Neya, S., & Morishima, I. (1981) *J. Biol. Chem. 256*, 793-798.
Neya, S., Hada, S., & Funasaki, N. (1983) *Biochemistry 22*, 3686-3691.
Neya, S., Hada, S., & Funasaki, N. (1985a) *Biochim. Biophys. Acta 828*, 241-246.
Neya, S., Hada, S., Funasaki, N., Umemura, J., & Takenaka, T. (1985b) *Biochim. Biophys. Acta 827*, 157-163.
Perutz, M. F., Heidner, E. J., Ladner, J. E., Bettlestone, J. G., Ho, C., & Slade, E. F. (1974) *Biochemistry 13*, 2187-2200.
Perutz, M. F., Sanders, J. K. M., Chenery, D. H., Noble, R. W., Pennelly, R. R., Fung, L. W.-M., Ho, C., Giannini, F., Pörschke, D., & Winkler, H. (1978) *Biochemistry 17*, 3640-3652.
Philo, J. S., & Dreyer, U. (1985) *Biochemistry 24*, 2985-2992.
Savicki, J. P., Lang, G., & Ikeda-Saito, M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 5417-5419.
Scholler, D. M., & Hoffman, B. M. (1979) *J. Am. Chem. Soc. 101*, 1655-1662.
Scholler, D. M., Wang, N.-Y. R., & Hoffman, B. M. (1978) *Methods Enzymol. 52*, 487-493.
Verloop, A., Hoogenstraaten, W., & Tipker, J. (1976) in *Drug Design* (Ariëns, E. J., Ed.) Vol. 7, pp 165-207, Academic Press, New York.
Winter, M. R. C., Johnson, C. E., Lang, G., & Williams, R. J. P. (1972) *Biochim. Biophys. Acta 263*, 515-534.
Wüthrich, K. (1970) in *Structure and Bonding* (Hemmerich, R., Jorgensen, C. K., Neilands, J. B., Nyholm, S. R. S., Reinen, D., Williams, R. J. P., Eds.) Vol. 8, pp 53-121, Springer-Verlag, West Berlin.
Yonetani, T., Iizuka, T., Asakura, T., Otsuka, J., & Kotani, M. (1972) *J. Biol. Chem. 247*, 863-868.

Cooperative and Noncooperative Binding of Protein Ligands to Nucleic Acid Lattices: Experimental Approaches to the Determination of Thermodynamic Parameters†

Stephen C. Kowalczykowski,‡,§ Leland S. Paul,‡,∥ Nils Lonberg,‡,⊥ John W. Newport,‡,# James A. McSwiggen,‡,○ and Peter H. von Hippel*,‡

Institute of Molecular Biology and Departments of Chemistry and Biology, University of Oregon, Eugene, Oregon 97403, and Department of Molecular Biology, Northwestern University Medical School, Chicago, Illinois 60611

Received August 2, 1985

ABSTRACT: Many biologically important proteins bind nonspecifically, and often cooperatively, to single- or double-stranded nucleic acid lattices in discharging their physiological functions. This binding can generally be described in thermodynamic terms by three parameters: n, the binding site size; K, the intrinsic binding constant; ω, the binding cooperativity parameter. The experimental determination of these parameters often appears to be straightforward but can be fraught with conceptual and methodological difficulties that may not be readily apparent. In this paper we describe and analyze a number of approaches that can be used to measure these protein–nucleic acid interaction parameters and illustrate these methods with experiments on the binding of T4-coded gene 32 (single-stranded DNA binding) protein to various nucleic acid lattices. We consider the following procedures: (i) the titration of a fixed amount of lattice (nucleic acid) with added ligand (protein); (ii) the titration of a fixed amount of ligand with added lattice; (iii) the determination of ligand binding affinities at very low levels of lattice saturation; (iv) the analysis of ligand cluster size distribution on the lattice; (v) the analysis of ligand binding to lattices of finite length. The applicability and limitations of each approach are considered and discussed, and potential pitfalls are explicitly pointed out.

The nonspecific binding of proteins to single- or double-stranded nucleic acid lattices is a central feature of many functional and regulatory biological processes. Theoretically, the problem can be viewed as the binding of large ligands to a one-dimensional lattice, where each ligand interacts with more than one lattice unit (nucleotide residue or base pair), and thus also covers more than one potential ligand binding site (i.e., ligand binding is of the "overlap" type). Overlap binding of ligands to a one-dimensional lattice complicates the analysis of titration curves because the binding sites on the lattice are not titrated independently. Because of overlap, the number of free lattice binding sites occluded per binding event *decreases* with increasing saturation of the lattice. As a consequence, overlap binding is effectively "negatively cooperative", and it becomes progressively more unfavorable to bind additional ligands as lattice saturation is approached. In addition, protein binding to nucleic acid lattices may also be (and generally is) positively cooperative, in that the binding

† These studies were supported by USPHS Research Grants GM-15792 and GM-29158 (to P.H.v.H.), as well as (in part) by Research Grant AI-18987 (to S.C.K.). L.S.P., J.W.N., and J.A.M. were predoctoral trainees on USPHS Institutional National Research Service Award GM-07759 during parts of these studies. Some of this work was also submitted by L.S.P. to the Graduate School of the University of Oregon in partial fulfillment of the requirements for the Ph.D. degree in chemistry.

‡ University of Oregon.
§ Northwestern University Medical School.
∥ Present address: Abbott Laboratories, North Chicago, IL 60064.
⊥ Present address: Sloan-Kettering Institute, New York, NY 10021.
Present address: Department of Biology, University of California at San Diego, La Jolla, CA 92093.
○ Present address: Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309.

0006-2960/86/0425-1226$01.50/0

of an additional protein adjacent to one that is already bound is favored over binding to an isolated site. This antagonizes the overlap effect, promotes the clustering of bound ligands, and facilitates lattice saturation. [For a complete theoretical discussion of overlap and cooperative binding to a linear lattice, see McGhee & von Hippel (1974).]

The effects of nonspecific protein–nucleic acid interactions on biological regulatory mechanisms are widespread and can be illustrated by the following examples. The processes of DNA replication, recombination, and repair require the binding of single-stranded DNA binding proteins to transiently single-stranded sequences formed as intermediates in these processes [for a review, see Kowalczykowski et al. (1981a)]; binding is nonspecific and often cooperative to force saturation of the lattice. Repressors, RNA polymerase, and other genome regulatory proteins use nonspecific binding to random sequences on the chromosome as an intermediate step in the location of specific DNA regulatory targets such as operators and promoters (Berg et al., 1982) and to provide modulation of the net affinity for the specific binding site through competitive binding (von Hippel et al., 1974; Lin & Riggs, 1975). Nonspecific protein–nucleic acid interactions have also been utilized in controlling the translation of messenger RNA through the cooperative binding of the regulator protein to the ribosome binding site on the RNA [see von Hippel et al. (1982)]. Analysis of the above systems (as well as other functional and regulatory systems involved in gene expression) has shown that quantitative measures of nonspecific protein–nucleic acid binding parameters are required to interpret (and to discriminate) competing molecular biological models of such processes, since such models can often be distinguished only in physical chemical terms.

A variety of mathematical approaches have been taken to solving the problem of nonspecific binding of large ligands to one-dimensional lattices. Latt and Sober (1967) used both a combinatorial method and the method of sequence-generating functions to investigate lattice binding of noninteracting ligands. Crothers (1968) and Schwarz (1970) used a matrix method to describe both the noninteractive and the interactive (cooperative) binding of ligands to nucleic acid lattices. Zasadetlev et al. (1971) developed a combinatorial procedure for the same purposes, while Schellman (1974) has applied the method of sequence-generating functions to the solution of this problem. McGhee and von Hippel (1974) used a conditional probability approach and were able to obtain solutions in closed analytic form for both cooperative and noncooperative binding.

Despite this multiplicity of theoretical approaches, the problems associated with applying these theories to real systems have often led to difficulty and confusion. In our laboratories, over a period of years, we have been developing a variety of experimental approaches to these protein–nucleic acid interaction systems and have been trying to sort out the interpretive complications associated with each. Some of these approaches are summarized here. For convenience, we have formulated our discussion in the context of the McGhee–von Hippel version of the binding theory, which has the advantage of easy computational manipulation. However, this presentation could easily be recast in terms of any of the other theoretical methods listed above [e.g., see Schwarz & Watanabe (1983) and Watanabe & Schwarz (1983)].

The simple model used in all these theories, as well as in this paper, describes the binding of proteins to nucleic acids in terms of three thermodynamic parameters: n, the binding site size of the protein (in units of nucleotide residues or base pairs); K, the intrinsic binding constant (in M^{-1});[1] ω, the cooperativity parameter (unitless) that specifies the relative affinity of an incoming ligand for a contiguous, as opposed to an isolated, binding site. These parameters are illustrated in Figure 1. Obviously a description in terms of single values of n, K, and ω provides an oversimplified model of many real biological systems. Even the paradigm system (T4 gene 32 protein binding to single-stranded DNA) fails to fit it exactly. Possible complications include heterogeneous binding to compositionally different nucleic acid sequences (more than one value of K), different modes of binding of the protein to the nucleic acid lattice (more than one value of n), and cooperative interactions between ligands beyond the nearest-neighbor level (more than one value of ω). Furthermore, most of the theoretical models (including McGhee–von Hippel) were developed to work with long (effectively infinite) lattices, while many real biological systems involve binding to short nucleic acid lattices. Theoretical ways of handling the finite lattice problem have been developed by Epstein (1978) and by J. A. Schellman (unpublished results); some experimental consequences of finite lattice binding are discussed in this paper.

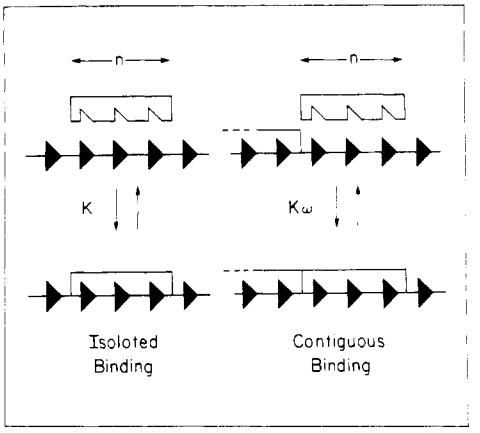


FIGURE 1: Definitions of the thermodynamic parameters that describe the interaction of a protein with a nucleic acid lattice. Each arrowhead represents a lattice site (i.e., a nucleotide residue), and the illustrated protein ligand covers three such sites ($n = 3$). K (M^{-1}) is the intrinsic association constant for protein binding to the lattice at an isolated site, and ω (dimensionless) represents the cooperativity of binding. ω is defined as the equilibrium constant for moving a protein from a given isolated site to a given contiguous binding site. Thus, $K\omega$ is the net binding constant per contiguously bound protein molecule. If contiguous binding is favored, $\omega > 1$, if the binding is noncooperative, $\omega = 1$, and if the binding is negatively cooperative, $\omega < 1$.

The descriptive limitations of the model we use should be clearly borne in mind. However, the "open-endedness" of assigning more detailed and specific interaction parameters to models of real biological systems, plus the errors involved in real measurements, generally limits us to the use of the simple model and the three thermodynamic parameters defined above. The determination and manipulation of these parameters is described in what follows. Knowing these parameters for a given system is generally useful in itself, and obviously they can be further refined or subdivided as additional knowledge of the system at issue is gathered by other techniques.

[1] The units of K, like those of n, can be expressed either as nucleotide residues or as base pairs. The former represents the obvious choice for the formation of the single-stranded nucleic acid–protein complexes that are primarily considered as examples in this paper. Either set of units is appropriate for the treatment of double-stranded nucleic acid–protein complexes, but the choice must be explicitly made and stated since the magnitude of the listed value of K (and of n) will depend on it. For additional discussion of this point, in connection with *E. coli lac* repressor–DNA complexes, see Butler et al. (1977).

Materials and Methods

The preparation and manipulation of the nucleic acids and proteins used in this work, as well as the details of experimental procedures, specific buffer systems, etc., have been described elsewhere (Kowalczykowski et al., 1981b; Newport et al., 1981; Lonberg et al., 1981). The fluorescence titrations described in this paper were performed either with a Schoeffel spectrofluorometer, as described by Kowalczykowski et al. (1981b), or with a SLM-8000 spectrofluorometer interfaced to a Hewlett-Packard HP-87 computer (J. A. McSwiggen et al., unpublished results). Fractionated samples of poly(riboadenylic acid) [poly(rA)] were obtained from Miles Biochemical Co.

