann, T. Kormoker, V. Iniguez, J. Guzman-Otazo, J. L. Tavares, F. G. D. Figueiredo, M. T. P. Razzolini, V. Dougnon, G. Gbaguidi, O. Traoré, J. M. Blais, L. E. Kimpe, M. Wong, D. Wong, R. Ntchantcho, J. Pizarro, G.-G. Ying, C.-E. Chen, M. Páez, J. Martínez-Lara, J.-P. Otamonga, J. Poté, S. A. Ifo, P. Wilson, S. Echeverría-Sáenz, N. Udikovic-Kolic, M. Milakovic, D. Fatta-Kassinos, L. Ioannou-Ttofa, V. Belušová, J. Vymazal, M. Cárdenas-Bustamante, B. A. Kassa, J. Garric, A. Chaumot, P. Gibba, I. Kunchulia, S. Seidensticker, G. Lyberatos, H. P. Halldórsson, M. Melling, T. Shashidhar, M. Lamba, A. Nastiti, A. Supriatin, N. Pourang, A. Abedini, O. Abdullah, S. S. Gharbia, F. Pilla, B. Chefetz, T. Topaz, K. M. Yao, B. Aubakirova, R. Beisenova, L. Olaka, J. K. Mulu, P. Chatanga, V. Ntuli, N. T. Blama, S. Sherif, A. Z. Aris, L. J. Looi, M. Niang, S. T. Traore, R. Oldenkamp, O. Ogunbanwo, M. Ashfaq, M. Iqbal, Z. Abdeen, A. O'Dea, J. M. Morales-Saldaña, M. Custodio, H. de la Cruz, I. Navarrete, F. Carvalho, A. B. Gogra, B. M. Koroma, V. Cerkvenik-Flajs, M. Gombač, M. Thwala, K. Choi, H. Kang, J. L. C. Ladu, A. Rico, P. Amerasinghe, A. Sobek, G. Horlitz, A. K. Zenker, A. C. King, J.-J. Jiang, R. Kariuki, M. Tumbo, U. Tezel, T. T. Onay, J. B. Lejju, Y. Vystavna, Y. Vergeles, H. Heinzen, A. Pérez-Parada, D. B. Sims, M. Figy, D. Good and C. Teta, *Proc. Natl. Acad. Sci. U. S. A.*, 2022, **119**, e2113947119.
12 B. Gworek, M. Kijeńska, J. Wrzosek and M. Graniewska, *Water, Air, Soil Pollut.*, 2021, **232**, 145.
13 W. Zheng and M. Guo, *Curr. Pollut. Rep.*, 2021, **7**, 510–523.
14 L. Wang, R. Govind and R. A. Dobbs, *Environ. Sci. Technol.*, 1993, **27**, 152–158.
15 S. Thiele-Bruhn, T. Seibicke, H. Schulten and P. Leinweber, *J. Environ. Qual.*, 2004, **33**, 1331–1342.
16 J. Gao and J. A. Pedersen, *Environ. Sci. Technol.*, 2005, **39**, 9509–9516.
17 A. A. MacKay and B. Canterbury, *J. Environ. Qual.*, 2005, **34**, 1964–1971.
18 Q. Wang, M. Guo and S. R. Yates, *J. Agric. Food Chem.*, 2005, **54**, 157–163.
19 T. L. ter Laak, W. A. Gebbink and J. Tolls, *Environ. Toxicol. Chem.*, 2006, **25**, 904–911.
20 S. T. Kurwadkar, C. D. Adams, M. T. Meyer and D. W. Kolpin, *J. Agric. Food Chem.*, 2007, **55**, 1370–1376.
21 W. Lertpaitoonpan, S. K. Ong and T. B. Moorman, *Chemosphere*, 2009, **76**, 558–564.
22 R. A. Figueroa-Diva, D. Vasudevan and A. A. MacKay, *Chemosphere*, 2010, **79**, 786–793.
23 A. Białk-Bielińska, J. Maszkowska, W. Mrozik, A. Bielawska, M. Kołodziejska, R. Palavinskas, P. Stepnowski and J. Kumirska, *Chemosphere*, 2012, **86**, 1059–1065.
24 J. Schwarz, S. Thiele-Bruhn, K.-U. Eckhardt and H.-R. Schulten, *ISRN Soil Sci.*, 2012, **2012**, 1–10.
25 T. Müller, I. Rosendahl, A. Focks, J. Siemens, J. Klasmeier and M. Matthies, *Environ. Pollut.*, 2013, **172**, 180–185.
26 W. J. Farmer, M. S. Yang, J. Letey and W. F. Spencer, *Soil Sci. Soc. Am. J.*, 1980, **44**, 676–680.
27 G. Deane, Z. Chroneer and W. Lick, *J. Environ. Eng.*, 1999, **125**, 689–696.
28 S. N. Meijer, W. A. Ockenden, A. Sweetman, K. Breivik, J. O. Grimalt and K. C. Jones, *Environ. Sci. Technol.*, 2003, **37**, 667–672.
29 Y.-X. Chen, H.-L. Chen, Y.-T. Xu and M.-W. Shen, *Environ. Int.*, 2004, **30**, 31–37.
30 J. D. Kubicki, *Environ. Sci. Technol.*, 2006, **40**, 2298–2303.
31 D. Tunega, M. H. Gerzabek, G. Haberhauer, K. U. Totsche and H. Lischka, *J. Colloid Interface Sci.*, 2009, **330**, 244–249.
32 A. A. Ahmed, O. Kühn and P. Leinweber, *Sci. Total Environ.*, 2012, **441**, 151–158.
33 H. Zhang, Y. Wang, C. Sun, M. Yu, Y. Gao, T. Wang, J. Liu and G. Jiang, *Environ. Sci. Technol.*, 2014, **48**, 1525–1531.
34 L. Böhm, P. Grančič, E. Scholtzová, B. J. Heyde, R.-A. Düring, J. Siemens, M. H. Gerzabek and D. Tunega, *Environ. Sci. Pollut. Res.*, 2022, **30**, 36824–36837.
35 J. Zhang, Z. Zhou, L. Zeng, C. Wang, R. Han, X. Ren, W. Wang, M. Xiang, S. Chen and H. Li, *Sci. Total Environ.*, 2024, **947**, 174657.
36 S. Zeng, Z. Dai, B. Ma, R. A. Dahlgren and J. Xu, *Earth Critical Zone*, 2024, **1**, 100002.
37 International Humic Substances Society, **https://humic-substances.org**.
38 J. Gerke, *Agronomy*, 2018, **8**, 76.
39 M. Kleber and J. Lehmann, *J. Environ. Qual.*, 2019, **48**, 207–216.
40 M. H. Hayes and R. S. Swift, in *Vindication of Humic Substances as a Key Component of Organic Matter in Soil and Water*, Elsevier, 2020, pp. 1–37.
41 T. C. de Aguiar, D. F. de Oliveira Torchia, T. A. van Tol de Castro, O. C. H. Tavares, S. de Abreu Lopes, L. de Souza da Silva, R. N. Castro, R. L. L. Berbara, M. G. Pereira and A. C. García, *Sci. Total Environ.*, 2022, **833**, 155133.

 © 2025 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 26 September 2025. Downloaded on 10/24/2025 2:41:56 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

42 D. C. Olk, P. R. Bloom, E. M. Perdue, D. M. McKnight, Y. Chen, A. Farenhorst, N. Senesi, Y. Chin, P. Schmitt-Kopplin, N. Hertkorn and M. Harir, *J. Environ. Qual.*, 2019, **48**, 217–232.
43 D. C. Olk, P. R. Bloom, M. De Nobili, Y. Chen, D. M. McKnight, M. J. M. Wells and J. Weber, *J. Environ. Qual.*, 2019, **48**, 1633–1643.
44 H. R. Schulten and M. Schnitzer, *Naturwissenschaften*, 1995, **82**, 487–498.
45 H.-R. Schulten and M. Schnitzer, *Soil Sci.*, 1997, **162**, 115–130.
46 H.-R. Schulten and P. Leinweber, *Biol. Fertil. Soils*, 2000, **30**, 399–432.
47 H.-R. Schulten, *Soil Mineral-Organic Matter-Microorganism Interactions and Ecosystem Health, Dynamics, Mobility and Transformation of Pollutants and Nutrients*, Elsevier, 2002, pp. 351–381.
48 A. J. A. Aquino, D. Tunega, G. Haberhauer, M. H. Gerzabek and H. Lischka, *Eur. J. Soil Sci.*, 2007, **58**, 889–899.
49 A. J. A. Aquino, D. Tunega, G. E. Schaumann, G. Haberhauer, M. H. Gerzabek and H. Lischka, *J. Phys. Chem. C*, 2009, **113**, 16468–16475.
50 A. A. Ahmed, O. Kühn, S. G. Aziz, R. H. Hilal and P. Leinweber, *Sci. Total Environ.*, 2014, **476–477**, 98–106.
51 A. A. Ahmed, S. Thiele-Bruhn, S. G. Aziz, R. H. Hilal, S. A. Elroby, A. O. Al-Youbi, P. Leinweber and O. Kühn, *Sci. Total Environ.*, 2015, **508**, 276–287.
52 A. A. Ahmed, S. Thiele-Bruhn, P. Leinweber and O. Kühn, *Sci. Total Environ.*, 2016, **559**, 347–355.
53 P. Gros, A. Ahmed, O. Kühn and P. Leinweber, *Sci. Total Environ.*, 2017, **586**, 527–535.
54 H. Feng, H. Zhang, H. Cao, Y. Sun, A. Zhang and J. Fu, *Environ. Sci. Technol.*, 2018, **52**, 14228–14234.
55 Q. Xue, Z. Jiao, X. Liu, W. Pan, J. Fu and A. Zhang, *Environ. Sci. Technol.*, 2023, **58**, 1531–1540.
56 A. Sündermann, R. Solc, D. Tunega, G. Haberhauer, M. H. Gerzabek and C. Oostenbrink, *J. Mol. Graphics Modell.*, 2015, **62**, 253–261.
57 Y. Escalona, D. Petrov and C. Oostenbrink, *J. Mol. Graphics Modell.*, 2021, **103**, 107817.
58 M. H. Gerzabek, A. J. A. Aquino, Y. I. E. Balboa, E. Galicia-Andrés, P. Grančič, C. Oostenbrink, D. Petrov and D. Tunega, *J. Plant Nutr. Soil Sci.*, 2022, **185**, 44–59.
59 V. A. Harmandaris, N. P. Adhikari, N. F. A. van der Vegt and K. Kremer, *Macromolecules*, 2006, **39**, 6708–6719.
60 W. G. Noid, *J. Chem. Phys.*, 2013, **139**, 090901.
61 S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid and A. Kolinski, *Chem. Rev.*, 2016, **116**, 7898–7936.
62 J. Jin, A. J. Pak, A. E. P. Durumeric, T. D. Loose and G. A. Voth, *J. Chem. Theory Comput.*, 2022, **18**, 5759–5791.
63 M. Orsi and J. W. Essex, *PLoS One*, 2011, **6**, e28637.
64 L. Darré, M. R. Machado, A. F. Brandner, H. C. González, S. Ferreira and S. Pantano, *J. Chem. Theory Comput.*, 2015, **11**, 723–739.
65 M. R. Machado, E. E. Barrera, F. Klein, M. Sóñora, S. Silva and S. Pantano, *J. Chem. Theory Comput.*, 2019, **15**, 2719–2733.
66 S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. de Vries, *J. Phys. Chem. B*, 2007, **111**, 7812–7824.
67 L. F. Dettmann, O. Kühn and A. A. Ahmed, *J. Chem. Theory Comput.*, 2024, **20**, 5291–5305.
68 L. F. Dettmann, O. Kühn and A. A. Ahmed, *J. Chem. Theory Comput.*, 2024, **20**, 10684–10696.
69 P. C. T. Souza, R. Alessandri, J. Barnoud, S. Thallmair, I. Faustino, F. Grünewald, I. Patmanidis, H. Abdizadeh, B. M. H. Bruininks, T. A. Wassenaar, P. C. Kroon, J. Melcr, V. Nieto, V. Corradi, H. M. Khan, J. Domański, M. Javanainen, H. Martinez-Seara, N. Reuter, R. B. Best, I. Vattulainen, L. Monticelli, X. Periole, D. P. Tieleman, A. H. de Vries and S. J. Marrink, *Nat. Methods*, 2021, **18**, 382–388.
70 R. Alessandri, J. Barnoud, A. S. Gertsen, I. Patmanidis, A. H. de Vries, P. C. T. Souza and S. J. Marrink, *Adv. Theory Simul.*, 2021, **5**, 2100391.
71 F. Grünewald, M. H. Punt, E. E. Jefferys, P. A. Vainikka, M. König, V. Virtanen, T. A. Meyer, W. Pezeshkian, A. J. Gormley, M. Karonen, M. S. P. Sansom, P. C. T. Souza and S. J. Marrink, *J. Chem. Theory Comput.*, 2022, **18**, 7555–7569.
72 C. Empereur-mot, K. B. Pedersen, R. Capelli, M. Crippa, C. Caruso, M. Perrone, P. C. T. Souza, S. J. Marrink and G. M. Pavan, *J. Chem. Inf. Model.*, 2023, **63**, 3827–3838.
73 L. Borges-Araújo, G. P. Pereira, M. Valério and P. C. Souza, *Biochim. Biophys. Acta, Proteins Proteomics*, 2024, **1872**, 141014.
74 A. F. Brandner, I. P. S. Smith, S. J. Marrink, P. C. T. Souza and S. Khalid, *J. Chem. Inf. Model.*, 2025, **65**, 1537–1548.
75 F. Grünewald, R. Alessandri, P. C. Kroon, L. Monticelli, P. C. T. Souza and S. J. Marrink, *Nat. Commun.*, 2022, **13**, 68.
76 R. Shrestha, R. Alessandri, M. Vögele, C. Hilpert, P. Souza, S. J. Marrink and L. Monticelli, *J. Chem. Theory Comput.*, 2025, **21**, 9035–9053.
77 C. Empereur-Mot, L. Pesce, G. Doni, D. Bochicchio, R. Capelli, C. Perego and G. M. Pavan, *ACS Omega*, 2020, **5**, 32823–32843.
78 C. H. Bennett, *J. Comput. Phys.*, 1976, **22**, 245–268.
79 Y. Escalona, D. Petrov, E. Galicia-Andrés and C. Oostenbrink, *Agronomy*, 2023, **13**, 1044.
80 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1–2**, 19–25.
81 C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke and T. E. Oliphant, *Nature*, 2020, **585**, 357–362.
82 J. D. Hunter, *Comput. Sci. Eng.*, 2007, **9**, 90–95.
83 N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, *J. Comput. Chem.*, 2011, **32**, 2319–2327.
84 R. Gowers, M. Linke, J. Barnoud, T. Reddy, M. Melo, S. Seyler, J. Domański, D. Dotson, S. Buchoux, I. Kenney and O. Beckstein, *Proceedings of the Python in Science Conference*, 2016.

Open Access Article. Published on 26 September 2025. Downloaded on 10/24/2025 2:41:56 PM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

85 F. Pérez and B. E. Granger, *Comput. Sci. Eng.*, 2007, **9**, 21–29.
86 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.
87 R. Alessandri, S. Thallmair, C. G. Herrero, R. Mera-Adasme, S. J. Marrink and P. C. T. Souza, *A Practical Guide to Recent Advances in Multiscale Modeling and Simulation of Biomolecules*, AIP Publishing LLC Melville, New York, 2023, pp. 1–34.
88 C. Hansch, A. Leo and D. Hoekman, *Exploring QSAR: Hydrophobic, Electronic, and Steric Constants*, American Chemical Society, Washington, D.C., 1995, vol. 2, pp. 15–22.
89 G. Jung, M. Hanke and F. Schmid, *J. Chem. Theory Comput.*, 2017, **13**, 2481–2488.
90 V. Klippenstein, M. Tripathy, G. Jung, F. Schmid and N. F. A. van der Vegt, *J. Phys. Chem. B*, 2021, **125**, 4931–4954.
91 T. Schilling, *Phys. Rep.*, 2022, **972**, 1–45.
92 M. S. Shell, *J. Chem. Phys.*, 2008, **129**, 144108.
93 M. S. Shell, *J. Chem. Phys.*, 2012, **137**, 084503.
94 Y. Rosenfeld, *Phys. Rev. A*, 1977, **15**, 2545–2549.
95 J. C. Dyre, *J. Chem. Phys.*, 2018, **149**, 210901.
96 G. G. Rondina, M. C. Böhm and F. Müller-Plathe, *J. Chem. Theory Comput.*, 2020, **16**, 1431–1447.
97 J. Jin, K. S. Schweizer and G. A. Voth, *J. Chem. Phys.*, 2023, **158**, 034103.
98 P. K. Depa and J. K. Maranas, *J. Chem. Phys.*, 2005, **123**, 094901.
99 D. Fritz, K. Koschke, V. A. Harmandaris, N. F. A. van der Vegt and K. Kremer, *Phys. Chem. Chem. Phys.*, 2011, **13**, 10412.
100 M. K. Meinel and F. Müller-Plathe, *J. Chem. Theory Comput.*, 2020, **16**, 1411–1419.
101 M. K. Meinel and F. Müller-Plathe, *J. Phys. Chem. B*, 2022, **126**, 3737–3747.
102 P. C. T. Souza, S. Thallmair, P. Conflitti, C. Ramírez-Palacios, R. Alessandri, S. Raniolo, V. Limongelli and S. J. Marrink, *Nat. Commun.*, 2020, **11**, 3714.
103 S. Redner, *A Guide to First-Passage Processes*, Cambridge University Press, 2001.
104 P. Danckwerts, *Chem. Eng. Sci.*, 1953, **2**, 1–13.
105 L. Rabiner, *Proc. IEEE*, 1989, **77**, 257–286.
106 F. J. Larney and D. A. Angers, *Can. J. Soil Sci.*, 2012, **92**, 19–38.
107 A. C. Stark, C. T. Andrews and A. H. Elcock, *J. Chem. Theory Comput.*, 2013, **9**, 4176–4185.
108 M. Javanainen, H. Martinez-Seara and I. Vattulainen, *PLoS One*, 2017, **12**, e0187936.
109 A. Majumder and J. E. Straub, *J. Chem. Theory Comput.*, 2021, **17**, 2513–2521.
110 P. S. Schmalhorst, F. Deluweit, R. Scherrers, C.-P. Heisenberg and M. Sikora, *J. Chem. Theory Comput.*, 2017, **13**, 5039–5053.
111 R. Alessandri, P. C. T. Souza, S. Thallmair, M. N. Melo, A. H. de Vries and S. J. Marrink, *J. Chem. Theory Comput.*, 2019, **15**, 5448–5460.
112 S. J. Marrink, L. Monticelli, M. N. Melo, R. Alessandri, D. P. Tieleman and P. C. T. Souza, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2022, **13**, e1620.
113 F. E. Thomasen, T. Skaalum, A. Kumar, S. Srinivasan, S. Vanni and K. Lindorff-Larsen, *Nat. Commun.*, 2024, **15**, 6645.
114 N. J. H. Dunn, T. T. Foley and W. G. Noid, *Acc. Chem. Res.*, 2016, **49**, 2832–2840.
115 J. Jin, A. J. Pak and G. A. Voth, *J. Phys. Chem. Lett.*, 2019, **10**, 4549–4557.
116 K. M. Kidder, R. J. Szukalo and W. G. Noid, *Eur. Phys. J. B*, 2021, **94**, 153.
117 Z. Jarin, J. Newhouse and G. A. Voth, *J. Chem. Theory Comput.*, 2021, **17**, 1170–1180.
118 T. D. Loose, P. G. Sahrmann, T. S. Qu and G. A. Voth, *J. Chem. Theory Comput.*, 2024, **20**, 9190–9208.
119 D. Petrov, D. Tunega, M. H. Gerzabek and C. Oostenbrink, *Eur. J. Soil Sci.*, 2020, **71**, 831–844.
120 V. P. Kalinitchenko, A. V. Swidsinski, A. P. Glinushkin, V. P. Meshalkin, S. V. Gudkov, T. M. Minkina, V. V. Chernenko, V. D. Rajput, S. S. Mandzhieva, S. N. Sushkova, A. A. Okolelova and A. A. Shestakova, *Environ. Geochem. Health*, 2023, **45**, 8967–8987.
121 V. P. Kalinitchenko, A. V. Swidsinski, A. P. Glinushkin, M. R. Overcash, V. P. Meshalkin, T. M. Minkina, V. D. Rajput, V. V. Chernenko, S. S. Mandzhieva, S. N. Sushkova, A. E. Rykhlik, G. S. Larin, S. V. Gudkov, M. G. Baryshev, M. A. Sevostyanov and D. A. Makarenkov, *Land Degrad. Dev.*, 2024, **35**, 2367–2381.
122 A. Kan, W. Chen and M. Tomson, *Environ. Pollut.*, 2000, **108**, 81–89.
123 J.-j. Zhang, B. Wen, X.-q. Shan, S. Zhang and S. U. Khan, *Environ. Pollut.*, 2007, **150**, 234–242.

 © 2025 The Author(s). Published by the Royal Society of Chemistry

Supplementary Information (SI) for Environmental Science: Advances.
This journal is © The Royal Society of Chemistry 2025

Supplementary Information

Coarse-Grained Simulations of Sulfanilamide and Hexachlorobenzene Mobility in Soil Organic Matter

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: oliver.kuehn@uni-rostock.de

S1 Details of the pollutant parametrization

S1.1 Modeling approach

The pollutants Sulfanilamide (SAA) and Hexachlorobenzene (HCB) were parametrized within the framework of the Martini 3 force field[1]. First, a mapping of the molecules was undertaken, which corresponds to a translation from the atomistic to the coarse-grained representation. In the Martini framework, between two and four non-hydrogen atoms are grouped into one bead, with the possibility of minor deviations. Among other factors, the mapping approach was chosen to reflect the symmetry of the two molecules. Predefined Martini beads were then applied to these atom groups. In general, a bead should represent the size and polarity

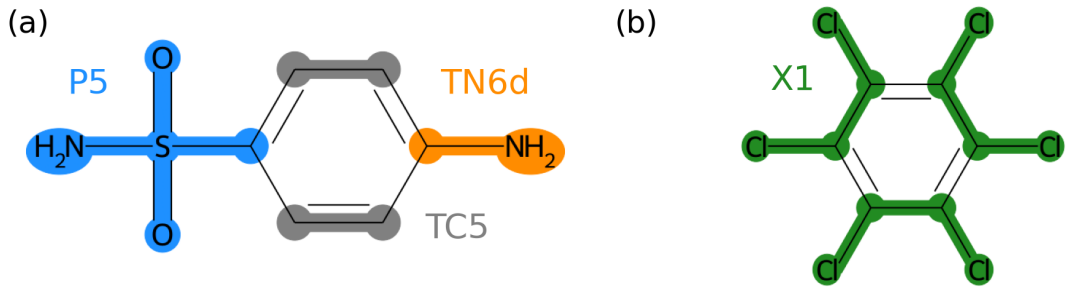


Figure S1: Mapping and bead types of the Martini models of SAA (a) and HCB (b).

of the underlying group, which is determined from the bead's label. A comprehensive explanation of these labels can be found in the supplementary information for the main Martini 3 publication[1]. The initial selection of a bead for a certain group is based on molecules that have already been parametrized within the Martini framework. If an atom group has not been parametrized before, or if the bead labels need to be fine-tuned, the choice of a bead is mainly determined by the reproduction of the (experimental) partitioning data for the molecule, especially the free energy of transfer ΔG. The selected mapping of HCB and SAA and the corresponding bead types are depicted in Figure S1. Traditionally, experimental octanol-water (OC-W) ΔG values have been used as a reference for the parametrization of a molecule in this framework. During the course of a reparametrization in version 3 of Martini, hexadecane-water (HD-W) and chloroform-water (CHL-W) partitioning data were utilized as well. Here, the experimental OC-W ΔG values were calculated from the corresponding partition coefficients $\log(P)$, using the following equation:

$$\Delta G = RT \cdot \ln(P) = RT \cdot \ln(10) \cdot \log(P), \tag{1}$$

with the universal gas constant R and the temperature $T = 298.15\,\mathrm{K}$. Furthermore, these values were supplemented with HD-W and CHL-W ΔGs, determined using the COSMO-RS method[2–4]. The corresponding ΔG values of the models were determined by free energy calculations. The setup for these calculations was identical to that used for the fragment molecules in Dettmann et al.[5]. It involved the utilization of the Bennett acceptance ratio[6]

method and the *gmx bar* tool. For further computational details, please refer to Section 2.3.3. of this study.

