Molecular models of natural organic matter and its colloidal aggregation in aqueous solutions: Challenges and opportunities for computer simulations
Andrey G. Kalinichev

To cite this version:

HAL Id: in2p3-00769189
https://hal.in2p3.fr/in2p3-00769189
Submitted on 7 Jan 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Molecular models of natural organic matter and its colloidal aggregation in aqueous solutions: Challenges and opportunities for computer simulations*

Andrey G. Kalinichev1,2,‡

1Laboratoire SUBATECH (UMR 6457), Ecole des Mines de Nantes, 44307 Nantes Cedex 3, France; 2Departments of Chemistry, Department of Geological Sciences, Michigan State University, East Lansing, MI 48824, USA

Abstract: Natural organic matter (NOM) is ubiquitous in soil and groundwater, and its aqueous complexation with various inorganic and organic species can strongly affect the speciation, solubility, and toxicity of many elements in the environment. Despite significant geochemical, environmental, and industrial interest, the molecular-scale mechanisms of the physical and chemical processes involving NOM are not yet fully understood. Recent molecular dynamics (MD) simulations using relatively simple models of NOM fragments are used here to illustrate the challenges and opportunities for the application of computational molecular modeling techniques to the structural, dynamic, and energetic characterization of metal–NOM complexation and colloidal aggregation in aqueous solutions. The predictions from large-scale MD simulations are in good qualitative agreement with available experimental observations, but also point out the need for simulations at much larger time- and length-scales with more complex NOM models in order to fully capture the diversity of molecular processes involving NOM.

Keywords: aggregation; environmental chemistry; molecular dynamics; natural organic matter; soil; solution chemistry.

BACKGROUND AND MOTIVATION

Natural organic matter (NOM) is ubiquitous in soil, groundwater, and sedimentary environments. Being the product of natural decay and weathering of plants and other organisms [1–5], NOM has a highly heterogeneous and complex composition, so that it is normally not possible to determine its unique molecular structure, even for NOM from one locality. The reported apparent molecular weights of NOM range from a few hundred to several hundred thousand Daltons (e.g., [6]), but whether it chemically represents a true macromolecular entity or is merely a supramolecular aggregate of smaller molecular fragments held together by relatively weak noncovalent interactions was still under discussion until recently [3,4,8,9], but most recent experimental evidence led to the general acceptance of the latter view [9–13].

‡E-mail: kalinich@subatech.in2p3.fr
The uncertainties about NOM composition and structure are limiting detailed quantitative characterization of NOM by many generally applicable methods, and these molecules (also often called humic substances, HS, or soil organic matter, SOM) are usually only operationally characterized as consisting of three separate fractions: (i) fulvic acids (FA), the fraction soluble under all pH; (ii) humic acids (HA), the fraction soluble under alkaline conditions but not under acidic conditions; and (iii) humin, the insoluble NOM fraction.

The soluble components of NOM play an extremely important geochemical and environmental role by forming aqueous complexes of widely differing chemical and biological stabilities with inorganic and organic species [14,15]. Metal–NOM interactions lead to very significant correlations between the concentration of NOM and the speciation, solubility, and toxicity of many elements in the environment.

Experimental results show that the extent of metal–NOM binding varies with the size, composition, and configuration of the NOM, the pH and the ionic strength of the solution, the chemical properties of the metal, and the metal–NOM compositional ratio [14–16]. There is also substantial experimental evidence that NOM can strongly influence the migration of radionuclides in the environment [17–23], and a comprehensive understanding of its mobility in natural and engineered geochemical barriers of nuclear waste repositories is one of the critical scientific challenges to ensure their long-term stability and safety [24–26].

In the technological processes of water purification and desalination, NOM plays a highly negative role as the major foulant of ultrafiltration, nanofiltration, and reverse osmosis membranes [27–32], either directly or by forming a surface conditioning layer for microbial attachment (so-called “membrane biofouling”).

However, despite significant geochemical, environmental, and technological interest, the molecular-level mechanisms and dynamics of the physical and chemical processes involving NOM, its complexation with metals and colloidal aggregation in aqueous are not yet well understood.

Computational molecular modeling has long been recognized as a very powerful tool to study such complex phenomena in true atomistic detail at the nanoscale in both space and time [33,34]. These methods are widely used in modern chemical, biological, and materials sciences, and their application in geo-, environmental, and soil chemistry research has also progressed rapidly in recent years. Classical molecular computer simulations are typically performed for a relatively small system of $10^3 < N < 10^6$ particles (atoms, ions, and/or molecules) confined in a box with so-called “periodic boundary conditions”. Using rigorous formalism of statistical mechanics to analyze the large number of computer-generated instantaneous molecular configurations, these methods can yield many important thermodynamic, structural, spectroscopic, and transport properties of the simulated systems [34].

