required to infer molecular level details. The broad substrate specificity of RNA ligase (Gumport & Uhlenbeck, 1981) should permit incorporation of a wide variety of nucleotide analogues at the contact sites identified here. Such variants should also be useful for understanding the energetic consequences of altering a contact, since the more subtle substitutions are expected to show smaller changes in K_a.

Registry No. **1**, 86784-73-8; **2**, 86784-63-6; **3**, 86784-59-0; **4**, 86784-57-8; **5**, 86784-66-9; **6**, 86784-62-5; **7**, 86784-61-4; **8**, 86784-60-3; **9**, 86784-58-9; **10**, 86784-74-9; **11**, 86784-68-1; **12**, 86784-67-0; **13**, 86784-78-3; **14**, 86784-75-0; **15**, 86784-77-2; **16**, 86784-76-1; **17**, 86784-72-7; **18**, 86784-64-7; **19**, 86784-65-8; **20**, 86784-81-8; **21**, 86784-79-4; **22**, 86784-70-5; **23**, 86784-71-6; **24**, 86784-69-2; ApApCpApUpG, 86669-63-8; UpGpCpCpC, 86669-64-9; UpCpCpCpC, 86669-65-0; UpUpCpCpC, 56931-11-4; GpApCpCpC, 86669-66-1; CpApCpCpC, 86669-67-2; ApApCpCpC, 86669-68-3; ApCpCpCpC, 86669-69-4; UpApCpC, 82604-49-7; UpApCpCpCpC, 86669-74-1; UpApUpU, 86669-75-2; UpApUpUpU, 86669-76-3; UpApApApA, 86669-70-7; UpApCpCpU, 86669-71-8; UpApUpUpC, 86669-72-9; ApApApApU, 70700-49-1; pGpGpApUp, 86669-73-0; ApApCpGpGpApUp, 86688-55-3.

References

Bernardi, A., & Spahr, P. F. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 3033–3037.

Betz, J., & Sadler, J. (1981) *Gene 13*, 1–12.

Carey, J., & Uhlenbeck, O. C. (1983) *Biochemistry 22*, 2610–2615.

Carey, J., Cameron, V., de Haseth, P. L., & Uhlenbeck, O. C. (1983) *Biochemistry 22*, 2601–2610.

Caruthers, M. H. (1980) *Acc. Chem. Res. 13*, 155–160.

Gilbert, W., Gralla, J., Majors, J., & Maxam, A. (1975) in *Protein–Ligand Interactions* (Sund, H., & Blauer, G., Eds.) pp 193–210, Walter de Gruyter, Berlin.

Goeddel, D. V., Yansura, D. G., Winston, C., & Caruthers, M. H. (1978) *J. Mol. Biol. 123*, 661–687.

Gumport, R. I., & Uhlenbeck, O. C. (1981) in *Gene Amplification and Analysis* (Chirikjian, J., & Papas, T., Eds.) Vol. 2, pp 313–345, Elsevier/North-Holland Biomedical Press, New York.

Jansone, I., Berzin, V., Gribanov, V., & Gren, E. J. (1979) *Nucleic Acids Res. 6*, 1747–1760.

Johnson, A., Meyer, B. J., & Ptashne, M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 5061–5065.

Krug, M., deHaseth, P. L., & Uhlenbeck, O. C. (1982) *Biochemistry 21*, 4713–4720.

Lodish, H. F., & Zinder, N. D. (1966) *J. Mol. Biol. 19*, 333–348.

Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y., & Matthews, B. W. (1982) *Nature (London) 298*, 718–723.

Peattie, D. A., Douthwaite, S., Garrett, R. A., & Noller, H. F. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 7331–7335.

Schulman, L. H., & Pelka, H. (1977) *Biochemistry 16*, 4256–4265.

Schulman, L. H., Pelka, H., & Susani, M. (1983) *Nucleic Acids Res. 11*, 1439–1446.

Shafranski, P., Zagorski, V., Khrobochek, Ya., & Zagorska, L. (1975) *Mol. Biol. (Engl. Transl.) 9*, 78–85.

Squires, C., & Carbon, J. (1971) *Nature (London), New Biol. 233*, 274–277.

Steitz, J. A. (1974) *Nature (London) 248*, 223–225.

Uhlenbeck, O. C., & Cameron, V. (1977) *Nucleic Acids Res. 4*, 85–98.

Weber, H. (1976) *Biochim. Biophys. Acta 418*, 175–183.

On the Determination of Deoxyribonucleic Acid–Protein Interaction Parameters Using the Nitrocellulose Filter-Binding Assay†

Charles P. Woodbury, Jr.,* and Peter H. von Hippel

ABSTRACT: We examine the effects of filter efficiency on DNA–protein binding data obtained by the popular nitrocellulose filter-binding assay. Graphical procedures for determining the efficiency parameter ϵ (for the efficiency of retention of DNA on the filter, per bound protein molecule) are established. Filter efficiency modified formulas for determining thermodynamic binding parameters are derived for four simple prototypes of DNA–protein binding systems. The effects of experimental error on discrimination between models are considered. Finally, we discuss conceptual errors often made in calculating binding stoichiometry from filter-binding experiments and suggest a general protocol for the analysis of filter-binding data.

The nitrocellulose filter assay is a popular tool for studying protein–nucleic acid interactions. As is well-known, double-stranded DNA binds very poorly to nitrocellulose while many proteins bind quite strongly; a protein molecule can thus retain on the filter (more or less efficiently) any double-stranded DNA molecule bound to it, since "free" DNA molecules are flushed through the filter. Typically, the DNA is radioactively labeled, and the extent of protein–DNA binding is determined by liquid scintillation counting of the filter.

A basic difficulty with interpreting results of the assay is that it monitors only the amount of DNA retained on the filter rather than the actual number of protein–DNA complexes in solution. Thus dissociation of weak complexes during the course of filtration may bias the results considerably. Further, since one protein molecule suffices (in general) to bind an entire DNA molecule to the filter, the assay does not discriminate among DNA molecules carrying one, two, or more

† From the Department of Medicinal Chemistry and Pharmacognosy, University of Illinois at Chicago, Chicago, Illinois 60680 (C.P.W.), and the Institute of Molecular Biology and Department of Chemistry, University of Oregon, Eugene, Oregon 97403 (P.H.v.H.). *Received February 16, 1983.* Supported by National Institutes of Health Postdoctoral Fellowship GM-05669 (to C.P.W.) and National Institutes of Health Grant GM-15792 (to P.H.v.H.).