Experimental Approaches

Equations

Because the equations derived by McGhee and von Hippel (1974) can be written in closed form, they provide a useful basis for the quantitative analysis of the binding of proteins to nucleic acids. For noncooperative systems, we write

$$\frac{\nu}{L} = K(1 - n\nu)\left[\frac{1 - n\nu}{1 - (n-1)\nu}\right]^{n-1} \quad (1)$$

and for cooperative systems, we write

$$\frac{\nu}{L} = K(1 - n\nu) \times \left[\frac{(2\omega - 1)(1 - n\nu) + \nu - R}{2(\omega - 1)(1 - n\nu)}\right]^{n-1}\left[\frac{1 - (n+1)\nu + R}{2(1 - n\nu)}\right]^{2} \quad (2)$$

where $R = [[1 - (n+1)\nu]^2 + 4\omega\nu(1 - n\nu)]^{1/2}$, ν is the binding density of the ligand on the lattice (in moles of ligand per mole of nucleotide residue or base pair), and L is the *free* ligand concentration (in moles per liter). The derivation of eq 1 and 2 assumes that ligand binding occurs on an infinite lattice. We use these equations throughout because of their ease of interpretation and evaluation. In a later section, we examine the limitations for real systems that result from the infinite lattice assumption.

Titration of Lattice with Ligand

The titration of a fixed amount of lattice with added ligand is probably the most direct and commonly used method of determining the thermodynamic parameters of a ligand–lattice (protein–nucleic acid) interaction. Typically, some lattice-dependent spectroscopic signal, reflecting a change in a lattice property that varies linearly with ligand binding, is monitored as a function of *total* ligand concentration. Less commonly, a ligand-based signal that is perturbed by the binding interaction is used, and the change in this signal is used to monitor the progressive saturation of the *lattice*. For our measurements of gene 32 protein binding, changes in the UV absorbance, circular dichroism, or fluorescence of the nucleic acid or of an appropriate nucleic acid derivative have been monitored as protein is added; alternatively, the deviation from a linear increase of the intrinsic fluorescence of the protein (due to quenching of the protein fluorescence in the bound state) as it is added to a lattice-containing solution may be followed [see Alma et al. (1983) for an example of this approach].

The result of a titration of lattice with ligand generally resembles the family of curves shown in Figure 2. In this figure we present model titration curves, which have been generated with eq 1 and 2, for noncooperatively and cooperatively binding ligands. The total apparent affinity (i.e., the

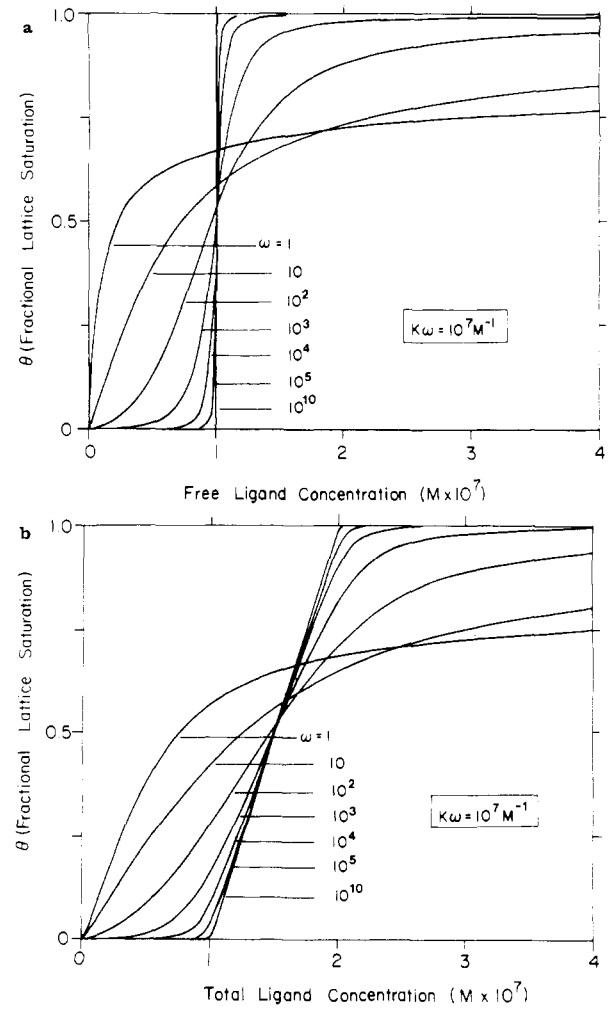


FIGURE 2: (a) Fractional lattice saturation vs. *free* ligand concentration for various values of K and ω; the product of K and ω has been held constant at 1×10^7 M^{-1}. Values of ω are indicated in the figure. The site size (n) is 10, and the concentration of lattice is 0. (b) Plot identical with (a), but the concentration of lattice units is 10^{-6} M, and the x axis becomes *total* ligand concentration.

product of K and ω) has been held constant in these calculations, while the individual values of K and ω are varied reciprocally. The most prominent feature of these plots is the progressive development of the classical sigmoid titration curve (indicative of cooperative ligand binding) as ω is increased. In addition, note that all of the curves with values of ω significantly larger than n pass through a common value that falls at the midpoint of each titration curve. As was pointed out by McGhee and von Hippel (1974), the value of the *free* protein concentration at the midpoint of such binding isotherms approaches $1/(K\omega)$ (for curves plotted in terms of *free* ligand concentration) for values of ω that are significantly larger ($\sim$10-fold) than n; the usefulness of this relationship will become apparent when the analysis of cooperatively binding systems is discussed.

The curves of Figure 2 have been presented in two equivalent ways in order to illustrate the difference between plotting fractional saturation as a function of *free* ligand and of *total* ligand concentration. Although the overall shapes of the plots are similar, the slopes of the steeply rising portions of the cooperative curves differ, due solely to the fact that in Figure 2b the total ligand concentration includes the ligand that is bound to the lattice. The amount of bound ligand is simply equal to νN, where N is the total concentration of lattice units (in nucleotide residues) present and ν (the binding density)

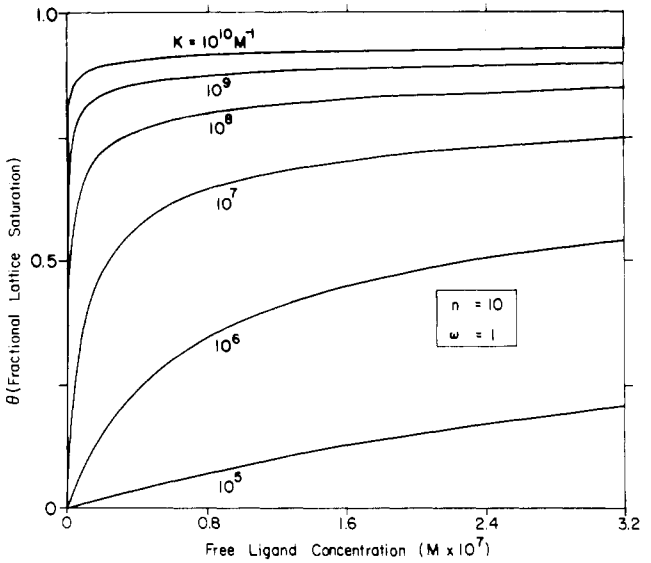


FIGURE 3: Fractional lattice saturation vs. free ligand concentration for a noncooperatively binding ligand, at various values of K, with $n = 10$ and $\omega = 1$.

is equal to the fractional saturation of the lattice (θ) divided by the ligand binding site size (n). Thus, the two sets of curves in Figure 2 are related through conservation of mass by

$$L_T = L + \nu N = L + \theta N/n \quad (3)$$

where L_T is the total ligand concentration and L is the free ligand concentration.

Since experimental data are obtained as a function of *total* ligand concentration and the theoretical curves generated with eq 1 and 2 are in terms of the *free* ligand concentration, it is necessary to convert one form to the other. However, in order to use eq 3 for this purpose, the binding site size (n) must be obtained, since θ, rather than ν, is generally the preferred dependent variable.[2]

The site size (n) for a protein can often be obtained by conducting a titration under stoichiometric binding conditions. In such a titration, an abrupt "break" is seen at the equivalence point, and the molar ratio of lattice residues (e.g., nucleotide residues or base pairs) to ligand concentration at the break is equal to the site size [for real examples, see Figure 3 in Kowalczykowski et al. (1981b) or Figure 1 in Newport et al. (1981)]. The problem of defining stoichiometric binding conditions will be addressed specifically below. However, such conditions can generally be defined experimentally as those under which there is no change in the stoichiometry at the equivalence point when the lattice concentration is varied over an approximately 10-fold range. Because the binding affinity of proteins for nucleic acids is often dependent on salt concentration [see Record et al. (1978, 1981) for reviews], one of the easiest experimental ways to increase the binding affinity of the protein for the nucleic acid is to lower the salt concentration. Measured site sizes, and the conditions required for tight binding for a number of single-stranded DNA binding proteins, have been summarized in Kowalczykowski et al. (1981a). In the next sections, we discuss approaches for the quantitative analysis of titrations of lattice with added ligand for systems in which binding is nonstoichiometric and either noncooperative or cooperative.

The Noncooperative Case. When ω has a value close to 1, no sigmoidicity of the binding isotherm is observed (see Figures 2 and 3). It is then a simple matter to model the titration curve if a value of n has been obtained from a stoichiometric binding experiment as described above. With this value of n, theoretical curves can be generated for various trial values of the binding constant (K) with eq 1 and compared with the experimental results. Alternatively, a nonlinear fitting procedure can be used (see Appendix). It is possible also to determine the value of K from the midpoint of the titration curve (i.e., at $\theta = {}^1/_2$). By substituting the value of n, the value of the free ligand concentration at the midpoint of the titration ($L_{1/2}$), and the value of ν at the midpoint [$\nu_{1/2}$, which equals $1/(2n)$], into eq 1, we obtain

$$K = \frac{\nu_{1/2}}{L_{1/2}}\left[(1 - n\nu_{1/2})\left[\frac{1 - n\nu_{1/2}}{1 - (n-1)\nu_{1/2}}\right]^{n-1}\right]^{-1} = nL_{1/2}(1 + 1/n)^{n-1} \quad (4)$$

This equation can be used to determine exact values for K if n is known.

If n is *not* known exactly, an estimate of the binding constant (K_{approx}) can be obtained by ignoring ligand overlap, thus reducing eq 4 to

$$K_{approx} = 1/(nL_{1/2}) \quad (5)$$

For example, for $n = 10$, eq 5 underestimates the true value of K approximately 2.3-fold, and for $n = 5$, it underestimates $K \sim$ 2-fold. Nevertheless, eq 5 is often helpful in estimating trial values of K for use in curve-fitting methods (see Appendix).

Despite the relative simplicity of determining K from noncooperative titration data once n is known, two experimental problems often make the analysis more difficult. The first is encountered when conditions cannot be found under which ligand binding is stoichiometric, thus precluding an independent determination of n. The second is related to the first, in that if binding is weak, or if n is large, saturation of the lattice cannot be achieved due to overlap problems [see McGhee & von Hippel (1974)] and consequently the end point of the titration cannot be accurately determined. In Figure 3, we have plotted a number of curves with correspondingly weaker binding affinities from left to right; such data might be encountered in protein–nucleic acid titrations as the salt concentration is increased [e.g., see Lonberg et al. (1981); Figure 6].

In the absence of an independent value of n, the simple one-parameter determination of K from a noncooperative titration curve becomes a three-parameter problem, i.e., to determine n, K, and the true end point. Other than a direct three-parameter fit, there is no easy solution to this situation. We have had some limited success by using a "bootstrap" approach whereby the experimental end point is *assumed* to represent some value of θ less than 1, and the curve is then fit to values of n and K. These first approximations to n and K are then used to calculate the actual value of θ at the end point, and this value is used rather than the initially assumed θ value to redetermine a "second-order" fit of n and K, and so forth. This procedure is used iteratively until n, K, and θ no longer change (see Appendix).

