

BPC 01011

A GENERAL METHOD OF DERIVING THE BEST BINDING SITE MODEL CONSISTENT WITH EXPERIMENTAL BINDING DATA

J. Boiden PEDERSEN and Susanne M. PEDERSEN

Fysik Institut, Odense Universitet, DK - 5230 Odense M, and Klinisk Kemisk afd., Odense Sygehus, DK - 5000 Odense C, Denmark

Received 24th July 1984

Revised manuscript received 25th June 1985

Accepted 28th June 1985

Key words: Protein-ligand binding; Binding data analysis; Binding model; Homotropic binding; Binding-polynomial roots

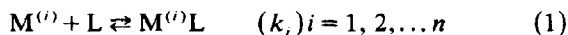
An analysis of binding data is presented which yields the best binding site model consistent with the experimental data. The analysis is applicable to homotropic binding and yields the number of independent sites, number of interacting sites (dimers and tetramers of sites), intrinsic association constants, and degree of interaction. The information is derived from the roots of a binding polynomial constructed by the fitted Adair constants.

1. Introduction

Most experimental investigations of the binding of a small molecule or ion, in the following called ligand L, to a macromolecule M are performed with the intention of identifying the sites on the macromolecule onto which the binding of L occurs and to learn about the direct or allosteric interactions between the sites. The desired information is usually obtained by fitting the binding data to an analytic expression derived from an assumed model. Many important models have been proposed [1–8] for various systems, in particular to describe the binding of oxygen to hemoglobin [2–8], but although both graphical [7,9,10] and iterative least-squares curve-fitting procedures [11–13] have been applied to the analysis of binding data there remain important unsolved problems, some of which will be discussed below.

Before the allosteric interaction was recognized as being of importance for the functioning of macromolecules it was often assumed that the binding sites were independent of each other. For the sake of simplicity this assumption is also quite popular in the current literature. For n indepen-

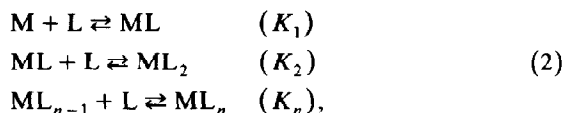
dent binding sites the binding is represented by n independent reaction schemes of the type



where superscript i indicates site i and k_i is the intrinsic association or binding constant for site i . The strategy for identifying the sites is to group together all sites with equal (or similar) binding constants into separate classes and then give the number of sites in the classes and the corresponding values of the intrinsic association constants. The idea of this representation is based on the assumption that the binding capacity of the functional groups that make up the sites is approximately independent of their surroundings. This implies that the value of the intrinsic binding constant is used to characterize a functional group and that the number of sites in the class gives the available number of this kind of functional group on the macromolecule. This is approximately true when the binding sites are not too close and when allosteric interactions are absent. The use of the independent site model (ISM) is, however, not recommended either for graphical or for computerized methods. It is difficult to test the validity

of the model, within the model, and when the ISM is not 'completely' consistent with the data then neither the derived intrinsic binding constants nor the values of the number of sites in the different classes will be reliable.

Contrary to the independent site description the thermodynamic description (TD) of the multiple, consecutive reaction steps



with association constant K_i for the i -th reaction step, is free of any assumptions concerning the nature of the binding. Note, however, that for both descriptions M must not form dimers, trimers, etc. Thus polysteric effects are excluded in both descriptions (eqs. 1 and 2), but allosteric effects are included in the thermodynamic description, eq. 2 [14]. The thermodynamic description is the most general description of the binding of L to M and therefore it gives rise to a better fit to the experimental data and thus to a better representation of the binding data in terms of a set of association constants. The thermodynamic description was originally introduced by Adair [8] in order to obtain a better description of oxygen binding to hemoglobin, and since then it has been applied to a variety of problems, ranging from the binding of various substances to macromolecules [3,7,11–14] to the formation of small inorganic complexes [15]. A simple iterative method to extract the binding constants from experimental data has been published [15], but is not generally applicable, especially not to macromolecular binding. Today's fitting procedures are all based on least-square nonlinear regression methods (e.g. refs. 11–13) and can only be effectively carried out on computers. The methods are currently being improved and statistical and other tests [14] of the reaction scheme (eq. 2) and of the goodness of a fit are being refined [11,12,16,17].