0006-2960/83/0422-4730$01.50/0

bound protein molecules. Lastly, for a number of reasons, even tight complexes of protein with DNA are known to be retained by the filter with less than 100% efficiency [a good example being *lac* repressor protein and *lac* operator DNA, with a binding constant of $\sim 10^{13}$ M^{-1} under typical binding conditions, where often only 30–40% of the total DNA is retained, even under stoichiometric binding conditions (Riggs et al., 1970)].

Quantitative corrections to filter-binding data were first discussed by Yarus & Berg (1967, 1970), in connection with the binding of tRNA by tRNA synthetase. Hinkle & Chamberlin (1972), in their study of RNA polymerase–promoter interactions, suggested a statistical correction to filter-assay data for the particular case of a DNA molecule containing multiple (10 or more) equivalent binding sites, when the complexes are inefficiently retained by the filter. More recently, Bailey (1979) has examined the applicability of Scatchard-type analyses of filter binding assay data and has emphasized the difficulties involved in dealing with systems with multiple equivalent binding sites on the DNA. Clore et al. (1982) have further explored the application of the theory for multiple binding sites on the DNA, using the binding-polynomial concept of Wyman (1948, 1964). They have also applied the theory to particular systems that show both non-specific and site-specific binding, as well as binding cooperativity.

Our purposes here are several. First, we wish to generalize previous theoretical treatments of multisite binding in the filter assay to include specifically the effects of less than 100%-efficient filter trapping of DNA–protein complexes. Second, we shall show through model calculations how filter inefficiency can confuse the interpretation of filter-assay results, resulting in incorrect interpretations of the data. Third, we shall set out procedures to determine (or at least to estimate) the efficiency parameter (ϵ), as well as standard thermodynamic quantities of interest (such as binding constants, number of binding sites, and binding cooperativity), under conditions where the filter retains complexes inefficiently. Along the way, we shall comment on the weakness of the filter binding assay method in determining binding stoichiometry, and in an examination of so-called "stoichiometric" titration procedures, we shall show how such titrations, though easily misinterpreted, can yield information on the number of binding sites.

Theory

Binding-Polynomial Approach. We base our approach on the concept of the binding polynomial for multisite binding (Wyman, 1948, 1964; Schellman, 1975). Consider a system containing a total molar concentration, D, of DNA molecules and a total molar concentration, P_t, of protein molecules. We assume a single DNA binding site per protein molecule but allow the DNA molecules to bind up to n protein molecules, so that the DNA exists in solution in the form of any of $n + 1$ species (having 0, 1, 2, ..., n protein molecules bound). The concentration of each of the $n + 1$ species obeys the relation:

$$D_i = D_0 K_i P^i \tag{1}$$

where P is the free protein activity (which we shall simply equate to the concentration), D_i is the concentration of the ith complex species, and K_i is the phenomenological association constant for formation of that species. The total DNA concentration (D) is related to the D_i's by

$$D = \sum_{i=0}^{n} D_i = D_0(1 + K_1P + K_2P^2 + \ldots + K_nP^n) = D_0\Sigma \tag{2}$$

where Σ is the binding polynomial:

$$\Sigma = 1 + K_1P + K_2P^2 + \ldots + K_nP^n \tag{3}$$

The fraction f_i of DNA molecules with i proteins bound is

$$f_i = K_iP^i/\Sigma \tag{4}$$

and the total fraction F of DNA molecules binding *one or more* protein molecules is

$$F = \sum_{i=1}^{n} f_i = \sum_{i=1}^{n} K_iP^i/\Sigma \tag{5}$$

This fraction F is generally not what is detected directly in the filter-binding assay. Due to inefficiencies in retention of DNA–protein complexes, some lesser fraction R is detected, and the whole problem now is to establish the relation of R to F.

To obtain a general expression for R, we consider first a single class of n binding sites on the DNA that are characterized by a common filter-binding efficiency, ϵ. The quantity ϵ can be thought of as the probability that a protein complexed to DNA at a particular site is held on the filter. (See Results and Discussion for further consideration of the ϵ parameter.) The quantity $\zeta = 1 - \epsilon$ is then the probability that this same complex (and *not* necessarily the entire DNA molecule) is *not* retained. We shall assume that the protein from "unretained" complexes, in whatever form it exists in the system (i.e., bound to the matrix, bound to DNA, or free in solution), plays no further role in DNA retention. For a DNA molecule binding i protein molecules with each complexed site having an independent probability ζ of *not* being retained, the overall probability of retaining the DNA molecule itself is just $1 - \zeta^i$. The fraction of all DNA molecules retained (with one or more sites occupied) is finally given by

$$R = \sum_{i=1}^{n} f_i(1 - \zeta^i) \tag{6a}$$

$$R = \sum_{i=1}^{n} K_iP^i(1 - \zeta^i)/\Sigma \tag{6b}$$

Note that while we have assumed that the effectiveness of *individual* protein molecules in holding DNA on the filter is constant (that is, ϵ is fixed), the *overall* efficiency of DNA retention (that is, $1 - \zeta^i$) increases rapidly with the number of protein molecules bound.

The concentration of *bound* protein, P_b, is given, in general, by

$$P_b = D\sum_{i=1}^{n} if_i = D_0\sum_{i=1}^{n} iK_iP^i \tag{7}$$

The *total* protein concentration, P_t, is then simply the sum of the free (P) and bound (P_b) protein concentrations:

$$P_t = P + P_b = P + D_0\sum_{i=1}^{n} iK_iP^i \tag{8}$$

Very commonly, filter-binding results are reported in the literature in the form of a graph of the extent of DNA retention (often as merely the amount of radioactivity detected, rather than the moles of DNA) as a function of added protein. Equations 6b and 8 provide the basis for modeling such titrations. In what follows, we shall refer to titrations where R is monitored as a function of total protein concentration as type I titrations; that is, ligand (protein) is added to fixed total lattice (DNA).

One frequently encounters a second type of titration where P_t is held constant while D is varied (i.e., lattice or DNA is added to fixed *total* ligand or protein). We shall refer to this as a type II titration. To model such titrations, we use the relation

$$D = \frac{(P_t - P)\Sigma}{\sum_{i=1}^{n} iK_iP^i} \quad (9)$$

Generally, in such titrations one follows the *apparent* concentration, C, of DNA–protein complexes. C is just the fraction of the total DNA that is retained on the filter and is the product of the retention factor, R, and the total DNA concentration, D:

$$C = RD = \frac{R(P_t - P)\Sigma}{\sum_{i=1}^{n} iK_iP^i} \quad (10)$$

The general formulas for R, D, and C in eq 6, 9, and 10 can be refined when one chooses a particular model for the binding system. Questions of site specificity and independence of binding naturally arise when deciding upon the model. We consider here four simple binding models showing varying degrees of binding independence and site specificity. The formulas for R, D, and C will, of course, vary from model to model.

