Environ. Sci. Technol. **1998**, *32,* 539−548

Application of Maximum Entropy Formalism in the Determination of the Affinity Spectrum in Macromolecular Complexation

JOSEP LLUÍS GARCÉS AND FRANCESC MAS*

Physical Chemistry Department, Facultat de Química, Barcelona University (UB), C/Martí i Franquès, 1, E-08028 Barcelona, Catalonia, Spain

JAUME PUY

Chemistry Department, Escola Tècnica Superior d'Enginyeria Agrària (ETSEA), Lleida University (UdL), Av. Rovira Roure 177, E-25198 Lleida, Catalonia, Spain

Binding of small molecules to heterogeneous ligands and/or adsorption on heterogeneous surfaces with a multiplicity of specific binding sites is of great importance in many fields, ranging from the analysis of interactions between receptors and ligands in biochemistry to environmental sciences. This paper presents the application of the Maximum Entropy (MaxEnt) formalism to the calculation of the Langmuirian affinity distribution function or affinity spectrum, describing the interactions between small molecules and a surface or a macromolecule. Discrete, continuous, and mixed affinity distribution functions are successfully reproduced using a very reduced number of simulated data of coverage versus free complexing agent concentration and without any a priori assumption on the number of kinds of sites. MaxEnt is also applied to experimental data previously reported in the literature. Finally, the advantages and drawbacks of applying MaxEnt formalism to the spectral analysis of coverage data are discussed.

1. Introduction

Heterogeneous complexation gives rise to significant difficulties in the study of the interactions between complexing agents and macromolecules with several kinds of binding sites. Interest in this field is broad, ranging from the analysis of interactions between receptors and ligands in biochemistry (*1*, *2*) to speciation studies in natural waters (*3*). In the latter case, additional complications are derived from the poor characterization of the macromolecules involved (fulvic and humic acids, proteins, polysacharides, cell walls, etc.), which generally have a large number of sites, and from the concurrence of highly distinct phenomena, such as polydispersity, conformational changes, polyelectrolyte effects, or steric effects (*3*). In the limit of very dilute solutions, the study of the interactions between small molecules and a macromolecule can be formulated in most cases as an adsorption problem in which the macromolecule provides the adsorptive surface (*4*, *5*). Thus, the terms heterogeneous adsorption and macromolecular (heterogeneous) complexation are used without distinction in this paper.

Since the pioneering work of R. Sips (*6*), many efforts have been made to describe the properties of adsorptive surfaces using a distribution function of microscopic binding energies. When thermodynamic equilibrium can be assumed, the most relevant information for heterogeneous systems is the affinity distribution function or affinity spectrum $p(k)$, which provides the proportion of sites with an adsorption coefficient laying between k and $k + \mathrm{d}k$. (k is also called the local or intrinsic affinity constant in the context of macromolecular complexation.)

Experimental data usually consist of a list of mean coverages θ (defined as the normalized mean number of small molecules bound to a macromolecule) corresponding to some concentrations c of free small molecules. In dealing with heterogeneity, the mean coverage is usually interpreted as a superposition of a set of local isotherms $f(c, k)$, each corresponding to a binding site, weighted by its probability of occurrence. Therefore, the calculation of the affinity spectrum is equivalent to solving, for $p(k)$, the integral equation

$$\theta(c) = \int_0^\infty p(k) f(c, k)\, \mathrm{d}k \tag{1}$$

Equation 1 is not the most general relationship, because it assumes that all binding sites have the same local isotherm and that they do not interact. Moreover, although the local equilibrium process could be assayed in terms of other isotherms, such as BET and Frumkin isotherms (*7*), the simplest assumption is to postulate a Langmuirian isotherm. So, we restrict our study to this case, i.e., the kernel of the integral eq 1 being

$$f(c, k) = \frac{kc}{1 + kc} \tag{2}$$

If, as is often the case, the other physical phenomena typically involved are not explicitly taken into account, only an apparent affinity spectrum can be determined. In this case, the physical meaning of the affinity spectrum $p(k)$ is usually lost. It could even be impossible to associate an affinity spectrum to some adsorption data, as Merz showed working with pyridine (*7*). Nevertheless, affinity spectra can be extremely useful in understanding the nature of the interactions between small molecules and a surface or a macromolecule in many systems.

Several methods have been proposed for solving the integral eq 1. The simplest way is to use an analytical function that reasonably fits the experimental coverage data and then to invert eq 1 using Stieljes integral transformation. This procedure, used by Sips (*6*), has been followed to obtain analytical $p(k)$ expressions for different heterogeneous systems (*8*). Although this is a rapid way of evaluating the heterogeneity of the system being studied, the assumption of an a priori expression for the coverage is equivalent to the fixing of the interaction model, so that some information about these interactions can be lost, despite experimental data being reasonably fitted.

Another family of methods seeks the analytical inversion of eq 1 after some approximations. The Local Isotherm Approximation (LIA) methods (*9*−*12*) are based on replacing the kernel of the integral eq 1 with a suitable function to allow an approximate analytical solution to be found. Depending on the approximation made to the kernel (eq 2), different methods arise: the CA (Condensation Approxima-

* To whom correspondence should be addressed. E-mail: francesc@daphne.qf.ub.es; fax: 34-3-402 12 31.

S0013-936X(97)00468-9 CCC: $15.00
Published on Web 01/08/1998

tion) method, based on replacing the Langmuirian local isotherm (eq 2) by a step function; the ACA (Asymptotically Correct Approximation) method that replaces eq 2 by a local linear isotherm with a slope equal to the initial slope of the Langmuirian isotherm; the LOGA (Logarithmic Symmetrical Approximation) method in which eq 2 is replaced by a potential function whose parameters are fitted to a Langmuirian isotherm. The Differential Equilibrium Function (DEF) (*3*, *13*) method could also be considered as a particular case of the LIA method, as Nederlof and co-workers showed (*14*). It is especially suitable in cases involving high stability constants with a very low probability (*3*). A recognized advantage in using such methods is their local character, i.e., the knowledge of total coverage versus free metal concentration within a certain concentration range allows the $p(k)$ involved in this range to be determined. A second advantage is its simplicity and the clear assumption that the local binding function is only approximated. Thus, even when no information about the local isotherm is available, valuable information on the apparent affinity distribution could easily be obtained. Major limitations arise from the introduction of an a priori approximation in eq 1 and from the need of performing numerical differentiation of the experimental data.

Finally, several numerical methods have also been used to solve the integral eq 1 without requiring any approximation of the kernel. This integral equation is ill-posed, i.e., a small change in the data (as produced, for instance, by experimental error) will produce a large error in the solution (*7*). Numerical methods designed to solve such equations try to overcome this problem by using additional conditions usually derived from some prior knowledge of the nature of the solution. Among these, linear regularization methods have been successfully applied to adsorption problems (*15*−*17*). These methods need coverage data for a very large range of concentrations, thus preventing their widespread use (*10*, *17*).