The complete generality of the thermodynamic description (eq. 2) implies that it does not give any direct information on the binding mechanism, which in fact is the main goal of a binding study. However, all homotropic binding models, includ-

ing allosteric but excluding polysteric ones, can be represented by the general reaction scheme (eq. 2), the problem being how to extract the best binding model. For several simple tetrameric models relations between the thermodynamic association constants K_1 – K_4 have been derived [4–7,11,18]. In addition, the best tetrameric model for a given set of appropriate binding data has been derived by comparing the experimental ratios of the thermodynamic association constants (i.e., K_i/K_{i+1} , etc.) with the ratios predicted by the known models [11,12,18]. This procedure, however simple, can only be applied to models for which the theoretical ratios of K_i/K_{i+1} have been evaluated, i.e., to a restricted class of models. This, for example, means that the above models can be tested only for systems where it is known that the number of binding sites is four. There are, however, many interesting systems for which the number of sites is different from four and where tetrameric interactions are less important but where dimeric interactions and site heterogeneity are important. For such systems a direct, best way of analysis has been absent.

It is shown below that from the thermodynamic association constants one can directly derive all relevant information concerning the binding of ligands to sites on the macromolecule and thus uniquely deduce the best binding site model, i.e., the number of monomers, dimers and tetramers of sites and the associated intrinsic association constants. The question of the applicability of a specific site model to a given set of experimental data is thus answered unambiguously. The analysis is based on a set of relations which connects the two descriptions (TD and ISM), and is related to the factorability of the binding polynomial of Wyman [19]. When a complete set of association constants is known in one description then these relations permit an evaluation of the association constants in the other description.

2. The relations connecting the descriptions

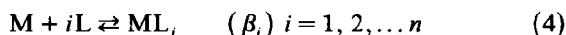
In order to facilitate the presentation we introduce the $n + 1$ gross association constants β_i de-

finied by

$$\beta_0 = 1$$

$$\beta_i = K_1 K_2 \dots K_i, \quad i = 1, 2, \dots, n \quad (3)$$

which are easily seen to be the equilibrium constants of the reactions



in which i ligands L bind to the macromolecule M .

It is well known (e.g., see ref. 20) and in fact readily shown that the binding isotherm for the reaction scheme (eq. 2), or its equivalent (eq. 4), is

$$\bar{v} = \frac{\sum_{i=1}^n i \beta_i c^i}{\sum_{i=0}^n \beta_i c^i} \quad (5)$$

where as usual \bar{v} is the average number of ligands bound to a single M , and $c = (L)$ the activity of free (unbound) L . Eq. 5 is sometimes called the Adair equation and is simply the mass action law applied to the reaction scheme (eq. 1) or its equivalent (eq. 4). As noted above, eqs. 4 and 5 are also valid when allosteric interactions are present [14].

The binding isotherm for n independent sites is also well known (e.g., see ref. 20), and can readily be derived from eq. 1. It is given by

$$\bar{v} = \sum_{i=1}^n \frac{k_i c}{1 + k_i c} \quad (6)$$

where k_i is the intrinsic association constant for site i (cf. eq. 1).

Eqs. 5 and 6 can both be written as

$$\bar{v} = c(d \ln P / dc) \quad (7)$$

where P is the so-called binding polynomial first introduced by Wyman. By comparing eqs. 6 and 7 it is seen that for the independent site model (ISM), P is given by

$$P_{\text{ISM}}(c) = \prod_{i=1}^n (1 + k_i c) \quad (8)$$

while the thermodynamic description (TD), used

in eq. 5, yields

$$P_{\text{TD}}(c) = \sum_{i=0}^n \beta_i c^i \quad (9)$$

Both eqs. 8 and 9 are polynomials of order n . By requiring these two polynomials to be identical one obtains a set of n criteria which furnishes the translation between the two descriptions. There is of course no a priori guarantee that the criteria for the TD \rightarrow ISM transformation can be fulfilled.

Note that all the coefficients in the binding polynomial are positive. This is evident from the interpretation of β_i as an equilibrium constant (cf. eq. 4) and may be obtained as well by regarding the binding polynomial as the grand canonical partition function for the macromolecule-ligand system, open with respect to ligands.

