Feature Article

Application of a constrained regularization method to extraction of affinity distributions: Proton and metal binding to humic substances

Silvia Orsetti, Estela María Andrade, Fernando V. Molina *

INQUIMAE, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, C. Universitaria, Pabellon II, Buenos Aires C1428EHA, Argentina

A R T I C L E   I N F O

Article history:
Received 12 March 2009
Accepted 15 April 2009
Available online 24 April 2009

Keywords:
Humic substances
Pollutant retention
Binding equilibrium
Soil organic matter
CONTIN

A B S T R A C T

The binding of proton and metal cations to humic substances has been analyzed with a regularized fitting procedure (using the CONTIN software package) to extract conditional affinity distributions, valid at a given ionic strength, from binding (titration) curves. The procedure was previously tested with simulated titration curves using a simple bi-Gaussian model, the NICA–Donnan model, and the Stockholm humic model. Application to literature data for proton binding shows that in several cases the affinity distribution found is bimodal (carboxylic and phenolic sites) as usually assumed; however in other cases, especially for fulvic acids, a trimodal distribution is clearly discerned, with a smaller peak between the two noted above attributed to the presence of vicinal carboxylic groups. The analysis of metal binding curves has been performed in a few cases where the available data could be reliably processed, separating the proton affinity distribution and obtaining the conditional affinity spectra. For Cd(II) and Pb(II) a bimodal distribution is found, attributed in principle to mono- and bidentate binding, based on spectroscopic data. In the case of Cu(II), a more complex affinity distribution is found showing 3–4 peaks; this is consistent with spectroscopic studies, where different binding modes, up to tetradentate, have been observed.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Humic substances (HSs) form the most important part of soil organic matter (SOM) and are also present in groundwaters. SOM has a fundamental role in nutrient availability for plants, and also has several other important properties [1,2]. HS show colloidal behavior and interact with other chemical species in the natural environment, including binding of inorganic cations; this ability, in particular for binding heavy metals, has an important role in the fate of such pollutants in the environment. Thus, this topic has attracted a high number of researchers [3–13] over the last decades. Among the different experimental studies, equilibrium binding data, usually in the form of titration curves, have been reported in the literature; most of them have been reviewed by Milne et al. [3,14], and several more have been published more recently [4,15–20]. The interpretation of such curves in terms of equilibrium binding reactions is complicated by the fact that HS are heterogeneous in nature, showing a high number of binding chemical groups (sites) with varying equilibrium constants (or affinity for the bound species). A fully detailed molecular description of HS is, at best, very difficult to achieve, so that simplified models have been widely employed to interpret the titration curves. These models generally include a description of the intrinsic binding constants in terms of a distribution, the affinity distribution or affinity spectrum, and an electrostatic contribution to account for the colloidal behavior of HS. The intrinsic affinity distribution has been assumed to be either discrete (effectively similar to a mixture of weak acids) or continuous, with some assumed distribution function, but in both cases two main types of groups or sites have been included: carboxylic sites, with $pK_a$ in the range 3–5, and phenolic ones, with $pK_a$ ranging between 8 and 10.

One of the first models was the Gaussian humic model, where a bimodal Gaussian function, including the two types noted, without electrostatic contribution, was applied [5]. The most widely used continuous distribution model is the NICA–Donnan (ND) model, developed by the Wageningen group [3,6,14,21], which has been applied with good results, from the point of view of fitting ability, to a number of problems involving natural organic matter, and humic substances in particular. It consists of a bimodal Sips distribution coupled with the Donnan volume electrostatic model. Among the discrete site models, Tipping models WHAM V and VI [4,22–25] have shown good predicting capabilities. It consists of two groups of four discrete sites (carboxylic and phenolic), and has been combined for the electrostatic effects either with a Donnan volume model or with a solid particle surface model [4]. More recently, the Stockholm humic model (SHM) was proposed and applied to several cases [26–30]. It is related to Tipping’s WHAM models V–VI, defining a discrete affinity distribution and introducing a particle-surface-type electrostatic model, with some similarity to the CD-MUSIC model for inorganic colloids.
Besides the application of specific models, work has also been devoted to extract affinity distributions directly from experimental data, including analytical [31–34] and numerical [35–40] methods. The extraction of affinity distribution from experimental data is an ill-posed problem [35,36,41] and as such common least-squares methods lead to a nearly infinite number of oscillating solutions [41]. One way to overcome this problem is the use of constrained regularized least-squares methods, which have been applied by several authors [35–37]. In a regularization procedure, the objective function to be minimized, usually the sum of quadratic differences, is modified by a regularizer term, which penalizes the unwanted solutions, in this case those showing rapidly changing behavior, favoring instead smooth solutions, with few peaks. Even when it is difficult to extract true exact distributions from titration data [36], at least without additional knowledge, useful information can be obtained with such procedures. Provencher has developed a regularization method, implemented in the CONTIN program code [41,42], which has been widely employed, with applications to relaxation studies [43], dynamic light scattering [44,45], and NMR spectroscopy [46,47] among many others [48–50]. In particular, several applications to adsorption studies have been reported [35,51–54]. In the application of this and other numerical methods, conditional affinity spectra are commonly obtained, valid under given conditions, usually fixed ionic strength and, in the case of metal binding, also fixed pH [31,39,55].

