ELSEVIER

A stochastic model for the synthesis and degradation of natural organic matter. Part III: Modeling Cu(II) complexation

Stephen E. Cabaniss [a,*], Patricia A. Maurice [b], Greg Madey [c]

[a] *Department of Chemistry, University of New Mexico, Albuquerque, NM 87131, USA*
[b] *Department of Geology and Civil Engineering, University of Notre Dame, USA*
[c] *Department of Computer Science, University of Notre Dame, USA*

Abstract

An agent-based biogeochemical model has been developed which begins with biochemical precursor molecules and simulates the transformation and degradation of natural organic matter (NOM). This manuscript presents an empirical quantitative structure activity relationship (QSAR) which uses the numbers of ligand groups, charge density and heteroatom density of a molecule to estimate Cu-binding affinity (K'_{Cu}) at pH 7.0 and ionic strength 0.10 for the molecules in this model. Calibration of this QSAR on a set of 41 model compounds gives a root mean square error of 0.88 log units and $r^2 = 0.93$. Two simulated NOM assemblages, one beginning with small molecules (tannins, terpenoids, flavonoids) and one with biopolymers (protein, lignin), give markedly different distributions of log K'_{Cu}. However, calculations based on these log K'_{Cu} distributions agree qualitatively with published experimental Cu(II) titration data from river and lake NOM samples.
© 2007 Elsevier Ltd. All rights reserved.

1. Introduction

Natural organic matter (NOM) is a naturally-occurring assemblage of molecules derived from plants and animals, often modified by environmental processes so that the original natural products synthesized by those organisms are barely recognizable. The key role played by NOM in environmental and biogeochemical process has motivated numerous studies, but unraveling the complex structures within the NOM mixture remains a challenging analytical problem (Schmitt-Koplin et al., 1998; Cook et al., 2003 Hatcher et al., 2001).

One of the principal ecological and biogeochemical roles of NOM is to influence the bioavailability and transport of metals through complexation. NOM complexation affects bioavailability of transition metals like Cu(II), Pb(II) and Fe(III), as well as Al(III), and alters the transport of radionuclide metals like Co, U, etc. (DiToro et al., 2001; Westall et al., 1995). Copper complexation has been studied more extensively than complexation of other metals, partly because its role in Cu toxicity to algae was discovered early (Sunda and Guillard, 1976) and partly because a variety of relatively sensitive and inexpensive electrochemical methods (ion-selective electrode, stripping voltammetry) are available for the detection of free Cu(II) in the presence of Cu(II) complexes (Cabaniss and Shuman, 1988a; Xue and

---
* Corresponding author.
E-mail address: cabaniss@unm.edu (S.E. Cabaniss).

Sigg, 1993; Bruland et al., 2000; Kogut and Voelker, 2001). Because of the substantial number of studies, NOM complexation of Cu is probably better known and quantified than NOM complexation of any other metal.

The reaction for Cu(II) complexation by a deprotonated ligand molecule L^{n-} can be written most simply and commonly as a 1:1 complex formation

$$Cu^{2+} + L^{n-} \rightleftharpoons CuL^{(2-n)} \qquad K_{Cu} = \frac{[CuL^{(2-n)}]}{[Cu^{2+}][L^{n-}]} \tag{1}$$
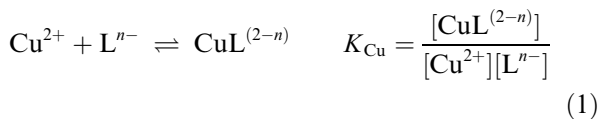
Frequently the complexation constants are determined as apparent constants at a given pH and ionic strength, denoted K'_{Cu}. Reported values for this constant vary from $\sim 10^{14}$ (Xue and Sigg, 1993) down to $\sim 10^5$ (Bhat et al., 1981). Although this simplistic model can fit data from single titrations well, it cannot extrapolate to solution pH and ionic strength which differ from the original experiment.

Sophisticated thermodynamic models of Cu-NOM complexation have been developed which consider ligand protonation, binding site heterogeneity and electrostatic effects (Tipping, 2002; Koopal et al., 2005). These models have been used successfully to predict pCu ($-\log[Cu^{2+}]$) and complexed Cu(II) at variable pH and ionic strength. However, these models require calibration with Cu-NOM binding data and have only limited ability to predict a priori how Cu binding will vary among NOM samples and how Cu binding will change in response to processes affecting the NOM (e.g., photolysis, sorptive fractionation, microbial degradation). A mechanistic kinetic model of NOM with the ability to predict pCu would be useful to ecologists and geochemists interested in Cu binding in a dynamic environment.

AlphaStep is an agent-based model which simulates the transformation and degradation of NOM using a stochastic kinetic algorithm (Xiang et al., 2004; Huang et al., 2005). As discussed in Parts I and II (Cabaniss et al., 2005; Cabaniss et al., submitted for publication), this approach treats NOM as a set of interacting molecules derived from given precursors – tannins, lignin, terpenoids, proteins, etc. Thus, a simulation which begins with only 2 or 3 different structures potentially can generate thousands of different structures during several months of reactions. Structural properties like

molecular weight can be calculated exactly for each molecule, but if the model is to be used to predict thermodynamic behavior of NOM, then it must have some mechanism for estimating thermodynamic properties from structural information.

The computationally simplest approach to predicting K'_{Cu} from molecular structure data is to establish a quantitative structure–activity relationship (QSAR). A typical QSAR calculates a thermodynamic quantity like ΔH^0, ΔG^0 or $\log K$ as the weighted sum of a set of structural and/or energetic terms (Carey and Sundberg, 2000; Schwarzenbach et al., 2003). This approach has been used to predict octanol–water partitioning and acidity constants using structural information available in AlphaStep (Cabaniss et al., submitted for publication). In the case of Cu(II) binding, the equation is of the form

$$\log K'_{Cu} = a_0 + \sum_{i=1}^{N} a_i x_i \tag{2}$$

where x_i are the N predictor variables (numbers of a particular structural fragment, for example), a_i are the N coefficients for these variables, and a_0 is the intercept. Devising a robust and useful QSAR requires first selecting a calibration data set with variability in $\log K'_{Cu}$ and molecular structure similar to the group of 'unknown' compounds and then determining appropriate values for a_i.

This paper presents a QSAR for predicting K'_{Cu} and shows how it may be used in to predict Cu(II) complexation by simulated NOM in AlphaStep both as a function of time and under differing reaction conditions. Model predictions are qualitatively consistent with published experimental results both for terrestrial and aquatic NOM.