In the first model, the DNA molecule has a single class of identical, independent binding sites. A relevant biochemical example would be the binding of RNA polymerase to identical promoter sites. In the second model, the DNA has several classes of binding sites, with the sites in each class being again identical and independent. As an example, one might consider a DNA molecule with several classes of promoter sites that are distinguished by their binding affinity for RNA polymerase. A third simple model allows binding to specific sites on the DNA, with interactions between neighboring (nonoverlapping) filled sites (the sites are no longer independent). An example here would be the cooperative binding of bacteriophage λ repressor protein to the λ right operator DNA sequence (Ptashne et al., 1976). Finally, in the fourth model, the protein binds nonspecifically to the DNA and may interact with neighboring bound protein molecules. In this model, each bound protein molecule also may cover more than a single unit (e.g., base or base pair) of the DNA strand(s) and so exclude from binding a number of potential sites (McGhee & von Hippel, 1974). The gene 32 protein of bacteriophage T4 binds to single-stranded DNA in such a manner (Kowalczykowski et al., 1981). Let us now consider the expressions for R, D, and C for these four models.

(*i*) *Equal and Independent Binding Sites.* Associated with each site is an intrinsic binding constant k so that the general phenomenological binding constant K_i is

$$K_i = \frac{n!}{i!(n-i)!}k^i \quad (11)$$

The binding polynomial Σ is simply

$$\Sigma = (1 + kP)^n \quad (12)$$

which leads to

$$R = 1 - \frac{(1 + kP\zeta)^n}{(1 + kP)^n} \quad (13)$$

$$D = \frac{(P_t - P)\Sigma}{nkP(1 + kP)^{n-1}} = \frac{(P_t - P)(1 + kP)}{nkP} \quad (14)$$

$$C = \frac{(1 + kP)(P_t - P)}{nkP}\left[\frac{(1 + kP)^n - (1 + kP\zeta)^n}{(1 + kP)^n}\right] \quad (15)$$

The case of a single binding site on the DNA (e.g., *lac* repressor binding to *lac* operator DNA) is treated simply by setting n equal to 1.

(*ii*) *Multiple Classes of Equal and Independent Sites.* In each class j, there are n_j sites, each of intrinsic binding constant k_j and efficiency $\epsilon_j = 1 - \zeta_j$, with a total of m classes. We have

$$R = 1 - \prod_{j=1}^{m}\left(\frac{1 + k_jP\zeta_j}{1 + k_jP}\right)^{n_j} \quad (16)$$

$$D = \frac{(P_t - P)}{\sum_{j=1}^{m}\frac{n_jk_jP}{1 + k_jP}} \quad (17)$$

$$C = \frac{(P_t - P)[\prod_{j=1}^{m}(1 + k_jP)^{n_j} - \prod_{j=1}^{m}(1 + k_jP\zeta_j)^{n_j}]}{\sum_{j=1}^{m}\frac{n_jk_jP}{1 + k_jP}\sum_{j=1}^{m}(1 + k_jP)^{n_j}} \quad (18)$$

This second model is a generalization of the first; it reduces to model i upon setting m (the number of classes) equal to 1.

(*iii*) *Specific Binding Sites with Cooperative Interactions between Nearest Neighbors.* We suppose that the n binding sites are arranged linearly along the DNA. Binding of a protein molecule to an isolated site (that is, the two flanking sites are unfilled) is characterized by an intrinsic binding constant, k. Binding of a protein to a site where one (and only one) of the flanking sites is filled is characterized by a binding constant, $k\omega$; ω is a positive (dimensionless) constant known as the cooperativity parameter and allows for favorable or unfavorable (in terms of free energy) interactions between neighboring sites. Finally, binding of a protein to a site flanked by two filled sites is characterized by $k\omega^2$. The extra factor of ω arises from the interactions of the bound protein with the *two* adjacent neighbors. For simplicity, we shall suppose that the filter-binding efficiency for each protein binding site on the DNA is ϵ (however, see Results and Discussion for consideration of the possible nonindependence of filter retention of adjacently bound protein molecules). The phenomenological binding constant for binding i protein molecules is then

$$K_i = k^i\sum_{j=1}^{i}\frac{(n-j-1)!}{(i-j)!(n-i-1)!}(n - i + 1)\omega^{j-1} \quad (19)$$

where j is the number of contiguous and cooperatively interacting protein molecules. The formulas for R, D, and C then follow immediately upon insertion for K_i in eq 6, 9, and 10.

(*iv*) *Cooperative Binding to a Linear Lattice with Potential Binding-Site Overlap.* As one of a number of equivalent descriptions of cooperatively binding protein–DNA systems, we choose the McGhee–von Hippel (1974) model. The basic parameters here are the length of the lattice, N (in lattice units), the number of consecutive lattice units covered by a protein molecule (the site size), $\bar{n}$, the intrinsic binding constant for an isolated protein molecule, k, and the cooperativity parameter ω, a dimensionless equilibrium constant for placing contiguously a pair of previously separated but lattice-bound protein molecules. We again assume a common filter-binding efficiency of ϵ for each site. Then, the phenomenological binding constant for binding i proteins is

$$K_i = k^i\sum_{j=1}^{i}\frac{[N - i(\bar{n} - 1) - j - 1]!}{(i-j)!(N - i\bar{n} - 1)!}(N - i\bar{n} + 1)\omega^{j-1} \quad (20)$$

where j is the number of *contiguous* and cooperatively interacting protein molecules. [The combinatorics of such systems have been worked out by McQuistan (1968).] The

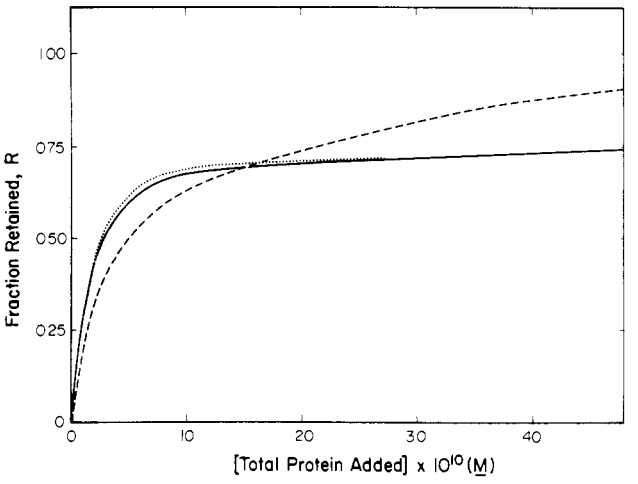

FIGURE 1: Effects of filter efficiency and types of binding on type I titrations: (—) $\epsilon = 0.75$, $n = 1$, $k = 1 \times 10^{10}$ M^{-1}; (···) $\epsilon = 0.50$, $n = 2$, $k = 1 \times 10^{10}$ M^{-1}; (--) $\epsilon = 0.50$, $n_1 = 1$, $k_1 = 1 \times 10^{10}$ M^{-1}, $n_2 = 10$, $k_2 = 1 \times 10^8$ M^{-1}.