The bonded interactions of the coarse-grained models were determined using the Swarm-CG tool[7]. More specifically, in water solvated HCB and SAA were simulated in their all-atom representation, using the GROMOS force field version 54A7, for 25 ns each. The edge length of the water box was 4 nm in both cases. Other parameters of the atomistic simulations were the same as for the parametrization of the fragment molecules of the Vienna Soil-Organic-Matter Modeler 2. Please refer to Section 2.3.1. of the corresponding study[5].

S1.2 Results of the parametrization

In addition to the reproduction of thermodynamic data, the parametrization of a new molecule involves comparing shape properties of the coarse-grained model with an atomistic reference. Among these are the solvent-accessible surface area (SASA) and the Connolly surface. These quantities were determined using the *gmx sasa* tool. Van der Waals radii from Rowland and Taylor's work[8] were used for the atomistic molecules. The comparison of the Connolly surfaces between the atomistic and coarse-grained representations of HCB and SAA is depicted in Figure S2. The coarse-grained model of SAA sufficiently reproduces the Connolly surface of the atomistic reference (Figure S2 (a)). For HCB, a larger deviation is visible in directions orthogonal to the ring (Figure S2 (b)). This is most likely connected to the 4:1 mapping, and therefore to the application of regular sized beads. HCB is a flat ring molecule and comprised of a relatively high number of atoms, which makes the reproduction of the Connolly surfaces in all directions using spherical beads challenging. An alternative approach to the mapping could involve the utilization of smaller beads with the addition of a virtual site bead at the center of the molecule, as was done for the parametrization of carbohydrates[9]. However, three regular sized beads were applied here because, with the introduction of an additional bead, the correct choice of bead types through a parametrization would require additional and more complex reference data. Furthermore, the bulk properties

of HCB are not considered in this study, where such shape mismatches would presumably become more relevant. To optimize the coarse-grained SASA of HCB (Figure S3) the bond distances were scaled by a factor of 12.7 %. The scaling has a minor influence on the Connolly surfaces (see Figure S2 (c) for the model with the scaled bond distances).

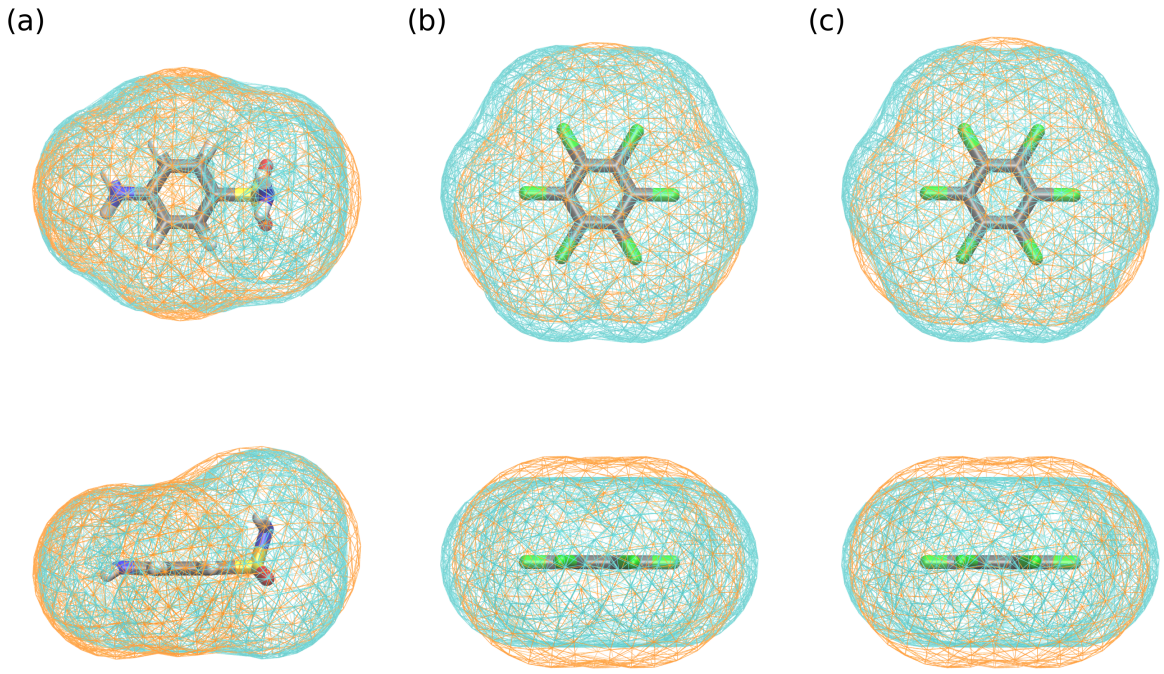


Figure S2: Connolly surfaces of atomistic (cyan) and coarse-grained (orange) molecules. Depicted are a view from the top and from the side respectively for SAA (a), HCB (b) and HCB with scaled bond distances (c).

In the case of SAA, the standard beads for the amino group (TN6d) and the -C=C- parts of the aromatic ring (TC5) were applied. These beads have been frequently used in other Martini models for the aforementioned groups. The bead for the sulfonamide group was chosen to reproduce, together with the other beads, the experimental OC-W ΔG of the molecule as closely as possible. This resulted in the P5 bead, which is a reasonable outcome considering the high polarity of this group. The sulfonamide group can be either a hydrogen bond acceptor or a donor. Therefore, no additional bead label (a/d) was applied to describe the preference of this group as an acceptor or donor, as only one of these labels

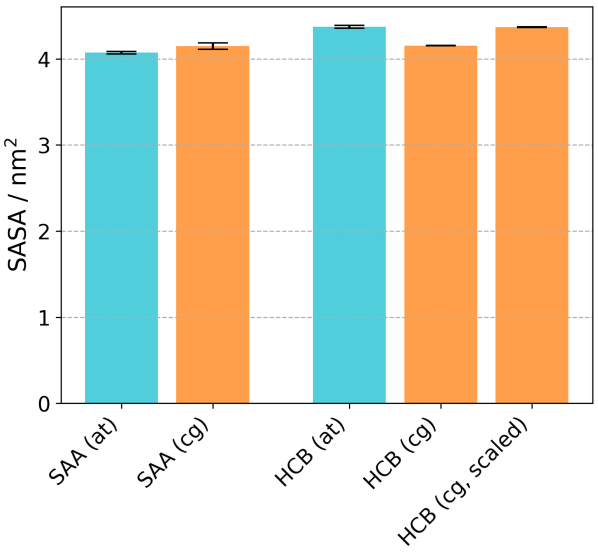


Figure S3: Comparison between the SASA values of the atomistic (at, cyan) and coarse-grained (cg, orange) versions of the pollutant molecules. Additionally, for HCB, the SASA value resulting from scaling the bond distances is depicted. The Martini models for SAA and HCB with the scaled bond distances show a good agreement with the atomistic reference.

can be applied. The difference between the experimental and calculated OC-W value is at around $1.73\,\mathrm{kJ\,mol^{-1}}$. Thus, the model's energy difference is below $k_b T \approx 2.5\,\mathrm{kJ\,mol^{-1}}$ and can therefore be considered to be in close agreement with the experimental value[10]. The calculated ΔG values deviate more from those determined by COSMO-RS. However, a higher difference is acceptable, as the values from COSMO-RS itself usually reproduce the experimental values with slight deviations. Therefore, higher deviations can be expected, as suggested by one of our previous investigations, in which free energy values of Martini models were compared to experimental values and to those determined using COSMO-RS[5]. It should be noted that as the experimental value is the most accurate reference, only this value was used to determine the optimal bead. The values determined with COSMO-RS were utilized for additional verification regarding the trends for the HD-W and CHL-W values.

Regarding the parametrization of HCB, Martini beads belonging to the halogen group were considered (bead labels are denoted with an X). These beads were designed to optimally describe the non-bonded interactions of organic halogen compounds. Out of all the regular

sized halogen beads available, the X1 bead, applied to the three HCB groups, reproduced its OC-W ΔG most accurately. As the most hydrophobic bead in this group, it accurately reflects the strong hydrophobicity of the HCB molecule. With a difference of 2.76 kJ mol^{-1} from the experimental value, the deviation is slightly above, but very close to k_bT, which could therefore still be regarded as a high level of agreement. The calculated HD-W and CHL-W values also correctly reproduce the trends for the values determined with COSMO-RS.

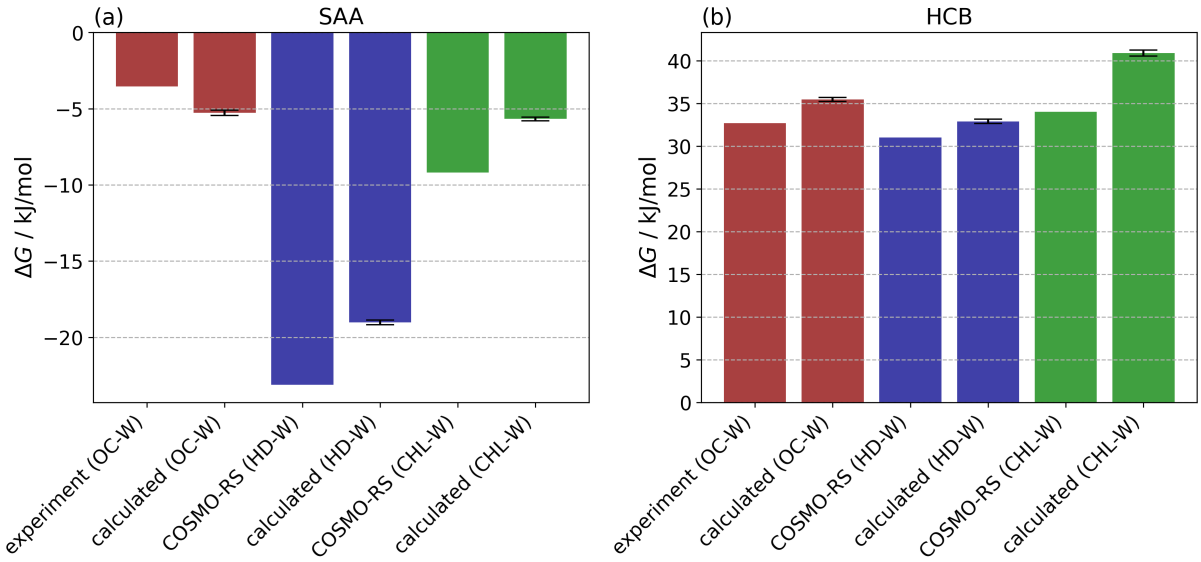


Figure S4: Comparison of ΔG values of SAA (a) and HCB (b) to reference values. OC-W values are compared to experimental ones[11,12], while HD-W and CHL-W values are compared to ΔGs determined with COSMO-RS.

Benchmarking the pollutant-SOM interactions against experimental sorption free energies or diffusion coefficients would ideally require well-defined SOM systems, such as IHSS reference materials, and specialized experimental setups. However, obtaining such data is inherently challenging due to the complex, amorphous, and chemically heterogeneous nature of SOM, which contains diverse functional groups and binding environments. Unlike crystalline materials, SOM lacks a regular structure, making interaction modeling and interpretation difficult. Environmental variables like moisture, pH, and competing solutes further complicate experimental measurements. While this represents a valuable direction for fu-

ture work, it lies beyond the scope of the current study. Instead, the parametrization in this study is based on the experimentally determined transfer free energies, which remain the most robust and transferable dataset available for modeling pollutant–SOM interactions within the Martini force field framework.

Table S1: Bead types that occur in the SOM models with the corresponding functional groups that they represent. The TN2a bead is used to represent the methoxy and furan functional groups.

bead	functional group	bead	functional group
Q5n	carboxylate	TN4	secondary amine
SQ5n	carboxylate	N4a	methyl acetate
P2	carboxylic acid	TN4a	aldehyde
SP2	carboxylic acid	TN3ar	ether
P2a	acid anhydride	TN2a	methoxy
P1	hydroxyl	TN2a*	furan
SP1	hydroxyl	SC6	thiol
TP1	hydroxyl	TC5	aromatic
SP1r	hydroxyl	TC5e	aromatic
SN6	phenol	C4h	alkene
TN6	phenol	C2	branched alkane
TN6a	quinone	SC2	branched alkane
SN6d	primary amine	C1	alkane
TN6d	primary amine		
N5a	ketone		
SN5a	ketone		
TN5a	ketone		

Table S2: Number of different bead types in the SOM systems. The functional groups represented by the respective beads are listed in Table S1.

Type	SOM I	SOM II	SOM III	SOM IV	SOM V
Q5n	40	251	13	106	310
SQ5n	1376	539	1093	1154	826
P2	3	31	1	12	20
SP2	108	69	89	126	95
P2a	43	257	184	265	247
P1	39	231	34	84	246
SP1	90	427	5	254	364
TP1	635	322	51	497	622
SP1r	0	90	5	59	5
SN6	183	330	820	452	307
TN6	45	299	153	130	49
TN6a	168	4	374	46	62
SN6d	3	82	16	34	8
TN6d	6	171	37	70	8
N5a	61	12	152	25	61
SN5a	563	189	144	323	244
TN5a	602	8	66	96	140
TN4	466	140	166	217	100
N4a	9	27	38	44	46
TN4a	31	1	137	33	15
TN3ar	60	294	121	190	111
TN2a	0	117	99	15	69
TN2a*	2	12	28	15	4
SC6	31	31	78	39	19
TC5	667	750	1289	777	432
TC5e	379	248	450	375	158
C4h	29	93	18	53	170
C2	50	55	38	34	38
SC2	53	15	60	52	81
C1	2	44	30	20	49

S2 Further computational details

The simulation parameters were identical to those used in our previous work[13], except for the production run, which lasted for 1.5 µs in the present study. Before the production run, an energy minimization and an NPT equilibration of 10 ns were conducted for each system. In order to integrate the equations of motion, the leap-frog integrator was employed in conjunction with a time step of 20 fs. The temperature was set to $T = 298.15$ K by utilizing the velocity rescale thermostat[14], with a coupling constant of $\tau_T = 1$ ps. Using a pressure of 1 bar as a reference value, the Berendsen barostat[15] with a coupling constant of $\tau_p = 8$ ps was applied for the NPT equilibration, whereas the Parrinello-Rahman barostat[16] with a coupling constant of 12 ps was used for the production run. The Lennard-Jones potentials were calculated with the Verlet cut-off scheme, using a straight cutoff of 1.1 nm. The reaction-field electrostatics were utilized for evaluating the Coulomb interaction, applying a cutoff of 1.1 nm and dielectric constant of $\epsilon_r = 15$. Constraints were applied through the use of the LINCS algorithm. All simulations were performed with periodic boundary conditions.

S3 Determination of the diffusion coefficients

The diffusion coefficients of the pollutant molecules were calculated using their mean squared displacements (MSDs). To obtain a statistically sound diffusion coefficient, for each pollutant and each SOM case, the trajectories of the pollutant molecules were unwrapped and merged into one trajectory using the Python package MDAnalysis[17,18]. The diffusion coefficients, shown in Figure 2 in the main manuscript, were determined from the merged trajectories using the GROMACS[19] tool *gmx msd*. For each pollutant and each SOM type, the MSD curves of the individual pollutant molecules, which were also determined using *gmx msd*, and the MSD of the merged trajectory (averaged) are plotted in Figure S5.

The apparent short-term decrease observed in individual MSD curves does not imply non-physical behavior, but rather reflects statistical fluctuations inherent to single-molecule trajectories. Molecules may transiently revisit regions near their starting positions, causing

temporary reductions in MSD values. Averaging over all molecules compensates for these local reversals, yielding an ensemble-averaged MSD (red curves) that exhibits the expected monotonic growth characteristic of diffusive motion.

In some cases, the averaged MSD curves exhibit a slight downward curvature near the end of the trajectory. This is a well-known artifact arising from the use of multiple time origins in the MSD calculation, where statistical noise increases at longer lag times due to the reduced number of available time origins. To avoid the influence of this artifact on the diffusion coefficient estimation, the linear fit is restricted to the 10 %–90 % range of the MSD curve.

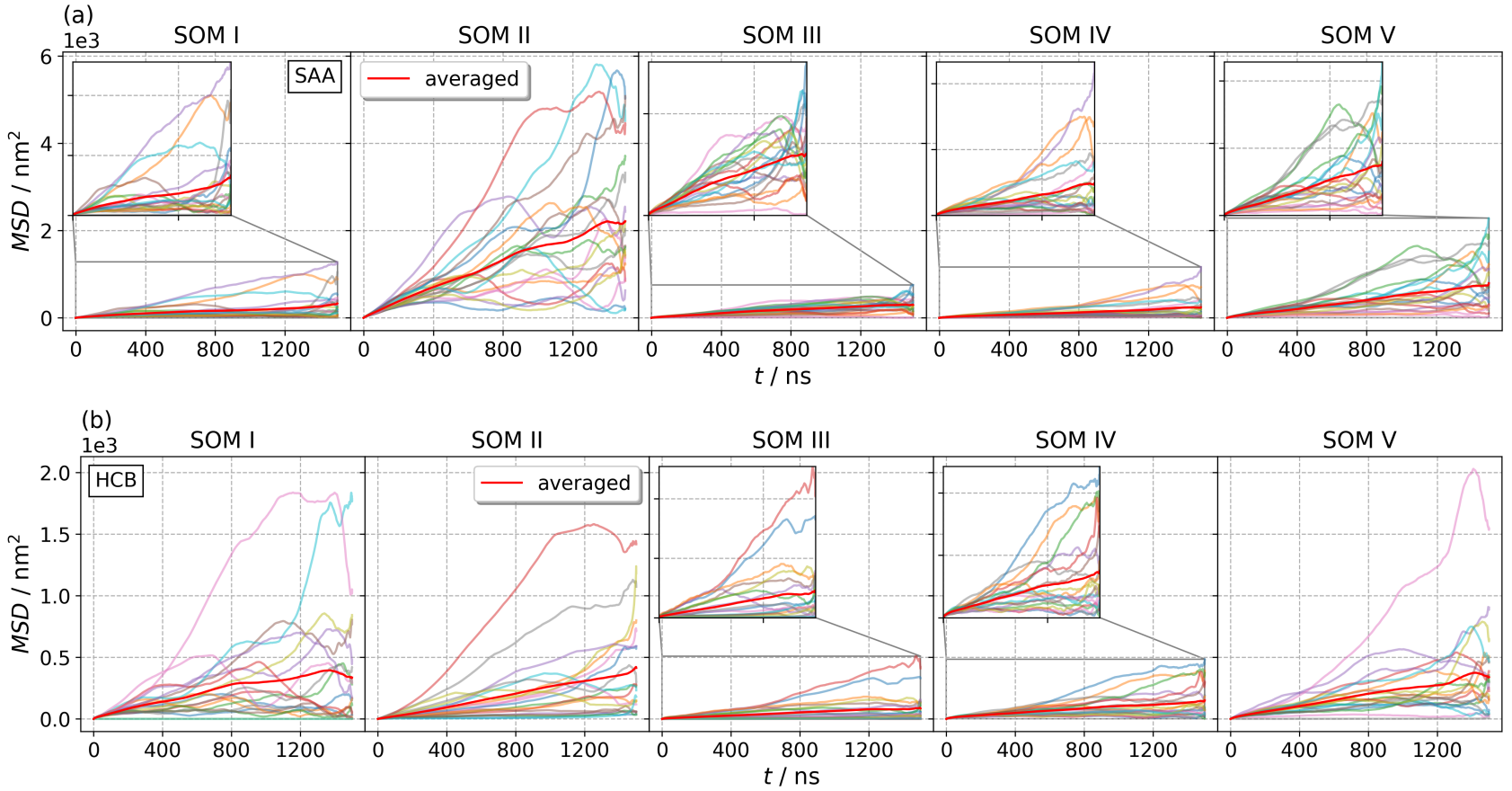


Figure S5: Mean squared displacements (MSDs) for SAA (a) and HCB (b) and each SOM type. In each case, the red MSD curve (averaged) shows the MSD of the merged trajectory of the 20 individual trajectories of the pollutant molecules. Please note the zoomed-in regions in the case of the smaller MSDs for clearer visibility.

S4 Determination of coordination numbers

For the coordination numbers, the *gmx rdf* tool with the *-cn* flag was used, considering the whole trajectory of each production run in the analysis. The relationship between the coordination number n_c and the radial distribution function $g(r)$ is given as follows:

$$n_c = 4\pi\rho \int_0^{R_c} g(r) r^2 dr \tag{2}$$

with ρ being the number density of the target beads. The cutoff R_c that was set to 1.5 times the van der Waals radii between a pollutant bead and a bead from the rest of the system, that was considered. For example, the van der Waals radius determined from the Martini bead sizes between a regular-sized bead of an HCB molecule and a small bead of an HS molecule is 0.43 nm (please compare to the Martini bead sizes illustrated in the supplementary information for the Martini 3 publication[1], Figure 2). If the two beads have a distance of 1.5 times 0.43 nm or smaller, the HS bead was taken into account in the calculation. Finally, the results were averaged over the 20 individual runs of a pollutant interacting with one SOM system. The coordination numbers of the full spectrum of HS beads of a certain type around a pollutant molecule is shown in Figure S7.

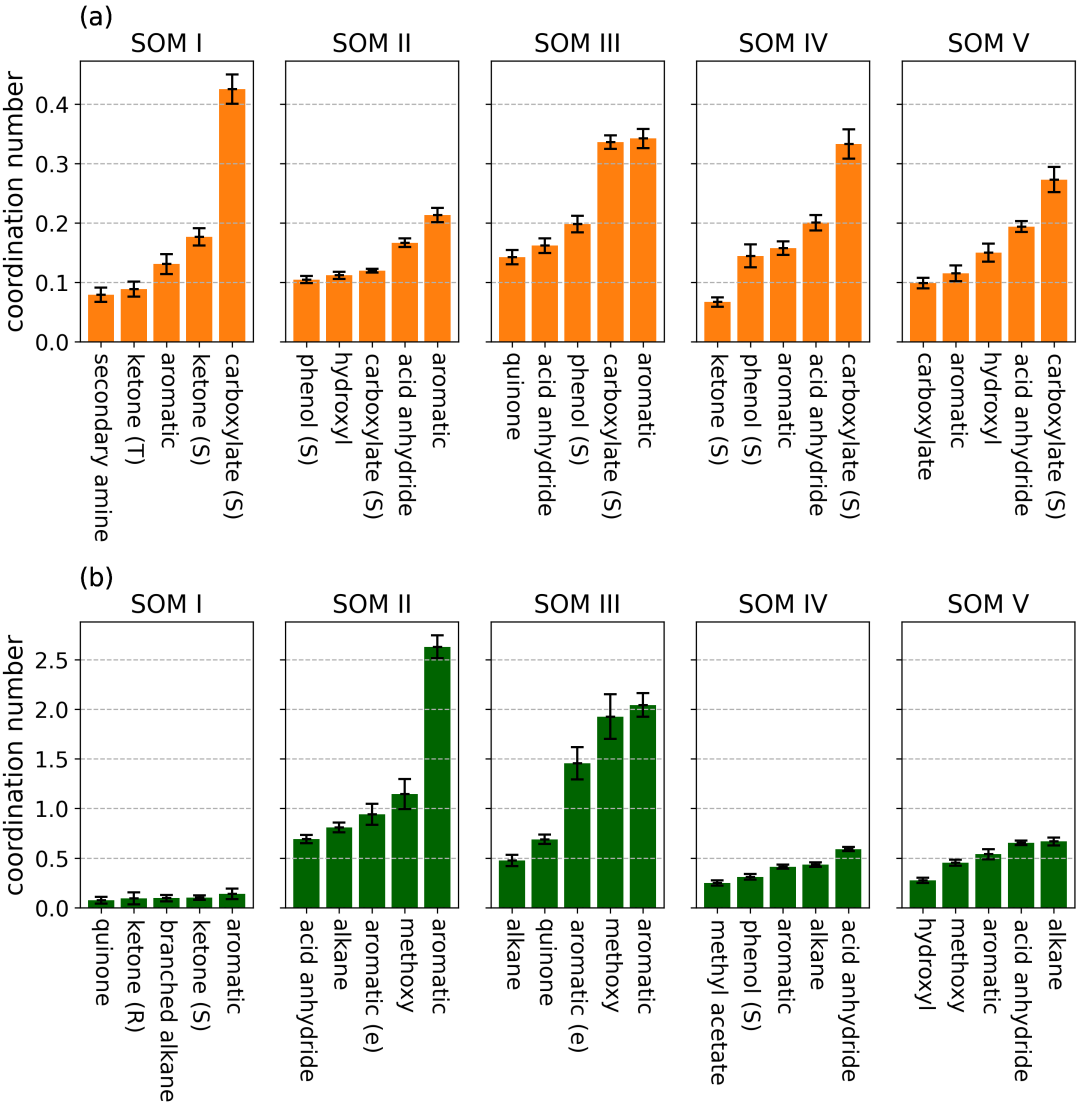


Figure S6: Coordination numbers of beads from different HS functional groups around SAA (a) and HCB (b), averaged over the 20 trajectories for each pollutant and each SOM type. Please note the different scales on the y-axes for SAA and HCB.



