Biophysical Chemistry 27 (1987) 59–65
Elsevier

BPC 01156

Binding of *n*-mers to one-dimensional lattices with longer than close-contact interactions

Yi-der Chen

Laboratory of Molecular Biology, National Institute of Diabetes, and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, U.S.A.

Received 9 December 1986
Accepted 4 February 1987

Ligand binding; Cooperativity; Scatchard plot; Binding isotherm

A general formalism is derived for the evaluation of binding isotherms of *n*-mers (ligands) to one-dimensional polymers in the presence of ligand-ligand interactions which extend over several binding sites with distance-dependent interaction energies (multi-parameter model). This is an extension of the usual *n*-mer binding theory developed by several investigators in which ligand-ligand interaction occurs only when two ligands are in close contact (one-parameter model). The difference in binding isotherms between a one-parameter model and a multi-parameter model is studied numerically using the present formalism.

1. Introduction

The cooperative binding of large ligands (*n*-mers) to long one-dimensional lattices (polymers) has been studied by a number of people using several different approaches. Using ingenious conditional-probability arguments, McGhee and Von Hippel [1] obtained a closed-form expression from which a Scatchard plot can be readily constructed. The same formula was obtained by Zasedatelev et al. [2] based on the combinatorial method and by Schellman [3] based on the sequence-generating function method of Lifson [4]. On the other hand, Tsuchiya and Szabo [5] have obtained the same formula using the standard matrix method. Recently, Chen et al. [6] have extended the sequence-generating function method to the binding of two species of long (*n*-mer) ligand to a one-dimensional lattice.

In these *n*-mer (and most other) binding studies, the effect of ligand-ligand interactions on binding isotherms is usually contained in a single cooperativity parameter defined as

$$y = \exp(-w/kT) \tag{1}$$

where w is the nearest-neighbor interaction energy between two bound ligands. That is, ligand molecules bound on the polymer are assumed to interact only when they are in close contact. In general, this 'one-parameter' model should work well if the distance between two binding sites (lattice spacing) on the polymer is relatively large (such as in F-actin molecules where the spacing is about 55 Å). On the other hand, if the lattice spacing is small (such as in DNA molecules where the spacing is about 3.4 Å), ligand-ligand interaction is expected to extend beyond the close-contact domain. That is, ligand molecules can interact even if they are separated by one or more empty binding sites (see fig. 1). Then, a rigorous binding theory should contain more than one cooperativity parameter in the formalism. This 'multi-parameter' binding problem has been treated before for the monomer ($n = 1$) case (see, e.g., ref. 3). In this

Correspondence address: Y. Chen, Laboratory of Molecular Biology, National Institute of Diabetes, and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, U.S.A.

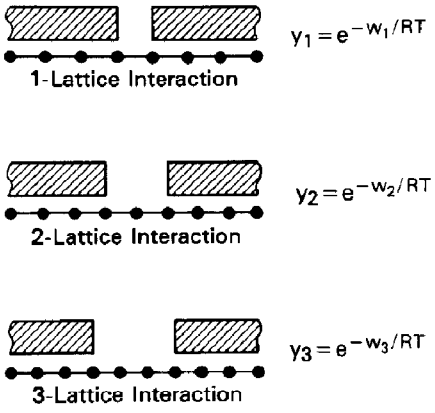

Fig. 1. Possible long-range interactions between two neighboring n-mer ligands. Each dot represents a binding site on the polymer and the space between two binding sites is called the lattice spacing of the polymer. When two ligands are separated by i lattice spacing (with $i-1$ binding sites in between), the interaction is referred to as the i-lattice interaction. The 1-lattice interaction here corresponds to the usual nearest-neighbor interaction used in most binding theories.

paper, we report the derivation of a general formalism for n-mer ligands with n larger than 1. The formalism is easy to apply and should prove useful in analyzing nonspecific binding of proteins to DNA molecules.

In section 2, the formalism is derived for systems with only two cooperativity parameters. In this case, both the matrix method and the sequence-generating function method can be used to derive the formalism. The extension of the derivation to cases with an arbitrary number of parameters of interactions using the matrix method is presented in section 3. In section 4, the formalism is applied to an illustrative binding system. The question of whether the binding isotherm of a multi-parameter model can be reproduced by the usual one-parameter model is discussed.

2. Theoretical formulation for cases with two cooperativity parameters

For simplicity and clarity, we will consider the case that only 1-lattice and 2-lattice interactions are present in the system in this section. The extension to longer range interactions will be given in section 3. For this simple two-parameter case, both the matrix method and the sequence-generating function method can be used to derive the binding isotherm.