But the application of computational molecular modeling approaches presents particular challenges in the case of NOM, because its extraordinary molecular diversity precludes the construction of any unique atomistic model. On the one hand, NOM can only be experimentally characterized as a complex mixture of many components with varying composition and structure depending on the sample origin, extraction procedure, and other experimental conditions [3–5,8,9]. On the other hand, independent of the source of NOM, its principal functional groups are quite well characterized and understood, and the proposed NOM models have many common features.

There have been several efforts to develop molecular models of NOM based on these common characteristics. Schulten and Schnitzer used analytical pyrolysis measurements to develop a series of NOM structural models [33,36,37]. Such models were later used in MD simulations of NOM interaction with hydrated Na$^+$ and Ca$^{2+}$ ions [38]. The MD of water interaction with FA [39] and lignin [40] has also been simulated. Shevchenko and Bailey [41,42] have modeled the NOM sorption on soil mineral particles using a NOM model based on an oxidized lignin-carbohydrate complex. Leenheer et al. [43] have proposed a molecular model of Suwannee River FA and used it for the interpretation of experimental data on metal–NOM binding. Kubicki and Apitz [34] used this model to compare the structures computed by classical molecular mechanics with quantum mechanical calculations and to test the effect...
of computational methodology on the predicted structure. This model was also later used in the molecular simulation of hydrogen bonding and clustering of neutral FA fragments in aqueous solution [39].

An algorithm of computer-assisted structure elucidation (CASE) was applied to a comprehensive set of spectroscopic and analytical data for Chelsea soil HA [44], resulting in the development of a series of representative 3D structural models for these substances. The Temple–North-eastern–Birmingham (TNB) model of a NOM building block [45] was successfully used in a study of NOM conformations in solution [46]. We have also recently used this model for the molecular simulations of metal–NOM complexation in bulk aqueous solutions [47–49] and for the molecular modeling of the effects of different cations on the NOM adsorption at the surfaces of polyethersulfone ultrafiltration membranes [32]. The results of these molecular simulations are in good qualitative agreement with available experimental observations of metal–NOM complexation and colloidal aggregation in aqueous solutions. In particular, they confirm that the cation binding occurs principally with the carboxylic groups of NOM, and to a much lesser extent with phenolic and other –R–OH groups. The contributions of other NOM functional groups appear to be minimal.

In a recent series of molecular simulations, Aquino et al. [52–56] have used several simplified molecular models to study the role of hydrogen bonding and cation bridges in stabilizing the NOM aggregates in aqueous solution. Polyacrylic acid (PAA), oligomers, conjugated olefinic chain with regularly distanced aliphatic tails terminated by a polar carboxyl group, and a tetramer of undecanoid fatty acids were successfully used to quantify many structural and thermodynamic parameters by quantum mechanical and classical simulations. Such a molecular approach finds support in experimental studies that indicate that the complexation behavior of NOM is often nearly indistinguishable experimentally from the similar behavior of the structurally much better defined molecules of PAA [17,57,58]. On the other hand, the experimental study of Eu³⁺ complex formation with HA and PAA revealed noticeable differences in the complexation behavior, strongly suggesting that the heterogeneous coexistence of the strong and weak binding sites in the actual HA may play a significant role [18].

However, taken together, recent computational molecular modeling efforts [13,38,48–50,52–56] clearly indicate that despite the significant uncertainties inherent to the molecular modeling of NOM, the results of various simulations are quite robust in terms of their qualitative and even quantitative reproduction of the most significant structural, dynamic, and energetic characteristics of metal–NOM complexation, hydration, and supramolecular aggregation.

This short paper is not intended to be a comprehensive review on the subject. Instead, it presents a quantitative comparison of results between several recent molecular simulations which used different models of intermolecular interactions, but the same TNB model of NOM [46] illustrates one important point that notwithstanding all the differences and uncertainties in the detailed atomistic characterization of NOM, such calculations can still be very useful and informative by offering additional opportunities for better molecular-scale understanding of the NOM-containing systems and processes.

