

# Interactions of Cationic Ligands and Proteins with Small Nucleic Acids: Analytic Treatment of the Large Coulombic End Effect on Binding Free Energy as a Function of Salt Concentration<sup>†</sup>

Irina A. Shkel,<sup>\*,‡</sup> Jeff D. Ballin,<sup>‡,§</sup> and M. Thomas Record, Jr.<sup>‡,||</sup>

Departments of Chemistry and Biochemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706

Received October 7, 2005; Revised Manuscript Received March 10, 2006

**ABSTRACT:** For nonspecific binding of oligopeptides and other cationic ligands, including proteins, to nucleic acid oligomers, we develop a model capable of quantifying and predicting the salt concentration dependence of the binding free energy ( $\Delta G_{\text{obs}}^{\circ}$ ) by way of an analytic treatment of the Coulombic end effect (CEE). Ligands, nucleic acids, and their complexes (species  $j$  of valence  $Z_j$ ) are modeled as finite lattices with  $|Z_j|$  charged residues; the CEE is quantified by its characteristic length  $N_e$  (specified in charged residues) and its consequences for the free energy and ion association of the oligomer. Expressions are developed for the individual site binding constants  $K_i$  as a function of position (site number  $i$ ) of a bound ligand on a nucleic acid and for the observed binding constant  $K_{\text{obs}}$  as an ensemble average of  $K_i$ . Analysis of  $\Delta G_{\text{obs}}^{\circ} = -RT \ln K_{\text{obs}}$  and  $S_a K_{\text{obs}} \equiv (\partial \ln K_{\text{obs}})/(\partial \ln a_{\pm})$  for binding of the oligopeptide KWK<sub>6</sub> ( $Z_L = +8$ ) to single-stranded (ss) dT(pdT)<sub>|Z<sub>D</sub>|</sub> oligomers (dT-mers) where  $Z_D = \{-6, -10, -11, -14, -15\}$  in the range 0.1–0.25 M Na<sup>+</sup> yields  $N_e = 9.0 \pm 0.8$  residues at each end, demonstrating that both KWK<sub>6</sub> and the above dT-mers are sufficiently short so that the CEE extends over the entire molecule. The dependences of  $K_{\text{obs}}$  and of  $S_a K_{\text{obs}}$  on  $|Z_D|$  for a given  $Z_L$  are determined by the difference between  $2N_e$  and the net number of charged residues  $Q$  in the complex ( $Q \equiv |Z_D| - Z_L$ ). For  $Q < 2N_e$ , characteristic of complexes of KWK<sub>6</sub> with this set of dT-mers, the distribution of binding free energies  $\Delta G_i^{\circ} = -RT \ln K_i$  for sites along the DNA oligomer is parabolic, and  $K_{\text{obs}}$  and  $S_a K_{\text{obs}}$  are strongly dependent on  $|Z_D|$ . For  $Q \geq 2N_e$ , the distribution of binding free energies  $\Delta G_i^{\circ}$  is trapezoidal, and the dependence of  $K_{\text{obs}}$  and  $S_a K_{\text{obs}}$  on  $|Z_D|$  is weaker. Application of the model to nonspecific binding of human DNA polymerase  $\beta$  to ssDNA demonstrates the significance of the CEE in determining  $K_{\text{obs}}$  and  $S_a K_{\text{obs}}$  of binding of a cationic site on a protein to a DNA oligomer.

Processes involving polymeric nucleic acids are generally extremely salt concentration ( $[\text{salt}]$ )<sup>1</sup> dependent, even at the lowest accessible  $[\text{salt}]$ . In particular, the equilibrium constant ( $K_{\text{obs}}$ ) for nonspecific binding of polyamines, oligopeptides, and other oligocationic ligands to a site on the polymeric DNA or RNA exhibits a strong negative power dependence on  $[\text{salt}]$  [ $S_a K_{\text{obs}} \equiv (\partial \ln K_{\text{obs}})/(\partial \ln [\text{salt}])$ ], typically determined below 0.3 M salt (1–9). These studies, performed as a function of oligocation charge,  $Z_L$ , found that  $S_a K_{\text{obs}}$  of binding to a polyanionic nucleic acid is proportional to  $Z_L$  and that the extrapolated value of the standard binding free energy ( $\Delta G_{\text{obs}}^{\circ} = -RT \ln K_{\text{obs}}$ ) at 1 M salt is small in magnitude.