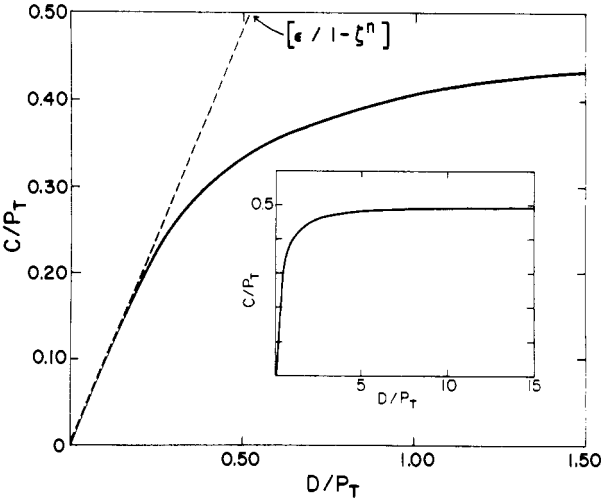

FIGURE 2: Type II titration (with reduced variables C/P_t and D/P_t) showing a linear extrapolation of curve to intercept asymptote for C/P_t. Here, $\epsilon = 0.5$, $n = 5$, $k = 1 \times 10^{10}$ M^{-1}, and $P_t = 1 \times 10^{-8}$ M.

formulas for R, D, and C in the case of cooperative binding then follow immediately upon insertion for K_i in eq 6, 9, and 10. We note that model iii above is simply model iv with the site size $\bar{n}$ equal to 1.

Results and Discussion

There are several parameters to consider in interpreting a filter-binding curve: the efficiency, ϵ; the number of binding sites, n; the binding constant, k (or set of constants, k_j); the binding site size, $\bar{n}$ (for nonspecific binding); and the binding cooperativity, ω. Our aim here will be to lay out relatively rapid routes to the determination (or, at least, the estimation) of a number of these parameters, on the basis of binding-curve asymptotes, limiting slopes, titration midpoints, etc. With these parameters in hand, one can then go back and, through curve fitting, obtain values for the other parameters of interest. In our discussion, we shall assume that the protein preparation is 100% active and that each protein molecule has only one site for binding DNA. We are concerned primarily with cases involving multiple binding of proteins to a single DNA molecule, where the differences between the filter assay and other methods (that measure directly the extent of binding saturation) are most pronounced. Simplification to the single binding site case is straightforward.

Determination of Efficiency Parameter, ϵ. Inefficiency in filter retention of DNA–protein complexes can be determined from either type I or type II titrations. In a type I titration, an apparent plateau below 1.0 in experimental values of R shows the loss or escape of complexes from the filter. From eq 6, one can easily show that for large values of P the asymptotic value of R is $1 - \zeta^n$ for n equal and independent sites (as before, we assume a common efficiency factor ϵ per binding site, as well). For example, with $n = 1$ and $\epsilon = 0.75$, the "plateau" for R values should be 0.75; for $n = 2$ and $\epsilon = 0.50$, this asymptotic value is again 0.75.

At this point, we digress briefly to remark on the possible similarities of type 1 binding curves for quite different binding models. In Figure 1 we show three different binding systems: (i) a single site per DNA molecule with $\epsilon = 0.75$ and $k = 10^{10}$ M^{-1}; (ii) two sites per DNA with $\epsilon = 0.50$ and $k = 10^{10}$ M^{-1}; and (iii) 11 sites per DNA with $\epsilon = 0.50$, where 10 of the 11 sites are 100-fold weaker in binding affinity ($k_2 = 10^8$ M^{-1}) than the first or primary site ($k_1 = 10^{10}$ M^{-1}). First, we see that the curve for the single binding site ($\epsilon = 0.75$) is virtually superimposable on the curve for two sites with $\epsilon = 0.50$; the filter assay cannot by itself distinguish between these two binding schemes. Second, the binding curve for the 11-site system with two distinct classes of sites lies very close to the curves for the one- and two-site systems; with the usual scatter in experimental data this alternative scheme might provide as good an apparent fit to a set of data as do the two simpler schemes.

Returning to the determination of ϵ, in a type I titration the asymptotic value for R of $1 - \zeta^n$ provides a way to obtain ζ or ϵ, if one assumes that all of the complexes formed are retained with equal per site efficiency. This method, however, is not a very satisfactory one. First, the calculation of ϵ depends on knowledge of the number of binding sites, n. Second, as R approaches 1, a small error in its determination can lead to a larger error in ϵ. As an example, let $R = 0.90 \pm 0.05$ and let $n = 2$. The calculated values of ϵ then range from 0.61 to 0.78, rather more than the (expected?) error of ~5% in ϵ. One would like to have a method of establishing ϵ that avoids these problems.

The use of a type II titration to obtain ϵ was discussed some time ago by Yarus & Berg (1970) in the context of a system with single binding sites on both the protein and the nucleic acid. However, it has perhaps not been widely appreciated that the type II titration approach is a very simple way to determine ϵ for systems where the nucleic acid has multiple sites for binding protein. From eq 10, one can readily show that for large values of D, under stoichiometric binding conditions, the asymptotic value of C (the *apparent* concentration of DNA–protein complexes) is just ϵP_t (see Figure 2). To show this, we first note that as P approaches 0, so do both R and the sum in the denominator of eq 10. This leaves C in an indeterminate form as P goes to 0. However, we can use L'Hôpital's rule to find the limit of this quotient by taking derivatives with respect to P. The limit of the derivative of the numerator is $P_t K_1(1 - \zeta)$, and that of the denominator is just K_1. Hence, the limit of C as P approaches 0 is just $P_t(1 - \zeta)$ or ϵP_t. Note that this result is independent of the number of sites, unlike the asymptotic value for R in a type I titration, and that ϵ is easily and directly calculated without magnification of the experimental error. (Our assumptions are that, per bound protein, all complexes are equally likely to be trapped by the filter, there is only one binding site per protein molecule, and, of course, all protein molecules are fully active with respect to binding.) Thus the use of the asymptote in a type II titration appears to be a better method of determining

ϵ than the use of R from a type I titration.

