Biophysical Chemistry 5 (1976) 319–326

BINDING ISOTHERMS AND THE INTERACTION BETWEEN PROFLAVINE AND A DNA OF HIGH G · C CONTENT

J. RAMSTEIN, M. LENG and N.R. KALLENBACH
Department of Biology, University of Pennsylvania, Philadelphia, Pa. 19174, USA
and Centre de Biophysique Moléculaire, La Source, Orléans, France

Received 9 October 1975

Binding isotherms corresponding to several situations of ligand binding to a linear polymer are calculated, including cases of cooperativity or anticooperativity between the bound ligand states, multiple binding modes that are competitive or non competitive, and possible exclusion of an arbitrary number of adjacent sites upon occupancy of a site by a single ligand. The sequence generating function method of Lifson and Bradley is used, requiring the assumption that no end effects are involved. The case of strong binding of the dye proflavine to a DNA of high G·C content, that of *M. lysodeikticus*, is considered in detail, and a single model capable of reconciling the available kinetic and equilibrium data on this system, involving two competing binding modes, is discussed.

1. Introduction

Associations between biological macromolecules and various agents capable of binding or complexing them are involved at all levels of biological activity, and thus have been intensively studied. In the case of DNA, the interactions of a variety of ions, intercalating dyes, antibiotics, mutagens and proteins with DNA or model helical polynucleotides have been investigated in vitro by determination of binding isotherms that relate the concentrations of bound to free ligand as the latter is varied over a suitable range. In principle these isotherms contain the essential thermodynamic information characterizing the equilibrium constant(s), K, and total available number of binding sites, n. Extraction of the underlying thermodynamic parameters requires postulating some model for the binding process, and subsequent fitting of the parameters characterizing the model. Thus, analysis of binding isotherms resembles kinetic analysis of a reaction in that it is difficult or impossible to establish a mechanism unequivocally, although numerous possible mechanisms can be excluded.

In the case of linear polymers, such as DNA and polynucleotides, the process of predicting isotherms for specific models of binding is generally simpler. If it is permissible to neglect ends of the polynucleotide chains, the method of sequence generating functions due to Lifson [1] furnishes a direct means of computing the isotherms. Accounts of applications of this method to the general problem of cooperative multisite binding of ligands to DNA have been previously published [2,3] and the closely related problem of the binding of complementary monomers or oligomers to polynucleotide chains has also been treated by Lifson's method [4]. After a brief review of the method we apply it to the specific problem of the intercalation process of proflavine in a high G · C content DNA.

In the range of our experimental conditions ($P/D > 5$, D being the concentration of dye and P the concentration of phosphate) different studies have provided several lines of evidence which show that acridines such as proflavine form with the DNA a complex called the intercalation complex. In this complex proflavine is intercalated between two base pairs, the planar dye being parallel to the base pairs. Li and Crothers [5] using a relaxation kinetic method have found that the intercalation of proflavine in the case of calf thymus DNA (42% G + C) proceeds in a two-stage mechanism,

$$\text{DNA(CT)} + \text{Proflavine} \rightleftharpoons C_1 \rightleftharpoons C_2 ,$$

where C_2 represents the intercalation complex and C_1

is an intermediate complex of unknown structure. The intermediate complex C_I contributes only a small fraction of the total binding (about 7% in 0.1 M Na^+).

In previous work [6,7] we have studied the intercalation of proflavine in the DNA of *Micrococcus lysodeikticus* (72% G + C) using a temperature-jump relaxation kinetic method and viscosity. Our results are also in agreement with the existence of two complexes but the bound dye is distributed almost equally between them. However the kinetic results alone did not permit us to give any precise mechanism for the interaction of proflavine and *M. lysodeikticus* DNA. It is then interesting to use this thermodynamical approach in the case of the interaction between proflavine and a G–C rich DNA in order to get information concerning the mechanism. More generally all previous discussions of the experimental binding isotherms of proflavine –DNA system assumed a single type of complex and are then basically incorrect. It thus seems appropriate to calculate isotherms for processes involving at least two binding modes.

2. Experimental methods

In this work we used *Micrococcus lysodeikticus* DNA (72% G + C). The purification of this DNA has been described elsewhere [8]. Proflavine was a gift of Dr. M. Charlier and was purified according to the method of Weill and Calvin [9].