2.1. The matrix method

The basic ingredient to calculate the fraction of bound sites (θ) in dilute solutions is to differentiate the grand partition function (Ξ) of the system with respect to the concentration (c) of the ligand:

$$\theta = \frac{1}{N}\frac{\partial \ln \Xi}{\partial \ln c} = \frac{1}{N}\frac{\partial \ln \Xi}{\partial \ln x} \tag{2}$$

where N is the number of binding sites on the lattice and

$$x = Kc. \tag{3}$$

The K in eq. 3 is the intrinsic binding constant of a ligand to an isolated lattice site(s).

In the matrix method, the grand partition function can be obtained from the 'transfer matrix' of the system (see refs. 5, 7 and 8 for details). To construct the transfer matrix, it is necessary to classify and label a binding site into states according to whether it is empty or bound with ligand and to which part of the ligand the site is in contact. As shown in fig. 2, a site is labelled as state 1 when it is empty and the site to its right is not covered by a ligand. State 2 is also an empty site, but it is just in front (to the left) of a bound

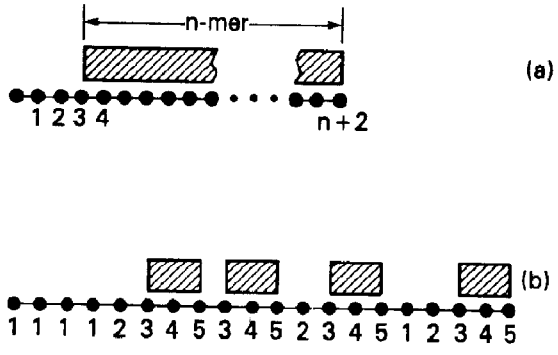

Fig. 2. (a) The labelling of the state of a site in systems with 1-lattice and 2-lattice interactions. For ligands that can cover n binding sites (n-mers), each binding site can exist in $n+2$ states. (b) The complete state specification of each binding site for tri-mer ligand systems.

ligand. States 3, 4, ..., $n+2$ are those sites covered by the same ligand. The reason that each site is labelled differently is that the sites are not equivalent to each other. For example, an empty site in state 2 can be followed (to its right) only by a site in state 3, not state 4 or others. Similarly, a site in state 3 can be followed only by a site in state 4. For example, the labelling of all sites of a tri-mer binding system with 1- and 2-lattice neighbor interactions is explicitly indicated in fig. 2b. With this label system for the sites, the transfer matrix $\boldsymbol{M}$ (of order $(n+2)\times(n+2)$) can be written down for an arbitrary n-mer binding system as

$$\boldsymbol{M}=\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & \cdots\cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \cdots\cdots & 0 \\ 0 & 0 & 0 & x^{\frac{1}{n}} & 0 & 0 & \cdots\cdots & 0 \\ 0 & 0 & 0 & 0 & x^{\frac{1}{n}} & 0 & \cdots\cdots & 0 \\ \vdots & & & & & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots\cdots & x^{\frac{1}{n}} \\ x^{\frac{1}{n}} & x^{\frac{1}{n}}y_2 & x^{\frac{1}{n}}y_1 & 0 & 0 & 0 & \cdots\cdots & 0 \end{pmatrix} \tag{4}$$

where x is defined in eq. 3 and y_1 and y_2 are the cooperativity parameters defined in fig. 1.

We would like to mention two points concerning the transfer matrix in eq. 4. First, we have assigned the value 1 for the weight of an empty site and $x^{\frac{1}{n}}$ for a site bound with ligand. Second, we have adopted the convention (see ref. 7 for details) that each M_{ij} in eq. 4 represents the weight of a site in state i when the site on the right is in state j. For example, M_{11} is the weight of an empty site when it is followed (to its right) by another empty site.

The grand partition function for a circular lattice of N sites can be expressed in terms of this transfer matrix as

$$\Xi(x)=\text{Trace}\,\boldsymbol{M}^N=\lambda_1^N+\lambda_2^N+\ldots+\lambda_{n+2}^N, \tag{5}$$

where $\lambda_1>\lambda_2>\ldots>\lambda_{n+2}$ are the $n+2$ eigenvalues of $\boldsymbol{M}$.

