Biophysical Chemistry, 45 (1993) 281–303
Elsevier Science Publishers B.V., Amsterdam

BIOCHE 01720

A comparative study of Scatchard-type and linear lattice models for the analysis of EPR competition experiments with spin-labeled nucleic acids and single-strand binding proteins

R.S. Keyes and A.M. Bobst

Department of Chemistry, University of Cincinnati, Cincinnati, OH 45221 (USA)

(Received 1 June 1992; accepted in revised form 11 September 1992)

Abstract

An EPR competition formalism is developed which provides relative affinities of proteins for nucleic acids. Two models for analyzing protein–nucleic acid interactions, one assuming independent binding sites (Model 1) and the other considering site overlap (Model 2), are examined with respect to their validity and limitations. The models are employed to derive affinity ratio relationships which are used to calculate the relative affinities of gene 32, gene 5, and SSB proteins for various nucleic acids. It is determined that although Model 2 must be used when determining absolute binding constants, by taking the ratio of binding constants the site overlap becomes unimportant under conditions of moderate to high cooperativity and relatively small site size. This allows Model 1 to considerably simplify binding analyses. Both models are applied to the single-strand binding proteins of bacteriophage T4 gene 32, bacteriophage fd gene 5, and the *Escherichia coli ssb* gene, and the results are compared.

Keywords: Competition formalism; Nucleic acid binding affinities; EPR measurements; Nucleic acid binding proteins; Linear lattice models

1. Introduction

Many cellular mechanisms involving nucleic acids, as well as alteration of these mechanisms through drug binding, proceed by way of nonspecific ligand–nucleic acid interactions. A variety of techniques have been used to study interactions of this type [1,2]. It has been found that although a ligand may bind nonspecifically, it may still exhibit base preference. For instance, gene 32 protein of bacteriophage T4 regulates it own synthesis by maintaining a difference in affinity for DNA and RNA. Gene 32 protein preferentially binds all available single-stranded DNA and then binds to its own mRNA, thus repressing its production [3]. EPR competition experiments with spin-labeled nucleic acids have been used to determine the relative affinities of various nucleic acids for the single-strand binding proteins of bacteriophage T4 gene 32, bacteriophage fd gene

Correspondence to: A.M. Bobst, Department of Chemistry, University of Cincinnati, Cincinnati, OH 45221 (USA).

5, and the *Escherichia coli ssb* gene [4–7]. These proteins perform various functions involved with nucleic acid replication, recombination, and repair by binding tightly to nucleic acids in a highly cooperative manner. In general, the EPR competition approach is applicable to other types of ligands as well. For example, oligonucleotides exhibit the same type of multiresidue binding that proteins do. The spin-labeled systems are analyzed empirically by taking advantage of their EPR signatures in the form of digitized spectral arrays. Linear combinations of EPR signatures of a spin-labeled nucleic acid with and without protein indicate the degree to which the labeled lattice is saturated.

Characterization of binding thermodynamics on a microscopic scale is generally accomplished by applying a model to the analysis of experimental binding data. Scatchard's method [8] has been useful for many systems consisting of independent binding sites which do not overlap. However, the potential binding sites overlap for large ligands which bind nonspecifically. The resulting multiple-contact binding requires more elaborate methods of analysis. A number of linear lattice models have been developed using various theoretical methods [9–18]. The conditional probability approach of McGhee and von Hippel [11] has been particularly popular due to its mathematical simplicity and to the intuitive insights it provides.

In contrast to methods using linear lattice models, the EPR competition approach described here is generally used in conjunction with a Scatchard-type model and provides the relative affinities of nucleic acids for a protein [5]. Because this method involves taking the ratio of apparent binding constants rather than attempting to obtain absolute binding constants, it will be shown that the assumption of independent binding sites can be retained under certain conditions. In particular, we show that the Scatchard-type method is valid for binding systems displaying moderate to high cooperativity and relatively small binding site sizes. The main strength of this approach is that the parameters needed for the calculations are closely tied to experimental results. Two models will be developed in order to contrast a Scatchard-type approach (Model 1) for determining the number of free binding sites on a lattice with the linear lattice model of McGhee and von Hippel (Model 2). It is determined that although Model 2 is more versatile than Model 1

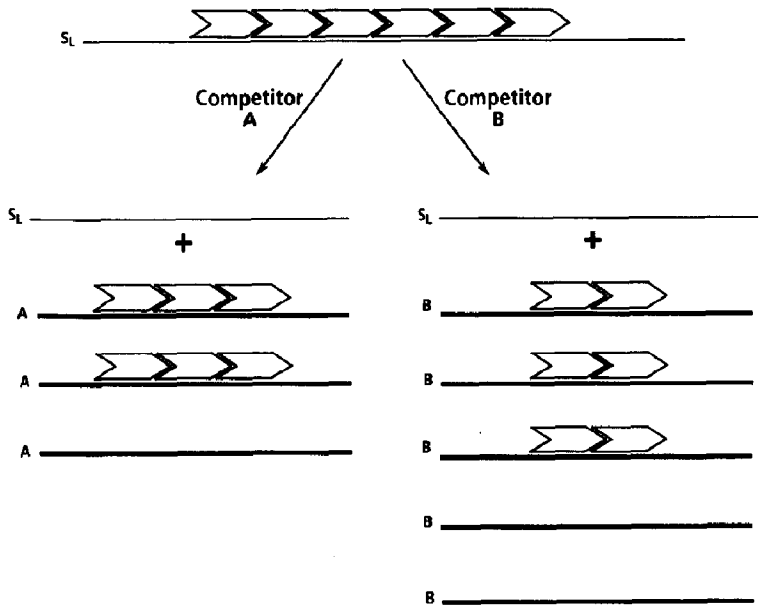

Fig. 1. Schematic representation of a pair of competition experiments where competitor lattices A and B strip the protein off of a spin-labeled lattice (S_L). The quantity of competitor needed is inversely related to its affinity for the protein.

in its ability to handle any degree of cooperativity, Model 1 is more versatile when either secondary structure or different classes of binding sites (i.e. sites with different binding constants) are present. Both situations can readily occur when performing binding experiments with single-strand binding proteins. After a general comparison, data from competition experiments with gene 32 [5], gene 5 [6], and SSB [7] proteins are used to compare results from both models. Whereas both models provide comparable analyses of gene 32 protein, there is a discrepancy when they are applied to gene 5 protein. This may be due to additional cooperative interactions resulting from helical formation of the complex. SSB protein has some unique characteristics that lead to limitations for both models.

2. EPR competition approach with a spin-labeled nucleic acid

The EPR competition approach consists essentially of two steps (Fig. 1). First, spin-labeled nucleic acid, S_L, is titrated with protein, P, to a specific initial fraction of saturation, F_i. Second, unlabeled competitor nucleic acid, A or B, is titrated into the sample. The sites on the unlabeled lattice compete with the sites on the labeled lattice for P. The result is that as more competitor lattice is added, P is pulled off of S_L and F decreases until a final fraction of saturation, F_f, is achieved. Since different nucleic acids vary in their affinities for a protein, the amount of each competitor required to reach a given F_f will reflect this difference. In other words, the quantity needed of a low affinity competitor will be greater than that of a high affinity competitor.

In order to use EPR to monitor protein–nucleic acid interactions, a correlation is needed which relates the EPR line shape to the fraction of protein bound to the labeled lattice. Ideally, the spectrum representing partial saturation will be a linear combination of the spectra representing free and fully complexed nucleic acid. On the molecular level, the question is whether or not the bound and free states of sites on the labeled lattice will be seen by the nitroxide. If the exchange rate of the protein on and off the lattice is slow with respect to nitroxide dynamics, then the two states will be resolved as distinct resonance peaks [19,20]. If not, then the two states will only be seen as an average resonance peak. Since nitroxide dynamics are on the order of nanoseconds and protein off rates are several orders of magnitude slower (e.g. gene 32 [21], gene 5 [22], and SSB [23–25]), it is expected that EPR will detect both states.

A straightforward approach to characterizing an EPR spectrum is to measure the ratios of the peak heights of the spectral lines. Two commonly used height ratios are h_+/h_0 (low field line/center field line) and h_-/h_0 (high field line/center field line). These ratios were plotted against the ratio of protein to nucleotide to observe the resulting relationships [5], and the curves obtained from both height ratios were nonlinear. Since the shape of the curves did not necessarily support a linear combination approach, another method of characterizing the spectra was attempted. Rather than using height ratios, the entire line shape was used. The spectra were digitized into arrays, and a fraction, corresponding to F, of the complexed array was added to the complementary fraction, $1-F$, of the free array to achieve a resultant array [26]. When F was plotted against the fraction of saturation determined by the amount of protein added, a linear relationship was observed at all but the highest fractions of saturation (Fig. 2).

The mathematical definition of F as a fraction of the complexed spectrum physically represents the fraction of spin-labeled residues along the nucleic acid that have been complexed with protein. This in turn represents the fraction of total nucleotide residues complexed with protein. This definition of F is independent of the configuration of the protein molecules on the lattice (Fig. 3). Therefore, the experimentally determined value of F will be the same regardless of the model used to define the protein binding sites. It is interesting that the EPR spectrum varies linearly with the amount of bound spin-labeled nucleic acid. On a microscopic level, one would expect a different signal when a protein is sitting on a spin-labeled residue than when the protein

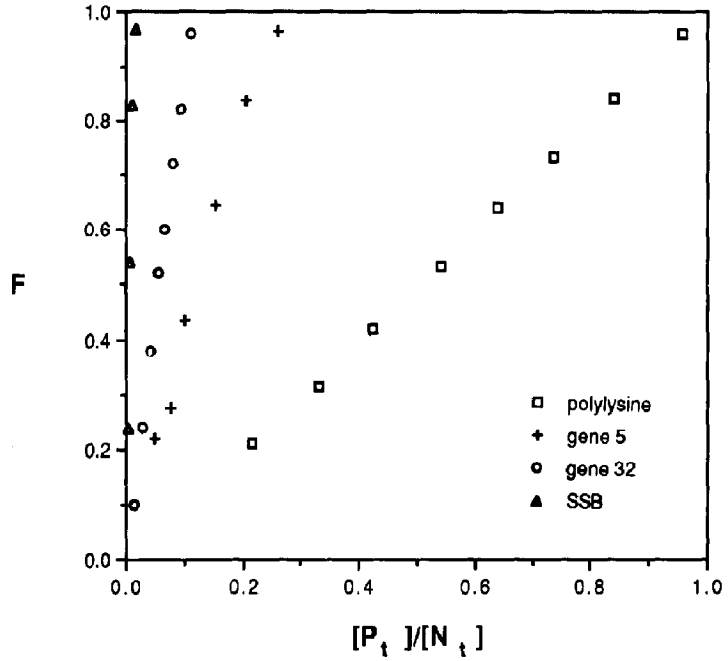

Fig. 2. Data from competition experiments with polylysine [5], gene 5 [6], gene 32 [5], and SSB [7] proteins illustrating the linear relationship between F and $[P_t]/[N_t]$.

is placed nearby. Thus, it would seem that the signal should depend upon protein location on the lattices. However, from a macroscopic standpoint, the EPR is detecting an average signal from an ensemble of lattices in the sample. Each lattice will have a somewhat different distribution of protein with respect to the label positions which are likewise randomly distributed. In addition, any protein translocation (e.g. sliding or hopping) that might take place will average out the spin-label signals.