In this work, the CONTIN package is employed to analyze proton and cation binding data to HS, including humic (HA) and fulvic (FA) acids, without applying any specific model. Literature data for proton and cation binding data to HS, including humic (HA) and fulvic (FA) acids, without applying any specific model. Literature data for metal binding, also fixed pH [31,39,55]. In particular, several applications to adsorption studies have been reported [35,51–54]. In the application of this and other numerical methods, conditional affinity spectra are commonly obtained, valid under given conditions, usually fixed ionic strength and, in the case of metal binding, also fixed pH [31,39,55].

In this work, the CONTIN package is employed to analyze proton and cation binding data to HS, including humic (HA) and fulvic (FA) acids, without applying any specific model. Literature data for metal binding, also fixed pH [31,39,55]. In particular, several applications to adsorption studies have been reported [35,51–54]. In the application of this and other numerical methods, conditional affinity spectra are commonly obtained, valid under given conditions, usually fixed ionic strength and, in the case of metal binding, also fixed pH [31,39,55].

2. Methods

2.1. Application of regularization methods to ion binding

2.1.1. Proton binding

Humic substances are highly heterogeneous compounds found in groundwaters and soils. They have acid base behavior which can be written as a binding process

$$A^- + H^+ \rightleftharpoons HA$$  \hspace{1cm} (1)

where $A^-$ stands for a single proton binding group (site) in the humic substance. For a simple acid with a homogeneous binding equilibrium constant, the process (1) can be described by a Langmuir-type equation

$$Q_{H} = Q_S \frac{K_{H}a_{H}}{1 + K_{H}a_{H}},$$  \hspace{1cm} (2)

where $Q_{H}$ is the amount (mol/kg) of $H^+$ bound, $Q_S$ is the maximum (saturation) amount of $H^+$, $K_{H}$ is the binding equilibrium constant, and $a_{H}$ is the $H^+$ activity in the solution. Because HSs are heterogeneous, there is not a single constant but a distribution (or affinity spectrum), so Eq. (2) becomes, as it is well known,

$$Q_{H} = Q_S \int_{-\infty}^{\infty} \frac{K_{H}a_{H}}{1 + K_{H}a_{H}} g(\log K_{H}) d \log K_{H},$$  \hspace{1cm} (3)

where the Langmuir expression is applied to the fraction of sites with the same $K_{H}$ value; this is the local isotherm.

$$Q_H = Q_S \frac{K_{H}a_{H}}{1 + K_{H}a_{H}},$$  \hspace{1cm} (4)

where $Q_S$ is the amount of sites with binding constant between $\log K_{H}$ and $\log K_{H} + d \log K_{H}$. In Eq. (3), $g$ is a conditional affinity distribution if the bulk $H^+$ activity is considered, as electrostatic effects are not taken into account. As often $Q_S$ is not known a priori, the product $Q_S g(\log K_{H}) = g(\log K_{H})$ is obtained as an unnormalized distribution; $Q_S$ can in principle be obtained through normalization of $g$. Moreover, from titration experiments, the absolute value of $Q_S$ is not determined, but instead the change of humic negative charge (i.e., unbound sites) $\Delta Q_S$ is obtained, so that the equation to be solved is

$$\Delta Q_{A} = Q_{A} + Q_{B} = -Q_{H} - Q_{D} = Q_{0} - Q_{S} = Q_{H} + \Delta Q_{B},$$  \hspace{1cm} (5)

where $Q_{B}$ is the initial charge before the titration.

2.1.2. Regularization procedure

Eq. (3) is a Fredholm integral equation of the first kind, and the inversion of such equations is an ill-posed problem because there are multiple oscillation solutions satisfying them. Thus, standard fitting procedures are not suitable for finding the affinity distribution. These problems can be solved numerically to estimate the distribution function $g(\log K_{H})$ using regularization algorithms, such as that implemented in the CONTIN package by Provencher [41,42]. In this procedure, Eq. (3) is discretized on a grid of $N_k$ points $\lambda_m$ as

$$Q_k = \sum_{m=1}^{N_k} c_{k}\delta(\lambda_m)s(\lambda_m) + L_{k} \beta,$$  \hspace{1cm} (6)

where $F_k(\lambda)$ is the kernel function, in the present case the local isotherm which is chosen as the Langmuir function

$$F_k(\lambda) = \frac{10^a_k}{1 + 10^{a_k}},$$  \hspace{1cm} (7)

with $a_k$ being the proton activity corresponding to $Q_k$: $a_k$ and $Q_k$ ($1 \leq k \leq N_k$) are the experimental points, $\beta$ is a constant background (here, $\Delta Q_{B}$). The discretized distribution $s(\lambda)$ and $k$ are to be determined. Eq. (6) can be written in matrix form