For a linear lattice of N sites, the grand partition function can be expressed as

$$\Xi(x)=(1\ 1\ 1\ 0\ \ldots\ 0)\boldsymbol{M}^N\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}. \tag{6}$$

For finite N, $\Xi(x)$ can be calculated by direct multiplication and then numerically differentiated to yield θ. For an infinite linear lattice, the end effect becomes negligible and eq. 5 can be used for its grand partition function. Furthermore, the contribution from the smaller eigenvalues (λ_2, $\lambda_3,\ldots\lambda_{n+2}$) in eq. 5 also becomes negligible. As a result, the fraction of bound sites can be expressed in terms of the largest eigenvalue as

$$\theta=\frac{\partial\ln\lambda_1}{\partial\ln x}. \tag{7}$$

To obtain the eigenvalues of $\boldsymbol{M}$, let us consider the eigenvalue problem:

$$\boldsymbol{M}\begin{pmatrix}1\\c_1\\c_2\\\vdots\\c_{n+1}\end{pmatrix}=\lambda\begin{pmatrix}1\\c_1\\c_2\\\vdots\\c_{n+1}\end{pmatrix}. \tag{8}$$

This matrix equation is equivalent to the set of linear equations:

$$1+c_1=\lambda, \tag{9a}$$

$$c_2=\lambda c_1, \tag{9b}$$

$$x^{\frac{1}{n}}c_{i+1}=\lambda c_i,\ (i=2,3,\ldots,n) \tag{9c}$$

$$x^{\frac{1}{n}}(1+c_1y_2+c_2y_1)=\lambda c_{n+1} \tag{9d}$$

From eq. 9c we have

$$c_{n+1}=\lambda^{n-1}x^{-\frac{n-1}{n}}c_2. \tag{10}$$

Substituting this into eq. 9d, we obtain

$$x^{\frac{1}{n}}(1+c_1y_2+c_2y_1)=\lambda^n x^{-\frac{n-1}{n}}c_2. \tag{11}$$

Eliminating c_1 and c_2 from eqs. 9a, 9b and 11, we find

$$x=\frac{\lambda^{n+1}(\lambda-1)}{1+(\lambda-1)y_2+\lambda(\lambda-1)y_1}. \tag{12}$$

This is the characteristic equation of $\boldsymbol{M}$. After differentiating eq. 12 with respect to x, we obtain (cf. eq. 7)

$$\theta=\frac{(\lambda-1)A}{(n+2)\lambda A-(n+1)A-\lambda y_2-\lambda(2\lambda-1)y_1} \tag{13}$$

where

$$A = (\lambda - 1)^{-1} + y_2 + \lambda y_1. \tag{14}$$

Eqs. 12 and 13 are the basic equations to calculate the binding isotherm, $\theta(x)$. In principle, the calculation should contain these two steps: (1) Obtain the largest λ from eq. 12; (2) θ is then calculated from eqs. 13 and 14 with this λ. In practice, the easiest way is to calculate x and θ from eqs. 12 and 13 separately as functions of λ.

2.2. *The sequence-generating function method*

For this simple two-parameter system, the binding isotherm can also be obtained by the sequence-generating function method [3,7]. In this method, the configuration of bound ligands on a lattice is divided into a combination of empty and bound sequences. For example, the configuration of the usual one-parameter system can be grouped into alternating sequences of empty sites and contiguously bound ligands. In systems with one and two lattice interactions, any configuration can be grouped into a combination of three types of sequences: (A) empty sites; (B) contiguously bound ligands; (C) units of one bound ligand with one empty site adjacent to its right (see fig. 3). Let $\psi_A(\gamma)$, $\psi_B(\gamma)$, and $\psi_C(\gamma)$ be the sequence generating functions of the three sequences. Then,

$$\psi_A(\gamma) = \sum_{i=1}^{\infty} \xi_A(i)\gamma^{-i}, \tag{15}$$

$$\psi_B(\gamma) = \sum_{j=1}^{\infty} \xi_B(j)\gamma^{-nj}, \tag{16}$$

$$\psi_C(\gamma) = \sum_{k=1}^{\infty} \xi_C(k)\gamma^{-(n+1)k}, \tag{17}$$

where $\xi_A(i)$ is the partition function (p.f.) of an A (empty sites) sequence of i units long, $\xi_B(j)$ is the p.f. of a B sequence of j units long (on $n \cdot j$ sites), and $\xi_C(k)$ is the p.f. of a C sequence of k units long (on $(n+1) \cdot k$ sites). Then, the grand partition function of the entire system at large N (the number of binding sites) is equal to

$$\Xi = \gamma_1^N \tag{18}$$

where γ_1 is the largest root of the following characteristic equation (see ref. 4)

$$\begin{vmatrix} 1 & -\psi_B(\gamma) & -\psi_C(\gamma) \\ 0 & 1 & -\psi_C(\gamma)y_1 \\ -\psi_A(\gamma) & -\psi_B(\gamma)y_2 & 1 \end{vmatrix} = 0. \tag{19}$$

Thus, if $\psi_A(\gamma)$, $\psi_B(\gamma)$ and $\psi_C(\gamma)$ are available, various thermodynamic functions of the binding system can be obtained from eqs. 18 and 19.