The amount of bound proflavine per base pair $(r_1 + r_2)$ was determined from the visible absorption spectra of the DNA–proflavine solution [10], assuming that complex C_1 and C_2 have the same molar extinction coefficient ϵ_B [5]. We used the procedure of Li and Crothers [5] for the determination of ϵ_B and K, molar extinction coefficient and intrinsic association constant respectively. We found $\epsilon_B = 15\,440\ \mathrm{cm^{-1}\ M^{-1}}$ at $\lambda = 430$ nm irrespective of DNA base composition and $K = 68\,000\ \mathrm{M^{-1}}$. K can be defined by the following relation

$$K = \lim_{(r_1+r_2)\to 0} \left(\frac{r_1 + r_2}{[1 - (r_1 + r_2)]\, C_F} \right),$$

where C_F stands for the concentration of free proflavin.

Absorption spectra were determined on a spectrophotometer Cary 15.

All solutions were prepared in a standard buffer consisting of 0.1 M sodium chloride, 0.01 M sodium acetate, 10^{-4} M EDTA.

The experiments have been carried out by successive additions of concentrated DNA solution to a constant amount of proflavine.

3. Theoretical models

The problem of main concern here involves binding of a ligand capable of occupying more than one kind of site on a DNA helix represented as a homogeneous lattice of binding sites. We consider cases of competition between various modes, which arise in discussing the effect of salts on the equilibrium between DNA and acridine dyes, for example. In essence, the SGF method determines the generating function of the grand canonical partition functions Z_N for chains including N binding sites. avoiding the complicated combinatorial or matrix calculations required to obtain Z_N directly. Average properties of the ensemble of chains and ligands can be determined from the SGF, without actually calculating Z_N itself. The method thus concenrs obtaining the sum

$$\Gamma(x) = \sum_{N=0}^{\infty} Z_N X^{-N},$$

where X represents an effective dummy variable, the significant value of which is determined by the condition that $\Gamma(X)$ just diverge [1]. The particular features of the binding process of interest are introduced by assigning statistical weights corresponding to the various modes and their equilibria, as well as the presence of cooperative (or anticooperative) interactions, according to the following set of rules:

(1) Each lattice site, regardless of state of occupancy receives a statistical weight factor X^{-1}.

(2) A site occupied by a single ligand, in a particular mode, m say, such that no neighboring sites are filled receives a factor

$$\omega_m = \exp(-\Delta G_m^0/RT), \qquad (1)$$

where ω_m is the statistical weight corresponding to the mode m, as for example the process A or B in fig. 1, related to the intrinsic binding constant K_m for that mode

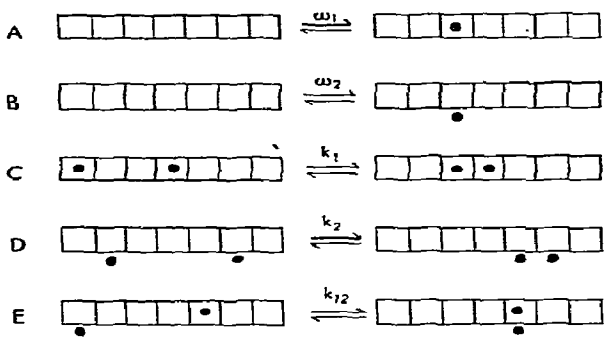


Fig. 1. Schematic of some possible binding modes on a helical duplex represented as a horizontal ladder, involving: (A) internal binding (intercalation) of a ligand between adjacent base pairs, (B) external binding (at the phosphate or in one of the grooves of the helix, (C) cooperative ($k_1 > 1$) or anti-cooperative ($k_1 < 1$) interaction between adjacent intercalated ligands, (D) similar interaction between adjacent externally bound ligands, (E) interaction between bound ligands in two different modes.

$$\omega_m = K_m a_f \approx K_m c_f , \tag{2}$$

where a_f denotes the thermodynamic activity and c_f the molar concentration of free ligand in solution.

