Analytic Binding Isotherms Describing Competitive Interactions of a Protein Ligand with Specific and Nonspecific Sites on the Same DNA Oligomer

Oleg V. Tsodikov,* Jill A. Holbrook,† Irina A. Shkel,* and M. Thomas Record, Jr.*†
*Departments of Chemistry and †Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin 53706 USA

ABSTRACT Many studies of specific protein-nucleic acid binding use short oligonucleotides or restriction fragments, in part to minimize the potential for nonspecific binding of the protein. However, when the specificity ratio is low, multiple nonspecifically bound proteins may occupy the region of DNA corresponding to one specific site; this situation was encountered in our recent calorimetric study of binding of integration host factor (IHF) protein to its specific 34-bp H′ DNA site. Here, beginning from the analytical McGhee and von Hippel infinite-lattice nonspecific binding isotherm, we derive a novel analytic isotherm for nonspecific binding of a ligand to a finite lattice. This isotherm is an excellent approximation to the exact factorial-based Epstein finite lattice isotherm even for short lattices and therefore is of great practical significance for analysis of experimental data and for analytic theory. Using this isotherm, we develop an analytic treatment of the competition between specific and nonspecific binding of a large ligand to the same finite lattice (i.e., DNA oligomer) containing one specific and multiple overlapping nonspecific binding sites. Analysis of calorimetric data for IHF-H′ DNA binding using this treatment yields enthalpies and binding constants for both specific and nonspecific binding and the nonspecific site size. This novel analysis demonstrates the potential contribution of nonspecific binding to the observed thermodynamics of specific binding, even with very short DNA oligomers, and the need for reverse (constant protein) titrations or titrations with nonspecific DNA to resolve specific and nonspecific contributions. The competition treatment is useful in analyzing low-specificity systems, including those where specificity is weakened by mutations or the absence of specificity factors.

INTRODUCTION

Nucleic acid binding proteins generally exhibit two modes of interaction with nucleic acids: specific and nonspecific. Even though in vitro studies of specific binding typically use short DNA oligomers, nonspecific binding can be a significant contributor to binding equilibria for proteins whose specificity ratio is not very large, as expected for protein or DNA variants with reduced specific binding affinity. In calorimetric measurements of binding of wild-type integration host factor (IHF) to a DNA oligomer (Holbrook et al., 2001), equal in size and sequence to its H′ site, we observed an unusual three-phase character of reverse titrations, caused by competition between nonspecific and specific IHF binding to the same DNA fragment. The analysis of the thermodynamics of competitive nonspecific and specific binding to the same DNA oligomer (finite lattice) is therefore important but has not previously been worked out. Nonspecific binding isotherms for binding one or more types of large ligands to infinite and finite lattices of overlapping sites have been obtained previously. McGhee and von Hippel (MvH) (1974) showed the crucial importance of the correct definition of a nonspecific site and the interplay between the binding constant and the overlap site size in determining occupancy of a polymeric (infinite) DNA lattice by noncooperative (or cooperative) nonspecific binding. Their analytic binding isotherm is an invaluable tool for analysis of nonspecific binding of ligands to macromolecular DNA (Jezewska et al., 2001; Lundback et al., 1998; Padmanabhan et al., 1997; McAfee et al., 1996; Veal and Rill, 1991; Mascotti and Lohman, 1990; Bujalowski et al., 1989). Epstein (1978) derived a series expression representing the binding isotherm for nonspecific binding of a large ligand to a finite lattice and discussed the approach to infinite lattice behavior with increasing lattice length. The Epstein isotherm is an exact result, but it requires explicit enumeration of all possible states of bound ligands on the lattice, which does not simplify to a closed-form expression. This finite lattice isotherm has been used in the analyses of binding of large ligands to short DNA oligomers where the infinite-lattice limit is not accurate (Zhang et al., 1999; Jezewska and Bujalowski, 2000). Although MvH and Epstein isotherms model only one mode of binding of a given ligand to a lattice, they also serve a basis for more complex binding models, where more than one binding mode (Schwarz and Stankowski, 1979; Bujalowski et al., 1989; Rajendran et al., 1998) or a different mode of binding near the ends of a finite lattice (Munro et al., 2000) are considered.

Here, starting from the MvH infinite-lattice noncooperative isotherm, we derive an analytic expression for the end effect and show that introduction of this end effect makes the analytic MvH isotherm an excellent approximation to finite lattice behavior at any binding density, even for lattices as short as twice the nonspecific (overlap) site size. The resulting analytic finite lattice isotherm demonstrates clearly the roles of lattice length, ligand size, and the bind-

Received for publication 21 May 2001 and in final form 5 July 2001.

Address reprint requests to Dr. M. Thomas Record, Jr., 433 Babcock Drive, Madison, WI 53706. Tel.: 608-262-5332; Fax: 608-262-3453; E-mail: record@chem.wisc.edu.

0006-3495/01/10/1960/10 $2.00

ing constant in determining the shape of the isotherm and the occupancy at given concentrations of ligand and lattice. It also provides a route to incorporate ligand binding to a finite nucleic acid lattice into an analytic description of ligand effects on processes involving DNA, RNA, and other biopolymers such as heparin. As an added benefit, without any significant loss of accuracy in most cases, this isotherm is much easier to use in iterative calculations and data analysis than the exact finite-lattice isotherm cast in the form of a finite sum of factorial-based terms (Epstein, 1978).

We apply the analytic isotherm to analyze ligand binding to a finite DNA lattice containing one specific site but multiple potential nonspecific sites. We derive binding isotherms for both specifically and nonspecifically bound populations of ligand species for two situations, in which either very short or long regions of nonspecific DNA flank the specific site. We show that conventional (forward) titrations at constant DNA concentration are not sufficient for the analysis of the binding parameters where nonspecific binding is significant. Only titrations at constant protein concentration (reverse titrations) provide sufficient information to deconvolute the thermodynamics (enthalpies and binding constants) of both specific and nonspecific binding and obtain the nonspecific site size. The analysis is applied to isothermal titration calorimetry (ITC) studies of binding of IHF to its 34-bp specific (H′) site on a DNA of this length performed in this laboratory (Holbrook et al., 2001). We show that both nonspecific and specific binding of IHF make significant contributions to the enthalpic effects observed in forward (constant DNA) and especially reverse (constant IHF) ITC titrations, even for a DNA oligomer with the length and sequence of the specific H′ site. This effect should be observed with any nucleic acid binding proteins for which the ratio of specific to nonspecific binding constants (specificity ratio, K_{sp}/K_{ns}) is moderate and does not greatly exceed the ratio of site sizes (s/n) for these binding modes. Systems with large deviations from the consensus nucleic acid sequence and/or the wild-type protein site should be especially prone to this behavior.