2. QSAR calibration

Writing the reaction for Cu(II) complexation by a ligand L^{n-} as in Eq. (1) above obscures several features of the binding process:

a. For larger molecules, more than one metal ion could conceivably bind to the ligand. This is well documented for proteins and polyelectrolytes like polyacrylic acid and polymaleic acid (Tanford, 1961).
b. The magnitude of K'_{Cu} is partly a function of electrostatic attraction, and thus varies with the salt content or ionic strength of the solution. This variation is usually less than threefold for small

molecules, but can be several orders of magnitude for larger, more highly charged molecules (Bartschat et al., 1992).

c. Because metal-binding groups are also Bronsted–Lowry bases (H^+ binding groups), the value of K'_{Cu} is a function of the ligand protonation, which in turn is a function of pH. Differences in pH can change the measured K'_{Cu} by many orders of magnitude (Harris, 1999).

In order to obtain a single K'_{Cu} for each molecule, 3 assumptions are made a′–c′:

a′. Assume only 1:1 complexation. This is environmentally reasonable since in general Cu levels are low in natural systems- typically $<10^{-6}$ M in freshwaters, for example. Furthermore, only the strongest binding sites are experimentally measurable, and binding becomes weaker as more Cu(II) ions are complexed to a single ligand. However, this assumption will lead to underestimation of Cu(II) binding at the elevated Cu concentrations attainable in some laboratory experiments.

b′. Assume ionic strength of 0.10. While the traditional 'reference state' for binding constants is infinite dilution (zero ionic strength), this condition is experimentally unobtainable. Most available binding constants for small ligands were obtained at higher ionic strength, typically 0.10. Much of the data available on Cu(II) binding to NOM was also collected at ionic strength 0.10 (e.g., Cabaniss and Shuman, 1988a; Brown et al., 1999). An activity correction will be required to apply the calculated K'_{Cu} values at other ionic strengths.

c′. Assume pH 7.0. This is convenient because many natural waters are nearly neutral, and thus many studies of Cu(II) complexation by NOM have used an experimental pH near 7. Reported K'_{Cu} values of standard compounds (typically given for the completely deprotonated form as K^0_{Cu}) can be corrected for pH using the α formalism.

$$K'_{Cu} = \alpha K^0_{Cu} \qquad (3)$$

where K'_{Cu} is the apparent binding constant at a given pH and α is the fraction of non-complexed ligand which is present in the fully deprotonated form at that pH (Harris, 1999). Although most carboxylic acids are deprotonated at pH 7.0, phenols and aliphatic amines are not, so that α

can be $\ll 1$ at pH 7.0 for molecules containing these ligand groups.

QSAR calibration proceeded in 3 steps – the selection of a data set of standard molecules, the selection of candidate predictor variables, and the optimization of the a_i values for those variables and that data set.

A calibration data set of 41 ligand molecules containing various combinations of carboxylic acids, phenols, alcohols, ethers, amides and amines was selected, ranging from very weak (2-methoxy benzoic acid, $\log K'_{Cu} = 1.60$) to very strong (ethylenediamine tetraacetic acid, EDTA, $\log K'_{Cu} = 16.4$) Cu binding. Binding constants for the 1:1 Cu(II)-ligand complex at ionic strength 0.10 were obtained from the NIST critical stability constants database (Smith et al., 1998). Most constants were determined at 25 °C, although a few were determined at 20 °C. All K'_{Cu} values discussed below were corrected for pH 7.0 using Eq. (3) and pK_a values obtained from the same database, and these values are listed in the Appendix.

Two types of candidate predictor variables were considered. The first type was the number of each functional group on a molecule which could potentially act as ligand groups: carboxylic acids, alcohols, phenols, ethers, amines and amides. The second type was a set of 'whole molecule' variables: molecular weight (MW), molecular charge (Z), charge density (ZD = Z/MW) and heteroatom density (HD = (#O + #N) divided by #C).

The procedure for obtaining a useful QSAR had two criteria: (1) to obtain a 'good' fit to the data set and (2) to exclude candidate predictor variables which made minimal contribution to the goodness of fit. A 'good' fit was indicated by a small root mean square error (RMSE) and high correlation coefficient (r^2), while a 'minimal' contribution was defined as a coefficient a_i which had an uncertainty larger than its absolute value. Eq. (2) was fitted to the calibration data set using multiple linear regression with all candidate variables, and the least significant variable (highest ratio of uncertainty (s_{a_i}) to absolute value for a_i) was discarded. This procedure was repeated until all remaining variables had uncertainties s_{a_i} less than the absolute value of a_i.

Four parameters were eliminated by this iterative procedure – the numbers of alcohol and ether groups, the molecular weight and the molecular charge. On the one hand, the former two groups are both known to act as Cu(II) ligands in certain

compounds, and citric acid (3-hydroxy pentane 1,3,5-tricarboxylic acid) does bind Cu(II) more strongly than tricarballylic acid (pentane 1,3,5-tricarboxylic acid); on the other hand, it seems unlikely that the O atoms in these groups have significantly more electron density to donate than do the O atoms in water molecules, and the difference between citric acid and tricarballylic acid $\log K'_{Cu}$ values is only 0.21 log units. Although both MW and total charge have been plausibly suggested to affect metal complexation through more complete complexation and electrostatic attraction, the use of ligand counts and charge density ZD represents these same effects in the final QSAR.

The final predictive QSAR uses the parameters shown in Table 1, and has an r^2 of 0.93, standard error of 0.96 log units and RMSE of 0.88 log units for the calibration data set. Residual errors are approximately randomly distributed, with a maximum absolute value of 1.88 log units (Fig. 1). The RMSE and standard error are larger than those for carboxyl pK_a or $\log K_{ow}$ prediction (Cabaniss et al., submitted for publication), which

was expected both because errors in the experimental values are larger for $\log K'_{Cu}$ and because the $\log K'_{Cu}$ data set covers a range of >14 log units, compared to ranges of 5–6 log units for the pK_a and $\log K_{ow}$ data sets (Cabaniss et al., submitted for publication).

Most values of the coefficients a_i are chemically reasonable, at least in a relative sense. The intercept a_0 is zero, which is sensible in that molecules with no ligand groups, charge or heteroatoms should not bind Cu(II) ($\log K'_{Cu}$ is 0 and K'_{Cu} is 1, corresponding to binding too weak to measure at pH 7). ZD has a large, negative coefficient (-78) which leads to stronger binding as the molecular charge density becomes more negative. For example, a molecule with MW of 1000 amu and a charge of -5 at pH 7.0 would have an 'electrostatic contribution' to $\log K'_{Cu}$ of $+0.39$ log units, while a smaller molecule (MW 500) with the same charge would have an 'electrostatic contribution' of $+0.78$ log units. HD has a small negative coefficient, which means that as the proportion of O and N functional groups increases, $\log K'_{Cu}$ will decrease. This seems counter-intuitive, and appears to arise as a counterbalance to overestimation of K'_{Cu} from summing several ligand group contributions. Since s_{HD} is nearly as large as a_{HD}, this coefficient may not be physically meaningful.