The partition function of empty sites $\xi_A(i)$ is equal to unity if the weighting factor for empty sites is assigned 1. Thus, $\psi_A(\gamma)$ in eq. 15 becomes

$$\psi_A(\gamma) = \frac{1}{\gamma - 1} \tag{20}$$

Since there are $(j-1)$ neighbors between j contiguous ligands, the partition functions of B and C sequences can be obtained as

$$\xi_B(j) = x^j y_1^{j-1}, \tag{21}$$

$$\xi_c(k) = x^k y_2^{k-1}. \tag{22}$$

After substituting eqs. 21 and 22 into eqs. 16 and 17 and carrying out the summation, $\psi_B(\gamma)$ and $\psi_C(\gamma)$ are obtained as

$$\psi_B(\gamma) = \frac{x}{\gamma^n - xy_1}, \tag{23}$$

$$\psi_C(\gamma) = \frac{x}{\gamma^{n+1} - xy_2}. \tag{24}$$

With eqs. 20, 23 and 24, the characteristic expression, eq. 19, can be shown to be identical to that in eq. 12.

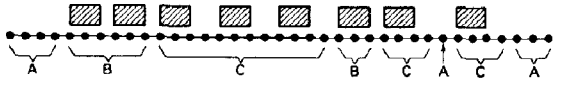

Fig. 3. Classification of binding sites into three types of sequences for systems with 1-lattice and 2-lattice interactions. Type A contains empty sites. Type B contains sites completely covered by ligands. Type C contains sites covered by ligands with one empty site on the right side of each ligand.

3. Generalization to an arbitrary range of interactions

The characteristic equation of the n-mer binding problem was derived only for systems with two cooperativity parameters in section 2. In this

section, we extend the derivation to a longer range of interaction. That is, we consider the general case that interaction can exist between two ligands separated by 1, 2,..., and p lattice spacings. As we are interested only in large ligand molecules, it is assumed that n is always larger than p. This eliminates the complicated second-neighbor interaction problem.

Since the number of sequences in the sequence-generating method becomes very large when p is large, it is impractical to derive the characteristic equation from the sequence-generating method. On the other hand, the problem can be handled easily using the matrix method as shown below.

Let the state of each site on the lattice be labelled as shown in fig. 4. That is, a site is labelled as state 1 if there are $p-1$ or more empty sites between this site and the left end of a bound ligand, labelled as state 2 if there are exactly $p-2$ empty sites between this site and the left end of a bound ligand, etc. The total number of states each site can have is $p+n$. Then, the transfer matrix of this binding system can be written down as

$$M = \begin{pmatrix} 1 & 1 & 0 & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & & & & & & & \\ 0 & 0 & & & & \boldsymbol{D} & & & \\ \cdot & \cdot & & & & & & & \\ \cdot & \cdot & & & & & & & \\ x^{\frac{1}{n}} & x^{\frac{1}{n}} y_p & & x^{\frac{1}{n}} y_{p-1} & \dots & x^{\frac{1}{n}} y_1 & 0 & \dots & 0 \end{pmatrix} \tag{25}$$

where $\boldsymbol{D}$ is a diagonal matrix of order $(p+n-2) \times (p+n-2)$ with its elements defined as

$$\begin{aligned} D_{ij} &= \delta_{ij} \qquad (i, j = 1, 2, \dots, p-1); \\ &= x^{\frac{1}{n}} \cdot \delta_{ij} \quad (i, j = p, p+1, \dots, p+n-2). \end{aligned} \tag{26}$$

The δ_{ij} in eq. 26 is the usual Kronecker delta function.

