Journal of Mathematical Chemistry 26 (1999) 139–155

Protonation microequilibrium treatment of polybasic compounds with any possible symmetry

Zoltán Szakács [a] and Béla Noszál [b]

[a] *Department of Inorganic and Analytical Chemistry, Loránd Eötvös University, Pázmány Péter sétány 1/a, H-1117 Budapest, Hungary*

[b] *Institute of Pharmaceutical Chemistry, Semmelweis University, Hőgyes Endre utca 9, H-1092 Budapest, Hungary*

E-mail: nosbel@hogyes.sote.hu

Received 14 June 1999

Microequilibrium treatment that has previously been limited to tridentate ligands is generalized to arbitrary number of functional groups in the molecule and the role of symmetry is also investigated. Cumulative microconstant, a new type of equilibrium parameter, is introduced, allowing an equivalent, but more compact mathematical treatment of large microequilibrium systems. The sufficient number of independent pieces of information for the unambiguous determination of all microconstants is deduced. It has been concluded that even if protonation mole fraction for all the basic sites is available, determinability of all the microconstants is rather the exception than the case, without *a priori* simplifying assumptions. It has been shown that all microconstants can only be determined from protonation mole fractions for molecules of up to three groups. For molecules of four groups and beyond, only specific symmetry and the concomitant simplification of the microequilibrium system make the strict, complete microspeciation feasible. As a case study, the protonation scheme and the complete microspeciation of a tetradentate ligand is analyzed in detail.

1. Introduction, definitions

Acid-base properties of polyprotic compounds are usually characterized in terms of proton-association or -dissociation equilibrium constants. These thermodynamic parameters can be classified in several ways. Two major aspects of their classification are the direction of the process, and thoroughness of the information provided by the equilibrium constant.

If the equilibrium process is regarded from the direction of association, the constant is a protonation one (the values are typically given in K or $\log K$ units), whereas those that refer to the proton liberation are acidity, or dissociation constants (K_a, $\mathrm{pK_a}$). Both of them can refer to one single step of hydrogen ion release or binding (stepwise or successive constants, $K_1, K_2, \ldots, K_n$; $K_{\mathrm{a}1}, K_{\mathrm{a}2}, \ldots, K_{\mathrm{a}n}$), or can accumulate all previous steps (cumulative constants, $\beta_1, \beta_2, \ldots, \beta_n$; $\beta_{\mathrm{a}1}, \beta_{\mathrm{a}2}, \ldots, \beta_{\mathrm{a}n}$). The dissociation- and association-oriented formulas can be interconverted, indicating that the related de-

finitions do not imply principal differences. For the simplicity of formulations (and in accordance with the widely accepted convention of complex equilibria [8]), all equilibria are treated hereinafter as association ones.

The second major aspect of classification is the thoroughness of the information provided by the equilibrium constant. In this sense, structure-unrelated and structure-related parameters are distinguished. The structure-irrespective ones are macroconstants, which are definite in the stoichiometry of the species participating in the equilibrium process, but not in any kind of their structure [9]. The vast majority of reported equilibrium constants is macroconstants. The prefix, macro- or macroscopic is therefore usually omitted. Protonation macroconstants, β_i ($i = 1, 2, \ldots, n$) characterize the proton-binding propensity of the polyfunctional molecule as a whole,

$$\mathrm{L} + i\mathrm{H}^+ \rightleftharpoons \mathrm{H}_i\mathrm{L}, \qquad \beta_i = \frac{[\mathrm{H}_i\mathrm{L}]}{[\mathrm{L}] \cdot [\mathrm{H}^+]^i} = \prod_{j=1}^{i} K_j = \prod_{j=1}^{i} \frac{[\mathrm{H}_j\mathrm{L}]}{[\mathrm{H}_{j-1}\mathrm{L}] \cdot [\mathrm{H}^+]}, \qquad (1)$$

and, in principle, they cannot be assigned to individual binding sites. For the sake of simplicity, charges of the species (except for that of H^+) are omitted in our study.

Site-specific proton-binding equilibria can be expressed in terms of group constants, microconstants and submicroconstants. In this order, the structural content of the equilibrium constant is increasingly profound. Group constants reflect the proton-binding capability of individual groups, when the protonation state and hence the inductive effects of the rest of the molecule can be disregarded. Their properties and applications have been described earlier [11,12]. The now classical microscopic (or micro-) constants depict the basicity of proton-binding sites, when the protonation state of the other moieties is also definite [2,4,6,9–12,14,16–18,20,22,24]. Submicroconstants quantitate the basicity of individual proton-binding sites as well, when not only the protonation state, but the conformational (rotational) status of the molecule is also definite [13]. The analysis of submicroequilibria is beyond the scope of this paper.

Figure 1 shows the microequilibrium scheme of a biprotic molecule, in which the basic sites are A and B. The proton-coordination to either of the sites is denoted by H. The overlapping protonation of the α- and β-carboxylates of aspartate between pH 1 and 6 can be taken as an example, hence in this pH region the proton-uptake of the more basic amino group is complete [13]. There are four microspecies (four different forms of protonation of the molecule): a, b, c and d in the solution. Their concentrations can be related to the hydrogen ion concentration and the microconstants k^{A}, k^{B}, $k^{\mathrm{A}}_{\mathrm{B}}$ and $k^{\mathrm{B}}_{\mathrm{A}}$:

$$k^{\mathrm{A}} = \frac{[b]}{[a] \cdot [\mathrm{H}^+]}, \qquad k^{\mathrm{B}} = \frac{[c]}{[a] \cdot [\mathrm{H}^+]}, \qquad (2a)$$

$$k^{\mathrm{A}}_{\mathrm{B}} = \frac{[d]}{[c] \cdot [\mathrm{H}^+]}, \qquad k^{\mathrm{B}}_{\mathrm{A}} = \frac{[d]}{[b] \cdot [\mathrm{H}^+]}. \qquad (2b)$$

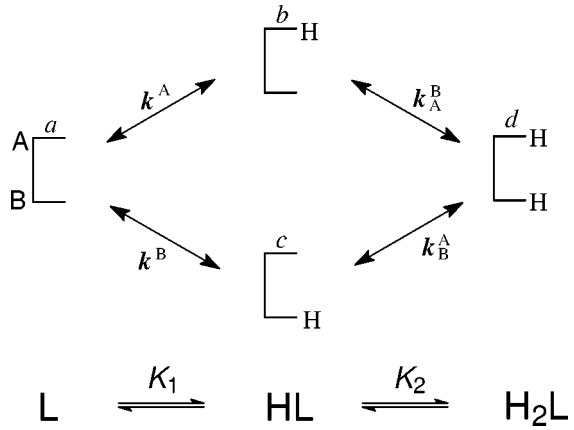

Figure 1. Protonation scheme of a molecule containing two functional groups, A and B.

The superscript on microconstant k denotes the functional group protonating in the given process, while the subscript (if any) denotes the group already protonated.