$$q = Ax,$$  \hspace{1cm} (8)

where $A$ is an $N_k \times N_k + 1$ matrix containing $c_{k}\delta(\lambda_m)$ and the $L_k$ in the present case set to 1; $q$ is a vector (dimension $N_k$) containing the $Q_k$ values and $x$ is a vector (dimension $N_k + 1$) containing the unknowns, $s_{\lambda_k}(\lambda)$ and $\beta$. The objective function minimized in CONTIN [41,42] is

$$V(x) = x^2 + \beta^2 R = \left\| M^{-1/2}(q - Ax) \right\|^2 + \beta^2 \left\| r - Bx \right\|^2,$$  \hspace{1cm} (9)

where $M$ is the covariance matrix of the experimental uncertainties and $\left\| \cdot \right\|$ indicates the Euclidean norm. The second term of the right-hand side of Eq. (9) is the regularizer; $x$ is the regularization parameter, its value controlling the weight of the regularization function $R$. This function in CONTIN is set as

$$R = \left\| r - Bx \right\|^2 = \int_{\lambda_1}^{\lambda_{\text{u}}} \frac{d^2 \beta}{d \lambda^2} \, d \lambda,$$  \hspace{1cm} (10)

where $\lambda_1$ and $\lambda_{\text{u}}$ are the upper and lower grid limits. This function, being essentially the curvature of $s$, favors smooth solutions. The regularization parameter $\beta$ can be automatically chosen by CONTIN or controlled through input parameters. These parameters are $\text{rsvmin}$ and $\text{rsvmax}$, such as $x$ lies in the range [41]

$$\text{rsvmin} \leq x \leq \text{rsvmax},$$  \hspace{1cm} (11)

where $s_1$ is an internal scale factor (see Appendix A.1 of [41]) and $p$ is the internal precision, automatically set, in the Windows
environment employed, to \( p = 1.49 \times 10^{-15} \). Other restrictions include nonnegativity of the solution, which is important in excluding many oscillating solutions, and also user-defined restrictions.

CONTIN uses as a means to select the optimum \( x \) value the parameter
\[
P_1(x) = F(f_1(x), N_{d_i}(x_0), N_p - N_{d_i}(x_0)),
\]
where \( F(f, n_1, n_2) \) is Fischer’s \( F \) distribution with \( n_1 \) and \( n_2 \) degrees of freedom, \( x_0 \) is the \( x \) value which minimizes \( V(x) \), \( N_{d_i}(x_0) \) is the number of freedom degrees for such value, and
\[
f_1(x) = \frac{V(x) - V(x_0)}{V(x_0)} \times \frac{N_p - N_{d_i}(x_0)}{N_{d_i}(x_0)}.
\]

The solution chosen by CONTIN is that with \( P_1 \) nearest 0.5. Thus, the convergence criterion in CONTIN is essentially based on the \( F \) distribution. As discussed later on in Section 3.1.4, on physical grounds restrictions to the \( x \) range were imposed thus modifying the above behavior.

### 2.1.3. Competitive metal-proton binding

HSSs bind metal cations such as Ca\(^{2+}\), Pb\(^{2+}\), Al\(^{3+}\), etc. As a first approximation, the metal binding can be written similarly to (1):
\[
A^+ + M^{z+} = AM^{z-}.
\]

However, in general both \( M^{z+} \) and \( H^+ \) bind to the same sites, so that the competitive Langmuir equation can be used as local isotherm for the amount of \( M^{z+} \) bound, \( q_M \). Assuming that each \( M^{z+} \) binds to a single site, it can be written
\[
q_M = \frac{K_{d_M} a_M}{1 + K_{d_M} a_M}.
\]

where \( K_{d_M} \) is the metal binding constant and \( a_M \) is the metal activity. Now for the heterogeneous case, Eq. (15) should be integrated for a bidimensional distribution \( p(\log K_M, \log K_H) \), which in the general case will include the correlations between \( H^+ \) and \( M^{z+} \) binding:
\[
Q_M = Q_s \int g(\log K_H, a_M) p(\log K_M, \log K_H) d\log K_H d\log K_M.
\]

As a way to simplify the problem we will assume that \( p \) can be written as the product of two distributions corresponding to \( K_H, g(\log K_H) \) and \( K_M, f(\log K_M) \). This assumption implies in principle that these distributions are different and independent (uncorrelated). In practice, if \( g \) is the distribution deduced for \( H^+ \) binding in the absence of any interfering ions, the affinity distribution \( f \) will carry the \( H^+ - M^{z+} \) correlations. Because binding of metal cations in the full absence of protons is, at best, difficult to achieve experimentally, the “pure” distribution for \( M^{z+} \) binding is experimentally not accessible. Thus, the amount of metal bound will be written as
\[
Q_M = Q_s \int \frac{K_{d_M} a_M}{1 + K_{d_M} a_M} g(\log K_H) f(\log K_M) d\log K_H d\log K_M.
\]