Fraction F can be equated to the fraction of sites occupied by protein assuming a Scatchard-type model of the binding sites:

$$F = PS_L/S_{L_t} \tag{1}$$

where PS_L represents the number of binding sites occupied by protein, and S_{L_t} is the total number of binding sites. However, for the purpose of performing competition experiments, F is only used to monitor the amount of protein that has been removed from the labeled lattice by any unlabeled competitor. Thus, this aspect of the experiment does not require taking into account the total number of potential binding sites as defined by overlap models.

3. Derivation of the relative affinity ratio

In order to analyze nucleic acid affinity differences quantitatively, it is necessary to derive the

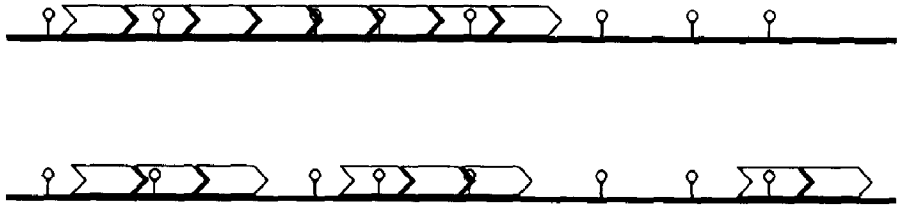

Fig. 3. Two possible configurations of protein on a randomly spin-labeled nucleic acid lattice having the same value of F.

equilibrium relationships between the components of the competition experiments. The equilibrium that occurs when a protein associates with a free binding site, such as:

$$P + S \rightleftharpoons PS \tag{2}$$

is described by an intrinsic binding constant, K_{int}. This reaction has a free energy change, $\Delta G^{\circ}_{int} = -RT \ln K_{int}$, associated with it. If cooperative effects occur that influence the equilibrium in the form of either interactions between protein molecules or conformational changes in the lattice, then the free energy of binding will include an additional term:

$$\Delta G^{\circ} = \Delta G^{0}_{int} + \Delta G^{\circ}_{coop} \tag{3}$$

Therefore, the equilibrium is characterized by an apparent binding constant which encompasses both the intrinsic binding constant and the effect of cooperativity:

$$K_{app} = K_{int} \exp(-\Delta G^{\circ}_{coop}/RT) \tag{4}$$

If cooperative effects are introduced in the form of a cooperativity parameter, ω [27], then:

$$K_{app} = K_{int}\omega = \frac{[PS]}{[P][S]} \tag{5}$$

When a competition experiment is performed, S_L is initially titrated with P to a specific F_i. The competitor, A or B, is then introduced which essentially strips P off of S_L until a particular F_f is reached. These equilibria are written for a case A as:

$$PS_{L_A} \rightleftharpoons P + S_{L_A} \tag{6}$$

$$P + S_A \rightleftharpoons PS_A \tag{7}$$

Each of these equations can be described by an intrinsic and an apparent binding constant. Summing reactions (6) and (7) gives:

$$PS_{L_A} + S_A \rightleftharpoons PS_A + S_{L_A} \tag{8}$$

An analogous equation can be written for a case B:

$$PS_{L_B} + S_B \rightleftharpoons PS_B + S_{L_B} \tag{9}$$

The equilibrium constants for reactions (8) and (9) are:

$$K_A = \frac{[PS_A][S_{L_A}]}{[PS_{L_A}][S_A]} = \frac{K_{app_A}}{K_{app_L}} = \frac{K_{int_A}\omega_A}{K_{int_L}\omega_L} \tag{10}$$

$$K_B = \frac{[PS_B][S_{L_B}]}{[PS_{L_B}][S_B]} = \frac{K_{app_B}}{K_{app_L}} = \frac{K_{int_B}\omega_B}{K_{int_L}\omega_L} \tag{11}$$

where the relationships to the appropriate apparent and intrinsic binding constants (eq. 5) have been included.

Because the affinities of nucleic acids for single-strand binding proteins are so high, the free protein concentration is negligible. This was verified experimentally by plotting F vs. $[P_t]/[N_t]$ where P_t is the total amount of protein in solution and N_t represents the total residues of nucleic acid. Figure 2 shows data obtained for polylysine [5], gene 32 [5], gene 5 [6], and SSB [7] proteins. Since F represents the ratio of bound nucleic acid residues to total nucleic acid residues and $[P_t]/[N_t]$ represents the ratio of bound protein to total nucleic acid residues, the plot should follow the linear relationship:

$$F = n\frac{[P_t]}{[N_t]} \tag{12}$$

where n is the number of residues occupied by a bound protein. Analysis of the data of Fig. 2 indicates that eq. (12) is obeyed. Thus, in solution, P distributes between the labeled sites and the competitor sites. The total protein for either case is:

$$P_t = PS_{A,B} + PS_{L_{A,B}} \tag{13}$$

The degree of association of P to each of the two lattices in case A is:

$$\theta_A = \frac{[PS_A]}{[P_t]}, \qquad 1 - \theta_A = \frac{[PS_{L_A}]}{[P_t]} \tag{14a,b}$$

Likewise, for case B:

$$\theta_B = \frac{[PS_B]}{[P_t]}, \qquad 1 - \theta_B = \frac{[PS_{L_B}]}{[P_t]} \tag{15a,b}$$

Rearranging eqs. (10) and (11) and substituting the degrees of association for each case results in:

$$\frac{[PS_A]}{[PS_{L_A}]} = K_A \frac{[S_A]}{[S_{L_A}]} = \frac{\theta_A}{1-\theta_A} \tag{16}$$

$$\frac{[PS_B]}{[PS_{L_B}]} = K_B \frac{[S_B]}{[S_{L_B}]} = \frac{\theta_B}{1-\theta_B} \tag{17}$$

Since P_t and F_f (i.e. $PS_{L_A} = PS_{L_B}$) are the same in both cases

$$1-\theta_A = 1-\theta_B \tag{18a}$$

$$\theta_A = \theta_B \tag{18b}$$

$$[S_{L_A}] = [S_{L_B}] \tag{19}$$

Therefore, assuming equal sample volumes for both cases, the ratio of the binding constants is obtained by dividing eq. (16) by eq. (17) and simplifying:

$$\frac{K_A}{K_B} = \frac{S_B}{S_A} \tag{20}$$

Equation (20) is a thermodynamic relationship that is model-independent. However, since the number of residues bound by a single protein is, in general, greater than one, the number of free binding sites available can become complicated by overlapping of sites. Therefore, two approaches will be discussed for describing the free binding sites as illustrated in Fig. 4.

3.1 Model 1

Consider a nucleic acid lattice consisting of N residues. If a collection of lattices in a sample is homogeneous with respect to size, then every lattice will contain approximately the same number of residues. A particular protein will have a binding site size of n residues. Therefore, a fully saturated lattice will contain N/n occupied sites. In other words, N/n represents the maximum number of proteins that can sit upon a particular lattice. If there are N_t total residues in the sample, then N_t/n represents the maximum number of bound proteins that can be present. If the binding sites are not considered to overlap, N_t/n also represents the total number of potential binding sites.

By using n as the number of residues occupied by a protein, it is generally assumed that all nucleic acid residues of a saturated lattice interact with protein. However, this will not always be the case. The number of residues, n, interacting

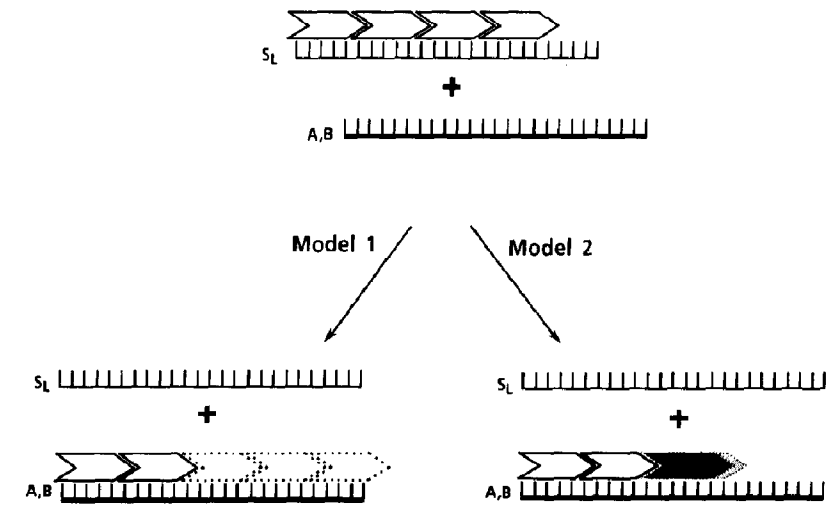

Fig. 4. Schematic illustration of the next three potential binding sites for contiguous binding of a protein with $n = 5$ as assumed by Model 1 and Model 2.

with a protein may not be the same as the number of residues, n', covered by a protein. More importantly, there may be secondary structure present which prevents binding along the entire lattice. Even in the presence of helix-destabilizing proteins, secondary structure may melt out very slowly. During the time of a competition experiment, which is normally one to two hours, these kinetic blocks will prevent complete binding. A way to circumvent these problems is to use an experimentally determined stoichiometry rather than a fixed site size n. The stoichiometry, s, represents the number of nucleotide residues per bound protein. The advantage of using s is that any unoccupied residues are divided among the bound proteins. Therefore, s will be larger than n by an amount proportional to the number of residues unavailable due to protein or nucleic acid structure. Since s is experimentally determined, it reduces the theoretical assumptions necessary. When s is substituted for n, the total number of binding sites becomes N_t/s.