Among the ligand group counts, the amine coefficient is the largest and the amide coefficient the smallest, while phenols make a larger contribution than carboxylic acids. These values are sensible in light of the trends observed for model compounds.

Table 1
QSAR parameters for $\log K'_{Cu}$ prediction

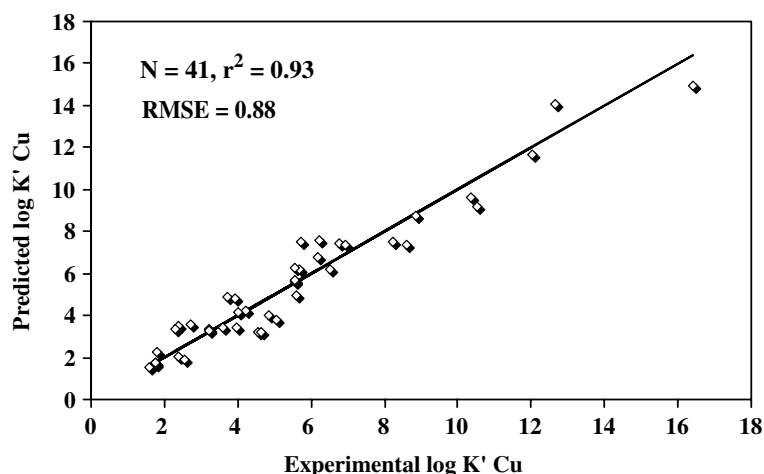| x_i	a_i	s_{a_i}
# COOH	1.37	(0.23)
# Phenols	2.51	(0.31)
# Amines	4.68	(0.32)
# Amides	0.61	(0.19)
HD	−0.36	(0.32)
ZD	−78	(26)

Fig. 1. $\log K'_{Cu}$ predicted using the parameters in Table 1 plotted versus experimental $\log K'_{Cu}$ for the 41 compounds in the calibration data set.

S.E. Cabaniss et al. / Applied Geochemistry 22 (2007) 1646–1658

For example, catechol (two phenols, $\log K'_{Cu} = 5.6$) binds Cu(II) more strongly than salicylic acid (1 COOH, 1 phenol, $\log K'_{Cu} = 4.2$) which in turn binds more strongly than phthalic acid (two COOH, $\log K'_{Cu} = 3.22$). Ethylenediamine, with two aliphatic amine groups, has a $\log K'_{Cu} = 6.5$ at pH 7, and all of the ligands with $\log K'_{Cu} > 10$ have at least one amine group.

A simple linear equation can be unrealistic when applied to Cu(II) binding by larger molecules. Although a very large molecule might have dozens of potential ligand groups, Cu(II) ions typically have a maximum coordination number of 6. In addition, the calibration data set contains no molecules which have more than two phenolic or amine groups in the actual complex. To circumvent this problem, the prediction algorithm used in Alpha-Step limits the number of ligand groups which can contribute to the estimated $\log K'_{Cu}$. The maximum group counts are two amines, two phenols and four amides with an overall maximum of six ligand groups on a single Cu(II). The full number of groups is used in the calculation of ZD and HD, however. For example, calculating K'_{Cu} for a molecule with six carboxylic acids and four phenols and MW 1000 would use Eq. (2) and the coefficients from Table 1 with the #phenols = 2, #carboxylic acids = 4, and ZD = −0.0060. If the molecule had the formula $C_{48}H_{40}O_{24}$, then heteroatom density HD = 0.5 and predicted $\log K'_{Cu} = 10.79$, comparable to nitrilotriacetic acid (NTA) which has $\log K'_{Cu} = 10.54$ at pH 7.0. This modified equation is more physically reasonable for large molecules and provides the same fit to the small-molecule calibration data set as the fully linear equation.

Note that although every molecule in the simulation is assigned a K'_{Cu} value, not every molecule would be detectable in a Cu(II) binding experiment. At pH 7.0 in equilibrium with atmospheric CO_2, the organic ligands must compete with hydroxide and carbonate ligands at a minimum, and perhaps other ligands as well. Since pCu values >5 are unattainable at pH 7.0 due to precipitation of solid $Cu(OH)_2$ ($K_{sp} = 10^{-19.32}$), ligands with $\log K'_{Cu} < 5$ are mostly not complexed. For purposes of discussion, ligands with $\log K'_{Cu} < 5$ at pH 7.0 are designated 'weak', having little effect on Cu(II) speciation. Ligands with $\log K'_{Cu} > 5$ but <10 are designated 'strong' and those with $\log K'_{Cu} > 10$ are designated 'very strong'.

If this QSAR is applied to the six precursor molecules in the present version of AlphaStep (Cabaniss et al., 2005), the results are plausible. Cellulose is predicted to have negligible binding ($\log K'_{Cu} = -0.30$), since it lacks good ligand groups (i.e., the alcohol groups are not better ligands than water). Lignin and abietic acid each have a single ligand group, but are weak ligands since the $\log K'_{Cu}$ values are 2.44 and 1.59, respectively. Fustin and meta-digallic acid have multiple phenols, and binding constants are higher ($\log K'_{Cu} = 4.88$ and $\log K'_{Cu} = 6.40$, respectively), consistent with catechol (two adjacent phenols, $\log K'_{Cu} = 5.6$). On the other hand, the protein fragment used a precursor molecule has multiple amine and acid groups leading to a very strong predicted binding constant of $\log K'_{Cu} = 14.6$.

3. Simulation results and discussion

Simulation of NOM assemblages under the biogeochemical conditions discussed here was described in Part I of this series (Cabaniss et al., 2005). Briefly, system 1 is a soil simulation which begins with small molecule precursors (abietic acid, meta-digallic acid and fustin) and incubates them at pH 5.0 in the dark. System 2 simulates the degradation of protein and lignin in sunlit water at pH 7.0. Although the initial precursor molecules of the two simulations were quite different in size and structure, the NOM assemblages produced at 5000 h have very similar molecular weight, aromaticity and percent H; on the other hand, the system 2 assemblage is higher in O and N content and has lower carboxyl acidity.