The relationships between the protonation micro- and macroconstants for a bidentate molecule are as follows [2]:

$$\beta_1 = K_1 = k^{\mathrm{A}} + k^{\mathrm{B}}, \tag{3a}$$

$$\beta_2 = K_1 K_2 = k^{\mathrm{A}} k^{\mathrm{B}}_{\mathrm{A}} = k^{\mathrm{B}} k^{\mathrm{A}}_{\mathrm{B}}. \tag{3b}$$

These equations can be incorporated into the total (analytical) concentration of the ligand:

$$\begin{aligned} C_{\mathrm{L}} &= [a] + [b] + [c] + [d] = [a] \cdot \left(1 + k^{\mathrm{A}}\left[\mathrm{H}^+\right] + k^{\mathrm{B}}\left[\mathrm{H}^+\right] + k^{\mathrm{A}} k^{\mathrm{B}}_{\mathrm{A}}\left[\mathrm{H}^+\right]^2\right) \\ &= [\mathrm{L}] + [\mathrm{HL}] + [\mathrm{H_2L}] = [\mathrm{L}] \cdot \left(1 + \beta_1\left[\mathrm{H}^+\right] + \beta_2\left[\mathrm{H}^+\right]^2\right). \end{aligned} \tag{4}$$

Microspecies b and c are of the same stoichiometric composition, but they hold the proton at different site, therefore called protonation isomers. Their concentration *ratio* (in certain cases called as zwitterion constant [6], which is not the general case) is independent of both the pH and the total ligand concentration:

$$k_{\mathrm{zw}} = \frac{[b]}{[c]} = \frac{k^{\mathrm{A}}[a][\mathrm{H}^+]}{k^{\mathrm{B}}[a][\mathrm{H}^+]} = \frac{k^{\mathrm{A}}}{k^{\mathrm{B}}} = \frac{[d] \cdot (k^{\mathrm{B}}_{\mathrm{A}}[\mathrm{H}^+])^{-1}}{[d] \cdot (k^{\mathrm{A}}_{\mathrm{B}}[\mathrm{H}^+])^{-1}} = \frac{k^{\mathrm{A}}_{\mathrm{B}}}{k^{\mathrm{B}}_{\mathrm{A}}}. \tag{5}$$

Since in most solvents the protonation processes are instantaneous, the protonation isomers occur exclusively in the presence of each other in the solution. Consequently, their individual spectroscopic or kinetic characteristics cannot be measured by direct methods [11]. It has been shown, however, that they act individually in specific biochemical reactions and the reactive species is not necessarily the major one [14]. The determination of microspecies concentrations and microconstants are interrelated and unified in the term of microspeciation.

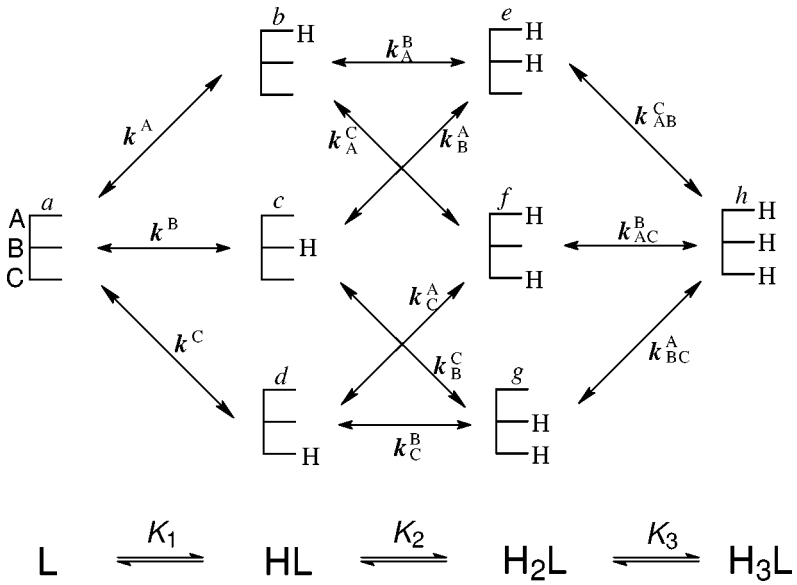

Figure 2. Protonation scheme of a molecule containing three functional groups, A, B and C.

Protonation at site A typically diminishes the basicity of site B and vice versa. This effect is expressed by the interactivity parameter [10,17,20]:

$$\log \Delta E_{\mathrm{AB}} = \log k^{\mathrm{A}}_{\mathrm{B}} - \log k^{\mathrm{A}} = \log k^{\mathrm{B}}_{\mathrm{A}} - \log k^{\mathrm{B}}. \tag{6}$$

In most cases, the value of $\log \Delta E_{\mathrm{AB}}$ is negative, indicating a negative cooperativity between the two sites, otherwise the cooperativity is positive. The interactivity parameter is a relatively invariant quantity. It is perturbed to a lesser extent by the protonation and the concomitant electron withdrawing effects of other groups than the microconstants themselves [20]. The interactivity parameter pertaining to the same molecular fragment and analogous pair of functional groups can therefore be transferred from small molecules to appropriate moieties of large molecules.

In fact, equation (5) states the mathematical constraint $k^{\mathrm{A}}k^{\mathrm{B}}_{\mathrm{A}} = k^{\mathrm{B}}k^{\mathrm{A}}_{\mathrm{B}}$ between the four microconstants, therefore only three of them are independent parameters. Thus, the degree of freedom in the two-group microequilibrium system is three. In other words, the knowledge of three arbitrarily chosen, mathematically independent basicity parameters (e.g., K_1, K_2, k^{A} or K_1, k^{B}, ΔE_{AB}, etc.) is a necessary and sufficient condition to compute all the microconstants.

The microequilibrium scheme for a three-group ligand is depicted in figure 2. The relations connecting macro- and microconstants [11] are given as follows:

$$\beta_1 = K_1 = k^{\mathrm{A}} + k^{\mathrm{B}} + k^{\mathrm{C}}, \tag{7a}$$

$$\beta_2 = K_1K_2 = k^{\mathrm{A}}k^{\mathrm{B}}_{\mathrm{A}} + k^{\mathrm{A}}k^{\mathrm{C}}_{\mathrm{A}} + k^{\mathrm{B}}k^{\mathrm{C}}_{\mathrm{B}} = k^{\mathrm{B}}k^{\mathrm{A}}_{\mathrm{B}} + k^{\mathrm{C}}k^{\mathrm{A}}_{\mathrm{C}} + k^{\mathrm{C}}k^{\mathrm{B}}_{\mathrm{C}}, \tag{7b}$$

$$\beta_3 = K_1K_2K_3 = k^{\mathrm{A}}k^{\mathrm{B}}_{\mathrm{A}}k^{\mathrm{C}}_{\mathrm{AB}} = k^{\mathrm{B}}k^{\mathrm{A}}_{\mathrm{B}}k^{\mathrm{C}}_{\mathrm{AB}} = \cdots . \tag{7c}$$

The species HL exists in the form of three nonidentical protonation isomers in the solution. It will be shown that only seven of the twelve microconstants are independent. The extended complexity of a triprotic over a diprotic system is self-explanatory.