Molecular dynamics simulations of aqueous metal–NOM systems

The TNB model of a NOM molecular fragment used in all three simulations discussed below has a molecular weight of 753 Da and contains three carboxylic groups, three carbonyl groups, two phenolic groups, two amine groups, and four other –OH alcohol groups [45,46]. The size, composition, and molecular structure of the TNB model of NOM fragment is illustrated in Fig. 1. In terms of the molecular weight, atomic composition, degree of aromaticity, and total charge density the TNB model fragment is in good agreement with available experimental characterizations of NOM [3,8,10,58], the results of stochastic modeling of 3D structure elucidation [44], and biogeochemical reconstructions [59]. The atomic composition of the TNB model (Table 1) also closely resembles the composition of Suwannee River NOM (SRNOM), which is often used in experiments as a typical representative of NOM (e.g., [32,47,58]).

© 2012, IUPAC Pure Appl. Chem., Vol. xx, No. x, pp. 0000–0000, yyyy
Table 1  TNB model and the composition of the composition of SRNOM.

<table>
<thead>
<tr>
<th></th>
<th>C wt %</th>
<th>H wt %</th>
<th>O wt %</th>
<th>N wt %</th>
<th>S wt %</th>
<th>P wt %</th>
<th>Carboxyl mol/kgC</th>
</tr>
</thead>
<tbody>
<tr>
<td>TNB [46]</td>
<td>57.4</td>
<td>4.9</td>
<td>34.0</td>
<td>3.7</td>
<td>–</td>
<td>–</td>
<td>7.0</td>
</tr>
<tr>
<td>Exp. (SRNOM) [16]</td>
<td>52.5</td>
<td>4.2</td>
<td>42.7</td>
<td>1.1</td>
<td>0.6</td>
<td>0.02</td>
<td>9.85</td>
</tr>
</tbody>
</table>

For the near-neutral pH conditions, it is reasonable to assume that the three carboxylic groups of the NOM fragments are completely deprotonated (pKₐ values between 4 and 5), whereas the hydroxyl groups are always protonated (pKₐ values of ~9). The deprotonated carboxylic groups of NOM are known to be the principal source of the NOM negative charge development at the near-neutral pH range [16], the most important binding sites for metal cations [15,60]. Ca²⁺ appears to be among the most strongly NOM-associating ions [17,19,30,43,61,62].

Three separate molecular dynamics (MD) simulations were performed for the same TNB model in aqueous solutions containing Na⁺, Mg²⁺, or Ca²⁺ ions at approximately the same concentrations, but using different system sizes, different force fields, and two different water models [48–51]. In the first MD simulation (A), a single TNB molecule was hydrated by 553 water molecules in a cubic box with standard periodic boundary conditions [35]. The negative charge of the NOM fragment was balanced by the presence of three Na⁺ ions (or two Mg²⁺/Ca²⁺ and one Cl⁻). The interatomic interactions among H₂O, dissolved ions, and NOM were described by the simple point charge (SPC) water model [63], SPC-compatible parameters for the ions [64], and the consistent valency force field (CVFF) for the NOM fragment [65]. For the MD simulations (B) and (C), the simulation cells were almost an order of magnitude larger.
magnitude larger and contained 8 TNB fragments and 24 Na\(^+\) (12 Mg\(^{2+}\)/Ca\(^{2+}\)) cations hydrated by more than 4000 H\(_2\)O molecules [50]. In the simulation (B), the CHARMM27 force field [66] was used to describe all interatomic interactions in the system, whereas in the simulation (C) the AMBER FF99 force field [67] was employed. In both of these larger simulations H\(_2\)O molecules were described by the TIP3P model [68].

In all three simulations, the Newtonian equations of atomic motions were numerically integrated with a time step of 1 fs and Ewald summation [35] was applied to calculate long-range electrostatic contributions to the total intermolecular potential energy of the simulated systems. A sufficient thermodynamic equilibration and initial NOM conformational relaxation was assured by performing pre-equilibration MD runs for each of the simulated system in several stages with the total simulation length of several hundreds of picoseconds. This pre-equilibration process resulted in simulated solution densities that corresponded well to an ambient pressure of 0.1 MPa. These optimized models were then used as the starting configurations for the production MD runs (A), (B), and (C) that were all performed in the NVT statistical ensemble at a constant temperature of 300 K and at constant volume.

The equilibrium duration of the simulation (A) was 100 ps, whereas the simulations for the larger systems (B) and (C) were nearly 100 times longer (10 ns each). The longer time scale of these runs allowed us not only to determine the structural characteristics of cation-NOM complexation at individual binding sites, but also to quantitatively estimate the degree of supramolecular aggregation of NOM fragments in solution due to their strong electrostatic interactions with metal ions.

To quantitatively assess the structural and dynamic effects of Ca\(^{2+}\)-NOM complexation, we calculated the radial distribution functions (RDFs) and running coordination numbers using standard procedures [35]. The running coordination numbers of species \(j\) around species \(i\) in the solution, \(n_{ij}(r)\), are calculated from the RDFs as

\[
n_{ij}(r) = 4\pi \rho_j \int_0^r g_{ij}(r) r^2 dr
\]

where \(\rho_j\) is the number density of species \(j\) in the system, \(g_{ij}(r)\) are the atom–atom RDFs.