The strong effects of  $[\text{salt}]$  on conformational transitions and oligocation binding to polyelectrolyte nucleic acids are widely recognized; theoretical and computational analysis show that those salt effects are primarily Coulombic in origin (10–17).

Although less widely recognized and the subject of some long-standing controversy (reviewed in ref 18), processes that involve short oligonucleotides or the ends of any nucleic acid (NA) generally exhibit a very different  $[\text{salt}]$  dependence than the corresponding processes involving the interior of a polymeric nucleic acid as a result of Coulombic end effects (CEE) (19–23). CEE are responsible for the experimentally observed reductions in the  $[\text{salt}]$  dependence of the melting temperature [ $ST_m = dT_m/(d \log [\text{salt}])$ ] for transitions of NA hairpin helices shorter than 30 base pairs (bp) and two-stranded helices shorter than 15 bp (18). Both  $K_{\text{obs}}$  and  $S_a K_{\text{obs}}$  for binding of an oligocationic ligand are much smaller in magnitude when the target site is on a short DNA oligomer than when on polymeric DNA (24–26). As the number of DNA charges,  $|Z_D|$ , is reduced from 22 to 10, the average per site equilibrium constant for binding of the peptide KWK<sub>6</sub> at 0.1 M salt decreases by more than 2 orders of magnitude. For longer DNA oligomers, binding of KWK<sub>6</sub> is far less sensitive to changes in DNA charge; the per site  $K_{\text{obs}}$

<sup>†</sup> This research was supported by the University of Wisconsin, Madison, and by NIH grants to M.T.R.

<sup>\*</sup> Corresponding author. Phone: (608) 262-3019. Fax: (608) 262-3453. E-mail: ishkel@wisc.edu.

<sup>‡</sup> Department of Chemistry.

<sup>§</sup> Present address: Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, 108 N. Greene St., Baltimore, MD 21201.

<sup>||</sup> Department of Biochemistry.

<sup>1</sup> Abbreviations: CEE, Coulombic end effect; NA, nucleic acid;  $[\text{salt}]$ , salt concentration; bp, base pair; ss, single stranded; ds, double stranded; nt, nucleotide; DBM, distributed binding model; CCM, central complex model; NLPB, nonlinear Poisson–Boltzmann; MC, Monte Carlo; pol  $\beta$ , human DNA polymerase  $\beta$ .

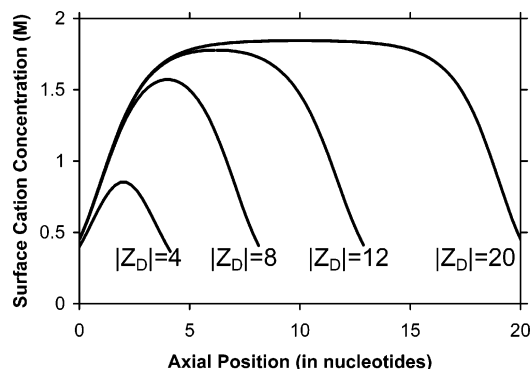


FIGURE 1: Trapezoidal (for oligomer charge  $|Z_D| = 20$ ) and parabolic (for oligomer charges  $|Z_D| = 4$  and  $|Z_D| = 8$ ) axial profiles of surface cation concentration calculated using Boltzmann relations from bulk concentration and the surface potential obtained from the cylindrical nonlinear Poisson–Boltzmann model of a ssDNA oligomer (as described in ref 18) at 0.1 M 1:1 salt. The curve for  $|Z_D| = 12$  illustrates the transition between parabolic and trapezoidal profiles. The axial position on the abscissa is counted from the left terminus of the DNA lattice.

increases by a factor of 3 at 0.1 M  $\text{Na}^+$  as the DNA length increases from 39 to 169 charges. This trend is mirrored in the [salt] dependences measured for these interactions:  $|SK_{\text{obs}}|$  at 0.1 M  $\text{Na}^+$  is  $\sim 30\%$  smaller for binding of  $\text{KWK}_6$  to a site on  $\text{dT}(\text{pdT})_{10}$  than to  $\text{dT}(\text{pdT})_{22}$  but only  $\sim 10\%$  smaller for binding of  $\text{KWK}_6$  to a site on  $\text{dT}(\text{pdT})_{39}$  than to  $\text{dT}(\text{pdT})_{169}$ .