When a protein binds to a homogeneous lattice, all sites should have the same intrinsic binding constant. If a protein is completely nonspecific when it binds, then the sites along a heterogeneous lattice should exhibit the same affinity also. However, in general, single-strand binding proteins do show some amount of base specificity as seen by affinity differences between lattices of varying composition [3–7,28–30]. Therefore, the intrinsic binding constants will not be equal. Since, according to eqs. (10) and (11), the affinity ratio is related to the apparent and intrinsic binding constants by:

$$\frac{K_A}{K_B} = \frac{K_{app_A}}{K_{app_B}} = \frac{K_{int_A}\omega_A}{K_{int_B}\omega_B} \tag{21}$$

analyses with Model 1 will indicate overall or average affinity ratios for heterogeneous lattices.

The number of free binding sites at any particular fraction of saturation equals the difference between the total number of sites, S_t, and the number of bound sites:

$$S = S_t - PS \tag{22}$$

Substituting eq. (22) into eq. (20) and expressing S_t and PS explicitly yields the affinity ratio relationship for nonoverlapping sites [5]:

$$\frac{K_A}{K_B} = \frac{\dfrac{N_{t_B}}{s_B} - P_t\left(1 - \dfrac{F_f}{F_i}\right)}{\dfrac{N_{t_A}}{s_A} - P_t\left(1 - \dfrac{F_f}{F_i}\right)} \tag{23}$$

where $(1 - F_f/F_i)$ represents the fraction of P stripped off of S_L. When $s = n$ and $n_A = n_B$ (conditions often met), eq. (23) can be expressed in terms of the binding densities, ν_A and ν_B (moles bound protein/moles nucleic acid residue):

$$\frac{K_A}{K_B} = \frac{\dfrac{1}{\nu_B} - n}{\dfrac{1}{\nu_A} - n} \tag{24}$$

While eq. (23) is generally used for analyzing experimental results, eq. (24) will be useful for comparing Model 1 with Model 2.

3.2 Model 2

In general, site overlap cannot be ignored for proteins having a site size greater than one. This problem has been examined in detail by McGhee and von Hippel [11] for an infinite one-dimensional lattice for both noncooperative and cooperative binding. For instance, the site for a particular protein that binds to a lattice with a site size of five may begin on residue number one. But, for the noncooperative case, it is just as likely to have begun on residues two, three, four, or five. Considering overlap, the number of potential binding sites on a naked lattice is $N - n + 1$. Therefore, the first protein to bind to a lattice may begin its binding site at any residue except the last $n - 1$ residues. This first protein will eliminate $2n - 1$ potential binding sites rather than just one as in Model 1. Any protein that binds to an isolated site (i.e. both ends of the site are at least $n - 1$ residues away from any other bound protein) will eliminate this same number of sites. Proteins that bind immediately adjacent to other bound pro-

teins (assuming that there are at least $n-1$ free residues at the other end of the protein) will remove n potential binding sites. Finally, if a protein binds in a gap of n residues between two bound proteins, only one site will be eliminated. Clearly, the number of binding sites eliminated by the binding of a single protein can range from $2n-1$ down to one depending upon the size of the gaps adjacent to the protein binding site. And the size and number of gaps will be a function of the cooperativity.

McGhee and von Hippel [11] derived formulas for the number of free binding sites on a lattice as a function of binding density, ν, and site size, n, for both the noncooperative and cooperative cases. These formulas may be substituted directly into eq. (20).

3.3 Model 2 – Noncooperative case

For the noncooperative case, where the cooperativity, ω, is equal to one, the average number of free potential binding sites per lattice, N_{Avg}, is:

$$N_{Avg} = N(1-n\nu)\left[\frac{1-n\nu}{1-(n-1)\nu}\right]^{n-1} \tag{25}$$

The number of lattices in a sample can be represented by N_t/N; therefore total

The total number of free binding sites

$$= N_t(1-n\nu)\left[\frac{1-n\nu}{1-(n-1)\nu}\right]^{n-1} \tag{26}$$

Substituting eq. (26) into eq. (20) for both cases A and B yields:

$$\frac{K_A}{K_B} = \frac{N_{t_B}(1-n\nu_B)\left[\frac{1-n\nu_B}{1-(n-1)\nu_B}\right]^{n-1}}{N_{t_A}(1-n\nu_A)\left[\frac{1-n\nu_A}{1-(n-1)\nu_A}\right]^{n-1}} \tag{27}$$

The total amount of protein is the same in either case. If the same F_f is achieved in both cases, then the same amount of protein will be bound to the competitor lattices in each case. If β represents the amount of protein bound per lattice, then:

$$\frac{\beta_A N_{t_A}}{N_A} = \frac{\beta_B N_{t_B}}{N_B} \tag{28}$$

Rearranging eq. (28) and recognizing that $\nu = \beta/N$ gives:

$$\frac{N_{t_A}}{N_{t_B}} = \frac{\beta_B/N_B}{\beta_A/N_A} = \frac{\nu_B}{\nu_A} \tag{29}$$

Substituting eq. (29) into eq. (27) results in the affinity ratio relationship for overlapping sites without cooperativity:

$$\frac{K_A}{K_B} = \frac{\nu_A(1-n\nu_B)\left[\frac{1-n\nu_B}{1-(n-1)\nu_B}\right]^{n-1}}{\nu_B(1-n\nu_A)\left[\frac{1-n\nu_A}{1-(n-1)\nu_A}\right]^{n-1}} \tag{30}$$

3.4 Model 2 — Cooperative case

When cooperative effects influence binding, this must be accounted for in the binding analysis. If proteins present on the lattice increase the affinity for binding of other proteins, this indicates positive cooperativity where $\omega > 1$. If bound proteins decrease further binding, negative cooperativity is operating and ω is between 0 and 1.

When cooperativity is considered, the average number of free sites on a lattice is:

$$N_{Avg} = N(1-n\nu)\left[\frac{(2\omega-1)(1-n\nu)+\nu-R}{2(\omega-1)(1-n\nu)}\right]^{n-1} \times \left[\frac{1-(n+1)\nu+R}{2(1-n\nu)}\right]^{2} \tag{31}$$

where $R = \{[1-(n+1)\nu]^2 + 4\omega\nu(1-n\nu)\}^{1/2}$

Reasoning analogously to the noncooperative

case provides the affinity ratio relationship for overlapping sites with cooperativity:

$$\frac{K_A}{K_B} = \left\{ \nu_A(1-n\nu_B) \times \left[\frac{(2\omega_B - 1)(1-n\nu_B) + \nu_B - R_B}{2(\omega_B - 1)(1-n\nu_B)} \right]^{n-1} \times \left[\frac{1-(n+1)\nu_B + R_B}{2(1-n\nu_B)} \right]^2 \right\} / \left\{ \nu_B(1-n\nu_A) \times \left[\frac{(2\omega_A - 1)(1-n\nu_A) + \nu_A - R_A}{2(\omega_A - 1)(1-n\nu_A)} \right]^{n-1} \times \left[\frac{1-(n+1)\nu_A + R_A}{2(1-n\nu_A)} \right]^2 \right\} \quad (32)$$

When protein structure (e.g. $n \neq n'$) or nucleic acid structure (e.g. hairpins) complicates binding, then the equations for Model 2 must be modified. If n' is different from n, then this adjusted value can be substituted for n in the equations with minimal error as long as end effects are negligible. However, if secondary structure of the lattice influences binding, then a different approach must be taken. The problem consists of determining how the binding of multiple proteins will be affected by sites being blocked by various conformations. The elimination of sites in Model 2 is based on the probabilities of gaps forming of various sizes which exclude protein binding. Therefore, the stoichiometry approach of Model 1 will not work for Model 2.

An additional concern is that Model 2 considers all binding sites to have the same K_{int}. As noted previously, this is not generally true for single-strand binding proteins. Heteropolymers contain different classes of binding sites with correspondingly different protein affinities. Consequently, the probability of binding to various sites will not be equal.

In order to resolve the problems of secondary structure and nonequivalent binding sites, an alternate method can be applied involving the calculation of macroscopic binding constants [31]. An outline of the approach will be given here.

Binding of multiple proteins to DNA can be expressed as:

$$NA_{ss} + mP \rightleftharpoons NA_{ss}P_m \quad (33)$$

for single-strand binding proteins. This equilibrium is described by an overall binding constant

$$K_{bind} = \frac{[NA_{ss}P_m]}{[NA_{ss}][P]^m} = \prod_{i=1}^{m} (K_{int}\omega)_i \quad (34)$$

K_{bind} represents the collective effect of all of the apparent binding constants of sites occupied by protein. If the individual apparent binding constants can be determined, then they can be combined to calculate K_{bind}.

When DNA is present in the double-stranded form, the net equilibrium is:

$$K_{net} = \frac{[NA_{ss}P_m]}{[NA_{ds}][P]^m} \quad (35)$$

The double strand to single strand conformational change is described by:

$$K_{conf} = [NA_{ss}]/[NA_{ds}] \quad (36)$$

Therefore, the net equilibrium constant is obtained from:

$$K_{net} = K_{bind}K_{conf} \quad (37)$$

4. Comparison of the two models

The main difference between the two models resides in the contrasting approaches taken to calculate the number of free binding sites. The applicability of the models for any given protein is mainly a function of the degree of cooperativity that occurs. Model 1 requires that there be moderate to high cooperativity acting in the system, while Model 2 can be used for systems displaying any degree of cooperativity.

There are a large number of proteins which bind to DNA cooperatively. The cooperative interactions result in binding sites adjacent to bound proteins having a higher apparent binding constant than isolated sites. This is the formal defini-

tion of cooperativity. However, difficulties arise when attempting to determine the magnitude of ω. Following from the formal definition and eq. (5), ω can be defined as the equilibrium constant describing two bound, isolated proteins becoming adjacent (Fig. 5). A direct approach for measuring ω is to determine the binding constants for protein binding to a polymer (where cooperativity can occur) and for binding to an oligomer (where only one protein can bind) and taking the ratio of the two. However, the binding mode is not necessarily the same for both of these cases. Thus, an ω determined by this method could be in error.