In Figure 2 we show a model titration where the total protein concentration is held constant (P_t fixed) while DNA is added (a type II titration, in our classification scheme). For convenience, we have plotted the reduced (dimensionless) variable C/P_t as a function of the reduced variable D/P_t. From the model efficiency (ϵ) chosen, the asympotic value for C/P_t is 0.5; the figure illustrates the approach of C/P_t to this asymptote. We have also indicated on the figure the intercept at the asymptote of the initial linear portion of the binding curve. This intercept may be useful in estimating binding stoichiometry (see next section).

Determination of Binding Stoichiometry. If one assumes now that ϵ has been determined, the next problem is to find the number of binding sites on the DNA and their associated binding constants. Unfortunately, the filter-binding method is especially weak in the experimental determination of binding stoichiometry. This is a reflection of the sensitivity of the method to the binding of the DNA to the *first* of n protein molecules and its relative insensitivity to the subsequent binding of more protein molecules.

If the efficiency is precisely known and there are only a few sites (e.g., five or less), then one can attempt to find n from the asymptotic value of R in a type I titration through the relation $R = 1 - \zeta^n$. (Of course, this assumes once again equal efficiencies, per bound protein, of trapping on the filter all possible complexes.) Given the usual experimental uncertainty in R, this approach will not give n accurately (e.g., to ± 1 sites) if n is greater than about 5 or if the efficiency ϵ is too high (e.g., greater than 0.9).

When experimental techniques are used that measure directly the degree of binding saturation, a common scheme for determining n under "stoichiometric" binding conditions ($kP \gg 1$; $P_b \simeq P_t$) is to extrapolate the initial tangent line of a type II titration to an intercept with P_t, the asymptote for S_b.[1] (One follows the *true* concentration of bound *sites*, S_b, *not* the *apparent* concentration given by the filter assay.) Since the equation of this straight line is $S_b = nK_{app}P_tD/(1 + K_{app}P_t)$, at the point of interception we have $n = (P_t + K_{app}^{-1})/D_{int}$, where D_{int} is the value of D at the intercept. When $P_t \gg K_{app}^{-1}$, we have $n = P_t/D_{int}$, and the binding stoichiometry is thus readily calculated. (A similar scheme can be worked out for type I titrations as well.)

It is *not* proper to apply this scheme naively to filter binding assay results. In this assay, we measure the concentration of *apparent* (or retained) complexes, leading to $C = R_iD$ as the equation of the initial tangent line, where R_i is the initial slope. Under stoichiometric binding conditions, we can approximate the binding polynomial Σ and the retention factor R_i by the major term (which is the last or nth term, for the models we have considered) in eq 3 and 6, respectively, so that $\Sigma \simeq K_nP^n$ and $R_i = 1 - \zeta^n$. Since $C = \epsilon P_t$, at the point of interception with the asymptote we have (see Figure 2)

$$\frac{D_{int}}{P_t} = \frac{\epsilon}{1 - \zeta^n} \tag{21}$$

Plainly, D_{int}/P_t does not yield the binding stoichiometry directly. Furthermore, we note from the form of eq 21 that, unless ζ (or ϵ) is precisely known, the error in calculating n may be substantial.

Determination of Binding Constants. One can attempt to fit the entire range of binding data with a single isotherm, making certain assumptions about the character of the binding system. In the case of a sufficiently large number of independent and identical sites, so that Poisson statistics govern the distribution of bound protein, eq 13 reduces to

$$R = 1 - e^{-\epsilon u} \tag{22}$$

where u is the average number of protein molecules bound per DNA molecule:

$$u = \frac{nkP}{1 + kP}$$

Equation 22 was first derived by Hinkle & Chamberlin (1972) and is an especially simple expression to use in curve-fitting routines to obtain k (always providing that n and ϵ are known). More generally, for systems with a *very* large number of *nonspecific* binding sites, Clore et al. (1982) have shown that, to a very good approximation, the series expressions for ζ and F can be truncated as fifth degree polynomials in P, even though cooperativity or site-specific binding may be present. The same should be true also of our expression for R, eq 6. These simplified expressions can then be employed throughout the entire range of binding data to obtain binding constants.

In the case of multiple classes of binding sites, we note that, qualitatively, a few high-affinity sites in a background of many low-affinity sites will produce a binding curve (in a type I titration) like that seen in Figure 1 for the 11-site/two-classes system. Depending on the binding efficiency, the relative magnitudes of the binding constants, and the number of binding sites, the curve may show two regions. The first is a sharp rise at relatively low P_t concentrations, corresponding principally to binding at the high-affinity sites. The second is a rather more gentle rise extending over a much wider P_t concentration range, which corresponds to titration of the lower affinity sites. However, for high efficiencies, this second region of the curve may be compressed against the $R = 1$ asymptote, while, for lower efficiencies, the gentle rise may not be discernible in the possibly scattered data available. (Consider whether the three curves of Figure 1, with uniform uncertainties in experimental values of R of $\pm 5\%$, would be easily distinguished.)

Probably, the best approach to take, when site-specific binding is to be characterized against a background of non-

---

[1] Consider the apparent equilibrium between "free" and complexed ("bound") sites on the DNA:

$$S_f + P \underset{}{\overset{K_{app}}{\rightleftharpoons}} S_b$$

Here S_f is the concentration of free sites, S_b the concentration of bound sites, P the free-protein concentration, and K_{app} the apparent equilibrium constant. Since $K_{app} = S_b/(S_fP)$ and $S_f = nD - S_b$ (where n is the number of sites and D the concentration of DNA molecules), we find

$$S_b = (1/2)(K_{app}^{-1} + nD + P_t - [(nD + P_t + K_{app}^{-1})^2 - 4nDP_t]^{1/2})$$

where P_t is the total protein concentration. The limiting slope of a plot of S_b vs. D is the derivative dS_b/dD, evaluated at $D = 0$:

$$\left(\frac{dS_b}{dD}\right)_{D=0} = \frac{nK_{app}P_t}{1 + K_{app}P_t}$$

We then have for the equation of the tangent line of the binding plot as D approaches 0:

$$S_b = \frac{nK_{app}P_tD}{1 + K_{app}P_t}$$

When all of the protein has been bound, $S_b = P_t$, and from the intercept of the initial tangent line with this asymptote, we find

$$D_{int} = (nK_{app})^{-1} + P_t/n$$

where D_{int} is the DNA concentration corresponding to the intercept. Then

$$n = P_t/D_{int} + (K_{app}D_{int})^{-1}$$

specific binding, is to compare binding to DNA molecules that do and do not contain the putative specific sites. [See, for example, Strauss et al. (1980) for an application to RNA polymerase–promoter interactions.] An alternative, unfortunately not often available, is to add a small molecule "inducer" that will reduce or eliminate site-specific binding; a well-known example here is the use of various galactosides with *lac* repressor to reduce operator-specific binding. For further quantitative considerations of the general problem, see Clore et al. (1982).