The transformation from the ISM to the TD is obtained by evaluating the coefficient to c^i in eq. 8 and equating this with β_i . This yields

$$\beta_i = \frac{1}{i!} \sum_{l_1} \sum_{l_2 \neq l_1} \dots \sum_{l_i} k_{l_1} k_{l_2} \dots k_{l_i}$$

$$= \sum_{l_1 < l_2} \dots \sum_{l_i} k_{l_1} k_{l_2} \dots k_{l_i} \quad (10)$$

These formidable looking expressions become simpler when written out explicitly, e.g.

$$\beta_1 = \sum_{i=1}^n k_i; \quad \beta_2 = \frac{1}{2} \sum_{i \neq j} k_i k_j = \left(\beta_1^2 - \sum_{i=1}^n k_i^2 \right) / 2$$

Eqs. 10 and 3 are closed-form expressions for the thermodynamic association constants K_i expressed in terms of a given set of intrinsic association constants k_i . These expressions are used in the proposed analysis of binding data and are also useful when comparing fits obtained by the two descriptions.

The transformation the opposite way, i.e., from the thermodynamic to the independent site description (TM \rightarrow ISM), is more complicated and not at all obvious. We require again the two polynomials in eqs. 8 and 9 to be identical. Since any polynomial of degree n has n roots and is uniquely determined by these roots (the constant term is always equal to 1) it follows that the two polynomials are identical if they have identical roots.

The roots of $P_{\text{ISM}}(c)$ are immediately seen to be equal to $-1/k_i$ ($i = 1, 2 \dots n$). Consequently, the independent site model is consistent with a given thermodynamic description if the independent site association constants (k_i) are determined from the thermodynamic constants (K_i or β_i) as the negative inverse of the roots of the binding polynomial

$$P_n(c) = \sum_{j=0}^n \beta_j c^j \quad (11)$$

Alternatively, the k_i terms must be chosen as the roots of

$$Q_n(c) = \sum_{j=0}^n (-1)^j \beta_j c^{n-j} \quad (12)$$

The acceptability of this choice of k_i is discussed in section 3.

3. The best binding site model

The best binding site model is obtained from the roots of the polynomial $Q_n(c)$ by arranging these into the smallest groups with physical significance. These groups are called irreducible linkage groups by Wyman.

A polynomial of degree n has always n roots. Thus, from an arbitrary set of thermodynamic association constants $\{\beta_i\}$, one can always define an associated set of constants $\{k_i\}$ as the roots of the characteristic polynomial $Q_n(c)$. However, only if the roots are real and positive can these k_i terms be interpreted as intrinsic association constants for independent sites. Since all β_j are positive it immediately follows that all real roots of P_n are negative which in turn implies that the real k_i terms determined this way are positive. We may thus conclude:

The independent site model is consistent with a given set of thermodynamic association constants $\{K_i\}$ if and only if all the roots of the characteristic polynomial $Q_n(c)$ are real. If this is the case then the intrinsic association constants $\{k_i\}$ of the independent sites are equal to the roots of $Q_n(c)$.

Complex roots will always occur in complex conjugated pairs, which ensures that eq. 6 remains real even when some of the k_i terms are complex.

However, a separate term in eq. 6 with a complex k_i is not real and therefore cannot be given a physical interpretation. The contribution to \bar{v} from a complex conjugated pair of k_i terms is real and may be written as

$$\begin{aligned} \bar{v}_p &= \frac{kc}{1+kc} + \frac{k^*c}{1+k^*c} \\ &= \frac{2k'c + 2|k|^2c^2}{1 + 2k'c + |k|^2c^2} \end{aligned} \quad (13)$$

where the asterisk denotes complex conjugation, k' and k'' are the real and imaginary parts of k , and $|k| = (k'^2 + k''^2)^{1/2}$ is the norm of k . The latter expression is identical with the contribution of \bar{v} from two interacting sites if $k' > 0$ (cf. eq. 5) with $n = 2$. Consequently:

An uncombined complex conjugated pair of zeroes of Q_n with a positive real part corresponds to a pair of interacting sites (a dimer of sites). The meaning of uncombined is given below. If we assume that the two interacting sites are identical then the interaction energy w between the sites may be determined as (e.g., see ref. 21)

$$w/kT = \ln((k')^2/|k|^2) \quad (14)$$

Instead of stating the interaction energy directly it might be convenient to give a 'degree of interaction' which could be defined as

$$d = 1 - \exp(w/kT) = (k'')^2/|k|^2 \quad (15)$$

Obviously, $d = k'' = 0$ for a pair of independent sites ($w = 0$), while for a pair of very strongly interacting sites ($-w/kT \gg 1$) one has $d = 1$ and $k' = 0$. It is interesting to note that from eq. 14 it follows that w is always negative or zero. This means that only attractive interactions between identical sites (positive cooperation) can be directly discovered. A repulsive interaction (negative cooperation) between the two identical sites will cause the two sites to show up as two nonidentical, noninteracting sites.