Analytical Theory and Applications

Extension of the noncooperative McGhee-von Hippel nonspecific binding isotherm to oligomers (finite lattices)

We consider a process of ligand binding to a one-dimensional homogeneous lattice of N monomer units (e.g., base pairs or bases of a nucleic acid, as appropriate). In an elegant derivation using conditional probabilities, McGhee and von Hippel (1974) obtained the general analytical expression for the isotherm describing noncooperative binding of a large ligand to an infinitely long lattice ($N\rightarrow\infty$). Here we derive an analytic expression to extend their infinite lattice isotherm to finite lattices by incorporating end effects. For a ligand that occludes n lattice units in nonspecific binding to an infinite lattice, the McGhee and von Hippel (1974) result for the dependence of the nonspecific binding density ν_{ns} on the free ligand concentration $[L]$ is:

$$\frac{\nu_{ns}}{K_{ns}[L]} = (1 - n\nu_{ns})(ff)^{n-1}, \quad (1)$$

where the conditional probability ff is defined as

$$ff = \frac{1 - n\nu_{ns}}{1 - (n-1)\nu_{ns}}. \quad (2)$$

In Eqs. 1 and 2, ν_{ns} is defined as the average number of bound ligands per lattice unit, and K_{ns} is the equilibrium constant for binding to any potential site on a lattice. In Eq.1, the quantity $1 - n\nu_{ns}$ is the probability that a lattice unit is unoccupied, and $(ff)^{n-1}$ is the conditional probability of finding $n - 1$ free lattice units next to a free lattice unit on an infinite lattice. For finite lattices, the dependence of this conditional probability on binding density is a function of the lattice size N. However, to a good approximation, especially at low enough lattice saturation (ν_{ns}), we find that the infinite-lattice conditional probability $(ff)^{n-1}$ is applicable to all interior units on a finite lattice (see below), where interior is defined as being at least n lattice units from either end. For residues less than n units from either end of the lattice, the number of ways to find $n - 1$ lattice units next to it is smaller than for an interior residue.

There are $n - 1$ terminal lattice units at each end and $N - 2(n - 1)$ interior lattice units on a finite lattice. On an unoccupied lattice, any interior lattice unit is a part of n differently positioned free sites. If end-binding modes in which ligands partially overhang the ends of the lattice are disallowed, then any terminal lattice unit is a part of fewer than n differently positioned free sites. For example, the first and last lattice units are a part of only one free site; relative to interior lattice units, the probability of finding $n - 1$ free lattice units adjacent to a free lattice unit at the end of the lattice is $(ff)^{n-1}/n$. For the second lattice unit from either end, this probability is $2(ff)^{n-1}/n$ and so on. Summation of these weighted probabilities over all N lattice residues yields the average conditional probability of finding $n - 1$ free lattice units next to a free lattice unit on a finite lattice:

$$\frac{(N - 2(n-1))(ff)^{n-1} + 2\sum_{i=1}^{n-1}(ff)^{n-1}\frac{n-i}{n}}{N} = \frac{N - n + 1}{N}(ff)^{n-1}.$$

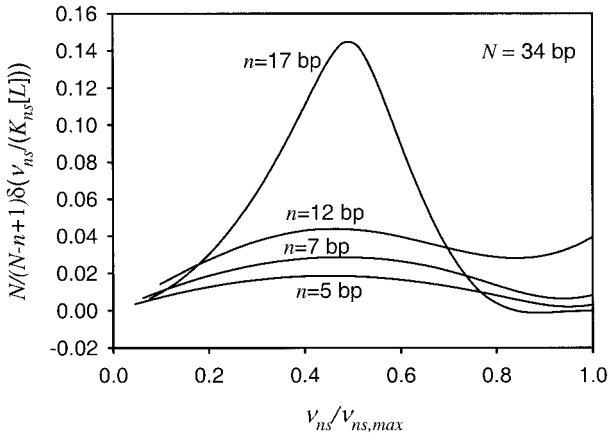

FIGURE 1 Difference between the analytic (Eq. 3) and exact numerical (Epstein, 1978) results for the average fraction of free lattice sites, $N/(N - n + 1)\delta(\nu_{ns}/K_{ns}[L]) \equiv N/(N - n + 1)\,[(\nu_{ns}/K_{ns}[L])_{Eq.3} - (\nu_{ns}/K_{ns}[L])_{Epstein}]$, for $N = 34$ bp and $n = 5$ (cf. Fig. 3), 7, 12, and 17 bp.

Therefore, the isotherm for nonspecific binding of a large ligand to finite lattices of N monomer units is

$$\frac{\nu_{ns}}{K_{ns}[L]} = (1 - n\nu_{ns})(ff)^{n-1}\left(\frac{N - n + 1}{N}\right), \qquad (3)$$

which approaches the infinite lattice result (Eq. 1) as N increases.

Using numerical calculations, Epstein (1978) confirmed the estimate of McGhee and von Hippel (1974) that the binding density calculated using the infinite lattice isotherm given by Eq. 1 differs from the finite lattice result by a factor of $(N - n + 1)/N$ in the limit $\nu_{ns} \rightarrow 0$. The above derivation shows how this factor arises from end effects on nonspecific binding to a finite lattice. Even for quite short lattices ($N/n > 2$), use of the $(N - n + 1)/N$ correction to the MvH isotherm yielded values of ν_{ns} that deviate by less than 10% from the numerical finite lattice value at $\nu_{ns} < 0.15/n$ (i.e., <15% saturation) (Epstein, 1978). Below, we demonstrate that Eq. 3 is an excellent approximation to the exact Epstein isotherm over the entire range of experimentally significant binding densities, even for the shortest possible lattices, and hence is of great practical significance.

For lattices of various sizes, Fig. 1 shows the difference between the average fraction of free lattice sites, $[N/(N - n + 1)][\nu_{ns}/(K_{ns}[L])]$, calculated using Eq. 3 and the exact value of this quantity calculated from the numerical Epstein isotherm. For all lattices, the deviation is small at low binding densities because Eq. 3 is exact at $\nu_{ns} = 0$. At high binding densities (where ν_{ns} approaches $1/n$), the relative deviation of $[N/(N - n + 1)][\nu_{ns}/(K_{ns}[L])]$ is large because the average number of free lattice sites is small, but the absolute deviation, which is the experimentally relevant quantity (see below), is small (Fig. 1). The deviation is most significant near half-saturation of the lattice, i.e., at $\nu_{ns} \approx 1/(2n)$. Fig. 1 shows that even for a very short finite lattice that can accommodate only two ligands per lattice (calculated for the case $N = 34$, $n = 12$, but generally observed for $N > 2.5n$), the largest difference does not exceed the usual experimental error (5%). For the other two lattices shown in Fig. 1 ($N/n = 5$ and $N/n = 7$), the difference in the fraction of free lattice sites is even smaller. For the situation $N = 2n$, corresponding to the shortest lattice where this analysis is required (e.g., $N = 34$, $n = 17$) and where the error in the analytic expression is largest, the absolute error in the fraction of free lattice sites does not exceed 15%. (For $N < 2n$, only one ligand can be bound to the lattice, and therefore one does not need to use a lattice isotherm analysis.) Even in this case, Eq. 3 remains a good approximation at low and high binding densities. This largest deviation for this lattice is experimentally significant only in titrations, where half-saturation of the lattice ($\nu_{ns} \approx 1/(2n)$) occurs in the midrange of the titration curve. This result is illustrated by simulated reverse titrations ($[L]_{tot}$ = constant) in Fig. 2. In a typical equilibrium binding experiment (Fig. 2 *A*), where $[L]_{tot} \ll (K_{ns})^{-1}$, measurable fraction of bound ligand is achieved only at low binding densities ($\nu_{ns} < 1/3n$), i.e., in large excess of lattice sites over ligand. In this regime, Eq. 3 is an excellent approximation even for the shortest lattice. In the tight binding regime (Fig. 2 *C*), where $[L]_{tot} \gg (K_{ns})^{-1}$, Eq. 3 is in very good agreement with the Epstein isotherm up to the region of high ligand saturation, where the discrepancy is difficult to observe experimentally. Most of this titration is at high binding densities, i.e., at $\nu_{ns} \approx 1/n$. In the intermediate regime (Fig. 2 *B*), where $[L]_{tot} \sim (K_{ns})^{-1}$, the midpoint of the titration corresponds to half-saturation of the lattice. Therefore, even for an extremely short lattice, the deviation between the analytic isotherm given by Eq. 3 and the exact theory is significant only in a narrow concentration regime and does not exceed 10–15%, as indicated above.