As the \( \log K_H \) distribution is determined in a separate experiment, Eq. (17) can be solved with CONTIN by grouping as
\[
Q_M = \int_{-\infty}^{\infty} \frac{K_{d_M} a_M}{1 + K_{d_M} a_M} g(\log K_H) f(\log K_M) d\log K_H d\log K_M.
\]

with the factor in brackets as the kernel. In the general case, multisite binding of metals should be considered, as there is spectroscopic evidence indicating that a fraction of the bound metal is in bidentate or even tetradentate form [56,57]. However, because different metal fractions will be bound

with different stoichiometries, it will be introduced here an average stoichiometric coefficient \( r \) such as
\[
rA^+ + M^{z+} = AM^{z-}.
\]

In this case, considering a Langmuir-type equilibrium for the local isotherm leads to the following equations:
\[
K_H a_M = \frac{q_M}{q_s - q_M - a_M},
\]
\[
K_M a_M = \frac{q_M}{q_s - q_H - a_M}.
\]

The system (20),(21) cannot be solved analytically for \( q_M \) (except for integer \( r \)) which can, instead, be found numerically, and used to find \( f \)
\[
Q_M = \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} q_M(\log K_M) g(\log K_H) f(\log K_M) d\log K_H d\log K_M.\right]
\]

Eq. (19) can be solved for \( f \) with CONTIN, with \( r \) as an external parameter and
\[
F_k = \int_{-\infty}^{\infty} q_M(\log K_M, \log K_H) f(\log K_H) d\log K_H.
\]

as the kernel function. The approach present here bears some resemblance to the conditional affinity spectrum (CAS) method proposed by Garcés and co-workers [31,55] where the CAS is the distribution \( p \) restricted to a fixed \( pH \) (constant \( a_H \)), \( p'(\log K_M) \) (constant). However, the CAS of [55] includes a proton contribution whereas Eq. (17) implies a further separation of the proton affinity spectrum in the absence of \( M \).

### 2.2. Data employed and numerical procedure

The CONTIN program, modified to use Eqs. (7) and (23) as kernels, was used to extract affinity distributions. The statistical weights of experimental data were generally set as all equal, unless otherwise noted. As a first test, synthetic titration curves were used, generated with Eq. (3) and a simple bimodal Gaussian distribution for proton binding; a second bimodal distribution was assumed for metal binding (\( r = 1 \)) and titration curves were simulated with Eq. (16). Different tests to study the effect of experimental errors were also conducted. Following that, titration curves simulated using the ND and the SHM models, at different ionic strengths, were processed with CONTIN and compared. ND simulations were conducted using Kinntburgh’s FIT program version 2.581 [58], and SHM calculations were done using Gustafsson’s Vminteq program [26]. In the case of metals, test with simulated titration curves using a simple bi-Gaussian model were followed by analysis of experimental data. Experimental data were mostly taken from the extensive gathering of Mline et al. [3,14] which is available as Supplementary Information of [3]; it contains acid base titration datasets of a number of fulvic and humic acids, and for titration with different metal cations for some of them. Some datasets were also taken from Fernández et al. [15]. However, the processing with CONTIN to recover the conditional affinity distribution has several requirements (wide \( pH \) or \( pM \) range, relatively high density of data points, constant ionic strength) which are not fulfilled in all cases, so some data could not be processed; moreover, some dataset were discarded when processing due to the presence of large peaks at either end of the calculated distributions which could not be eliminated. All that is especially true for metal binding data, where only metal-PPHA data could be used [7,59]. The datasets selected are summarized in Table 1.
3. Results and discussion

3.1. Proton binding

3.1.1. Test with data simulated with a simple model

Fig. 1 shows the bimodal Gaussian distribution used as the \( g(\log K_H) \) distribution. With this distribution (no electrostatic effects), a simulated titration curve was calculated with Eq. (3). Besides the exact one, other curves were also calculated with \( \pm 1\% \), \( 3\% \), and \( 5\% \) random deviation. Fig. S1 (supplementary information) shows the \( Q_H \) vs \( \text{pH} \) curves calculated with \( Q_S = 1 \). Fig. 2 shows the recovered affinity distributions. The algorithm appears to be quite robust regarding random dispersion, as the original distribution is recovered reasonably well up to \( 5\% \) added random dispersion. Fig. 3, on the other hand, shows the result of changing the \( \text{pH} \) range of the binding curve. For a \( \text{pH} \) range of 4–9 a similar result as for \( \text{pH} \) 3–10 is found, but by further restricting to the range to 5–8 the resulting distribution is severely distorted compared with the original. The number of peaks could not be restricted, even by setting the parameter \( r_{sv,\text{min}} \) as high as \( 10^{11} \). These results are in agreement with previous observations by Ruf and co-workers, that these methods are more sensitive to systematic than random deviations [44,60].