If a complex pair of roots ($k'_1 \pm ik''_1$) have a negative real part ($k'_1 < 0$) then they cannot be interpreted as an isolated pair of interacting sites since the negative sign of k'_1 will cause \bar{v}_p to be negative for small values of c . It is thus necessary

to combine this complex pair of roots with another complex pair of roots with a positive real part or with real roots in order to obtain physically acceptable results. We may thus immediately conclude:

A complex conjugated pair of roots of Q_n with a negative real part indicates that three or more sites are mutually interacting.

In order to specify the exact number of interacting sites it is necessary to discuss in more detail what is meant by a physically acceptable combination of sites (roots). Obviously, if a set of roots $\{k_i\}$ $i = 1, \dots, n_\alpha$ is to be considered an independent subunit (group) of the total system, then the average number of ligands bound to that group must be real and positive. Since $\bar{\nu}_\alpha$ is given by eq. 6 with n replaced by n_α and we have shown that eqs. 5 and 6 are identical when β_i is expressed in terms of the roots k_i by eq. 10, then it follows that $\bar{\nu}_\alpha$ is real and positive if and only if β_i is real and positive for $i = 1, \dots, n_\alpha$. Note that the summation in eq. 10 runs over members of the investigated group only. Since we are interested in the smallest possible groups we start with β_1 which we have already seen is equal to

$$\beta_1 = \sum_{i=1}^{n_\alpha} k_i \quad (16)$$

From this expression and the condition that β_1 must be real and positive it immediately follows that a single root represents a physically meaningful system, an independent site, if k_i is real and positive; in accordance with the discussion above. A complex root $k' + ik''$ becomes physically meaningful only when combined with its complex conjugate $k' - ik''$ and when $k' > 0$ since for these two roots $\beta_1 = 2k'$. A complex root $(k'_1 + ik''_1)$ with a negative real part must be combined with its complex conjugated root in order to make β_1 real and with other roots in order to make β_1 positive. A combination with a real root k_2 yields the condition

$$\beta_1 = 2k'_1 + k_2 > 0 \quad (17)$$

while a combination with another complex conjugated pair of roots $(k'_2 \pm ik''_2)$ yields

$$\beta_1 = 2k'_1 + 2k'_2 > 0 \quad (18)$$

The corresponding conditions for β_2 are

$$\beta_2 = |k_1|^2 + 2k_2k'_1 > 0 \quad (19)$$

$$\beta_2 = |k_1|^2 + |k_2|^2 + 4k'_1k_2 > 0 \quad (20)$$

For manual work some useful alternatives to eqs. 19 and 20 are the sufficient conditions $k_2 < k''_1$ and $k''_1 > k''_2$. However, such explicitly written conditions are useful only for small binding groups and when the total number of sites is small. For more complex systems one should let a computer, which is used anyway to determine the Adair constants and the roots of Q_n , find the combination of roots that are acceptable physical groups, i.e., where all β_i are real and positive. Such a search is easily performed by application of eq. 10.

Since we are only interested in the smallest possible linkage groups we proceed as follows. If there is no complex root with a negative real part then the system is simple and consists of only single sites (real roots) and pairs of interacting sites (complex roots) and no computer search is necessary. If there is a complex conjugated pair of roots with a negative real part then this pair is first combined with a real root in order to see whether the associated values of β_1 , β_2 and β_3 are all positive which indicates a trimer of sites. If no trimer is found we continue to examine tetrameric combinations with either two real roots or one complex conjugated pair with a positive real part. The search is continued with an increasing number of combining sites until all complex roots with negative real part have been combined with other roots into physically acceptable linkage groups. The remaining (uncombined) real roots and complex conjugated pair of roots with positive real part represent single independent sites (monomers) and pairs of interacting sites (dimers), respectively.

Perhaps it is relevant to denote the independent sites, the pair of sites, the tetramer of sites, etc., defined by the roots of Q_n - as normal sites, normal pair of sites, etc. These normal sites may be a combination of the local physical sites just like the vibrational normal modes of a molecule may be a combination of local modes. On a macromolecule the local sites (the functional groups) are generally far apart and interact only weakly by direct interaction. Allosteric interaction may, how-

ever, be long range. Consequently, it may be expected that the normal sites on a macromolecule are closely related to the physical sites. There is no direct way of getting around the problem of determining the normal sites rather than the local physical sites. The results of an analysis of binding data should be judged with this in mind, and comparison with independent observations (e.g., spectroscopic) may thus be essential for a conclusive identification of the binding sites.