Maximum Entropy (MaxEnt) formalism is proposed here as a complement to the more classical methods mentioned above for the determination of the affinity distribution function $p(k)$ from coverage data. MaxEnt is specially designed to solve inverse problems (*18*) in which the unknown object is a probability distribution function. MaxEnt formalism can be considered as a method of statistical induction (the so-called Bayesian interpretation) but also as a nonlinear regularization method (*18*, *19*). Although an interesting debate can be found in the literature regarding the respective merits of interpretation (*19*), here we have chosen to use both so as to facilitate understanding of the results. It has been claimed that MaxEnt requires fewer data to obtain a better resolution of the probability distribution that is sought (*20*). This enhanced resolution power, in comparison with other methods, seems to be related to the nonlinear nature of the equation (*19*).

MaxEnt has been successfully applied in very different fields, such as crystallography (to determine the electron density in crystallized macromolecules using X-ray diffraction data) (*20*), in the deconvolution of Scanning Tunneling Microscopy (STM) images (*21*, *22*), and in the interpretation of light scattering experiments (*23*). Moreover, in the study of macromolecule−small molecule dissociation kinetics, MaxEnt has been succesful in obtaining the distribution of dissociation constants using fluorescence amplitude data (*24*−*27*). MaxEnt has also been used to obtain the two-dimensional distributions of activation enthalpies and entropies from chemical kinetics results (*28*) and in the deconvolution of dielectric and impedance data in electrodic processes (*29*).

To facilitate understanding, section 2 of this paper includes a short introduction to MaxEnt formalism, together with its implementation in adsorption phenomena. In section 3, the method is tested with different sorts of affinity spectra: discrete, continuous affinity spectra, and combinations of the two. The results obtained in each case are fully discussed. Section 4 applies the method to a set of experimental data concerning the complexation of Cu(II) with chestnut leaf litter extract, previously reported in the literature (*30*). Finally, section 5 discusses the application of MaxEnt in heterogeneity studies.

2. Application of Maxent Formalism in Obtaining the Affinity Spectrum

2.1. Operational Formulation of the Integral Equation for the Affinity Distribution Function. Let us consider the experimental system (macromolecule and complexing agent) as a "black box" attaining a value of the coverage, θ, for a given value, c, of the free complexing agent in the solution. Let $\{c_i, \theta_{\mathrm{exp},\,i};\ i = 1, ..., n\}$ be the experimental data set.

The first problem encountered in applying numerical methods to solve eq 1 is that the variable k runs very different orders of magnitude. In order to overcome this problem, we use the changes of variables $x \equiv \log k$ and $\mu \equiv \log c$, and we assume that $p(x)$ takes non-null values only in some restricted integration domain $[a, b]$, being $a \equiv \log k_{\min}$ and $b \equiv \log k_{\max}$, defined by the experimental window. Then, for each experimental value, the integral eq 1, written as a superposition of local Langmuirian isotherms, becomes

$$\int_a^b p(x)\,\frac{1}{1+10^{-x-\mu_i}}\,\mathrm{d}x = \theta_{\mathrm{exp},\,i} \qquad i = 1, ..., n \quad (3)$$

where $\mu_i = \log c_i$.

We denote the cumulative distribution function by

$$F(x) \equiv \int_{-\infty}^{x} p(x')\,\mathrm{d}x' \quad (4)$$

Although $p(x)$ is implicitely considered as a continuous function in eq 4, discrete spectra can easily be included if $p(x)$ includes Dirac delta functions

$$p(x) = \sum_{i=1}^{m} p_i\,\delta(x - \log(k_i)) \quad (5)$$

where $\{p_i\}$ $(i = 1, ..., m)$ are the weights of the m different intrinsic stability constants involved. In this case, the cumulative distribution function, $F(x)$, consists of a set of step functions.

2.2. MaxEnt Formalism. Let us consider that each macromolecule has a large total number, N, of coordinating sites, and let n_i be the number of sites having the same intrinsic stability constant, k_i. The distribution $\{p_i\}$ can be written as

$$p_i = \frac{n_i}{N}\ (i = 1, ..., m) \qquad \sum_{i=1}^{m} n_i = N \quad (6)$$

The MaxEnt formalism tries to find the most probable distribution $\{p_i\}$ of sites, i.e., the most probable set $\{n_i\}$, compatible with the set of the n experimental data obtained. If $\{p_i\}$ is the so-called a priori distribution, where q_i indicates the a priori probability of a given type i of sites, the probability of finding a given configuration $\{n_j\}$ of the N sites among all the possible sets is given by the polynomial distribution

$$P[\{n_j\}] = \frac{N!}{n_1!\ldots n_m!}\prod_{i=1}^{m}(q_i)^{n_i} \quad (7)$$

Taking logarithms and using the Stirling approximation, eq 7 becomes

$$\ln P[\{n_i\}] = -N\sum_{i=1}^{m} p_i \ln \frac{p_i}{q_i} \tag{8}$$

The Shannon entropy (*18*, *31*) of the distribution $\{p_i\}$ is defined as: $S[\{p_i\}] \equiv \ln P[\{n_i\}]/N$. Extending the MaxEnt formalism to continuous distribution functions defined in a domain $[a,b]$, the problem reduces to the computation of the maximum of the functional

$$S[p(x)] \equiv -\int_a^b p(x) \ln \frac{p(x)}{q(x)}\,\mathrm{d}x \qquad \int_a^b p(x)\,\mathrm{d}x = 1 \tag{9}$$

subject to the fulfilment of the n constraints given by (eq 3), and to the normalization condition for $p(x)$.

This problem is equivalent to a restricted maximization that can be solved applying the method of undetermined Lagrange multipliers. For the Langmuirian heterogeneous complexation, this functional reads

$$\Phi[p] \equiv -\int_a^b p(x) \ln \frac{p(x)}{q(x)}\,\mathrm{d}x - \sum_{j=1}^{n}\lambda_j\left(\int_a^b p(x)\frac{\mathrm{d}x}{1+10^{-x-\mu_j}} - \theta_{\mathrm{exp},\,j}\right) - \lambda_0\left(\int_a^b p(x)\,\mathrm{d}x - 1\right) \tag{10}$$

where $\{\lambda_0, \lambda_1, ..., \lambda_n\}$ are the $n+1$ undetermined Lagrange multipliers, each associated to one constraint.

The role of the entropy functional can be better understood if the maximum of $S[\{p_i\}]$ is calculated without constraints. In such a case one obtains

$$\frac{\delta S[p(x)]}{\delta p(x)} = 0 \Rightarrow p(x) = q(x) \tag{11}$$

that is, $p(x)$ equals the a priori distribution $q(x)$ if no experimental data are taken into account. So, the entropy functional (eq 9) is a useful measure of the difference between the probability functions $p(x)$ and $q(x)$. If the set of constraints (eq 3) is taken into account, the distribution $p(x)$ obtained by maximizing eq 9 will be as similar as possible to $q(x)$ while constraints (eq 2) are satisfied. As a consequence, $p(x)$ will deviate from $q(x)$ just to reproduce the data, and no more will be added to $q(x)$ if it is not necessary. Then, if $q(x)$ is chosen equiprobable, i.e., $q(x) = 1/(b-a)$, $p(x)$ will be as flat as possible. In this sense, MaxEnt can also be understood as a nonlinear regularization method resulting in a solution that is most similar to the a priori solution and especially designed to avoid spurious peaks due to experimental errors (*32*).