For the longest lattice considered in Fig. 1 (where $N = 34$ bp and $n = 5$ bp; $N/n \approx 7$), which are the parameters in the study of nonspecific binding of IHF to the 34-bp H′ site at low [salt] (Holbrook et al., 2001), the Scatchard plot shown in Fig. 3 illustrates that Eq. 3 is an excellent approximation to the numerical Epstein isotherm over the entire range of binding densities. Without correction for the end effect, the infinite lattice isotherm is not accurate in this case (cf. Fig. 3).

As an example, we apply the finite lattice isotherm described by Eq. 4 to the analysis of previously reported reverse titration fluorescence data for oligocation (L^{8+}) binding to single-stranded DNA oligonucleotides of three different sizes ($N/n \approx 8$, $N/n \approx 5$, and $N/n \approx 2$) (Zhang et al., 1999). These data were originally analyzed numerically with the Epstein isotherm to quantify coulombic end effects on oligocation-oligoanion binding (Zhang et al., 1999). Nonlinear regression analysis of the data (Johnson and Frasier, 1985) using Eq. 3 shown in Fig. 4, yields best-fit parameters that are in excellent agreement with those per-

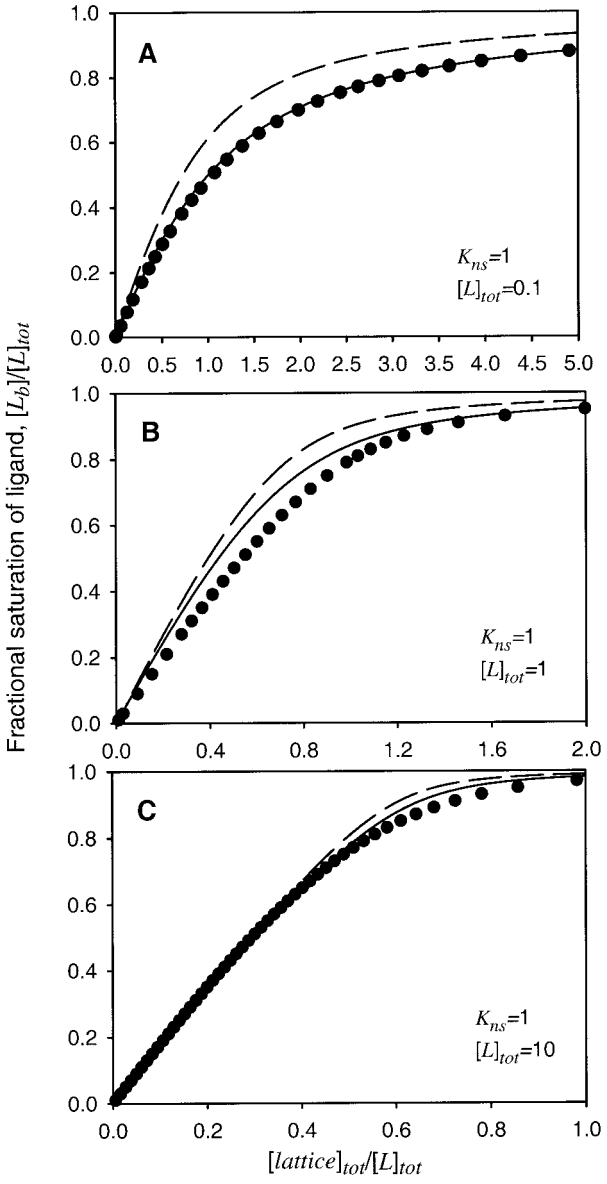

FIGURE 2 Simulated reverse (constant $[L]_{tot}$) titrations for nonspecific ligand binding to the shortest lattice. Fraction saturation of ligand ($[L_b]/[L]_{tot}$) is plotted versus the excess of DNA lattices over ligand ($[lattice]_{tot}/[L]_{tot}$) using the exact Epstein isotherm (●), Eq. 3 (——), and the MvH infinite lattice isotherm (— — —) for (N = 34, n = 17) with the binding constant K_{ns} = 1 (arbitrary units) and total ligand concentrations $[L]_{tot}$ = 0.1 (*A*); $[L]_{tot}$ = 1 (*B*); and $[L]_{tot}$ = 10 (*C*) (arbitrary units).

formed using the exact Epstein isotherm (Zhang et al., 1999). As expected, for $N/n \approx 8$ and $N/n \approx 5$ (Fig. 4, *A* and *B*), Eq. 3 is a good approximation over the entire range of binding densities. For the shortest oligonucleotide ($N/n \approx 2$; Fig. 4 *C*), where the data are at low binding densities, a value of n cannot be obtained from the fit even to the exact numerical Epstein isotherm. By fixing $n = 8$ as determined from experiments with larger oligomers, we obtain the fit shown in Fig. 4 *C*, which yields a value of K_{ns} that is in a