In the general case, a molecule containing n basic sites has 2^n microspecies and $n \cdot 2^{n-1}$ microconstants [9,11]. The exponentially increasing complexity of the microequilibrium system is certainly the reason why the vast majority of microspeciation studies considers bidentate molecules. Correct microconstants of tridentate ligands are sporadic in the literature [10,24 and references therein] and as far as we are concerned just one paper appeared on microequilibria of a tetraprotic system [20]. This is especially surprising in view of the fact that recent papers in prominent journals [6,10,24] indicate the obvious need of an improved, general treatment of microspeciation, since most biomolecules are polyprotic ones, and their interactions take place via specific microforms.

Concerning, however, the general treatment of microequilibria, several principal questions arise, which as yet have not, or have only partially been answered. What is the impact of intramolecular symmetry elements on the number of microspecies and microconstants? What is the sufficient number of pieces of information to determine all the microconstants? Are there theoretical limits of the complete microequilibrium treatment and in which cases?

Here, the number of independent parameters describing unambiguously an arbitrary microequilibrium system is studied first, followed by the feasibility and limitations of parameter calculations from the experimental data. In order to avoid ill-conditioned mathematical models during the parameter estimation procedure, suitable combinations of microconstants are presented. The theoretical findings are illustrated by a detailed analysis of the protonation sequence of a symmetrical, tetraprotic compound.

2. Theory

2.1. The number of independent microconstants

Let A, B, C, ..., etc. basic sites be parts of a molecule and let their number be n. Assuming no symmetry relation between the sites, the number of two-proton-containing isomers is $\binom{n}{2}$ and that of the i-proton-containing protonation isomers is $\binom{n}{i}$. Consequently, the total number of microspecies $N_{\text{msp}}^{\max}$ can be given [9,11] as

$$N_{\text{msp}}^{\max} = \sum_{i=0}^{n} \binom{n}{i} = 2^n. \tag{8}$$

(The superscript max will be explained below.) In a microspecies holding i protons, the number of unoccupied sites is $n-i$, and this is also the number of its possible protonation processes and the corresponding microconstants, as figures 1 and 2 illustrate. For example, a molecule of A, B, C, D and E sites, protonated at sites A and D, can bind a third proton at sites B, C and E. The related microequilibria are characterized

in terms of the microconstants $k^{\mathrm{B}}_{\mathrm{AD}}$, $k^{\mathrm{C}}_{\mathrm{AD}}$ and $k^{\mathrm{E}}_{\mathrm{AD}}$, respectively. In the general case, $N^{\max}_{\mathrm{msp}}$, the total number of microconstants in the system can be expressed as

$$N^{\max}_{\mathrm{msp}} = \sum_{i=0}^{n} \binom{n}{i} \cdot (n-i) = n \cdot 2^{n-1}. \tag{9}$$

Thus, the number of microconstants can be enormously large, even in molecules of medium size. Not all the microconstants are, however, independent, due to relations like equation (5). One of the objectives of this work is the derivation of the number of independent microconstants for an arbitrary molecule.

If the molecule contains symmetry elements, certain basic sites become equivalent, which can be envisioned by a proper C_n rotation, which transform them into one another. If the number of equivalent A sites (denoted by A, A′, A″, etc.) is n_1, the multiplicity of site B is n_2, etc. and the multiplicity of the νth site is n_ν, the summation of these numbers leads to the number of all the functional groups,

$$n = \sum_{k=1}^{\nu} n_k. \tag{10}$$

Those protonation isomers, which hold the same number of protons on their equivalent moieties, have the same constitution and represent physically indistinguishable species in the solution (see also section 4). Hence, symmetry diminishes the number of microspecies with different constitution, as compared to the general case: $N_{\mathrm{msp}} < N^{\max}_{\mathrm{msp}}$. Since identical microspecies must also be identical in their basicity [20,22], the corresponding microconstants should also have the same value. Thus the number of different microconstants diminishes as well, $N_{\mathrm{mc}} < N^{\max}_{\mathrm{mc}}$. At ligands of no symmetry, ν equals to n and the complexity of the microequilibrium system is maximum.

Microspeciation aims at determining the concentration of *all* microspecies in the solution, including the most inferior ones. This requires the knowledge of at least one microconstant for each microspecies, corresponding to a microequilibrium step, in which that microspecies forms. In complete analogy to the cumulative stability constant β_i defined for the macrospecies $\mathrm{H}_i\mathrm{L}$, we introduce a new unifying parameter, the cumulative Hessian microconstant κ for each microspecies. This parameter is cumulative, since all previous steps of protonation are incorporated, and can be called Hessian, because it expresses its thermodynamic nature, being independent of the intermediate stages and ways of the microspecies formation. It can be expressed as the product of the corresponding (stepwise) k microconstants. For example, κ_e for microspecies e in figure 2 can be formulated as

$$\kappa_e = k^{\mathrm{A}} k^{\mathrm{B}}_{\mathrm{A}} = k^{\mathrm{B}} k^{\mathrm{A}}_{\mathrm{B}}. \tag{11}$$

The concentration of this microspecies can be calculated as

$$[e] = \alpha_e \cdot C_{\mathrm{L}} = \frac{\kappa_e [\mathrm{H}^+]^2}{D} \cdot C_{\mathrm{L}}, \tag{12}$$

where α_e is the mole fraction of microspecies e, C_{L} is the total ligand concentration, and D is expressed below:

$$D = \frac{C_{\mathrm{L}}}{[a]} = \frac{[a] + [b] + \cdots + [z]}{[a]} = \frac{[\mathrm{L}] + [\mathrm{HL}] + [\mathrm{H_2L}] + \cdots + [\mathrm{H}_n\mathrm{L}]}{[\mathrm{L}]}$$
$$= 1 + \sum_{i=1}^{n} \beta_i \cdot \left[\mathrm{H}^+\right]^i. \tag{13}$$

In the concentration expression of a, the completely deprotonated microspecies no unknown κ cumulative microconstant occurs:

$$[a] = \alpha_a \cdot C_{\mathrm{L}} = \frac{\kappa_a}{D} \cdot C_{\mathrm{L}} = \frac{1}{D} \cdot C_{\mathrm{L}}. \tag{14}$$

Thus, the calculation of concentrations for the remaining $N_{\mathrm{msp}}-1$ microspecies requires

$$DF_{\mathrm{mc}} = N_{\mathrm{msp}} - 1 \tag{15}$$

microscopic protonation parameters. DF_{mc} is the number of degrees of freedom in the microspeciation system. This is, actually, the number of mathematically independent microconstants. The other microconstants can be calculated from this minimum set of parameters using constraints like

$$k^{\mathrm{A}} \cdot k^{\mathrm{B}}_{\mathrm{A}} = k^{\mathrm{A}} \cdot k^{\mathrm{B}}_{\mathrm{A}} \quad \text{or} \quad k^{\mathrm{C}}_{\mathrm{AA'}} \cdot k^{\mathrm{D}}_{\mathrm{AA'C}} = k^{\mathrm{D}}_{\mathrm{AA'}} \cdot k^{\mathrm{C}}_{\mathrm{AA'D}} \quad \text{etc.} \tag{16}$$

Concerning the protonation scheme in graph representation, the vertices of the graph can be identified as the microspecies and the edges are the microconstants. The protonation scheme is always a *connected graph*, since each vertex can be reached from all other vertices. Constraints like equation (5) correspond to *circles* in the graph. A *spanning tree* of the graph means a tree with minimum edges, which covers all the vertices. The enumeration of the independent microconstants is equivalent to counting the edges of the spanning tree. In graph theory it is known that a walk reaching all the N vertices of a connected graph consists of $N - 1$ edges. The same result is stated in equation (15).