(3) Two adjacent occupied sites of a particular mode m receive an additional statistical weight factor, k_m, representing the equilibrium constants

$$k_m = \exp(-\Delta G_m^E/RT) \tag{3}$$

of the processes depicted in fig. 1C or D, where ΔG_m^E, is the *excess* free energy involved in bringing a second ligand bound in the same mode from a remote lattice site adjacent to a reference occupied site. The case $\Delta G_m^E < 0$ or $k_m > 1$ leads to cooperatively interacting sites; conversely $\Delta G_m^E > 0$ or $k_m < 1$ leads to anti-cooperativity among the sites.

(4) Two adjacent occupied sites of two *different* modes, m,m' say, receive a similar statistical weight factor

$$k_{mm'} = \exp(-\Delta G_{mm'}^E/RT) = k_{m'm} , \tag{4}$$

$k_{mm'}$ being the equilibrium constant for bringing two remote occupied sites, in modes m and m', to adjacent sites, and $\Delta G_{mm'}^E$ the excess molar free energy, of this process. An example is illustrated schematically in fig. 1E.

(5) Site *exclusion* between a particular occupied site in a given mode can be introduced by requiring a set of l lattice sites adjoining either side of the occupied site to remain unfilled. The overall weight of such a bound site then is given by

$$\omega_m X^{-(l+1)} .$$

since $l + 1$ lattice sites are required for binding a single ligand. This case is of interest when large size ligands are involved, for example, such as oligonucleotides [11] repressor proteins, histones, polymerases or oligopeptides [12,13].

The above rules, together with the SGF procedure, permit rapid calculation of the binding isotherms for any of a wide choice of cases of interest. Note that whenever these rules lead to excessive complexity the relevant constants k_m or $k_{mm'}$ above can be set to unity, corresponding to lack of any interaction among the bound states of the sites.

4. Calculation of isotherms

The sequence generating functions corresponding to the unoccupied sites, as well as the occupied sites for each binding mode, are constructed from the above rules by summing the contributions of adjacent sets of 1,2,3, ..., etc., sequences of sites in each state. Thus, the SGF for unoccupied sites of the lattice is given by

$$U_0(X) = \sum_{i=1}^{\infty} X^{-i} = \frac{1}{X-1} \tag{5}$$

since according to the first rule above, each unoccupied site contributes a factor X^{-1}. Similarly the SGF for occupied sites of the mth mode assuming the presence of interactions between adjacent bound ligands without site exclusion is

$$U_m(X) = \omega_m X^{-1} + k_m \omega_m^2 X^{-2} + k_m^2 \omega_m^3 X^{-3} + \ldots$$

$$= \sum_{i=1}^{\infty} k_m^{i-1} (K_m c_f)^i X^{-i} \tag{6}$$

$$= K_m c_f/(X - k_m K_m c_f) . \tag{7}$$

The condition that the generating function $\Gamma(X)$ just diverge determines the critical value X_1 by solving the secular equation (eq. (13) of Lifson [1]):

$$f(X) = \begin{vmatrix} -1 & U_0(X) & U_0(X) & \dots & U_0(X) \\ U_1(X) & -1 & k_{21}U_1(X) & \dots & k_{m1}U_1(X) \\ U_2(X) & k_{12}U_2(X) & -1 & & \\ \vdots & \vdots & & \ddots & \\ U_m(X) & k_{1m}U_m(X) & k_{2m}U_m(X) & \dots & -1 \end{vmatrix} = 0 , \quad (8)$$

where the $k_{mm'}$'s are given by eq. (4), for the maximum root $x = X_1 > 0$. This value X_1 then has the property

$$Z_N = X_1^N$$

according to Lifson's theorem. As Schellman has clearly recognized [3], it is only necessary to treat eq. (8) as an implicit function, so that the average fraction of ligand bound per site according to each binding mode, m, appears as

$$r_m = \left(\frac{\partial \ln f(X)}{\partial \ln K_m}\right)\left(\frac{\partial \ln f(X)}{\partial \ln X^{-1}}\right)_{x=X_1} , \quad (9)$$

where K_m is the intrinsic binding constant for the mth mode and the factor X^{-1} is common to all lattice sites. Typically, both the numerator and denominator of eq. (9) diverge since chains can achieve unlimited size, while the ratio r_m remains finite.

Since a number of special cases included within the scheme we have presented have already been dealt with, we present here a tabulation of scheme of some examples of interest.