4. Discussion

The present analysis deals with the binding of only one type of ligand to a non-associating macromolecule. Thus, competitors are either absent or present with a constant activity. For such systems the analysis determines the best binding site model consistent with the experimental data and its application is restricted only by the numerical method used to fit the data to eq. 5, which yields the gross association (Adair) constants β_i . The analysis is therefore particularly well suited to systems with a relatively small number of binding sites, which fortunately include many biologically interesting system.

The usefulness of binding polynomials and some of their properties have been discussed previously by others. In a series of papers, Bardsley and

Table 1

Adair constants (β_i)^a and associated roots (k_i) of the polynomial Q_4

	I	II	III
β_1	9.77×10^{-2}	1.00×10^{-2}	3.98×10^{-2}
β_2	1.32×10^{-3}	4.68×10^{-5}	5.01×10^{-3}
β_3	1.32×10^{-7}	9.55×10^{-6}	3.31×10^{-5}
β_4	9.77×10^{-5}	9.55×10^{-6}	6.61×10^{-5}
k_1, k_2	$-0.05 \pm i0.07$	$-0.04 \pm i0.04$	$-0.04 \pm i0.07$
k_3, k_4	$0.10 \pm i0.06$	$0.04 \pm i0.04$	$0.06 \pm i0.07$

^a The three sets of 'redundant' Adair constants are from ref. 16.

Waight [22] have considered in great detail the relationships between the magnitude of Hill plot slopes, apparent binding constants and factorability of binding polynomials and their Hessians. Closely related to the present method of analysis are the postulates of Gill and Wyman and a discussion of these by Bardsley et al. [22]. Whitehead [23] has derived conditions for various types of cooperativity allowed by the Adair equation and their relation with the algebraic character of binding polynomials. The previous approach [23] considered the factorability of the binding polynomial by finding explicit conditions for the various possibilities in terms of the Adair constants. Such conditions are very complex, are different for different values of n the total number of sites, and the complexity of the conditions in-

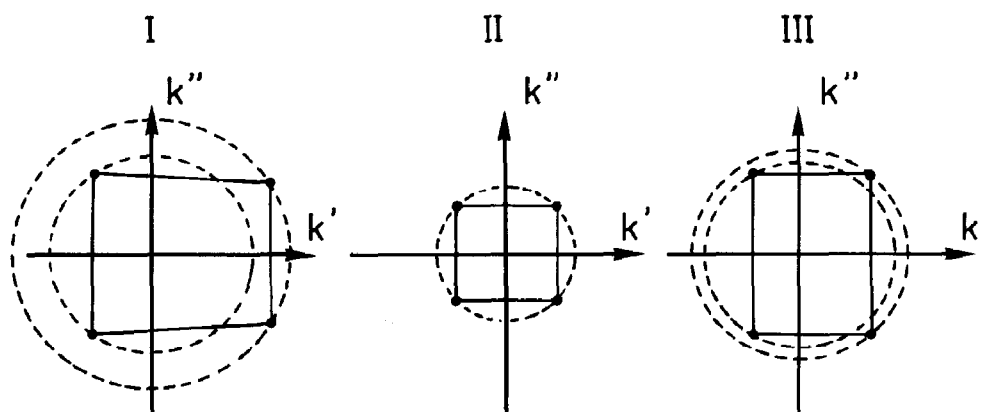


Fig. 1. Geometric root pattern in the complex plane of the three sets of roots of Q_4 corresponding to Reich's three set of redundant Adair constants (cf. table 1).

creases drastically with increasing n . The present approach is in some respects directed the opposite way. Here the binding polynomial is first completely factorized into (possibly unphysical) monomers by finding all the roots of Q_n and then combining these roots into physically acceptable linkage groups. Due to eq. 10 this combination of roots into linkage groups is easily done on a computer and can be performed manually for a small number of sites. Furthermore, the method of combination of roots is independent of the total number of sites.

It has been argued [16] that the gross association (or Adair) constants determined by fitting the experimental data to the Adair equation (eq. 5) are highly redundant, i.e., the constants cannot be determined very precisely, since very different sets of constants yield very similar fit. Our preliminary investigations (see below) show that this redundancy of the Adair constants does not carry over to the intrinsic association constants found as the zeroes of the characteristic polynomial. This means that different sets of Adair constants, which are equivalent with respect to the goodness of the fit, give rise to essentially identical microscopic binding site models when analyzed as proposed above.