The Cooperative Case. When binding is highly cooperative (i.e., titration curves are noticeably sigmoidal, which indicates

[2] In principle, it is possible to determine n, K, and ω simultaneously from a three-parameter fit of the experimental data to eq 1 and 2. While the titration curve itself should contain sufficient information to define these parameters uniquely, we have found that in practice an unambiguous (uncorrelated) fit generally cannot be obtained, due to the experimental uncertainty of the data. Thus, we generally determine the site size (n) first and then fit the experimental data to trial values of K and ω. In this way, only a two-parameter fit is required, which (as will be demonstrated below) can often be reduced to a one-parameter fit under special circumstances.

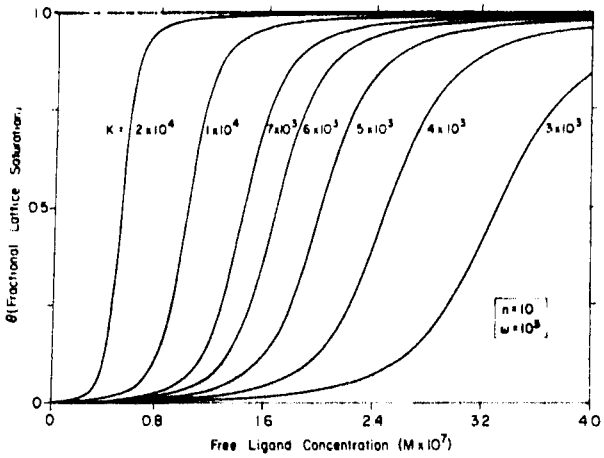


FIGURE 4: Fractional lattice saturation vs. free ligand concentration for a cooperatively binding ligand with $\omega = 10^3$, $n = 10$, and K varying as indicated.

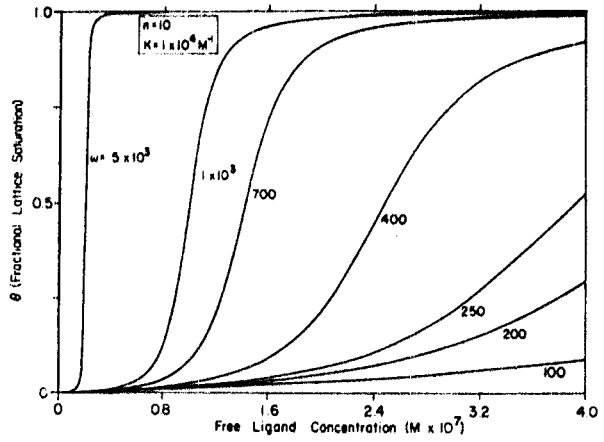


FIGURE 5: Same as Figure 4 except that K is fixed at 10^4 M^{-1} and ω varies as indicated.

that $\omega \geq 10n$), the determination of n, K, and ω is relatively straightforward. First, n is determined from a stoichiometric titration curve as described previously. In the absence of such data, a good estimate of n can be obtained from the region of the binding curve that is steeply increasing. This portion of the curve basically represents the amount of protein needed to titrate the lattice. As can be seen in Figure 2b, that binding (for the most cooperative case) begins at $\sim 1 \times 10^{-7}$ M ligand and reaches saturation at 2×10^{-7} M. Since the total lattice concentration in these model plots was 1×10^{-6} M, with a site size (n) of 10, this method of estimating bound ligand is clearly quite good for highly cooperative titrations. It becomes less accurate with smaller values of ω, but as can be seen in Figure 2b, it is still $\sim$50% accurate with values of ω as low as 10^3.

As binding conditions become more nonstoichiometric (e.g., as the salt concentration is raised), three different classes of changes may occur: (i) K may decrease with ω remaining constant; (ii) ω may decrease with K remaining constant; (iii) both K and ω may decrease. Class i is the most common for systems so far studied and has been observed with T4 gene 32 protein (Kowalczykowski et al., 1981b; Newport et al., 1981), fd gene 5 protein (Alma et al., 1983; Porschke & Rauh, 1983), and *Escherichia coli* recA protein (Menetski & Kowalczykowski, 1985); qualitatively, a family of curves such as that of Figure 4 is observed. Notice that the shape of each curve is about the same, but that they are shifted to the right as K decreases. For class ii, behavior such as that of Figure 5 is observed. Here, each curve displays a "lag" region, but the slope of the steep portion of each decreases as ω is decreased. The observed behavior for class iii will be a mixture of classes i and ii; i.e., the curves shift to the right *and* the slopes decrease. The actual extent of each effect will depend on which parameter is changing more rapidly with the independent variable that is being manipulated.

In order to extract values of K and ω from curves such as these of Figures 4 and 5, we take advantage of the fact (see above) that when ω is much larger than n, the value of the product of K and ω is defined by the value of $1/L$ (the reciprocal of the free ligand titration) at the midpoint of a nonstoichiometric titration (see Figure 2a). Note that if the binding isotherm is plotted in terms of *total* (rather than *free*) ligand concentration, then $K\omega$ is related to the total ligand concentration at the midpoint, $L_{T,1/2}$, by

$$1/(K\omega) = L_{T,1/2} - N/(2n) \qquad (6)$$

This equation is obtained by rearranging eq 3 and setting $\theta = {}^1/_2$; the second term on the right in eq 6 is simply the concentration of ligand that is bound at the midpoint of the titration. Once the product $K\omega$ is known, separate values of K and ω can be determined by generating theoretical titration curves, with eq 2, in which K and ω are varied individually.

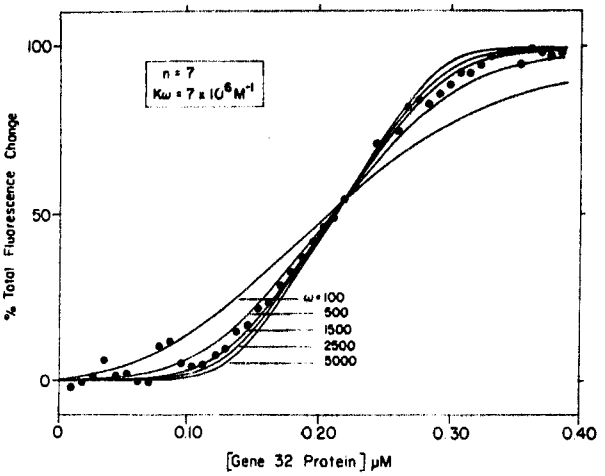


FIGURE 6: Titration of poly(riboethenoadenylic acid) (10^{-6} M residues) with gene 32 protein in buffer B [see Kowalczykowski et al. (1981b)] containing 0.45 M NaCl at 25 °C. The solid lines are theoretical curves for various values of ω (calculated with $K\omega$ held constant at 7.0×10^6 M^{-1} and $n = 7$).

An example of this procedure is shown in Figure 6, with data on the binding of T4 gene 32 protein to poly[ribo(etheno)adenylic acid]. The solid lines in Figure 6 represent theoretical curves that have been obtained by incrementing values of ν in eq 2 and solving for L in terms of given values of n, K, and ω. Note that the data points (unlike the theoretical curves) are not perfectly symmetrical about the midpoint of the plot. Thus, one model curve cannot be used to fit both the lower and upper portions of the experimental data. This follows because the polynucleotide lattices are not really of infinite length (as assumed in the binding theory), so that "end effects" and lattice-length heterogeneity effects become important when the lattice is more than half-saturated. More will be said of finite lattices below, but here we emphasize that if the data are *not* symmetrical about the midpoint of the titration, the lower "half" of the titration should be emphasized in fitting procedures to minimize such problems.

From a comparison of the data and model curves in Figure 6, one can see that values of $\omega < 500$ produce binding isotherms that are too shallow to fit the data, while binding

isotherms corresponding to values of ω greater than 1×10^4 are too steep. Thus, a range of fits can be readily established, with the "best" fit obtained for this system at $\omega = 1.5 \times 10^3$. We note that this procedure loses resolution when values of ω become very large; under these conditions, only a minimum estimate of ω is obtained.

The raw titration data can be replotted in terms of free ligand concentration by simply subtracting the amount of ligand bound at each point from the total ligand concentration, with eq 3. This has the advantage of making the slope of the steeply rising portion of the observed titration curve independent of polynucleotide concentration (see Figure 4; Kowalczykowski et al., 1981b), and therefore, data sets obtained at different concentrations of lattice can be presented on the same plot. Either method of graphical presentation of the data is suitable for analysis.

Instances where the cooperativity parameter is only on the order of 1–10-fold greater than n are the most difficult to fit accurately, because eq 6 does not apply. Here, at least a two-parameter fit is required, even if the site size is independently known. Although eq 6 may not provide an accurate approximation of the product $K\omega$, as seen in Figure 2a for the curves representing values of ω of 1 and 10, it may still be useful in obtaining a starting point for further numerical approximation. If the data are quite good, a two-parameter fit will be reasonably accurate, although difficulty in assigning unique values of K and ω may be encountered because the two parameters are correlated. An independent measure of the binding constant can be obtained by performing binding measurements at low ligand saturation densities (see below).

Titration of Ligand with Lattice

Another relatively common way of performing a titration—often made necessary by solubility or stability limits of the protein in the absence of nucleic acid—is to add increasing amounts of lattice to a fixed concentration of ligand. Again, spectroscopic changes in ligand properties can be used to measure interaction, but in such titrations the data are usually plotted in terms of fractional saturation of *ligand* (instead of lattice as in the previous section), since here a fixed amount of ligand is being titrated.

As an example, the binding of polynucleotide by gene 32 protein results in quenching of the intrinsic fluorescence of the protein, and thus, this signal can be used to monitor the extent of ligand binding. Although this is an experimentally straightforward procedure, the analysis of the results obtained from such titrations for cooperatively binding systems is *not* straightforward and can easily result in significant misinterpretations. The reason for this is qualitatively that while the free ligand concentration continuously increases in a ligand-to-lattice titration, in a lattice-to-ligand titration one begins with a great excess of ligand and ends with a great excess of lattice. This means, for highly cooperative systems, that as lattice is added and free ligand concentration in the titration cell is depleted a point will be reached at which the concentration of ligand that remains free is only $\simeq 1/(K\omega)$. As more lattice is added beyond this point, additional ligand will only bind noncontiguously (i.e., with a net affinity of K rather than of $K\omega$; Figure 1). As a result, an apparent plateau in the titration will be reached at a level that will depend on the total ligand concentration used in the titration. This plateau will *not* correspond to saturation.

To illustrate this effect quantitatively, in Figure 7 we present the results of some model calculations that are based on eq 1 and 2 and use concentration values that might be encountered in a typical fluorescence titration. In order to present

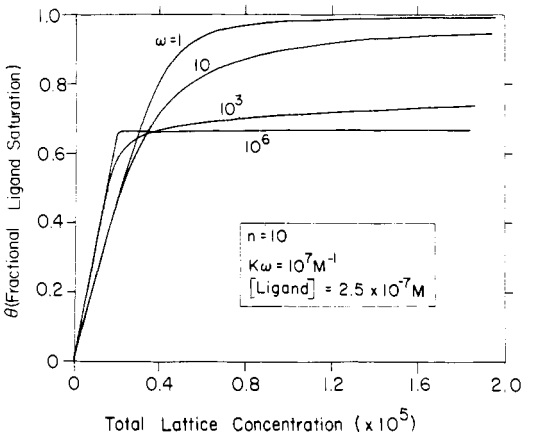


FIGURE 7: Fractional *ligand* saturation vs. total lattice concentration for various values of K and ω; the product of K and ω was held constant at 10^7 M^{-1}. The values of ω are indicated in the figure. The site size is 10, and the concentration of total ligand is 2.5×10^{-7} M.

curves with a wide range of ω values in this figure, the *net* affinity ($K\omega$) is held constant at 10^7 M^{-1}, but the individual values of K and ω are varied as in Figure 2. It is apparent that although the same values for n, K, and ω were used as in the ligand-to-lattice titration of Figure 2, the curves in Figure 7 bear no resemblance (except for $\omega = 1$) to those of Figure 2. Note that as the value of ω is increased (and K correspondingly decreased) two effects on the titration curves are observed: (i) the initial regions of the curve become sharper, as expected; (ii) the curves seem to approach a plateau value that corresponds to less than 100% ligand saturation. Although the more cooperative curves *appear* to be saturating at a plateau value representing less than 100% binding, this "plateau" is, in fact, gradually increasing and eventually approaches 100% saturation at very high concentrations of lattice (data not shown). However, this "real saturation" region may be experimentally inaccessible and thus may never be reached.