4.1. Single binding mode with interaction between adjacent sites

Here we have only the SGF's $U_0(X)$ and $U_1(X)$, so that eq. (8) reduces to:

$$f(X) = U_0(X)\,U_1(X) - 1 = 0$$

giving the quadratic secular equation with eqs. (5) and (7):

$$f(X) = X^2 - (1 - kKc_f)X + (k-1)kc_f = 0 , \quad (10)$$

and from eq. (9) we obtain the standard result:

$$r = \tfrac{1}{2}[1 + (kKc_f - 1)/\sqrt{(1 - kKc_f)^2 + 4Kc_f}] , \quad (11)$$

corresponding to the infinite one-dimensional Ising model with nearest neighbor interactions. Note that for the special case $k = 1$ eqs. (10) and (11) simplify to the ideal case

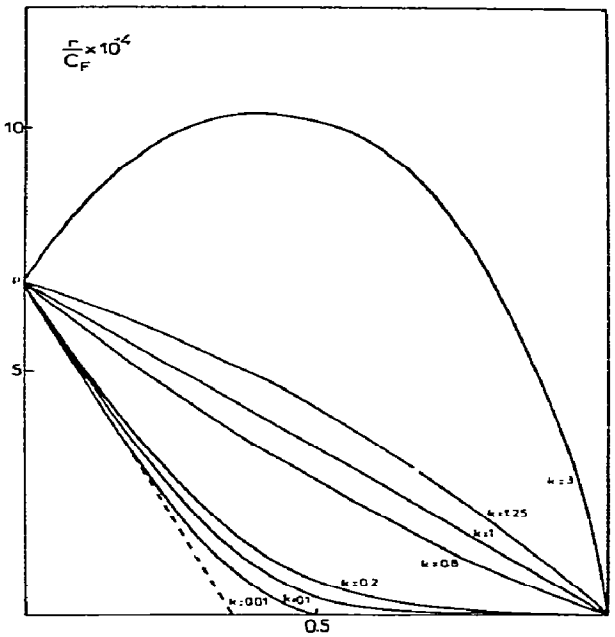


Fig. 2. Isotherms calculated for a binding model with a single binding mode and interaction between adjacent sites. The intrinsic association constant for the site: $K = 6.8 \times 10^{-4}$ M^{-1}; k is the interaction coefficient between two ligands bound on adjacent sites.

$$r = Kc_f/(1 + Kc_f) . \quad (12)$$

In fig. 2 we have illustrated the binding isotherms for this case, using the value $K = 6.8 \times 10^4$ corresponding to the strong binding constant for the profiavine–DNA of the *Micrococcus lysodeikticus* case.

The results are graphed according to the classical Scatchard plot of r/c_f versus r, which exhibits the non-interacting isotherm of eq. (12) as a straight line with intercepts $(0, K)$ on the ordinate and $(1,0)$ on the abscissa, the binding constant and fraction of available binding sites at saturation respectively. For $k \neq 1$, note that both these points are common to all curves, the family being humped for $k > 1$, and bowed for $k < 1$. While these results again are well known [3] it is worth observing that for strongly anticooperative interactions the isotherms become nearly linear for small r, and this linear region has slope of $3K$ and an apparent intercept on the abscissa of $n/3$. Thus the strongly anticooperative model is indistinguishable for small r from a model of independent sites with the above parameters.

4.2. Site exclusion models

In these cases, binding of a ligand at a site according to one mode inhibits further binding in that mode for l adjacent sites. Such a situation has been postulated by Bauer and Vinograd [14] in the case of binding of the dye ethidium bromide to DNA, for example where each intercalated molecule appears to exclude one adjacent site. A similar site exclusion appears to operate in the case of actinomycin D binding to DNA [15]. Using rule 5, the SGF for occupied sites becomes