2.2. Description of the microequilibrium system in terms of interactivity parameters

The interactivity parameter (equation (6)) establishes a relationship between two particular sites in a molecule. It quantitates a protonation-initiated, reciprocal change (usually drop) of basicity in a pair of sites. In a molecule containing n basic sites, the number of such pairs of sites can be given as

$$N_{\mathrm{ip}} = \binom{n}{2} = \frac{n \cdot (n-1)}{2}. \tag{17}$$

For the trifunctional system (figure 2), the interactions among the three moieties can be quantitated in terms of the ΔE_{AB}, ΔE_{AC} and ΔE_{BC} interactivity parameters. For

instance, the basicity of site C in microspecies e, characterized by $k^{\mathrm{C}}_{\mathrm{AB}}$, can be decomposed into the k^{C} *intrinsic* basicity and the two modifying effects upon protonation of sites A and B:

$$\log k^{\mathrm{C}}_{\mathrm{AB}} = \log k^{\mathrm{C}} + \log \Delta E_{\mathrm{AC}} + \log \Delta E_{\mathrm{BC}}. \tag{18}$$

In a similar, additive manner, every microconstant in the protonation scheme can be expressed using n independent microconstants and N_{ip} interactivity parameters. Thus, the degree of freedom DF_{ip} expressed in terms of intrinsic microconstants and interactivity parameters can be formulated as

$$DF_{\mathrm{ip}} = n + N_{\mathrm{ip}}. \tag{19}$$

It can be concluded that the interactivity parameter treatment provides the description of the system with the same thoroughness but using fewer parameters than the full microconstant treatment outlined in the previous section. The missing part of the information is supplied here by the *a priori* assumption that interactivity parameters are invariant at various protonation stages of the rest of the ligand.

2.3. The number of experimentally available parameters

The most frequently used experimental arrangement for the determination of microconstants is a combination of pH potentiometry and a pH-dependent set of spectra. The former technique provides the macroconstants, the latter one selectively monitors the protonation of one, or some of the basic sites [4,6,9,17,18,20]. If a spectroscopic method is sufficiently sensitive to the protonation changes of all sites, the pH-dependent spectral data can serve as exclusive source for the determination of all constants. Such "spectral–pH" techniques are ideally suited for the determination of microconstants if the protonation of the particular sites can exclusively be assigned to some spectral changes. Our treatment below is focused on such a case. Among the various methods applied to date, NMR–pH [10,17,18,20,22,24] and UV–pH titrations [6,15,16] have been most widely employed. In both techniques, a spectral parameter (chemical shift or absorbance) is recorded as a function of pH. The so-called *protonation fraction* curves can be calculated from the experimental titration curves [10,17]. The protonation fraction of a particular site, k ($k = 1, 2, \ldots, n$) is defined as

$$f_k = \frac{\text{sum of concentrations of microspecies protonated on site } k}{\text{sum of concentrations of all microspecies}}. \tag{20}$$

The summation in the denominator gives C_{L}, the total ligand concentration. For example, the protonation fraction of site B for the triprotic molecule (figure 2) is given in equation (21):

$$f_{\mathrm{B}} = \frac{[c] + [e] + [g] + [h]}{[a] + [b] + [c] + [d] + [e] + [f] + [g] + [h]}. \tag{21}$$

Expressing the microspecies concentrations in terms of cumulative microconstants (see equations (12) and (14)) and simplifying with $[a]$ yields

$$f_{\mathrm{B}} = \frac{\kappa_c[\mathrm{H}^+] + \kappa_e[\mathrm{H}^+]^2 + \kappa_g[\mathrm{H}^+]^2 + \kappa_h[\mathrm{H}^+]^3}{1 + \kappa_b[\mathrm{H}^+] + \kappa_c[\mathrm{H}^+] + \kappa_d[\mathrm{H}^+] + \kappa_e[\mathrm{H}^+]^2 + \kappa_f[\mathrm{H}^+]^2 + \kappa_g[\mathrm{H}^+]^2 + \kappa_h[\mathrm{H}^+]^3}. \tag{22}$$

Compressing the κ coefficients of the same power of $[\mathrm{H}^+]$ into a common Q value yields a fractional rational function of the third degree in $[\mathrm{H}^+]$

$$f_{\mathrm{B}} = \frac{Q_{\mathrm{B},1}[\mathrm{H}^+] + Q_{\mathrm{B},2}[\mathrm{H}^+]^2 + Q_{\mathrm{B},3}[\mathrm{H}^+]^3}{1 + \beta_1[\mathrm{H}^+] + \beta_2[\mathrm{H}^+]^2 + \beta_3[\mathrm{H}^+]^3}. \tag{23}$$

In this relation $Q_{\mathrm{B},2}$ is the sum of the κ constants of those microspecies, which hold proton on site B. Note that the corresponding cumulative parameter, the β_2 macroconstant incorporates *all* protonation isomers with the formula H_2L, irrespective of the site of protonation.

In general, a molecule with n basic sites has n protonation fraction functions, each being a fractional rational function of the nth degree in $[\mathrm{H}^+]$,

$$f_{\mathrm{A}}\big([\mathrm{H}^+]\big) = \frac{Q_{\mathrm{A},1}[\mathrm{H}^+] + Q_{\mathrm{A},2}[\mathrm{H}^+]^2 + \cdots + Q_{\mathrm{A},n-1}[\mathrm{H}^+]^{n-1} + Q_{\mathrm{A},n}[\mathrm{H}^+]^n}{1 + \beta_1[\mathrm{H}^+] + \beta_2[\mathrm{H}^+]^2 + \cdots + \beta_{n-1}[\mathrm{H}^+]^{n-1} + \beta_n[\mathrm{H}^+]^n}, \tag{24a}$$

$$f_{\mathrm{B}}\big([\mathrm{H}^+]\big) = \frac{Q_{\mathrm{B},1}[\mathrm{H}^+] + Q_{\mathrm{B},2}[\mathrm{H}^+]^2 + \cdots + Q_{\mathrm{B},n-1}[\mathrm{H}^+]^{n-1} + Q_{\mathrm{B},n}[\mathrm{H}^+]^n}{1 + \beta_1[\mathrm{H}^+] + \beta_2[\mathrm{H}^+]^2 + \cdots + \beta_{n-1}[\mathrm{H}^+]^{n-1} + \beta_n[\mathrm{H}^+]^n}, \tag{24b}$$

$$\vdots$$

The Q coefficients always build up from κ cumulative microconstants of one or several microspecies. In symmetrical molecules, the protonation fractions of equivalent sites are equal [22]. Hence, only ν different protonation fractions can be measured, where ν is the number of non-equivalent sites.