As pointed out above, the basis of this apparent "plateau effect" in cooperative titrations is easily understood when one realizes that when lattice is added to ligand, titrations such as those in Figure 2 are being performed "in reverse". That is, the binding density of the lattice (ν) is being *decreased* as lattice is added, so that in Figure 2 one is proceeding along the binding isotherm from right to left. When the free ligand concentration is depleted to the point where there is little ligand binding in Figure 2a [leftmost side of the isotherms where the concentration of free protein is less than $1/(K\omega)$], an apparent end point is reached. This reasoning suggests that the percentage bound at the "apparent" plateau can be increased simply by increasing the input (total) ligand concentration, so that when the apparent end point at a ligand concentration of $1/(K\omega)$ is reached, the bound ligand represents a smaller percentage of the total ligand concentration.

Figure 8 shows the effect of varying the total concentration of ligand on the level of the apparent "plateau". All the parameters of Figure 8 are identical with those of Figure 7, except that the initial protein concentration is varied. Comparison of these figures shows that increasing the input concentration of ligand also increases the apparent plateau level for these cooperatively binding systems. The curves begin to approach the true value of saturation (100%) only when the input ligand concentration exceeds the value of $1/(K\omega)$ by approximately 10-fold. Note in Figure 8 that these effects are not seen for a noncooperatively binding ligand or for one that binds moderately cooperatively ($\omega \approx n$); this can be understood by observing that when $\omega = 1$–10, there is no distinct

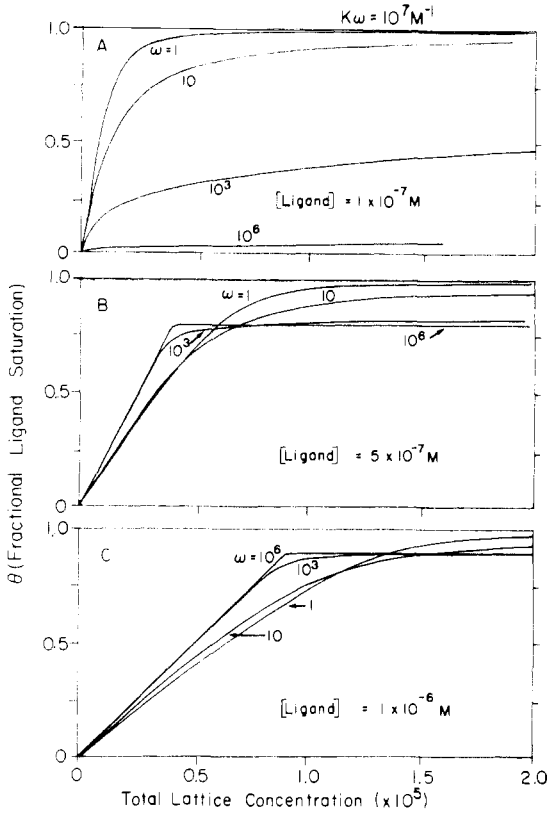


FIGURE 8: Same as Figure 7 except the concentration of total ligand is different in each figure: (A) $L_T = 10^{-7}$ M; (B) $L_T = 5 \times 10^{-7}$ M; (C) $L_T = 10^{-6}$ M.

region of ligand concentration (e.g., see Figure 2) where effectively no ligand binds.

A consequence of the phenomenon described above is that it is potentially very easy to misinterpret the results of a titration in which lattice is being added to ligand. An estimate of the site size, for example, based on the apparent end point, represents an underestimate of the true site size and, in addition, depends on the initial concentration of ligand used. If the experiment were a fluorescence titration in which the ligand fluorescence is quenched, the plateau value could correspond to an "apparent" value of the maximum quenching (Q_{max}) that represents a (possibly major) underestimate of the true Q_{max}. By using the apparent (incorrect) value of Q_{max}, any quantitative analysis of an experimental titration curve would be meaningless and would yield incorrect binding parameters. Finally, if one were only drawing qualitative conclusions from such titration curves, their shapes might suggest that the system contains two types of binding sites with differing affinities, one with high affinity, corresponding to the steep region in Figures 7 and 8, and the other with low affinity, corresponding to the gradually increasing plateau region.

In Figure 9 the effects of initial ligand concentration on the fractional saturation of the ligand, for a very highly cooperative system ($\omega = 10^7$), are shown to illustrate more clearly the unique properties of these types of cooperative titration curves. Each line represents a calculated titration curve performed at a different initial ligand concentration. Because we have made ω a very large value, the apparent plateaus seem very flat, but each curve actually does approach 100% saturation very slowly. On the basis of the previous discussion, the level of apparent saturation can be easily calculated as follows: since $1/(K\omega) = 1 \times 10^{-7}$ M, this concentration of ligand will remain *unbound* at these lattice concentrations. Therefore, if the initial ligand concentration is 1×10^{-6}, then only $\approx(1 \times 10^{-7})/(1 \times 10^{-6}) = 90\%$ of the total ligand will be bound to the lattice, which is the value of the apparent saturation plateau seen in Figure 9; at 2.5×10^{-7} initial ligand concentration, $(1 \times 10^{-7})/(2.5 \times 10^{-7}) = 60\%$ will bind, etc.

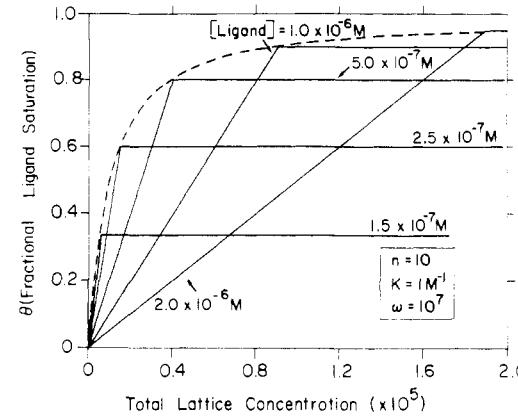


FIGURE 9: Effect of varying ligand concentration in a titration of ligand by lattice for a highly cooperative system; $n = 10$, $K = 1$ M^{-1}, and $\omega = 10^7$. The concentration of ligand is indicated in the figure. The breakpoints of the titration curve lie on a curve (dashed line) defined by the equation (fraction ligand bound) = $[N]/[N] + n/(K\omega)]$, where [N] is the total lattice concentration.

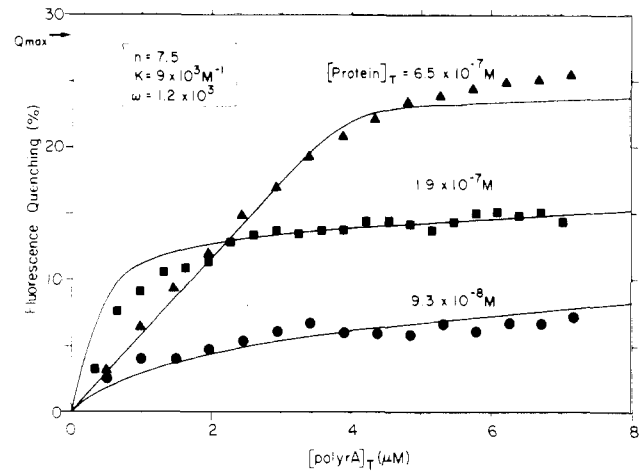


FIGURE 10: Titration of gene 32 protein with poly(rA) for different concentrations of protein; percent fluorescence quenching of the protein is plotted against total poly(rA) concentration. These nonstoichiometric titrations were carried out in buffer C (Kowalczykowski et al., 1981b) containing 220 mM NaCl at 25 °C. Concentrations of gene 32 protein were 9.26×10^{-8}, 1.85×10^{-7}, and 6.48×10^{-7} M; the fitted curves were generated from eq 2 with parameters $n = 7.5$, $K = 9 \times 10^3$ M^{-1}, and $\omega = 1.2 \times 10^3$ by using a value of $Q_{max} = 28.5\%$ that represents 100% saturation of the ligand by poly(rA).

To illustrate that this type of behavior is actually observed in a "real" cooperative binding system, experimental titrations are presented in Figure 10 in which the quenching of the intrinsic protein fluorescence of gene 32 protein upon nucleic acid binding is monitored as poly(rA) is added to fixed concentrations of gene 32 protein. By changing the initial total concentration of gene 32 protein in the cuvette, curves with different apparent end points and different apparent values of Q_{max} are obtained (as expected from Figures 8 and 9). The true values of n and Q_{max} were determined by performing titrations under stoichiometric conditions at low salt and were found to be 7.5 nucleotide residues and 28.5% (of the uncomplexed protein fluorescence), respectively. From comparison of the titration data with model curves, K was determined to be 9×10^3 M^{-1} and ω to be 1.2×10^3. These values of K and ω are in good agreement with those obtained previously with ligand-to-lattice titration procedures (Jensen

et al., 1976; Kowalczykowski et al., 1981b; Newport et al., 1981).

Thus, the proper analysis of this type of cooperative titration data requires that one know whether the observed plateau represents a true or an apparent end point; this is particularly important when indirect methods (such spectroscopic signals) are used to detect complex formation.

The easiest way to determine whether a true end point has been reached in a lattice-to-ligand titration is to vary the initial *ligand* concentration (as in Figure 9) in order to see whether the apparent end-point plateau value is constant (when normalized for the difference in ligand concentrations) and whether the apparent site size is unchanged. If the value of n is invariant in each titration, then conditions are stoichiometric and the value of n is defined. If the apparent plateau values are invariant, but the experimentally determined values of n are *not* constant, then ligand is binding noncooperatively but binding conditions are not stoichiometric. Finally, if neither n nor the plateau value is invariant with initial ligand concentration, then binding is probably cooperative and results such as those of Figure 9 are expected.

Once the value of n is known, the experimental titration curves can be analyzed to yield values of K and ω by an approach similar to that described for ligand-to-lattice titrations. If the system is noncooperative, the value of K is varied until a good fit is obtained. If the system is highly cooperative ($\omega > 10^3$), then one can take advantage of the fact that at the apparent "breakpoint" of a nonstoichiometric lattice-to-ligand titration the concentration of *free ligand* (not lattice) is approximately equal to $1/(K\omega)$. Then, as was done previously, model curves can be generated with a fixed value of $K\omega$, but reciprocally changing values of K and ω, until an adequate fit to the data is achieved. As with the ligand-to-lattice titration data, the experimental points may deviate from the calculated curves due to the fact that the experimental lattices are of finite size. However, in lattice-to-ligand titrations this deviation will manifest itself mostly at low lattice concentrations (the left-hand side of plots such as Figure 10). This is the region of the curve where the binding density is the greatest and where finite lattice effects are most pronounced (see below). Unfortunately, this is also the region of the titration curve that carries the most experimental "information". Thus, lattice-to-ligand titrations will tend to yield underestimated values of ω.

Experiments at Very Low Binding Densities

In addition to the titration methods that we have discussed, one can obtain values of K and ω by any other method that allows the determination of the free ligand concentration. The approach to be described in this section takes advantage of the fact that at very low binding densities the amount of ligand bound to the lattice depends *only* on the value of K and is independent of ω. An additional benefit of working at such low binding densities is that the "overlap" problem becomes insignificant, so that the amount bound will be independent of the site size, n. This is useful if stoichiometric binding conditions cannot be found for the system of interest.

This effect is illustrated in Figure 11, in which we plot eq 2 with K held constant and calculate free ligand concentrations (L) as a function of ν for different values of ω. Note that at low binding densities (i.e., $\nu \leq 10^{-5}$) all of the curves converge to a common value. In this region the free ligand concentration is independent of ω, and from Figure 11 we see that this situation applies for binding densities (ν) less than $1/(10\omega)$.