In the statistical thermodynamic approach of McGhee and von Hippel, the cooperativity was defined as [11]:

$$\omega = \frac{\text{probability of a bound protein being adjacent to another bound protein}}{\text{probability of a bound protein being in an isolated position}} \quad (38)$$

By fitting their equation (eq. 15 in ref. [11]) to experimental binding data, a value for ω can be determined. However, difficulties have arisen using this approach as well since the binding equation contains three parameters (K_{int}, n, and ω) and a unique solution is not necessarily achieved [32]. A survey of the literature indicates that there is no general agreement on the magnitude of ω. A range of values has been reported for gene 32 (200–10^4) [32–35], gene 5 (50–1000) [36,6,22], and SSB (50–10^5) [37,38] proteins. In addition, linear lattice models have failed to provide a correct analysis of some systems possibly as a result of the form of cooperativity assumed [36,39,40]. From a practical standpoint, Model 1 allows for the application of Occam's razor to reduce the theoretical assumptions needed as long as there is minimal gap formation and n is not too large.

When positive cooperativity is present, it is expected that the lowest free energy state will occur when the proteins bind contiguously along the DNA strands. An example from a real system will indicate the magnitude of ω beyond which gaps should become insignificant. Competition experiments were performed to determine the relative affinity of gene 32 protein for various nucleic acid lattices [5]. The molecular weight of the lattices was between 100,000 and 200,000 daltons (i.e. 250–500 residues). If a typical lattice (ca. 500 residues) is saturated with gene 32 protein ($n = 10$), it will require about 50 protein molecules. A cooperativity of 25 would indicate that there were 25 contiguously bound protein molecules for every gap. In other words, there would be one gap per lattice. A cooperativity greater than 25 would indicate that there was less than one gap per lattice. This is an important point, since most linear lattice models assume an infinite lattice. For an infinite lattice, it is possible that gap formation will be predicted as a result of the contribution of configurational en-

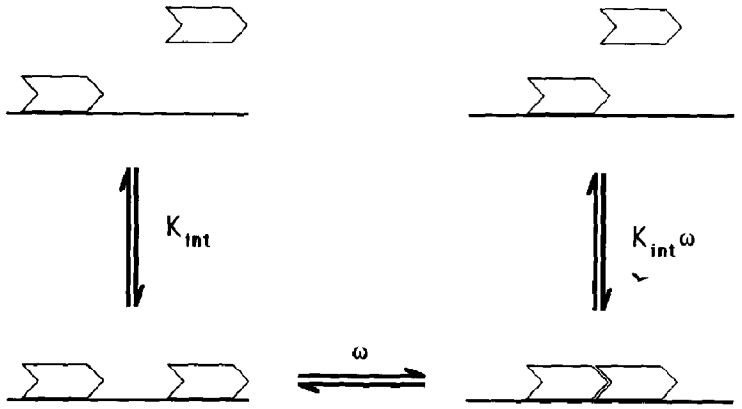

Fig. 5. Equilibria illustrating the definitions of the parameters K_{int}, ω, and $K_{int}\omega = K_{app}$.

tropy. For proteins binding to an infinite lattice, the process can be described by three steps [41,42]. The first is a noncooperative nucleation step where individual protein molecules bind to isolated binding sites along the lattice. Second, as the protein concentration increases, protein molecules will begin to cooperatively bind to singly contiguous sites adjacent to proteins which are already bound. This will generate clusters of protein distributed along the lattice. Finally, the clusters will redistribute on the lattice to form the lowest energy state. This scenario indicates that there is the possibility of gaps forming. However, for the finite lattices generally used in our competition experiments, it is unlikely that more than one or two clusters of any significant size would be present. Gaps that might be present after initial nucleation are not observed upon further addition of protein. Electron microscopy indicates that under appropriate salt conditions a variety of single-strand binding proteins form long clusters without gaps [43–46]. These clusters occur on some lattices while other lattices remain free. This is in contrast to noncooperative systems where the protein is equally distributed among all lattices. From a functional standpoint, gap formation between single-strand binding proteins would diminish their effectiveness in maintaining strand separation, preventing secondary structure formation, and guarding against nuclease digestion of the single-stranded nucleic acid.

Under conditions where gaps are insignificant, both Model 1 and Model 2 are applicable and give approximately the same results. In terms of

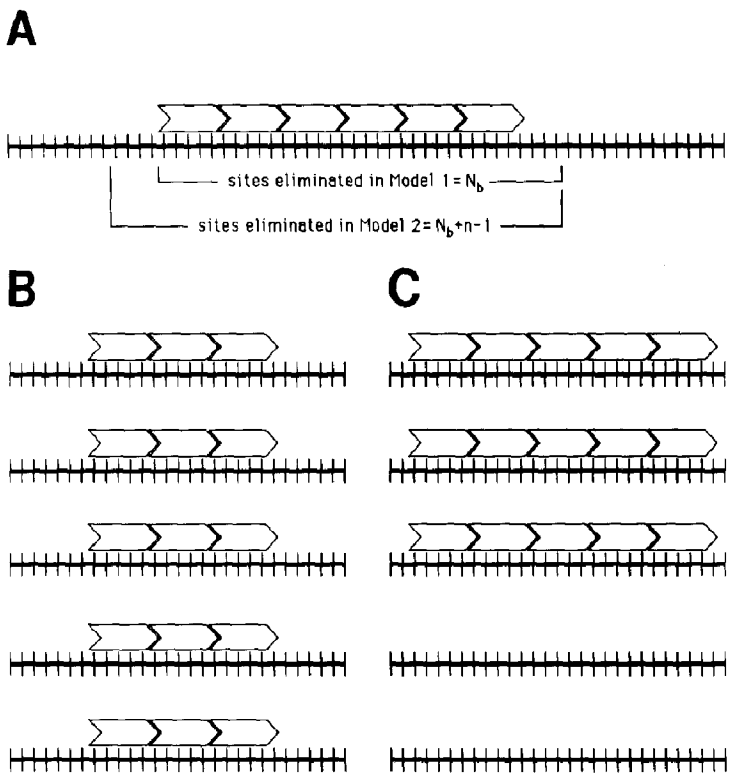

Fig. 6. (A) Comparison of the number of sites eliminated by Model 1 and Model 2. (B) Illustration of the statistical approach of Model 2 considering each lattice to have the same average degree of saturation. (C) Variable lattice saturation as experimentally observed by electron microscopy.

model parameters, this is demonstrated as follows (Fig. 6). The number of free binding sites is described by:

Free sites = Total sites
– Sites eliminated by protein binding (39)

The sample will consist of a mixture of strands some of which possess clusters of protein. Inserting the parameters for Model 1 gives

$$S = \frac{N_t - N_{b_t}}{n} \tag{40}$$

where N_{b_t} represents the total number of residues bound by protein in the sample. Equation (40) determines the free sites for the entire sample in terms of total free and total bound residues, not on an individual lattice basis. Substituting eq. (40) into eq. (20) yields:

$$\frac{K_A}{K_B} = \frac{(N_t - N_{b_t})_B}{(N_t - N_{b_t})_A} \tag{41}$$

It should be noted that by taking the ratio of free binding sites for Model 1, the site size n cancels out. Thus, eq. (41) can be considered to represent sites for Model 1 that overlap in a manner similar but not equivalent to Model 2 (see below).

According to Model 2, the number of sites eliminated by a cluster flanked by free residues will be $N_b + n - 1$, where N_b is the number of residues of an average lattice bound by protein. One way to obtain this formula is by first considering the sites eliminated on a lattice N_b residues long (i.e. $N_b - n + 1 = N_b - (n-1)$), and then adding the sites eliminated for two ends when this lattice is placed in a longer lattice:

$$N_b - (n-1) + 2(n-1) = N_b + n - 1 \tag{42}$$

Therefore, the analogous equation for free binding sites according to Model 2 is:

$$S = [(N - n + 1) - (N_b + n - 1)](N_t/N) \tag{43}$$

where N_t/N represents the number of lattices in the sample. Substituting eq. (43) into eq. (20) results in:

$$\frac{K_A}{K_B} = \frac{[(N - n + 1) - (N_b + n - 1)]_B (N_t/N)_B}{[(N - n + 1) - (N_b + n - 1)]_A (N_t/N)_A} \tag{44}$$

Rearranging eq. (44) gives:

$$\frac{K_A}{K_B} = \frac{[(N - N_b) - 2(n-1)]_B (N_t/N)_B}{[(N - N_b) - 2(n-1)]_A (N_t/N)_A} \tag{45}$$

The main difference between eqs. (41) and (45) is that there are sites eliminated at one end of a cluster that Model 1 does not take into account (Fig. 6A). Therefore, Model 1 overestimates the number of free binding sites. However, Model 2 also contains an approximation that must be considered. Since Model 2 performs a statistical treatment of an average lattice, it considers all bound protein to be equally distributed among all the lattices (Fig. 6B) rather than some lattices being complexed with protein and some lattices being free (Fig. 6C) as observed in binding experiments with gene 32, gene 5, and SSB proteins [43–46]. For a collection of lattices consisting of equivalent sites, there is an equal probability of a protein occupying any site. When statistical thermodynamics is applied to such a system, the ensemble will consist of all lattices since every lattice is expected to contribute to the ensemble average. If interaction occurs between bound protein or if the sites become non-equivalent, then this will influence the site probabilities, and the protein distribution will no longer be random. Under these circumstances, the correct statistical thermodynamic treatment is to consider the free lattices as part of the surroundings and the ensemble to only include lattices occupied by protein. Violating this principle results in an overestimation of the number of sites eliminated since the actual number of clusters is lower due to fewer lattices being complexed. Therefore, Model 2 will underestimate the number of free binding sites.