3.1. Evolution of $\log K'_{Cu}$ distributions over time

The Cu-complexing ability of the NOM assemblages evolves over time from very simple to more complex distributions of $\log K'_{Cu}$. Fig. 2 shows snapshots in the development of Cu(II) binding in the NOM assemblage from simulation 1. The three precursor values initially are present as sharp spikes in the distribution (Fig. 2a), but after 2000 h (~12 weeks) of reaction time, the phenolic compounds have reacted (mostly oxidation) and/or been consumed, so that few compounds with high $\log K'_{Cu}$ remain (Fig. 2b). However, as the simulation continues other reactions including condensations become more important, and the distributions after 4000 h and 6000 h (Fig. 2c and d) show increasing numbers of strong ligands as the average MW in the assemblage increases. The multiple 'peaks' in
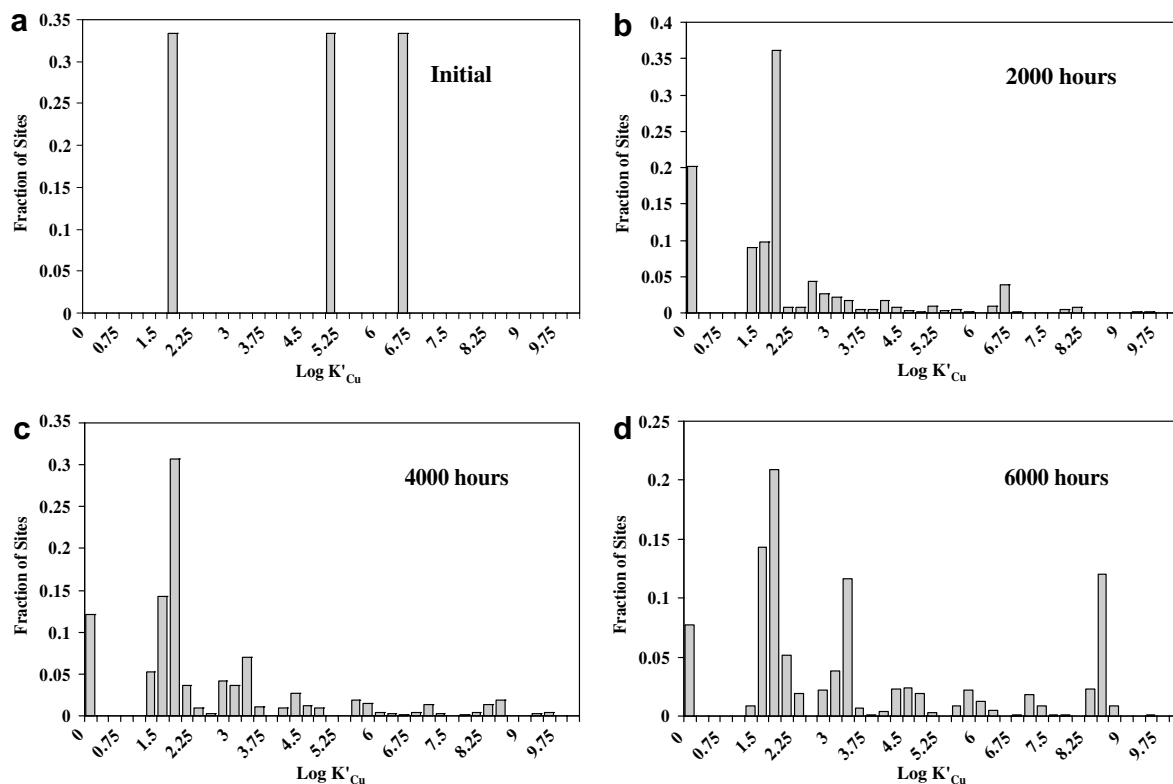
Fig. 2. Time dependent frequency distribution of numbers of molecules with various values of $\log K'_{Cu}$ for simulation 1. (a) Initial distribution, reaction time 0 h, (b) 2000 h reaction time, (c) 4000 h reaction time, (d) 6000 h reaction time.

these distributions correspond roughly to different numbers of ligand groups. Simulation 2 shows more consistent changes over this time period, with two initial spikes at very high and very low $\log K'_{Cu}$ gradually changing to a more diverse distribution between $\log K'_{Cu}$ 0 and 15.

The simulations in Cabaniss et al. (2005) examined the composition of the NOM assemblages after 5000 h simulated reaction time. At that point, the phenolic compounds in simulations 1 have been largely modified, and the resulting NOM assemblage contains 1154 molecules with bulk composition 54% C, 41% O and 5.3% H and number-average molecular weight (M_n) of 612 amu and weight-average molecular weight (M_w) of 1374 amu. After 5000 h of simulation 2, the NOM assemblage contains 4804 molecules with bulk composition 44% C, 49% O, 5.1% H and 2.3% N and M_n of 717 amu and M_w of 1173 amu. Calculated distributions of $\log K'_{Cu}$ after 5000 hare multimodal (Fig. 3) with patterns intermediate to those at 4000 and 6000 h (Fig. 2). In each simulation, most of the binding constants are too weak to serve as effective

complexing agents for Cu(II) at pH 7.0; however, simulation 2 produced many more high K'_{Cu} binding sites than simulation 1. While only 17.6% of the binding sites are "strong" and 0.4% are "very strong" in simulation 1, 29.6% of the sites are "strong" and 10% are "very strong" in the simulation 2 data set.

One advantage of a heterogeneous, agent-based model over models based on average or bulk properties is that variability within the NOM assemblage is easily represented. A related advantage is that the relationships among molecular variables can be explored within a single assemblage. Thus, while experimental studies or bulk models might compare or correlate Cu(II) binding and molecular weight (or some other property) among different NOM samples, agent-based models allow this type of comparison within a single NOM assemblage. This approach may be especially useful for properties for which heterogeneity (variability within a single assemblage) is greater than sample-to-sample variability. One can take advantage of this information on individual molecules to consider how and why
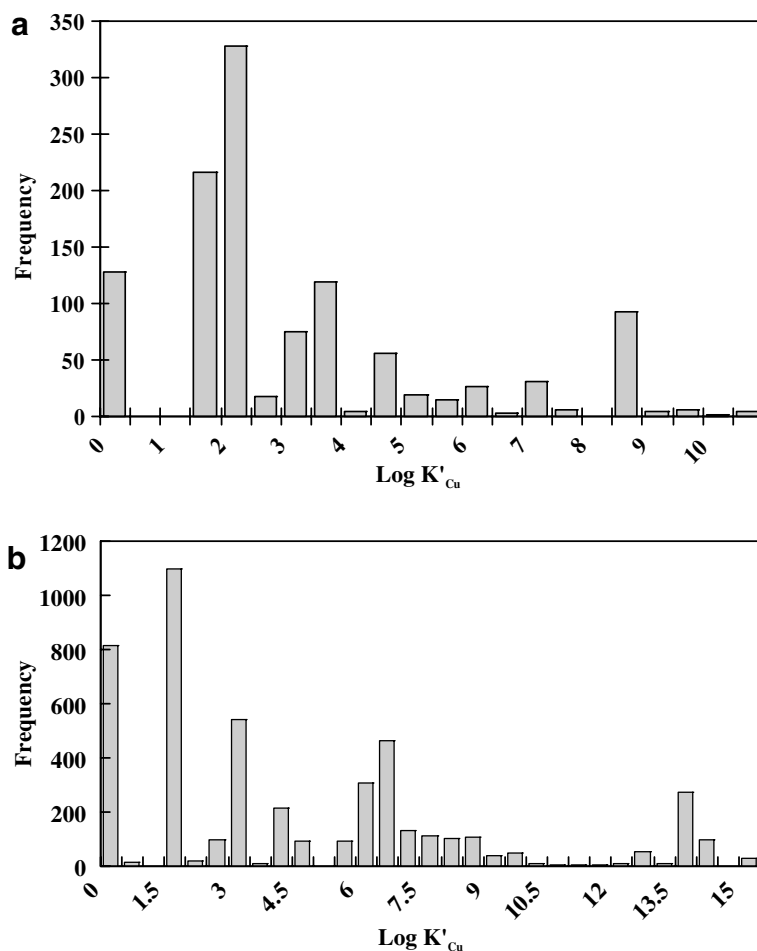
Fig. 3. Frequency distribution of numbers of molecules with various values of $\log K'_{Cu}$ after 5000 h of simulation. (a) Simulation 1, (b) simulation 2.