Summation of all the protonation fractions yields the so-called Bjerrum function [3],

$$\sum_{i=1}^{n} f_i = \sum_{k=1}^{\nu} (n_k \cdot f_k) = \overline{n}_{[\mathrm{H}^+]}, \tag{25}$$

which gives the average mole of protons associated with the ligand at a given pH. This quantity can be measured in separate experiments by well-known, sophisticated methods, like pH potentiometry [19]. Consequently, if one measures the $\overline{n}$ function under exactly the same experimental conditions, it suffices to record the protonation of $(\nu - 1)$ moieties with site-specific techniques [12].

If the coefficients of $[\mathrm{H}^+]$ are compared in Bjerrum's classical formula

$$\overline{n}_{[\mathrm{H}^+]} = \frac{\beta_1[\mathrm{H}^+] + 2 \cdot \beta_2[\mathrm{H}^+]^2 + \cdots + n \cdot \beta_n[\mathrm{H}^+]^n}{1 + \beta_1[\mathrm{H}^+] + \beta_2[\mathrm{H}^+]^2 + \cdots + \beta_n[\mathrm{H}^+]^n} = \frac{\sum_{i=1}^{n} i\beta_i[\mathrm{H}^+]^i}{1 + \sum_{i=1}^{n} \beta_i[\mathrm{H}^+]^i} \tag{26}$$

and in equations (24)–(25), the general relations between macro- and microconstants can be derived:

$$\beta_i = \frac{1}{i} \cdot \sum_{j=1}^{n} Q_{j,i} = \frac{1}{i} \cdot \sum_{k=1}^{\nu} n_k Q_{k,i}, \quad i = 1, 2, \ldots, n. \tag{27}$$

This result clearly shows that macroconstants cannot be regarded as further independent parameters over the microconstants. Thus, they do not influence the number of degrees of freedom.

Equations (24) have a common denominator and both their numerator and denominator are uniquely determined by the Q coefficients at a particular pH. The last Q coefficients in all these functions refer to the fully protonated microspecies. Thus, they are equal in equations (24) and (26):

$$Q_{\mathrm{A},n} = Q_{\mathrm{B},n} = \cdots = \beta_n. \tag{28}$$

Every (24)-type equation contains a $Q_{k,n}$ coefficient, as common parameter, plus $n-1$ Q-type unknown parameters. Accordingly, N_{exptl}, the number of experimentally available parameters from the ν different f_k-type functions, the most sophisticated relationships of the field, can be given as in equation (29):

$$N_{\mathrm{exptl}} = \nu(n-1) + 1. \tag{29}$$

On this basis, we postulate the following statement. If the number of unknown equilibrium constants DF_{mc} is greater than the experimentally available cumulative parameters ($DF_{\mathrm{mc}} > N_{\mathrm{exptl}}$), not *all* the microspecies concentrations can be necessarily calculated from the experimental data. In fact, this is rather the case than the exception if $n > 3$, as will be shown in section 3. In other words, the determination of all the microconstants for molecules of more than three basic sites can be not only technically difficult, but also theoretically impossible. This statement seems unquestionably true at the present status of state-of-the-art instrumentation. If, however, sometime in the future, for example, a spectroscopy is invented, of which the timescale is shorter than the lifetime of individual microspecies, the above statement can become obsolete. Such progress, at the moment, is hard to foresee. The above mentioned theoretical impossibility, perhaps, with the more easily comprehensible complexity of microspeciation on molecules with more than three sites could be the burden that prevented microequilibrium studies on such systems so far. It must also be mentioned that less strict treatment (importing, for example, interactivity parameters from related, simplified molecules) may result in the estimation of all microconstants.

Our statement was confirmed by model studies and computational experience for various molecular symmetry: when the number of unknown parameters ($\log Q$ values) was greater than N_{exptl} during the least-squares fitting of equations (24), some parameters were highly correlated and statistically insignificant. In addition, the effective dimensionality of the parameter space (which equals to the rank of the estimated correlation matrix of the parameters, determined by principal component analysis [5]) was always found exactly equal to N_{exptl}.

Table 1
The number of functional groups (n); the number of groups of different symmetry (ν); the number of different microspecies (N_{msp}) and microconstants (N_{mc}); the number of degrees of freedom (DF_{mc} and DF_{ip}) and the number of experimentally available parameters (N_{exptl}) for various symmetries of polydentate molecules. NTA = nitrilo-triacetate, Ins(1, 2, 6)P_3 = D-*myo*-inositol-1, 2, 6-tris(phospate), GSSG = oxidized glutathione, DTPA = diethylenetriaminepentaacetate, TTHA = triethylenetetraminehexa-acetate. For further information, see text.

n	Symmetry	ν	N_{msp}	N_{mc}	DF_{mc}	N_{exptl}	DF_{ip}	Example
2	A_2	1	3	2	2	2	$1+1$	succinate carboxylates
2	AB	2	4	4	3	3	$2+1$	lysine amino groups
3	A_3	1	4	3	3	3	$1+1$	NTA carboxylates
3	A_2B	2	6	7	5	5	$2+2$	citrate carboxylates
3	ABC	3	8	12	7	7	$3+3$	Ins(1, 2, 6)P_3 phosphates
4	A_4	1	5	4	4	4	$1+1$	EDTA carboxylates
4	A_3B	2	8	10	7	7	$2+2$	–
4	A_2B_2	2	10	16	9	7	$2+4$	GSSG carboxylates
4	A_2BC	3	12	20	11	10	$3+4$	–
4	ABCD	4	16	32	15	13	$4+6$	trilysine amino groups
5	A_5	1	6	5	5	5	$1+1$	–
5	A_4B	2	10	13	9	9	$2+2$	DTPA carboxylates
5	ABCDE	5	32	80	31	21	$5+10$	tobramycin amino groups
6	A_6	1	7	6	6	6	$1+1$	mellitate carboxylates
6	A_5B	2	12	16	11	11	$2+2$	–
6	A_4B_2	2	15	48	14	11	$2+3$	TTHA carboxylates
6	ABCDEF	6	64	192	63	31	$6+15$	corticotropin
⋮								
n	A_n	1	$n+1$	n	n	n	$1+1$	–
n	$A_{n-1}B$	2	$2n$	$3n-2$	$2n-1$	$2n-1$	$2+2$	–
n	ABCD...	n	2^n	$n\cdot 2^{n-1}$	2^n-1	$\nu(n-1)+1$	$n+\binom{n}{2}$	–

3. Strategies to compute microconstants at various molecule symmetries

In order to exemplify, visualize and analyze relationships among denticity, symmetry and parameter determinability, we have constructed microscopic protonation schemes and models for molecules of two to six donor groups with every possible symmetry. The number of necessary parameters (DF_{mc} or DF_{ip}) for the complete microspeciation treatment, as well as the number of experimentally available parameters (N_{exptl}) are shown in figure 1 and table 1, together with representative ligands. Comparing these two quantities, every polyfunctional molecule can be classified into one of the following three categories:

All microconstants can be calculated from the macroconstants

This is the case, when all binding sites are equivalent ($A_2, A_3, A_4, \ldots, A_n$ symmetry).