If the cluster end effects are negligible, then both models will give similar results and eqs. (41) and (45) will become equivalent. This condition

will hold for Model 1 when the binding density of protein on the competitor lattices is not too high. Since Model 1 neglects sites at one end of a cluster (Fig. 6A), this means that $N_t - N_{b_t}$ (i.e. the total free binding sites according to eq. (41)) must be large with respect to $(n-1)$ times the number of clusters (i.e. the total number of sites at the end of clusters overlooked by Model 1). This scenario is reasonable in light of the low binding densities generally used in competition experiments. For Model 2, the effect of cooperativity or of non-equivalent binding sites (both resulting in clustering) must be taken into account and the infinite lattice assumption must be valid. When the difference between the actual number of clusters and the number of clusters calculated by Model 2 is minimized (e.g. at high binding density) or when n is small enough, then cluster end effects will become insignificant for Model 2. In summary, the application of Model 1 and Model 2 must be coupled with a knowledge of their respective assumptions. Both models should give reasonable results when cluster end effects are minimal.

Ribonuclease, which was the first melting protein discovered [47], is an example of a protein

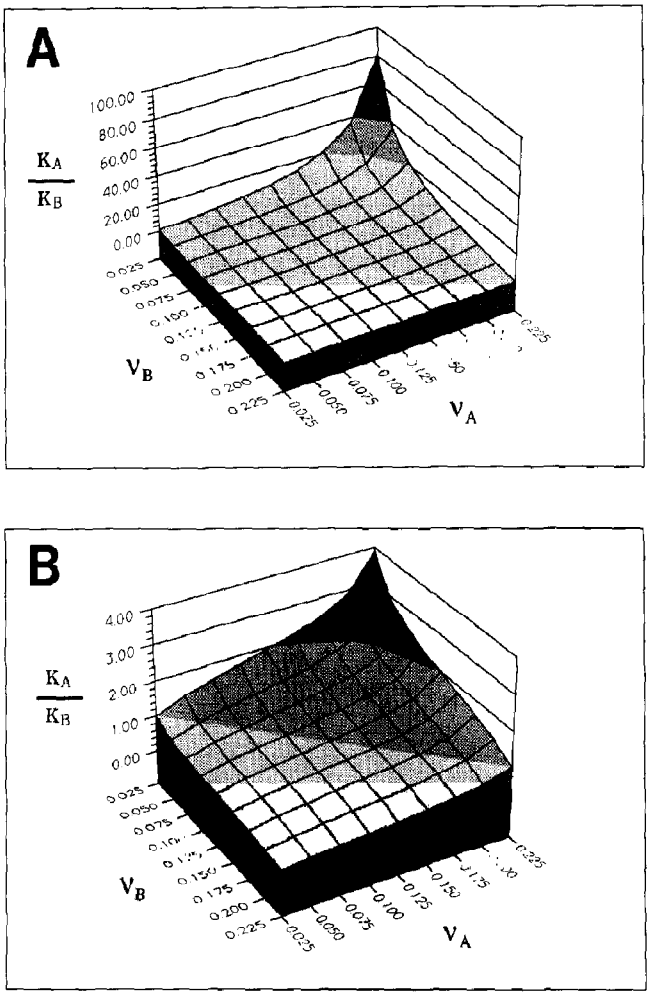

Fig. 7. 3-D graphs showing the behavior of the affinity ratio as a function of the binding densities of two competitors for (A) Model 1 and (B) Model 2. The data shown are for gene 5 protein binding to fd DNA (competitor A) and poly A (competitor B) (see Table 1).

that binds to DNA noncooperatively. Proteins of this type do not seem to bind in such a manner as to affect the binding of other proteins. Therefore, gaps of various sizes will occur between bound proteins which do not exhibit cooperative interactions. If a particular gap is smaller than the binding site size of a protein, then the protein molecule will not fit into the gap. Model 1 includes such gaps as part of the total number of residues available for binding. This overestimation of free binding sites for both cases A and B will result in errors in the calculated affinity ratios. Model 2 does not include undersized gaps as part of the number of free binding sites. In addition, when positive cooperativity is not present, protein will be distributed on all nucleic acid lattices to the same degree assuming that all binding sites are equivalent. Consequently, the statistical approach of calculating the free sites on an average lattice is valid. Thus, eq. (30) should be used to determine affinity ratios for this type of binding.

5. Results and discussion

In order to compare results obtained with the two models, data from previously published com-

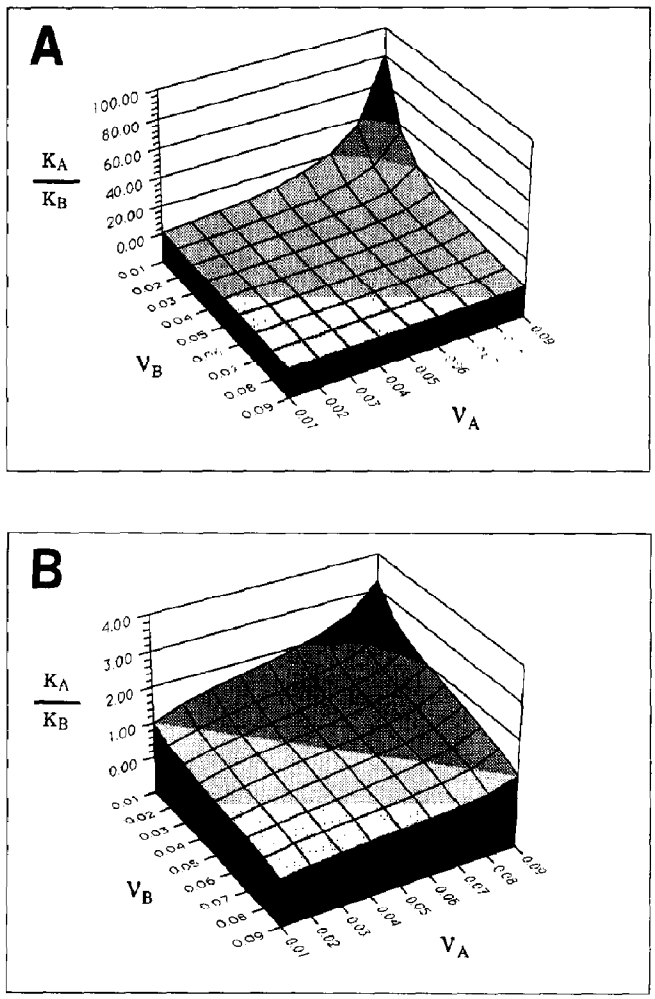

Fig. 8. Competition data analyzed according to (A) Model 1 and (B) Model 2 for gene 32 protein binding to poly dT (competitor A) and poly dI (competitor B) (see Table 1).

Table 1

Summary of analyses of experimental data based on Model 1, Model 2, and "Limited" ω Model

Competition system (case A: case B)	$N_{t_A}:N_{t_B}$ (nmol)	P_t (nmol)	$\nu_A:\nu_B$	Model 1 K_A/K_B	Model 2 ("Unlimited" ω Model)		"Limited" ω Model	
					$\omega_A=\omega_B$	K_A/K_B	$\omega_A=\omega_B$	K_A/K_B
Gene 5 [a] *(n = 4)*								
poly dT:fd DNA	9:3000	1.5	$0.17:5.0\times10^{-4}$	998	60 [d]	41	N/A	N/A
fd DNA:poly A	9:85	1.3	0.14:0.015	21	60	2.7	N/A	N/A
Gene 32 [b] *(n = 10)*								
poly dT:poly dI	130:440	2.4	0.018:0.0055	3.9	200 [e]	1.6	N/A	N/A
poly dI:poly I	440:1270	2.4	0.0055:0.0019	3.0	200	1.8	N/A	N/A
SSB [c] *(n = 65)*								
poly dT:fd DNA	26:88	0.14	0.0054:0.0016	4.7	50 [f]	4.3	400 [g]	4.3
fd DNA:poly A	12:225	0.13	$0.011:5.8\times10^{-4}$	61	50	57	400	92

[a] Ref. [5]; [b] Ref. [4]; [c] Ref. [6]; [d] Ref. [5]; [e] Ref. [31]; [f] Ref. [49]; and [g] Ref. [17].

petition experiments will be analyzed by each model. Model 1 does not separate the cooperativity from the intrinsic binding constant; however, Model 2 requires an ω value. Therefore, cooperativity values will be obtained from the literature and the results from the two models will be compared.

The conditions of no gaps and small n hold for both gene 5 and gene 32 proteins since they are highly cooperative and have binding site sizes of about 4 and 10, respectively. Figures 7A and 7B are three-dimensional graphs that show the relative affinity ratio for Model 1 (eq. 24) and Model 2 (eq. 32) as a function of ν_A(fd DNA) and ν_B(poly A) for gene 5 protein. The graphs for gene 32 protein (Figs. 8A and 8B) are very similar to those for gene 5 protein. The two functions result in surfaces with a similar shape; however, the affinity ratios for Model 2 are smaller. The largest differences in the ratios of the two models

Table 2

Comparison of experimentally obtained lattice saturation and calculated affinity ratios for Model 1 and Model 2

Competition system (case A:case B)	$\nu_A:\nu_B$	ν_{max}	Percent saturation (case A:case B)	$\frac{K_A/K_B \text{ (Model 1)}}{K_A/K_B \text{ (Model 2)}}$
Gene 5 [a]				
poly dT:fd DNA	$0.17:5.0\times10^{-4}$	0.25	68:0.2	24
fd DNA:poly A	0.14:0.015	0.25	56:6	7.8
Gene 32 [b]				
poly dT:poly dI	0.018:0.0055	0.1	18:5.5	2.4
poly dI:poly I	0.0055:0.0019	0.1	5.5:1.9	1.7
SSB [c]				
poly dT:fd DNA	0.0054:0.0016	0.0154	35:10	1.1
fd DNA:poly A	$0.011:5.8\times10^{-4}$	0.0154	71:3.8	1.1

[a] 20 m*M* Tris·HCl (pH 8.1), 1 m*M* sodium EDTA, 0.1 m*M* dithiothreitol, 10% (w/v) glycerol, 0.05% Triton, and 125 m*M* NaCl.
[b] 20 m*M* Tris·HCl (pH 8.1), 1 m*M* sodium EDTA, 0.1 m*M* dithiothreitol, 10% (w/v) glycerol, and 100 m*M* NaCl.
[c] 20 m*M* Tris·HCl (pH 8.1), 1 m*M* sodium EDTA, 0.1 m*M* dithiothreitol, 10% (w/v) glycerol, 0.05% Triton, and 125 m*M* NaCl.

occur when the competitors have large differences in binding density. These extremes are usually avoided in competition experiments by using a "bridge" nucleic acid having an intermediate affinity (e.g. fd DNA).