3.1.2. Test with data simulated with the NICA–Donnan model

Using the FIT 2.582 program [58], proton binding curves were simulated with the ND model for different ionic strengths, \( I \) from 0.01 to 1 M, using generic average parameters (default parameters of the FIT program). Also, a curve without electrostatics (obtained by setting the Donnan volume to 0) was simulated for comparison. Fig. 4 shows the normalized conditional distributions recovered with CONTIN, along with the theoretical one, the bimodal Sips distribution. As expected, the distribution deduced without electrostatics is coincident with the original. For the other cases it is observed that, as the ionic strength decreases, the distribution departs from the theoretical one both in position and in shape, shifting to higher \( K_H \) values and becoming broader. This is a consequence of the enhanced adsorption at low \( I \) values due to

![Fig. 1](image1.png)  
**Fig. 1.** Simulated bimodal Gaussian affinity distribution used for tests.

![Fig. 2](image2.png)  
**Fig. 2.** Distributions resulting from simulated binding curves calculated with the distribution shown in Fig. 1, exact and with random errors as indicated.

![Fig. 3](image3.png)  
**Fig. 3.** Distributions obtained for different pH ranges in the simulated binding curves.

### Table 1

Datasets analyzed with CONTIN.

<table>
<thead>
<tr>
<th>Code</th>
<th>Material</th>
<th>Ionic strengths (M)</th>
<th>pH range</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FH-04</td>
<td>Bersho fulvic acid</td>
<td>0.1</td>
<td>3.1–7.2</td>
<td>[3]</td>
</tr>
<tr>
<td>FH-06</td>
<td>Satilla River fulvic acid</td>
<td>0.1</td>
<td>3.2–10.8</td>
<td>[3]</td>
</tr>
<tr>
<td>FH-14</td>
<td>Laurentian soil fulvic acid</td>
<td>0.05</td>
<td>4.2–9.8</td>
<td>[3]</td>
</tr>
<tr>
<td>FH-23</td>
<td>Laurentian soil fulvic acid</td>
<td>0.1</td>
<td>3.6–10.4</td>
<td>[3]</td>
</tr>
<tr>
<td>HH-09</td>
<td>Purified peat humic acid (PPHA)</td>
<td>0.01; 0.08; 0.09; 0.31</td>
<td>3.5–10.5</td>
<td>[3]</td>
</tr>
<tr>
<td>HH-16</td>
<td>Kinosole OH humic acid</td>
<td>0.003; 0.02; 0.12</td>
<td>3.3–10.3</td>
<td>[3]</td>
</tr>
<tr>
<td>HH-18</td>
<td>Purified Aldrich humic acid</td>
<td>0.1</td>
<td>3.4–10.4</td>
<td>[3]</td>
</tr>
<tr>
<td>HH-23</td>
<td>PUHA</td>
<td>0.01; 0.03; 0.1; 0.3</td>
<td>3.0–9.8</td>
<td>[3]</td>
</tr>
<tr>
<td>TolS-FA</td>
<td>Toledo soil FA</td>
<td>0.3</td>
<td>3.5–10.3</td>
<td>[15]</td>
</tr>
<tr>
<td>TolS-HA</td>
<td>Toledo soil HA</td>
<td>0.3</td>
<td>3.6–10.4</td>
<td>[15]</td>
</tr>
<tr>
<td>HC2-03</td>
<td>PPHA + Cd(^{2+})</td>
<td>0.1</td>
<td>4.6; 8</td>
<td>[3]</td>
</tr>
<tr>
<td>HCU-04</td>
<td>PPHA + Cu(^{2+})</td>
<td>0.1</td>
<td>4.6; 8</td>
<td>[3]</td>
</tr>
<tr>
<td>HPb-05</td>
<td>PPHA + Pb(^{2+})</td>
<td>0.1</td>
<td>4.6; 8</td>
<td>[3]</td>
</tr>
</tbody>
</table>
nonspecific (electrostatic) interactions. The shift of the peak positions is linear in $I$ (see Fig. S2, supplementary information). In order to compare with other results later on, it is worth noting that the curves of Fig. 4 can be fitted very satisfactorily with two Sips-shaped peaks (Fig. 5b), whereas fitting to Gaussian peaks are less satisfactory (Fig. 5a). The curve shown is for a high ionic strength, where nonspecific effects are less pronounced.

### 3.1.3. Test with data simulated with the Stockholm humic model

Employing Gustaffson's Vminteq program [26], titration curves with standard parameters were obtained. Fig. 6a shows the exact curve obtained for $I = 1$ M, and the same with added random deviations, whereas in Fig. 6b the deduced distributions are shown. It is observed that a discrete distribution is only recovered using the exact curve; introduction of random errors, even small, causes that a discrete distribution is no longer found, but a continuous one, similar to those predicted by the ND model, is obtained. The same result is found by restricting the regularization parameter $\alpha$ to the maximum value compatible with data fitting (see Section 3.1.4). Furthermore, the exact discrete distribution is not recovered, especially for low $\log K_H$ values, where only three peaks are found. The titration curve should be extended to lower pH values in order to observe the full four $\log K_H$ values.