The microconstants can be computed from the stepwise macroscopic constants, using relationships [21] as shown in equations (30a)–(30c):

$$\log k^{\mathrm{A}} = \log K_1 + \log \frac{1}{n}, \qquad \log k_{\mathrm{A}}^{\mathrm{A}'} = \log K_2 + \log \frac{2}{n-1}, \tag{30a}$$

$$\log k_{\mathrm{AA}'\ldots\mathrm{A}^{(n-2)}}^{\mathrm{A}^{(n-1)}} = \log K_{n-1} + \log \frac{n-1}{2}, \tag{30b}$$

$$\log k_{\mathrm{AA}'\ldots\mathrm{A}^{(n-2)}\mathrm{A}^{(n-1)}}^{\mathrm{A}^{(n)}} = \log K_n + \log \frac{n}{1}. \tag{30c}$$

In these cases, the application of site-specific (spectroscopic) techniques is not necessary, the high symmetry allows the calculation of submolecular basicity from the macroconstants directly [11].

The number of microconstants and that of the experimentally available parameters are equal

This is the case when all but one binding sites are equivalent (AB, A_2B, A_3B, A_4B, ..., $A_{n-1}B$ symmetry) and, interestingly, the general trifunctional case, ABC also belongs here.

Table 1 shows that $DF_{\mathrm{mc}} = N_{\mathrm{exptl}}$, hence all microconstants can be obtained from the experimental protonation fraction curves with simultaneous least-squares fitting. Depending on the fitting functions used, the calculation procedure may result in

(a) the logarithms of independent microconstants (direct method),

(b) the logarithms of the unknown κ cumulative microconstants,

(c) the logarithms of the Q coefficients.

The three methods lead to the same result (within the limits of calculation error), since the problem is not overparametrized. No simplifying chemical assumptions are needed.

Not all the microconstants can be extracted

This is the case of all molecules and symmetries not listed in the above two categories.

Unfortunately, almost all multibasic natural or synthetic ligands (e.g., peptides, chelators, drug molecules, etc.) are typical compounds of low symmetry and they therefore belong to this class. Since $DF_{\mathrm{mc}} > N_{\mathrm{exptl}}$, *all* microconstants cannot be computed unambiguously from the experimental data. There are, however, reasonable approaches and efficient strategies for the computation of a large number of microconstants even for such cases.

(a) The functional form of equations (24) *always* allows the fitting of the log Q parameters (unlike the fitting of the microconstants themselves). From these quantities, it is possible to calculate the cumulative microconstants of certain microspecies, hence their concentration can be computed. For the rest of the microspecies, only the sum of concentrations can be obtained unambiguously.

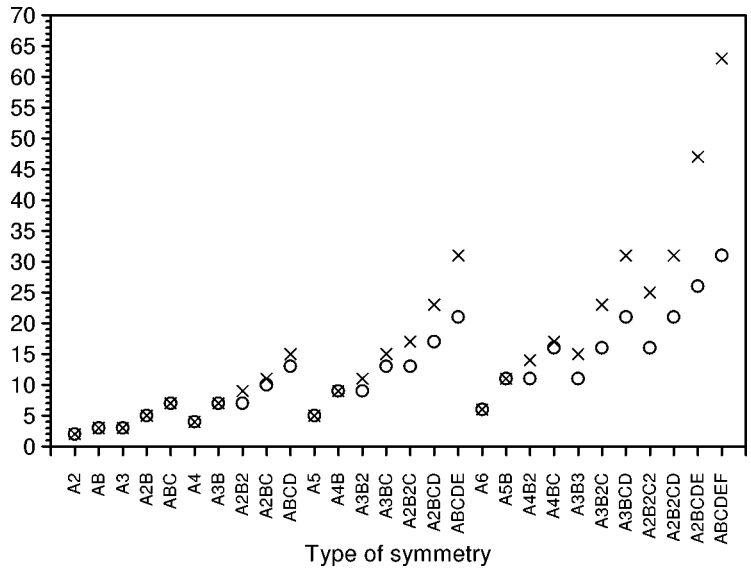

Figure 3. The number of independent microconstants DF_{mc} ($\times$) and the number of experimentally available parameters N_{exptl} (◯) for polyfunctional molecules of various symmetry.

(b) With *chemical preconditions*, of course, it is possible to unravel the complete protonation scheme. Assuming the additivity and transferability of the interactivity parameters (see equation (18)), only $DF_{\mathrm{ip}} < DF_{\mathrm{m}}$ parameters are required for the complete description, which can be more readily obtained by parameter estimation methods. This approach has recently been used to characterize the basicity of the four carboxylates of oxidized glutathione (GSSG) [23].

(c) If the binding sites are separated by more than three sigma bonds, their inductive communication can be neglected. The basicity of the groups can be quantified in terms of $n < DF_{\mathrm{ip}} < DF_{\mathrm{m}}$ group constants [11], indicating a further decrease in the number of degrees of freedom to the debit of a more stringent *a priori* chemical precondition.

The information content on the basicity-modifying interactions of the sites as well as on the pH-dependent fine structure of the ligand gradually decreases from (a) to (c).

4. A case study on a tetrabasic acid of A_3B symmetry

As a particular example, the microscopic protonation equilibria of a four-group ligand are discussed (figure 4). A molecule of this type can be found, for example, in two affinity capillary electrophoresis studies [1,7]. Three carboxylate groups (A, A′ and A″) are equivalent, thus the ligand belongs to the A_3B case of symmetry. The protonation scheme (figure 5) contains all the sixteen microspecies ($a, b, \ldots, p$). There

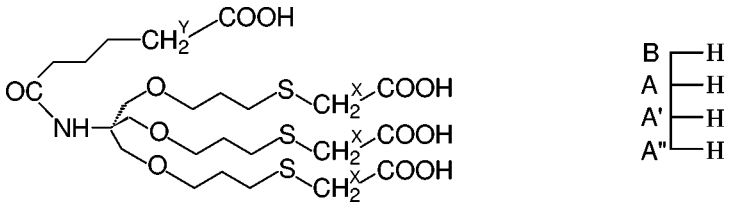

Figure 4. Structure of a tetraprotic molecule H_4L belonging to the A_3B symmetry and its symbol in the fully protonated state.