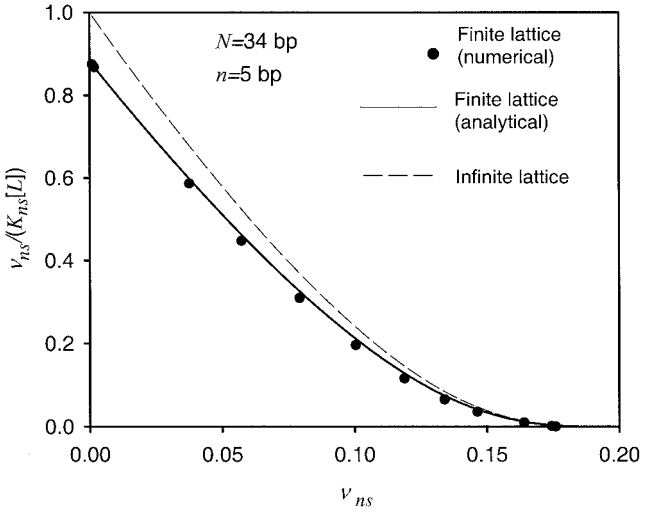

FIGURE 3 Scatchard plot comparing the infinite lattice McGhee-von Hippel isotherm (Eq. 1; — — —), analytic finite lattice isotherm (Eq 3; ——), and the exact numerical finite-lattice isotherm (Epstein, 1978; ●) for nonspecific binding of IHF to 34-bp DNA (N = 34 bp, n = 5 bp) (Holbrook et al., 2001).

very good agreement with that reported by Zhang et al. (1999) (see Fig. 4), as expected for this regime ($[L]_{tot} \ll (K_{ns})^{-1}$, Fig. 3 *A*). Most of the data for the shortest oligonucleotide fall in the range $\nu_{ns} < 1/3n$ and therefore are well modeled by Eq. 3.

Therefore, Eq. 3 is applicable to the analysis of experimental data even for lattice lengths that do not greatly exceed the size of the ligand. Because Eq. 3 does not require that n be an integer, it is of more practical use in nonlinear regression data analysis, where n is a fitting parameter, than the exact factorial-based Epstein isotherm, which requires that n be an integer. (Nearest-neighbor anticooperativity or low levels of positive cooperativity are parameterized in the MvH isotherm by noninteger n. Where cooperativity is significant, one should use cooperativity analysis (see Appendix), which separates the cooperativity factor ω from n.) In the following section, we incorporate Eq. 3 into the analysis of the situation in which a ligand binds both specifically and nonspecifically to the same lattice. In the study by Holbrook et al. (2001) we apply this analysis to ITC studies of binding of IHF to the H′ site and to nonspecific DNA.

Binding of ligand to a finite homogeneous lattice containing a specific site

Here we obtain an approximate analytic isotherm for ligand binding to a finite lattice containing one specific site and multiple nonspecific sites and develop two mathematically tractable applications of particular relevance for experimental studies of protein binding to synthetic oligonucleotides or restriction fragments.

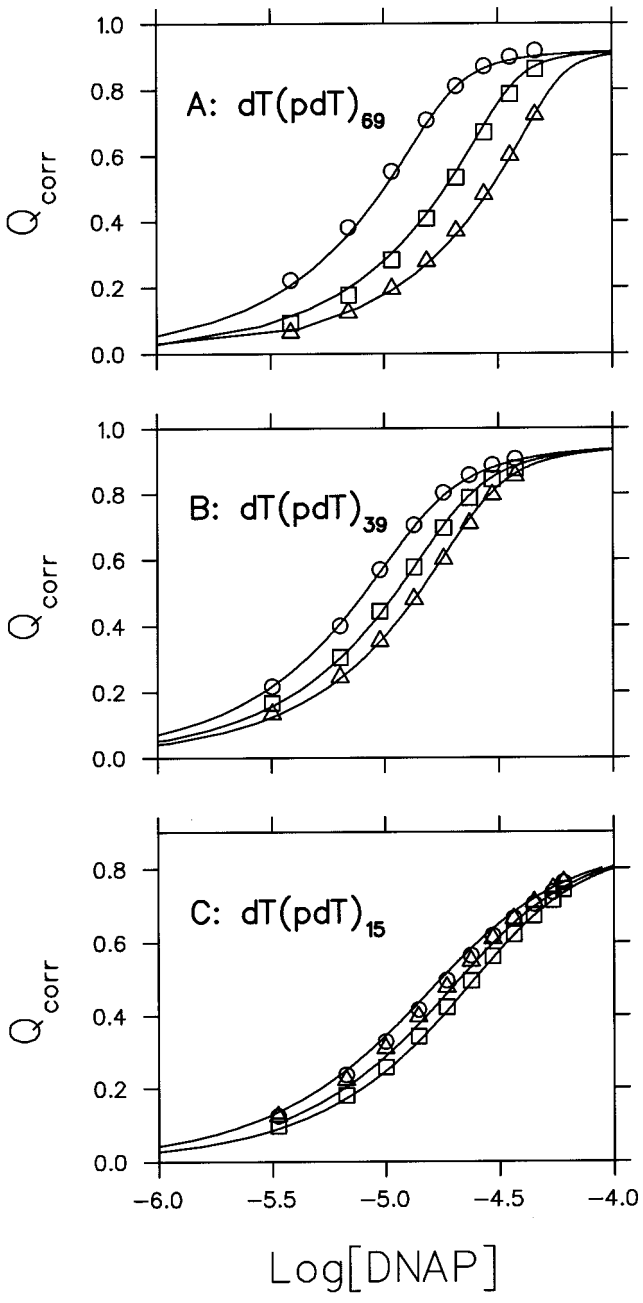

FIGURE 4 Binding isotherms obtained at constant ligand (KWK_6-NH_2) concentrations obtained by monitoring the fluorescence quenching (Q_{corr}) as a function of the DNA phosphate concentrations (data previously reported in Zhang et al., 1999). (*A*) KWK_6-NH_2 binding to $dT(pdT)_{69}$ at initial oligopeptide concentrations of 1.49 μM (○), 2.97 μM (□), and 4.95 μM (△) (correction of the erroneously reported values from Zhang et al., 1999). The solid lines are the best-fit curves to Eq. 3 with $K_{ns} = (2.0 \pm 0.5) \times 10^6\ M^{-1}$, $n = 8.5 \pm 0.2$, and $Q_{max} = 0.92 \pm 0.02$. (*B*) KWK_6-NH_2 binding to $dT(pdT)_{39}$ at initial oligopeptide concentrations of 0.99 μM (○), 1.49 μM (□), and 1.98 μM (△). The solid lines are the best-fit curves to Eq. 3 with $K_{ns} = (9.5 \pm 1.6) \times 10^5\ M^{-1}$, $n = 8.5 \pm 0.2$, and $Q_{max} = 0.95 \pm 0.02$. (*C*) KWK_6-NH_2 binding to $dT(pdT)_{15}$ at initial oligopeptide concentrations of 0.99 μM (○), 1.49 μM (□), and 1.98 μM (△). The solid lines are the best-fit curves to Eq. 3 with $K_{ns} = (2.5 \pm 0.3) \times 10^5\ M^{-1}$ and $Q_{max} = 0.88 \pm 0.02$ (due to low binding density range of the data in this panel, n could not be resolved by the fit and was fixed at $n = 8$ to facilitate comparison with Zhang et al., 1999). Fitting of this data set to the Epstein isotherm yielded $K_{ns} = (2.1 \pm 0.2) \times 10^5\ M^{-1}$ (Zhang et al., 1999).