$$U_1(X) = \sum_{i=1}^{\infty} k^{i-1}(Kc_f)^i X^{-(l+1)i}$$

$$= Kc_f/(X^{l+1} - kKc_f), \qquad (13)$$

so that from eqs. (5) and (8) we have

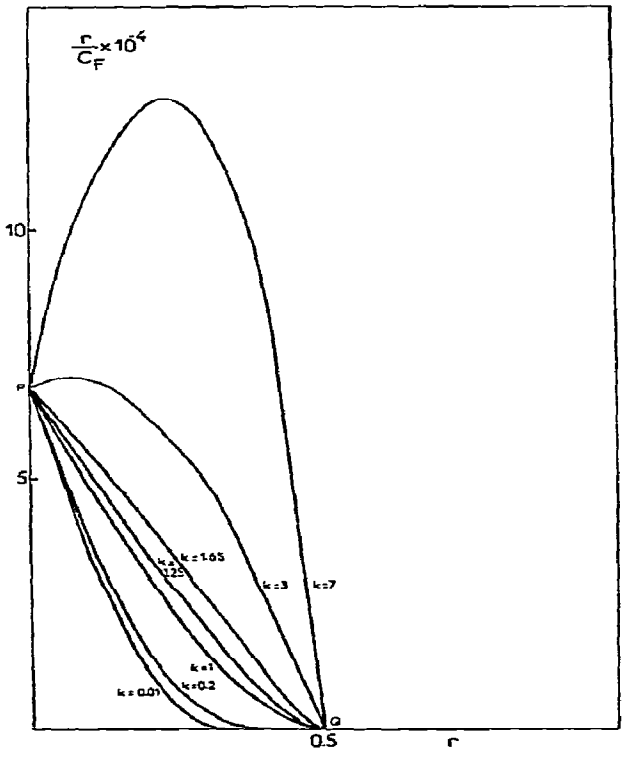


Fig. 3. Isotherms calculated for the site exclusion models. The intrinsic association constant for the site: $K = 6.8 \times 10^{-4}$ M^{-1}; k is the interaction coefficient between two ligands separated by an empty site.

$$f(X) = X^{l+2} - X^{l+1} - kKc_fX + (k-1)Kc_f = 0 \qquad (14)$$

and

$$r = \left.\frac{kKc_fX - (k-1)Kc_f}{(l+2)X^{l+2} - (l+1)X^{l+1} - kKc_fX}\right|_{X=X_1} . \qquad (15)$$

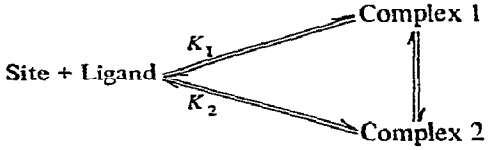
Eqs. (14) and (15) are sufficient to define the binding isotherm for the site exclusion model. The solution of eq. (14) for its maximum root $X = X_1$ requires a numerical procedure in general, so that the isotherms are most easily obtained by adjusting the parameters and comparing the results to experiment. On the other hand, McGhee and von Hippel [13] have presented a convenient alternative to eqs. (14) and (15) in discussing the binding of large ligands to DNA.

Our formulation is such that attachment of the ligand excludes $l + 1$ total lattice sites, one representing the ligand binding site as in the case of intercalation. Thus, our constant $l = n - 1$, where n is the size of the ligand according to McGhee and von Hippel [13]. It can be verified that the two approaches yield identical results; for Scatchard plots, the closed form of their eq. (15) is especially useful.

For the case of a single excluded site, $l = 1$, we have plotted the Scatchard isotherms in fig. 3 for $K = 6.8 \times 10^4$. Note that the isotherm for $k = 1.65$ is linear for almost all r, so that this case is indistinguishable from a model of independent binding sites with association constant $2K$ and $n = 0.5$. The isotherms for $k > 1$ are concave downwards, typical of a cooperative tive process, even though the binding to an adjacent lattice site is excluded. The intercepts on the ordinate and abscissa remain invariant.