Potentials of mean force (PMFs) for the interaction between the cations and the carboxylic groups of the NOM were also calculated. These functions characterize the change in the free energy of the system due to the changes in its configuration [69]. If the free energy of a system in thermodynamic equilibrium in the NVT statistical ensemble is

\[
F = -k_B T \ln Z
\]

where \(Z\) is the canonical partition function and \(k_B\) is the Boltzmann constant, then the potential of mean force, \(W_{ij}(r)\), for two interacting species is defined as the potential that would generate the mean force between the two species, averaged over all orientations, for each separation distance \(r\). Thus, the PMF represents the free energy profile of the system as a function of distance, and it can be shown [70] that

\[
W_{ij}(r) = -k_B T \ln g_{ij}(r)
\]

where \(g_{ij}(r)\) is the corresponding RDF for this pair of species with the standard normalization for large separations, \(g_{ij}(r) \rightarrow 1\) at \(r \rightarrow \infty\). Thus, \(W_{ij}(r)\) asymptotically approaches zero with large separation distance. The PMF calculations for the metal cation complexation with the carboxylic groups of NOM and several other molecules and the resulting estimates for the metal–NOM association constants are discussed in detail elsewhere [49].

RESULTS AND DISCUSSION

The TNB molecular model of NOM contains three structurally distinct carboxylic groups (C23, C25, and C29 in Fig. 1). Metal cations can associate with these carboxylic groups via several typical and rel-
atively stable coordination geometries [49]. Two contact ion pair (CIP) coordinations are possible. In a bidentate inner-sphere coordination (C29 in Fig. 1), the metal ion is coordinated simultaneously with the two oxygen atoms of the carboxylic group and stays approximately equidistant from both of them predominately near the bisector plane orthogonal to the plane of the carboxylic group. In a monodentate inner-sphere coordination (C23 and C25 in Fig. 1), the metal ion is coordinated with only one of the carboxylic oxygen atoms. Thus, in both bidentate and monodentate CIP cases, the water molecules in the first coordination shell of the ions are partially replaced by the carboxylic oxygens. In contrast, in an outer-sphere coordination, the metal ions are separated from the NOM carboxylic oxygens by a mono-molecular layer of H$_2$O, as fully hydrated cations only weakly associated with the carboxylic group and forming a solvent-separated ion pair (SSIP).

A comparison of PMFs calculated for simulations (A), (B), and (C) according to eq. 3 for Ca$^{2+}$ interaction with the carboxylic groups of NOM is presented in Fig. 2. All three simulations consistently result in a potential well of about $-13.0 \pm 2.0$ J/mol deep for the bidentate CIP coordination with an interatomic separation of about 0.3 nm and in a shallower and broader potential well of $-3.0 \pm 0.8$ kcal/mol at about 0.5 nm corresponding to an SSIP coordination (not visible for the simulation (A) because of its smaller size and shorter duration). However, the potential barrier between the CIP and SSIP coordinations is higher by almost 1 kcal/mol for (B) than for (C), indicating a stronger preference for inner- vs. outer-sphere coordination for the latter model. Consistent with the structural results [50], such model dependence between (B) and (C) is even more visible in the energy of monodentate-coordinated CIPs. For the former, the corresponding potential well (dashed line in Fig. 2 at about 0.35 nm) is located in the region of positive energies, indicating that such coordinations, although relatively stable, are generally unfavorable [49].

A detailed analysis of Ca$^{2+}$ dynamics around the carboxylic groups of NOM shows that an average residence time of a bidentate CIP is about 0.5 ns, compared with only about 0.1 ns for a monodentate CIP coordination, and about 0.2 ns for an SSIP coordination [49]. These values provide useful estimates for the time scale for the processes of Ca-NOM association in aqueous solutions.