Table 1 lists the affinity ratios obtained with Models 1 and 2 using data from previously published competition experiments. The values for ω have been taken from the literature as noted. The two models yield affinity ratios for gene 5 protein that differ by a factor of 7.8–24 (Table 2), whereas the values for gene 32 protein only differ by 1.7–2.4. One possible reason for the, respectively, bad and good agreement between the two models is that the two competitors for the gene 5 protein experiments had a much larger difference in binding density than did those for the gene 32 protein experiments. Competitor A in the examples for gene 5 protein had binding densities that were 56–68% of the maximum binding density ($\nu_{max} = 0.25$) while competitor B had percentages of saturation of only 0.2–6%. However, competitor A in the gene 32 protein experiments only had 5.5–18% of $\nu_{max} = 0.1$ compared to 1.9–5.5% for competitor B. The regions of the 3-D plots containing the observed affinity ratios for the gene 5 proteins (Figs. 9A and 9B) and gene 32 protein (Figs. 10A and 10B) experiments illus-

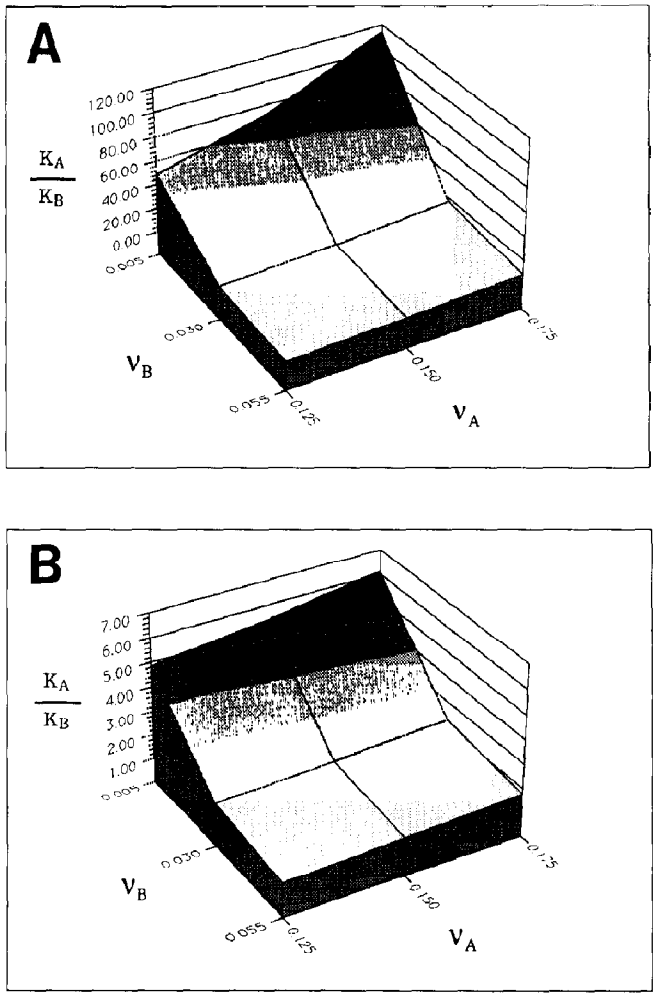

Fig. 9. The region of the 3-D graphs of Fig. 7 that is relevant to the competition experiments for (A) Model 1 and (B) Model 2.

trate the differences. A second possibility for the discrepancy is that gene 5 protein's unique manner of binding to single-stranded DNA may not be adequately described by Model 2 [36,39].

Figures 11A and 11B show the dependence of the affinity ratio on the cooperativity for Model 2. It is clear from the graphs that the dependence of the affinity ratio on the cooperativity is greatest for small values of ω. As the cooperativity becomes very large, the affinity ratio approaches one. This is a result of protein interactions becoming dominant enough to mask different protein–nucleic acid affinities. An example of a highly cooperative system is the protein coat in tobacco mosaic virus which can exist even without the nucleic acid lattice present [48]. As pointed out by McGhee and von Hippel [11], at very high cooperativities the binding isotherm becomes two state where a small increase in protein concentration will convert a lattice from completely unsaturated to completely saturated. Electron microscopy indicates that neither gene 5 [44] nor gene 32 [43] protein forms gaps when binding to DNA. Thus, if gaps are not significant for proteins of even moderate cooperativity, Model 1 should be valid for a large range of cooperativity values. This conclusion is further confirmed by the flat region of the curves in Figs. 11A and 11B for cooperativities larger than 50.

The situation with SSB protein is complicated

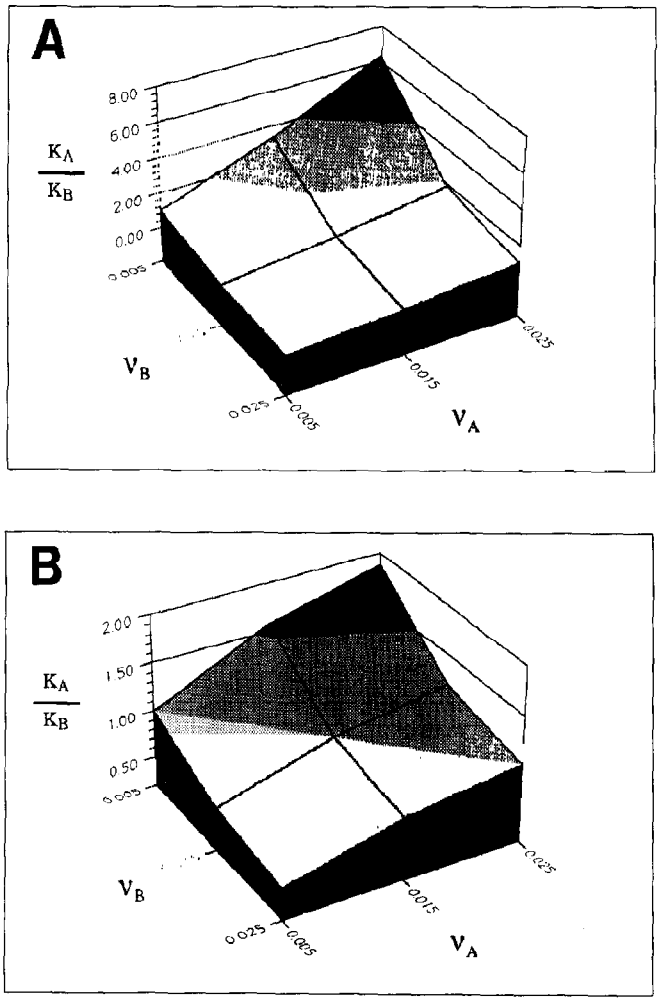

Fig. 10. The region of the 3-D graphs of Fig. 8 containing the observed affinity ratios for (A) Model 1 and (B) Model 2.

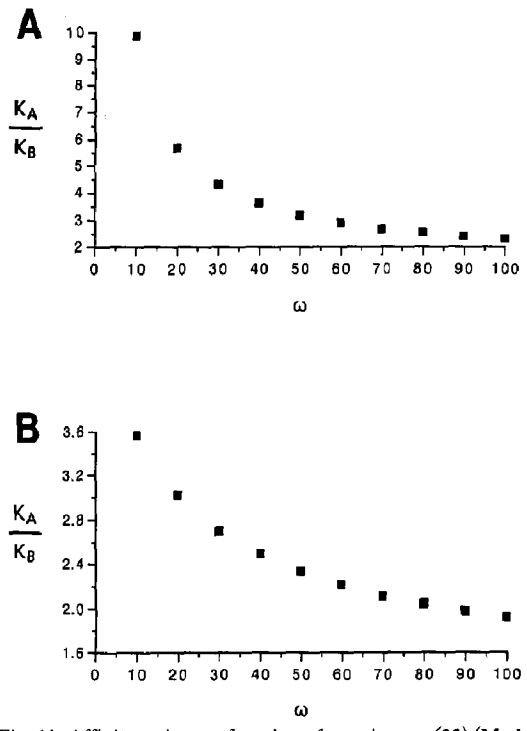

Fig. 11. Affinity ratio as a function of ω using eq. (32) (Model 2) with $\omega_A = \omega_B$. (A) gene 5 protein with competitors fd DNA and poly A; (B) gene 32 protein with competitors poly dT and poly dI.

by a number of additional features. One is that this protein can bind with a site size of 65. Therefore, a lattice of 500 residues would only be capable of occupying seven protein molecules. Because of the large n, the cluster end effects may become significant and both Model 1 and Model 2 may contain systematic errors that must be considered. Since one of the lattices used in the competition experiments contains different classes of binding sites (i.e. fd DNA), Model 2 should be modified to account for this. However, this would involve introducing variable probabilities into the equations which would drastically increase the complexity of the analysis. Therefore, Model 2 will be used in its standard form with the understanding that this approximation is being made. Another consideration with SSB is that data from electron microscopy have suggested that two morphologies are present in SSB complexed to nucleic acid [45,46]. The Lohman group has interpreted these morphologies as being representative of two distinct binding modes [49] corresponding to two types of cooperative interactions [50]. One type is an "unlimited" cooperativity where protein molecules bind contiguously along lattices in typical fashion. The other type is a "limited" cooperativity where interactions occur between properly oriented pairs of tetramers but not between the resulting octamers [18]. Chrysogelos et al. [45] have proposed a nucleosome-like structure consisting of SSB octamers wrapped in 145 base stretches of DNA separated by 30 base linkers. Bujalowski and Lohman [18] have used the sequence generating function (SGF) method to develop a binding equation for a linear lattice model taking this "limited" cooperativity into account. The number of potential free binding sites can be extracted from the binding equation resulting in:

$$N_{Avg} = N\left\{\frac{\left[q^2 - 2\nu q + (1-\omega)\nu^2\right]^n}{q^{2n-1}}\right\} \qquad (46)$$

where $q = [1-(n-1)\nu] + \{[1-(n-1)\nu]^2 - \nu(1-\omega)[2-(2n-1)\nu]\}^{1/2}$. The corresponding affinity ratio relationship for overlapping sites with "limited" cooperativity is:

$$\frac{K_A}{K_B} = \frac{\nu_A\left\{\frac{\left[q_B^2 - 2\nu_B q_B + (1-\omega_B)\nu_B^2\right]^n}{q_B^{2n-1}}\right\}}{\nu_B\left\{\frac{\left[q_A^2 - 2\nu_A q_A + (1-\omega_A)\nu_A^2\right]^n}{q_A^{2n-1}}\right\}} \qquad (47)$$

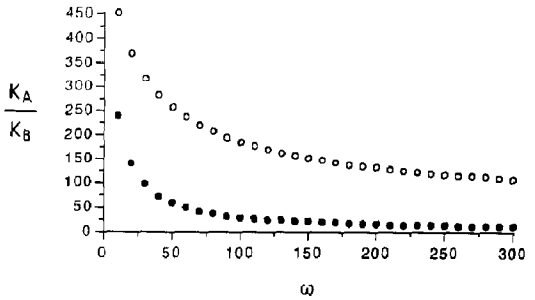

Fig. 12. Affinity ratio as a function of ω using eq. (32) (Model 2) (●) and eq. (47) ("Limited" ω Model) (○) with $\omega_A = \omega_B$ for SSB with competitors fd DNA and poly A.