4.3. Two competitive binding modes

We now consider a special case which may be relevant to the interaction between proflavine and DNA or synthetic polynucleotides. Suppose that at each site a ligand can form either of two complexes, so that there are two competing modes available. This model can be described by the following reaction scheme:



(16)

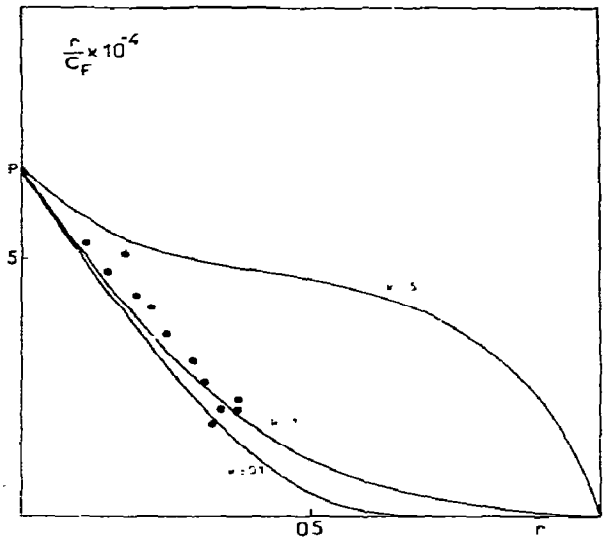


Fig. 4. Isotherms calculated for a model with two competitive binding modes. *Complex II:* Site exclusion model, with an intrinsic association constant $k_{II} = 46\,800\ M^{-1}$. *Complex I:* Model with interaction between adjacent sites (k is the interaction coefficient) and with an intrinsic association constant $K_I = 21\,200\ M^{-1}$. (•) Experimental points for the binding of proflavine on *Micrococcus lysodeikticus* DNA. K_I and K_{II} were estimated from kinetic measurements [12].

We assume that bound ligands of mode 2 exclude binding at the nearest neighbor sites ($l = 1$) while bound sites of mode 1 interact with a constant k, and with $k_{12} = 1$. Combination of eqs. (5) to (8) yields the secular equation:

$$X^3 - (1+\alpha)X^2 - [(K_1+K_2)c_f - \alpha]X + K_2c_f(\alpha - K_1c_f) = 0,$$

where α is defined as $\alpha = kK_1c_f$. We derive from this, using eq. (9)

$$r_1 = \frac{K_1}{X}\,\frac{kc_fX^2 + c_f(1-k)X + K_2c_f^2(1-k)}{3X^2 - 2(1+\alpha)X - [(K_1+K_2)c_f - \alpha]},$$

$$r_2 = \frac{K_2}{X}\,\frac{c_fX + K_2c_f^2 - c_f}{3X^2 - 2(1+\alpha)X - [(K_1+K_2)c_f - \alpha]}, \qquad (17)$$

for total binding obviously $r = r_1 + r_2$. Fig. 4 shows the isotherms, graphed according to Scatchard, for three values of k, using values for K_1 and K_2 estimated from relaxation studies [12] of proflavine–DNA. For

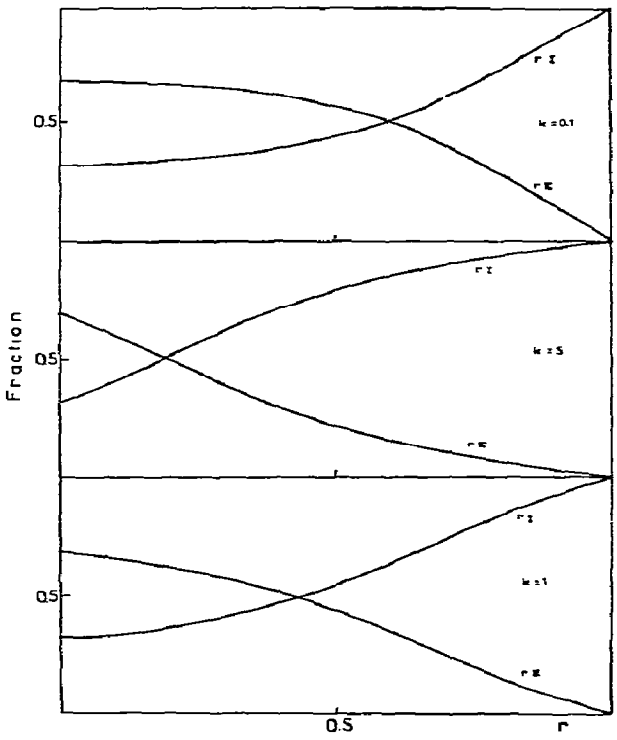


Fig. 5. Model with two competitive binding modes. Variation of the fraction r_I/r and r_{II}/r as a function of r.

sufficiently small r, neither competition nor interaction between bound ligand states has yet occurred so that

$$r/c_f = r_1/c_f + r_2/c_f \approx K_1(1-r_1) + K_2(1-r_2)$$

whence

$$\lim_{r\to 0} r/c_f = K_1 + K_2$$

as can be seen in the figure. Saturation is reached for $r = 1$ in this case. The fractional occupation of mode 1 and 2 is illustrated in fig. 5, where it can be seen that at saturation complex 1 is the dominant species, due to the anticooperativity of the second complex.