Figure S7: Coordination numbers of HS beads of a certain type around SAA and HCB for the different SOM systems. The results were averaged over the 20 individual runs for each pollutant and each SOM type. The translations of the bead types into the respective SOM functional groups are shown in Table S1. Please note that calcium and water beads are not taken into account here, and therefore the information about the extent to which the respective pollutant is found in the HS phase is not shown. This information can be obtained from the main manuscript (see Figure 3). Furthermore, please note the different color scales for SAA and HCB.

S5 Interaction energies

The interaction energies provide insight into the degree of attraction or repulsion between pollutants and SOM functional groups, or SOM as a whole. They potentially shed light on which functional groups contribute the most to pollutant accumulation in SOM, as these energies, studied in the present work, generally correlate with the corresponding coordination numbers between the compounds. The interaction energy between two chemical compounds A and B can be defined as:

$$E_{int} = E_{A,B} - (E_A + E_B) \tag{3}$$

where E_A and E_B are the total energies of the isolated compounds, and $E_{A,B}$ is the total energy of their interacting complex[20]. In the case of the classical potential energy function used for the coarse-grained simulations, the interaction energy reduces to the sum of the Lennard-Jones and Coulomb potential energies between the two compounds. The GROMACS tool *gmx energy* was utilized to determine the time-averaged interaction energies between the pollutant and the SOM compounds.

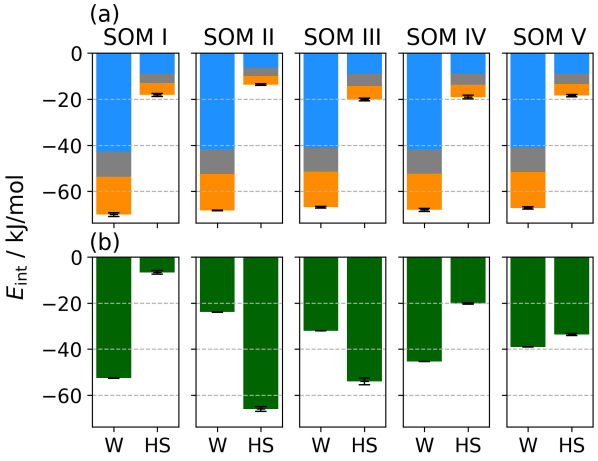
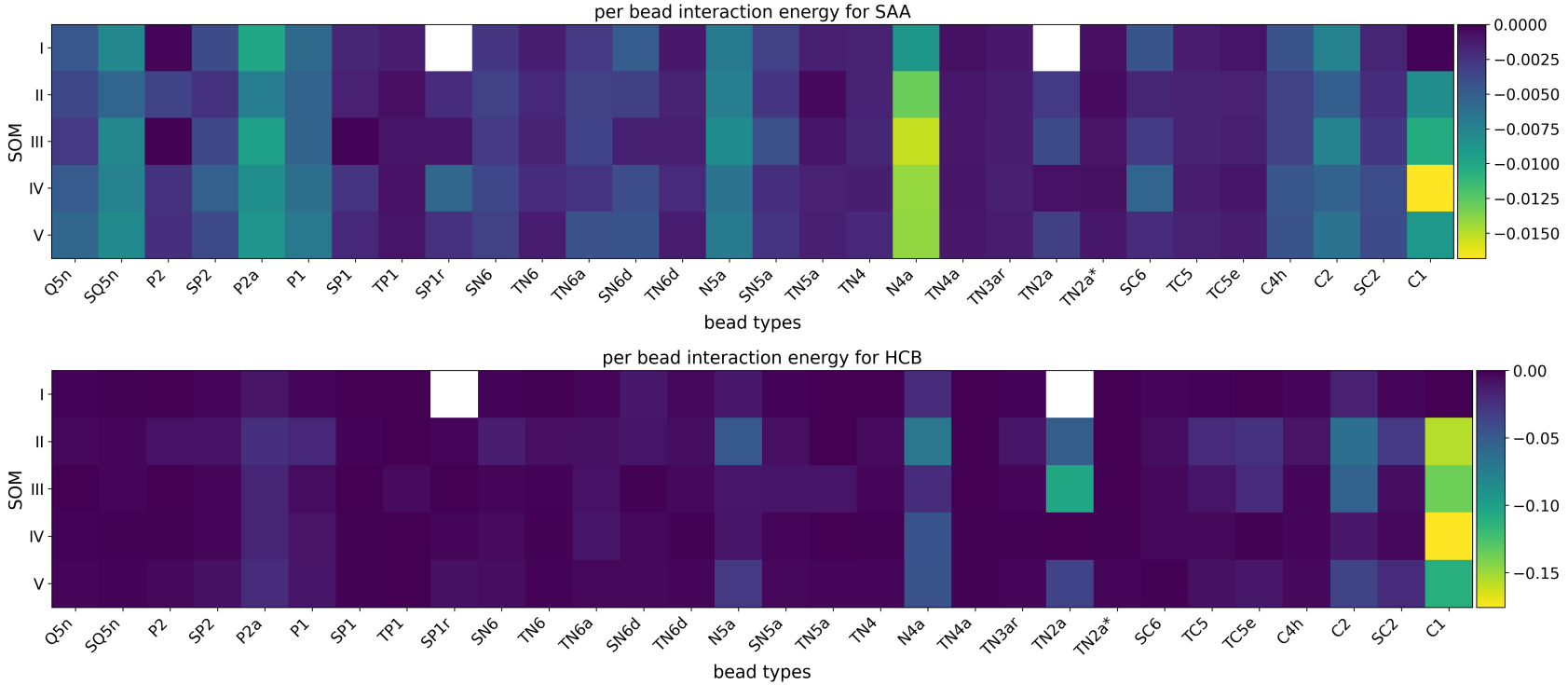


Figure S8: Interaction energies between the pollutant molecules and water (W) or HS molecules, averaged over the 20 trajectories for each pollutant and each SOM type. The values of SAA (a) are further distinguished in proportions from the sulfonamide group (blue), aromatic ring (gray) and the amino group (orange). HCB values (b) are depicted in green.

S6 Per bead interaction energies



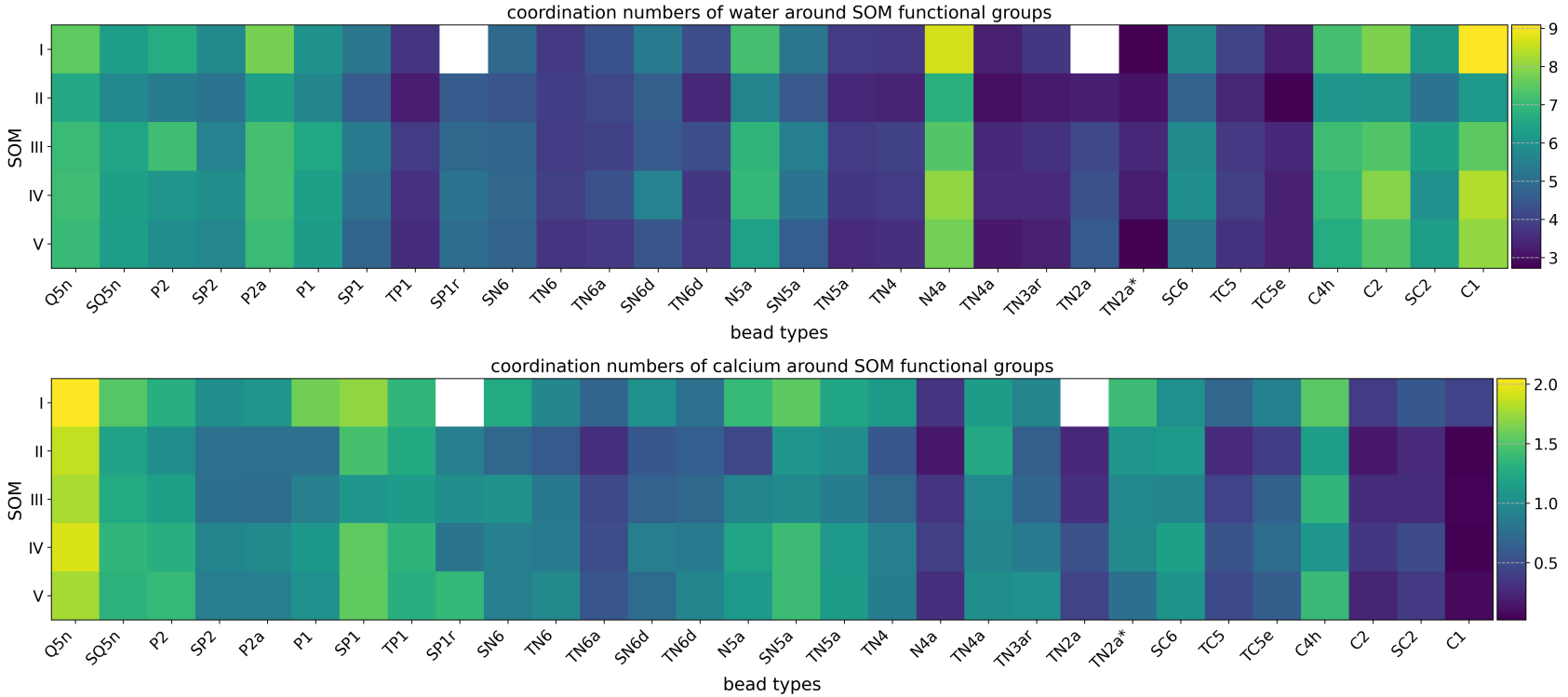


Figure S9: Per bead energies between the pollutants and SOM beads with a certain type (x-axis), for the different SOM systems (y-axis). Each energy value was normalized on the number of a respective bead type in the corresponding SOM system. Furthermore, coordination numbers of water and calcium around SOM functional groups are depicted, respectively. The results were averaged over the 20 individual runs for each pollutant and each SOM type. The translation of the bead types into the respective SOM functional groups is shown in Table S1.

The absolute values of the interaction energies between the pollutants and SOM functional groups depend on the number of the respective functional groups in the system. Normalizing by using the number of functional groups reveals the contribution of each functional group to the pollutant-HS interaction. Figure S9 shows the normalized interaction energies and coordination numbers for the different SOM systems. For SAA, certain functional groups such as N4a (methyl acetate) and C1 (alkane) exhibit the highest per bead interaction energies. Moreover, SQ5n (carboxylate), P2a (acid anhydride) and N5a (aliphatic ketone) show higher values. As expected, the interaction energies are generally more similar for a certain functional group after the normalization, when different SOM types are compared. The per bead energy for HCB is highly concentrated in the C1 functional group, with partial contributions from other groups, such as N4a again. The prominent behavior of N4a and C1 can be explained less by the direct interaction of the pollutant with these groups, but rather by information about their positioning within the SOM matrix, which is provided by the coordination numbers of water and calcium ions around SOM functional groups. The high coordination numbers of water around the functional groups N4a and C1 indicate that these have a higher contact with the water phase, providing more interaction sites for the diffusing pollutant molecules. Conversely, the low values for the calcium case suggest that these groups are not masked by the structural arrangement of the SOM matrix, formed by the calcium-complexes. This in turn explains the high per bead energies visible for SAA and HCB for these functional groups. In general, these are effects caused by the fragment design and the structural arrangement of the HS molecules. However, this provides an insight into which parameters could also play a role in the pollutant-SOM interaction in terms of the available functional groups. If functional groups are freely available in the water phase and are not masked by the formation of complexes with the anions in the system, the probability of interaction with the pollutants increases. The normalization of the energies by the number of functional groups leads to the conclusion that they are not directly comparable to monomer-calculations[20,21] because the contributions of the available functional

groups, are influenced by the structural arrangement of the SOM molecules, which affects the results. On the absolute scale, these effects do not significantly contribute to the energies, but only become relevant after normalizing. Furthermore, the per bead values of the relevant functional groups, such as N4a or C1, could change if their concentration within the system was higher, possibly leading to a different structuring of the HS molecules and altering the availability of the groups for the water phase. Future investigations could focus on the connection between the composition of SOM and the functional group availability for pollutant-SOM interactions.

S7 Determination of the trapping times

S7.1 Quantification via displacements and a hidden Markov model

In order to gain a deeper understanding of the pollutant dynamics, the time intervals during which the pollutant molecules were fixed at a specific position in the SOM matrix were identified. Since a visual inspection is not suitable for determining such time intervals due to the large number of trajectories (200 in total), the times at which the pollutants were trapped were determined using a hidden Markov model, implemented in a Python package (https://github.com/hmmlearn/hmmlearn). Such models have been used before to investigate confinement and distinct diffusion states of particles[22–24]. The advantage of such models is the almost parameter-free and unbiased determination of discrete states, independent of predefined thresholds.

For the model training, the absolute value of the spatial displacement of the center of geometry of the pollutant molecules between different lag times Δt was used:

$$\Delta r(t) = |\vec{r}_{\mathrm{COG}}(t + \Delta t) - \vec{r}_{\mathrm{COG}}(t)| \tag{4}$$

Two hidden states were employed in the model definition, which are interpreted as trapped and diffusing states of the pollutant molecules. Based on a convergence analysis of the

trapping probabilities for different lag times, a lag time 15 ns, i.e., 150 frames, was set. For the SOM II system, the values of the displacements along the trajectory are shown in Figure S10, together with the trapping states identified by the Markov model in red. These intervals indicate a fixed pollutant molecule. The trapping probabilities, which are shown in Figure 7 in the main manuscript, were calculated from the ratio of the number of time steps where trapping was identified to the total number of time steps, and by averaging over the 20 production runs for each pollutant and each SOM type.

Figure S11 shows the time-resolved coordination numbers of HS beads (black) or water beads (purple) surrounding each individual pollutant molecule. The typical confinement event is characterized by an increase in HS contacts and a simultaneous decrease in water contacts, consistent with trapping caused primarily by HS-pollutant interactions. Visual inspection of the trajectories revealed distinct trapping motifs of the pollutant molecules, which can roughly be divided into three categories: 1. Residence within an SOM cavity, characterized by high HS and low water coordination numbers. 2. Encapsulation in a water-filled cavity, a rare event, characterized by low HS and high water coordination numbers. 3. Trapping in a water-filled cavity with simultaneous adsorption to HS, leading to intermediate coordination levels for both species. More detailed analyses are needed to determine the molecular mechanisms and functional groups that favor each type of trapping, and to evaluate the influence of water, in particular, how its presentation at different coarse-grained resolutions (regular, small or tiny beads) modulates these processes.

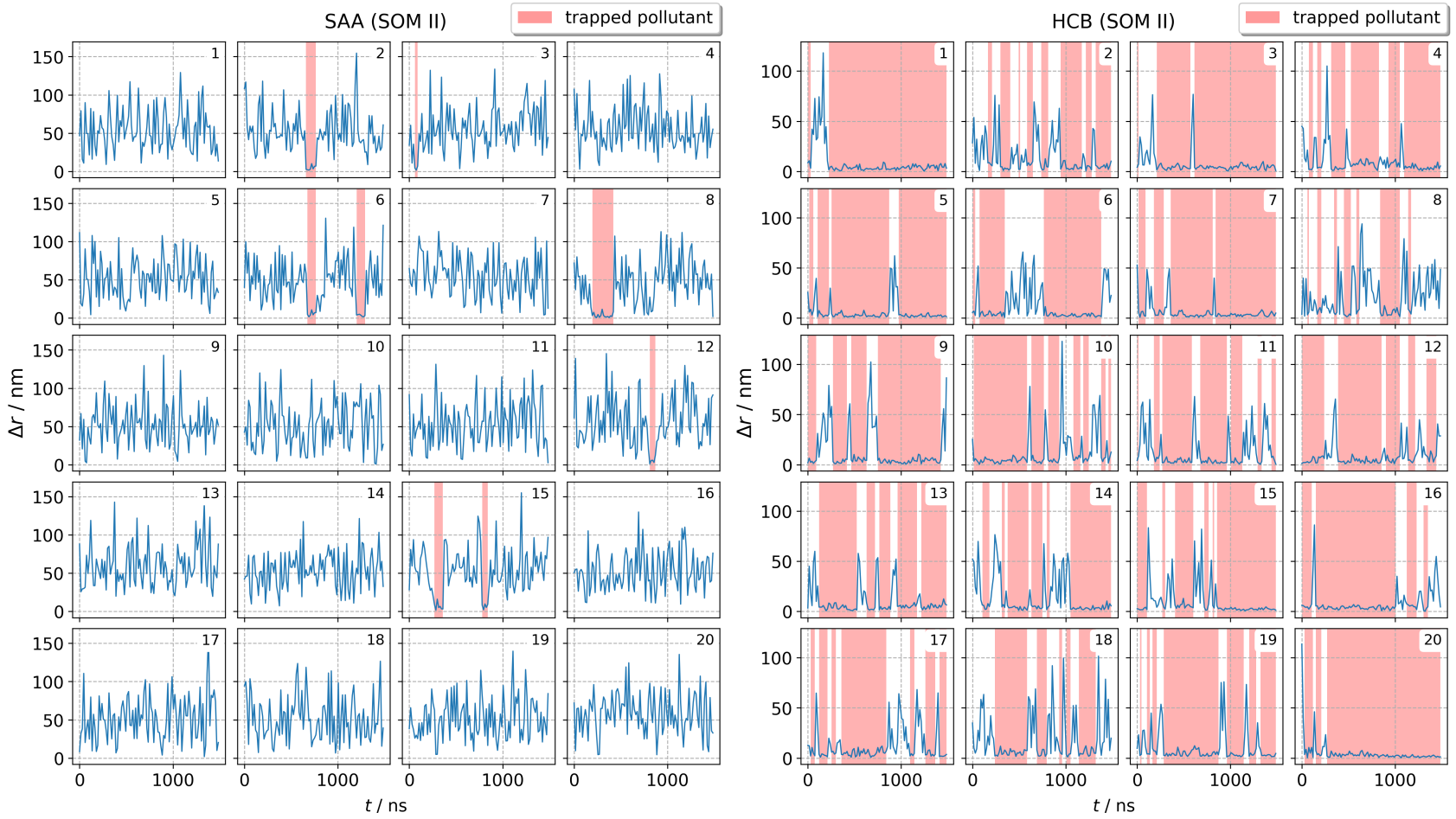


Figure S10: Values of the spatial displacements Δr (blue) defined in equation 4 for different time steps t. As an example, the results of the SOM II system are shown. Times, marked in red, indicate a trapped pollutant molecule.

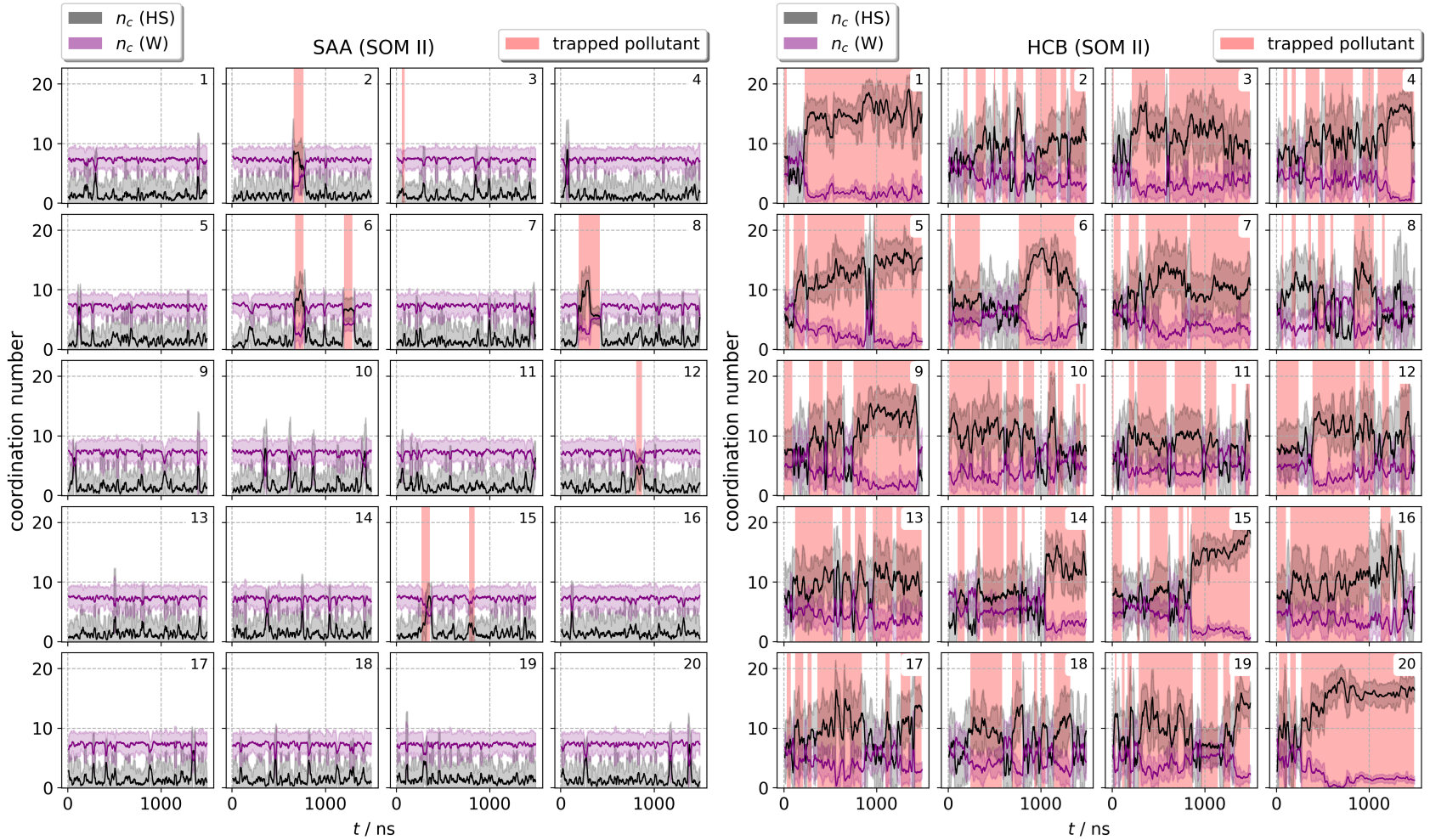


Figure S11: Per-frame coordination numbers of HS (black) and water (purple). The lines and shaded areas respectively represent the rolling average and the corresponding standard deviation. Times, marked in red, indicate a trapped pollutant molecule. These are equivalent to the ones shown in Figure S10.

S7.2 Quantification via a position autocorrelation function

An alternative approach to distinguish between different trapping states and to calculate the trapping probabilities was implemented using a normalized position autocorrelation function:

$$C(t;n) = \frac{1}{n}\sum_{k=1}^{n}\frac{\vec{q}(t)\cdot\vec{q}(t+k)}{|\vec{q}(t)|\,|\vec{q}(t+k)|} \tag{5}$$

Each value of the correlation function C at a reference time step t was calculated using the next n coordinate vectors of the unwrapped trajectory. To account for the arbitrary lengths of the absolute position vectors of a pollutant molecule during the trajectory, which influences the response of the correlation function to displacements of the molecule, shifted coordinate vectors $\vec{q}(t)$ were used, defined as the difference between the center of geometry of the pollutant $\vec{r}_{\mathrm{COG}}(t)$ and a shift vector $\vec{t}$: $\vec{q}(t) = \vec{r}_{\mathrm{COG}}(t) - \vec{t}$. Here, $\vec{t}$ is designed for each reference time step t considered, such that the length $|\vec{q}(t)|$ of the shifted position vector is scaled to a fixed distance (3 nm). This corresponds to a shift in the coordinate center for the start of each analysis of a new set of position vectors and ensures a standardized sensitivity of the correlation function to displacements of the pollutant molecules, independent of their absolute position in the system. If the pollutant was trapped for n time steps, the correlation is retained (close to 1) which indicates a trapping. In contrast, if it underwent a diffusion, the correlation is not retained, leading to values of the correlation function being lower than 1. Based on a comparison between the values of the autocorrelation functions and the pollutant trajectories, and on a comparison to the values from the hidden Markov model, the threshold value C_{thres} for the correlation function, which defines the trapping of a pollutant, was set to 0.98 for SAA and 0.97 for HCB. This means that if $C \geq C_{thres}$, the molecule is considered to be fixed. In order to minimize the impact of considering the subsequent n time steps for a reference time t, which mainly affects the values of the correlation function at the beginning and end of a trapping interval, a relatively small value for n was chosen ($n = 10$). In addition, a trapping time interval was only considered if the correlation function fulfilled

the trapping criterion ($C \geq C_{thres}$) for at least 150 consecutive frames, similar to the lag time chosen for the hidden Markov model analysis.