In situations where the specific binding sites on the DNA are of (roughly) equal efficiency but differ considerably in their affinity for protein, it is difficult to resolve the resulting filter-binding curve into contributions from the different classes of sites. (Such a situation might arise, for example, in the binding of RNA polymerase to a mixed set of strong and weak promoters.) In the course of a titration, the highest affinity sites will tend to be filled before any lower affinity sites. Because of the assay's insensitivity to the binding to the DNA of a second, third, etc. protein molecule, the experimental binding curve will not be much affected by the titration of the lower affinity sites. A good alternative to the filter-binding assay in such cases might be "protection" experiments of the sort described by Johnson et al. (1979) for the bacteriophage λ repressor–operator system. [A similar type of experiment has been described by Schmitz & Galas (1979, 1980).] A very nice thermodynamic analysis of the results of this sort of experiment has been given by Ackers et al. (1982).

Often enough, the binding constant k may not be known beforehand, yet some information may be available on the binding stoichiometry. For example, in the case of a restriction endonuclease, the number of cleavage sites on a given piece of DNA may be known and so provide a first estimate of the number of (specific) binding sites. Then the binding constant k can be calculated from a type I titration curve, by using the relations given above and assuming that the sites are identical in their binding behavior. Alternatively, one might follow Bailey's (1979) method, on the basis of a Scatchard plot of the data. One can also obtain k from a type I titration by using the relation (for n identical, independent sites)

$$k = \frac{2^{1/n} - 1}{(2^{1/n}\zeta + 1)P_{1/2}} \tag{23}$$

where $P_{1/2}$ is the free protein concentration corresponding to $R = 0.5$. This relation is a generalization (to include efficiency effects) of a result previously obtained by Clore et al. (1982).

Determination of Binding Cooperativity. The filter assay is moderately sensitive to cooperative effects in protein binding. Evidence for cooperativity is a sigmoid curve in a type I titration; however, the cooperative character of the binding may not be apparent (the curve not appreciably sigmoid) if the cooperativity is very low or if the efficiency is very high. In Figure 3, we show the behavior of a system where the binding is nonspecific and in which there is overlap between binding sites. Curves a and b show that high cooperativity alone ($\omega = 10^3$) can lead to nearly hyperbolic (as opposed to sigmoid) binding curves; however, the lower efficiency ($\epsilon = 0.2$) for curve b does result in a slightly more sigmoid curve than for curve a ($\epsilon = 0.8$). Curve c shows that moderate cooperativity ($\omega = 50$) in combination with high efficiency ($\epsilon = 0.8$) results in an apparently hyperbolic curve, with none of the sigmoid character expected for a cooperative binding system. Finally, curve d shows that when the efficiency is low ($\epsilon = 0.2$), the cooperative nature of the system ($\omega = 50$) is readily apparent from the sigmoid curve.

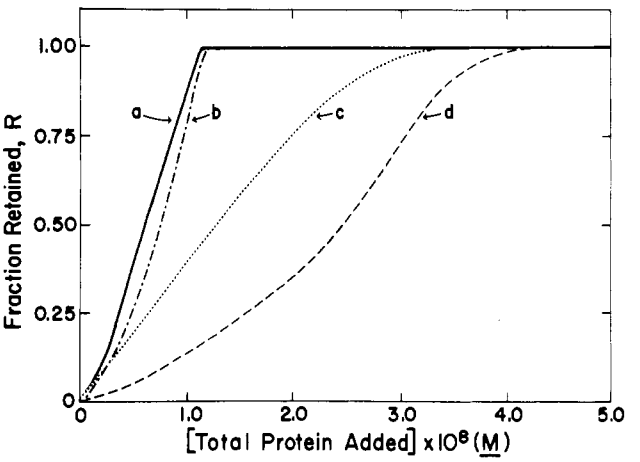

FIGURE 3: Effects of cooperativity and filter efficiency on type I titration curves: $N = 75$, $\bar{n} = 3$, $k = 1 \times 10^6\ M^{-1}$. (a) $\omega = 10^3$, $\epsilon = 0.8$; (b) $\omega = 10^3$, $\epsilon = 0.2$; (c) $\omega = 50$, $\epsilon = 0.8$; (d) $\omega = 50$, $\epsilon = 0.2$.

Unfortunately, there appear to be no simple relations to be derived between k, n, ϵ, and the cooperativity parameter ω. Complicating the picture is the possibility that the cooperativity may arise through protein–protein association in solution *before* binding. An example of this sort is the cI repressor protein of phage λ, which appears to dimerize before binding to the DNA (Pirotta et al., 1970). [A further complication in this particular system is the intrinsic cooperativity seen in binding this repressor to operator sites; for a review, see Johnson et al. (1981).] Cooperativity may also arise through the interaction of neighboring protein molecules after they are bound to the DNA lattice [e.g., the binding to single-stranded DNA of the gene 32 protein of phage T4; see the recent review by Kowalczykowski et al. (1981)].

We shall focus our attention here on the second possibility, where the binding cooperativity is to be interpreted in terms of the McGhee–von Hippel (1974) model. Since there is no simple way to obtain ω from either type I or type II titrations (using asymptotes, limiting slopes, etc.), iterative curve fitting must be employed; once $\bar{n}$, k, and ϵ are established (as described above), then ω is varied until suitable agreement between the theoretical curve and the experimental data is achieved. The most difficult situation to deal with in theory is that in which the lattice length N is only moderately large compared to the site size $\bar{n}$ of the protein. Under these circumstances, the series expressions for K_i, R, and Σ cannot be truncated (as they can for $N \gg \bar{n}$) without serious error, yet the full expressions are extremely cumbersome to handle. For simpler computations, a good experimental approach would be to use oligonucleotides that can bind only a few protein molecules at saturation and to use the full series expressions for K_i, etc. [see Draper & von Hippel (1978) for an analysis of the thermodynamics of binding two proteins to an oligonucleotide]. An alternative is to use *very* long polynucleotides, so that truncated series expressions for R, Σ, etc. can be applied in curve fitting.