4.4. Sequential binding or "dimer" model

Consider the alternative extreme case, in which binding at a site of one ligand is required for attachment of a second molecule, according to the scheme:

$$\text{site} + \text{ligand} \overset{K_M}{\rightleftarrows} C_M + \text{ligand} \overset{K_D}{\rightleftarrows} C_D ,$$

where the subscript M refers to binding of monomer, and the subscript, D to dimer. Such a process has been suggested for the binding of ethidium bromide to DNA, by Waring [16]. Each step is characterized by an association constant, K_M or K_D and we put

$$K_D = kK_M , \tag{18}$$

where k is the interaction constant reflecting the influence of the bound monomeric ligands on attachment of the second. Let us also suppose that in addition the monomer binding process requires site exclusion of adjacent lattice sites. According to fig. 1E, it is evident that we need to introduce one new sequence generating function, corresponding to dimer binding. This is given according to the above rules by

$$U_0(X) = \sum_{i=1}^{\infty} k^i (K_M c_f X^{-1})^{2i}$$

$$= k(K_M c_f)^2 / [X^2 - k(K_M c_f)^2] .$$

Again there are three possible states for a lattice site, so that eq. (8) reduces to

$$f(X) = \begin{vmatrix} -1 & U_0(X) & U_0(X) \\ U_M(X) & -1 & U_M(X) \\ U_D(X) & U_D(X) & -1 \end{vmatrix} = 0 ,$$

giving the quadratic:

$$f(X) = X^2 - X - (kK_M + 1) K_M c_f = 0 ,$$

with

$$X_1 = \tfrac{1}{2}(1 + \sqrt{1 + 4(kK_M + 1)K_M c_f}) .$$

Since X_1 is explicitly given, we find, for the overall binding,

$$r = \partial \ln X_1 / \partial \ln c_f$$

$$= (c_f / X_1)(2kK_M c_f + 1) K_M / (2X_1 - 1) .$$

As can be seen in fig. 6, for $k > 1$, favoring dimer formation, the isotherms are humped and for $k < 1$, they are bowed. The case $k = 1$ again produces a linear isotherm.

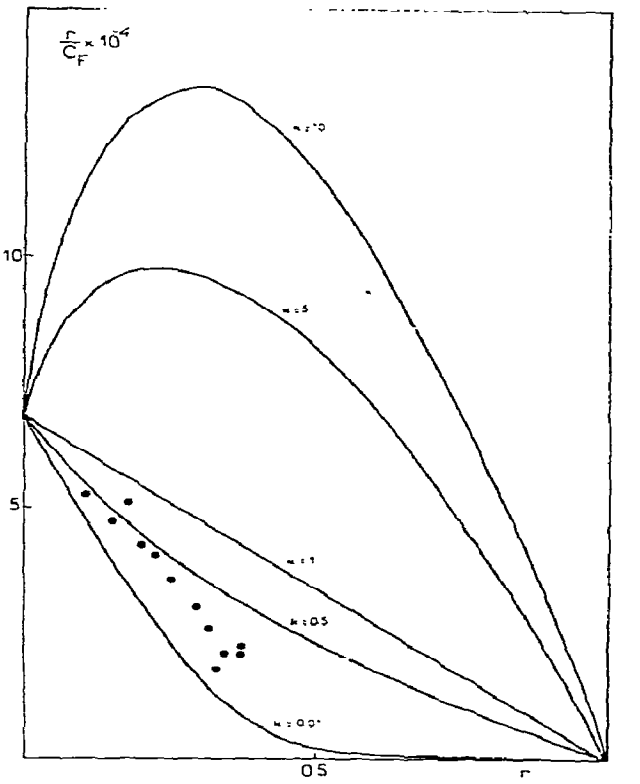


Fig. 6. "Dimer" model-calculated isotherms. The intrinsic association constant for binding of the monomer $K_M = 68\,000\ M^{-1}$. The intrinsic association constant for the binding of the dimer: $K_D = kK_M$. (•) Experimental points for the binding of proflavine on *Micrococcus lysodeikticus* DNA.