these two simulated NOM assemblages have such different Cu(II) binding distributions.

3.2. Molecular aspects of Cu(II) complexation by NOM

To a first approximation, the greater binding strength of the NOM in simulation 2 can be attributed to the presence of amine groups, which are entirely lacking from the NOM in simulation 1. The $\log K'_{Cu}$ of molecules in simulation 2 is strongly correlated with the number of N atoms in those molecules, with a correlation coefficient of 0.86 ($r^2 = 0.74$). Fig. 4 shows that as the number of N atoms increases, the likelihood of very strong binding increases even after the first 3 or 4 N atoms; this is because most of the N-containing molecules contain amide as well as amine linkages. Molecules in

simulation 1 are composed entirely of C, H and O and have an optimal ligand combination of two phenols and four carboxylic acids which gives a $\log K'_{Cu}$ contribution of 10.5 (not including ZD and HD effects); in contrast, a N-containing molecule in simulation 2 might have two amines and four carboxylic acids contributing 14.84 log units (and the 'optimal' ligand set of two amines, two phenols and two carboxylic acids would be even higher). Interestingly, no correlation is observed for $\log K'_{Cu}$ with ZD, MW, or the number of phenols (all $r^2 < 0.02$) in the simulation 2 assemblage.

In the absence of the strong amine ligands, binding by molecules in simulation 1 has very different characteristics. $\log K'_{Cu}$ correlates weakly with both ZD ($r^2 = 0.19$) and with MW ($r^2 = 0.59$). The latter relationship is illustrated by Fig. 5, which shows $\log K'_{Cu}$ plotted against MW for all the molecules
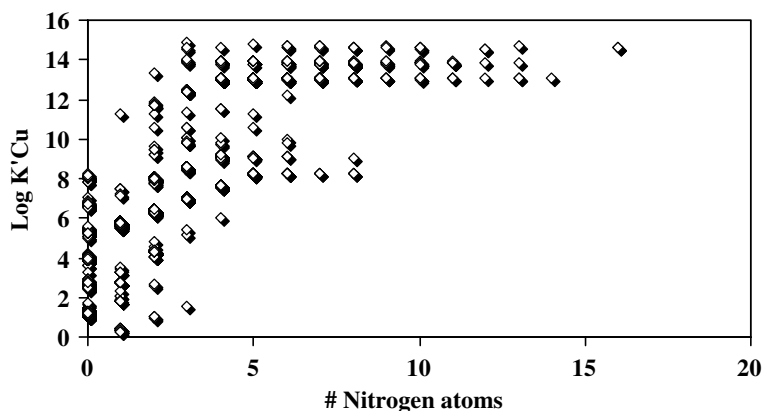
Fig. 4. $\log K'_{Cu}$ plotted versus the number of N atoms for each molecule of the simulation 2 NOM assemblage.
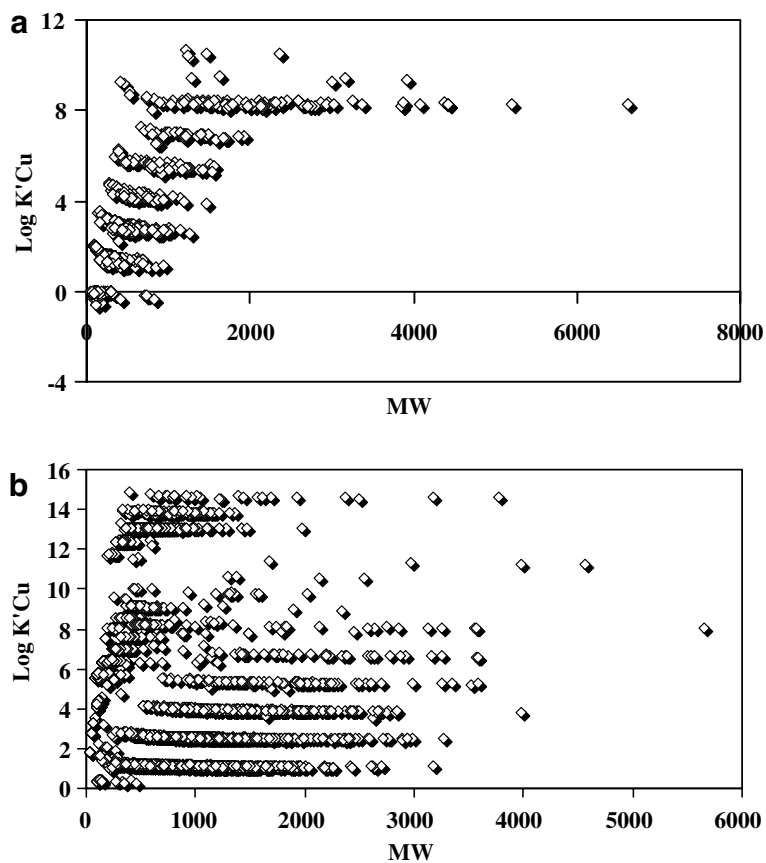

Fig. 5. $\log K'_{Cu}$ plotted versus the molecular weight for each molecule of the simulation 1 (a) and simulation 2 (b) NOM assemblages.

in each assemblage. Seven 'steps' are apparent in Fig. 5a, corresponding to the presence of 0–6 ligand groups. Fig. 5b shows two groups of steps, corresponding to molecules with O ligands only (lower $\log K'_{Cu}$) and both N and O ligands (higher $\log K'_{Cu}$).