Case 1

A synthetic DNA oligomer with the length and the sequence of the specific binding site is investigated. In this case, the lattice size (N) is the same as the specific binding site size (s), whereas the nonspecific site size (n) is much less than both N and s ($n \ll s = N$). In this case, specific binding of one ligand eliminates all nonspecific binding sites on that lattice. On the other hand, when the specific site is vacant, multiple ligands can be bound nonspecifically. In this case, the binding density of nonspecifically bound ligand (ν_{ns}) is

$$\nu_{ns} = \frac{[L_{b,ns}]}{[D_{ns}]_{tot} N}, \tag{4}$$

where $[L_{b,ns}]$ is the concentration of nonspecifically bound ligand. In Eq. 4, $[D_{ns}]_{tot} = [D]_{tot} - [D_{b,sp}]$ is the concentration of DNA molecules in the form of finite lattices of nonspecific sites, i.e., all DNA molecules except those occupied by the specifically bound ligand ($D_{b,sp}$). The relationship between the binding density (ν_{ns}) and the free ligand concentration ([L]) for a sufficiently large finite lattice is given by Eq. 3. The concentrations of specifically bound species are given by the specific 1:1 binding equilibrium,

$$[L_{b,sp}] = [D_{b,sp}] = K_{sp}[L][D_{sp}], \tag{5}$$

where $[D_{sp}]$ is the concentration of free specific sites. The mass balance equation for ligand is

$$[L]_{tot} = [L] + [L_{b,ns}] + [L_{b,sp}]. \tag{6}$$

The final required equation is the relationship between the concentration of free specific DNA sites ($[D_{sp}]$) and the average binding density of nonspecifically bound ligand. In this case, a specific site is free if no ligands are bound to the lattice (either specifically or nonspecifically). The fraction of DNA lattices without nonspecifically bound ligands is the product of the probability that a randomly chosen lattice unit is free ($1 - n\nu_{ns}$) and the weighted probability that there are $N - 1$ free lattice units adjacent to it $(ff)^{N-1}$:

$$[D_{sp}] = [D_{ns}]_{tot}(1 - n\nu_{ns})(ff)^{N-1}. \tag{7}$$

In deriving Eq. 7, the weighting coefficient in front of $(ff)^{N-1}$ for each lattice unit is $1/N$ because there is only one free site of size N on an unoccupied lattice. This coefficient cancels when this result is expressed per lattice, i.e., after multiplying by the total number of lattice units, N.

From Eqs. 3–7, we obtain a closed-form expression relating ν_{ns} to $[L]_{tot}$ and $[D]_{tot}$ in terms of n, K_{ns}, and K_{sp} (Eq.

TABLE 1 Binding isotherm describing competition between specific and nonspecific binding to a short DNA oligomer ($N \approx s$)

Equation number	Equation
8	$[L]_{tot} = \frac{N\nu_{ns}}{K_{ns}(N-n+1)(ff)^{n-1}(1-n\nu_{ns})} + N[D]_{tot}\nu_{ns}\left(\frac{K_{ns}(N-n+1)+K_{sp}(ff)^{N-n}}{K_{ns}(N-n+1)+NK_{sp}\nu_{ns}(ff)^{N-n}}\right)$
9	$[L_{b,ns}] = N\nu_{ns}[D_{ns}]_{tot} = \frac{N\nu_{ns}[D]_{tot}}{1+\frac{NK_{sp}}{K_{ns}(N-n+1)}\nu_{ns}(ff)^{N-n}} = \frac{N\nu_{ns}[D_{sp}]}{(1-n\nu_{ns})(ff)^{N-1}}$
10	$[L_{b,sp}] = \frac{K_{sp}[D]_{tot}N\nu_{ns}(ff)^{N-n}}{K_{ns}(N-n+1)+NK_{sp}\nu_{ns}(ff)^{N-n}}$

8 in Table 1). In a data-fitting procedure one solves Eq. 8 for ν_{ns} in each fitting iteration for a given trial set of fitting parameters (n, K_{ns}, and K_{sp}) and experimental input values ($[L]_{tot}$ and $[D]_{tot}$). This ν_{ns} is then substituted in Eqs. 5–7 to yield the concentrations of specifically and nonspecifically bound ligand and DNA sites (Eqs. 9 and 10 in Table 1). These concentrations are used to analyze experimentally observed quantities such as heats of binding in calorimetric measurements (Holbrook et al., 2001) or fluorescence changes upon binding (Zhang et al., 1999).

Case 2

A restriction fragment containing the specific site embedded in flanking nonspecific DNA is investigated. In this case, the lattice size (N) is significantly greater than the specific site size (s) so that even when a ligand is specifically bound, the flanking regions behave as sufficiently long finite lattices for nonspecific binding ($n \ll N - s$). The specific site is centrally located with flanking sequences of lengths N_{f1} and N_{f2}, each much larger than s. Here, for the specific site to be free, a stretch of s or more consecutive free lattice units should be positioned correctly to contain the specific site. If an arbitrary free lattice residue is chosen i residues from one end, then the conditional probability that the gap that contains this residue also contains the whole specific site is $(ff)^{(N_{f1}-i)+s-1}$, $0 \le i \le N_{f1}$. Analogously, for the other end, this probability is $(ff)^{(N_{f2}-i)+s-1}$, $0 \le i \le N_{f2}$. Therefore, the estimate of the concentration of free specific sites (analogous to Eq. 7 in case 1) should be recast as a summation along the length of the lattice:

$$[D_{sp}] = [D_{ns}]_{tot}\frac{1}{N}\left[\sum_{i=0}^{N_{f1}-1}(1-n\nu_{ns})(ff)^{(N_{f1}-i)+s-1} + \sum_{i=0}^{N_{f2}-1}(1-n\nu_{ns})(ff)^{(N_{f2}-i)+s-1}\right] = [D_{ns}]_{tot}\lambda, \quad (11)$$

where summation of the geometric series in Eq. 11 yields the closed-form result for λ given in Table 2. The mass balance for DNA is $[D_{ns}]_{tot} = [D]_{tot} - [D_{b,sp}]$, as in case 1. The mass balance equation for ligand species in this case is the same as Eq. 6. The difference is in the interpretation of the concentration of nonspecifically bound species, because in addition to the nonspecific binding in the region of the specific site, binding also occurs to flanking sequences on fragments with a specifically bound ligand:

$$\nu_{ns} = \frac{[L_{b,ns}]}{[D_{ns}]_{tot}N + (N_{f1}+N_{f2})[D_{b,sp}]}. \quad (12)$$

The nonspecific binding density ν_{ns} (Eq. 12) is related to the nonspecific site size and the binding constant by Eq. 3. In

TABLE 2 Binding isotherm describing competition between specific and nonspecific binding to a restriction fragment ($N = s + N_{f1} + N_{f2} \gg s$)

Equation number	Equation
13	$[L]_{tot} \approx \frac{N\nu_{ns}}{K_{ns}(N-n+1)(ff)^{n-1}(1-n\nu_{ns})} + [D]_{tot}\frac{N\nu_{ns}+(N\nu_{ns}+1)K_{sp}\lambda[L]}{1+\lambda K_{sp}[L]}$ $\lambda = \frac{1}{N}(1-n\nu_{ns})\frac{2-(ff)^{N_{f1}}-(ff)^{N_{f2}}}{1-(ff)}$
14	$[L_{b,ns}] \cong \nu_{ns}N[D]_{tot}$
15	$[L_{b,sp}] = \frac{K_{sp}\lambda[L][D]_{tot}}{1+K_{sp}\lambda[L]}$

many studies, using restriction fragments, the length of DNA is much longer than the specific site size ($N - s \approx N$ and $N_{f1} + N_{f2} \approx N$). Solving the system in this case yields Eq. 13 (Table 2). As in case 1, this equation can be solved numerically to obtain ν_{ns}. When ν_{ns} is known, the concentrations of various species can be found from the above expressions in the reverse order (Eqs. 14 and 15 in Table 2) to be used in a data-fitting procedure.