Since the form of the cooperativity underlying eq. (47) is such that there is no cooperativity between octamers, the distribution of octamers on nucleic acid lattices should be random. In general, there should be no uncomplexed lattices; therefore, calculating the number of free binding sites on an average lattice is valid. However, as with Model 2, the SGF method contains the infinite lattice assumption. Thus, the large site size of SSB may result in end-effect errors.

In the "unlimited" binding mode, both Model 1 and Model 2 are applicable since gaps are insignificant. However, each model may contain a systematic error due to the large site sizes involved. If "limited" binding occurs, then Model 2 will not provide a correct analysis, since it assumes that any cooperativity present will occur equally between any bound protein. If gaps are insignificant, then Model 1 will work for the "limited" binding mode, since it does not assume any form for the cooperativity. In addition, if gaps of constant size occur, then the stoichiometry, s, will include them correctly in the Model 1 analysis. If in the "limited" cooperativity mode gaps of variable size occur, then eq. (47) should be used; however, there may be end-effect errors due to the large n values.

The "unlimited" and "limited" cooperativity

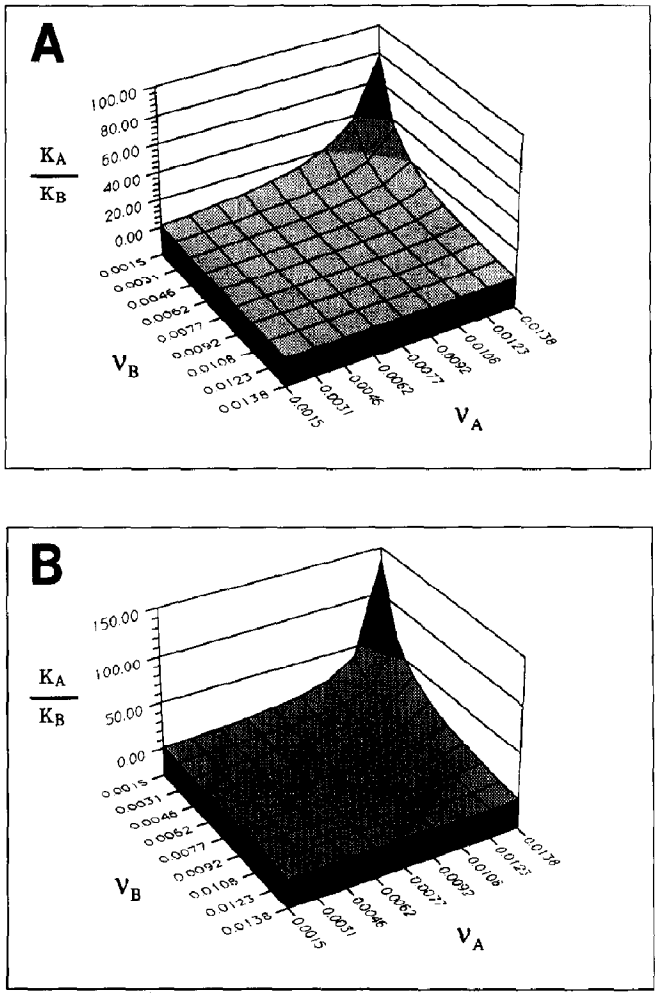

Fig. 13. Competition data analyzed according to (A) Model 1 and (B) Model 2 for SSB binding to fd DNA (competitor A) and poly A (competitor B) (see Table 1).

modes are thought to correspond to $n = 33$ and $n = 65$ binding modes, respectively [49]. Since the $n = 65$ binding mode is considered to be the preferred mode under high salt conditions, this value was used for calculations. Figure 12 shows the dependence of Model 2 on cooperativity for SSB when $\omega_A = \omega_B$. This approach is valid for the case of "unlimited" cooperativity. When the "limited" model (eq. 47) is applied, the resulting curve is similar to that of the "unlimited" model; however, the affinity ratios are shifted to higher values (Fig. 12). The higher cooperativity values obtained with this "limited" model reflect the assumption that cooperativity only occurs between corresponding tetramers. Table 1 summarizes the results for all three models. The fact that all three models give similar results is intriguing. From eq. (45), one would expect a site size of 65 to result in large differences between models. This is particularly true for the fd DNA:poly A system due to the high binding density of 71% on fd DNA (Table 2). The 3-D

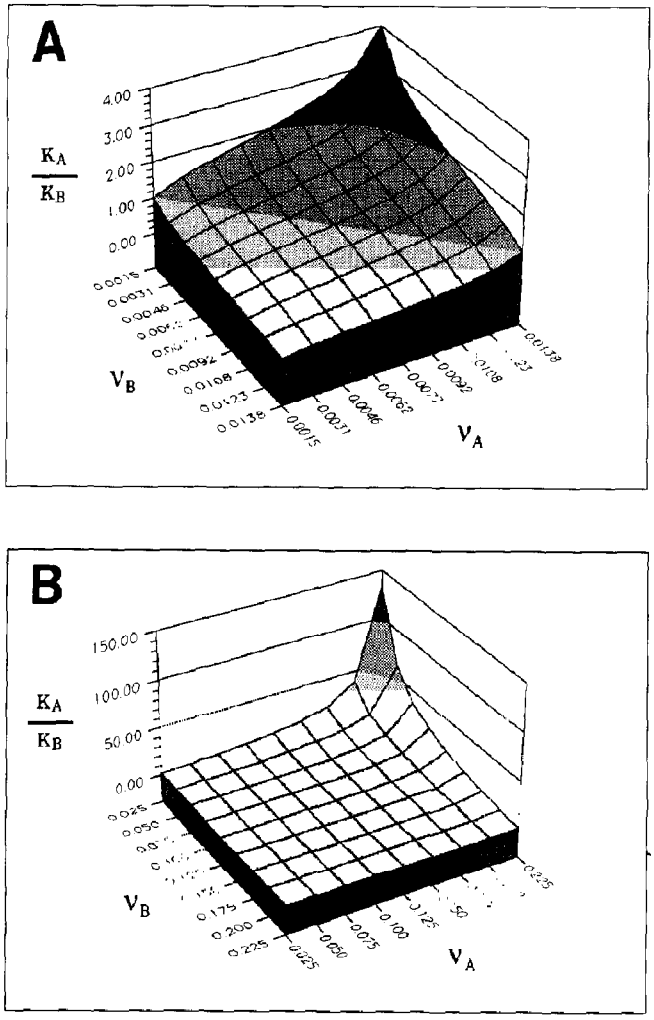

Fig. 14. Illustration of the effect of cooperativity on the 3-D graphs of SSB and gene 5 protein. (A) At an $\omega = 900$, the affinity ratios for SSB (Fig. 13B) become very similar to those for gene 5 protein (Fig. 7B). (B) Likewise, when $\omega = 3.5$, the plot for gene 5 protein (Fig. 7B) becomes similar to that of SSB (Fig. 13B).

graphs for this system are shown in Figs. 13A and 13B. Clearly, the plots have similar affinity ratios in the range of binding densities obtained in the experiments. Why does a high binding density in the case of SSB not lead to very different affinity ratios for the two models as with gene 5 protein particularly when SSB has such a large site size? Comparison of the 3-D graphs of Model 2 for SSB (Fig. 13B) and gene 5 (Fig. 7B) proteins indicates that the surface of the gene 5 plot is flattened out with respect to that of the SSB plot. In other words, at high ν_A and low ν_B, the plot for SSB gives much larger affinity ratios.

For a given protein, the shape of the 3-D plot's surface is a strong function of the cooperativity. Increasing ω for SSB would result in a plot similar to that of gene 5 protein (Fig. 14A). Likewise, decreasing ω for gene 5 protein would yield a larger peak in the affinity ratio plot (Fig. 14B). Clearly, by choosing the appropriate ω value, the plots for Model 1 and Model 2 for any protein can be made very similar. Also, if ω is lattice dependent, as has been found for gene 5 protein [51], then ω_A and ω_B may differ in the affinity ratio equations for Model 2. If $\omega_A \neq \omega_B$, then Models 1 and 2 can be brought into close agreement by choosing the appropriate pairs of ω values. Therefore, comparison of the affinity ratios for Models 1 and 2 must consider the value of ω used for Model 2. However, as previously discussed, reported values of ω for a given protein differ by several orders of magnitude.

6. Conclusion

Electron paramagnetic resonance competition experiments with spin-labeled nucleic acids form the basis of a simple, yet powerful approach for determining the relative affinities of protein–nucleic acid systems. In order to obtain quantitative information, the free binding sites must be defined in terms of a model. As described in this paper, a competition formalism based on Model 1 and Model 2 both result in similar affinity ratios provided their respective conditions are met.

Both Models 1 and 2 are applicable to systems displaying moderate to high cooperativity. Model 1 has the advantage of containing only experimentally determined parameters thereby simplifying the analysis. For situations involving kinetically blocked secondary structures or sites with different affinities, calculations become more elaborate for Model 2 and additional parameters must be introduced. Model 1 readily accounts for different stoichiometries and different classes of binding sites. Proteins that bind with little or no cooperativity require use of Model 2, since gap formation will become important. Under these circumstances, protein molecules are more randomly distributed and application of statistical thermodynamics will be more straightforward.