$\log K'_{Cu}$). It is interesting that for simulation 1 the correlation with MW is stronger than that with ZD, given that ZD is a variable in the QSAR while MW is not. The positive correlation with MW is apparently due to the increasing probability of find-

ing multiple ligand groups on a molecule as MW increases; this is much less important in simulation 2, since only a few amine groups are required to produce large $\log K'_{Cu}$ values.

3.3. Comparison with experimental data

Although this agent-based simulation is intended to provide mechanistic insight rather than to fit experimental data, it is nonetheless important to compare the model predictions with knowledge of Cu(II) binding by NOM. Qualitative agreement between simulation and experiment can provide confidence in the principles of the current model, although of course it does not constitute 'verification'. On the other hand, qualitative inconsistencies or gross quantitative discrepancies between simulation and experiment point out potential deficiencies in the conceptual model and/or its implementation.

Titration curves of simulated NOM were calculated by adding together contributions from all molecules in the NOM assemblage. First, a series of pCu values was selected to cover the range of experimental data. Then, for each ligand in the mixture the fraction of that ligand complexed by Cu(II), f_i, is calculated from

$$f_i = \frac{[CuL_i]}{L_{Ti}} = \frac{K'_{Cui}[Cu^{2+}]}{K'_{Cui}[Cu^{2+}] + 1} \qquad (4)$$

where L_{Ti} is the total concentration of the ith ligand, $[CuL_i]$ is the concentration of Cu(II) bound to the ith ligand L_i and K'_{Cui} is the binding constant for the ith ligand. Next, the total amount of NOM (as C) in the simulation is obtained by summing the C atoms in all molecules, C_T. Then, for a given NOM concentration in moles C L^{-1} (DOC), the total concentration of complexed Cu(II) (CuL$_T$) for each pCu value can be calculated from

$$CuL_T = \frac{(DOC)N_A \sum_{i=1}^{N} f_i}{C_T} \qquad (5)$$

where N_A is Avogadro's number. The total Cu Cu$_T$ is simply the sum of CuL$_T$ and $[Cu^{2+}]$, and the titration curve can be conveniently displayed as pCu versus pCu$_T$ ($= -\log(Cu_T)$).

Copper (II) binding titrations of NOM tend to be featureless and lacking a clear endpoint, but can vary significantly as a function of NOM source, chemical conditions (pH, salt, etc.) and the ratio of Cu(II) to NOM employed. On the one hand, titrations of some terrestrial NOM samples at low

pH and high Cu(II):NOM ratios have indicated <99% of the Cu(II) is bound (Bhat et al., 1981; Cabaniss and Shuman, 1988a), while titrations of aquatic and terrestrial NOM at very low levels of Cu(II) and pH > 7 have indicated >99.99% of the Cu(II) is complexed (Cabaniss and Shuman, 1988a; Xue and Sigg, 1993; Kogut and Voelker, 2001).

Suwannee River fulvic acid (SRFA) is a well-characterized extract from a black-water river draining the Okefenokee Swamp. It is low in N content and has acidity, molecular weight, aromaticity and other properties similar to many terrestrial fulvic acid isolates (Averett et al., 1989; Perdue and Ritchie, 2003). In addition to structural characterization studies, several independent Cu(II) binding studies on SRFA have reported mutually consistent results using Cu^{2+} ion selective electrodes (ISEs) (McKnight et al., 1983; Cabaniss and Shuman, 1988a,b; Brown et al., 1999). Copper (II) binding by SRFA appears to be generally similar (within a factor of 2–3) to Cu(II) binding by unfractionated terrestrial NOM (McKnight et al., 1983; Cabaniss and Shuman, 1988b).

Calculated titration data for simulation 1 (soil simulation) is in good qualitative agreement with experimental data for SRFA at low levels of Cu$_T$. Fig. 6 compares experimental (ISE) titration data on SRFA at pH 7.0 and ionic strength 0.10 with model calculations based on simulation 1 (at 5000 h reaction time, as in Fig. 3a). For pCu$_T$ > 5.5 (Cu$_T$ < 3 μM) the model slightly overestimates pCu, agreeing with experimental data within 0.42 log units, with a typical difference of only 0.2 log units! Although this is larger than experimental error, the agreement is remarkable for a model which was not 'fitted' to the data set. At higher Cu$_T$ (>4 μM), the model underestimates pCu by 0.5–0.9 log units. This underestimation is likely due to the assumption (a′ above) that only one Cu(II) binds to a given ligand molecule. Additional comparisons with SRFA titrations at lower DOC levels show a similar pattern – good agreement at lower Cu$_T$ and increasing underestimation of Cu(II) binding at higher Cu$_T$. However, the change from 'lower' to 'higher' Cu$_T$ behavior occurs at lower values of Cu$_T$ as the DOC decreases, which is sensible if the discrepancy is caused by assumption a′.

In contrast to SRFA, some aquatic NOM is believed to derive from algal and other autochthonous sources, and precursor compounds may contain significant quantities of proteinaceous matter
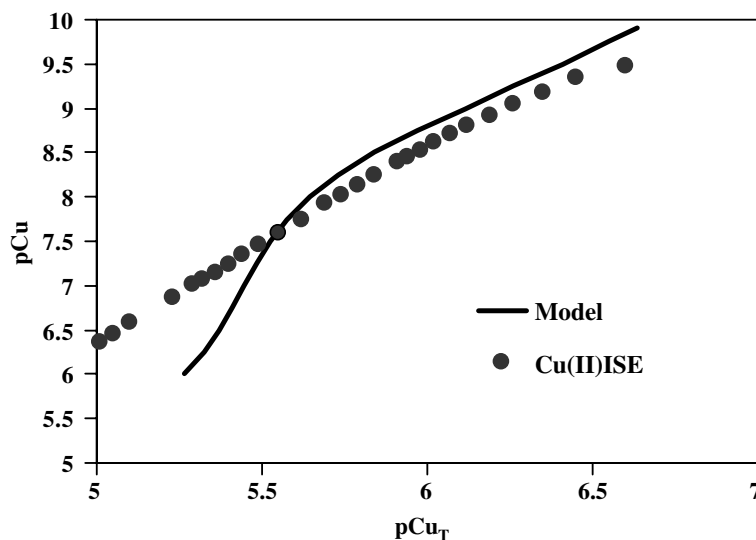
Fig. 6. Comparison of simulated Cu(II)-NOM titration curves with experimental data from a terrestrial NOM sample. Titration of 10 mg/L Suwannee River FA, pH 7.0 and 0.10 M NaClO$_4$, monitored by Cu^{2+} ion selective electrode. Experimental data from Kogut and Voelker (2001) (X) . Model calculations (shown as curve) using soil NOM distribution in Fig. 3a shown as curve.

and extracellular polysaccharides. For example, Xue and Sigg (1993) obtained Cu(II) complexation data on NOM from Lake Greifen, Switzerland, which is believed to contain a substantial proportion of autochthonous NOM. This material is present in low concentration (<5 mg C/L), but using cathodic stripping voltammetry (CSV) the authors observed very strong Cu(II) binding and very low values of [Cu^{2+}].