The fact that the curves for constant values of K with variable values of ω are superimposable at very low values of

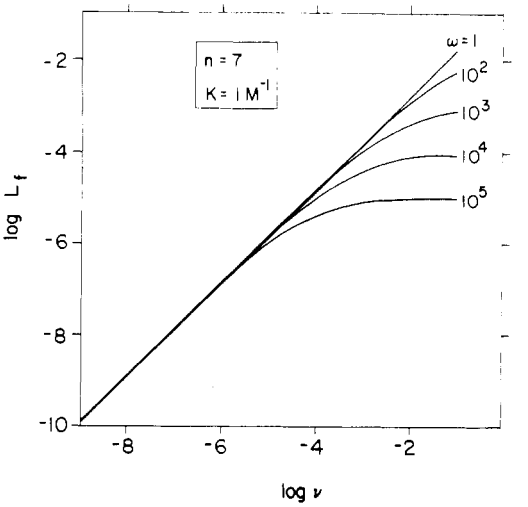


FIGURE 11: Logarithm of the ligand concentration (L_f) vs. logarithm of binding density (ν), for values of $K = 1\ M^{-1}$ and $n = 7$.

ν allows one to determine a value of K alone, even for highly cooperative systems. By evaluating eq 2 in the limit as $\nu \rightarrow 0$, one obtains

$$\lim_{\nu \to 0} \nu/L = K \tag{7}$$

Thus, at very low values of ν ($\nu < 0.1/\omega$), ω and n drop out of eq 2, and L depends only on K.

This approach can be exploited by any method in which low binding densities can be attained and free ligand concentration determined. For studies with T4 gene 32 protein, a DNA–cellulose column technique (deHaseth et al., 1977) was used, in which the concentration of bound DNA was approximately 10^{-2} M (in nucleotide residues) and free protein concentrations were determined by intrinsic protein fluorescence, allowing values of ν as low as 10^{-4} to be attained [see Kowalczykowski et al. (1981b) for details]. In addition, any other technique that enables one to work at high lattice-to-ligand ratios will be applicable [e.g., sedimentation (Draper & von Hippel, 1979; Yamamoto & Alberts, 1974) or ultrafiltration], coupled with any sensitive method for detecting the free ligand concentration, such as radioactive labeling, enzyme assay, or immunological detection; the lower limits of ν and L attainable will depend only on the method used.

Since eq 7 allows us to determine K independently of the effects of cooperativity, ω cannot be determined directly from these low binding density experiments. However, the results can be coupled with an independent measurement of $K\omega$ to determine ω as well. A method for determining $K\omega$ that differs from any described in this paper has been presented elsewhere (Kowalczykowski et al., 1981b; Newport et al., 1981) and referred to as the "salt back-titration" method (see references for details). Briefly, protein–nucleic acid complexes are formed and then titrated with NaCl to dissociate the complex. At salt concentrations at which half of the initially bound ligand is dissociated, $K\omega = 1/L$ (if $n \ll \omega$). Since half of the ligand is bound and half is free, $L = (1/2)L_T$, resulting in $K\omega = 2/L_T$, which defines the value of $K\omega$ at that salt concentration. Separate values of K and ω cannot be readily determined by the salt back-titration method, but if K is measured at a particular salt concentration with the low binding density method, then ω is also known.

This combined approach can also be used directly to partition the salt dependence of the $K\omega$ product between K and ω. This is shown for gene 32 protein binding to single-stranded DNA in Figure 12, in which we plot the free protein concentration data points as determined by the DNA–cellulose

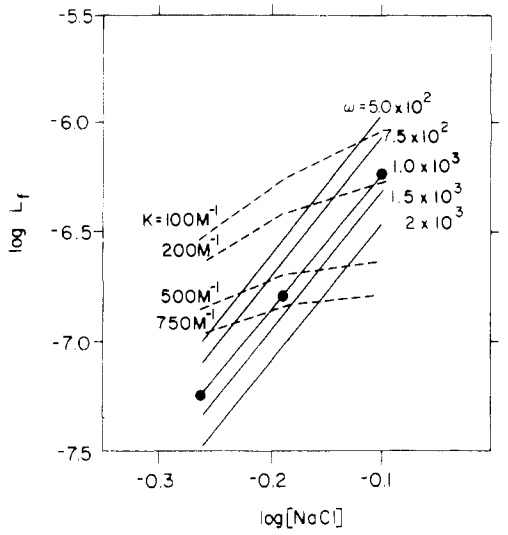


FIGURE 12: Logarithm of free ligand concentration vs. logarithm of salt concentration. The data points were obtained by the DNA–cellulose column method with gene 32 protein as a function of log [NaCl]. [DNA] = 1.3×10^{-2} M, and $\nu = 1.3 \times 10^{-4}$ [see Kowalczykowski et al. (1981b) for details]. The theoretical lines were calculated with eq 2 and values of $K\omega$ determined previously for ϕX174 DNA (Newport et al., 1981). The dashed lines are for fixed values of K, with ω changing with [NaCl] as indicated. The solid lines are for fixed values of ω, with K changing with [NaCl] as indicated.

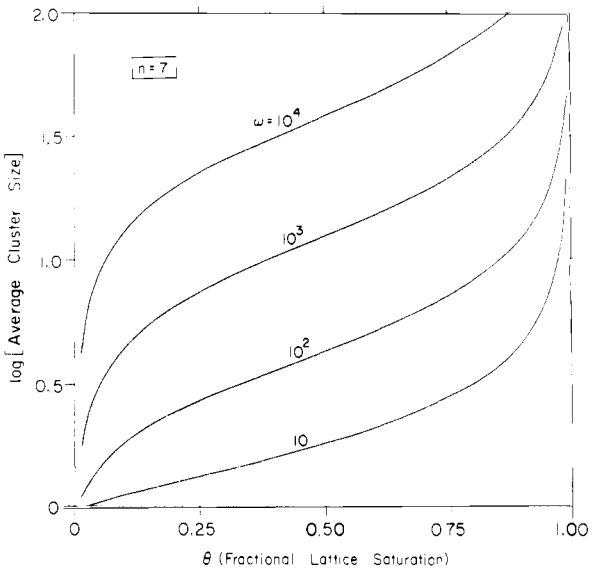


FIGURE 13: Logarithm of the average cluster size of bound ligands vs. fractional lattice saturation for the values of ω indicated and $n = 7$.

column method at various salt concentrations. The salt back-titration method was used to determine $K\omega$ at each salt concentration (data not shown), and this value of $K\omega$ was inserted into eq 2 to calculate the expected values of ν at the various salt concentrations for various trial values of K and ω (see legend). Since $K\omega$ is very salt dependent [$\delta \log (K\omega)/\delta \log [NaCl] = -6.3$; Newport et al., 1981], the results presented in Figure 12 can be used to demonstrate that the salt dependence of $K\omega$ is entirely attributable to K and not to ω, which stays constant at 10^3 as the salt concentration is varied. This follows because if K *were* salt-independent, L would not change with salt concentration in these low binding density experiments and the data points would fall on one of the dashed lines. Instead, we see that the experimental points fall on one of the solid lines, corresponding to a value of ω that is independent of salt concentration.

Cluster Size Distributions

Another way of determining the cooperativity of binding is to take advantage of the facts that ligands bind to lattices in "clusters" containing one or more contiguously bound ligands and that the size distribution of these clusters depends on the values of n, K, and ω at which the complexes are formed. Thus, n, K, and ω can be determined by measuring the average cluster size or, ideally, the cluster size distribution of the binding system. A method that is, in principle, ideally suited to such analysis is electron microscopy, where each cluster can be "frozen" and sized through direct visualization. This method has been used to determine an estimate of ω for the binding of *E. coli* SSB protein to single-stranded DNA (Ruyechan & Wetmur, 1975) and to estimate ω for *E. coli* transcription termination protein rho binding to RNA lattices (D. G. Bear et al., unpublished results). Alternatively, cluster sizes of proteins bound to nucleic acids can also be determined (in situations where protein dissociation is slow or binding equlibria can be frozen—perhaps by rapid cross-linking techniques) by utilizing a method such as the partial nucleolytic digestion of uncomplexed DNA, followed by a determination of the size (average or distribution) of the protein-protected regions of the DNA.

In order to use such methods to determine binding parameters, the theoretical distribution of cluster sizes must first be known, and for this, we again use the McGhee–von Hippel formalism [see Schellman (1974) for an alternate approach]. With the McGhee–von Hippel (1974) approach, the average cluster size ($\bar{C}$) is simply the average number of ligands bound per lattice (B) divided by the average number of bound ligands that are free on the right of these ligands [$B - B(b_n b_1)$]. The term $b_n b_1$ is the conditional probability that a given bound ligand will be located next to another bound ligand on the right (and so will *not* represent the end of a cluster); $b_n b_1$ is thus defined as $[1 - (n - 2\omega + 1)\nu - R]/[2\nu(\omega - 1)]$ [see McGhee & von Hippel (1974)]. Thus

$$\bar{C} = \frac{1}{1 - b_n b_1} = \frac{2\nu(\omega - 1)}{(n - 1)\nu - 1 + R} \qquad (8)$$

(see eq 2 for definition of variables). Equation 8 shows that the average cluster size is dependent on n, ω, and ν (which is dependent on K). In Figure 13, we have plotted $\bar{C}$ as a function of the fractional lattice saturation, θ ($=n\nu$), for various values of ω [see Schellman (1974) for a similar plot]. As expected intuitively, the average cluster size at a given level of lattice saturation is larger for greater values of the cooperativity parameter; i.e., cooperatively binding ligands tend to cluster more. In addition, as the binding density increases the average size of the bound clusters increases, approaching infinity as the fractional saturation approaches 100%. Because real lattices are of finite length, Figure 13 suggests that experimental deviations from the theory will be observed when the lattice length becomes less than the calculated cluster size and that this deviation will become pronounced at lower binding densities for ligands that bind with higher cooperativity. Figure 13 also shows that as $\theta \rightarrow 0$ all cluster sizes approach unity and that this limiting value is approached more rapidly for less cooperative systems (see also the previous section on low binding density experiments).

A more informative approach is to calculate the actual size distribution of bound clusters (i.e., the fraction of all of the ligands bound that are in a given sized cluster) as a function of the fractional saturation of the lattice. The fraction of bound ligand (F_c) in clusters of size c can be calculated from

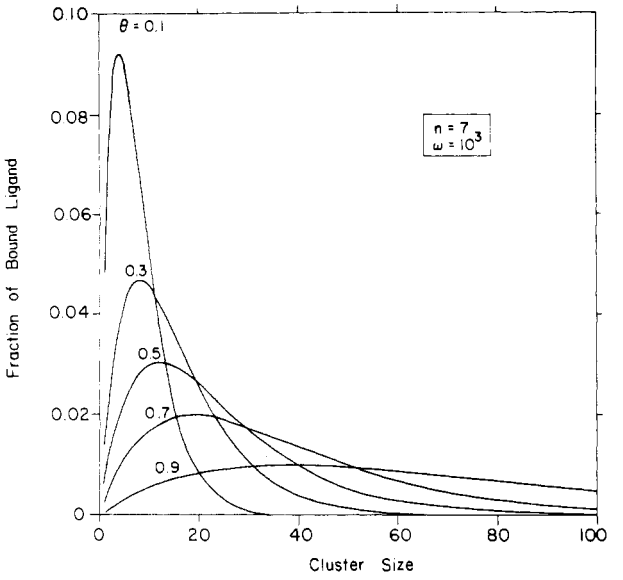


FIGURE 14: Fraction of bound ligand at a given cluster size vs. cluster size for various values of fractional lattice saturation, with $n = 7$ and $\omega = 10^3$.

the conditional probabilities defined in eq 13 of McGhee and von Hippel (1974). The probability of finding a cluster of c ligands next to a given free lattice position is

$$P_c' = ff \qquad c = 0$$

$$P_c' = (fb_1)(b_nb_1)^{c-1}(bnf) \qquad c > 0$$

If only clusters of size 1 or greater are considered, then the above probability must be normalized by $1 - P_0 = 1 - ff$ so that it will sum to 1:

$$P_c = \frac{P_c'}{1 - ff} = \frac{(fb_1)(b_nb_1)^{c-1}(1 - b_nb_1)}{fb_1} = (b_nb_1)^{c-1}(1 - b_nb_1)$$

where eq 2 and 3 of McGhee and von Hippel (1974) are used to simplify the expression. Finally, F_c can be obtained by multiplying P_c by the cluster size and normalizing by the sum over all cluster sizes:

$$F_c = \frac{cP_c}{\sum_{c=1}^{\infty} cP_c} = c(1 - b_nb_1)^2(b_nb_1)^{c-1} \qquad (9)$$

where the summation is seen to be simply the average cluster size as defined in eq 8.

