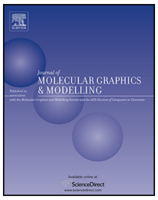
Journal of Molecular Graphics and Modelling 62 (2015) 253–261

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM



Vienna Soil-Organic-Matter Modeler—Generating condensed-phase models of humic substances

Axel Sündermann[a], Roland Solc[b], Daniel Tunega[b], Georg Haberhauer[b], Martin H. Gerzabek[b], Chris Oostenbrink[a,*]

[a] Department of Material Sciences and Process Engineering, Institute of Molecular Modeling and Simulation, University of Natural Resources and Life Sciences Vienna, Muthgasse 18, A-1190 Vienna, Austria
[b] Department of Forest- and Soil Sciences, Institute of Soil Research, University of Natural Resources and Life Sciences Vienna, Peter-Jordan-Straße 82, A-1190 Vienna, Austria

ARTICLE INFO

Article history:
Received 20 July 2015
Received in revised form 12 October 2015
Accepted 13 October 2015
Available online 19 October 2015

Keywords:
Soil-Organic-Matter (SOM)
Humic substances
Natural organic matter (NOM)
Molecular models
Molecular dynamics simulations

ABSTRACT

Humic substances are ubiquitous in the environment and have manifold functions. While their composition is well known, information on the chemical structure and three-dimensional conformation is scarce. Here we describe the Vienna Soil-Organic-Matter Modeler, which is an online tool to generate condensed phase computer models of humic substances (http://somm.boku.ac.at). Many different models can be created that reflect the diversity in composition and conformations of the constituting molecules. To exemplify the modeler, 18 different models are generated based on two experimentally determined compositions, to explicitly study the effect of varying *e.g.* the amount of water molecules in the models or the pH. Molecular dynamics simulations were performed on the models, which were subsequently analyzed in terms of structure, interactions and dynamics, linking macroscopic observables to the microscopic composition of the systems. We are convinced that this new tool opens the way for a wide range of *in silico* studies on soil organic matter.

1. Introduction

Soil Organic Matter (SOM) is a key part in the composition of soils and sediments, playing a significant role in adsorption and distribution of plant nutriments, pesticides, and pollutants, and in organic carbon stabilization [1]. However, the SOM composition varies spatially, due to geological and climatic conditions as well as plant cover and is heavily dependent on the utilization of the soil and the age of the soil layer [2–4]. Agriculture [5–8], and especially cover crop [9,10] has a significant effect on the quantity and quality of the SOM. Humic substances (HS) are the dominant constituents in SOM [11,12]. They are structurally very diverse, flexible molecules possessing multiple reactive sites and are crucial for numerous biogeochemical processes [13]. HS have been studied extensively regarding environmental, biochemical and therapeutic properties [14–22], and may also have an impact on atmospheric chemistry through the carbon cycle [11]. Challenges in the extraction and analysis complicate experimental molecular characterizations of HS [3,4]. Nuclear magnetic resonance (NMR) and infrared (IR) spectroscopy are frequently used to characterize and quantify HS in soil samples [23–26]. As complete experimental data on the chemical and three-dimensional structures of HS are lacking, molecular modeling is an important potential source of such information [27]. Little molecular modeling has been performed on these substances, mostly because of their seemingly random composition [28,29]. However, it has been suggested that HS are formed during a controlled biological degradation of a limited number of compounds or abiotic reactions of biological degradation products [20,30]. The current consensus is that HS are supramolecular structures composed of relatively small molecules with similar characteristics and varying intermolecular interactions [31–33]. Some model compounds have been proposed [13,30,34–37], attempting to represent the HS ensemble by a single average structure. More flexible approaches to simulate humic molecules, taylored to experimental data, were not described so far.

Two automated building tools have previously been described: The SIGNATURE program [38] and a self-assembly algorithm for single, branched 3D molecules based on a fragment database [39]. Both approaches have in common that a single representative structure is to be created, with which computational studies can be conducted. So far, HS were mostly studied in vacuum simula-

---

* Corresponding author. Fax: +43 1/47654 8309.
E-mail address: chris.oostenbrink@boku.ac.at (C. Oostenbrink).

http://dx.doi.org/10.1016/j.jmgm.2015.10.007

tions, or with very few water molecules attached [13,27,35,40–42]. Furthermore, quantum mechanical studies have been performed on proposed models representing typical functional groups or fragments of HS [35,43–46].

Here, we demonstrate the use of the GROMOS force field, optimized for condensed phase use, for molecular dynamics simulations of SOM. We have created an online tool, which uses a building-block database and assembles condensed-phase models representative of compositional data from experimental analyses of HS samples. Depending on the amount of available experimental data on the chemical composition, the modeler can aim to reproduce only the fraction of carbon *vs.* nitrogen in the models, or it can reproduce the experimentally determined abundances of various functional groups [3,47,48]. Condensed-phase systems are more representative of the aggregation state of HS than vacuum models [13]. The created model contains a user-defined number of HS molecules, water, and ions. The assembly of the selected building blocks is performed randomly; thus a multitude of systems can be created, which are all representative of the same experimental data. The initial building blocks were created based on previously proposed models [13,35,37] and additional building blocks can be readily added to the database. The generated models can subsequently be used to study *e.g.* thermodynamic properties at a molecular level in a multitude of ways.