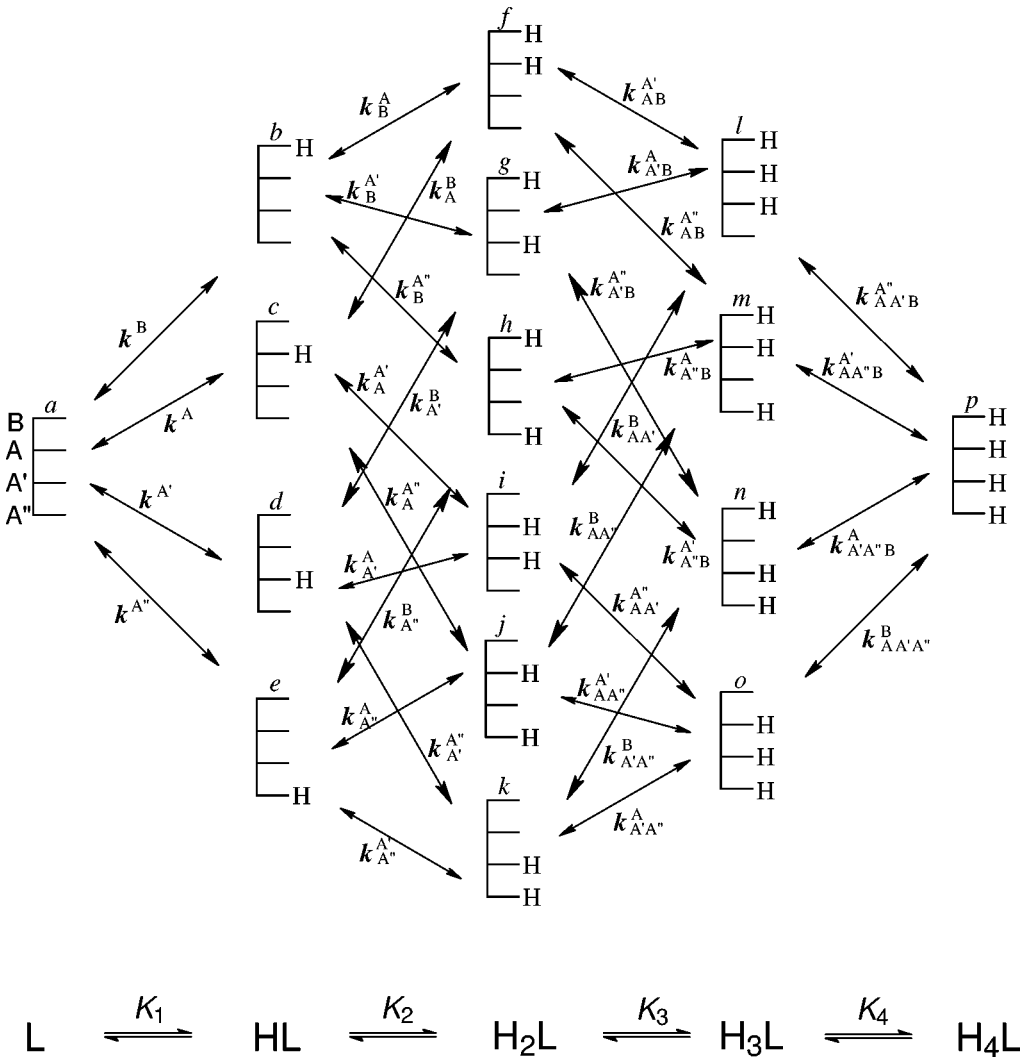

Figure 5. Protonation scheme of an A_3B ligand. For a particular molecule of this symmetry, see figure 4.

are, however, eight constitutionally distinguishable microspecies in the solution only, since certain chemical entities are identical: $c \equiv d \equiv e$, $f \equiv g \equiv h$, $i \equiv j \equiv k$ and $l \equiv m \equiv n$. Consequently, some of the 32 microconstants have also identical values, the number of different microconstants is $N_{\mathrm{mc}} = 10$. Table 1 shows that the number of independent microconstants is $DF_{\mathrm{m}} = 7$. Hence, the cumulative microconstants of the microspecies can be expressed in terms of seven independent microconstants (k):

$$\kappa_a = 1, \qquad \kappa_b = k^{\mathrm{B}}, \qquad \kappa_c = k^{\mathrm{A}}, \tag{31a}$$

$$\kappa_f = k^{\mathrm{A}} \cdot k^{\mathrm{B}}_{\mathrm{A}}, \qquad \kappa_i = k^{\mathrm{A}} \cdot k^{\mathrm{A}'}_{\mathrm{A}}, \qquad \kappa_l = k^{\mathrm{A}} \cdot k^{\mathrm{A}'}_{\mathrm{A}} \cdot k^{\mathrm{B}}_{\mathrm{AA}'}, \tag{31b}$$

$$\kappa_o = k^{\mathrm{A}} \cdot k^{\mathrm{A}'}_{\mathrm{A}} \cdot k^{\mathrm{A}''}_{\mathrm{AA}'}, \qquad \kappa_p = k^{\mathrm{A}} \cdot k^{\mathrm{A}'}_{\mathrm{A}} \cdot k^{\mathrm{B}}_{\mathrm{AA}'} \cdot k^{\mathrm{A}''}_{\mathrm{AA}'\mathrm{B}}. \tag{31c}$$

Applying the (20)–(23) protonation fraction equations for the A and B type carboxylates of this ligand yields

$$f_{\mathrm{A}} = \frac{[c]+[f]+[i]+[j]+[l]+[m]+[o]+[p]}{[a]+[b]+[c]+[d]+[e]+[f]+[g]+[h]+[i]+[j]+[k]+[l]+[m]+[n]+[o]+[p]} = \frac{Q_{\mathrm{A},1}[\mathrm{H}^+] + Q_{\mathrm{A},2}[\mathrm{H}^+]^2 + Q_{\mathrm{A},3}[\mathrm{H}^+]^3 + Q_{\mathrm{A},4}[\mathrm{H}^+]^4}{1 + \beta_1[\mathrm{H}^+] + \beta_2[\mathrm{H}^+]^2 + \beta_3[\mathrm{H}^+]^3 + \beta_4[\mathrm{H}^+]^4}, \tag{32a}$$

$$f_{\mathrm{B}} = \frac{[b]+[f]+[g]+[h]+[l]+[m]+[n]+[p]}{[a]+[b]+[c]+[d]+[e]+[f]+[g]+[h]+[i]+[j]+[k]+[l]+[m]+[n]+[o]+[p]} = \frac{Q_{\mathrm{B},1}[\mathrm{H}^+] + Q_{\mathrm{B},2}[\mathrm{H}^+]^2 + Q_{\mathrm{B},3}[\mathrm{H}^+]^3 + Q_{\mathrm{B},4}[\mathrm{H}^+]^4}{1 + \beta_1[\mathrm{H}^+] + \beta_2[\mathrm{H}^+]^2 + \beta_3[\mathrm{H}^+]^3 + \beta_4[\mathrm{H}^+]^4}. \tag{32b}$$