The resulting trapping probabilities for the autocorrelation method are depicted in Figure S12. Apart from minor discrepancies, the probabilities closely match those obtained

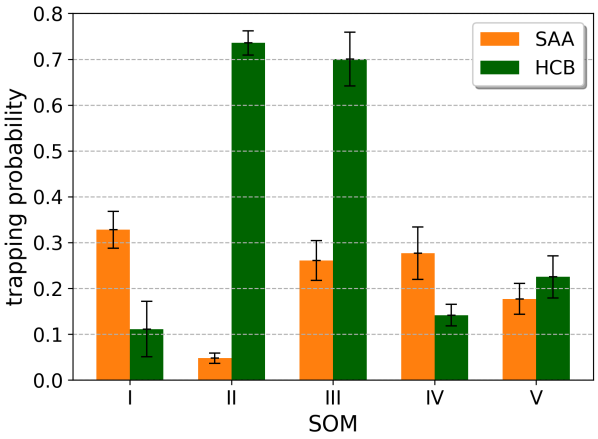


Figure S12: Trapping probabilities of SAA (orange) and HCB (green) for the different SOM types, determined using the position autocorrelation function.

from the hidden Markov model. While the hidden Markov model requires fewer number of parameters to identify the trapping states, the autocorrelation-based approach remains a readily interpretable alternative that also provides an independent validation for the results.

S8 $\pi - \pi$ stacking configurations

Table S3: Geometric definitions of the $\pi - \pi$ stacking configurations. For the two configurations, edge-to-face and face-to-face, criteria are defined for the distance between the geometric centers of the rings $|\vec{r}|$, the angle between the normal vectors of the rings $\sphericalangle(\vec{n}, \vec{n}')$, and for the angle between the centroid vector and one normal vector $\sphericalangle(\vec{r}, \vec{n})$. Please note that the angle ranges are designed to additionally account for the opposite directions of the normal vectors. The vectors and angles can be obtained from the visual representation of the stacking configurations in Figure S13.

	$\lvert\vec{r}\rvert$	$\sphericalangle(\vec{n}, \vec{n}')$	$\sphericalangle(\vec{r}, \vec{n})$
edge-to-face	$0.45 - 0.60\,\mathrm{nm}$	$75 - 105°$	$0 - 20°$, $160 - 180°$
face-to-face	$0.33 - 0.45\,\mathrm{nm}$	$0 - 20°$, $160 - 180°$	$0 - 20°$, $160 - 180°$

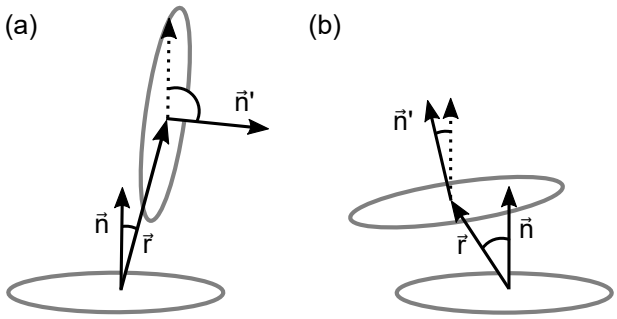


Figure S13: Visual representation of edge-to-face (a) and face-to-face (b) stacking configurations. A dotted vector represents the direction of the normal vector $\vec{n}$ of the corresponding lower ring.

To yield an in-depth insight into the interactions of the pollutants with aromatic beads of the HSs and to determine binding motifs induced by the local SOM environment, $\pi - \pi$ stacking configurations were calculated. Using MDAnalysis, the configurations between a pollutant and the rings of HS molecules were determined, utilizing the criteria presented in Table S3. Figure S15 shows the number of stacking configurations that occur in each frame of a trajectory. The results of all the trajectories of the SOM II system are shown, and a distinction is made between the two configurations. Multiple stacking configurations between pollutant and HS molecules can occur in one trajectory frame, even for face-to-face configurations. In the latter case, it is possible for two stacking configurations to occur simultaneously, if either two HS rings are interacting with the pollutant ring from both sides, or if both rings of a naphthalene-like fragment, such as HS29, are interacting with the pollutant and fulfilling the stacking criteria. For an overview of the fragments, please refer to the Supporting Information for Dettmann et al. [5]. In a relatively rare case of three face-to-face configurations, the interactions of a naphthalene-like fragment on the one side and of another HS ring from the other side take place (see e.g., HCB for SOM II with index 2).

Using the coarse-grained model, the $\pi - \pi$ stacking interactions are determined by dispersive forces. This leads to the fact that nuanced attractive and repulsive forces originating from displacements of two aromatic rings are not resolved, which can be explained, for exam-

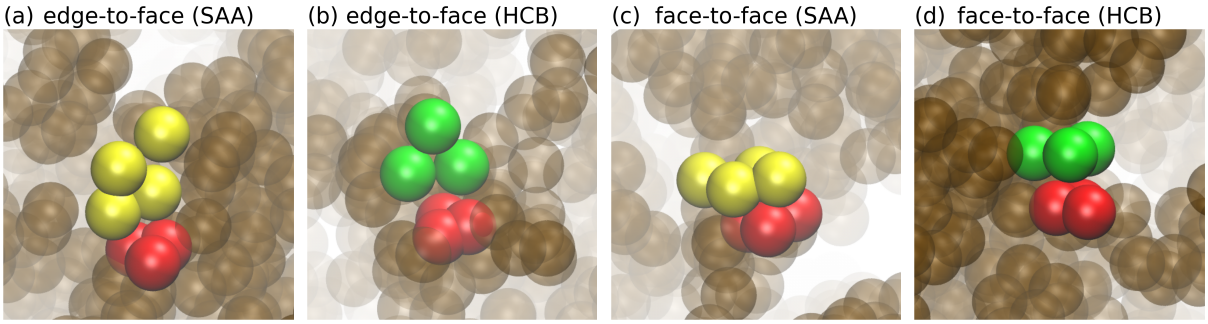


Figure S14: Snapshots from SOM III trajectories with $\pi - \pi$ stacking configurations. Edge-to-face configurations ((a) and (b)), and face-to-face configurations ((c) and (d)) are depicted for each pollutant. SAA (yellow), HCB (green) and HS molecules (transparent brown) are shown without the water beads. The ring of the HS molecule with which the respective pollutant molecule forms a configuration is shown in red.

ple, by the quadrupoles of the real molecules. However, the overall adsorption behavior of the pollutants on HS molecules is captured, which is of particular interest in this investigation.

The total count of $\pi - \pi$ stacking configurations, differentiated into the edge-to-face and face-to-face configurations, is shown in Figure S16. Relatively low counts are observed for SAA, with the highest values observed for the SOM III case. This is an expected result, given the high content of aromatic rings for this SOM type. For every case, the edge-to-face configurations occur more often than the face-to-face configurations. For certain cases, such as SOM II, III, and V, HCB exhibits a much stronger preference for forming $\pi - \pi$ stacking configurations with the HS molecules. It can be assumed that this is associated with a higher probability of finding HCB in the HS phase. Relatively low values are observed for the SOM IV case, where a low trapping probability was also detected. Especially in the case of HCB, a correlation between the trapping behavior (see Figure 7 in the main manuscript) and the formation of stacking configurations can be identified.

A further analysis revealed that the occurrence of stacking configurations was linked with the trapping times of the pollutant molecules. For each stacking type, the number of configurations that occurred during a trapping time interval was counted and summed up for all 20 cases, considering one pollutant and one SOM type. With regard to Figure S15, this would correspond to adding up the number of stacking configurations occurring in the red

marked areas. This sum is denoted as k_{trap}. Similarly, this was done for the configurations that occurred during times when the molecule was not trapped, which would correspond to summing up the number of stacking configurations outside the red marked areas, thus yielding k_{free}. The quantities P_{trap} and P_{free}, which represent the stacking probabilities of a trapped or free molecule, were calculated as follows:

$$P_{trap} = \frac{k_{trap}}{N_{trap}} \qquad P_{free} = \frac{k_{free}}{N_{free}} \tag{6}$$

Here, N_{trap} and N_{free} represent the total number of frames of the 20 trajectories, if a pollutant is trapped or not trapped, respectively. For each case, these two types of probabilities are plotted in Figure S17. Comparing the two configuration types for each case, the edge-to-face probabilities are generally higher, which can be explained by the fact that there are more possibilities of forming edge-to-face configurations than face-to-face configurations. For both pollutants, a notable exception is the SOM III system for the trapped case, where the edge-to-face and face-to-face probabilities have roughly the same value. The high aromatic content of this system could favor the formation of face-to-face configurations. Please note that the exact probabilities may depend on the geometric criteria, defined in Table S3. However, the feature that stands out, is that stacking is more likely to occur together with the trapping of the pollutant molecule. Although in this approach, the interactions were modeled using only dispersive forces, it is apparent that stacking is an important component in the pollutant-HS interaction and may be even more relevant when the calculations are performed using a higher level of theory. This would also be consistent with investigations, highlighting the importance of $\pi - \pi$ interactions occurring between pollutants and SOM monomer components[21]. The findings of the present study have the potential to serve as a starting point for further investigations that provide insight into the correlation between the adsorption configurations of pollutants in the SOM matrix and their trapping.

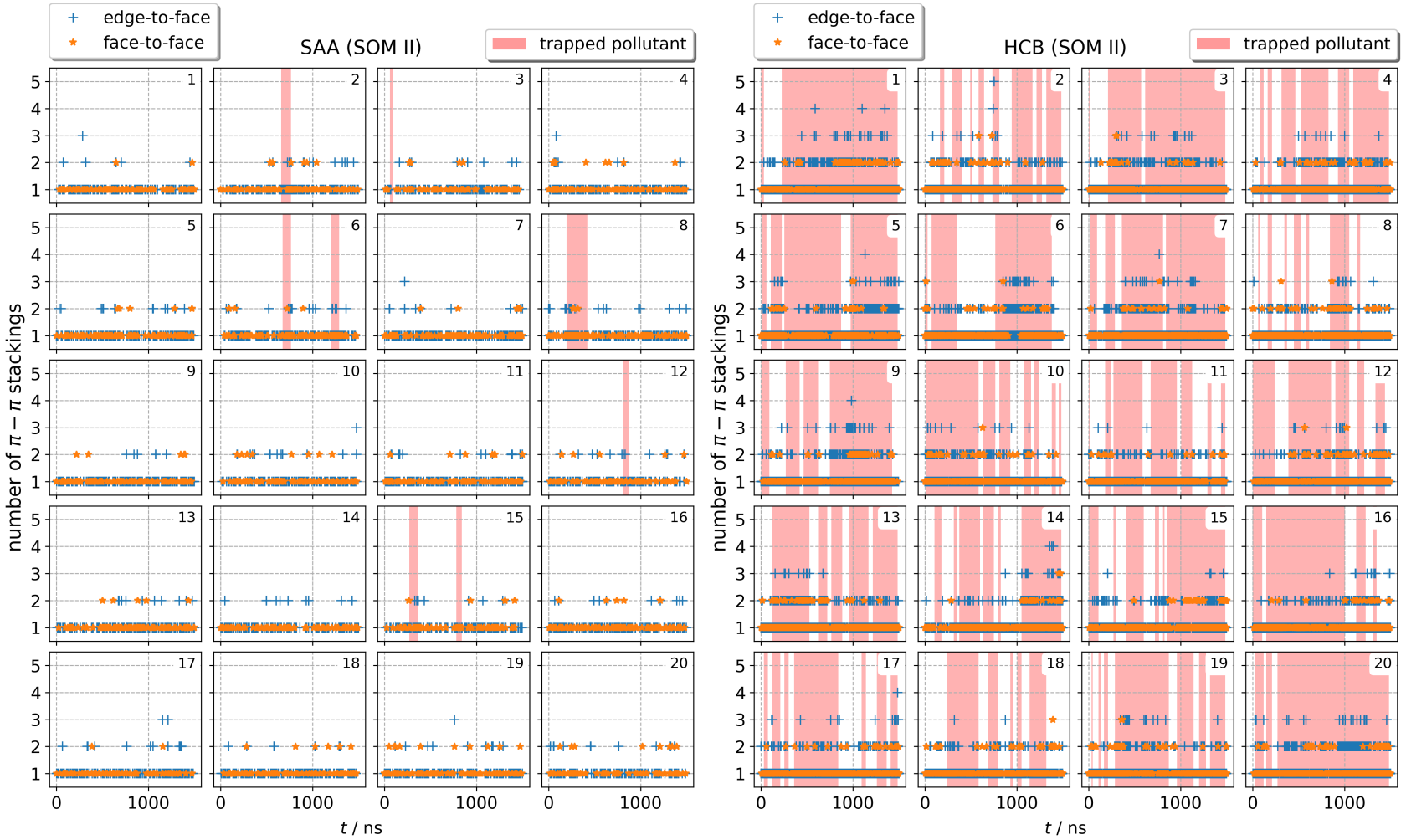


Figure S15: Number of $\pi - \pi$ stacking configurations for each frame of the trajectories. As an example, the results of the SOM II system are shown. The red marked times correspond to the respective pollutant molecule being trapped.

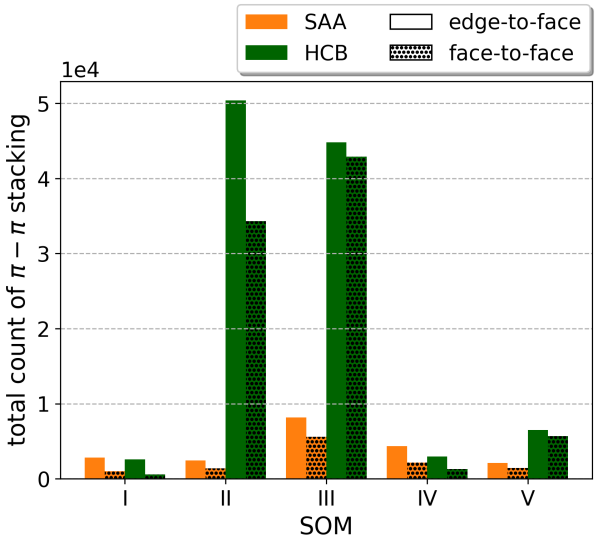


Figure S16: Total count of $\pi-\pi$ stacking for SAA (orange) and HCB (green) for the different SOM types. A distinction is also made between edge-to-face (blank color) and face-to-face (dotted pattern) configurations.

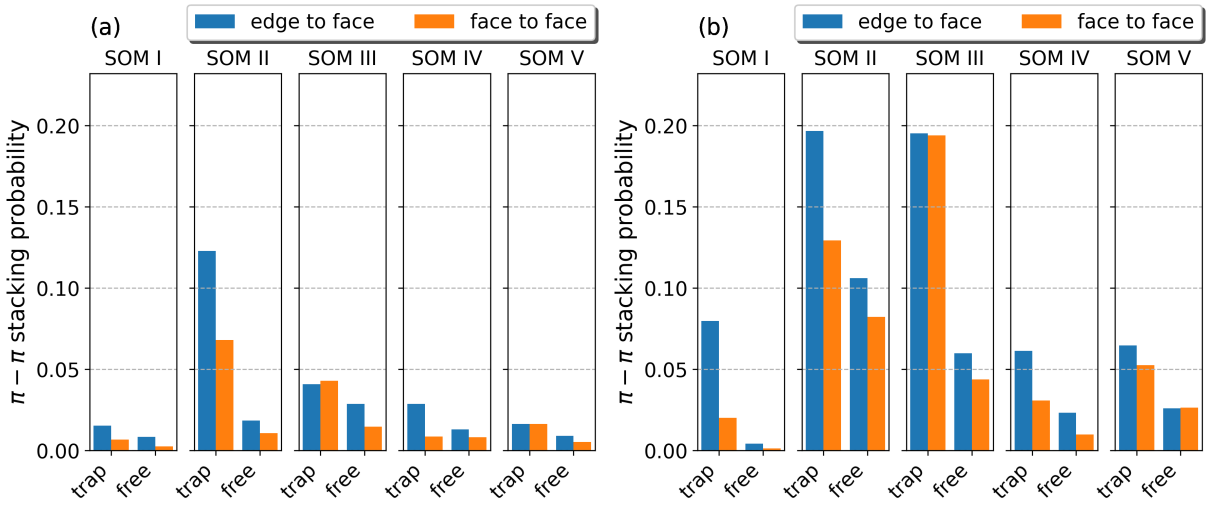


Figure S17: Probabilities of observing $\pi-\pi$ stacking for SAA (a) and HCB (b), differentiated into edge-to-face (blue) and face-to-face (orange) configurations. The probabilities are distinguished for time intervals with a trapped (trap) and free pollutant.

S9 Higher pollutant concentration

Simulations with multiple pollutants were conducted in the SOM III system to explore cooperative interactions at higher concentrations. Systems containing 20 HCB or 20 SAA molecules were set up with initially uniform distributions. Figure S18 shows snapshots of the SOM III system with SAA (a) and HCB (b) after a 5 µs production run. SAA molecules remained well-dispersed, exhibiting no aggregation. In contrast, nearly all HCB molecules formed a hydrophobic cluster over time, despite their initial separation. This behavior is consistent with the strong hydrophobicity of HCB and highlights that self-aggregation can influence accessibility to the SOM matrix. Future studies varying pollutant concentrations and water content may provide deeper insight into these cooperative effects on pollutant–SOM interactions. However, it should be noted that the concentrations used here generally exceed the typical environmental pollutant levels in soils, and that the single-molecule simulations more realistically reflect the limited occurrence of cooperative interactions under such conditions.

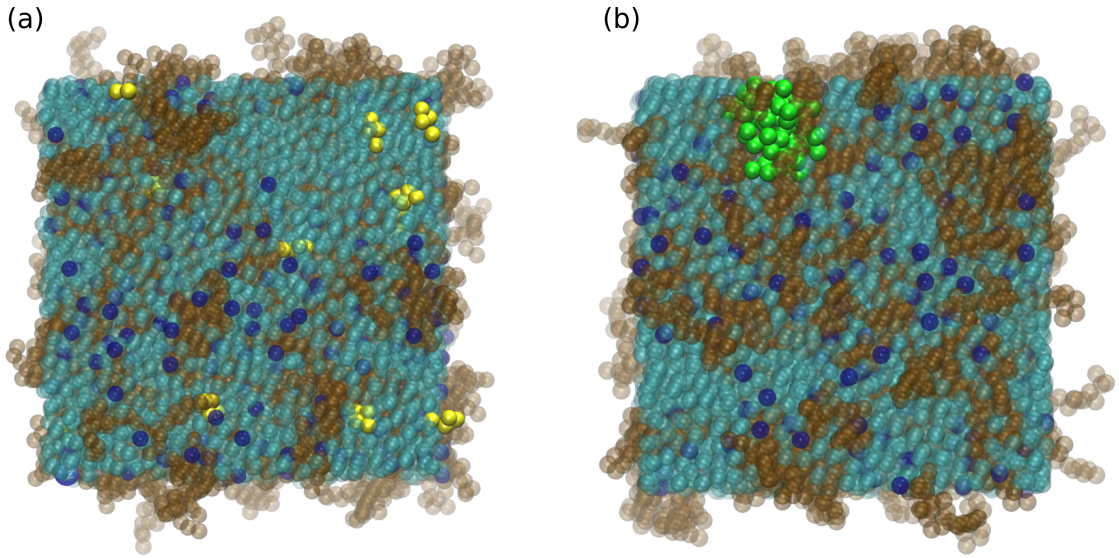


Figure S18: Snapshots from the SOM III system with multiple pollutants, containing either 20 SAA molecules (a), represented in yellow, or 20 HCB molecules (b), represented in green. HS molecules are shown in brown, water in cyan, and calcium ions in blue.

References

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

(2) 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.

(3) 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.

(4) 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.

(5) 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.

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

(7) 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.

(8) 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.

(9) 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.

(10) 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.

(11) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR: Hydrophobic, electronic, and steric constants*; ACS Professional Reference Books; American Chemical Society: Washington, D.C., 1995; Vol. 2; p 22.

(12) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR: Hydrophobic, electronic, and steric constants*; ACS Professional Reference Books; American Chemical Society: Washington, D.C., 1995; Vol. 2; p 15.

(13) Dettmann, L. F.; Kühn, O.; Ahmed, A. A. Automated Parametrization Approach for Coarse-Graining Soil Organic Matter Molecules. *Journal of Chemical Theory and Computation* **2024**, *20*, 10684–10696.

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

(15) 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.

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

(17) 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.

(18) 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.

(19) 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.

(20) 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.

(21) 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.

(22) Das, R.; Cairo, C. W.; Coombs, D. A Hidden Markov Model for Single Particle Tracks Quantifies Dynamic Interactions between LFA-1 and the Actin Cytoskeleton. *PLoS Computational Biology* **2009**, *5*, e1000556.

(23) Slator, P. J.; Cairo, C. W.; Burroughs, N. J. Detection of Diffusion Heterogeneity in Single Particle Tracking Trajectories Using a Hidden Markov Model with Measurement Noise Propagation. *PLOS ONE* **2015**, *10*, e0140759.

(24) Slator, P. J.; Burroughs, N. J. A Hidden Markov Model for Detecting Confinement in Single-Particle Tracking Trajectories. *Biophysical Journal* **2018**, *115*, 1741–1754.

THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Coarse-Grained Simulations of the Nanoplastic Interaction with Soil Organic Matter

Published as part of The Journal of Physical Chemistry B special issue "Physical Chemistry of Microplastics and Nanoplastics".

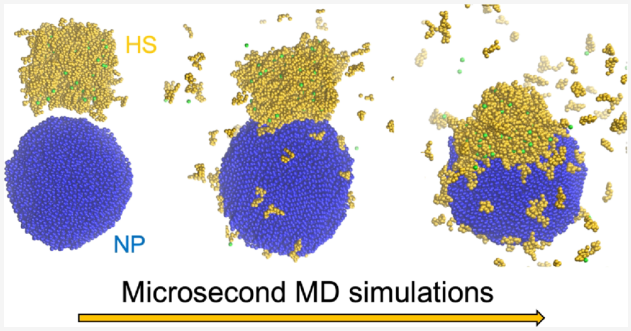
Lorenz F. Dettmann, Oliver Kühn, and Ashour A. Ahmed*

Cite This: J. Phys. Chem. B 2025, 129, 11605−11615

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Nanoplastics (NPs) in terrestrial ecosystems represent an emerging ecological threat, yet their molecular-level interactions with soil organic matter (SOM) remain difficult to probe experimentally. Molecular simulations can bridge this gap if they capture relevant time and length scales. Here, microsecond coarse-grained molecular dynamics simulations were employed with the Martini 3 force field using humic substances (HSs) as SOM representatives. We pursued two complementary modeling perspectives: (i) comparing the behavior of poly(ethylene oxide) (PEO) and polyethylene (PE) chains within SOM matrices and (ii) examining SOM association with a PE nanoparticle across varying pH values. PEO exhibited strong hydration and high mobility, whereas PE interacted weakly with acidic moieties but bound strongly to hydrophobic groups under low-acidity conditions, leading to reduced diffusivity. In the PE nanoparticle simulations, a low pH promoted dispersion of HS aggregates and enhanced surface coverage of the plastic particle. Together, these results provide molecular-scale insights into how the polymer polarity, SOM composition, and environmental pH govern NP−SOM interactions.