In Fig. S3 the influence of ionic strength on the distribution found with CONTIN (using the “exact” titration curves) is plotted. The shift of the distribution peaks with $I$ is not linear, at variance with the case of the ND model. This difference is a consequence of the respective electrostatic models.

### 3.1.4. Analysis of experimental data

In the following, we will show the application of CONTIN to experimental data of acid–base titrations (i.e., proton binding) of humic substances, using data published by Milne et al. [3,14] and Fernández et al. [15]. When processing several of these datasets with CONTIN with default conditions, often the resulting distributions show a relatively large number of peaks, which would have little physical justification. Thus, the following procedure was employed, illustrated by Figs. 7 and 8 for the dataset HH-16 from

---

**Fig. 4.** Distribution curves recovered from simulated NICA-Donnan titration curves.

**Fig. 5.** Fitting of the distribution curve, deduced for a NICA-Donnan simulated titration curve with $I = 1$, to Gaussian peaks (a) and Sips peaks (b).

**Fig. 6.** (a) Titration curve simulated with the SHM model, exact (line) and with added random deviations, of 1% and 3% magnitude; (b) the SHM assumed $pK$ discrete distribution (gray bars) and CONTIN obtained distributions.
Milne et al., corresponding to Kinshozan OH humic acid. Fig. 7a shows the fitting obtained (with the residuals plot in the inset) with default conditions, and Fig. 7b the corresponding distribution; this presents 6 peaks, with a predominance of low log $K_H$ values (relatively strong acidity). This distribution is unlikely from a physical point of view. Thus, further restrictions were introduced in the regularization. After several attempts, the following procedure gave consistent results in this and the following cases: the
parameter \( r_{sv,\text{min}} \) (limiting the lowest \( x \) value in the first analysis) was progressively increased, leading to smoother and less featured distributions as shown in Fig. 7c–h. \( r_{sv,\text{min}} < 10^6 \) did not cause significant changes in the results; for \( r_{sv,\text{min}} \) about \( 3.0 \times 10^6 \)–\( 1.0 \times 10^7 \) the distribution obtained shows essentially two peaks, in agreement with the usual assumptions, and still resulting in substantially good fitting. Beyond these values, the distribution shows still two peaks, but broader, and the fitting quality decreases noticeably, as revealed by the residuals plot (Fig. 7g). Table 2 shows the final values found for \( x, P_1 \), and \( \chi^2 \).

Fig. 9a compares the conditional affinity distributions found at different ionic strengths for the purified peat humic acid, and Fig. 9b shows the observed dependence with \( I \) of the two main peaks, compared with the dependence found from titration curves predicted (with optimized parameters for this HA) by the ND and SHM models. It is found that the experimental data show a smaller dependence with the ionic strength than the models, especially for the phenolic peak. Fig. S4 compares the affinity distributions found for several HAs with ND model predictions. Here, a general observation is that the ND model appears to predict less change of the distribution shape on the ionic strength than found from experimental data: the resulting effect in model predictions is mainly the displacement of the distribution curve along the \( \log K \) axis, whereas experimental data result in curves which change both position and shape.

The general shape of the distributions found in Fig. 7f and other datasets examined (later on) is similar to that described in the literature, interpreted in terms of carboxylic sites (lower proton affinity, the higher area peak) and phenolic ones (high proton affinity, the broad peak near \( \log K_H = 9 \)). Thus, one can attempt to deconvolute the peaks accordingly. Fig. 10a shows the deconvolution of the affinity distribution from Fig. 7f in terms of Gaussian functions. It is found that three Gaussian peaks are needed to obtain a satisfactory fit, with a small peak in the middle of the two main ones. Fig. 10b shows the same using Sips function peaks, the function underlying the widely used NICA–Donnan model. It is observed that, even using three peaks, the fitting is noticeably less satisfactory. As it has been discussed by several authors, the difference is probably not important regarding generic data fitting, since only the main moments are important for the description of the titration (binding) curve.

Figs. S5 and S6 show other examples; in several cases, the curve found should be fitted with three peaks whereas in others with two peaks only, but it is generally observed that Gaussian peaks give better fit than Sips ones. Another observation is that in some cases the curves found are similar to those predicted from NICA–Donnan fitting [14] but in other cases there are differences, most notably a higher and narrower phenolic-like site peak.

In the case of fulvic acids (see Figs. S7 and S8), it is noticeable that in most cases three well-differentiated peaks are found, which are satisfactorily fitted with Gaussian functions; when comparing humic and fulvic acids from the same origin, it can be observed in the example shown in Fig. S7 (Toledo soil [15]) that the fulvic acid shows three well-defined peaks whereas the HA fraction shows only two. Furthermore, the examples shown in Fig. S8a and S8b correspond to the same material (Laurential soil FA) reported by different authors; the distributions are qualitatively similar, albeit not quantitatively coincident. It should be noted that the NICA–Donnan parameters found by Milne et al. for these datasets are widely different, so the differences should be attributed to experimental differences between the respective authors, such as FA handling and purifying procedures, and titration technique. It should also be noted that in the case of Fig. S8c, the dataset spans...
only to pH about 7, so that the 2-peak distribution found for FH-04 could be conditioned by that fact.