All isotherms developed above can be readily generalized for cooperative binding to a lattice, using the MvH cooperative isotherm. This generalization is described in the Appendix.

DISCUSSION

The analytic finite lattice isotherm describing nonspecific binding of a large ligand, developed here, is particularly useful in analyses of binding data obtained under conditions where nonspecific binding to a DNA lattice competes with specific binding to the same DNA. One example of such a low-specificity system is binding of IHF to DNA containing the H′ site, investigated by Holbrook et al. (2001).

Analysis of forward and reverse titrations for low-specificity systems: application to IHF-H′ DNA interactions

Holbrook et al. (2001) performed ITC measurements of IHF protein binding to a 34-bp duplex DNA containing a specific (H′) site. Both specific and nonspecific interactions of IHF are exothermic, with characteristically different enthalpies of specific and nonspecific binding. Consequently, the competition between specific and nonspecific binding can be observed directly in reverse titrations, i.e., when adding DNA to protein. The calorimetric signature of this competition is the characteristic shape of the calorimetric profile of the reverse titration, which contains three regimes and is much more complex (Fig. 5 *A*) than that expected for a simple 1:1 binding isotherm. The species distribution (Fig. 5 *B*) for this isotherm demonstrates how specific and nonspecific binding contributions vary with addition of DNA in a reverse titration. The initial phase corresponds to the formation of nonspecifically bound protein-DNA complexes, because the effect of the difference in site sizes ($n \ll s$) overwhelms the fact that nonspecific binding is intrinsically weaker ($K_{ns} \ll K_{sp}$) in large excess of protein over DNA. The specificity ratio (K_{sp}/K_{ns}) is indeed found by this analysis to be ~100 and at high $[IHF]_{tot}/[D]_{tot}$ is dominated by the ratio of nonspecific to specific sites on this DNA fragment (found to be 30:1 in base pairs or 60:1 in phosphates, for $n = 8$ and $N = 68$ phosphates). Therefore, at high enough saturation of protein with DNA, both specific and nonspecific contributions are measurable, as shown in Fig. 5 *B*. Even though the concentration of free

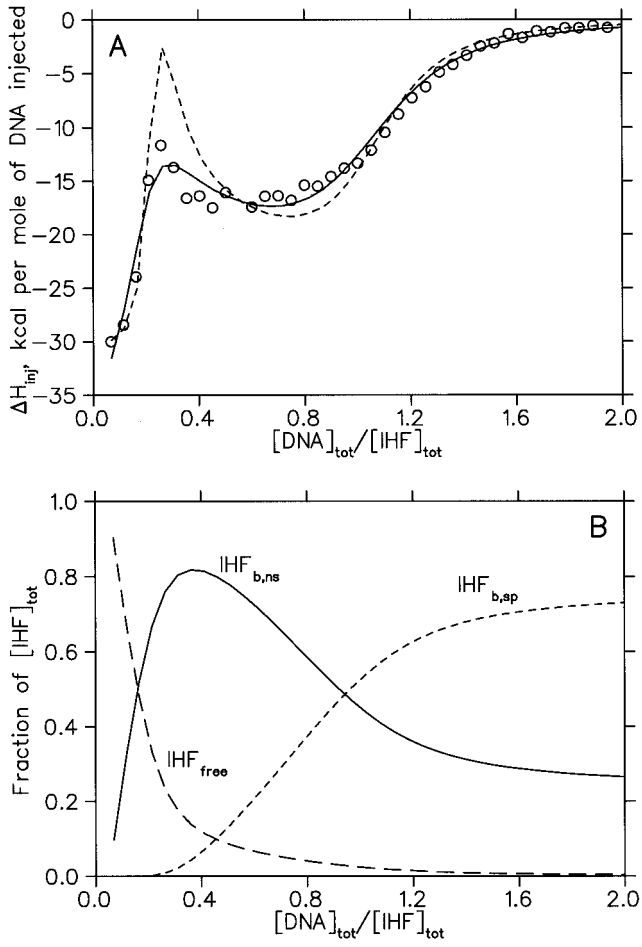

FIGURE 5 (*A*) Reverse ITC titration of 2.26×10^{-5} M IHF with 34-bp H′ DNA oligonucleotide. Heat of injection normalized per mole of injected DNA strands is plotted versus the molar ratio of DNA molecules to IHF. The titration is performed at 60 mM K^+ in phosphate buffer at 20°C (Holbrook et al., 2001). The solid curve is the best fit of these data to the model described by Eqs. 9–11 of this paper ($K_{ns} = 9.0 \times 10^4$ M^{-1}, $K_{sp} = 9.3 \times 10^6$ M^{-1}, and $n = 4.5$ bp (or 9 nucleotides; for details of the fitting see Holbrook et al., 2001)). The dashed line is the best fit to the data obtained by fixing $K_{sp} = 10^9$ M^{-1} and allowing the other fitting parameters to float. (*B*) Distribution of protein species (shown as a fraction of total IHF concentration in the reaction cell) in the titration shown in *A*.

DNA-specific sites is negligible throughout the experimentally achievable concentration range, the concentration of free nonspecific sites and the contribution of nonspecific binding are significant. As the ratio of DNA to protein concentrations increases, some fraction of protein molecules dissociates from nonspecific sites and binds to specific sites.