The condition of maximum is given by the functional equation $\{\delta\Phi[\mathrm{p}]\}/\{\delta p(x)\} = 0$, which, for eq 10, leads to

$$p(x) = \frac{q(x)\exp\left(-\sum_{j=1}^{n}\frac{\lambda_j}{1+10^{-x-\mu_j}}\right)}{Z} \tag{12}$$

where $Z = Z[\lambda_1, ..., \lambda_n]$ is also called the partition function and can be identified with

$$Z[\lambda_1, ..., \lambda_n] \equiv \mathrm{e}^{\lambda_0+1} = \int_a^b q(x)\exp\left(-\sum_{j=1}^{n}\frac{\lambda_j}{1+10^{-x-\mu_j}}\right)\mathrm{d}x \tag{13}$$

as is easily seen using the normalization condition for $p(x)$. This equation reduces the number of undetermined Lagrange multipliers to n, because λ_0 can be expressed in terms of the other multipliers.

Note that the exponential involved in eq 12 ensures the positivity of $p(x)$. In order to find the undetermined Lagrange multipliers λ_i, the set of constraints (eq 3) must be fulfilled

$$\int_a^b q(x)\left\{\frac{\exp\left(-\sum_{j=1}^{n}\frac{\lambda_j}{1+10^{-x-\mu_j}}\right)}{Z[\lambda_1, ..., \lambda_n]}\right\}\left(\frac{1}{1+10^{-x-\mu_i}}\right)\mathrm{d}x = \theta_{\mathrm{exp},i} \qquad i=1, ...,n \tag{14}$$

that form a system of n nonlinear equations for the n unknowns $\{\lambda_i\}$. Once the set $\{\lambda_i\}$ is obtained by numerically solving eq 14, $p(x)$ is obtained by replacing $\{\lambda_i\}$ in eq 12.

2.3. Computational Details. A multi-dimensional Newton−Raphson method has been used to solve eq 14. Iterations proceed until the Lagrange multipliers, replaced in the set of constraints (eq 3), reproduce the experimental data within the required accuracy. Integrals involved in the calculus of Z and its derivatives have been computed in the following way: values between $-\log c_{\min} - \Delta$ and $-\log c_{\max} + \Delta$, with Δ high enough to ensure that $p(k)$ has fallen to zero, are used for the integration limits a and b of $\log k$ (a wider interval of integration does not change the results); a number of 500−800 intervals for the $\log k$ unit is used to discretize the integrals involved in eq 14 (a large number of intervals are necessary to reproduce narrow peaks as those arising from discrete spectra). Finally, a Gaussian quadrature method of 10 points has been used to compute the integral within each integration interval.

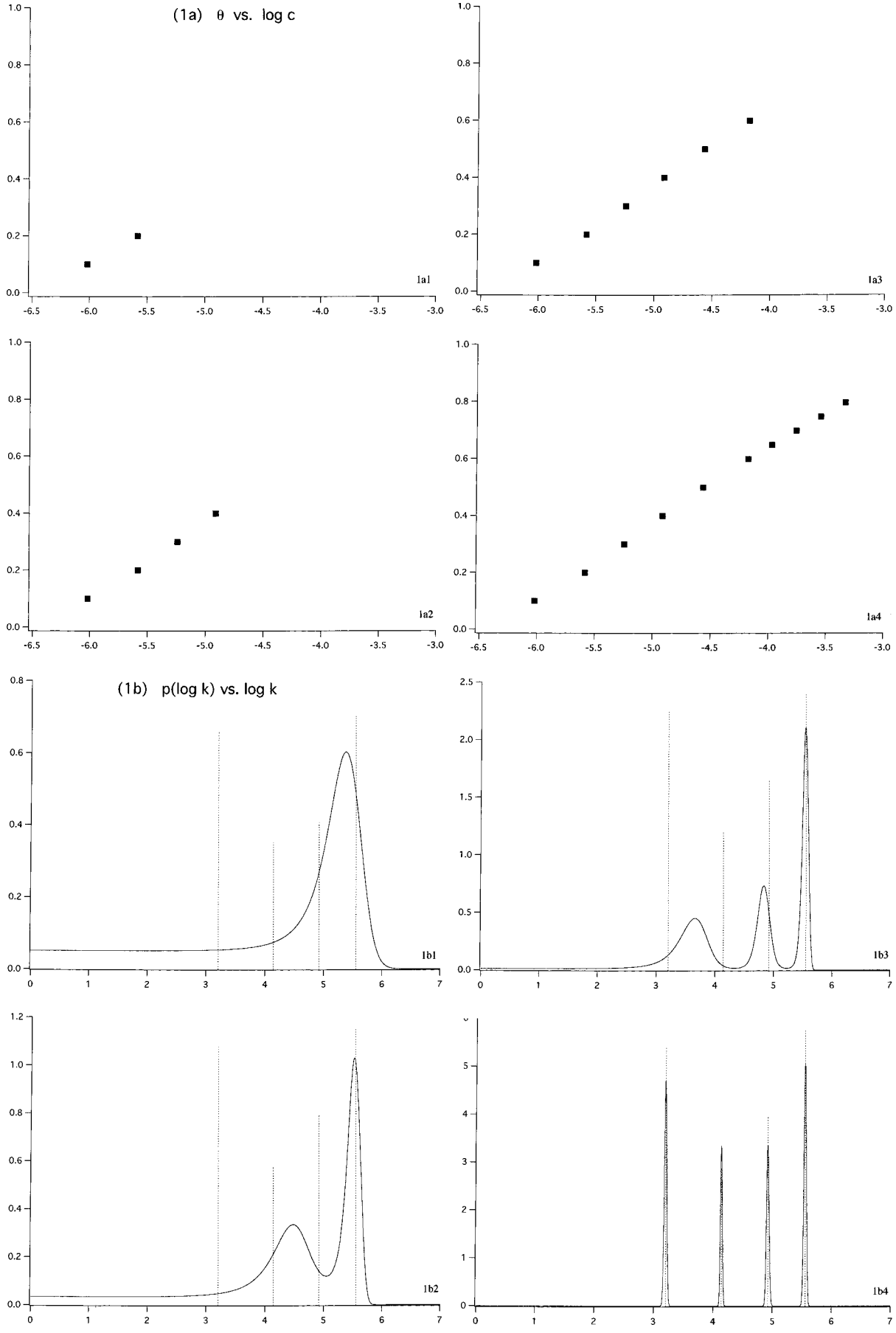
3. Application to Simulated Cases

In order to test the application of the MaxEnt formalism in the interpretation of coverage data, several simulations with discrete, continuous, and mixed (composed of both discrete and continuous) spectra were performed. Exact simulated data of total coverage vs free concentration were further used as experimental data to assess the goodness of recovering the affinity spectrum with the MaxEnt procedure. Direct extrapolation of these results to those obtained from experimental values requires a certain degree of caution as the real data may not be sufficiently accurate. In this case, a smoothing process should previously be applied. Koopal and co-workers have developed a spline smoothing method specially designed for this problem (*33*).

3.1. Discrete Affinity Distribution Functions. In many experimental systems, such as proteins or enzymes, specific interactions between the macromolecule and small molecules seem to be determining factors in their biochemical properties. Binding data are sometimes interpreted in terms of a weighted addition of Langmuirian isotherms, as in eq 5, which are derived from the assumption of a discrete affinity distribution function. In such cases, the most common method used to obtain the stability constants, together with their corresponding weights, is based on the Scatchard, Ruzic, or Hill plots (*1*, *2*, *34*, *35*). Geometrical and functional properties of such plots have been widely studied (*36*).