The characteristic equation of $\boldsymbol{M}$ in eq. 25 can be derived using the same procedure as shown in section 2. The final expression is

$$x = \left[\lambda^{n+p-1}(\lambda-1)\right]\left[1 + (\lambda-1)y_p + \lambda(\lambda-1)y_{p-1} + \dots + \lambda^{p-1}(\lambda-1)y_1.\right]^{-1} \tag{27}$$

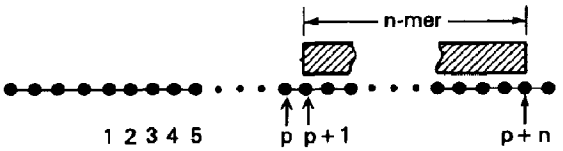

Fig. 4. The labelling of the state of a binding site in systems with ligand-ligand interactions extending over $1, 2, \dots, p$ lattices.

With eq. 7, eq. 27 can be used to derive the fraction of bound sites:

$$\theta = \frac{(\lambda-1)B}{(n+p)\lambda B - (n+p-1)B - C} \tag{28}$$

where

$$B = (\lambda-1)^{-1} + y_p + \lambda y_{p-1} + \lambda^2 y_{p-2} + \dots + \lambda^{p-1} y_1. \tag{29}$$

$$C = \lambda\left\{ y_p + (2\lambda-1)y_{p-1} + (3\lambda^2 - 2\lambda)y_{p-2} + \dots + \left[p\lambda^{p-1} - (p-1)\lambda^{p-2}\right] y_1 \right\}. \tag{30}$$

4. Illustrative numerical calculations

In this section, we will calculate the binding isotherms of a few three-parameter models using the formalism derived in section 3 and compare them with those from the ordinary one-parameter model. The purpose is to illustrate the differences in binding isotherms between a one-parameter and a multi-parameter model.

The basic equations for calculating the binding isotherm of an arbitrary multi-parameter model are given in eqs. 27–30. In general, the Scatchard plot is obtained by first calculating x and θ as a function of λ from eqs. 27 and 28 separately and then plotting θ/x as a function of θ.

In fig. 5, the calculated Scatchard plots of a penta-mer ($n=5$) system with (i) $y_1 = y_2 = y_3 = 10$ and (ii) $y_1 = 15$, $y_2 = 10$, $y_3 = 5$ are presented along with those calculated from the one-parameter model. As shown in the figure, it is impossible to fit the entire Scatchard plot of the two three-parameter models with a one-parameter model, no matter how the n and y values of the latter are

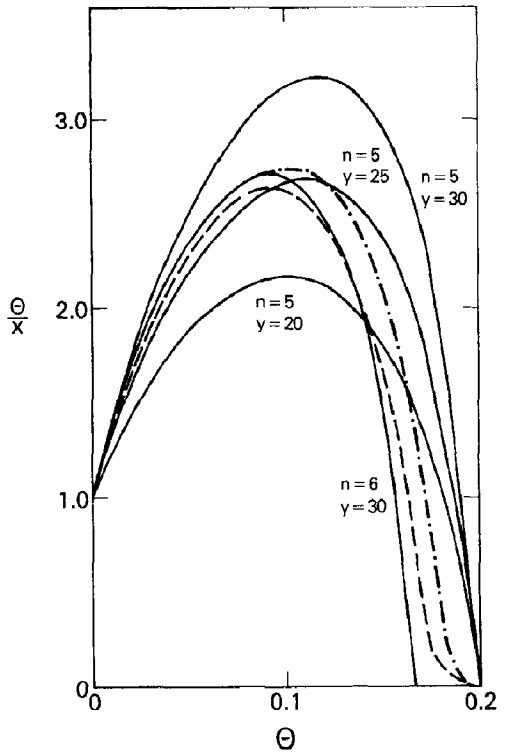

Fig. 5. The scatchard plot of a three-parameter model of $n = 5$ with (i) $y_1 = y_2 = y_3 = 10$ (- - - -) and (ii) $y_1 = 15$, $y_2 = 10$, $y_3 = 5$ (· – · – ·). The solid curves are those from the one-parameter model. The n and y for each one-parameter model are indicated by each curve.