Finally, it should be pointed out that sigmoid curves may be the result of excessive washing of the filter; an example is the early report by Freeman & Jones (1967) on the binding of RNA polymerase to T7 DNA. When sigmoid filter-binding curves occur, as a final check on artifactual "cooperativity", the binding system should be analyzed by an independent technique [such as the "protection" method of Johnson et al. (1979) mentioned above].

Further Considerations with Respect to Filter-Binding Efficiency. It is well-known that low to moderate amounts

of protein are quantitatively retained by nitrocellulose filters. Two good examples are *E. coli* RNA polymerase, at up to 20 μg/13-mm filter (Hinkle & Chamberlin, 1972), and *E. coli* isoleucyl-tRNA synthetase, at up to 100 μg/24-mm filter (Yarus & Berg, 1967). The exact mechanism by which the protein molecules are held by the filter is not known. The pores of the filter (typically 0.45 μm in diameter) are too large to hold back the protein by simple mechanical means. Changes in salt concentration or in pH will affect the level of protein retention (Yarus & Berg, 1967, 1970; Riggs et al., 1970; Hinkle & Chamberlin, 1972; Strauss et al., 1981), so presumably the protein adheres to the nitrocellulose matrix through ionic, hydrogen-bonding and hydrophobic interactions. Conceivably, partial or complete unfolding of a protein molecule on the filter might play a role in its retention. There is evidence, however, that the protein often retains at least some of its native conformation. For example, Thang et al. (1968) have found that filter-bound polynucleotide phosphorylase retains about 30% of the activity of free enzyme. In another example, Gilbert & Maxam (1973) have used a wash containing the synthetic inducer isopropyl β-D-thiogalactoside (IPTG) to release operator-containing DNA from filter-bound *lac* repressor–operator complexes. Their success implies that the filter-bound *lac* repressor retains at least those portions of its native conformation responsible for IPTG binding and for operator binding, as well as the functional connection of these regions.

Although a protein may be quantitatively retained by nitrocellulose, in many cases the DNA that it has bound will pass through the filter, resulting in the low apparent retention efficiencies seen in the *lac* and other systems. This inefficiency derives from the nonequilibrium nature of the filter assay, which has been commented on before by various workers [see, for example, Riggs et al. (1970) or Johnson & Geiduschek (1977)]. We simply note here that the relative rates of dissociation and association of complexes in solution and on the filter, as well as the rate of filtration and the volume filtered, all may be involved in inefficient complex retention.

Suppose that in a given system the protein were *quantitatively* retained. Then perhaps the simplest explanation of inefficient DNA retention is that the filter-bound complexes dissociate, and the released DNA then passes through the filter (we tacitly ignore recapture of DNA by filter-bound protein; this is very likely a negligible effect). It seems less likely to us that any protein bound to DNA is somehow "protected" against adhering to the filter, with such protected complexes escaping from the filter. We have found little or no evidence in the literature to suggest such protection, so long as the filter is not saturated with protein.

It also seems reasonable to us to suppose that, in general, complex dissociation at one site on a given DNA molecule will occur independently of any other site on that molecule. Of course, this would probably not hold if the DNA-bound protein molecules were close enough together to interfere with one another in the release of DNA (this might happen, for example, in cooperative, nonspecific binding systems or in cases where specific binding sites are very close together). Our supposition of the independence of DNA release then leads directly to the factor of $1 - \zeta'$ in eq 6 and in the consequent relations.

Conclusions

For simplicity, we have not considered here systems where the protein has multiple binding sites for DNA. Analysis of such systems would of course require reconsideration of the various formulas we have presented here for ϵ, n, k, etc. We have also left aside the problems posed by protein preparations that are less than 100% active in binding. Our emphasis has been on the effects of filter efficiency in systems where the DNA (not the protein) has multiple binding sites.

For such systems, the analysis of filter-binding experiments should start with the determination of the efficiency ϵ, since this quantity can drastically affect the shape of the curve in a type I titration (see Figure 1 and 3) and is also closely tied to the calculation of the binding parameters n, k, $\bar{n}$, and ω. The most direct and accurate means of finding ϵ is through a type II titration, where the asymptotic value of C (the apparent concentration of DNA–protein complexes) can be used.

The filter assay is relatively sensitive to cooperativity in binding, and a type I titration curve should show the sigmoid curve characteristic of cooperative systems, provided that ω is large enough and that the assay is less than 100% efficient. The filter assay is thus often useful as a diagnostic tool for cooperative binding. However, the determination of ω via filter binding will involve curve fitting rather than simple calculation from a formula.

The strength of the filter-binding assay is its sensitivity to the first protein molecule bound; its weakness is its relative insensitivity to the binding of the second, third, etc. proteins. Hence, the assay is not well suited to the determination of binding stoichiometry, either through type I or type II titrations. It is worth stating once again that stoichiometric titrations by filter binding do *not* yield the number of binding sites directly; instead, for example, in a type II titration the ratio D_{int}/P_t is equal to $(1 - \zeta)/(1 - \zeta')$, *not* simply n. However, if the binding stoichiometry is known (or guessed) by other means, then, at least for simple systems, the filter assay can yield reliable values for binding constants and can be a helpful tool for characterizing cooperativity as well.

Acknowledgments

We thank Dr. David Draper for helpful discussions on the theory of stoichiometric binding experiments and Dr. Robert Winter for helpful conversations on *lac* filter-binding experiments. We are grateful to Dr. Steve Kowalczykowski for a critical reading of the manuscript.

Registry No. Nitrocellulose, 9004-70-0.

References

Ackers, G. K., Johnson, A. D., & Shea, M. A. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 1129–1133.

Bailey, J. M. (1979) *Anal. Biochem. 93*, 204–206.

Clore, G. M., Gronenborn, A. M., & Davies, R. W. (1982) *J. Mol. Biol. 155*, 447–466.

Draper, D. E., & von Hippel, P. H. (1978) *J. Mol. Biol. 122*, 339–359.

Freeman, E. J., & Jones, O. W. (1967) *Biochem. Biophys. Res. Commun. 29*, 45–52.