The middle peak found in several distributions is centered about $pK_a = 6–7$, which lies in the range of the $pK_a$ values for adjacent carboxylic groups (such as phthalic or oxalic acids); a similar result was found by fluorescence measurements [32]. Thus, this finding would indicate that some carboxylic groups are vicinal.

Although the presence of carboxyl groups in humic substances is well established (see, for example, [61]), vicinal carboxyl groups are more difficult to detect. Benzene polycarboxylic groups, detected by analytical pyrolysis, have been reported in a number of publications (for example, see [62]). However, as it has been recently pointed out [63], the subject remains controversial as several studies attribute such findings to artifacts of the methylating pyrolysis procedure [64,65]. On the other hand, the presence of vicinal aliphatic carboxyl groups has been reported in some fulvic acids [8].

3.2. Competitive metal binding

3.2.1. Test with data simulated with a simple model

To deduce the conditional affinity (binding constant) distribution for cation binding, the $H^+$ affinity distribution found at the same ionic strength was employed in Eqs. (17) and (21).

Tests conducted with simulated data were conducted with satisfactory results as shown in the Supplementary Information. Fig. S9a shows (open circles) the simulated constant distribution for metal binding. Binding curves were calculated with Eq. (17) for different $r$ values, using the proton affinity distribution shown in Fig. 1. Following, the distributions were recovered using CON- TIN. Fig. S9a compares the results obtained using the exact simulated curve, with those using a curve with random errors introduced. It is observed that the same shape is obtained, but with some deviations. Fig. S9b shows the influence of the parameter $r$ on the distribution recovered. As observed, there are some differences, mainly in the low affinity part of the distribution; however, it should be noted that the presence of a spurious peak at the low end is not a common case, and can be suppressed by limiting the regularization parameter $\alpha$.

3.2.2. Analysis of experimental data

Literature data of Cd(II), Pb(II), and Cu(II) binding to PPHA are now analyzed, and the results are discussed in comparison with available spectroscopic information on binding of these cations to HS. Although information for exactly the same systems (the same HA) is not available, from a qualitative point of view the analysis is meaningful, as humic substances from different origin show a similar behavior, even those from paleosols [9].

Fig. 11 shows the conditional affinity distributions found for Cd(II) and Pb(II) binding to purified peat humic acid, using data published by Milne et al. [3] from [7,59]. As it is observed, the results are quite similar, showing relatively narrow peaks, the most prominent being that at low log $K_M$ values, with one or two smaller peaks for higher stability binding sites. There is a pH dependence of the distributions, shifting to higher values as the pH increases. The assumption of separability introduced in Eq. (17) implies that $f(\log K_M)$ should be pH independent to be consistent. However, it should be taken into account that those plotted in Fig. 11 are conditional distributions; even when they correspond to the same ionic strength, the HA will have a different charge at the different pHs and so the electrostatic effect will be different. The
displacement observed in Fig. 11 is consistent with the ionic strength effect shown in Fig. 9: as the electrostatic effect becomes more pronounced, the affinity curves shift to higher log K values. These results will be now discussed in terms of the available spectroscopic evidence.

Cd(II) binding to HS has been quite intensely studied [10,16,66–73] and, owing to the relative abundance of the 113Cd isotope, NMR studies have provided important structural information on the binding sites [66,67,74,75]. Otto et al. [66,67] studied Cd(II) binding to fulvic acids using 113Cd chemical shift and relaxation measurements, concluding that there are two clearly differentiated sites: one strong binding site best represented by a polydentate carboxylic model (albeit an oxo-carboxylate site could also be possible), and another weaker site, attributed to monocarboxylate binding. This last type of site was also considered as the more abundant by Grassi and Daquino [74]. Thus, the results of Fig. 11 could, in principle, be interpreted in terms of these two type of sites: a weaker, more abundant binding site with log K_{Cd} about 2–4 and a stronger, less abundant site with log K_{Cd} near 6; however, more evidence is required to confirm this interpretation. There is a third, very small peak observed at pH 8, which could be due to binding to sites with high proton affinity, not observed at lower pH values because its very low occupation by Cd ions would make them not detectable within experimental error. In the case of Pb(II), the curves are similar, albeit only one peak is found at pH 4; again, this can be attributed to the second type of site having higher proton affinity, thus displacing completely Pb(II) ions at pH 4. Lead sorption by humic substances has been also widely studied [11,17,18,31,68,76–83]; however, there is less spectroscopic evidence available. Xia et al. [76] studied the interaction of Pb(II) with a humic extract from a silt loam soil suggesting the presence of two C atoms in the second coordination shell, which would indicate bidentate binding. Thus, an interpretation similar to Cd(II) could in principle be given for lead: the higher peak at low log K_{Pb} values attributed to monodentate binding to carboxylic groups, and the peak at higher log K_{Pb} values to bidentate binding. However, the higher affinity sites could also be of the monodentate type: recently, Puy et al. [31] obtained conditional affinity spectra, based on the fitting parameters to the NICA model (i.e., bimodal Sips distribution), for Pb^{2+} binding to purified Aldrich HA at a fixed ionic strength of 0.1 M and several pH values. The CAS curves found show broadening and a shift to higher log K_{Pb} values as the pH increases. These curves are considerably broader than those found here, and interpreted in terms of two contributions: carboxylic at lower log K_{M} values, having higher area, and phenolic, at higher log K_{M} values but having smaller area. This interpretation could be consistent with that given above if the phenolic types were mainly involved in bidentate binding, in salicylate-type groups.