Calculated titration data for simulation 2 (water simulation) are in reasonably good qualitative

agreement with the experimental data of Xue and Sigg (1993) at the lowest levels of Cu$_T$. Fig. 7 compares experimental CSV titration data on Lake Greifen at pH 8.0 with 3.5 mg C/L (Fig. 4a in Xue and Sigg, 1993) with model calculations based on simulation 2 (at 5000 h reaction time, as in Fig. 3b). The sample was collected 5 m below the surface on April 23, 1990, and appears to be fairly typical of other samples collected by the authors. For pCu$_T$ > 7.0 (Cu$_T$ < 0.1 µM) the model overestimates pCu by 0.35–1.0 log units, with the smallest
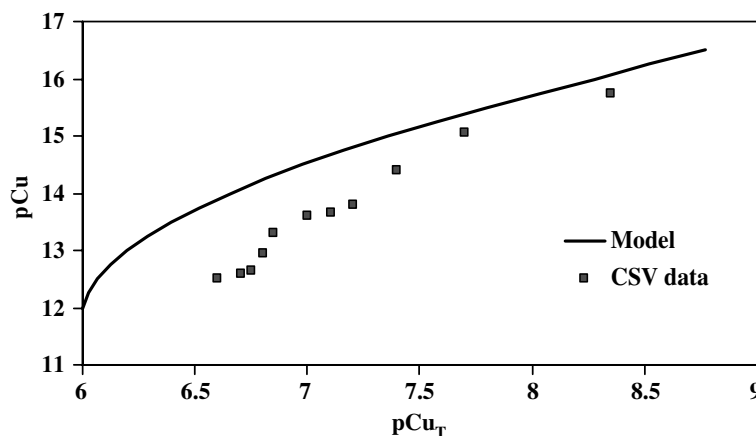

Fig. 7. Comparison of simulated Cu(II)-NOM titration curves with experimental data from an aquatic NOM sample. Titration of Lake Greifen NOM (3.5 mg C/L, pH 8.0) with Cu^{2+} monitored by cathodic stripping voltammetry. Experimental data from Xue and Sigg, 1993 (X). Model calculations (shown as curve) using water NOM distribution shown in Fig. 3b.

difference at lowest concentrations. At higher levels of Cu_T, the discrepancy is as large as 1.6 log units.

Although the quantitative agreement between the model and experimental data is not as good in simulation 2 as in simulation 1, the qualitative trend towards stronger binding by N-rich, autochthonous NOM is clearly predicted by the model. The fact that the predicted values of pCu (for a given pCu$_T$) are always <1 unit for the SRFA data and usually <1 log unit for the Lake Greifen data is also encouraging. Although it would be possible to improve the 'fit' to the data by changing the simulations (for example, by starting with less than 50% protein in simulation 2), optimal 'fit' is not the point here. Rather, this demonstrates that a first-principles model beginning only with precursor compounds and with K'_{Cu} estimation calibrated only on small molecules can make predictions that are frequently better than 'order of magnitude'.

4. Conclusions and future work

The stochastic, agent-based approach to NOM evolution from precursor natural products provides a molecular-level model of Cu(II) binding. It reproduces Cu-NOM experimental data in a qualitative or semi-quantitative sense. Approximately portraying the generally featureless but fairly strong binding curves and the differential binding properties of NOM from different precursors using such an approach represents a significant advance in understanding the biogeochemical processes that influence metal–organic complexation in natural systems. This approach could now be used to explore more specific questions about Cu(II)-NOM complexation: e.g., How does irradiation affect Cu(II) complexation?

This modeling effort is a work in progress, rather than a finished product, and several planned modifications in the general model will affect its ability to predict Cu(II) complexation. These include continuous additions of precursor molecules (as opposed to the current batch mode), seasonal and diurnal variations in temperature and more sophisticated enzyme activity controls, including feedback from the microbial community. An important planned addition to the data model is the inclusion of organic S in the form of thiols and thio-ethers. Organosulfur could be derived either from S-containing amino acids like cysteine or from nucleophilic attack by HS^- in reducing environments. Although typically present in low concentrations,

these functional groups are expected to bind transition metals like Cu(II) strongly, and their inclusion may greatly enhance predicted Cu complexation at very low Cu(II):DOC ratios. Other possible improvements which could affect Cu(II) binding predictions but which are not currently scheduled for coding include the addition of redox chemistry and inter-molecular associations.

The metal complexation portion of the current model is restricted in several notable ways. First, it is calibrated only for Cu(II), although additional calibration processes are underway for Ca(II), Cd(II), Ni(II), Pb(II), Zn(II) and Al(III). Other metals may be added as time and resources permit. Secondly, the K'_{Cu} values are appropriate only for pH 7.0 and ionic strength 0.1. Presumably, the range of pH applicability can be expanded to 3–9 by using estimated pK_a constants for carboxylic acids currently available (Cabaniss et al., submitted for publication) and making some reasonable assumptions about the pK_a values of phenols and amines in these molecules. Ionic strength corrections could be performed either using a simplistic Debye–Huckel approach appropriate for small ligands or a more sophisticated small sphere solution to the Poisson–Boltzmann equation (Bartschat et al., 1992). Finally, it may be appropriate to avoid assumption a′ by calculating multiple sets of K'_{Cu} values for ligand molecules which might complex more than one Cu(II) at a time.