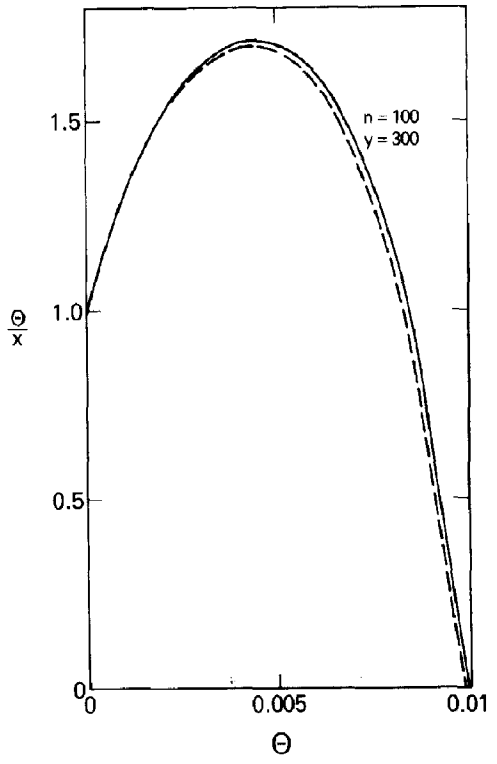

Fig. 6. The Scatchard plots of one-parameter (———) and three-parameter models (— — —) with $n = 100$. The cooperativity parameters are $y_1 = 150$, $y_2 = 100$, $y_3 = 50$ for the three-parameter model and $y = 300$ for the one-parameter model. The Scatchard plot of $n = 100$ and $y_1 = y_2 = y_3 = 100$ is also calculated. It is indistinguishable from the dashed curve.

varied. In general, a good fitting between a three-parameter model and a one-parameter model can be found at low θ values (e.g., $\theta = 0.1$ in fig. 5), but not at high θ values.

The Scatchard plots of a three-parameter model with $n = 100$ and (i) $y_1 = y_2 = y_3 = 100$ and (ii) $y_1 = 150$, $y_2 = 100$, $y_3 = 50$ are also calculated and shown in fig. 6. In contrast to the $n = 5$ case, the binding isotherms of both cases (i) and (ii) can be reproduced fairly well by a one-parameter model with $y = 300$ and $n = 100$.

5. Discussion

The main purpose of this paper is to present the derivation of a general formalism for the calculation of binding isotherms for n-mers (ligands) with long-range interactions. It is shown that both the matrix method and the sequence-generating function method can be used to derive the formalism if the range of ligand-ligand interactions is relatively short. On the other hand, if very long-range interactions are present, only the matrix method can be applied easily. This illustrates the advantage of using the matrix method in studying n-mer binding problems, as first pointed out by Tsuchiya and Szabo [5]. In another paper (Chen and Szabo, in preparation), the use of the matrix method in the study of simultaneous binding of two or more species of n-mers (with different sizes and different cooperativities) to one-dimensional lattices will be reported.

In contrast to the usual one-parameter case (the McGhee-Von Hippel case), a closed expression for the Scatchard plot is not available for n-mer ligands with long-range interactions. However, the binding isotherm or the Scatchard plot of the system can be evaluated easily from the characteristic equations in eqs. 27 and 28 (along with eqs. 29 and 30). We would like to emphasize that, after this work was finished, equations equivalent (but not identical) to eqs. 27 and 28 have been derived by Nechipurenko and Gursky [9] based on probabilistic arguments.

As discussed in section 4, when the ligand size (n) is small, marked differences between multi-parameter and one-parameter models can be observed in the Scatchard plot at large θ values. Thus, if the entire binding curve is available, a Scatchard plot may be used to differentiate between a one-parameter and a multi-parameter model. However, if only partial binding isotherms at low saturation are available, then model differentiation becomes rather difficult. Similarly, if n is very large (much larger than the range of ligand-ligand interactions, p), the Scatchard plot of a multi-parameter model is indistinguishable from that of a one-parameter model (see fig. 6). As a result, model differentiation based on binding isotherms is also impractical in this case.

References

1 J.D. McGhee and P.H. von Hippel, J. Mol. Biol. 86 (1974) 469.
2 A.S. Zasedatelev, G.V. Gurskii and M.V. Volkenstein, Mol. Biol. 5 (1971) 245.
3 J.A. Schellman, Isr. J. Chem. 12 (1974) 219.
4 S. Lifson, J. Chem. Phys. 49 (1964) 3705.
5 T. Tsuchiya and A. Szabo, Biopolymer 21 (1982) 979.
6 Y. Chen, A. Maxwell and H.A. Westerhoff, J. Mol. Biol. 190 (1986) 201.
7 T.L. Hill, Cooperativity theory in biochemistry (Springer-Verlag, New York, 1985).
8 D. Poland and H.A. Scheraga, Theory of helix-coil transitions in biopolymers (Academic Press, New York, 1970).
9 Y.D. Nechipurenko and G.V. Gursky, Biophys. Chem. 24 (1986) 195.