In Figure 14, we have plotted the cluster size distribution predicted by eq 9 for different values of fractional saturation for a ligand binding with a cooperativity value of $\omega = 10^3$ (e.g., T4-coded gene 32 protein) [see Lohman (1983) for a similar plot]. As expected, the distribution of bound ligands shifts to progressively larger sized clusters as the fractional saturation increases. Also note that though ligand is bound in clusters of all possible sizes (we have truncated the graph at $c = 100$, but it actually continues to infinite cluster length), at a fractional saturation of 10% very little ligand is bound in clusters greater than 100 ligands in length, while at a fractional saturation of 90%, many ligands would be expected to be in clusters of length greater than 100. As mentioned previously, experimental observations will deviate from theoretical behavior due to loss of the potential cooperative interactions that are unable to form at the ends of the finite-length lattices. Thus, with gene 32 protein as an example, if the actual length of the polynucleotide lattice is 400 residues, the maximum-sized cluster attainable would be approximately 60 protein molecules in length. Consequently, significant deviations from ideal behavior would be expected when a large proportion of the bound molecules occurs in clusters greater than this length (i.e., at fractional saturations greater than 70%). This is, in fact, observed in experimental titration curves (see Figure 6) and is the basis of the nonsymmetrical appearance of the experimental data and the reason that we emphasize the fit to the lower half of the titration curve in the analysis of ligand-to-lattice titration curves.

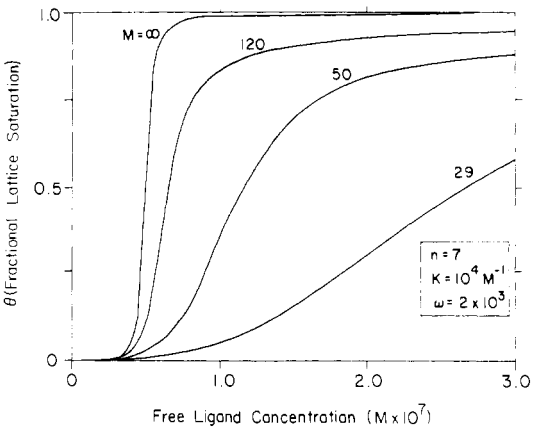


FIGURE 15: Theoretical (Epstein, 1978) plots of fractional saturation of lattice vs. free ligand concentration for various values of lattice length (M), with $n = 7$, $K = 10^4$ M^{-1}, and $\omega = 2 \times 10^3$.

As indicated above, the cluster size distribution approach to determining K and ω is ideally suited to an experimental method where the actual (equilibrium) number and size of bound ligand clusters can be counted (e.g., by electron microscopy). To use this approach, a histogram of cluster sizes is determined at various values of fractional saturation of the lattice and then compared with theoretical curves generated by eq 9.

While this approach is very straightforward, practical considerations may intervene. Difficulties will be encountered if the method of sample preparation does not ensure the "freezing" of the original cluster equilibrium. Obviously, the migration of ligands during sample preparation can seriously perturb the cluster size distribution obtained. Careful controls should eliminate errors introduced by nonequilibrium effects, or at least allow extrapolation to the true (initial) equilibrium distribution.

Effects of Finite-Lattice Length

Up to this point, the discussion has implicitly assumed that we always deal with a lattice of infinite length, since the derivation of eq 1 and 2 requires this. However, experimentally all lattices are of a finite length and usually exhibit a heterogeneous length distribution. We must then ask, at what length does a lattice become essentially infinite, and if a lattice is too short to be considered infinite, what effect does this have on the experimental results? The answer to the first question has been addressed in the previous sections, and from Figure 13 we can see that the point at which a lattice can be considered to be of "infinite" length will depend on the cooperativity of binding, the ligand site size, and the extent of lattice saturation; as each of these factors increases, experimentally longer lattices will be needed in order to utilize the approaches described previously without modification.

If the experimental situation does require the use of a lattice that is of "finite" length, a mathematical treatment that has been developed by Epstein (1978) can be employed. In Figure 15 we have used this method to generate theoretical curves that demonstrate the effects of finite lattice length on titration

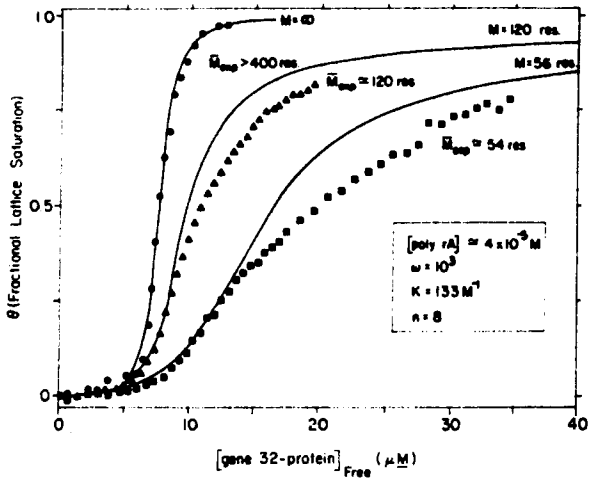


FIGURE 16: Titrations of poly(rA) of lengths (M_{exp}) indicated with gene 32 protein. The titrations were performed as described in Kowalczykowski et al. (1981b). The solid lines for $M = 56$ and 120 represent the theoretical titration curves expected from the finite-lattice theory of Epstein (1978) and values of $K = 133\ M^{-1}$, $\omega = 1 \times 10^3$, $n = 8$, and M as indicated. The solid line for $M = \infty$ was calculated with eq 2 and identical values of n, K, and ω. The titrations were performed in buffer C (Kowalczykowski et al., 1981b) containing 0.5 M NaCl and polynucleotide concentrations of $[poly(rA)_\infty] = 36.2\ \mu M$, $[poly(rA)_{120}] = 40.2\ \mu M$, and $[poly(rA)_{54}] = 47.6\ \mu M$.

curves. Here, the values of n, K, and ω are held constant, while the length of the lattice (M, in units of nucleotide residues) is varied. The effects of decreasing the lattice length are readily seen: (i) the apparent affinity of the ligand decreases with lattice size; (ii) the steepness of the cooperative transition decreases; (iii) complete saturation becomes more difficult to achieve; (iv) the deviation from the infinite length curve increases as the binding density increases. Most of these effects are attributable to the fact that with smaller lattices a greater percentage of cooperative interactions are lost due to the greater ratio of "ends" to "middles" of the lattice. This effect becomes more serious for larger values of ω, since the loss of cooperative interactions at the ends of the lattice will have a larger effect on the binding interaction. There is also some effect due to the loss of (overlapping) binding sites as the lattices become smaller. See Epstein (1978) for further discussion of these points and von Hippel et al. (1982) for a practical application to a biological control system.

An experimental demonstration of these finite lattice effects is presented in Figure 16. Three size fractions of poly(rA) were titrated with gene 32 protein under identical conditions. The solid lines represent theoretical plots with the Epstein equations for a lattice of the indicated length. The agreement with theory is good, though not perfect; the deviations at larger values of fractional saturation for the smaller lattices are due to the fact that the lengths of the *lattices* are also heterogeneous, with a weight-average size as indicated. The presence of lattices of smaller than average length will make the mixture more difficult to saturate and, hence, will broaden the titration near the end point and cause the experimental data to deviate in the observed manner. As mentioned in the preceding discussion of theoretical curves, the presence of finite lattices introduces a larger degree of "roundedness" in the titration curves in the region of fractional saturation greater than 0.5. Furthermore, this effect is most pronounced for cooperatively binding ligands. As seen in Figure 16, the experimental data in the upper region of the titration curves do not fit the theoretical curve that passes through the data in the lower half of the titration. For this reason, greater weight in fitting is placed on the data points at lower levels of saturation.

These problems with finite lattices become most severe when data are interpreted in which titrations are nearly stoichiometric. In this case, the data will appear to be "rounded" near the equivalence point, instead of showing a distinct break. If one interprets this curvature in the data as arising from a binding constant that is slightly smaller than needed to bring about totally stoichiometric binding, then the real binding constant may be grossly underestimated. Kelly et al. (1976) presented a simple formula for estimating the apparent binding constant (K_{app}) from a nearly stoichiometric titration, on the basis of the apparent roundedness of the titration close to the stoichiometric point. They also made it very clear that K_{app} represents a *lower limit estimate* for the true binding affinity, $K\omega$, and that the method assumes "infinite" lattices of (by definition) homogeneous length. A number of authors have used the Kelly et al. expression to obtain estimates of protein–nucleic acid binding constants. In some cases, because of finite lattice or lattice size heterogeneity effects, these estimates of $K\omega$ appear to be too low by several orders of magnitude. *Extreme caution* must be used when analyzing titration curves in this manner.

Conclusions

We have presented methods and analyses that enable one to determine separately the values of n, K, and ω that characterize the nonspecific binding of large ligands to a linear array of binding sites (a lattice). By analyzing binding data in terms of the McGhee–von Hippel equations and by making use of several different approaches to the binding measurements, it is possible to determine the values of the thermodynamic parameters for the interaction of proteins with nucleic acid, and we have analyzed, as an example, the interaction of T4-coded gene 32 protein with single-stranded nucleic acids under a variety of conditions. The usefulness of this type of thermodynamic information in understanding the molecular details of the binding of this protein to DNA and RNA (as well as for understanding the biological and functional significance of this binding) is very clear and has been discussed elsewhere [e.g., Kowalczykowski et al. (1981b), Newport et al. (1981), Lonberg et al. (1981), and von Hippel et al. (1982)].

The approaches described in this paper fall into five general categories: (i) titration of lattice with ligand; (ii) titration of ligand with lattice; (iii) experiments at very low binding density; (iv) determination of bound ligand cluster sizes; (v) finite lattice effects. The first two categories of approaches have been most commonly used and are, perhaps, the most generally applicable. They both rely on the monitoring of some signal (e.g., a spectrophotometric, enzymatic, or hydrodynamic change) that can originate in either the ligand or the lattice and is proportional to the amount of complex formed. The other three categories are more specialized and are particularly suited to techniques where very low ligand-to-lattice ratios are attainable (e.g., DNA–cellulose chromatography), where the actual size of bound protein clusters can be determined (e.g., electron microscopy, nuclease protection) or where actual lattice length sizes and distributions are known or can be determined.

The analysis of data from both types of titration methods uses the general scheme that has been developed to eliminate the need for a three-parameter fit of the data. First, the site size, n, is determined from a stoichiometric titration curve. Second, the nonstoichiometric binding data are fitted to a value of K if the binding is noncooperative. If binding is cooperative, values of $K\omega$ are determined, and then K and ω are varied

reciprocally until a fit is obtained. The specific method used to determine $K\omega$, and to fit the data, depends on whether the signal being used to monitor complex formation tracks the addition of lattice to ligand or of ligand to lattice. In either case, the values of n, K, and ω can often be determined without the use of multivariable fitting routines (although see Appendix). The two direct titration methods are the most straightforward approaches to the determination of these thermodynamic parameters. However, due to the apparently unusual behavior of cooperatively binding ligands in the lattice to ligand type of titrations, Figures 7–9 should be studied carefully when such data are analyzed, since the appearance of these titration curves is deceptive and can lead to major misinterpretations (e.g., erroneously small site sizes or the assumed presence of two different types of binding sites).

The strength of the low binding density approach is that it allows the determination of the intrinsic binding constant, K, independently of any knowledge of n and ω, provided that high enough ratios of lattice to ligand can be attained. In addition, this method, when used in conjunction with a procedure such as salt back-titration, allows the separate determinations of K and ω (Kowalczykowski et al., 1981b; Newport et al., 1981).

The cluster size approach is potentially powerful and well suited to the analysis of electron microscopy data or of nuclease digestion experiments on partially protein-protected lattices. Knowledge of the cluster size distribution of the bound ligands provides the necessary information to define K and ω, provided that the bound ligand clusters can be studied in a true (or "frozen") equilibrium distribution; i.e., there must be no loss, gain, or migration of ligand due to the sample preparation methods employed.