Continuous analytical methods have been also used. Although some advantages have been recognized, difficulties are still found in the detection of close peaks in the affinity spectrum and in the computation of derivatives from experimental data.

As the selection of data is arbitrary, we will use MaxEnt to obtain weights and stability constants included in an example that was previously selected by Gamble and Langford (*37*) and Buffle et al. (*38*) to illustrate the DEF method and by Koopal et al. (*14*) to compare semianalytical methods


(1a) θ vs. log c
1a1
1a3
1a2
1a4
(1b) p(log k) vs. log k
1b1
1b3
1b2
1b4

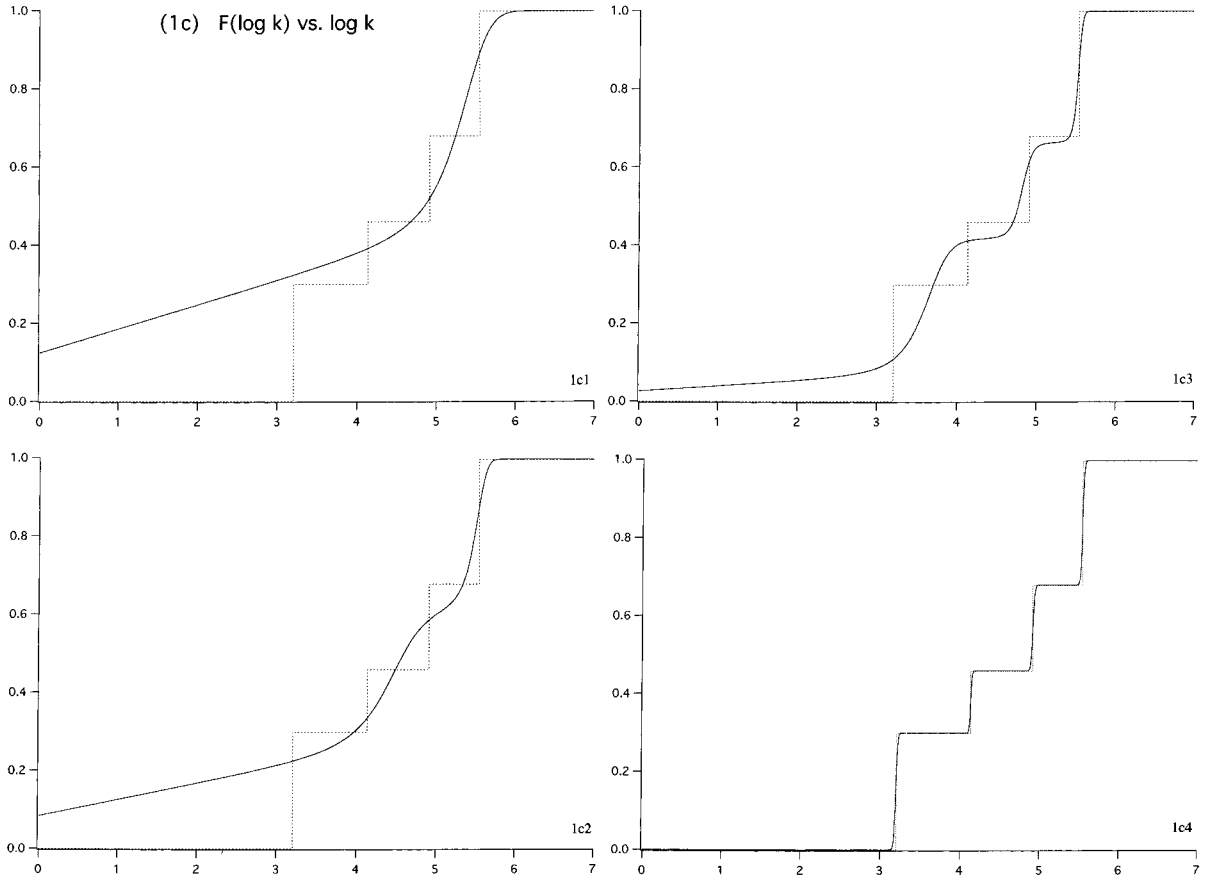


FIGURE 1. Discrete affinity spectrum corresponding to the following intrinsic stability constants and respective weights: log $k_1 = 5.55$, $p_1 = 0.32$; log $k_2 = 4.92$, $p_2 = 0.22$; log $k_3 = 4.14$, $p_3 = 0.16$; log $k_4 = 3.20$, $p_4 = 0.30$. Panels labeled "a" show simulated mean coverage data versus the logarithm of the free complexing agent concentration. Panels labeled "b" show p(log k) versus log k. Panels labeled "c" show F(log k) versus log k (units of c and k are M and M^{-1} units, respectively). Number of simulated data of mean coverage used in the recovering of the spectrum: 2 data (panels labeled 1); 4 data (panels labeled 2); 6 data (panels labeled 3); and 10 data (panels labeled 4). Dotted lines (···) indicate, in panels b, the exact k values with a height proportional to the probability of each kind of site. The normalization factor is arbitrary and is not relevant. In panels c, the normalization factor is the usual one, F (log (k_{max})) = 1.

with the DEF method. In the application of MaxEnt, special attention is devoted to analyze the use of a very reduced number of experimental data.

Figure 1 shows the improvement of the fitting when the number of simulated data in a system with four kinds of sites is increased. Figure 1a shows the data of mean coverage (θ) vs the logarithm of the free concentration of the complexing agent used in the MaxEnt procedure to obtain the affinity spectrum, and Figure 1b,c shows the corresponding results in terms of p(log k) and F(log k) versus log k, respectively. The third index used in the identification of the figures refers to the number of coverage data used (*2*, *4*, *6*, *10*).

Only two data for low values of the coverage were used (Figure 1a1) to obtain the results plotted in Figures 1b1 and 1c1, in which the highest affinity peak can be clearly detected. However, still nothing can be said about the rest of the spectrum, because for low values of the concentration c only the highest affinity sites are significantly occupied. In other words, these two data contain information about the highest affinity sites only, and therefore, the low affinity part of the spectrum appears flat. If the number of data is increased and data for higher coverages are used, the low-intensity part of the affinity spectrum begins to manifest itself because such points contain information about the stability constants of lower value. In Figures 1b2 and 1c2, which use four data (see Figure 1a2), the highest affinity peak is more clearly defined (log $k_1 = 5.55$, $p_1 = 0.32$), and a new peak at a position that is an average of the remainig intrinsic stability constants appears. An analogous behavior is seen in Figure 1b3 for six data. For 10 data (Figure 1a4), the four peaks of the spectrum with their respective weights appear clearly defined [(log $k_2 = 4.92$, $p_1 = 0.22$), (log $k_3 = 4.14$, $p_3 = 0.16$), and (log $k_4 = 3.20$ $p_4 = 0.30$)], and clearly tend to Dirac delta functions as well as F(log k) to step functions. The resolution of the spectrum is also remarkable despite the peaks being separated by less than 1 log k unit. In conclusion, MaxEnt provides a very high resolutive method, even when a low number of experimental data are used.