As an example of the application of the method to an allosteric system we consider the three binding curves constructed by Reich [16] to illustrate the redundancy of the Adair constants. The three sets of Adair constants used to construct the binding curves are displayed in table 1. The sets differ by orders of magnitude in the individual constant and yet give almost identical binding curves (cf. fig. 3 of ref. 16). The roots of the corresponding polynomials Q_4 were calculated and are displayed in table 1. In all three cases the polynomial Q_4 has two pairs of complex conjugated roots with one pair having a negative real part. Thus, all three sets represent a system of four mutually interacting sites (a tetramer). The similarities of the three sets of roots is more clearly seen from the graphical illustration in fig. 1, where the roots are displayed in the complex plane. It is evident that both the magnitude of the roots and the root pattern are very similar for the different sets. Set 2 represents a perfect square while sets 1 and 3 represent distorted squares with a slightly higher

binding affinity. Thus, although the three sets of roots are not identical they represent binding site systems that are very similar and certainly much more similar than the three sets of Adair constants would seem to imply. It should be noted that these three sets of Adair constants have extreme differences. The Adair constants determined by Roughton et al. [11] for four different samples of sheep hemoglobin show a much smaller variation and thus a corresponding small variation in the root pattern.

The illustration used in fig. 1 is related to a recent work by Briggs [24], who showed that for a symmetric binding polynomial the Hill coefficient is conveniently expressed in terms of the roots of the polynomial and that this cooperativity measure is related to the geometric pattern of the roots in the complex plane. It should be noted that Briggs uses the roots of the binding polynomial P_n while we use the roots of the associated polynomial Q_n , since the latter are more directly connected to the binding constants. The two set of roots are each other's negative inverse.

In summary the proposed method of analysis of experimental binding data yields the best binding site model consistent with the experimental data without any assumptions of a specific mechanism, or other knowledge of the system. The method is particularly well suited for a computerized analysis.

References

- 1 L. Michaelis and M. Menten, *Biochem. Z.* 49 (1913) 333.
- 2 A.V. Hill, *J. Physiol.* 40 (1910) IV.
- 3 W.H. Forbes and F.J.W. Roughton, *J. Physiol.* 71 (1931) 229.
- 4 L. Pauling, *Proc. Natl. Acad. Sci. U.S.A.* 21 (1935) 186.
- 5 J. Wyman, *Adv. Protein Chem.* 4 (1948) 407.
- 6 J. Monod, J. Wyman and J.-P. Changeux, *J. Mol. Biol.* 12 (1965) 88.
- 7 D.E. Koshland, Jr., G. Némethy and D. Filmer, *Biochemistry* 5 (1966) 365.
- 8 G.S. Adair, *J. Biol. Chem.* 63 (1925) 529.
- 9 G. Scatchard, *Ann. N.Y. Acad. Sci.* 51 (1949) 660.
- 10 I.M. Klotz, F.M. Walker and R.B. Pivan, *J. Am. Chem. Soc.* 68 (1946) 486.
- 11 F.J.W. Roughton, F.R.S., A.B. Otis and R.L.J. Lyster, *Proc. R. Soc. B* 44 (1955) 29.

- 12 A. Cornish-Bowden and D.E. Koshland, Jr, *Biochemistry* 9 (1970) 3325.
- 13 J.G. Reich, G. Wangermann, M. Falck and K. Rohde, *Eur. J. Biochem.* 26 (1972) 368.
- 14 G. Weber and S.R. Anderson, *Biochemistry* 4 (1965) 1942.
- 15 J. Bjerrum, Metal ammine formation in aqueous solution (P. Haase & Son, Copenhagen, 1941).
- 16 J.G. Reich, *Stud. Biophys.* 42 (1974) 165.
- 17 J.G. Reich and I. Zinke, *Stud. Biophys.* 43 (1974) 91.
- 18 J.T.-F. Wong and L. Endrenyi, *Can. J. Biochem.* 49 (1971) 568.
- 19 J. Wyman, *Biophys. Chem.* 14 (1981) 135.
- 20 C. Tanford, *Physical chemistry of macromolecules* (Wiley, New York, 1961).
- 21 T.L. Hill, *An introduction to statistical thermodynamics* (Addison-Wesley, Reading, MA, 1962).
- 22 W.G. Bardsley and R.D. Waight, *J. Theor. Biol.* 72 (1978) 321; *J. Theor. Biol.* 85 (1980) 247.
- 23 E.P. Whitehead, *J. Theor. Biol.* 93 (1981) 547.
- 24 W.E. Briggs, *Biophys. Chem.* 18 (1983) 67.