2. Methods

2.1. Building blocks

The modeler uses a database of currently 34 organic fragments called building blocks, to generate the SOM models. The concept of building blocks is commonly used in biomolecular simulations, which use amino acids as building blocks for proteins or nucleotides for DNA. The SOM building blocks were derived by splitting up model compounds described in the literature [13,35,37]. A list of the building blocks and their chemical structures can be found in the Supplementary material. They vary strongly in size and chemical composition. Some include aromatic ring systems, some include nitrogen and/or sulfur atoms. They can have one or more of the following functional groups: Carbonyl-C, Carboxyl-C, O-aryl-C, Aryl-C, Di-O-Alkyl-C, O-Alkyl-C, Methoxyl-C and Alkyl-C. The parameters that are needed to describe the interactions and dynamics of the building blocks, such as atom mass, partial charge and van der Waals interaction parameters, as well as covalent parameters such as bond length, angle width, and torsional profiles were derived from the GROMOS force field, parameter set 54A7 [49]. This force field was parameterized to reproduce the free energy of solvation in water and hydrophobic solvents for a variety of functional groups that are also included in the SOM building blocks [50]. As such, it is expected to appropriately describe both hydrophobic and hydrophilic interactions in heterogeneous systems. The set of building blocks can readily be extended in future work.

2.2. Modeler

The Vienna Soil-Organic-Matter Modeler consists of two parts. The frontend web application is written in javascript and uses the AJAX technique to communicate with php scripts on the server without reloading the page. It is used to submit the input parameters for the model generation. The input is stored in a MySQL database and used by the modeler backend to generate the corresponding model, which will then be displayed in the results section of the web application. If a name and password are provided, the model can be accessed for two weeks. For the generated model, a GROMOS coordinate file as well as the corresponding topology file can be downloaded. Additionally, a pdb coordinate file of the model and a pdf file with a modeling report can be downloaded. The report shows the user-specified input in comparison to the actual properties of the generated model and gives a score on how well the modeler was able to represent the input with the available building blocks. The files can also be viewed online, together with a 3D representation of the generated model, excluding the ions and water.

The backend modeler is a php application which takes the input from the MySQL database and produces a model representing the input as well as possible. The modeling algorithm is as follows. Until the maximum amount of building blocks is reached, for each building block in the database, the difference between the input fractions and the fractions calculated for a model upon addition of the building block is determined and the building block with the lowest difference is taken. The user-specified pH is reflected by selecting the proper fraction of protonated and deprotonated carboxylic groups, assuming a uniform pK_a value of 4.7. After all building blocks have been selected, they are randomly connected to assemble the desired number of HS molecules and these molecules are subsequently energy minimized using GROMOS [51]. After the energy minimization, they are randomly placed in a cubic, periodic simulation box, avoiding overlap between the individual molecules. Additionally, the user-defined amount of simple point charge (SPC) water molecules [52] and a number of water molecules corresponding to the amount of counter ions required is placed in the simulation box. Finally, if counter ions are desired, the excess water molecules are replaced by the selected ion type to create a system with an overall neutral charge. The GROMOS coordinate file, the GROMOS topology file as well as the modeling report are then stored in the MySQL database and ready to be retrieved via the web application.

2.3. Molecular dynamics simulations

Using the Vienna Soil-Organic-Matter Modeller, 18 models were created as outlined in Table 1 and subsequently subjected to molecular dynamcs simulations. All simulations were performed using the GROMOS11 molecular simulation package [51]. The initial models represent the requested systems under periodic boundary conditions and at an artificially low density of approximately $500\,\mathrm{kg\,m^{-3}}$. An offline equilibration was performed in 18 discrete steps during which the temperature was slowly increased and the box volume reduced. In the first step, random velocities were assigned to all constituting particles, sampled from a Maxwell–Boltzmann distribution at 60 K. The temperature was subsequently increased by 30 K in following steps. At every temperature a 20 ps simulation at constant volume was followed by a 500 ps simulation at a constant pressure of 1 atm, during which the box volume could adjust. For the last two equilibration steps, the temperature was kept at 300 K and for the last step the simulation time was increased to 2 ns. Production simulations were subsequently performed for 10 ns, during which the coordinates were stored to disc every 0.5 ps. A weak coupling with a relaxation time of 0.1 ps for the temperature and 0.5 ps for the pressure was used to keep the temperature and pressure constant at 300 K and 1 atm, respectively [53]. The isothermal compressibility was estimated at $4.575 \times 10^{-4}\,(\mathrm{kJ\,mol^{-1}\,nm^{-3}})^{-1}$ [54], and pressure scaling was applied anisotropically, *i.e.* using separate scaling factors for the *x*-, *y*- and *z*-directions. The SHAKE algorithm was used to constrain the bond lengths to their optimal values with a relative geometric accuracy of 10^{-4}, allowing for a time step of 2 fs [55]. A molecular pair-list was generated using a triple-range cutoff [56]. Nonbonded interactions up to a short range of 0.8 nm were calculated at every time step from a pair-list that was updated every 5 steps. Interactions up to a long-range cutoff of 1.4 nm were calculated at pair-list