Finally, we have shown that the effects of finite-length (i.e., of "real") lattices is to introduce "artifactual" curvature into titration data, particularly at high values of fractional lattice saturation. Such curvature has the effect of making the apparent cooperativity of the system less than that expected for the infinite-length lattice and of making stoichiometric titration data seem to be "nonstoichiometric", potentially resulting in a severe underestimate of the apparent binding affinity of the system. Since the size of the experimental lattice may be intrinsic to the system being investigated, little can sometimes be done to alleviate the problem, and it is best avoided (if the objective is to determine thermodynamic parameters) by using lattices that are as large as possible or by focusing on data obtained at lower levels of binding density.

Although not discussed quantitatively in this paper, another source of difficulty that could be encountered in all the procedures presented arises from another characteristic of "real" experimental lattices, i.e., *compositional* heterogeneity of the lattice. Since eq 1 and 2 are strictly applicable only to homogeneous lattices, any heterogeneity will manifest itself as a deviation from theoretical predictions. The extent of the deviation will depend on the actual composition (and/or sequence) of the lattice used, as well as on the differences in values of K and ω that result from these differences in composition. For example, if the lattice is compositionally heterogeneous in a random manner, the values of K and ω will represent approximately "average" values for the entire lattice. One might then expect that values of ω determined on such systems would represent underestimates, due to the broadening of the binding transition because of heterogeneity in K, just as melting transitions of native DNA are broadened relative to transitions involving homopolymer duplexes. At the other extreme, if the lattice consists of large distinct domains of a given composition (much like a block copolymer), the binding of a ligand to this lattice would be approximately equivalent to binding to a mixture of lattices, each of which has the composition of a given domain. If ω were large, then the ligands would tend to bind and cluster to that region of the lattice for which the intrinsic binding constant, K, is the largest and the various regions of the lattice would be saturated in order of increasing K. In any event, the approaches outlined here are still generally applicable; however, the values of K and ω obtained will represent some weighted average that is dependent on the lattice composition and sequence. Such compositional effects have been observed in the binding of gene 32 protein to nucleic acids (Newport et al., 1981).

In summary, our results show that one can apply a theory based on the binding of ligands to homogeneous linear lattices to the binding of proteins to real nucleic acids and proceed to extract thermodynamic information about complex formation. It is important to be aware of the difficulties that can be encountered with the interpretation of such experimental data, particularly when binding is measured by titrating ligand with lattice. In addition, it is necessary to realize that binding theories have usually been developed assuming homogeneous lattices of infinite length, and thus deviations due to the breakdown of this assumption will occur in predictable ways. As mentioned previously, the success of these methods has already been demonstrated with the phage T4-coded gene 32 protein–nucleic acid interaction system, and the results obtained with each method we have described are consistent with one another. The methods presented here should be generally applicable to all proteins that bind nonspecifically to either single- or double-stranded nucleic acids and should be useful in many types of quantitative studies of protein–nucleic acid interactions.

ACKNOWLEDGMENTS

We are very grateful to Jean Parker for typing the many versions of the manuscript and to Mary Gilland for preparing the figures.

APPENDIX

Nonlinear Least-Squares Procedures for Fitting Titration Curves. Titrations can be modeled directly, or they can be converted into Scatchard or related binding plots by linearization techniques [see McGhee & von Hippel (1974)]. In general, we prefer direct modeling, since this minimizes bias in the fit due to transformation of errors in the data. The wide-spread availability of computers and graphics devices allows visual "fits" to the experimental data to be readily performed by most investigators. However, not all titrations are amenable to the visual fitting techniques described in the body of this paper; for these titrations, a numerical fitting approach may be more useful. Such a procedure may be needed if stoichiometric binding conditions cannot be found that can provide an independent estimate of n. Similarly, if binding cooperativity falls in the "intermediate" range ($\omega < 10n$), a visual separation of K and ω may not be possible. Other reasons for turning to a numerical fitting technique might include a desire to process titrations more quickly and with less subjectivity, though we emphasize that some visual fits should be tried first to assess the overall "reasonableness" of the outcome.

The usual approach to finding the best fit of a model curve to observed data is to employ the method of least squares. This method involves finding the minimum with respect to all unknown parameters (i.e., finding $\delta SD/\delta q = 0$ for each parameter q) of the function

$$SD = \sum(Y_i - y_i)^2 \qquad (A1)$$

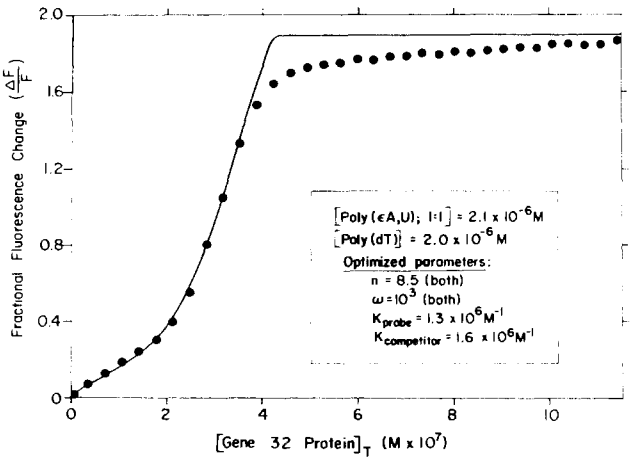


FIGURE A1: Nonlinear least-squares fit for a competition titration of a 1:1 mixture of poly(εA,U) (1:1) and poly(dT) with gene 32 protein in 0.1 M KCl. The concentration of the lattices is indicated in the figure, and the monitored signal is the increase in poly(εA,U) fluorescence [which plateaus at $\Delta F/F = 1.8$; F is the initial value of the fluorescence of the poly(εA,U) lattice]. The optimized parameters obtained for gene 32 protein binding were $n = 8.5$ and $\omega = 10^3$ for both lattices, $K_{probe} = 1.3 \times 10^6$ M^{-1}, and $K_{competitor} = 1.6 \times 10^6$ M^{-1}.

where SD is the square deviation, y_i is the *observed* value of the dependent variable at a given value of the independent variable (x_i), and Y_i is the *predicted* value of the dependent variable at the same point, calculated from the model for a given set of parameters. If the model curve can be expressed in a form that is linear in x_i, the process of minimizing eq A1 is quite simple. Unfortunately, eq 1 and 2 cannot be expressed as linear functions of the independent variable (L); in fact, they cannot even be solved *explicitly* for the dependent variable (ν or θ). Nonlinear least-squares algorithms have been developed that can accomodate both of these limitations, but they also require *explicit* forms of $\delta Y/\delta q$ for each parameter q that is to be fit; these also are not conveniently available from eq 1 and 2.

Our approach has been to develop a simple computer program that searches a restricted portion of "parameter space" for values that give a minimum in eq A1. This brute force approach is made possible by the observation that eq 1 and 2 are both monotonically increasing functions and that the effects of varying the parameters n, K, and ω are thus equally predictable. The process can be divided into two parts: (i) given values for n, K, and ω, calculate the square deviation (SD) between that model and the observed data; (ii) systematically vary the parameters to find the smallest value of SD. This discussion will focus on titrations in which lattice is added to a fixed concentration of ligand, although a similar approach can be used with the reverse type of titrations.

Calculation of SD. The most common types of titration data involve total ligand (L_T) as the independent variable and some signal (S) that is assumed or demonstrated to be directly proportional to fractional saturation (θ) as the dependent variable. Equation A1 then becomes

$$SD = \sum(S_i - s_i)^2 = S_{max}^2\sum\theta_i^2 - 2S_{max}^2\sum\theta_i s_i + \sum s_i^2 \quad (A2)$$

Here, s_i represents the observed signal for each experimental value of L_T, $S_i = \theta_i S_{max}$ represents the calculated signal for that same point, and S_{max} is the maximum signal observed when $\theta = 1$.

Since eq 1 and 2 cannot be solved explicitly for θ, it is necessary to work backward to find θ_i values. The θ axis is divided into equally spaced intervals (usually 50–100), and the value of L_T is calculated from eq 3 for each value of θ. Since these curves are all monotonically increasing, it is then possible to find a unique pair of neighboring L_T values that bracket the observed point. The corresponding θ values will then also bracket the predicted value of θ_i for that point, and θ_i can be estimated by interpolation.

The best fit value of S_{max} can be calculated directly from eq A2 by setting $\delta(SD)/\delta S_{max} = 0$ and solving for S_{max}:

$$S_{max} = \sum\theta_i s_i / \sum\theta_i^2 \quad (A3)$$

Equation A3 can then be substituted back into eq A2 to obtain

$$SD = \sum s_i^2 - \sum(\theta_i s_i)^2/\sum\theta_i^2 = \sum s_i^2 - S_{max}\sum\theta_i s_i \quad (A4)$$

Thus, given the predicted values of the fractional saturation for each data point, one can use eq A4 to calculate SD.

Minimization of SD. Minimization of the square deviation is simply a matter of varying one parameter at a time and checking to see if the new value of SD is smaller than the previous one. If all parameters are treated as unknowns, this could mean a great deal of searching and poor estimates of the unknowns, especially if S_{max} is added to the list of unknowns (as would be the case if conditions that permit complete saturation of the lattice are not attainable). Any information that will help to place boundaries on the parameters should be used in the search. Thus, the techniques described in the body of this paper are applicable here as well.

The following is a list of general guidelines that can limit the search and are applicable to most standard titrations for which lattice concentrations (in nucleotide residues) range from 10^{-5} to 10^{-9} M: (i) Values of $K\omega > 10^{12}$ will give curves that are too sharp to distinguish one from another, while values of $K\omega < 10^4$ will probably not result in measurable binding under standard spectroscopic conditions. Thus, one may hold the search area within these limits. (ii) Because of limitations of the model for systems displaying negative cooperativity, values of site size for $\omega < 1$ will be indistinguishable from a value of n (for $\omega \geq 1$) that is larger than the true site size by something less than one lattice unit (i.e., $n < N_{app} < n + 1$ for $\omega < 1$). Thus, the search can be limited to $\omega \geq 1$. (iii) We have found that varying K or ω by factors of 10 is an efficient way to approach the minimum in SD. The final approach is carried out in steps of $10^{0.1}$, and finally in steps of $10^{0.05}$, to obtain values of K and ω good to $\sim$0.1 log unit. (iv) Because K, ω, and n are so tightly coupled, a great deal of searching can be eliminated by adjusting each variable in rotation until the global minimum is achieved. For example, one first minimizes SD with respect to n while keeping K and ω fixed; then, keep n and ω fixed while making small adjustments to K for a better fit, etc. (v) The tight coupling of n, K, and ω also results in a fairly flat minimization function. That is, changes in one parameter can be compensated by changes in another parameter so that reasonably good fits to a *range* of parameter combinations are possible. For this reason, it is important that any independent information that can be used to constrain some of the parameters be utilized fully. It is also recommended that any fitting routine be set up to maintain a listing of intermediate results in the search procedure. This gives the operator a feeling for how accurately a particular parameter can be specified. (vi) If titrations are carried out under conditions for which finite lattice effects may be a problem, it is useful to limit the initial fitting process to points that are located on the lower binding density portions of the titration curve. Subsequently, such fits can be compared with fits to the entire curve to detect finite lattice problems. In this case, the mean square deviation [msd = SD/(number of data points)] may be a more convenient minimization

function since it permits comparison of fits made with different numbers of data points.

Competition Titrations. In some cases, the ligand and lattice under study may not show intrinsic signals that permit the direct monitoring of their interactions. For such systems it is still possible to obtain binding information through the use of competition titrations. The lattice of interest (i.e., the competitor) is mixed with a "probe" lattice for which the binding *can* be monitored, and ligand is added to the mixture. If the competitor binds ligand poorly compared to the probe (and the two are present in similar amounts), it will appear as though the competitor is not present during the titration. If the competitor lattice binds ligand *as well as* the probe, the resulting titration will appear to contain an amount of probe equal to the sum of probe and competitor concentrations. Finally, if the competitor binds ligand *more strongly* than the probe, the resulting titration will show a lag as each succeeding aliquot of ligand is partitioned preferentially onto the competitor. Thus, by comparing the titration of probe alone with the titration of a mixture of probe and competitor, it is possible to determine binding parameters for the competitor lattice.