Acknowledgements

Appendix

Calibration data for $\log K'_{Cu}$ QSAR. Values from NIST database (Smith et al., 1998) adjusted for pH 7.0

Compound	$\log K'_{Cu}$
Acetic acid	1.79
Malonic acid	5.04
Succinic acid	2.7
Tartaric acid	3.2

Appendix (*continued*)

Compound	log K'_{Cu}
Pentanedioic acid	2.37
Hexanedioic acid	2.3
Malic acid	3.6
Methylamine acid	1.75
Ethylenediamine	6.5
Benzoic acid	1.76
Salicylic acid	4.2
Phthalic acid	3.22
Catechol	5.6
Citric acid	3.91
Tricarballylic acid	3.7
NTA	10.54
EDTA	16.4
EDMA	10.37
Glycine	5.57
Leucine	5.56
Aspartic	6.23
Glutamic	5.74
Glutamine	5.7
3-Hydroxynaphthoic acid	4.01
3,4-Dihydroxy benzoic acid	6.18
Glycolic acid	2.4
Ditartronic acid	5.54
Lactic acid	2.54
4-Hydroxy butanoic acid	1.76
2-Methoxy benzoic acid	1.6
Oxalic acid	4.85
Phenyl malonic acid	4.54
Cyclohexane 1,1-dicarboxylic acid	4.62
Diglycolic acid	3.95
2-Aminoethyl *N,N*-diacetic acid	12.02
Gly-Pro-Gly-Pro-Glutamine	8.87
Imino 3-propanoic acid	6.75
Iminodiacetic acid (IDA)	8.22
cis-2,6-dicarboxy piperidine	10.92
N-Benzyl imino diacetic acid	8.6
Diethyl triamine acetic acid	12.66

References

Averett, R.C., Leenheer, J.A., McKnight D.M., Thorn K.A., 1989. Humic Substances in the Suwannee River, Georgia: Interactions, properties and proposed structures. US Geol. Surv. Open File Report 87-557, Washington, DC.

Bartschat, T., Cabaniss, S.E., Morel, F.M.M., 1992. An oligo-electrolyte model for cation binding by humic substances. Environ. Sci. Technol. 26, 284–294.

Bhat, G.A., Saar, R.A., Smart, R.B., Weber, J.H., 1981. Titration of soil-derived fulvic acid by Cu(II) and measurement of free Cu(II) by anodic stripping voltammetry and Cu(II) selective electrode. Anal. Chem. 53, 2275–2280.

Brown, G.K., Cabaniss, S.E., MacCarthy, P., Leenheer, J.H., 1999. Cu(II) binding by a pH-fractionated fulvic acid. Anal. Chim. Acta 402, 183–193.

Bruland, K.W., Rue, E.L., Donat, J.R., Skrabal, S.A., Moffett, J.W., 2000. Intercomparison of voltammetric techniques to determine the chemical speciation of dissolved copper in a coastal seawater sample. Anal. Chim. Acta 405, 99–113.

Cabaniss, S.E., Shuman, M.S., 1988a. Copper binding by dissolved organic matter: I. Suwannee River fulvic acid equilibria. Geochim. Cosmochim. Acta 52, 185–193.

Cabaniss, S.E., Shuman, M.S., 1988b. Copper binding by dissolved organic matter: II. Variation in type and source of organic matter. Geochim. Cosmochim. Acta 52, 195–200.

Cabaniss, S.E., Madey, G., Leff, L., Maurice, P.A., Wetzel, R.G., 2005. A stochastic model for the synthesis and degradation of natural organic matter. Part I: Data structures and reaction kinetics. Biogeochemistry 76, 319–347.

Cabaniss, S.E., Madey, G., Leff, L., Maurice, P.A., Wetzel, R.G., submitted for publication. A stochastic model for the synthesis and degradation of natural organic matter. Part II: Molecular property distributions. Biogeochemistry.

Carey, F.A., Sundberg, R.J., 2000. Advanced Organic Chemistry Part A: Structure and Mechanisms, fourth ed. Kluwer Academic/Plenum Publishers, New York.

Cook, R.L., McIntyre, D.D., Langford, C.H., Vogel, H.J., 2003. A comprehensive liquid-state heteronuclear and multidimensional NMR study of Laurentian fulvic acid. Environ. Sci. Technol. 37, 3935–3944.

DiToro, D.M., Allen, H.E., Bergman, H.L., Meyer, J.S., Paquin, P.R., Santore, R.C., 2001. Biotic ligand model of the acute toxicity of metals 1. Technical basis. Environ. Toxicol. Chem. 20, 2383–2396.

Harris, D.C., 1999. Quantitative Chemical Analysis, fifth ed. WH Freeman.

Hatcher, P.G., Dria, K.J., Kim, S., Frazier, S.W., 2001. Modern analytical studies of humic substances. Soil Sci. 166, 770–794.

Huang, Y., Xiang, X., Madey, G., Cabaniss, S., 2005. Agent-based scientific simulation. Comput. Sci. Eng. 7, 22–29.

Kogut, M.B., Voelker, B.M., 2001. Strong copper binding behavior of terrestrial humic substances in seawater. Environ. Sci. Technol. 35, 1149–1156.

Koopal, L.K., Saito, T., Pinheiro, J.P., van Riemsdijk, W.H., 2005. Ion binding to natural organic matter: general considerations and the NICA-Donnan model. Colloid Surf. A: Physicochem. Eng. Aspects 265, 40–54.

McKnight, D.M., Feder, G.L., Thurman, E.M., Wershaw, R.L., Westall, J.H., 1983. Complexation of copper by aquatic humic substances from different environments. Sci. Total Environ. 28, 65–76.

Perdue, E.M., Ritchie, J.D., 2003. Dissolved organic matter in freshwaters. In: Drever, J.I. (Ed.), Treatise on Geochemistry, Surface and Ground Water, Weathering, and Soils, vol.5. Elsevier, San Diego, CA, USA, pp. 273–318 (Chapter 10).

Schmitt-Koplin, P., Garrison, A.W., Perdue, E.M., Freitag, D., Kettrup, A., 1998. Capillary electrophoresis in the analysis of humic substances – facts and artifacts. J. Chromatogr. A 807, 101–109.

Schwarzenbach, R.P., Gschwend, P.M., Imboden, D.M., 2003. Environmental Organic Chemistry. Wiley-Interscience, Hoboken, NJ.

Smith, R.M., Martell, A.E., Motekaitis, R.M., 1998. NIST Critically Selected Stability Constants of Metal Complexes Database. US Department Commerce, Gaithersburg, MD.

Sunda, W., Guillard, R.L.L., 1976. The relationship between cupric ion activity and the toxicity of copper to phytoplankton. J. Water Res. 34, 513–529.

Tanford, C., 1961. Physical Chemistry of Macromolecules. Wiley, NY.

Tipping, E., 2002. Cation Binding By Humic Substances. Cambridge University Press, Cambridge, UK.

Westall, J.C., Jones, J.D., Turner, G.D., Zachara, J.M., 1995. Models for associations of metal ions with heterogeneous environmental sorbents 1. Complexation of Co(II) by Leonardite humic acid as a function of pH and $NaClO_4$ concentration. Environ. Sci. Technol. 29, 951–959.

Xiang, X., Huang, Y., Madey, G., Cabaniss, S., 2004. A web portal for supporting environmental science research. In: Scharl, A. (Ed.), Environmental Online Communication. Springer, New York, pp. 113–126.

Xue, H.B., Sigg, L.L., 1993. Free cupric ion concentration and Cu(II) speciation in a eutrophic lake. Limnol. Oceanogr. 38, 1200–1213.