Downloaded via UNIV ROSTOCK on November 8, 2025 at 16:00:00 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

INTRODUCTION

Plastics are man-made polymeric structures with long-chain molecular architecture, possessing unique properties, such as low density, chemical resistance, and moldability into almost any form.[1] Through modification with functional groups, incorporation of tailored additives, and copolymer design, plastics can be adjusted to achieve improved flexibility and persistence, broadening their use from single-use packaging to aerospace components or medical devices.[2] Their versatility places plastic into an essential role for modern sustainability strategies, including the construction of lightweight vehicles or components for renewable-energy technologies.[3] Plastic production has grown at an accelerating rate since the 1950s.[4] Approximately half of all plastics were manufactured within the past two decades,[5] and it is anticipated that this amount will increase two to threefold by 2050.[6]

The durability that makes plastics valuable has caused a worldwide problem of persistent waste.[7] Polymer debris has been detected in various regions of the world,[8−10] including the Arctic[11] and remote, pristine mountain catchments,[12] which is connected to their versatile transportation paths in air, water, or sediments.[13] Once in the environment, ultraviolet radiation, thermal or mechanical stress, as well as microbial processes can generate smaller plastic pieces,[14] termed microplastics (5 mm − 1 μm), and, with continued degradation, nanoplastics (NPs, 100−1 nm), which are particles that are significantly more mobile and bioavailable.[15] Terrestrial ecosystems are estimated to hold 4−23 times more microplastics than the oceans,[16] and agricultural soils are emerging as a large sink and source of NP pollution.[17]

Soil is a complex, heterogeneous matrix consisting of mineral particles, various organic residues, pore-filling water, and gases, creating an environment that encompasses a wide range of redox conditions, pH values, and adsorption-specific surfaces.[18,19] In order to understand how man-made pollutants and newly emerging particles interact within a multicomponent system, scientists must systematically unravel the physicochemical processes occurring in soils. The organic fraction, termed soil organic matter (SOM), typically makes up only 1−5% of the soil but plays a significant role in the sorption of chemical compounds due to its large surface area and diverse

Received: August 30, 2025
Revised: October 15, 2025
Accepted: October 17, 2025
Published: October 23, 2025



ACS Publications

© 2025 The Authors. Published by American Chemical Society

functional groups.[20] Since SOM itself spans a continuum from plant and animal residues from various sources in different stages of degradation, operational isolates known as humic substances (HSs) have long been used to obtain model fractions for systematic studies.[21,22] The International Humic Substances Society (IHSS) provides rigorously extracted and globally selected reference materials that enable cross-laboratory comparability and thus facilitate fundamental insights into the structure and reactivity of SOM.[23]

Despite recent progress, quantitative detection of NPs in soils remains extremely challenging. Conventional spectroscopic and thermoanalytical measurement techniques rarely provide the sensitivity or size resolution required to analyze polymer particles in the nanometer range within the spectrally noisy soil background.[24,25] The analytical challenges are amplified by NPs' low mass and concentrations, as well as intrinsic heterogeneity and tendency to agglomerate with soil components. The utilization of thermal-desorption proton-transfer-reaction mass spectrometry[26] or the combination of pyrolysis with spectroscopic techniques, such as gas chromatography-mass spectrometry,[27] are among the most recent methodological advancements. Nevertheless, estimates of NP concentrations, transport rates, and heteroaggregation in soils remain largely uncertain, highlighting the need for further research.

Given these analytical limitations, molecular dynamics (MD) simulations have recently proven to be a powerful tool for elucidating NP interactions with organic materials,[28−34] revealing mechanisms at the molecular level that cannot be directly observed in experiments. For example, Chen et al.[33] investigated the aggregation of polystyrene NPs and humic acids in the presence of metal ions and bisphenol A via experiments and MD simulations, revealing competitive interaction schemes and the role of NP functionalization. Li et al.[28] demonstrated that NP toxicity is dictated by the extent and configurations of adsorbed humic acid molecules. In such studies, coarse-graining is commonly employed to extend the accessible time and length scales of simulations. This approach reduces computational cost by grouping atoms into larger interaction sites, thereby decreasing the number of pairwise interactions and enhancing performance.[35,36] However, no study has so far investigated the interaction between NPs and SOM of varying composition at the microsecond time scale, accessible through coarse-grained simulations, nor examined the role of SOM functional groups and the pH-dependent behavior, which represents a critical knowledge gap.

Therefore, NP−SOM interactions were investigated in this study from two complementary modeling perspectives. First, short polymer chains were embedded in HS matrices to probe their partitioning, dominant interaction sites, and mobility. Polyethylene (PE), the most widely produced plastic, and its polar analogue, poly(ethylene oxide) (PEO), were chosen to represent hydrophobic and hydrophilic behaviors, respectively. Second, a PE nanoparticle was modeled to examine its association with SOM molecules, explicitly incorporating pH-dependent protonation states. Together, these simulations provide molecular-scale insights into how polymer polarity, SOM composition, and environmental acidity govern NP−SOM interactions.

MATERIALS AND METHODS

Modeling Single Polymer−SOM Interactions. In this study, the coarse-grained Martini force field[37,38] was utilized to describe the interactions between polymers and SOM.[39,40] Two model polymer chains, PE and PEO, generated with polyply[41] were investigated to represent a purely hydrophobic and a polar polymer, respectively. The coarse-grained structures exhibit distinct mapping schemes, 4:1 for PE and 3:1 for PEO, so that one coarse-grained bead replaces four or three heavy atoms (Figure 1). As a result, the equilibrium

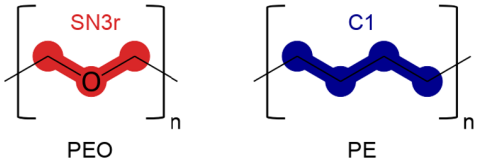


Figure 1. Atomistic structures and applied mappings to the coarse-grained representation for PEO (3:1 mapping, red) and PE (4:1 mapping, blue). The labels refer to the Martini bead types used for the monomers of the corresponding polymer.

coarse-grained bond lengths of the two polymers are not identical. To ensure comparable spatial extents before the study of confinements in the SOM structures, the radius of gyration (R_g) for chains of different numbers of monomers was determined by explicit simulations in bulk water. Because the hydrophobic PE tends to collapse in water while PEO remains more extended, the chain length alone was not used as a matching criterion. Systematic R_g analysis revealed that polymer chains containing seven coarse-grained beads in PE and PEO provided nearly identical R_g values of around 0.53 nm under aqueous conditions and were therefore used for all simulations.

Real plastic chains are typically much longer than the ones investigated here. However, such long chains would make extensive sampling of the heterogeneous interaction sites in the SOM matrix computationally prohibitive. Therefore, these shorter chains were chosen as a tractable model that preserves the local chemical characteristics while allowing for efficient sampling of polymer−SOM interactions.

Two structurally distinct SOM systems were generated with the Vienna Soil Organic Matter Modeler 2 (VSOMM2)[42] and subsequently converted to a coarse-grained representation by the SOM2CG tool.[43] This tool converts atomistic VSOMM2 systems into coarse-grained representations compatible with Martini 3. The underlying models were developed following the standard Martini mapping and validated against various reference properties, including solvent-accessible surface areas, partition data, and densities of atomistic SOM systems.[39] Previous simulations further demonstrated the capability of this approach to efficiently investigate the structural and compositional diversity of SOM samples from the IHSS across different water contents.[40]

The VSOMM2 builds SOM assemblies from basic molecular fragments using user-defined elemental and organic fractions that emulate HS samples from the IHSS. This approach enables the investigation of diverse chemical compositions. Guided by a principal-component analysis conducted by Escalona et al.,[44] which identified the relative abundance of carboxylate and carbonyl functionalities of IHSS samples, one HS model rich in acidic functional groups and one depleted in these moieties were selected. Each SOM system consisted of 320 HS molecules with five fragments per molecule. Further details on the composition of the chosen samples and on the simulation setups are provided in Tables 1 and 2, respectively.

https://doi.org/10.1021/acs.jpcb.5c06072
J. Phys. Chem. B 2025, 129, 11605−11615

Table 1. IHSS Samples Modeled and Analyzed in This Work Are Rich and Depleted in Acidic Functional Groups, Respectively[a]

			elemental and organic fractions							
identifier	HS sample	characteristics	C	N	carbonyl	carbox.	arom.	acetal	heteroaliph.	aliph.
SOM I	Elliot Soil I (FA)	high carboxyl and carbonyl content	0.5057	0.0272	0.1212	0.2525	0.3031	0.0101	0.0909	0.2222
SOM II	Elliot Soil IV (HA)	low carboxyl and carbonyl content	0.5951	0.0390	0.0100	0.1100	0.4100	0.0600	0.1400	0.2700

[a]The corresponding compositions were used as input parameters for the VSOMM2 model generation. Depicted are the relative fractions of carbon (C), nitrogen (N), carbonyl, carboxylic (carbox.), aromatic (arom.), acetal, heteroaliphatic (heteroaliph.), and aliphatic (aliph.) groups.

Table 2. Box Length (a) and Number of HS Beads (N_{HS}) for the Considered SOM Systems

	a/nm	N_{HS}
SOM I	9.25	5744
SOM II	9.34	5139

The models were simulated at pH = 5, realized by adapting the ratio of carboxylate to carboxylic acid groups to this pH level, consistent with the pH modeling of the VSOMM2. It should be noted that constant pH simulations in Martini can also be achieved by a different approach, utilizing titratable beads.[45]

Charge neutrality was maintained by adding calcium counterions. While soils contain a diverse spectrum of inorganic cations, calcium is typically one of the most abundant divalent species and is well-known to contribute to the structural stability of SOM.[46] In addition, its use facilitates comparability with previous computational studies, where calcium was commonly employed as a counterion.[44,47]

To probe single polymer interactions with SOM and avoid cooperative polymer−polymer effects, each system contained one PE or PEO chain inserted into a preequilibrated SOM matrix. Polymers were initially placed ≥0.75 nm from any HS bead to ensure unrestricted diffusion before first contact with the HS phase (Figure 2), thereby favoring interaction from the aqueous phase. The hydrated SOM boxes contained water at a heavy-atom fraction of 0.567, sufficient to form continuous water domains. Although this setup represents a dilute regime compared to environmental conditions, it enables quantification of partitioning, binding preferences, and diffusivity at the molecular level. Ensemble statistics were derived from 20 independent replicas per polymer−SOM combination.

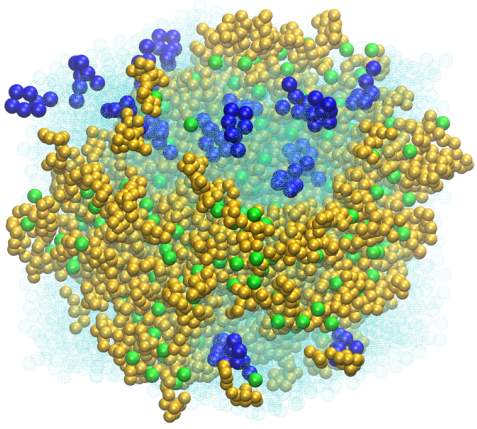


Figure 2. All initial positions of the individual polymer chains within the SOM system, in the case of PE in SOM II. Depicted are PE, HS molecules, and calcium ions in blue, orange, and green, respectively. Water beads are represented as dotted spheres in cyan. Please note that the actual simulations are carried out using only one polymer chain in each system and each using one of the different starting positions depicted here.

Modeling NP−SOM Interactions. Recent advances in chemically specific spectro-microscopies allow single-particle detection of polymer NPs in the regime of tens of nanometers.[48−50] To investigate NP−SOM association configurations near the experimentally accessible regime, a single, large nanoparticle consisting of PE chains was constructed. The required numbers of monomers were estimated from the Flory equation, which links these to the expected R_g of a specific polymer in aqueous solution, using empirical parameters.[51] Eighteen independent PE chains of 500 coarse-grained beads were built with polyply and placed in a cubic water box with an edge length of 60 nm, providing sufficient space for folding into a compact aggregate without artifacts, which could arise from periodic boundary conditions. The number of water beads was adjusted to obtain a water density of around 1g cm^{-3}. After equilibration, the chains aggregated into an almost spherical nanoparticle with a diameter of around 9.5 nm. Further information about the generation and equilibration of the NP particle is given in Section S2.

The resulting PE nanoparticle was resolvated in a 30 nm box, and an equilibrated SOM II model system, previously used for the single polymer−SOM interaction simulations, was positioned adjacent to the particle to promote interfacial contact (see the initial configuration in Figure 3). The water density was again set to around 1g cm^{-3}. SOM II was selected because the single polymer−SOM simulations (described in the Results and Discussion) revealed a strong affinity of PE for SOM II, making it an appropriate choice for probing potential adsorption and heteroaggregation motifs on the nanoparticle surface. To assess the influence of the protonation state of

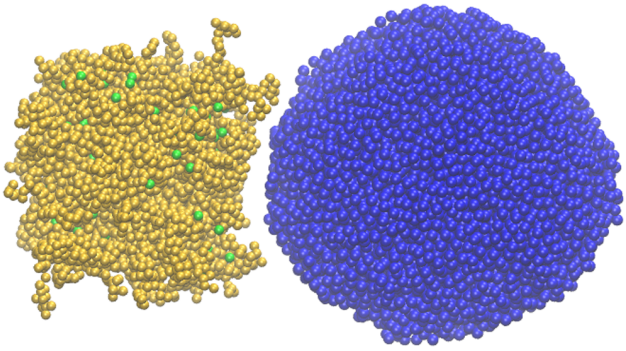


Figure 3. Initial configuration of the HS cluster and NP particle for the system at pH = 5. For the cases at other pH levels, a similar positioning was utilized. Depicted are PE in blue, HS molecules in orange, and calcium ions in green. The representation of water beads is left out for clarity.

https://doi.org/10.1021/acs.jpcb.5c06072
J. Phys. Chem. B 2025, 129, 11605−11615

acidic functional groups, the procedure was repeated at different levels of pH, ranging from 5 to 8, modeled by altering the ratio of carboxylate to carboxylic acid groups of the HS molecules while leaving all other compositional features unchanged (see Table S1 for details of the modeled ratios). No chemical modifications were applied to the nanoparticle surface, e.g., to mimic aging.

Computational Details. All MD simulations were carried out with GROMACS version 2019.4.[52] The analyses were conducted using a combination of GROMACS tools and Python 3 packages, including NumPy,[53] MDAnalysis,[54,55] Matplotlib,[56] and IPython.[57] The visualizations and trajectory inspections were performed with VMD.[58]

All models first underwent energy minimization followed by 10 ns NPT equilibration. For the single polymer−SOM systems, each replica was simulated for 1.5 μs of production time. For the nanoparticle systems, 10 μs long production runs were conducted, of which the last 4 μs were used for the analysis, ensuring the consideration of equilibrated HS molecules after adsorption for every pH case. All simulations employed the leapfrog integrator with a 20 fs time step. The temperature was maintained at T = 298.15 K using the velocity-rescale thermostat[59] (τ_T = 1 ps). Pressure coupling was achieved using the Berendsen barostat[60] during equilibration (τ_p = 8 ps) and the Parrinello−Rahman[61] barostat during production runs (τ_p = 12 ps), with the reference pressure set to 1 bar. Lennard-Jones interactions were evaluated with the Verlet cutoff scheme and a cutoff of 1.1 nm. For the Coulomb interaction, the reaction-field method using the same cutoff and a relative dielectric constant of ε_r = 15 was used. Constraints were realized with the LINCS algorithm,[62] and periodic boundary conditions were applied in every direction.

RESULTS AND DISCUSSION

Single Polymer−SOM Interactions. *Coordination Numbers.* The overall partitioning behavior of PEO and PE in the SOM systems was quantified from the coordination numbers of water (W) and HS beads surrounding the polymer chains, as depicted in Figure 4. The coordination number denotes the

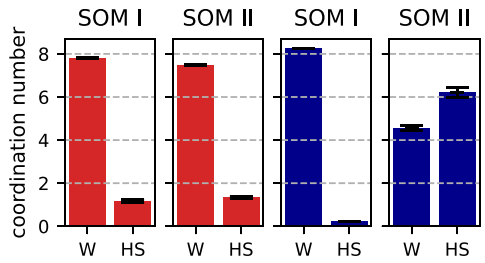


Figure 4. Coordination numbers of water (W) and HS beads around PEO (red) and PE (blue) averaged over 20 trajectories for each polymer and each SOM type.

average count of beads of a given type within a cutoff of $1.5\sigma_{ij}$, where σ_{ij} is the Lennard-Jones sigma value between beads i and j. This quantity offers direct insight into the localization and interactions of the polymer chains within the SOM matrix.

PEO chains exhibit a high coordination number with water and a lower coordination number with HS beads, largely independent of the two SOM compositions considered. Although HS provides polar binding sites for PEO, which are present to different extents in the respective systems, the polymer's stronger affinity for water dictates its partitioning. The diffusive drag imposed by water beads likely further suppresses contacts with HS molecules so that, on average, PEO chains penetrate only to a small extent into the HS phase.

In contrast to PEO, the partitioning of PE differs notably between the SOM compositions. In SOM I, PE is surrounded by many water beads but very few HS beads, indicating weak interaction with HS. Although counterintuitive for a hydrophobic polymer, this behavior arises from the high concentration of carboxylate and other polar functional groups in SOM I, which attract water more strongly than PE and thereby hinder PE−HS contacts. As a result, PE chains predominantly occupy water-filled voids within the SOM matrix. In SOM II, the substantially lower carboxylate content removes this competition between water and PE chains, allowing them to integrate into the HS phase, visible by the significantly higher HS coordination number. Thus, unlike that of PEO, the partitioning of PE is more sensitive to the SOM composition.

Interaction Energies. To further resolve the polymer−SOM contacts, interaction energies between the polymer chains and the individual SOM functional groups were calculated. The five groups exhibiting the strongest contributions in every system are depicted in Figure 5.

Although interaction energies are straightforward to compute, binding free energies between the polymer chains and individual SOM functional groups would provide a more rigorous thermodynamic measure. Such free energy calculations were not conducted here due to the high computational cost and the pronounced heterogeneity of the SOM systems. Instead, to quantify the polymers' local environment, coordination numbers of each functional group around the polymer chains were additionally determined (see Figure S1).

For the PEO in SOM I, polar functional groups dominate the interactions. The carboxylate groups show the highest contribution, which is connected to their strong polarity and high abundance in SOM I. Ketone, amine, and phenol moieties also interact notably with PEO. This behavior is consistent with the atomistic picture in which the ether oxygens of PEO can interact strongly with polar HS interaction sites, resulting in a pronounced affinity toward these groups.

However, carboxylate groups, like PEO monomers, act exclusively as hydrogen-bond acceptors. Therefore, the high energies observed here can be rationalized only by water-mediated interactions, which are not explicitly resolved in the coarse-grained representation. As shown by the coordination numbers in Figure 4, PEO is predominantly surrounded by water beads, indicating that water implicitly mediates the interactions between PEO and HS groups. This interpretation is further supported by the radial distribution functions comparing PEO−carboxylate and PEO−water interactions (Figure S2), which reveal a stronger and more direct association of carboxylates with water. When the interaction energies are interpreted, it is thus essential to consider both the overall polymer partitioning within the SOM systems (Figure 4) and the relative abundance of the functional groups, such as the high carboxylate content in the SOM I system.

Although aromatic moieties are more hydrophobic, they still interact to a considerable extent with PEO. The delocalized electron clouds of the aromatic structures can undergo slight polarization, producing a moderate attraction. In the coarse-grained model employed here, this mechanism is represented implicitly by the Lennard-Jones interaction parameters.

https://doi.org/10.1021/acs.jpcb.5c06072
J. Phys. Chem. B 2025, 129, 11605−11615

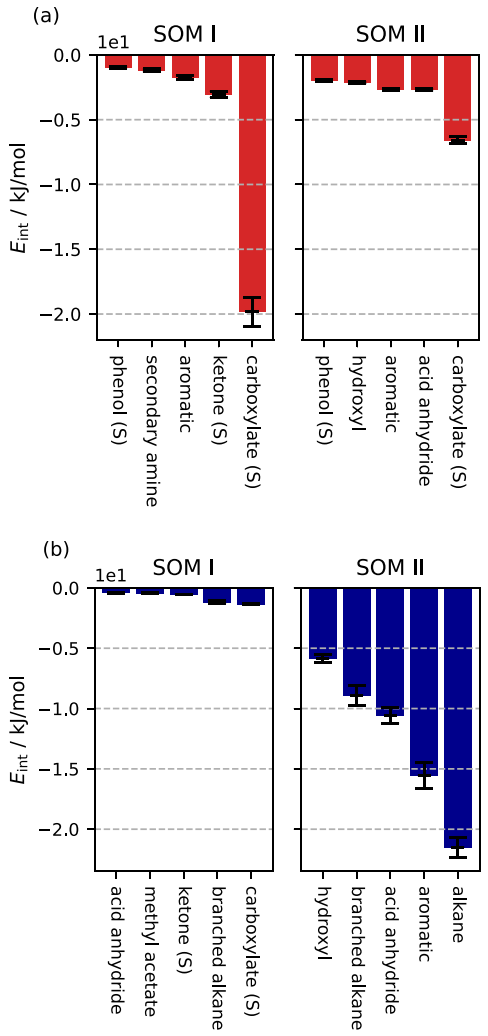


Figure 5. Interaction energies (time-averaged) between PEO (a) or PE (b) and HS functional groups. The values are further averaged across 20 independent trajectories for every polymer and SOM type. Since similar functional groups within the HS molecules can be represented by beads of different sizes, an "S" denotes the representation of the respective functional group by a Martini bead with a small size.

Similarly, polar moieties in SOM II dominate in the corresponding interaction energy profile of PEO. Although the absolute value arising from the carboxylate groups is lower due to their reduced concentrations in SOM II, they remain the largest single contributor. Aromatic fragments show a similar interaction energy as in SOM I, which is consistent with their comparable abundance in the two SOM systems (see also Table 1).

For PE in SOM I, the largest interaction energy, yet small in absolute value, arises with the highly abundant carboxylate groups. Other polar functional groups follow, and branched-alkane fragments appear as the only noteworthy hydrophobic contributor. These functional groups reflect the contacts with the HS molecules that occur during the competition with the water particles for interaction sites. The overall magnitudes remain low, consistent with the results from the coordination number analysis (Figure 4).

In SOM II, the affinity of PE for the HS phase, indicated by the high coordination numbers, is reflected in the attraction to mostly hydrophobic functional groups. Nonetheless, polar interaction sites provided by acid anhydride and hydroxyl groups also contribute. In comparison to PEO for SOM II, the functional-group spectrum of PE shows pronounced differences, indicating that besides the concentration of the present groups, the chemical nature of the considered polymer dictates the interaction motifs.

SOM samples whose acidic and ketone concentrations resemble those investigated here are likely to induce comparable polymer−SOM interactions and partitioning behavior. Nevertheless, natural soils are considerably more heterogeneous, containing a broad spectrum of chemical constituents that could modulate these effects. Therefore, the conclusions presented here represent a step toward elucidating the full range of molecular interaction mechanisms in real soil.

Comparable investigations have been reported by Zhang et al.,[34] who analyzed the aggregation mechanisms of leonardite humic acid (LHA) with polymer chains using atomistic MD and density functional theory calculations. They observed that PE and polystyrene formed homogeneous aggregates, while LHA molecules adsorbed onto their outer surfaces. In this configuration, LHA clusters are bound to PE via hydrophobic functional groups, leaving hydrophilic groups exposed to the aqueous phase, consistent with our finding that PE interacts predominantly through hydrophobic contacts, as shown by the functional-group contributions in Figure 5.

Diffusion Coefficients. Diffusion coefficients were determined to assess the mobility of the polymer chains (Figure 6).

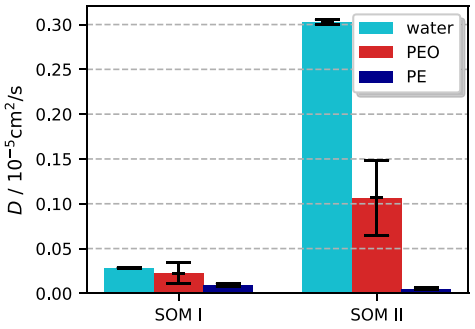


Figure 6. Diffusion coefficients of water beads, PEO, and PE for different SOM types.

Because Martini coarse-grained models generally underestimate friction, the correct absolute values cannot be reproduced. Therefore, the interpretation is restricted to relative trends and comparisons with the water diffusion. The corresponding mean-squared displacement (MSD) curves can be taken from Figure S3.