In addition, the strong competition between specific and nonspecific binding to the same fragment makes it possible to obtain a value (instead of a lower bound) for the specific binding constant. Despite the fact that the concentrations of both reactants are in the micromolar range, and equivalence is obtained at 23 μM, Fig. 5 *A* demonstrates that the specific

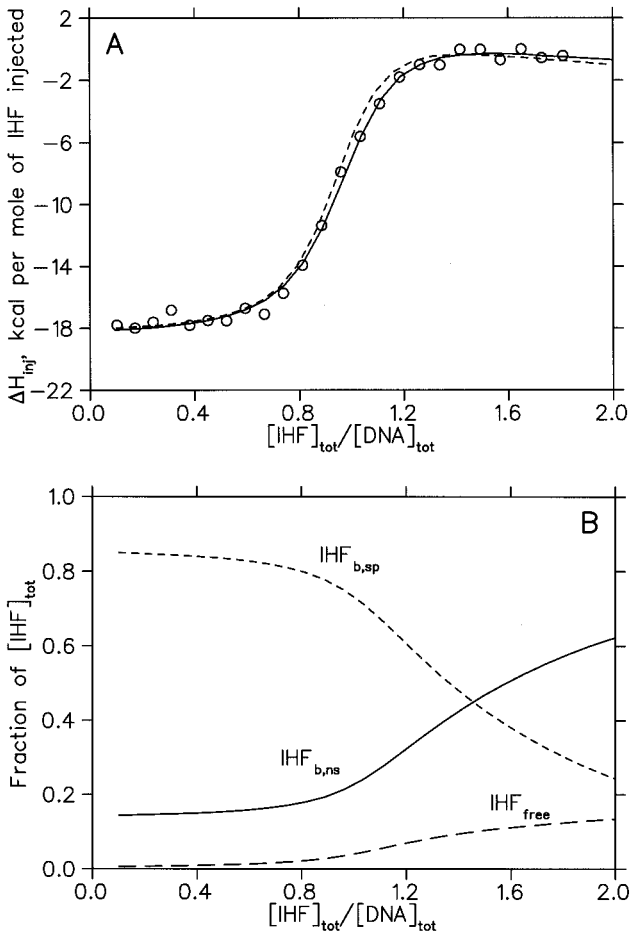

FIGURE 6 (*A*) Forward ITC titration of 10^{-5} M 34-bp H′ DNA with IHF. Heat of injection normalized per mole of injected active IHF is plotted versus the molar ratio of IHF to DNA molecules. Conditions are the same as in Fig. 3. The solid curve is the best fit of these data to the model described by Eqs. 9–11 of this paper ($K_{ns} = 8.0 \times 10^4$ M^{-1}, $K_{sp} = 1.4 \times 10^7$ M^{-1}, and $n = 4.5$ bp; for other parameters and the details of the fitting see Holbrook et al., 2001). The dashed line is the best fit to the data obtained by fixing $K_{sp} = 10^9$ M^{-1} and allowing the other fitting parameters to float. (*B*) Distribution of protein species (shown as a fraction of total IHF concentration in the reaction cell) in the titration shown in *A*.

binding constant ($K_{sp} = (9.2 \pm 0.9) \times 10^7$ M^{-1}; $K_{sp}^{-1} = 0.011 \pm 0.001$ μM) is indeed well determined by the fitting procedure. The full range of the reverse titration contains both a nonspecific binding titration and a competitive nonspecific-specific binding titration, which is a consequence of the relatively small specificity ratio (~100) and to the differences in site size ($N/n \approx 7$).

Fig. 6 *A* shows the forward titration (addition of protein to DNA) at the same conditions as Fig. 5. Although the concentrations of reactants are comparable (micromolar) and competition from nonspecific binding is therefore significant, the signature of competition is absent from these data. Therefore, if the information from a reverse titration is not available, one can erroneously fit these data to a 1:1 specific binding isotherm. The species distribution (Fig. 6 *B*) predicted from the fit to the data in Fig. 6 *A* demonstrates that the fraction of free protein (the titrant) is negligibly small in the range where the calorimetric enthalpy changes are observable. An equally good fit to the forward titration data (Fig. 6 *A*) is obtained when both binding constants are increased by a factor of 100, maintaining the same specificity ratio. This fit yields the same species distribution (with an undetectable free protein concentration). Therefore, fitting only forward titration data yields only lower bounds on the binding constants. Thus, the fit corresponding to the larger specific and nonspecific binding constants is statistically distinguishable from the best fit in the reverse but not in the forward titration. This illustrates the requirement for exploring the widest possible range of fractional saturation of both protein and DNA and shows that reverse titrations are necessary to extract both specific and nonspecific binding parameters from the data. Wong and Lohman (1995) showed that reverse titrations contain more thermodynamic information than forward titrations in the case when protein oligomerization is coupled to DNA binding. Our study shows that reverse titrations also contain more information than forward ones when nonspecific binding contributes significantly to the equilibrium species distribution.

Concluding discussion

The above analysis quantifies the two ways in which nonspecific binding can contribute in experiments where one measures specific binding.

1) If the DNA fragments used in the experiment are sufficiently short so that the sequences flanking the specific sites are very short or absent ($s \approx N$), and if the nonspecific site size is much smaller than the length of the fragment ($n \ll s$), then competition between nonspecific and specific binding for occupancy of the specific site is the dominant contribution. For two systems studied to date, the ratio of specific (s) to nonspecific (n) site size has been shown to be large: for *lac* repressor (Revzin and von Hippel, 1977), $s/n \approx 2$; and for IHF (Holbrook et al., 2001), $s/n \approx 8$ at low [salt]. We expect that other specific binding proteins will exhibit moderate-to-large site size (s/n) ratios, especially where large-scale protein or DNA conformational changes are coupled only to specific binding. For such systems, nonspecific binding will significantly affect the thermodynamics of specific protein-DNA binding even on short DNA fragments if the specificity ratio is relatively low. A related, extensively studied example is the binding of *Escherichia coli* single-strand binding protein (SSB) to a polynucleotide, where different modes with site sizes that differ by a factor of two are observed. In studies of SSB binding to $(dA)_{70}$, Ferrari et al. (1994) showed that $(dA)_{70}$ can bind either one SSB tetramer in a mode that occludes 65 nucleotides or one or two SSB tetramers in a mode that occludes only 35

nucleotides per tetramer. The binding mode switches from the larger to the smaller occluded site size with increasing binding density.

2) If the DNA sequences flanking the specific site are relatively long, as may be the in vivo situation in prokaryotes, then the effect of nonspecific binding is primarily to reduce the concentration of free protein; i.e., here the competition is reversed, and it is the DNA sites that compete for free protein. In a prokaryotic cell, the concentration of most DNA-binding proteins in the free state is thought to be very small; most of the protein is bound either specifically or nonspecifically. Most of intracellular *lac* repressor (Kao-Huang et al., 1977) and IHF (Murtin et al., 1998) exist as DNA-bound species. Therefore, the competition between specific and nonspecific binding modes probably plays an important regulatory role (von Hippel et al., 1974; Bremer and Dennis, 1996). In eukaryotes, chromatin structure constrains the average size of a DNA segment available for binding of trans-acting factors. Such small DNA targets are probably most closely approximated by finite lattices.

The analysis developed in this paper is generally applicable to binding studies in which the specificity ratio K_{sp}/K_{ns} is modest and/or where the site size ratio s/n is large. Mutations in proteins or specific DNA sites, the absence of specificity factors, or changes in solution conditions can greatly affect specificity or site size. Therefore, we predict that studies of mutations that significantly weaken protein-DNA interactions will require the analysis developed in this paper. Wild-type proteins such as IHF, which function in vivo in both specific and nonspecific contexts, naturally exhibit relatively low specificity, so that studies examining their binding to DNA also require the above analysis.