3.2. Continuous Affinity Distribution Functions. Continuous spectra are usually associated with the so-called secondary effects occurring as a result of the interactions between a small molecule and a macromolecule. Such effects can be produced by very different phenomena, such as the repulsive interaction between neighboring occupied sites (negative cooperativity), steric effects, chelate complexation, etc. (*3*, *5*). Although the resulting spectra have lost their physical meaning (*5*), it can be useful to know the characteristics of the complexation. Macromolecules exemplifying this kind of affinity spectrum are homopolysaccharides and peptides with a large number of sites (*3*).

As in the preceeding section, simulated data of mean coverage vs free concentration of complexing agent, corresponding to an affinity spectrum consisting in two Gaussian functions with centers differing in 3 log k units, were used as experimental data to test the ability of the MaxEnt procedure to obtain a continuous spectrum. Figure 2b,c shows a good repoduction of the theoretical spectrum when the 10 coverage data depicted in Figure 2a are used in the

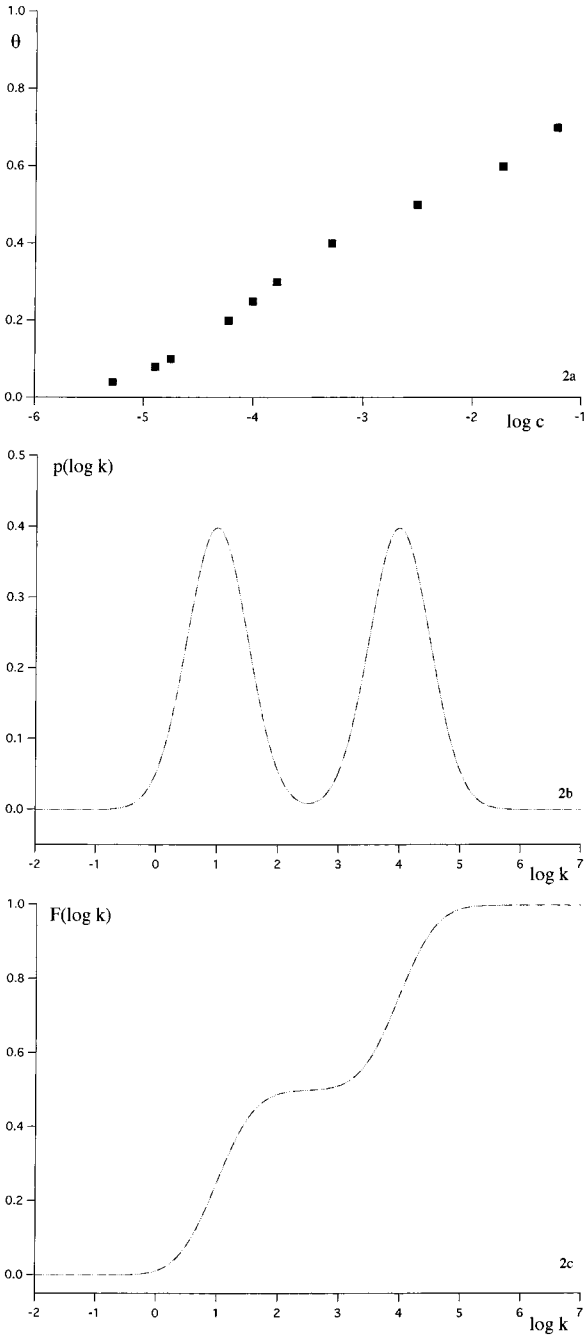


FIGURE 2. Continuous affinity spectrum. A bi-Gaussian distribution has been used to simulate the mean coverage data. The used parameters are as follows: $k_1 = 10^4$ M^{-1}, $\sigma_1 = 0.5$ M^{-1}, $p_1 = 0.5$; $k_2 = 1$ M^{-1}, $\sigma_2 = 0.5$ M^{-1}, $p_2 = 0.5$. The labeling of panels follows the same conventions as in Figure 1. In panels b and c, the dotted line (···) corresponds to the exact solution; the discontinuous line (− −) corresponds to MaxEnt solution.

optimizing procedure. Although not detailed here for the sake of brevity, an increasing number of data improves the fit.

3.3. Affinity Distribution Functions with Continuous and Discrete Components. The affinity distribution function can also exhibit both a continuous and a discrete component for a given macromolecule. This is the case of macromolecules or surfaces having specific and non-interacting sites together with other sites responsible for the continuous part of the spectrum. Usually, the latter type of site has lower affinity for the complexing agent. In biochemical studies, "non-specific binding" usually refers to this part of the spectrum. In natural waters, the continuous part of the spectrum is of less importance, provided that the concentration of metals available for complexation is low enough to consider only high affinity sites (*3*).

Figure 3 shows the affinity spectra obtained by the MaxEnt procedure using the coverage data depicted in Figure 3a1 (10 data) and Figure 3a2 (22 data). The spectrum has a discrete part made up by three kinds of site with peaks at $k_1 = 10^4$ M^{-1}, $k_2 = 10^5$ M^{-1}, $k_3 = 10^6$ M^{-1}. In order to test the resolution power of MaxEnt, very small weights were chosen for these stability constants ($p_1 = 0.05$, $p_2 = 0.01$, and $p_3 = 0.005$). Despite the low weights involved, this part of the spectrum is often responsible for the most interesting environmental properties of the system. The largest part of the spectra (with a weight of 0.935) consists of a continuous Sips distribution function (*39*) with $m = 0.75$ and mean value $\bar{k} = 1$ M^{-1}.

As can be seen in Figure 3, MaxEnt produces a very high resolution of the three discrete peaks. Relevant information for the discrete part is contained in very low coverage data, as is to be expected from the values of both the discrete and continuous parts of the theoretical spectrum. Hence, using data from Figure 3a1 (10 values), the high region of the affinity spectrum is clearly defined (Figures 3b1 and 3c1). If higher coverages are included (Figure 3a2), the continuous spectrum also appears. Figure 4a,b magnifies the high affinity region of the spectrum, showing that MaxEnt accurately reproduces the expected value of both weights and positions of the peaks. It should be emphasized that the resolution of the discrete peaks is very good (very high and narrow peaks) even when the weights of these peaks are very small and they are separated by only 1 unit of log k.

However, the case of a mixed spectrum is the most subtle one from a numerical point of view. The main critical region lies at the intersection of the discrete and the continuous components, since spurious peaks can appear unless many experimental data are taken into account. Using 22 coverage data, small oscillations can be observed in that part of the spectrum close to the discrete part (Figures 3b2 and 4a). Nevertheless, $F(\log k)$ presents good behavior (Figures 3c2 and 4b), since the summation involved in the computation of $F(\log k)$ mitigates oscillations. Generally, these problems appear when highly dispersed data are used, i.e., when a relevant number of data are missing. In this case, the number of local Langmuirian isotherms involved in the expression for $p(x)$ (5) cannot reproduce the real spectrum.