Due to the equivalence of the three A sites in the compound,

$$f_{\mathrm{A}'} = f_{\mathrm{A}''} = f_{\mathrm{A}}. \tag{33}$$

In the above equations, the coefficients Q are composed of the cumulative microconstants as follows:

$$Q_{\mathrm{A},1} = \kappa_c, \qquad Q_{\mathrm{A},2} = \kappa_f + 2 \cdot \kappa_i, \qquad Q_{\mathrm{A},3} = \kappa_o + 2 \cdot \kappa_l, \tag{34a}$$

$$Q_{\mathrm{A},4} = \kappa_p, \qquad Q_{\mathrm{B},1} = \kappa_b, \qquad Q_{\mathrm{B},2} = 3 \cdot \kappa_f, \qquad Q_{\mathrm{B},3} = 3 \cdot \kappa_l. \tag{34b}$$

The expressions of the cumulative *macroconstants* clearly reflect the symmetry of the three A sites:

$$\beta_1 = (3 \cdot Q_{\mathrm{A},1} + Q_{\mathrm{B},1})/1, \qquad \beta_2 = (3 \cdot Q_{\mathrm{A},2} + Q_{\mathrm{B},2})/2, \tag{35a}$$

$$\beta_3 = (3 \cdot Q_{\mathrm{A},3} + Q_{\mathrm{B},3})/3, \qquad \beta_4 = (3 \cdot Q_{\mathrm{A},4} + Q_{\mathrm{B},4})/4 = Q_{\mathrm{A},4}. \tag{35b}$$

If the protonation of the carboxylates is monitored, e.g., by ^{1}H NMR spectroscopy, the protonation fractions can be obtained from the experimental data using the following expressions [10,17,22]:

$$\delta_{\mathrm{X}} = \delta_{\mathrm{Xp}} \cdot f_{\mathrm{A}} + \delta_{\mathrm{Xd}} \cdot (1 - f_{\mathrm{A}}) \quad \text{and} \quad \delta_{\mathrm{Y}} = \delta_{\mathrm{Yp}} \cdot f_{\mathrm{B}} + \delta_{\mathrm{Yd}} \cdot (1 - f_{\mathrm{B}}), \tag{36}$$

where δ_{X} and δ_{Y} are the pH-dependent chemical shifts of the covalently bound hydrogens adjacent to the carboxylates (see figure 4), δ_{Xp}, δ_{Xd}, δ_{Yp} and δ_{Yd} are their limiting

values in the fully protonated and deprotonated state of the ligand, respectively. Table 1 shows that *all* microconstants can be calculated from the f_A and f_B functions by nonlinear parameter estimation methods.

5. Concluding remarks

The earlier, general misbelief that microspeciation of molecules beyond three sites is only a matter of simple extension of existing relationships and methods is confuted. In fact, it is shown here that rigorously consistent microspeciation of more than tridentate ligands is feasible under certain symmetry conditions only. The relationships among the number and symmetry of sites, the number of experimentally available parameters and the degrees of freedom are itemized for molecules of up to six sites. Molecules with the corresponding $\binom{6}{2} = 15$ or more real inter-site interactions hardly exist in practice, since such a large number of sites quite necessarily involves some simplifying molecular conditions, such as a large number of isolating bounds. Consequently, itemization of considerations for molecules of $n > 6$ is unnecessary. There can be, however, real molecules with 4 or 5 sites, in which microequilibria of high complexity is the case, and the principles established in this paper can be put into real chemistry practice. We will publish such cases for the carboxylates of oxidized glutathione, a molecule of A_2B_2 symmetry and ovothiol, a natural compound of ABCD symmetry.

Acknowledgements

Financial support from the grants OTKA T 17570, MKM FKFP 1126/1997 and NjM 447/96 is gratefully acknowledged. Z.Sz. is indebted to Prof. Ernö Keszei and to Csaba Megyeri for valuable discussions.

References

[1] L.Z. Avila, Y.-H. Chu, E.C. Blossey and G.M. Whitesides, J. Med. Chem. 36 (1993) 126–133.
[2] N. Bjerrum, Z. Phys. Chem., Stoechiom. Verwandtschaftl. 106 (1923) 209–214.
[3] J. Bjerrum, *Metal Ammine Formation in Aqueous Solution* (Haase, Copenhagen, 1941).
[4] G.M. Bodner, J. Chem. Educ. 63 (1986) 246–247.
[5] C. Chatfield and A.J. Collins, *Introduction to Multivariate Analysis* (Chapman and Hall, 1980).
[6] J.C. D'Angelo and T.W. Collette, Anal. Chem. 69 (1997) 1642–1650.
[7] F.A. Gomez, L.Z. Avila, Y.-H. Chu and G.M. Whitesides, Anal. Chem. 66 (1994) 1785–1791.
[8] A.E. Martell and R.M. Smith, *Critical Stability Constants* (Plenum Press, New York, 1974).
[9] R.B. Martin, *Introduction to Biophysical Chemistry* (McGraw-Hill, New York, 1964).
[10] K. Mernissi-Arifi, L. Schmitt, G. Schlewer and B. Spiess, Anal. Chem. 67 (1995) 2567–2574.
[11] B. Noszál, J. Phys. Chem. 90 (1986) 4104–4110.
[12] B. Noszál, J. Phys. Chem. 90 (1986) 6345–6349.
[13] B. Noszál and P. Sándor, Anal. Chem. 61 (1989) 2631–2637.
[14] B. Noszál, V. Scheller-Krattiger and R.B. Martin, J. Am. Chem. Soc. 104 (1982) 1078–1081.

[15] J. Ollig, G. Hägele, Computers Chem. 19 (1995) 287–294.
[16] L. Pillai, R.D. Boss and M.S. Greenberg, J. Solution Chem. 8 (1979) 635–646.
[17] D.L. Rabenstein, J. Am. Chem. Soc. 95 (1973) 2797–2803.
[18] D.L. Rabenstein and T.L. Sayer, Anal. Chem. 48 (1967) 1141–1146.
[19] F. J. Rosotti and H. S. Rosotti, *The Determination of Stability Constants* (McGraw-Hill, New York, 1961).
[20] M.A. Santos, M.A. Esteves, M.C. Vaz, J.J.R. Fraústo da Silva, B. Noszál and E. Farkas, J. Chem. Soc. Perkin Trans. 2 (1997) 1977–1983.
[21] B. Sen, Anal. Chim. Acta 27 (1962) 515–518.
[22] J.L. Sudmeier and C.N. Reilley, Anal. Chem. 36 (1964) 1698–1706.
[23] Z. Szakács and B. Noszál, in preparation.
[24] O. Zekri, P. Boudeville, P. Genay, B. Perly, P. Braquet, P. Jouenne and J.-L. Burgot, Anal. Chem. 68 (1996) 2598–2604.