In Fig. 12 the results found for Cu(II) binding to PPHA at two pHs are shown. Unfortunately, from the data at pH 6 reliable results for the affinity distribution could not be obtained. Here, the behavior is more complex. Also, changing the value of r did not show important differences in the results, only the peaks are better defined with r = 1.5. This value was used because Kimnburgh et al. [6] measured the exchange ratio during the titration, finding values close to 1.5. As Cu(II) is known to bind more strongly that Pb(II) or Cd(II), and in a multidentate way, a more complex behavior is to be expected. There is an important number of studies on Cu(II) binding to HS [9,19,56,59,76,84–86]. In particular, spectroscopic studies show evidence of chelate structures. Senesi and Calderoni [9] studied copper binding to paleosol humic acids; in ESR studies, the spectral parameters are consistent with Cu(II) ions present as inner-sphere complexes of tetragonal—or Jahn-Teller distorted octahedral—configuration, the coordinating sites involving on average three oxygen and one nitrogen atoms. Xia et al. [76] in X-ray absorption studies of Cu(II) binding to HS in the pH range 4–6 found the Cu ions in a tetragonally distorted octahedral environment with four O atoms in the first coordination shell at 0.194 nm and four C atoms at an average distance of 0.313 nm forming the second coordination shell. Karlsson et al. [56] found EXAFS evidence of chelate structures in Cu(II) complexion to soil and natural dissolved organic matter, finding four O/N atoms at about 0.193 nm, a second coordination shell with an average of 2–3.8 C atoms at about 0.28 nm, and a third coordination shell with 2–3.8 O/C atoms at about 0.4 nm. They suggested that Cu(II) is forming one or two 5-membered chelate rings. It should be taken into account in analyzing those results that part of the O atoms found in the first coordination shell could belong to water molecules or, as pointed out by Álvarez-Puebla et al. [84,87], OH− ions. In the CONTIN analysis results of Fig. 12a 3–4 modal distribution is observed, depending on pH; because the complexity of the spectra, and that the above spectroscopic results are for different HS, an assignment is more difficult. However, the affinity spectra show that the more complex behavior of Cu(II) binding, including monodentate and multidentate binding, is reflected.

3.3. Concluding remarks

Conditional affinity distributions were obtained from experimental titration data using a regularization numerical procedure. The results found for proton binding in some cases agree with the bimodal distribution usually assumed, but in several cases, especially for fulvic acids, a third Gaussian peak at pH 6–7 is found.
This is attributed to the presence of carboxylic acid groups on adjacent C atoms. Because electrostatic effects, which are not considered here, should also be present, coupling the present treatment with a suitable electrostatic model is required to separate the effect. Nevertheless, the observed behavior is consistently found at the different ionic strengths examined.

Analysis of cation–HA titration data for PPHA binding Cd(II), Pb(II), and Cu(II) at a relatively high ionic strength, separating the H⁺ affinity spectrum deduced at the same ionic strength, shows results which are, in principle, consistent with literature spectroscopic studies of similar systems. Comparison between titration data analysis and spectroscopic measurements for the same cation–HS system is required for a better understanding of cation binding and confirmation of the affinity distributions found. Again, including a suitable electrostatic model will allow a more quantitative treatment of such distributions.

The regularized regression analysis of CONTIN has been found as a useful tool in the study of binding to humic substances, and suggested the presence of more complex affinity spectra than usually considered for proton and metal cation binding. Coupled with spectroscopic studies and including electrostatic modeling could bring more insight into these important processes in natural systems. Work is in progress in that direction.

Acknowledgments

The authors gratefully acknowledge financial support from the Universidad de Buenos Aires (UBACYT 2004-2007 X105 and 2008-2010 X148), the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, PIP 05216), and the Agencia Nacional de Promoción Científica y Tecnológica (Grant No. 06–12467). E.V.M. is a member of the Carrera del Investigador Científico of CONICET.

Appendix A. Supplementary data


References