Gilbert, W., & Maxam, A. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 3581–3584.

Hinkle, D. C., & Chamberlin, M. J. (1972) *J. Mol. Biol. 70*, 157–185.

Johnson, G. G., & Geiduschek, E. P. (1977) *Biochemistry 16*, 1473–1485.

Johnson, A. D., Meyer, B. J., & Ptashne, M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 5061–5065.

Johnson, A. D., Poteete, A. R., Lauer, G., Sauer, R. T., Ackers, G. K., & Ptashne, M. (1981) *Nature (London) 294*, 217–223.

Kowalczykowski, S. C., Bear, D. G., & von Hippel, P. H. (1981) *Enzymes, 3rd Ed. 15*, 373–444.

McGhee, J. D., & von Hippel, P. H. (1974) *J. Mol. Biol. 86*, 469–489.
McQuistan, R. B. (1968) *Nuovo Cimento 58*, 86–92.
Pirrotta, V., Chadwick, P., & Ptashne, M. (1970) *Nature (London) 227*, 41–44.
Ptashne, M., Backman, K., Humayun, M. Z., Jeffrey, A., Maurer, R., Meyer, B., & Sauer, R. T. (1976) *Science (Washington, D.C.) 194*, 156–161.
Riggs, A. D., Suzuki, H., & Bourgeois, S. (1970) *J. Mol. Biol. 48*, 67–83.
Schellman, J. A. (1975) *Biopolymers 14*, 999–1018.
Schmitz, A., & Galas, D. J. (1979) *Nucleic Acids Res. 6*, 111–137.
Schmitz, A., & Galas, D. J. (1980) *Nucleic Acids Res. 8*, 487–506.
Strauss, H. S., Burgess, R. R., & Record, M. T., Jr. (1980) *Biochemistry 19*, 3496–3504.
Strauss, H. S., Boston, R. S., Record, M. T., Jr., & Burgess, R. R. (1981) *Gene 13*, 75–87.
Thang, M. N., Graffe, M., & Grunberg-Manago, M. (1968) *Biochem. Biophys. Res. Commun. 31*, 1–8.
Wyman, J. (1948) *Adv. Protein Chem. 4*, 407–531.
Wyman, J. (1964) *Adv. Protein Chem. 19*, 223–286.
Yarus, M., & Berg, P. (1967) *J. Mol. Biol. 28*, 479–490.
Yarus, M., & Berg, P. (1970) *Anal. Biochem. 35*, 450–465.

Cooperative, Excluded-Site Binding and Its Dynamics for the Interaction of Gene 5 Protein with Polynucleotides†

Dietmar Pörschke* and Hannelore Rauh

ABSTRACT: The binding of gene 5 protein to various single-stranded polynucleotides is investigated by fluorescence titrations and stopped-flow measurements. The association state of gene 5 protein itself is analyzed by equilibrium sedimentation: the monomer–dimer equilibrium found in the micromolar concentration range is described by a stability constant of 8×10^5 M^{-1}. The fluorescence quenching upon binding to polynucleotides, studied over a broad concentration range and analyzed in terms of a cooperative excluded-site binding model, provides binding constants for "isolated" and for "cooperative" sites. The cooperativity for various ribo- and deoxyribopolymers is between 400 and 800 and is virtually independent of the ionic strength. The binding to isolated sites is strongly dependent upon the ionic strength; analysis in terms of polyelectrolyte theory indicates the compensation of 4 ± 0.5 charges upon complex formation. The number of nucleotide residues covered by one protein molecule is also found to be 4 ± 0.5 units. The affinity of gene 5 protein for polynucleotides increases in the series poly(C) < poly(dA) < poly(A) < poly(U) ≪ poly(dT); the binding constant for poly(dT) is roughly a factor of 1000 higher than that for the other polymers. Model studies with Lys-Tyr-Lys and Lys-Trp-Lys suggest that the preferential interaction with poly(dT) is not simply due to enhanced stacking interactions between the aromatic amino acids and the thymine residues. Stopped-flow reaction curves obtained by mixing of gene 5 protein with poly(dT) in the micromolar concentration range show three relaxation processes with time constants between 1 ms and 1 s. From the concentration dependence, the fast process is assigned to the formation of isolated complexes with a rate constant of 8.1×10^9 M^{-1} (polynucleotide) s^{-1}. The medium and the slow process observed at moderate degrees of binding are partly due to excluded-site effects; the medium and slow process observed at low degrees of binding are assigned to some conformational transition and rearrangement of ligands on the lattice, respectively. This interpretation is supported by Monte Carlo simulations of the binding process using the approach developed by Epstein [Epstein, I. R. (1979) *Biopolymers 18*, 2037–2050]. Stopped-flow data obtained for the "weakly" binding poly(A) are consistent with a nucleation mechanism. The formation of cooperatively bound protein clusters can be explained without postulating a special translocation mechanism of the protein along the polynucleotide chain. A possible misinterpretation of data for the kinetics of nucleation with respect to fast translocation is discussed. However, evidence for some translocation process is obtained from the kinetics of binding close to saturation of the polymer lattice. The redistribution of cooperative clusters is faster than expected from Monte Carlo simulations. According to these results, the clusters are translocated in the time range of approximately a second.

Gene 5 protein is one of the best characterized DNA binding proteins (Kowalczykowski et al., 1981). It has been investigated by numerous techniques including NMR (O'Connor & Coleman, 1982; Alma et al., 1982), fluorescence spectroscopy (Pretorius et al., 1975; Veiko et al., 1981), CD (Day, 1973; Anderson et al., 1975; Tyaglov et al., 1980), ultracentrifugation (Cavalieri et al., 1976; Pretorius et al., 1975), X-ray crystallography (McPherson et al., 1980), and neutron scattering (Torbet et al., 1981; Gray et al., 1982). From these investigations, it is known that gene 5 protein binds cooperatively to single-stranded polynucleotides and due to this binding reduces the melting temperature of DNA by about 40 °C. The association seems to be mainly driven by stacking interactions between tyrosine and phenylalanine residues of the protein with base residues of the polynucleotides as well as electrostatic interactions between lysine and arginine residues of the protein with phosphate residues of the polynucleotides. The structure of the complex is not yet known in detail; various models have been suggested from NMR and chemical modification data (Coleman & Armitage, 1978), from the crystal structure of the free protein together with some data on oligonucleotide–

† From the Max Planck Institut für biophysikalische Chemie, 3400 Göttingen, West Germany. *Received February 22, 1983.*

0006-2960/83/0422-4737$01.50/0