From the theoretical analysis of Models 1 and 2, it can be concluded that under conditions of minimal gap formation and negligible cluster end effects both models yield similar affinity ratios. Equations (41) and (45) demonstrate the similarity in the two approaches mathematically. Indeed, there is little difference in affinity ratio between the two models when experimental data for gene 32 and SSB proteins is analyzed. There is a larger difference for gene 5 protein that may be attributable to the competitor binding densities obtained in the competition experiments as well as to a breakdown in the linear lattice assumptions of Model 2.

It is important to remember that the affinity ratio values for Model 2 depend both on the magnitude and the form of cooperativity operating. Current methods do not allow for the direct determination of either of these aspects of cooperativity. Model 1 has a distinct advantage of needing neither a definite value of ω nor the form of cooperativity since it requires only experimentally determined parameters that account for these factors.

Thus, although Model 1 appears to be highly simplified, closer analysis reveals that taking the ratio of binding constants introduces a statistical element similar to that of Model 2. As a result of this, in the absence of reliable cooperativity data, competition analyses of systems displaying moderate to high cooperativity should preferably be done with Model 1.

The advantages and limitations of both models are resumed in the following subsections.

6.1 Model 1

6.1.1 Advantages

(1) The variables in the affinity ratio equation for Model 1 (Eq. 23) are all directly obtained by experiment. Thus, Model 1 follows the operational principle of P.W. Bridgman which requires that all symbols represent purely physical operations and observations.
(2) Model 1 does not assume nearest-neighbor cooperativity. This allows it to readily accommodate longer range interactions (e.g. gene 5 protein) and "limited" cooperativity (assuming constant gap size).
(3) Model 1 does not require a cooperativity value.
(4) Base preferences exhibited by single-strand binding proteins resulting in different K_{int} values along the DNA strand are automatically accounted for by Model 1.
(5) The problem of secondary structure is resolved by using the stoichiometry, s, rather than the site size, n.

6.1.2 Limitations

(1) Model 1 requires that proteins bind in long clusters resulting in minimal gap formation between bound ligands. In other words, the ligands must bind with moderate to high cooperativity. Alternatively, the proteins must bind with constant gap size such that s can accommodate the gap into the number of residues per protein molecule.
(2) Since Model 1 neglects potential binding sites eliminated at one end of a cluster, proteins with large site sizes will introduce end-effect errors.

6.2 Model 2

6.2.1 Advantages

(1) Model 2 is based upon a statistical thermodynamic approach which attempts to be comprehensive in its applicability. It allows separate K_{int} and ω values to be determined rather than only a composite K_{app}.
(2) Model 2 is valid for any cooperativity value and, therefore, can handle proteins that bind with gaps in between binding sites.
(3) Theoretically, Model 2 should be valid for any site size; however, see limitation (4) below.

6.2.2 Limitations

(1) Preferential binding to different sequences of bases significantly increases the difficulty of determining the relative affinities between two competitors. The analysis requires the determination of multiple K_{int} values (eq. 34).
(2) Model 2 requires a cooperativity value as a parameter for calculating the relative affinity ratio.
(3) The cooperativity cannot significantly deviate from the nearest-neighbor assumption for the "unlimited" ω model or the tetramer pairwise-interaction assumption for the "limited" ω model.
(4) Since a sample of nucleic acid will contain a collection of finite lattices rather than an infinite lattice, there will be free lattices in solution at all but the highest fractions of saturation. These free lattices should not be included in the ensemble when determining the statistical distribution of ligand on the complexed lattices. This, in effect, leads to end-effect errors. Thus, the magnitude of the error will increase with increasing site size.
(5) Secondary structure must be accounted for by introducing an additional equilibrium constant, K_{conf}.

Acknowledgments

We would like to thank Dr. Frank Meeks for helpful discussions concerning statistical thermodynamics. This work was in part supported by NIH grant GM27002.

References

1 A. Revzin, in: The biology of nonspecific DNA–protein interactions, ed. A. Revzin (CRC Press, Boca Raton, FL, 1990) p. 5.
2 N.R. Kallenbach, Chemistry and physics of DNA–ligand interactions (Adenine Press, Schenectady, NY, 1990).

3 M. Russel, L. Gold, H. Morrissett and P.Z. O'Farrell, J. Biol. Chem. 251 (1976) 7263.
4 A.M. Bobst and Y.-C.E. Pan, Biochem. Biophys. Res. Commun. 67 (1975) 562.
5 A.M. Bobst, P.W. Langemeier, P.E. Warwick-Koochaki, E.V. Bobst and J.C. Ireland, J. Biol. Chem. 257 (1982) 6184.
6 A.M. Bobst, J.C. Ireland and E.V. Bobst, J. Biol. Chem. 259 (1984) 2130.
7 E.V. Bobst, F.W. Perrino, R.R. Meyer and A.M. Bobst. Biochim. Biophys. Acta 1078 (1991) 199.
8 G. Scatchard, Ann. N.Y. Acad. Sci. 51 (1949) 660.
9 S.A. Latt and H.A. Sober, Biochemistry 6 (1967) 3293.
10 A.S. Zasedatelev, G.V. Gurskii and M.V. Volkenshtein, Mol. Biol. 5 (1971) 245.
11 J.D. McGhee and P.H. von Hippel, J. Mol. Biol. 86 (1974) 469.
12 J.A. Schellman, Isr. J. Chem. 12 (1974) 219.
13 G. Schwarz, Biophys. Chem. 6 (1977) 65.
14 I.R. Epstein, Biophys. Chem. 8 (1978) 327.
15 I.R. Epstein, Biopolymers 18 (1979) 2037.
16 G. Schwarz and S. Stankowski, Biophys. Chem. 10 (1979) 173.
17 G. Schwarz and F. Watanabe, J. Mol. Biol. 163 (1983) 467.
18 W. Bujalowski and T.M. Lohman, J. Mol. Biol. 195 (1987) 897.
19 D. Marsh and L.I. Horváth, in: Advanced EPR: Applications in biology and biochemistry, ed. A.J. Hoff (Elsevier, New York, 1989) p. 707.
20 D. Marsh, in: Spin labeling: Theory and applications, eds. L.J. Berliner and J. Reuben, Biological Magnetic Resonance, vol. 8 (Plenum, New York, 1989) p. 255.
21 T.M. Lohman, Biochemistry 23 (1984) 4656.
22 D. Pörschke and H. Rauh, Biochemistry 22 (1983) 4737.
23 G. Krauss, H. Sindermann, U. Schomburg and G. Maass, Biochemistry 20 (1981) 5346.
24 R.J. Schneider and J.G. Wetmur, Biochemistry 21 (1982) 608.
25 R. Römer, U. Schomburg, G. Krauss and G. Maass, Biochemistry 23 (1984) 6132.
26 A.M. Bobst, T.K. Sinha, P.W. Langemeier and R.L. Prairie, Comput. Chem. 4 (1980) 45.
27 R.C. Kelly, D.E. Jensen and P.H. von Hippel, J. Biol. Chem. 251 (1976) 7240.
28 J.W. Newport, N. Lonberg, S.C. Kowalczykowski and P.H. von Hippel, J. Mol. Biol. 145 (1981) 105.
29 J.H. Weiner, L.L. Bertsch and A. Kornberg, J. Biol. Chem. 250 (1975) 1972.
30 I.J. Molineux, A. Pauli and M.L. Gefter, Nucleic Acids Res. 2 (1975) 1821.
31 P.H. von Hippel, S.C. Kowalczykowski, N. Lonberg, J.W. Newport, L.S. Paul, G.D. Stormo and L. Gold, J. Mol. Biol. 162 (1982) 795.
32 F. Watanabe, FEBS Lett. 242 (1989) 444.
33 S.C. Kowalczykowski, N. Lonberg, J.W. Newport and P.H. von Hippel, J. Mol. Biol. 145 (1981) 75.
34 T.M. Lohman, Biochemistry 23 (1984) 4665.
35 M.E. Kuil, H. van Amerongen, P.C. van der Vliet and R. van Grondelle, Biochemistry 28 (1989) 9795.
36 N.C.M. Alma, B.J.M. Harmsen, E.A.M. de Jong, J.V.D. Ven and C.W. Hilbers, J. Mol. Biol. 163 (1983) 47.
37 J. Greipel, G. Maass and F. Mayer, Biophys. Chem. 26 (1987) 149.
38 W.T. Ruyechan and J.G. Wetmur, Biochemistry 14 (1975) 5529.
39 N.C.M. Alma, B.J.M. Harmsen, J.H. van Boom, G. van der Marel and C.W. Hilbers, Eur. J. Biochem. 122 (1982) 319.
40 R.L. Karpel, in: The biology of nonspecific DNA–protein interactions, ed. A. Rev2in (CRC Press, Boca Raton, FL, 1990) p. 103.
41 G. Schwarz, Ber. Bunsenges. Phys. Chem. 76 (1972) 373.
42 T.M. Lohman and S.C. Kowalczykowski, J. Mol. Biol. 152 (1981) 67.
43 H. Delius, N.J. Mantell and B. Alberts, J. Mol. Biol. 67 (1972) 341.
44 B. Alberts, L. Frey and H. Delius, J. Mol. Biol. 68 (1972) 139.
45 S. Chrysogelos and J. Griffith, Proc. Natl. Acad. Sci. USA 79 (1982) 5803.
46 J.D. Griffith, L.D. Harris and J. Register III, Cold Spring Harbor Symp. Quant. Biol. 49 (1984) 553.
47 G. Felsenfeld, G. Sandeen and P.H. von Hippel, Proc. Natl. Acad. Sci. USA 50 (1963) 644.
48 H. Fraenkel-Conrat, Design and function at the threshold of life: The viruses (Academic Press, New York, 1962).
49 T.M. Lohman and L.B. Overman, J. Biol. Chem. 260 (1985) 3594.
50 T.M. Lohman, L.B. Overman and S. Datta, J. Mol. Biol. 187 (1986) 603.
51 H. Bulsink, B.J.M. Harmsen and C.W. Hilbers, J. Biomol. Struct. Dyn. 3 (1985) 227.