5. Results and discussion

We have reported the experimental results on fig. 4 and fig. 6. Let us compare first the experimental data with the calculated isotherms of the dimer model. As seen in fig. 6 there is no difficulty reconciling the experimental data with a dimer model if the parameter k in eq. (18) has a value of $0.01 < k < 0.5$. However our viscosity studies require that the binding of the second dye molecule be positively cooperative, with a value of $k \approx 70$ [17]. This process would evidently lead to a humped isotherm (fig. 6) and cannot account for the binding data without gross modification of the overall binding constants.

The model with two competitive binding modes

leads to calculated isotherms in excellent agreement with experimental data if $k \approx 1$. This model can equally well account for the quantities of intermediate (C_1) to internal binding (C_2) estimated from viscosity data [6] and moreover, with the values of the equilibrium constants given in fig. 4, satisfies the kinetic data [7]. From this there is a suggestion at least that the C_1 complex behaves in a fashion intermediate between the anticooperative intercalation complex (C_2) and the cooperative electrostatic weak binding process which takes place at low ionic strength and at low P/D.

6. Conclusion

We have presented an explicit calculation of the Scatchard isotherms corresponding to a number of cases of association between a ligand and a linear polymer using the general procedure of Lifson [1] and Lifson and Bradley [2] which requires that the mean polymer chain length be great enough to exclude end effects. In this limit, we can state that independently of the details of the model, positive cooperative interactions ($k > 1$) among bound ligand states produce characteristic humped isotherms, while anticooperativity ($k < 1$) leads to bowed isotherms. Beyond this we have found repeated instances in which linear Scatchard isotherms can result from complex underlying processes, so that the hazard of inferring a particular mechanism by simple inspection of the isotherms is obvious. This thermodynamical approach can be very useful when used in the frame of a general study including other techniques.

Acknowledgement

This research was supported by National Science Foundation grant GB 29210. Dr. Ramstein was the recipient of a grant from the Phillips Fund of the University of Pennsylvania.

References

[1] S. Lifson, J. Chem. Phys. 49 (1964) 3705–3734. Reprinted in: D. Polland and Scheraga, Theory of the Helix–Coil Transition (Academic Press, New York, 1970).
[2] D. Bradley and S. Lifson, in: Molecular Associations in Biology, ed. B. Pullman (Academic Press, New York, 1968) p. 261–270.
[3] J.A. Schellman, Israel J. Chem. 12 (1974) 219–238.
[4] S.A. Latt and H.A. Sober, Biochemistry 6 (1967) 3293–3306.
[5] H.J. Li and D. Crothers, J. Mol. Biol. 39 (1969) 461–477.
[6] J. Ramstein, M. Dourlent and M. Leng, Biochem. Biophys. Res. Commun. 47 (1972) 874–882.
[7] J. Ramstein and M. Leng, Biophys. Chem. 3 (1975) 234–240.
[8] J. Ramstein, C. Hélène and M. Leng, Eur. J. Biochem. 21 (1971) 125–136.
[9] G. Weill, Biopolymers 3 (1965) 567–572.
[10] A. Blake and A.R. Peacocke, Biopolymers 6 (1968) 1225–1253.
[11] V.N. Damle, Biopolymers 9 (1970) 353–372.
[12] P.H. Von Hippel and J.D. McGhee, Ann. Rev. Biochem. 41 (1972) 231–300.
[13] J.D. McGhee and P.H. Von Hippel, J. Mol. Biol. 86 (1974) 469–489.
[14] W. Bauer and J. Vinograd, J. Mole. Biol. 47 (1970) 419–435.
[15] W. Müller and D.M. Crothers, J. Mol. Biol. 35 (1968) 251–290.
[16] M.J. Waring, J. Mol. Biol. 54 (1970) 347–279.
[17] J. Ramstein, Thesis, University of Orléans (1974).