APPENDIX

Generalization of the analysis for cooperative binding

The formalism developed in the Results can be easily generalized to treat competition of specific and cooperative nonspecific binding. According to the treatment and the nomenclature of the cooperative case in McGhee and von Hippel (1974), in Eq. 3 and Eqs. 7–12, $(ff)^k$ should be multiplied by b^2, where

$$(ff) = \frac{(2\omega - 1)(1 - n\nu_{ns}) + \nu_{ns} - R}{2(\omega - 1)(1 - n\nu_{ns})},$$

$$b = \frac{1 - (n + 1)\nu_{ns} + R}{2(1 - n\nu_{ns})},$$

$$R = \sqrt{[1 - (n + 1)\nu_{ns}]^2 + 4\omega\nu_{ns}(1 - n\nu_{ns})},$$

and ω is the cooperativity parameter defined in McGhee and von Hippel (1974).

We thank Tim Lohman and the referees for their valuable comments on the manuscript and editor Paul Hagerman for suggesting the relevance of the finite lattice model for eukaryotic gene regulation.

This work was supported by National Institutes of Health grant GM23467.

REFERENCES

Bremer, H., and P. P. Dennis. 1996. Modulation of chemical composition and other parameters of the cell by growth rate. *In Escherichia coli* and Salmonella, 2nd ed, Vol. 2. ASM Press, Washington, D.C. 1553–1570.

Bujalowski, W., T. M. Lohman, and C. F. Anderson. 1989. On the cooperative binding of large ligands to a one-dimensional homogeneous lattice: the generalized three-state lattice model. *Biopolymers.* 28: 1637–1643.

Epstein, I. R. 1978. Cooperative and non-cooperative binding of large ligands to a finite one-dimensional lattice. A model for ligand-oligonucleotide interactions. *Biophys. Chem.* 8:327–339.

Ferrari, M. E., W. Bujalowski, and T. M. Lohman. 1994. Co-operative binding of *Escherichia coli* SSB tetramers to single-stranded DNA in the $(SSB)_{35}$ binding mode. *J. Mol. Biol.* 236:106–123.

Holbrook, J. A., O. V. Tsodikov, R. M. Saecker, and M. T. Record, Jr. 2001. Binding of integration host factor to specific and nonspecific DNA: thermodynamic evidence for disruption of multiple IHF surface salt bridges coupled to DNA binding. *J. Mol. Biol.* 310:379–401.

Jezewska, M. J., and W. Bujalowski. 2000. Interactions of *Escherichia coli* replicative helicase PriA protein with single-stranded DNA. *Biochemistry.* 39:10454–10467.

Jezewska, M. J., S. Rajendran, and W. Bujalowski. 2001. Interactions of the 8-kDa domain of rat DNA polymerase β with DNA. *Biochemistry.* 40:3295–3307.

Johnson, M. L., and S. G. Frasier. 1985. Nonlinear least-square analysis. *Methods Enzymol.* 117:301–342.

Kao-Huang, Y., A. Revzin, A. P. Butler, P. O'Conner, D. W. Noble, and P. H. von Hippel. 1977. Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: measurement of DNA-bound *Escherichia coli lac* repressor in vivo. *Proc. Natl. Acad. Sci. U.S.A.* 74:4228–4232.

Lundback, T., H. Hansson, S. Knapp, R. Ladenstein, and T. Hard. 1998. Thermodynamic characterization of non-sequence-specific DNA-binding by the Sso7d protein from *Sulfolobus solfataricus*. *J. Mol. Biol.* 276:775–786.

Mascotti, D. P., and T. M. Lohman. 1990. Thermodynamic extent of counterion release upon binding oligolysines to single-stranded nucleic acids. *Proc. Natl. Acad. Sci. U.S.A.* 87:3132–3136.

McAfee, J. G., S. P. Edmondson, I. Zegar, and J. W. Shriver. 1996. Equilibrium DNA binding of Sac7d protein from the hyperthermophile *Sulfolobus acidocaldarius*: fluorescence and circular dichroism studies. *Biochemistry.* 35:4034–4045.

McGhee, J. D., and P. H. von Hippel. 1974. Theoretical aspects of DNA-protein interactions: cooperative and noncooperative binding of large ligands to a one-dimensional homogeneous lattice. *J. Mol. Biol.* 86: 469–489.

Munro, P. D., C. M. Jackson, and D. J. Winzor. 2000. Consequences of the nonspecific binding of a protein to a linear polymer: reconciliation of stoichiometric and equilibrium titration data for the thrombin-heparin interaction. *J. Theor. Biol.* 203:407–418.

Murtin, C., M. Engelhorn, J. Geiselmann, and F. Boccard. 1998. A quantitative UV laser footprinting analysis of the interaction of IHF with specific binding sites: reevaluation of the effective concentration of IHF in the cell. *J. Mol. Biol.* 284:949–961.

Padmanabhan, S., W. T. Zhang, M. W. Capp, C. F. Anderson, and M. T. Record, Jr. 1997. Binding of cationic (+4) alanine- and glycine-containing oligopeptides to double-stranded DNA: thermodynamic analysis of effects of coulombic interactions and α-helix induction. *Biochemistry.* 36:5193–5206.

Rajendran, S., M. J. Jezewska, and W. Bujalowski. 1998. Human DNA polymerase β recognizes single-stranded DNA using two different binding modes. *J. Biol. Chem.* 47:31021–31031.

Revzin, A., and P. H. von Hippel. 1977. Direct measurement of association constants for the binding of *Escherichia coli lac* repressor to non-operator DNA. *Biochemistry.* 16:4769–4776.

Schwarz, G., and S. Stankowski. 1979. Linear cooperative binding of large ligands involving mutual exclusion of different binding modes. *Biophys. Chem.* 10:173–181.

Veal, J. M., and R. L. Rill. 1991. Noncovalent DNA-binding of Bis (1,10-phenanthroline)copper(i) and related compounds. *Biochemistry.* 30:1132–1140.

von Hippel, P. H., A. Revzin, C. A. Gross, and A. C. Wang. 1974. Non-specific DNA binding of genome regulating proteins as a biological control mechanism. I. The *lac* operon: equilibrium aspects. *Proc. Natl. Acad. Sci. U.S.A.* 71:4808–4812.

Wong, I., and T. M. Lohman. 1995. Linkage of protein assembly and protein-DNA binding. *Methods Enzymol.* 259:95–126.

Zhang, W., H. Ni, M. W. Capp, C. F. Anderson, T. M. Lohman, and M. T. Record, Jr. 1999. The importance of coulombic end effects: experimental characterization of the effects of oligonucleotide flanking charges on the strength and salt dependence of oligocation (L^{8+}) binding to single-stranded DNA oligomers. *Biophys. J.* 76: 1008–1017.