4. Application of MaxEnt to Experimental Data

Analysis of the Complexation of Cu(II) in Organic Dissolved Matter. The complexation of Cu(II) by the chestnut leaf litter extract (LLE) was studied by J. Lüster and co-workers at some pH values between 4 and 6 (*30*). Based on previously available information of this system, these authors assumed two kinds of independent binding sites. The first corresponds to long-chain aliphatic amino acids such as lysine (ligands L1, complexes Cu(L1)) and the second corresponds to saturated aliphatic carboxylic acids such as oxalic acid (ligands L2 and complexes Cu(L2)). Thus

$$\theta(c) = \frac{C_1}{C_1 + C_2}\frac{k_1 c}{1 + k_1 c} + \frac{C_2}{C_1 + C_2}\frac{k_2 c}{1 + k_2 c} \qquad (15)$$

where C_1 and C_2 are the binding capacities corresponding to ligands L1 and L2. Therefore, a discrete affinity spectrum is proposed for these data. The values obtained by these authors at pH = 6 are $C_1 = 0.132$ mol/(kg of C) and $C_2 = 2.72$ mol/(kg of C) for the complexing capacities and $\log k_1 = 8.19$ and $\log k_2 = 6.06$ for the stability constants. So, the weights of the respective kinds of sites, given by $p_i = C_i/(C_1 + C_2)$;

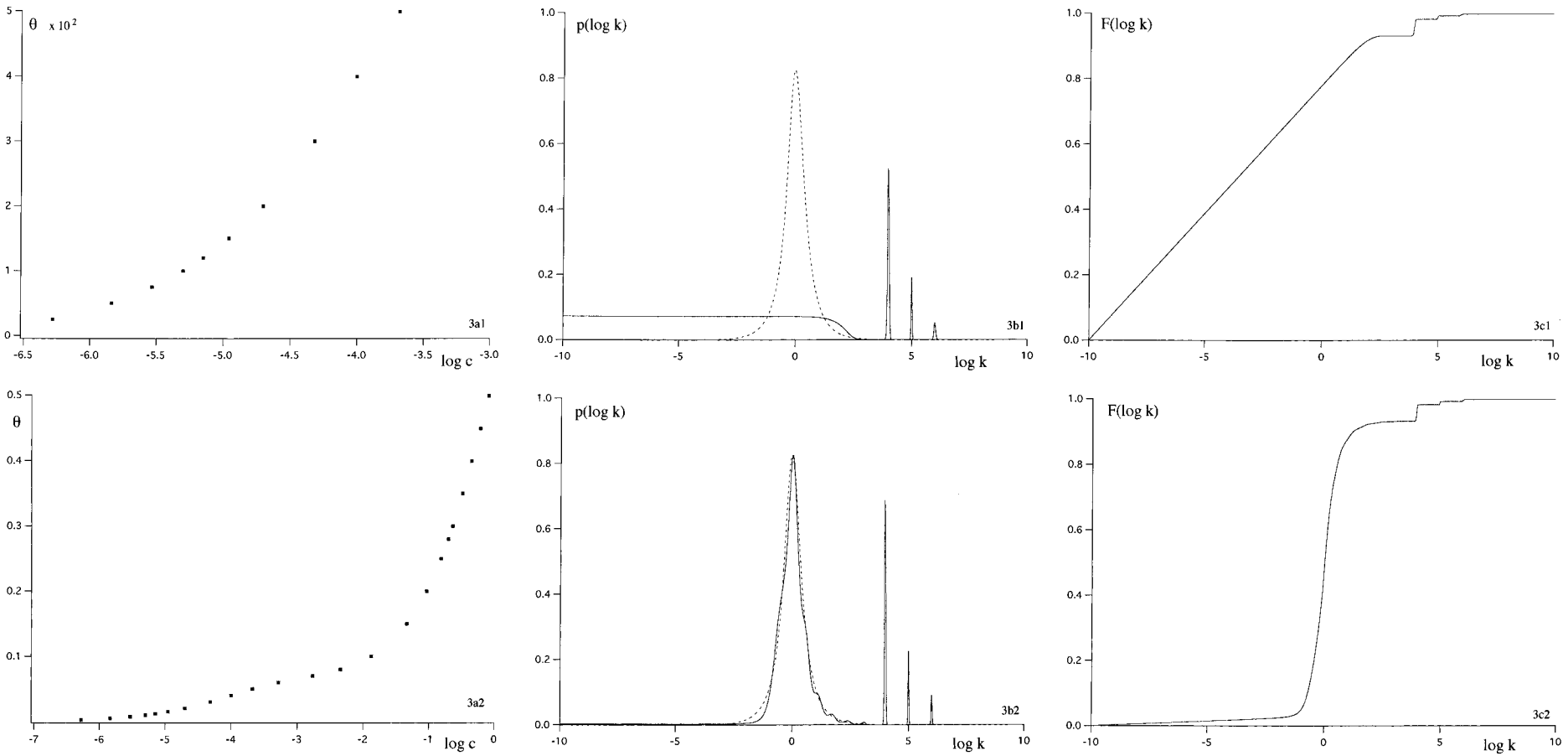


FIGURE 3. Affinity distribution function with continuous and discrete parts. Parameters used are log $k_1 = 4$, $p_1 = 0.05$; log $k_2 = 5$, $p_2 = 0.01$; log $k_3 = 6$, $p_3 = 0.005$ for the discrete part at the high affinity region. For the continuous (and low) affinity part, a Sips distribution with m = 0.75 and k = 1 M^{-1} has been used. The labeling of panels is the same as in Figure 1. Number of simulated data of mean coverage used in the recovering of the spectrum: 10 data (panels labeled 1) and 22 data (panels labeled 2). In panels b and c, the discontinuous line corresponds to the exact solution, and the continuous line corresponds to MaxEnt solution. Height of discrete *k* values is normalized as in Figure 1.

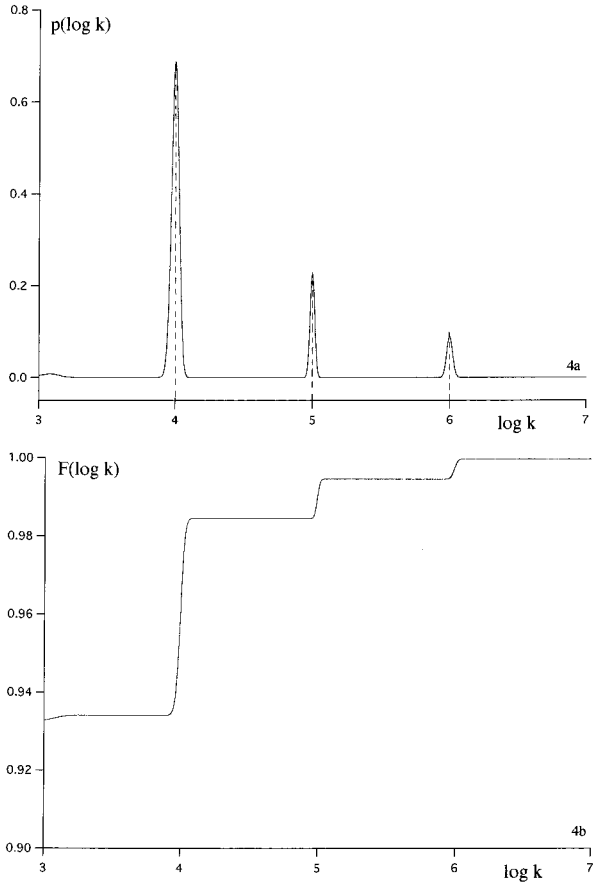


FIGURE 4. Detail of the discrete part of the spectrum plotted in Figure 3. Panel a shows an amplification of *p*(log *k*) versus log *k* from Figure 3b2; panel b shows an amplification of *F* (log *k*) versus log *k* from Figure 3c2.

where $i = 1$ and 2, reach the values $p_1 = 0.0462$ and $p_2 = 0.9537$. Figure 5a shows some experimental data of coverage versus free copper concentration obtained from Figure 4 of ref 30.