Ca$^{2+}$ is known to cause a particularly strong supramolecular aggregation of relatively small NOM fragments into larger colloidal particles [17,19,30,43,61,62,71]. The simulations (B) and (C) were already large enough to at least qualitatively probe the molecular mechanisms responsible for this behavior [49–51] during the 10 ns of equilibrium dynamical evolution. A snapshot from the simulation (C) is presented in Fig. 3, to illustrate the aggregation process. Three different aggregating NOM fragments are shown in different colors, Ca$^{2+}$ ions are shown as the light blue balls, while all water molecules are removed for clarity. One Ca$^{2+}$ ion (Ca4) is found in a strong inner-sphere bidentate coordination with the carboxylic groups of two NOM molecules (1C25 and 2C29 in Fig. 3) and, simultaneously, in a weaker monodentate or outer-sphere coordination to a third NOM molecule (6C23 in Fig. 3). A simultaneous coordination of Ca$^{2+}$ ion by two carboxylic groups of the same NOM molecule was also observed, confirming an earlier hypothetical view of Ca$^{2+}$ capable of accepting up to four NOM carboxylic groups in its inner-sphere coordination shell [43]. This observation also suggests two different mechanisms by which Ca$^{2+}$ may affect the supramolecular colloidal aggregation of NOM: Ca$^{2+}$ ions can directly affect aggregation by bridging carboxylic groups of different NOM molecules, effectively bringing and holding them together. In addition, simultaneous Ca$^{2+}$ coordination with two carboxylic groups of the same NOM molecule effectively reduces the net negative charge of this metal–NOM complex, thus allowing such complexes to approach each other in aqueous solutions more readily and, consequently, to coordinate with each other via weaker hydrogen-bonding interactions.

Hydrogen bonding also plays a major role in a weaker spontaneous aggregation of NOM when only Na$^+$ and Mg$^{2+}$ are present in solution as background cations [51]. This promotes the formation of loosely held spherically shaped NOM clusters in contrast to the effect of Ca$^{2+}$, where stronger cation bridging between carboxylic groups of different NOM fragments favors the formation of longer linear and branched aggregate structures. This picture appears to be qualitatively compatible with the results.
of dynamic light scattering (DLS) and small angle neutron scattering (SANS) experiments [51,72], which point to the formation of a wide range of supramolecular structures with sizes up to 100 and 1000 nm when Ca\(^{2+}\) ions are present in an SRNOM solution, in contrast to Na\(^{+}\) and Mg\(^{2+}\), which do not affect the aggregation of SRNOM as strongly.

**Fig. 2** PMFs for carboxyl carbon – Ca\(^{2+}\) IPs calculated from the three MD simulations: (A) black dash-dotted line; (B) red dashed line; (C) blue solid line.

**Fig. 3** A snapshot from the MD simulation (C) illustrating the NOM aggregation in the presence of Ca\(^{2+}\) ions. All water molecules and other ions are removed for clarity. Different NOM fragments are shown in different colors, whereas calcium ions are represented by blue spheres (not in scale). One Ca\(^{2+}\) (marked Ca4) is simultaneously coordinated by two carboxylic groups in bidentate configurations and by one carboxylic group in a monodentate configuration, whereas another Ca\(^{2+}\) (Ca2) is seen in a distorted bidentate inner-sphere coordination to yet another carboxylic group.
Comparison of the simulations discussed above makes it perfectly clear that the size and the duration of the MD simulations should be increased at least another order of magnitude before statistically meaningful quantitative estimates for the relative importance of different aggregation mechanisms and the relative effects of different ions could be provided.

CONCLUSIONS AND OUTLOOK

NOM is a good example of a complex chemical substance that cannot be fully characterized in full atomistic compositional and structural detail for completely natural reasons [73]. This makes it particularly challenging to approach a quantitative investigation of NOM behavior with traditional molecular modeling techniques. Nevertheless, despite the significant inherent uncertainties, the results of NOM molecular modeling are all quite robust in terms of their qualitative and even quantitative reproduction of the most significant structural, dynamic, and energetic characteristics of metal–NOM complexation, hydration, and supramolecular aggregation. This includes bidentate vs. monodentate configuration of the complexes, inner-sphere (CIP) vs. outer-sphere (SSIP) coordination of the IPs and their relative stability. The modeling results presented here shows some dependence on the size of the simulated system and on the parameters of the force field models used, but are yet consistent with available experimental data and other results of molecular modeling. However, these MD simulation results also clearly point out further opportunities for better quantitative molecular-scale understanding of metal–NOM complexation, colloidal aggregation, and mobility in aqueous environments, which would require, on the one hand, the development and application of more complex and diverse NOM models, and, on the other hand, the extension of such computer simulations to much larger time- and length-scales.

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences (Grant No. DE-FG02-08ER-15929) and by the industrial chair “Storage and Management of Nuclear Waste” at the Ecole des Mines de Nantes, funded by ANDRA, Areva, and EDF. The supercomputing resources of the NSF TeraGrid (Grant No. TG-EAR0000002) and of the U.S. DOE National Energy Research Scientific Computing Center (NERSC) were used for the simulations.

REFERENCES


© 2012, IUPAC