We have also modified our computer program so that competition titrations can be fitted numerically. The calculation and minimization procedures are essentially the same as outlined above. The main difference is that the method for determining L_T for each value of fractional saturation must be modified. In this case, eq 3 becomes

$$L_T = L + (\theta N/n)_{probe} + (\theta N/n)_{competitor} \qquad (A5)$$

and it is necessary to determine the fractional saturation of competitor as well as probe. This is done by first determining the free ligand concentration corresponding to a given fractional saturation of probe and then searching along the competitor binding curve in a like manner until two values of L are found that bracket the free ligand concentration determined for the probe. The fractional saturation of competitor is then determined by interpolation, and eq A5 is used to calculate L_T.

Figure A1 shows an example of a fitted competition titration in which the ligand is T4 gene 32 protein and the probe and competitor are poly(ϵA,U)[3] and poly(dT), respectively (the curve has been fitted to the lower 80% only, due to severe finite lattice effects). The net binding affinity of gene 32 protein for poly(dT) is known from separate studies to be only slightly higher than the binding affinity for poly(ϵA,U) under the conditions of the experiment. This is confirmed by a fit to the titration in which log ($K\omega$) is 9.11 for the poly(ϵA,U) probe and 9.20 for the poly(dT) competitor. That the curve shows a sigmoidal character in spite of the fact that the values of $K\omega$ are so close for the two lattices can be explained by the large cooperativity of binding of gene 32 protein. At *very low* fractional saturations, cooperativity has no effect, so that gene 32 protein will partition equally between probe and competitor. However, as more gene 32 protein is added, it binds contiguously next to protein molecules that are already bound. Thus, the high cooperativity magnifies the difference in binding constants between probe and competitor, favoring saturation of the competitor before full saturation of the probe can be achieved. For this reason, the titration curve shows a short initial rise, followed by a lag as added gene 32 protein partitions preferentially onto the competitor, followed by a second rise as additional gene 32 protein finally saturates the probe lattice. Under these conditions of high cooperativity *and* high binding constant, the shape of the curve is determined almost solely by the *ratio* of affinities for probe and competitor lattice rather than by the absolute affinities. Thus, a *proportionate* increase or decrease in these binding constants by as much as 20-fold has little effect on the shape of the curve below 80% saturation. However, a change in *one* binding constant relative to the other by as little as 5% results in a curve that no longer contacts the data points. Note that the ratio $K_{competitor}/K_{probe}$, estimated from Figure A1, may be low due to possible finite-lattice effects in the competitor as well as the probe lattice. These effects would result in premature upward curvature in the middle portion of the titration because ligand would begin binding to probe lattice prior to completely saturating the finite competitor lattice. The result would be an underestimate of $K_{competitor}$ relative to K_{probe}.

Registry No. Poly(riboethenoadenylic acid), 41911-88-0.

[3] Poly(ϵA,U) is a random copolymer of ethenoadenosine monophosphate and uridine monophosphate residues. Such fluorescent copolymers have proved very useful for the determination of titration curves of protein with nucleic acid. Details of the preparation and use of these copolymers to study the binding of *E. coli* transcription termination factor rho to nucleic acids will be presented elsewhere (J. M. McSwiggen, D. G. Bear, and P. H. von Hippel, unpublished results).

REFERENCES

Alma, N. C. M., Harmsen, B. J. M., deJong, E. A. M., Ven, J. V. D., & Hilbers, C. W. (1983) *J. Mol. Biol. 163*, 47–62.

Berg, O. G., Winter, R. B., & von Hippel, P. H. (1982) *Trends Biochem. Sci. (Pers. Ed.) 7*, 52–55.

Butler, A. P., Revzin, A., & von Hippel, P. H. (1977) *Biochemistry 16*, 4757–4768.

Crothers, D. M. (1968) *Biopolymers 6*, 575–584.

deHaseth, P. L., Gross, C. A., Burgess, R. R., & Record, M. T., Jr. (1977) *Biochemistry 16*, 4777–4783.

Draper, D. E., & von Hippel, P. H. (1979) *Biochemistry 18*, 753–760.

Epstein, I. R. (1978) *Biophys. Chem. 8*, 327–339.

Jensen, D. E., Kelly, R. C., & von Hippel, P. H. (1976) *J. Biol. Chem. 251*, 7215–7228.

Kelly, R. C., Jensen, D. E., & von Hippel, P. H. (1976) *J. Biol. Chem. 251*, 7240–7250.

Kowalczykowski, S. C., Bear, D. G., & von Hippel, P. H. (1981a) *Enzymes (3rd Ed.) 14*, 373–444.

Kowalczykowski, S. C., Lonberg, N., Newport, J. W., & von Hippel, P. H. (1981b) *J. Mol. Biol. 145*, 123–138.

Latt, S. A., & Sober, H. A. (1967) *Biochemistry 6*, 3293–3306.

Lin, S.-Y., & Riggs, A. D. (1975) *Cell (Cambridge, Mass.) 4*, 107–112.

Lohman, T. M. (1983) *Biopolymers 22*, 1697–1713.

Lonberg, N., Kowalczykowski, S. C., Paul, L. S., & von Hippel, P. H. (1981) *J. Mol. Biol. 145*, 123–138.

McGhee, J. D., & von Hippel, P. H. (1974) *J. Mol. Biol. 86*, 469–489.

Menetski, J. P., & Kowalczykowski, S. C. (1985) *J. Mol. Biol. 181*, 281–295.

Newport, J. W., Lonberg, N., Kowalczykowski, S. C., & von Hippel, P. H. (1981) *J. Mol. Biol. 145*, 105–121.

Porschke, D., & Rauh, H. (1983) *Biochemistry 22*, 4737–4745.

Record, M. T., Jr., Anderson, C. F., & Lohman, T. M. (1978) *Q. Rev. Biophys. 11*, 103–178.

Record, M. T., Jr., Mazur, S. J., Melancon, P., Roe, J.-H., Shaner, S. L., & Unger, L. (1981) *Annu. Rev. Biochem. 50*, 997–1024.

Ruyechan, W. T., & Wetmur, J. G. (1975) *Biochemistry 14*, 5529–5534.

Schellman, J. A. (1974) *Isr. J. Chem. 12*, 219–238.

Schwarz, G. (1970) *Eur. J. Biochem. 12*, 442–453.
Schwarz, G., & Watanabe, F. (1983) *J. Mol. Biol. 163*, 467–484.
von Hippel, P. H., Revzin, A., Gross, C. A., & Wang, A. C. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 4808–4812.
von Hippel, P. H., Kowalczykowski, S. C., Lonberg, N., Newport, J. W., Paul, L. S., Stormo, G. D., & Gold, L. (1982) *J. Mol. Biol. 162*, 795–818.
von Hippel, P. H., Bear, D. G., Morgan, W. D., & McSwiggen, J. A. (1984) *Annu. Rev. Biochem. 53*, 389–446.
Watanabe, F., & Schwarz, G. (1983) *J. Mol. Biol. 163*, 485–498.
Yamamoto, K. R., & Alberts, B. M. (1974) *J. Biol. Chem. 249*, 7076–7086.
Zasedetlev, A. S., Gurskii, G. V., & Volkenshtein, M. V. (1971) *J. Mol. Biol. 5*, 245–251.

Peroxidase-Catalyzed Covalent Binding of the Antitumor Drug N^2-Methyl-9-hydroxyellipticinium to DNA in Vitro†

Christian Auclair,*,‡ Bernard Dugué,§ Bernard Meunier,§ and Claude Paoletti‡

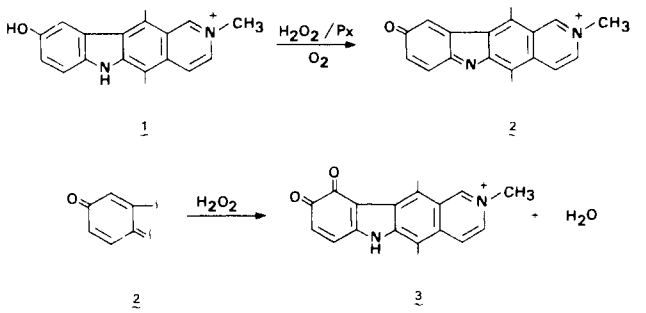
Laboratoire de Biochimie Enzymologie, INSERM U 140, CNRS LA 147, Institut Gustave Roussy, 94805 Villejuif Cedex, France, and Laboratoire de Pharmacologie et Toxicologie Fondamentales, Centre National de la Recherche Scientifique, 31400 Toulouse, France

Received June 25, 1985

ABSTRACT: In the presence of DNA, the antitumor drug N^2-methyl-9-hydroxyellipticinium (elliptinium; NMHE) [Le Pecq, J. B., Gosse, C., Dat-Xuong, N., & Paoletti, C. (1975) *C. R. Seances Acad. Sci., Ser. D 281*, 1365–1367] is oxidized by the horseradish peroxidase–hydrogen peroxide (HRP–H_2O_2) system to the quinone imine derivative N^2-methyl-9-oxoellipticinium (NMOE) [Auclair, C., & Paoletti, C. (1981) *J. Med. Chem. 24*, 289–295], which interacts with DNA according to the intercalation mode. When excess H_2O_2 was used, the major part of the quinone imine was further oxidized to the *o*-quinone N^2-methyl-9,10-dioxoellipticinium [Bernadou, J., Meunier, G., Paoletti, C., & Meunier, B. (1983) *J. Med. Chem. 26*, 574–579]. In the presence of stoichiometric amounts of H_2O_2 (H_2O_2/NMHE = 1), NMOE reacts with DNA, yielding a fluorescent compound irreversibly linked to the nucleic acid, which is related to the covalent binding of the ellipticinium chromophore. Under optimal reaction conditions, NMHE binding occurs according to a first-order process ($k = 4.3 \times 10^{-3}$ min^{-1}) with a linear increase with respect to drug to nucleotide ratio up to a maximum binding of 1 NMHE per 20 base pairs ($r = 0.05$). The fluorescence spectra (ex, 330 nm; em, 548 nm) of NMHE bound to DNA, the occurrence of energy transfer from the DNA to the drug, and the DNA length increase of the DNA–NMHE adduct suggest that the binding occurs at the intercalating site with limited denaturation of the DNA helix. The fluorescence properties of the ellipticine chromophore covalently bound to DNA are consistent with linkage between the C10 of NMHE and a primary amine of DNA [Auclair, C., Meunier, B., & Paoletti, C. (1983) *Biochem. Pharmacol. 32*, 3883–3886].

The oxidative bioactivation of the antitumor drug N^2-methyl-9-hydroxyellipticinium (NMHE)[1] has been found to occur through peroxidase and oxidase reactions leading to the generation of the reactive quinone imine N^2-methyl-9-oxoellipticinium (NMOE) (Scheme I) (Auclair & Paoletti, 1981; Auclair et al., 1983a; Bernadou et al., 1983). In the presence of suitable nucleophiles such as O, N, or S donor containing compounds, including amino acids, proteins, glutathione, and ribonucleosides, the oxidation of NMHE to NMOE results in the formation of covalent adducts in vitro (Auclair et al., 1983b,c, 1984; Meunier et al., 1983; Bernadou et al., 1984). Evidence for the occurrence of the oxidative bioactivation of NMHE in vivo was provided by the detection of the glutathione–NMHE adduct in biological fluids such as bile and urine of patients and animals treated with the drug (Monsarrat et al., 1983; Maftouh et al., 1984). Consequently, hypotheses concerning the mechanism of the cytotoxicity of NMHE should take into account its possible covalent binding to biological nucleophiles. Among the various cellular nucleophiles, double-stranded nucleic acids can be considered as preferential targets for NMHE since the drug interacts with nucleic acids through an intercalating mode of a high apparent affinity (Le

Scheme I



† This work was financially supported by CNRS and INSERM.
‡ Institut Gustave Roussy.
§ Centre National de la Recherche Scientifique.

[1] Abbreviations: NMHE, N^2-methyl-9-hydroxyellipticinium; NMOE, N^2-methyl-9-oxoellipticinium; NMDOE, N^2-methyl-9,10-dioxoellipticinium; Gly-NMHE, 7,10,12-trimethyl-6*H*-[1,3]-oxazolo[5,4-*c*]pyrido-[3,4-*c*]carbazole; Px, peroxidase; H_2O_2, hydrogen peroxide; HRP, horseradish peroxidase; r_i, initial drug to nucleotide ratio; ex, excitation; em, emission.

0006-2960/86/0425-1240$01.50/0