MaxEnt formalism was applied to these data to obtain the affinity spectrum corresponding to pH = 6, as an example. As mentioned earlier, the present development of this method does not require any assumption regarding the number of sites types or the continuous or discrete character of the affinity spectrum. To obtain the total binding capacity, an extrapolation of the data of high coverages was carried out. The affinity spectrum and the cumulative distribution function obtained are presented in Figure 5b,c respectively. The two peaks predicted by the authors are clearly reproduced using 6 data (log $k_1 = 8.13$, $p_1 = 0.037$; log $k_2 = 6.15$, $p_2 = 0.963$), although the spectrum obtained by using 4 or 5 data does not differ essentially.

5. Use of MaxEnt in Heterogeneous Complexation: A Discussion

The concentrations of maximum variation of the kernel (eq 2), for a given stability constant k, are those around $kc \sim 1$, that is, $c \sim 1/k$. Hence, the values of the affinity distribution function close to a certain value k are mainly dependent on coverage data corresponding to concentrations near $c \sim 1/k$, this conferring a certain local character to the fitting of the affinity spectrum. Oscillations in the distribution function can appear in those parts of the spectrum not included in the data set, as was shown in Figure 3b2.

The following empirical rule can be used to estimate the minimum number N of data necessary to obtain a good description of the affinity spectrum, using the MaxEnt procedure developed here

$$N \approx M \ln M \tag{16}$$

where M is the total number of parameters involved in the affinity distribution function. The number of data predicted by eq 16 corresponds to the number of data used for a good reproduction of the affinity spectrum in exemples presented above. As has been reported (*18*), the small number of data required by the MaxEnt method for a good fit of the affinity spectrum is related to the nonlinearity involved in the definition of the maximum entropy functional, eq 9. The formulation of a general regularization procedure for sufficiently accurate experimental data is based on maximizing

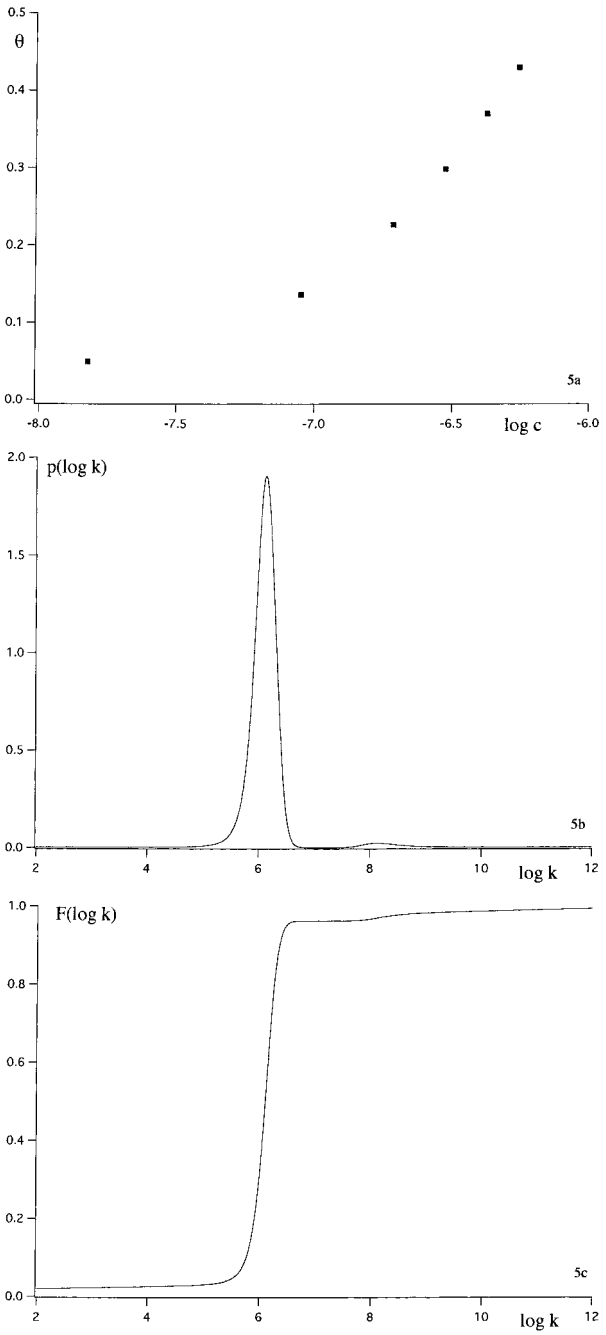


FIGURE 5. Affinity distribution function corresponding to the experimental complexation of Cu(II)−LLE. Mean coverage data used are taken from Figure 4 of ref 30. Labeling of panels is the same as in Figure 1.

a functional of the sort

$$H[p(k)] - \sum_{i=1}^{n}\lambda_i\left(\int_a^b f(x,c_i)p(x)\,dx - \theta_{\exp,i}\right) \qquad (17)$$

where $H[p(k)]$ is a quadratic functional in the linear regularization methods (*7*, *18*) and the functional given by eq 9 in the MaxEnt method. The conditions of maximum on eq 17 ensure that the resulting $p(k)$ fits the experimental data exactly (which we have assumed as being accurate enough in this formulation), and through functional differentiation with respect to $p(x)$

$$\frac{\delta H[p(x)]}{\delta p(x)} + \sum_{i=1}^{n}\lambda_i f(x,c_i) = 0 \Rightarrow p(x) = F\left(\sum_{i=1}^{n}\lambda_i f(x,c_i)\right) \qquad (18)$$

Equation 18 shows that if F is linear (as occurs when $H[p(x)]$ is a quadratic functional), then the solution $p(x)$ contains only a linear combination of the basis functions $f(x,c_j)$, corresponding to actual measurements $j = i$. This is equivalent to setting the contribution of the unmeasured concentrations equal to zero. Nevertheless, F is an exponential function in MaxEnt, and the expression of F in terms of a linear combination of the basis functions $f(x,c_j)$ will involve any concentration. That is, the nonlinearity acts to 'create' non-zero values for the unmeasured concentration, so as to suppress the low intensity ripple and sharpen the point sources. In this sense, it has been said that the nonlinearity acts to take maximum advantage of the information contained in each experimental datum (*19*).

Nevertheless, in order to achieve a wider application of MaxEnt, further work is currently being undertaken to include the total binding capacity as an unknown and to rewrite the functional to be maximized together with the corresponding numerical procedure to treat data within the usual experimental error.

6. Conclusion

When accurate experimental data are available, the MaxEnt formalism can be used to obtain the affinity distribution function that describes the heterogeneous adsorption onto a surface or the heterogeneous complexation of a small complexing agent to a macromolecule.