In SOM I, the diffusion coefficient of PEO is comparable to that of water beads, reflecting its predominant localization in the water phase, as indicated by the coordination numbers. Its mobility is therefore governed by frequent collisions with water. In SOM II, the absolute diffusion of PEO is higher but remains below that of water. Since the overall partitioning between water and HS is nearly identical to SOM I, the reduced mobility must arise from matrix-induced effects, which do not alter the coordination numbers. There are three likely reasons for this slowdown: (i) momentum transfer from surrounding water beads does not scale linearly beyond a certain diffusion threshold; (ii) the polarity of PEO promotes the formation of a solvation shell that alters its dynamics; and

https://doi.org/10.1021/acs.jpcb.5c06072
J. Phys. Chem. B 2025, 129, 11605−11615

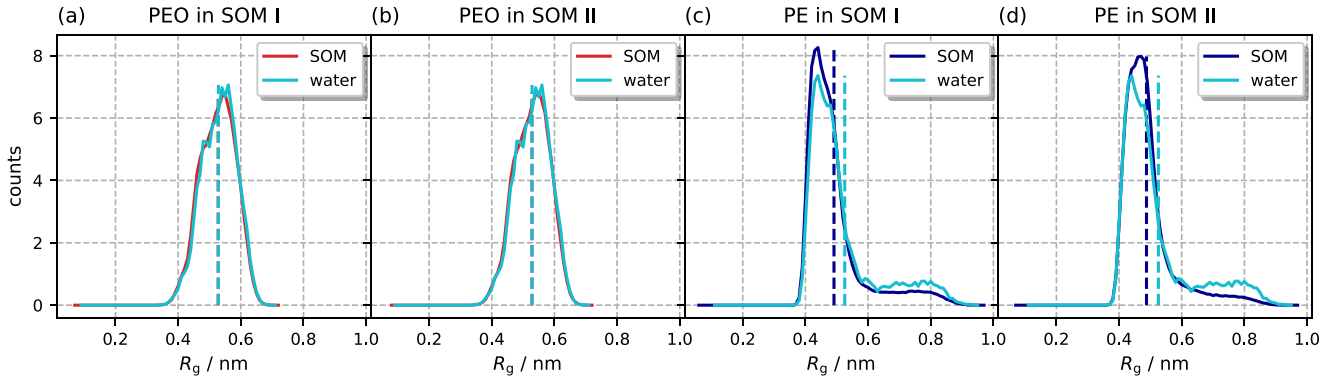


Figure 7. (a–d) Averaged R_g distributions for each polymer in SOM systems (20 different trajectories) vs bulk water for different systems as indicated. Dotted vertical lines represent the mean R_g of each distribution.

(iii) interactions with the heterogeneous SOM environment impose additional constraints on polymer motion.

In contrast, PE diffuses far more slowly than PEO, with a coefficient in SOM II being slightly lower than that in SOM I. Two factors are decisive. First, in SOM I, PE is trapped within water-filled voids, limiting its ability to move through the SOM system. Second, in SOM II, the strong PE–HS attractions lead to the formation of PE–SOM aggregates, effectively fixing the chains within the matrix. Figure 8 illustrates a PEO chain in SOM II that partitions from the HS into the continuous aqueous phase, facilitating its diffusion, whereas PE in SOM I remains trapped within a water-filled void of the SOM matrix, which restricts its mobility. Unlike PEO, these direct interactions and steric constraints almost immobilize PE, an effect that would probably be even more pronounced for longer chains.

Radii of Gyration. To evaluate how SOM modifies the polymer conformations, the distribution of the R_g was examined, which holds information about the frequency of folded and extended configurations. The investigated cases in SOM are compared to those of PE and PEO situated in bulk water (shown in Figure 7).

The R_g distributions of PEO in both SOM systems are virtually indistinguishable from those obtained in bulk water, indicating that the SOM matrix exerts little influence on PEO conformations. Because the chains spend most of the time in the water phase and diffuse at rates comparable to those of the water beads, this finding is consistent with the weak structural impact suggested by the coordination number analysis. Although PEO exhibits a strong energetic affinity for carboxylate groups, these contacts neither induce chain accumulation in the SOM nor alter its conformation. However, please note that the present systems contain a high water concentration and that at lower water contents the SOM matrix might have a stronger effect on PEO.

For PE, the R_g distributions deviate slightly from the bulk-water reference, similarly for SOM I and SOM II. In both cases, large R_g becomes less probable, and the distributions shift toward smaller values, indicating a greater tendency for the chains to adopt folded states. In SOM I, this shift likely arises from steric confinement within SOM voids (Figure 8b), combined with the inability of the PE chains to disrupt the solvation shell around polar functional groups, consistent with their very low interaction energies and coordination numbers. In SOM II, the strong PE–HS attractions favor chain aggregation within the SOM matrix. Therefore, the folding

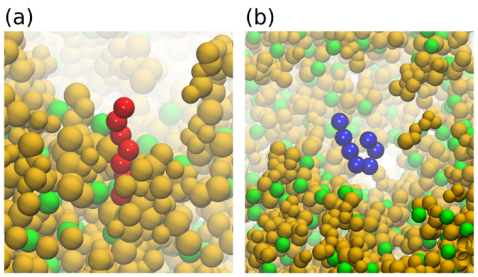


Figure 8. Snapshots of polymer chains in SOM systems. (a) PEO (red) in SOM II, with its lower segment enveloped by HS beads and the upper segment extending into water. (b) PE (blue) in SOM I being inside a water void. HS beads are represented in orange, and calcium ions are in green. Water has been omitted for clarity.

may accompany transfer into an energetically preferred stable state.

Unlike PEO, PE is sensitive to both SOM functional groups and steric restriction by voids. While PE ordinarily folds in a polar solvent such as water, the chain length used here permits occasional unfolding via collisions with water beads, resulting in higher R_g values. Different chain lengths or other void-to-R_g ratios may yield distinct outcomes, and PE could undergo unfolding in the SOM under altered conditions.

Plastic polymer chains occurring in terrestrial environments are generally longer than those investigated in this study. To assess the influence of polymer size on the results, polymers with twice the radius of gyration were also simulated on a trial basis. While qualitative trends, such as partitioning within the SOM systems, remained unchanged, the diffusivity decreased substantially, preventing adequate sampling of various SOM interaction sites. Consequently, considerably longer polymers would require enhanced sampling techniques (e.g., replica exchange MD or simulated annealing) to properly capture the heteroaggregation between polymer chains and SOM matrices.

Because soils exhibit a wide hierarchy of pore sizes, the relationships observed here could be transferred to other scales by considering the relative size ratios between the extent of the polymers and the voids they encounter. For example, HS contains nanopores of about 0.5 nm that dominate its internal surface area,[63] while organomineral complexes can contain pores from 2 to 10 nm at mineral interfaces.[64] At larger scales, soil aggregates stabilized by SOM exhibit micropores from 5 to 30 μm and mesopores from 30 to 75 μm,[65] and biological activity such as root growth creates pores that can extend to several millimeters.[66] This hierarchical pore network suggests

https://doi.org/10.1021/acs.jpcb.5c06072
J. Phys. Chem. B 2025, 129, 11605–11615

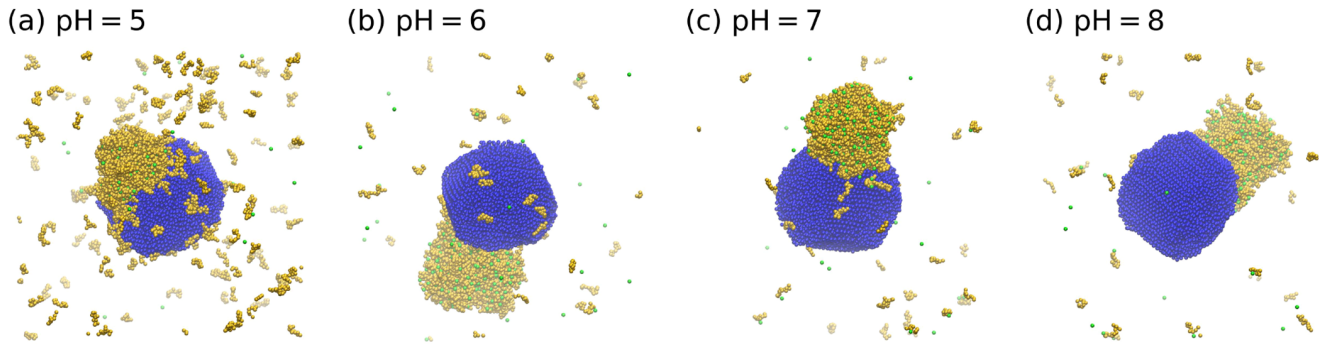


Figure 9. Snapshots of the final trajectory frame at pH = 5 (a), 6 (b), 7 (c), and 8 (d). Pronounced dissociation of the HS cluster is evident at pH = 5. The PE particle is depicted in blue, HS beads in orange, and calcium ions in green. Water is omitted for clarity.

that polymer−pore interactions studied at the nanoscale may be extrapolated to higher levels of soil organization, although further research is needed to validate these cross-scale relationships under environmental conditions.

NP−SOM Interactions. The following subsections present the results of NP−SOM simulations conducted at different pH values. Standard deviations are shown as error bars in the plots to indicate variability, and the corresponding trajectory data are provided in Section S3.

HS Cluster Size. Figure 9 shows representative snapshots at pH = 5−8. At all investigated pH levels, the HS cluster adsorbs to the PE nanoparticle surface and remains bound over the rest of the simulation time. Although full desorption was not observed, it cannot be excluded at longer time scales. The adsorption occurs either immediately or after a short delay (please refer to the trajectory-resolved coordination numbers for pH = 6 in Figure S10). The snapshots also indicate different degrees of HS dissolution and distinct interactions with the PE nanoparticle. A part of the initial HS cluster dissolves into the water phase, leaving isolated HS molecules and a single, larger residual cluster adsorbed onto the PE surface[1]. The isolated HS molecules can also be classified as "dissolved organic matter".

To quantify HS dissolution, the GROMACS tool *gmx clustsize* was used to determine both the number of molecules in the largest residual HS cluster and the total number of clusters in the system. These quantities serve as indicators of the degree of dissolution, i.e., the number of dissolved HS molecules (the fraction of HS molecules dispersed from the main cluster). The results for the various pH levels are listed in Figure 10. Subsequent analyses distinguish between the largest cluster and the remaining HS molecules in the system. To justify this separation, trajectory-resolved sizes of the largest and second-largest clusters are provided in Figure S8, which shows that the former consistently contains significantly more molecules. Please note that in each simulation, the largest cluster corresponds to the one attached to the NP.

In Figure 10a, a larger number of clusters indicates that more HS molecules have entered the solution phase because every isolated HS molecule is counted as a single cluster. Hence, at pH 5 markedly more HS molecules are dissolved, whereas this behavior is progressively suppressed as the pH increases.

The opposite trend is observed for the number of molecules in the largest cluster, as shown in Figure 10b. When more HS molecules migrate into solution, the residual aggregate necessarily shrinks. These findings demonstrate that the

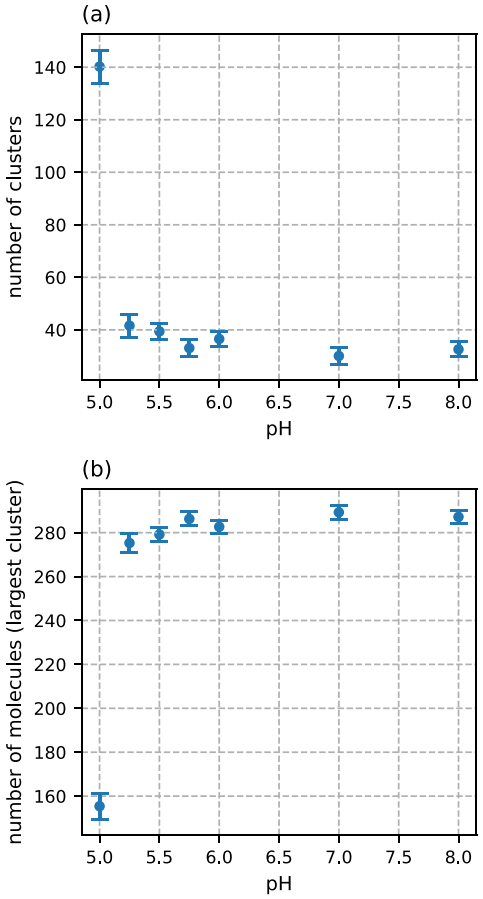


Figure 10. Results of the cluster analysis, including the number of HS clusters in the system (a) and the number of molecules in the largest HS cluster (b) for different pH values.

intermolecular forces stabilizing the largest HS cluster diminish under low pH conditions.

The corresponding trajectory-resolved curves for both observables are presented in Figure S7, showing a convergence. Note, however, that in cases with substantial HS dissolution (pH = 5), the periodic boundary conditions impose a saturated concentration of HS in the simulation box. In real systems, dissolution of HS clusters could therefore be stronger.

Borisover et al.[67] investigated the dilution of HS aggregates in aquatic systems using atomistic simulations. The authors found that a decisive amount of the stabilizing forces arose from the strong interactions between the carboxylate groups

https://doi.org/10.1021/acs.jpcb.5c06072
J. Phys. Chem. B 2025, 129, 11605−11615

and metal ions in the soil systems, especially from calcium ions. The aggregates were further stabilized by hydrogen-bonding networks, a feature that the coarse-grained model can capture only implicitly. Further, with the lack of bidentate motifs between the coarse-grained carboxylate groups and calcium ions (see Dettmann et al.[39]), the stability of the coarse-grained HS structures is expected to be lower against dilution compared to the atomistic models. Therefore, the pH scale, which links the ratio of carboxylate to carboxylic acid groups to the level of HS dilution, should be interpreted qualitatively.

NP Coverage. The NP-surface coverage by HS molecules was estimated as the ratio of NP-surface beads in contact with HS to the total number of NP-surface beads, effectively being the fraction of the surface covered by HS. Figure 11 displays these results, showing the contributions from the largest HS cluster and the dissolved HS molecules. The trajectory-resolved coverage curves are provided in Figure S9.

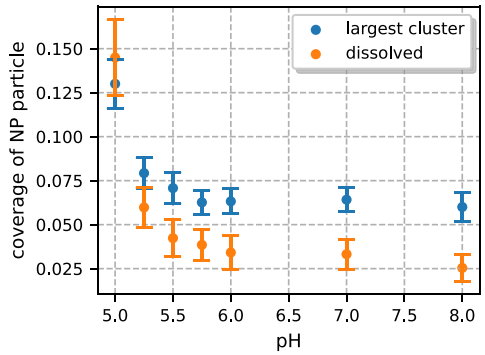


Figure 11. Percentage of NP-surface beads covered by HS molecules at different pH levels. The coverage is distinguished between contributions from the largest HS cluster (blue) and the dissolved HS molecules (orange).

Lower pH values promote higher HS adsorption on the NP surface. At pH = 5, the coverage contributed by dissolved HS molecules slightly exceeds that from the largest cluster, whereas at higher pH levels, the largest cluster dominates the coverage. For both molecular groups, the coverage decreases with an increasing pH and converges toward a similar value.

In Figure 10a, the curve for the number of HS clusters follows the same trend as that for the coverage. A higher HS concentration in solution supplies more molecules that can adsorb on the NP surface, thereby explaining the similar behavior found for the coverage contributed by the dissolved molecules. By contrast, a lower number of molecules in the largest HS cluster (Figure 10b) would imply reduced coverage, yet the opposite is observed. This apparent contradiction is presumably linked to a pH-dependent change in the shape of the largest cluster, most clearly visible in the trajectory snapshot for pH = 5 (Figure 9a), where the cluster deforms and increases its contact area with the NP particle.

The deformation was quantified via sphericity indices for the largest HS cluster, which measures its deviation from spherical symmetry (Figure S12). These indices show that at lower pH values, the cluster becomes less spherical, a change that can be related to a spreading of the cluster on the NP surface and the higher coverage measured under these conditions. Such spreading is likely connected to the reduced cluster stabilization emerging from the absence of Coulombic interactions due to the protonation of carboxylate groups. Nevertheless, the pH-dependent polarity of the HS molecules may also influence the interaction strength with the NP particle and, therefore, the overall coverage.

Please note that the coverage is a measure of the contact area and does not state the actual amount of HS molecules adsorbed on the particle. This is expected to be an interplay between the stability of the largest HS cluster and the changing polarity of the functional groups, dictated by pH, which elevates attraction to the NP particle.

Interaction Energies. To gain further insight into HS−PE interactions and HS cluster stability, interaction energies between the relevant molecular groups were calculated (Figure 12). For the HS−water Lennard-Jones (LJ) term, the energy of the largest cluster becomes more negative as the pH rises, whereas the dissolved molecules show the opposite trend. This behavior reflects the pH-dependent redistribution of HS: when the cluster dissolves at low pH, it contains fewer HS molecules and therefore forms fewer HS−water contacts, making its LJ contribution less pronounced. Conversely, the larger population of the dissolved molecules yields more interaction sites with water and a lower LJ energy. On the other hand, the HS−PE LJ energies follow the surface coverage (Figure 11). More HS−PE contacts (higher coverage) lead to more negative LJ energies.

LJ and Coulomb interactions with the calcium ions exhibit the same pH dependence with the LJ term contributing roughly twice as much energy. For the largest cluster, both energies become more negative as the pH increases, underscoring the stabilizing role of calcium ions. Fewer protonated HS groups at low pH weaken the attractive forces between HS and calcium, which promotes cluster dissociation. This trend is

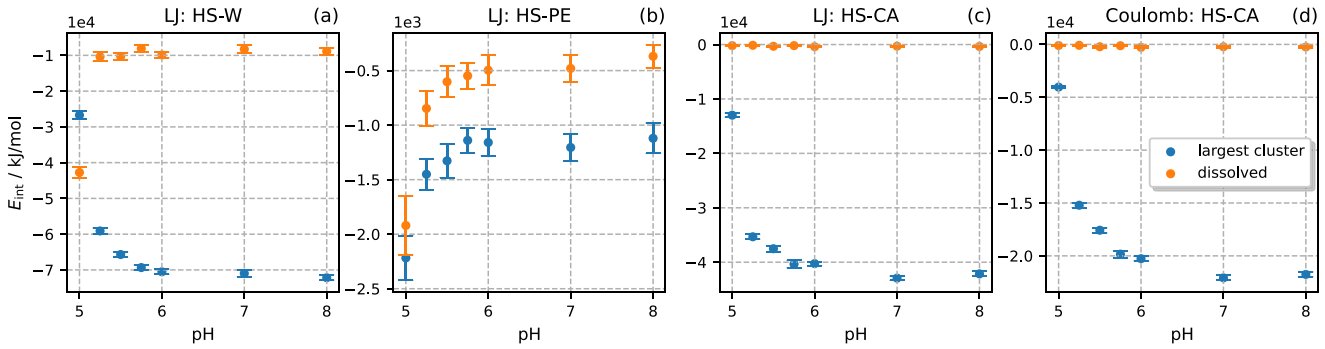


Figure 12. Interaction energies at different pH levels between the largest HS cluster (blue) or dissolved HS molecules (orange) and water (a), PE (b), and calcium ions, thereby differentiating into Lennard-Jones (c) and Coulomb interactions (d). Please note the different scales on the *y*-axis.

https://doi.org/10.1021/acs.jpcb.5c06072
J. Phys. Chem. B 2025, 129, 11605−11615

anticorrelated with the cluster size shown in Figure 10 and explains the high dilution of the HS cluster at the lowest pH by the low number of carboxylate groups present in that case (compare with Table S1). Interaction energies between dissolved HS molecules and calcium are practically negligible, presumably because their dilution in the aqueous phase reduces the probability of contacts.

An analysis of the number of protonated and unprotonated carboxylate groups in both the largest cluster and the dissolved fraction at pH = 5 revealed that predominantly molecules with protonated groups left the initial cluster (Figure S13). Consequently, almost exclusively protonated HS molecules were found in solution. This behavior is associated with the weaker attraction of protonated groups to calcium ions compared to unprotonated ones, leading to their easier detachment from the cluster. This interpretation is consistent with the interaction energies shown in Figure 12c, which indicate that at pH = 5, where most groups are protonated, the interaction with calcium is weakest.

CONCLUSIONS

NPs in terrestrial ecosystems are an increasing ecological issue, yet their molecular-level interaction with SOM is still poorly understood. By employing coarse-grained MD simulations, first, the behavior of polymer chains within SOM matrices at two levels of acidity was investigated. The simulations showed that the polymers' polarity governs key partitioning and diffusion behavior. PEO experienced a strong influence of water beads, expressed in a relatively pronounced diffusion, exhibiting a clear affinity for charged and polar SOM functional groups. On the other hand, PE showed a distinct, acidity-dependent partitioning. In more acidic SOM, it became trapped in water-filled voids due to competition with water particles for adsorption sites, while under less acidic conditions, it interacted preferentially with hydrophobic groups such as alkane and aromatic moieties, resulting in overall low diffusivity in both cases.

In a second set of simulations, SOM association with a PE nanoparticle was investigated at varying pH. At lower pH, reduced carboxylate−calcium ion interactions destabilized HS aggregates, enabling their dispersion and promoting more extensive NP-surface coverage. These findings provide a molecular-scale picture of how SOM composition and acidity regulate NP partitioning, diffusivity, and adsorption/hetero-aggregation processes.

Future studies could incorporate aged or surface-functionalized NPs bearing polar or charged groups, which would introduce new interaction motifs. A more accurate treatment of pH effects should also be pursued. In the current framework, the pH is approximated by altered protonation states, which may overemphasize pH sensitivity. Furthermore, bidentate calcium−carboxylate motifs that stabilize HS aggregates are not explicitly represented, likely underestimating cluster stability at low pH. Beyond pH modeling, the approach readily enables exploration of diverse SOM compositions, such as varying aromatic content, and can be extended to include biologically relevant macromolecules within the Martini 3 framework. Expanding the scope to other polymer types, coadsorption with pollutants, and variations in ionic strength and ion identity would further enhance the environmental relevance of the simulations.

Altogether, this work demonstrates that coarse-grained MD simulations provide a robust framework for unraveling NP−SOM interactions on the molecular scale. By connecting polymer polarity, SOM composition, and pH-dependent behavior to specific partitioning and aggregation motifs, the simulations establish a mechanistic basis for interpreting experimental data and guiding the development of analytical tools to assess the fate of NPs in soils.

ASSOCIATED CONTENT

Data Availability Statement

The topology files required to reproduce the simulations from this study are available in Zenodo: https://zenodo.org/records/15977520.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.5c06072.

Details on the construction of the PE particle model, additional trajectory-resolved quantities from the NP−SOM heteroaggregation simulations, and the sphericity indices of the largest HS clusters at different pH values (PDF)

AUTHOR INFORMATION

Corresponding Author

Ashour A. Ahmed − *University of Rostock, Institute of Physics, D-18059 Rostock, Germany; Leibniz Institute for Catalysis (LIKAT), D-18059 Rostock, Germany;* orcid.org/0000-0002-5933-5249; Email: ashour.ahmed@uni-rostock.de

Authors

Lorenz F. Dettmann − *University of Rostock, Institute of Physics, D-18059 Rostock, Germany;* orcid.org/0000-0001-8407-4469

Oliver Kühn − *University of Rostock, Institute of Physics, D-18059 Rostock, Germany;* orcid.org/0000-0002-5132-2961

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.5c06072

Notes

The authors declare no competing financial interest.

ADDITIONAL NOTE

[1]In this work, the term adsorption refers to the association between the largest HS cluster or dissolved HS molecules and the PE particle. Nonetheless, depending on the relative proportions of the interacting species, the process could also be described as heteroaggregation.

REFERENCES

(1) Callister, W.; Rethwisch, D. G. Materials science and engineering: An introduction. 2018.

(2) Hahladakis, J. N.; Velis, C. A.; Weber, R.; Iacovidou, E.; Purnell, P. An overview of chemical additives present in plastics: Migration, release, fate and environmental impact during their use, disposal and recycling. *J. Hazard. Mater.* **2018**, *344*, 179−199.

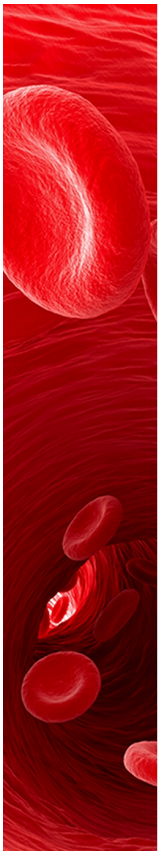
(3) de Sousa, F. D. B. The role of plastic concerning the sustainable development goals: The literature point of view. *Cleaner Responsible Consumption* **2021**, *3*, No. 100020.

(4) Abrahms-Kavunenko, S. Toward an anthropology of plastics. *J. Mater. Cult.* **2023**, *28*, 3−23.

(5) Allen, S.; Allen, D.; Moss, K.; Le Roux, G.; Phoenix, V. R.; Sonke, J. E. Examination of the ocean as a source for atmospheric microplastics. *PLoS One* **2020**, *15*, No. e0232746.