MaxEnt can have a stronger resolution power than other numerical techniques used to solve this problem. The examples analyzed here show that only a few number of data are required to obtain a good resolution of the affinity spectrum. It is shown that the major influence in obtaining $p(k)$ around some k values comes from the concrete range of concentration $c \sim 1/k$, but the nonlinearity involved in MaxEnt increases the interval of the spectrum obtained from a fixed range of the binding curve.

MaxEnt does not require any a priori assumption about the number of site types or about the continuous or discrete character of the affinity spectrum. On the other hand, the positivity of $p(k)$ is guaranteed by the concrete expressions of the method. The resulting spectrum is the most similar one to that introduced a priori, while fulfilling the experimental information introduced. Thus, the effect of spurious peaks, due to accidental errors in the data, can be mitigated. Finally, no computation of derivatives from the experimental data is required in MaxEnt procedure.

Acknowledgments

One of us (J.L.G.) benefited from a grant from the Generalitat de Catalunya (Catalonia, Spain). The authors gratefully acknowledge the support granted to their research by the Spanish Ministry of Education and Science (DGICYT: Projects PB 93-0641 and PB96-0379), by the Generalitat de Catalunya, and by the Ajuntament de Lleida.

Literature Cited

(1) Cantor, C. R.; Schimel, P. R. *Biophysical Chemistry, Part III: The behavior of biological macromolecules*; W. H. Freeman and Company: New York, 1980; Chapter 15.

(2) Di Cera, E. *Thermodynamic theory of site-specific binding processes in biological macromolecule*; Cambridge University Press: New York, 1995; Chapter 2.

(3) Buffle, J. *Complexation reactions in aquatic systems*; Ellis Horword Series in Analytical Science; Ellis Horwood Limited: Chichester, 1988.

(4) Galceran, J.; Salvador, J.; Puy, J.; Mas, F.; van Leeuwen, H. P. *J. Electroanal. Chem*. **1995**, *391*, 29−40.

(5) Garcés J. L.; Mas F.; Puy J.; Galceran J.; Salvador, J. Submitted for publication.

(6) Sipps, R. *J. Chem. Phys.* **1948**, *16*, 490−495.

(7) Merz, P. H. *J. Comput. Phys.* **1980**, *38*, 64−85.

(8) Koopal, L. K.; van Riemsdijk, W. H.; de Wit, C. M.; Benedetti, M. F. *J. Colloid Interface Sci.* **1994**, *166*, 51−60.

(9) van Riemsdijk, W. H.; Koopal, L. K. In *Environmental Particles*, Vol. 1; Buffle, J., van Leeuwen, H. P., Eds.; Lewis Publishers: Chelsea, MI, 1992; Chapter 12.

(10) Nederlof, M. M.; van Riemsdijk, W. H.; Koopal, L. K. In *Heavy Metals in the Environment, Trace Metals in the Environment*, Vol. 1; Verner, J. P., Ed.; Elsevier: Amsterdam, 1991; pp 365−396.

(11) de Witt, J. C. M.; van Riemsdijk, W. H.; Koopal, L. K. *Environ. Sci. Technol.* **1993**, *27*, 2015−2022.

(12) Nederlof, M. M.; de Witt, J. C. M.; van Riemsdijk, W. H. *Environ. Sci. Technol.* **1993**, *27*, 846−856.

(13) Gamble, D. S.; Underdown, A. W.; Langford, C. H. *Anal. Chem*. **1980**, *52*, 1901−1908.

(14) Nederlof, M. M.; van Riemsdijk, W. H.; Koopal, L. K. *Environ. Sci. Technol.* **1992**, *26*, 763−771.

(15) House, W. A. *J. Colloid Interface Sci.* **1978**, *67*, 166−180; *Chem. Phys. Lett.* **1978**, *60*, 169−174.

(16) Janoriec, M.; Bräuer, P. *Surf. Sci. Rep.* **1986**, *6*, 65−117.

(17) Koopal, L. K.; Vos, C. H. W. *Langmuir* **1993**, 2593−2605.

(18) *Maximum entropy and bayesian methods*, Vol. 1; Erickson, G., Smith, C. R., Eds.; Kluwe Academic Pub.: Dordrecht, 1988.

(19) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in Fortran*; Cambridge University Press: New York, 1992; Chapter 18.

(20) Bricogne, G. *Acta Cryst.* **1984**, *40*, 410−445.

(21) Fuchs, G. M.; Prohaska T.; Friedbacher G.; Hutter H.; Grasserbauer M. *Fresenius J. Anal. Chem*. **1995**, *351*, 143−147.

(22) Kokaran, A. C.; Persad N.; Lasenby J.; Fitzerald, W. J.; McKinnon, A.; Welland, M. *Appl. Opt.* **1995**, *34*, 5121−5132.

(23) Bryant G.; Thomas, J. C. *Langmuir* **1995**, *11*, 2480−2485.

(24) Shaver, J. M.; McGown, L. B. *Anal. Chem*. **1995**, *68*, 9−17.

(25) de Mello, A. J.; Crystall, B.; Rumbles, G. *J. Colloid Interface Sci*. **1995**, *169*, 161−167.

(26) Bruggeman, Y. E.; Schoenmakers, R. G.; Schots, A.; Pap, E. H. W.; van Hoek, A.; Visser, A. J. W. G.; Hilhorst, R. *Eur. J. Biochem*. **1995**, *234*, 245−250.

(27) Stanley, B. J.; Topper, K.; Marshall, D. B. *Anal. Chim. Acta* **1994**, *287*, 25−34.

(28) Steinbach, P. J. *Biophys. J.* **1996**, *70*, 1521−1528.

(29) VanderNoot, T. J. *J. Electroanal. Chem*. **1995**, *386*, 57−63.

(30) Lüster, J.; Blaser, P.; Magyar, B. *Talanta* **1994**, *41*, 1873−1880.

(31) Rényi, A. *Probability Theory*; North Holland Series in Applied Mathematics and Mechanics; North Holland Publishing Company: Amsterdam, 1970; Chapter 9.

(32) Cornwell, T. J.; Evans, K. F. *Astron. Astrophys.* **1985**, *143*, 77−83.

(33) Nederlof, M. M.; van Riemsdijk, W. H.; Koopal, L. K. *Environ. Sci. Technol.* **1994**, *28*, 1037−1047.

(34) Scatchard, G.; Scheinberg, I. K.; Armstrong, S. H. *J. Am. Chem. Soc.* **1950**, *27*, 535−540.
(35) Klotz, I. M.; Hunston, D. L. *Biochemistry* **1971**, *10*, 3065−3069.
(36) Hunston, D. L. *Anal. Biochem.* **1975**, *52*, 99−109.
(37) Gamble, D. S.; Langford, C. H. *Environ. Sci. Technol.* **1988**, *22*, 1325−1336.
(38) Buffle, J.; Altmann, R. S.; Filella, M.; Tessier, A. *Geochim. Cosmochim. Acta* **1988** *52*, 1505−1519.
(39) de Witt, J. C. M.; van Riemsdijk, W. H.; Nederlof, M. M.; Kinniburgh, D. G.; Koopal L. K. *Anal. Chim. Acta* **1990**, *232*, 189−207.

Received for review May 27, 1997. Revised manuscript received November 7, 1997. Accepted November 10, 1997.

ES970468A