(6) Dokl, M.; Copot, A.; Krajnc, D.; Fan, Y. V.; Vujanović, A.; Aviso, K. B.; Tan, R. R.; Kravanja, Z.; Čuček, L. Global projections of plastic use, end-of-life fate and potential changes in consumption, reduction, recycling and replacement with bioplastics to 2050. *Sustainable Prod. Consumption* **2024**, *51*, 498−518.
(7) LI, W.; Tse, H.; Fok, L. Plastic waste in the marine environment: A review of sources, occurrence and effects. *Sci. Total Environ.* **2016**, *566*−*567*, 333−349.
(8) Sadri, S. S.; Thompson, R. C. On the quantity and composition of floating plastic debris entering and leaving the Tamar Estuary, Southwest England. *Mar. Pollut. Bull.* **2014**, *81*, 55−60.
(9) Zbyszewski, M.; Corcoran, P. L.; Hockin, A. Comparison of the distribution and degradation of plastic debris along shorelines of the Great Lakes, North America. *J. Great Lakes Res.* **2014**, *40*, 288−299.
(10) Free, C. M.; Jensen, O. P.; Mason, S. A.; Eriksen, M.; Williamson, N. J.; Boldgiv, B. High-levels of microplastic pollution in a large, remote, mountain lake. *Mar. Pollut. Bull.* **2014**, *85*, 156−163.
(11) Bergmann, M.; Collard, F.; Fabres, J.; Gabrielsen, G. W.; Provencher, J. F.; Rochman, C. M.; van Sebille, E.; Tekman, M. B. Plastic pollution in the Arctic. *Nat. Rev. Earth Environ.* **2022**, *3*, 323−337.
(12) Allen, S.; Allen, D.; Phoenix, V. R.; Le Roux, G.; Durántez Jiménez, P.; Simonneau, A.; Binet, S.; Galop, D. Atmospheric transport and deposition of microplastics in a remote mountain catchment. *Nat. Geosci.* **2019**, *12*, 339−344.
(13) Zhang, M.; Xu, L. Transport of micro- and nanoplastics in the environment: Trojan-Horse effect for organic contaminants. *Crit. Rev. Environ. Sci. Technol.* **2022**, *52*, 810−846.
(14) Chamas, A.; Moon, H.; Zheng, J.; Qiu, Y.; Tabassum, T.; Jang, J. H.; Abu-Omar, M.; Scott, S. L.; Suh, S. Degradation Rates of Plastics in the Environment. *ACS Sustainable Chem. Eng.* **2020**, *8*, 3494−3511.
(15) Gigault, J.; El Hadri, H.; Nguyen, B.; Grassl, B.; Rowenczyk, L.; Tufenkji, N.; Feng, S.; Wiesner, M. Nanoplastics are neither microplastics nor engineered nanoparticles. *Nat. Nanotechnol.* **2021**, *16*, 501−507.
(16) de Souza Machado, A. A.; Kloas, W.; Zarfl, C.; Hempel, S.; Rillig, M. C. Microplastics as an emerging threat to terrestrial ecosystems. *Global Change Biol.* **2018**, *24*, 1405−1416.
(17) Ng, E.-L.; Huerta Lwanga, E.; Eldridge, S. M.; Johnston, P.; Hu, H.-W.; Geissen, V.; Chen, D. An overview of microplastic and nanoplastic pollution in agroecosystems. *Sci. Total Environ.* **2018**, *627*, 1377−1388.
(18) Weil, R. R.; Brady, N. C. *The Nature and Properties of Soils*, 15th ed.; Pearson: Upper Saddle River, NJ, 2016.
(19) Kleber, M.; Bourg, I. C.; Coward, E. K.; Hansel, C. M.; Myneni, S. C. B.; Nunan, N. Dynamic interactions at the mineral-organic matter interface. *Nat. Rev. Earth Environ.* **2021**, *2*, 402−421.
(20) Blume, H.-P.; Brümmer, G. W.; Fleige, H.; Horn, R.; Kandeler, E.; Kögel-Knabner, I.; Kretzschmar, R.; Stahr, K.; Wilke, B.-M. *Scheffer/Schachtschabel Soil Science*; Springer: Berlin Heidelberg, 2015; pp 55−86.
(21) Schnitzer, M.; Khan, S. U. Humic Substances in the Environment. In *Books in Soils and the Environment*; Marcel Dekker, Inc.: New York, 1972; p 327.
(22) Sutton, R.; Sposito, G. Molecular Structure in Soil Humic Substances: The New View. *Environ. Sci. Technol.* **2005**, *39*, 9009−9015.
(23) International Humic Substances Society https://humic-substances.org (accessed Oct 6, 2025).
(24) Ivleva, N. P. Chemical Analysis of Microplastics and Nanoplastics: Challenges, Advanced Methods, and Perspectives. *Chem. Rev.* **2021**, *121*, 11886−11936.
(25) 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. *Environ. Pollut.* **2023**, *317*, No. 120788.
(26) ten Hietbrink, S.; Materić, D.; Holzinger, R.; Groeskamp, S.; Niemann, H. Nanoplastic concentrations across the North Atlantic. *Nature* **2025**, *643*, 412−416.
(27) Li, P.; Li, Q.; Hao, Z.; Yu, S.; Liu, J. Analytical methods and environmental processes of nanoplastics. *J. Environ. Sci.* **2020**, *94*, 88−99.
(28) Li, S.; Wang, S.; Yan, B.; Yue, T. Surface Properties of Nanoparticles Dictate Their Toxicity by Regulating Adsorption of Humic Acid Molecules. *ACS Sustainable Chem. Eng.* **2021**, *9*, 13705−13716.
(29) Feng, H.; Liu, Y.; Xu, Y.; Li, S.; Liu, X.; Dai, Y.; Zhao, J.; Yue, T. Benzo[a]pyrene and heavy metal ion adsorption on nanoplastics regulated by humic acid: Cooperation/competition mechanisms revealed by molecular dynamics simulations. *J. Hazard. Mater.* **2022**, *424*, No. 127431.
(30) Rama, P.; Gallego-Urrea, J. A.; Abbas, Z. Interfacial interactions of humic acids with polystyrene nano-plastics in aqueous/ionic environments: a molecular dynamics exploration. *Environ. Sci.: Nano* **2023**, *10*, 1385−1393.
(31) Chen, Y.; Tang, H.; Cheng, Y.; Huang, T.; Xing, B. Interaction between microplastics and humic acid and its effect on their properties as revealed by molecular dynamics simulations. *J. Hazard. Mater.* **2023**, *455*, No. 131636.
(32) Chen, Y.; Li, H.; Yin, Y.; Shan, S.; Huang, T.; Tang, H. Effect of microplastics on the adherence of coexisting background organic contaminants to natural organic matter in water. *Sci. Total Environ.* **2023**, *905*, No. 167175.
(33) Chen, Y.; Tang, H.; Li, H.; Yin, Y.; Song, W.; Guo, H.; Huang, T.; Xing, B. Molecular-level insight into the behavior of metal cations and organic matter during the aggregation of polystyrene nanoplastics. *J. Hazard. Mater.* **2024**, *473*, No. 134665.
(34) Zhang, C.; Zhou, Z.; Xi, M.; Ma, H.; Qin, J.; Jia, H. Molecular modeling to elucidate the dynamic interaction process and aggregation mechanism between natural organic matters and nanoplastics. *Eco-Environ. Health* **2025**, *4*, 100122.
(35) Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.* **2013**, *139*, No. 090901.
(36) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-grained protein models and their applications. *Chem. Rev.* **2016**, *116*, 7898−7936.
(37) 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. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(38) 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.; et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **2021**, *18*, 382−388.
(39) Dettmann, L. F.; Kühn, O.; Ahmed, A. A. Martini-Based Coarse-Grained Soil Organic Matter Model Derived from Atomistic Simulations. *J. Chem. Theory Comput.* **2024**, *20*, 5291−5305.
(40) Dettmann, L. F.; Kühn, O.; Ahmed, A. A. Automated Parametrization Approach for Coarse-Graining Soil Organic Matter Molecules. *J. Chem. Theory Comput.* **2024**, *20*, 10684−10696.
(41) Grünewald, F.; Alessandri, R.; Kroon, P. C.; Monticelli, L.; Souza, P. C.; Marrink, S. J. Polyply; a python suite for facilitating simulations of (bio-) macromolecules and nanomaterials. *Nat. Commun.* **2022**, *13*, No. 68.
(42) Escalona, Y.; Petrov, D.; Oostenbrink, C. Vienna soil organic matter modeler 2 (VSOMM2). *J. Mol. Graphics Modell.* **2021**, *103*, No. 107817.
(43) Dettmann, L. F. SOM2CG: Coarse-graining Soil-organic-matter Models *GitHub Repository*, 2025 https://github.com/LorenzDettmann/SOM2CG (accessed Aug 8, 2025).
(44) Escalona, Y.; Petrov, D.; Galicia-Andrés, E.; Oostenbrink, C. Exploring the Macroscopic Properties of Humic Substances Using Modeling and Molecular Simulations. *Agronomy* **2023**, *13*, No. 1044.

(45) Grünewald, F.; Souza, P. C. T.; Abdizadeh, H.; Barnoud, J.; de Vries, A. H.; Marrink, S. J. Titratable Martini model for constant pH simulations. *J. Chem. Phys.* **2020**, *153*, No. 024118.
(46) 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*, No. 115237.
(47) 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*, No. 1138.
(48) Hu, R.; Zhang, K.; Wang, W.; Wei, L.; Lai, Y. Quantitative and sensitive analysis of polystyrene nanoplastics down to 50 nm by surface-enhanced Raman spectroscopy in water. *J. Hazard. Mater.* **2022**, *429*, No. 128388.
(49) Ruan, X.; Xie, L.; Liu, J.; Ge, Q.; Liu, Y.; Li, K.; You, W.; Huang, T.; Zhang, L. Rapid detection of nanoplastics down to 20 nm in water by surface-enhanced raman spectroscopy. *J. Hazard. Mater.* **2024**, *462*, No. 132702.
(50) Li, Y.; Zhang, C.; Tian, Z.; Cai, X.; Guan, B. Identification and quantification of nanoplastics (20−1000nm) in a drinking water treatment plant using AFM-IR and Pyr-GC/MS. *J. Hazard. Mater.* **2024**, *463*, No. 132933.
(51) Dettmann, L. F.; Kühn, O.; Ahmed, A. A. Coarse-grained molecular dynamics simulations of nanoplastics interacting with a hydrophobic environment in aqueous solution. *RSC Adv.* **2021**, *11*, 27734−27744.
(52) 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.
(53) 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.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357−362.
(54) 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.
(55) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011**, *32*, 2319−2327.
(56) Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90−95.
(57) Pérez, F.; Granger, B. E. IPython: a System for Interactive Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 21−29.
(58) Humphrey, W.; Dalke, A.; Schulten, K. VMD − Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(59) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101, DOI: 10.1063/1.2408420.
(60) 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. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(61) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(62) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(63) de Jonge, H.; Mittelmeijer-Hazeleger, M. C. Adsorption of CO2 and N2 on Soil Organic Matter: Nature of Porosity, Surface Area, and Diffusion Mechanisms. *Environ. Sci. Technol.* **1996**, *30*, 408−413.
(64) Kögel-Knabner, I.; Guggenberger, G.; Kleber, M.; Kandeler, E.; Kalbitz, K.; Scheu, S.; Eusterhues, K.; Leinweber, P. Organo-mineral associations in temperate soils: Integrating biology, mineralogy, and organic matter chemistry. *J. Plant Nutr. Soil Sci.* **2008**, *171*, 61−82.
(65) Dal Ferro, N.; Stevenson, B.; Morari, F.; Müller, K. Long-term tillage and irrigation effects on aggregation and soil organic carbon stabilization mechanisms. *Geoderma* **2023**, *432*, No. 116398.
(66) Kautz, T. Research on subsoil biopores and their functions in organically managed soils: A review. *Renewable Agric. Food Syst.* **2015**, *30*, 318−327.
(67) Borisover, M.; Petrov, D.; Oostenbrink, C.; Galicia-Andrés, E. Diluting humic substances in water in molecular dynamics simulations: Are aggregates stable? *Colloids Surf., A* **2025**, *704*, No. 135507.


CAS BIOFINDER DISCOVERY PLATFORM™
CAS BIOFINDER HELPS YOU FIND YOUR NEXT BREAKTHROUGH FASTER
Navigate pathways, targets, and diseases with precision
Explore CAS BioFinder
CAS
A division of the American Chemical Society

https://doi.org/10.1021/acs.jpcb.5c06072
J. Phys. Chem. B 2025, 129, 11605−11615

Supporting Information

Coarse-Grained Simulations of the Nanoplastic Interaction with Soil Organic Matter

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

S1 Additional results of the single polymer simulations

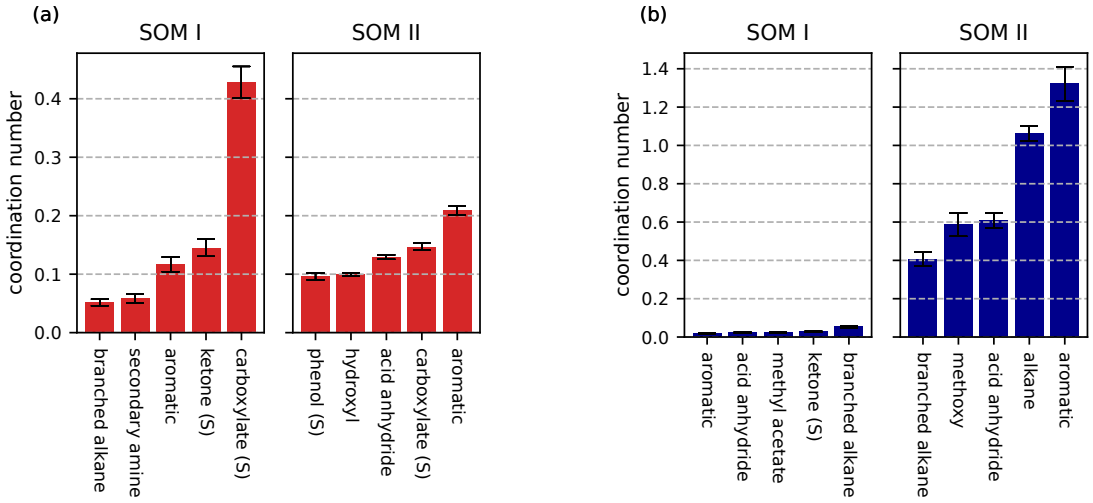


Figure S1: Coordination numbers of the five most contributing SOM functional groups around PEO (a) and PE (b).

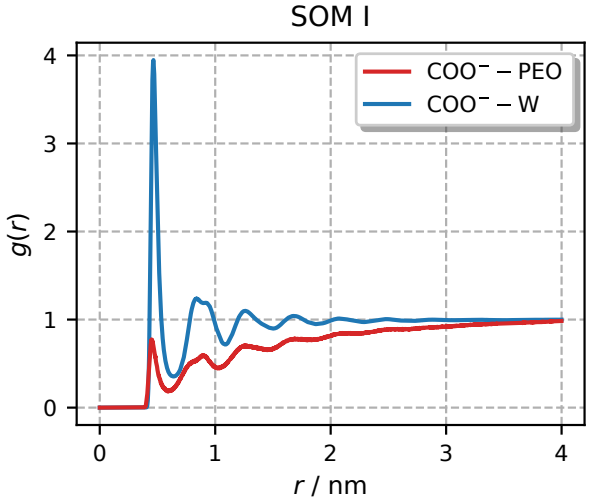


Figure S2: Radial distribution functions between carboxylate groups and PEO (red) and between carboxylate groups and water beads (blue). The stronger association of carboxylate groups with water, compared to PEO, is evident from the higher first peak of the corresponding curve.

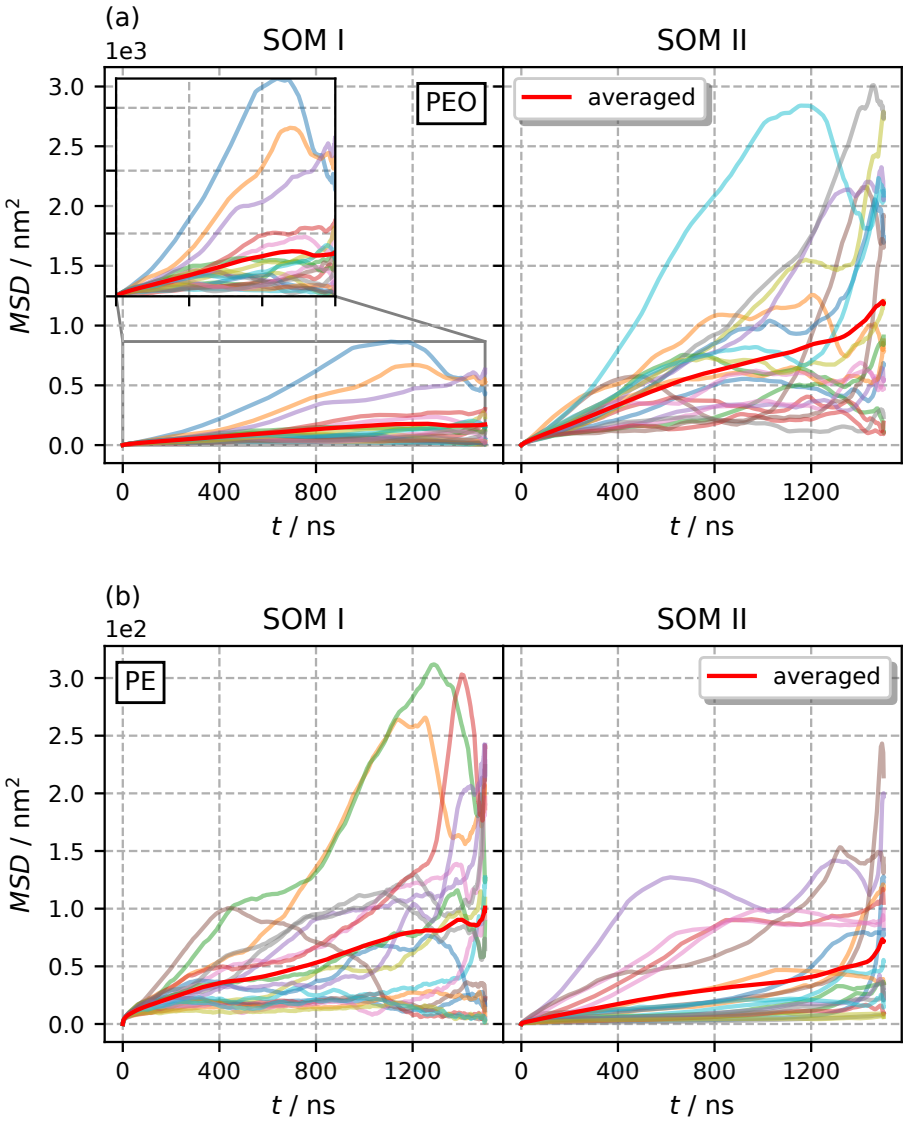


Figure S3: Mean squared displacements (MSDs) for PEO (a) and PE (b). In each case, the red MSD curve shows the averaged MSD of the 20 individual trajectories, which was used to calculate the diffusion coefficients. To account for nonlinearities at the beginning and the end of these curves, arising from using multiple time origins, the linear fit was restricted to the 10 %–90 % range. Please note the zoomed-in region in the case of the smaller MSDs for clearer visibility.

S2 Generation and equilibration of the PE nanoplastic particle

The PE nanoparticle was created aiming to approach a diameter of around 10 nm. The required number of beads was estimated using the Flory equation for polymers and calculated fit parameters from our previous work[1]. The estimated total number of around 9000 was split up into 18 polymer chains, each containing 500 coarse-grained beads. Note that one bead represents four carbon atoms and the attached hydrogens.

The following procedure was used to compact the relatively long polymer chains in an aqueous environment. A single polymer chain was first created in a vacuum using the polyply tool[2]. 18 of these chains were then inserted into a cubic box with an edge length of 18.75 nm, using the GROMACS tool *gmx insert-molecules*. To account for artifacts that can arise from periodic boundary conditions, the chains were centered in a larger box having an edge length of 60 nm and filled with water using the *gmx solvate* tool, with a total number of 1786910 coarse-grained water beads. After an energy minimization and NVT equilibration of 10 ns, an NPT equilibration of 1.5 µs using the Parrinello-Rahman barostat was conducted. The rest of the parameters were equivalent to the ones stated in the *Computational Details* section of the main manuscript. R_g and potential energy curves of the NPT equilibration are depicted in Figures S4 and S5.

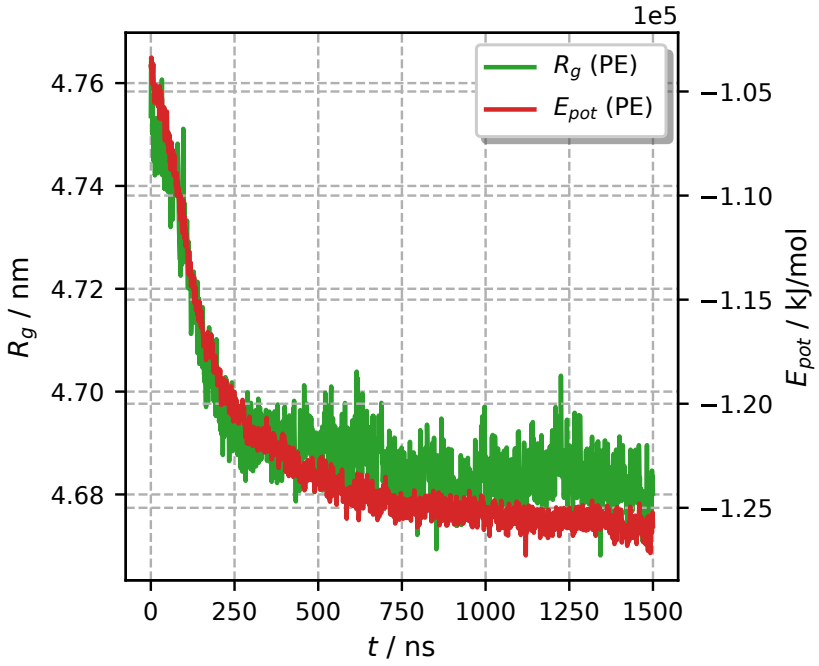


Figure S4: R_g (green) and potential energy (red) of the sole PE nanoparticle during the NPT equilibration as a function of the simulation time t. Both quantities decrease rapidly in the first 250 ns and reach a plateau, indicating the equilibrated formation of the sphere-like nanoparticle. The potential energy displayed here contains bond, angle, dihedral angle, and PE-PE Lennard-Jones interactions, which are separately depicted in Figure S5.

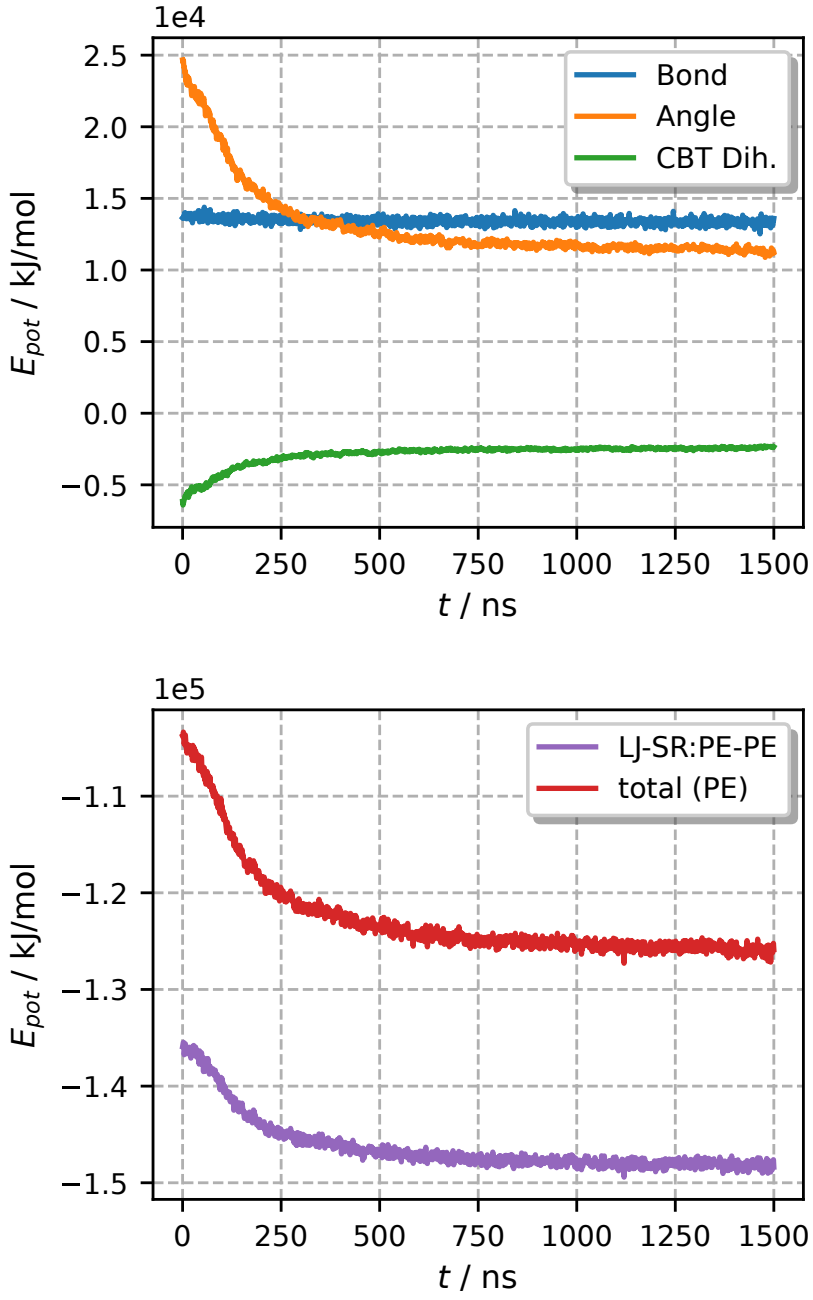


Figure S5: Potential energy contributions of the sole PE particle. The total part (red) is the sum of the other displayed contributions, equal to the potential energy curve in Figure S4. The slowest converging types are the angle potentials (orange) and Lennard-Jones potentials (violet).

S3 Additional information about the nanoparticle simulations

Table S1: Ratios of carboxylate (COO^-) to carboxylic acid (COOH) functional groups as a function of the modeled pH. "Input" refers to the theoretical ratios, while "Output" denotes the values generated by the VSOMM2 during model construction. At pH = 8.0, no COOH functional groups were present in the system, and therefore no ratio could be defined. The default pKa value of the VSOMM2 of 4.76 was used, corresponding to the one of carboxylic acid.

pH	5.0	5.25	5.5	5.75	6.0	7.0	8.0
Input	1.7	3.1	5.5	9.8	17.4	173.8	1737.8
Output	0.4	2.9	5.6	11.5	21.4	212.0	-

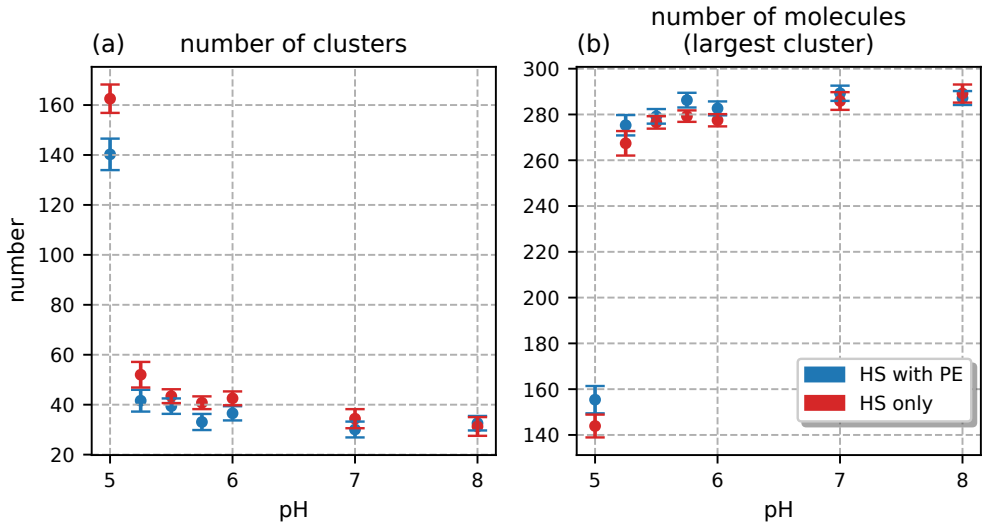


Figure S6: Results of the cluster analysis, including the number of HS clusters in the system (a) and the number of molecules in the largest HS cluster (b) for different pH values. The red data points represent results from additional simulations performed without the PE nanoparticle.

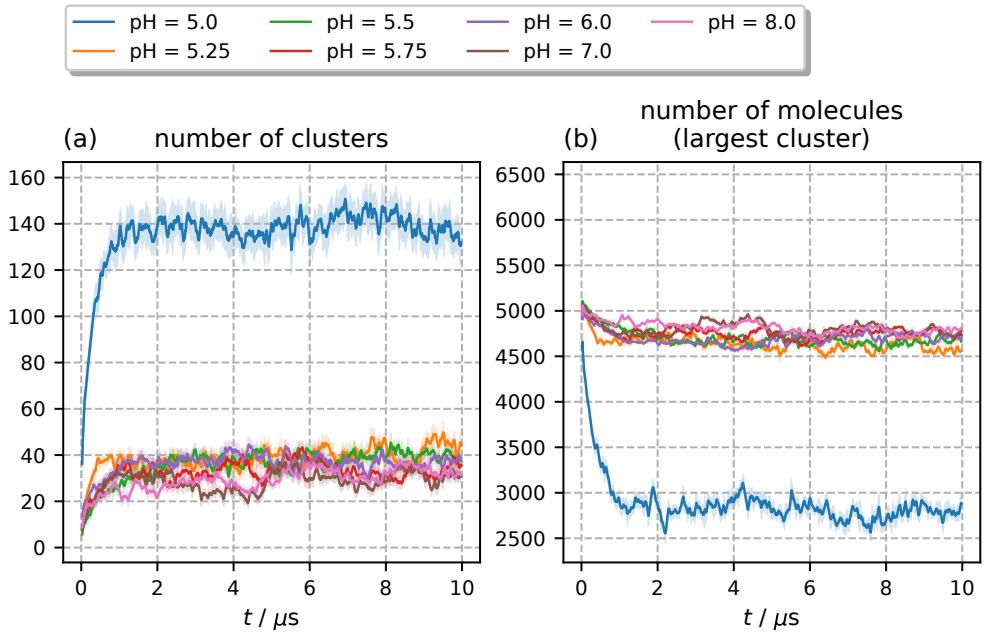


Figure S7: Trajectory-resolved number of clusters (a) and number of molecules in the largest cluster (b). The time averaged values correspond to the blue data points in Figure S6, using the last 4 µs for the averaging.

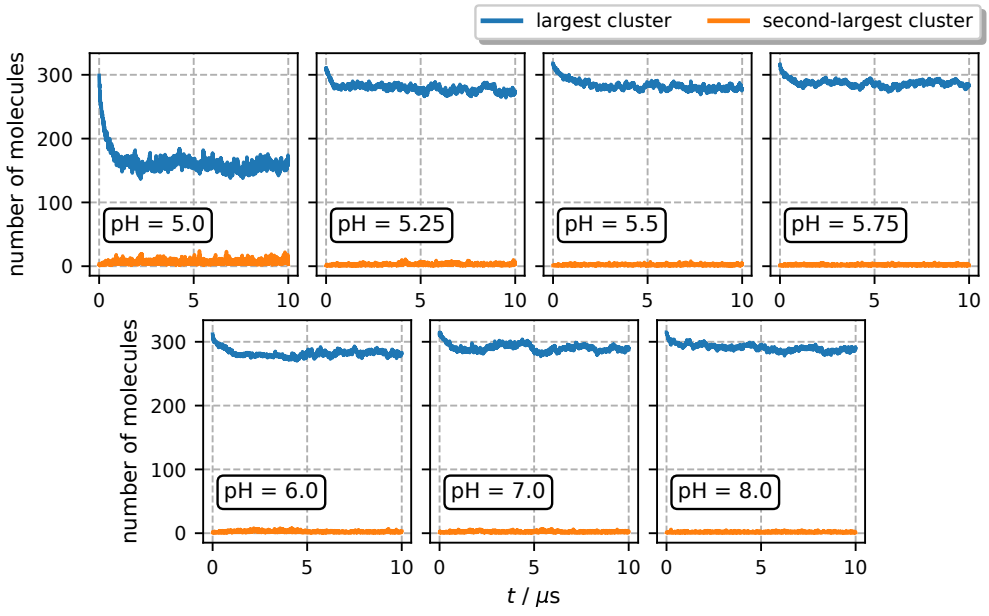


Figure S8: Number of molecules in the largest (blue) and second-largest (orange) cluster in the system, as a function of simulation time t and for the different pH values. The visibly large difference between the curves for each time in each case justifies the exclusive consideration of the largest cluster in the analyses.

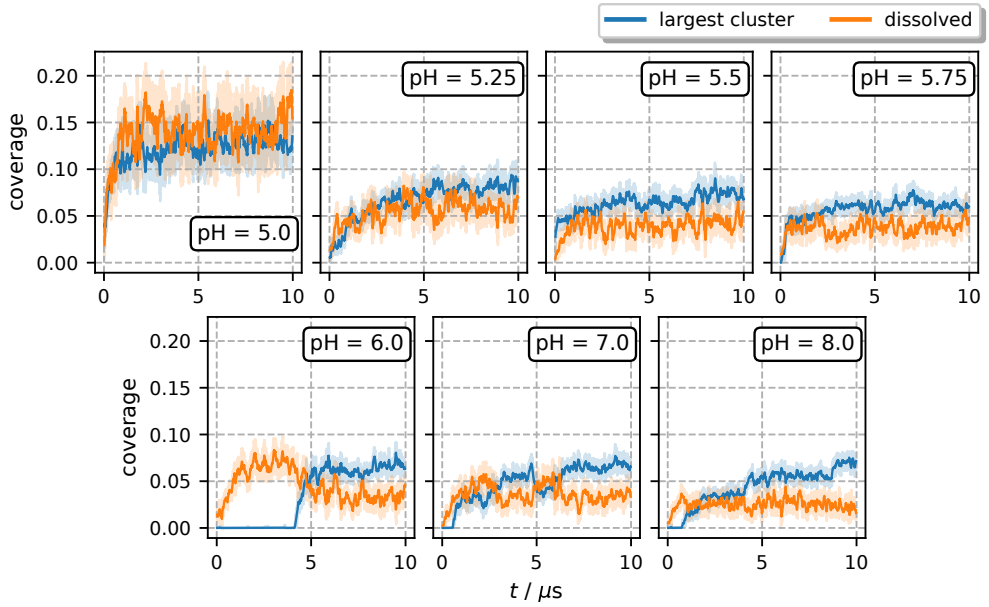


Figure S9: Trajectory-resolved coverage of the PE nanoparticle surface by HS molecules, resolved for the largest cluster (blue) and the dissolved HS molecules (orange). The averaged values using the last 4 µs are displayed in Figure 11 of the main manuscript.

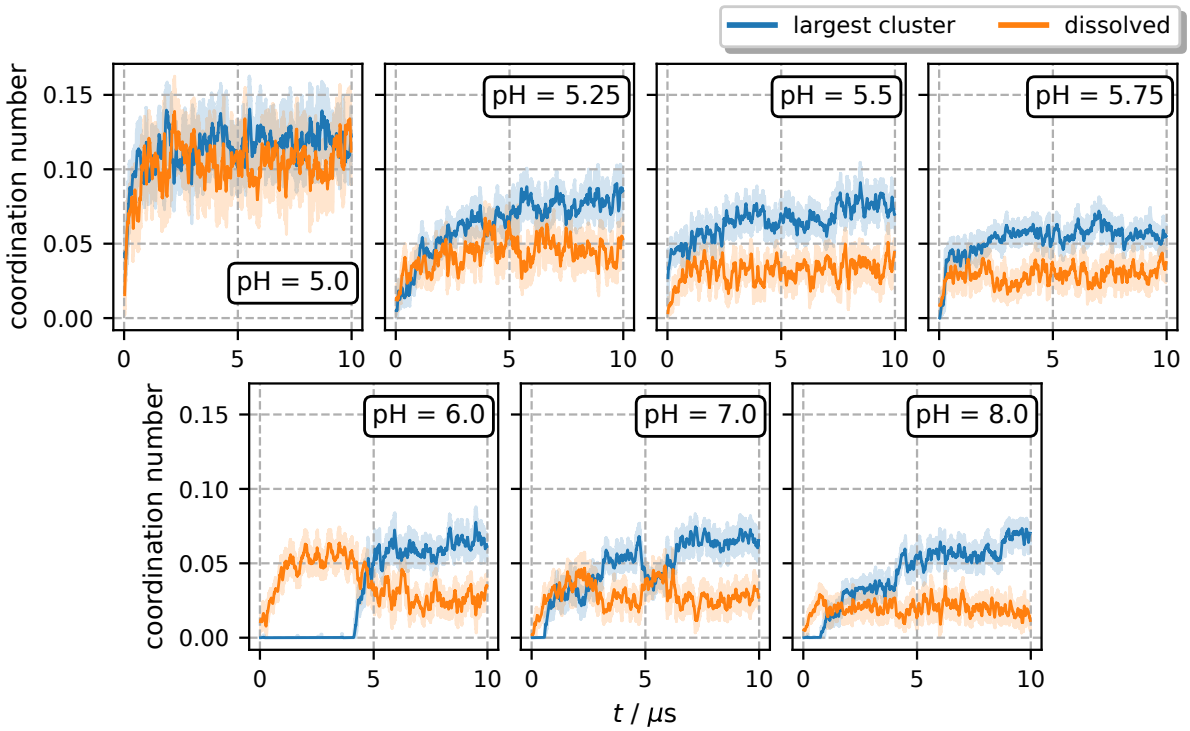


Figure S10: Trajectory-resolved coordination numbers of HS beads around the PE nanoparticle, resolved for the largest cluster (blue) and the dissolved HS molecules (orange).

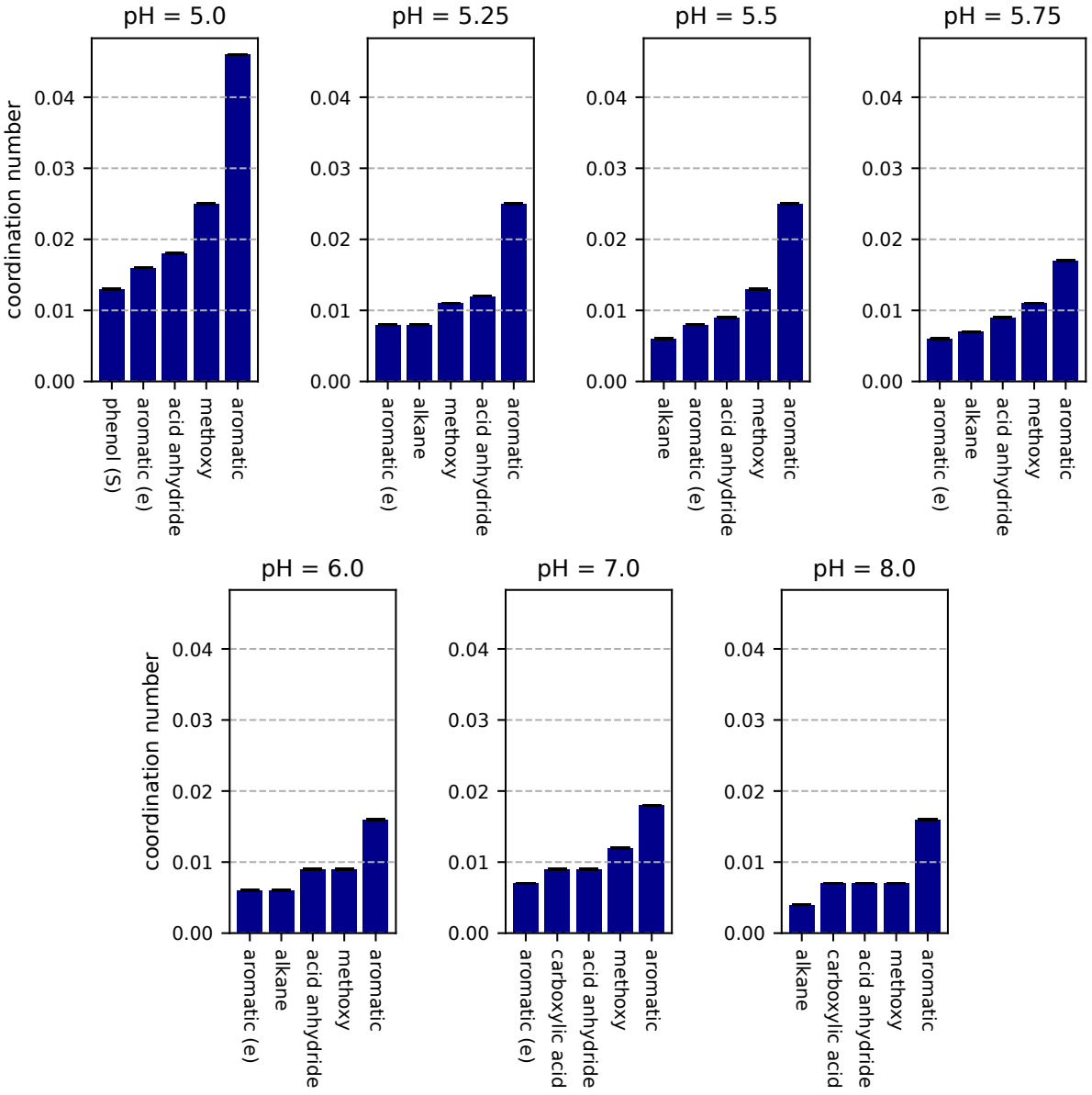


Figure S11: Coordination numbers of the five most contributing SOM functional groups around the PE nanoparticle for different pH values.

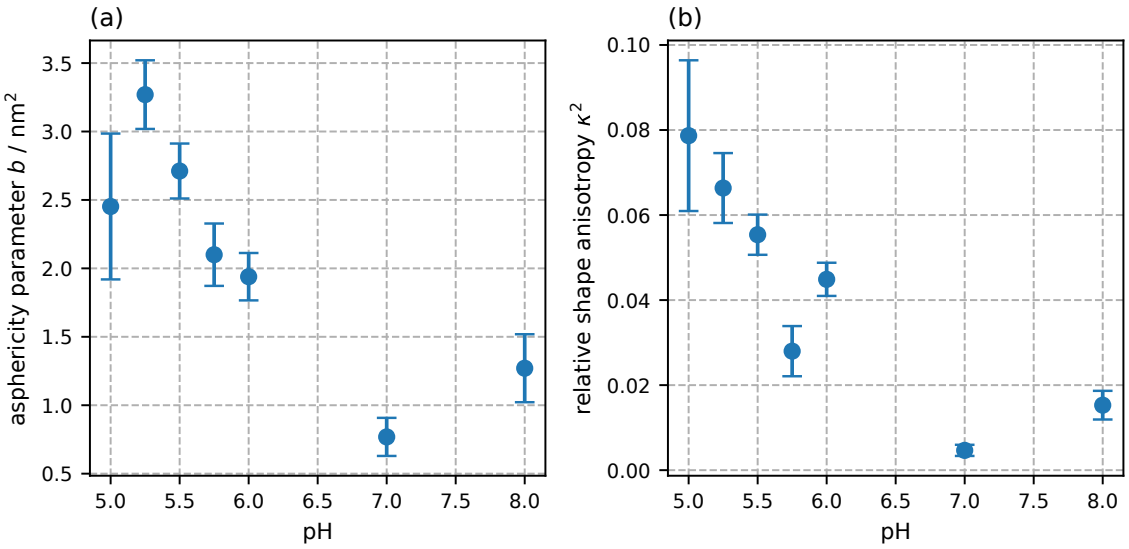


Figure S12: Sphericity indices calculated for the largest cluster at different pH values, including the asphericity parameter b (a) and the relative shape anisotropy κ^2 (b). Error bars denote standard deviations. The calculation of the sphericity indices was based on determining the eigenvalues of the gyration tensor $\bar{S}$ using the coordinates of the largest HS cluster over time. An overview of the corresponding equations can be found in the work by Arkın and Janke[3].

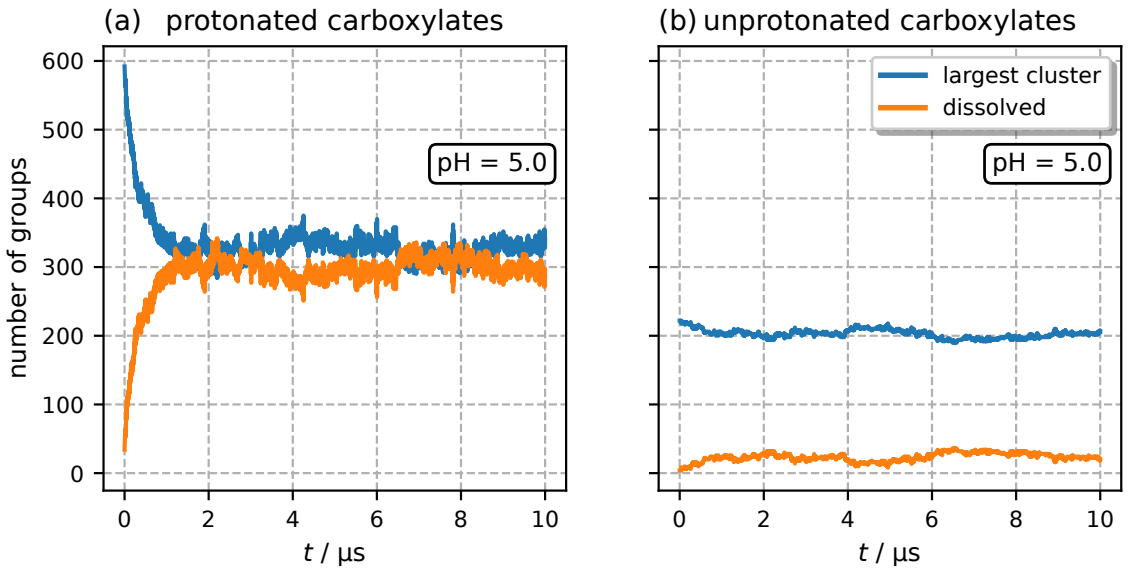


Figure S13: Trajectory-resolved number of protonated (a) and unprotonated (b) carboxylate groups belonging to the largest cluster (blue) and dissolved HS molecules (orange). The rapid increase of protonated groups in solution mirrors their behavior in the largest cluster, suggesting that HS molecules with protonated groups tend to dissolve, while those with unprotonated groups remain within the largest cluster.

References

(1) 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.

(2) Grünewald, F.; Alessandri, R.; Kroon, P. C.; Monticelli, L.; Souza, P. C.; Marrink, S. J. Polyply; a python suite for facilitating simulations of (bio-) macromolecules and nanomaterials. *Nature Communications* **2022**, *13*, 68.

(3) Arkın, H.; Janke, W. Gyration tensor based analysis of the shapes of polymer chains in an attractive spherical cage. *The Journal of Chemical Physics* **2013**, *138*.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisors, Prof. Oliver Kühn and Dr. Ashour Ahmed, for their continuous support and guidance throughout my studies. I am deeply thankful to Prof. Kühn for the comprehensive discussions and his valuable conceptual ideas, as well as for the financial support that made this work possible. My special thanks go to Dr. Ahmed, whose guidance and patience have accompanied me since my Bachelor's studies. He introduced me to the fascinating world of molecular modeling and has been an invaluable mentor ever since.

I would also like to thank all members of the Molecular Quantum Dynamics group for creating a friendly and stimulating research environment. Special thanks go to Andy for the productive collaboration on administrative and IT-related tasks, to Sadegh, Tolib, and Amine for many enjoyable discussions, and to Tobias for his helpful insights and advice on the operation of our group's computing cluster.

I am very grateful to Prof. Chris Oostenbrink and Dr. Edgar Galicia-Andrés from BOKU University for the opportunity to integrate the developed tool into the VSOMM2 website. The scientific discussions we had were highly inspiring, and it was a great honor to collaborate with them.

My heartfelt thanks go to Lisa McKee for the professional proofreading of this thesis. I truly enjoyed our communication, and her thoughtful corrections and feedback significantly improved the text.

Finally, I would like to express my deepest gratitude to my family and friends for their constant support. I am particularly grateful to my parents for their steadfast encouragement and to my sister Christine, whose empathetic and thoughtful perspective provided me with valuable guidance.