Coulombic end effects on binding of oligocations to NA oligomers are the result of the strong *axial* dependence of the surface electrostatic potential and of the surface salt ion concentration, both for the oligocation and for the 10–20 phosphates at each end of the NA molecule (21–24). Calculations for representative oligomers, modeled as charged cylinders, at 0.1 M salt are shown in Figure 1. In particular, Figure 1 shows that sufficiently long NA oligomers exhibit a polyelectrolyte-like interior resulting in a trapezoidal axial profile of the surface counterion concentration, tapering from the polyion value to a much smaller value at each end. By contrast, for short NA oligomers, Figure 1 shows that even at the center of the NA the surface counterion concentration is less than that of polyanionic NA and the axial profile is parabolic instead of trapezoidal (20–22). Oligocationic ligands with small numbers of positive charges also behave as short oligoelectrolytes, characterized by surface anion concentration profiles which are parabolic and everywhere smaller than the polymeric value. The electrostatic potential calculated for a cylindrical model of the nucleic acid (23) exhibits an axial dependence which parallels those of the surface counterion concentration. Since Coulombic interactions account for the majority of the binding free energy in nonspecific interactions of oligocations with nucleic acids, the CEE exhibited in surface counterion concentration profiles in Figure 1 are predicted to also characterize binding behavior of oligocationic ligands, which therefore are predicted to bind preferentially to the central region of the NA lattice, especially at low [salt].

In this paper, we develop a simple predictive model of Coulombic end effects and their consequences for the axial distribution of complex formed between two oppositely charged oligoions of different lengths. The model introduces a set of coefficients describing the dependences of the excess

chemical potential of the oligoion and its salt derivative on oligoion length (number of charged monomers). The coefficients are evaluated by fitting experimental data (26) to analytical expressions relating binding free energy ( $\Delta G_{\text{obs}}^\circ = -RT \ln K_{\text{obs}}$ ) and its salt derivative [ $S_a K_{\text{obs}} \equiv (\partial \ln K_{\text{obs}}) / (\partial \ln a_{\pm})$ ] to the numbers of charged residues on the NA ( $|Z_D|$ ) and oligocationic ligand ( $Z_L$ ). (Although properties of the nucleic acid are designated by the subscript D, our analysis is equally applicable to DNA and RNA lattices.) The model predicts that the trapezoidal and parabolic axial profiles of surface salt ion concentration (Figure 1) lead directly to trapezoidal and parabolic axial profiles of site binding free energy  $\Delta G_i^\circ$  for oligocationic ligands and that the transition from parabolic to trapezoidal behavior of  $\Delta G_i^\circ$  occurs when the net number of charges in the complex  $Q \equiv |Z_D| - Z_L$  increases beyond  $2N_e$ , where  $N_e$  is equal to the number of charges in each terminal region of the trapezoidal distribution of  $\Delta G_i^\circ$ .

The distributed binding model (DBM) relates axial profiles of binding free energy  $\Delta G_i^\circ$  for ligand to the ensemble average values of both  $K_{\text{obs}}$  and  $S_a K_{\text{obs}}$ . As an approximation to the DBM, the central complex model (CCM) approximates the ensemble-averaged  $K_{\text{obs}}$  and its salt dependence,  $S_a K_{\text{obs}}$ , by the binding constant of the central complex,  $K_c$ , and its salt dependence,  $S_a K_c = (\partial \ln K_c) / (\partial \ln a_{\pm})$ . Experimental values of  $K_{\text{obs}}$  for binding of  $\text{KWK}_6$  ( $\text{L}^{8+}$ ) are well described by the CCM for both short ( $|Z_D| < 16$ ) and long ( $|Z_D| > 27$ ) dT-mers. (For intermediate length dT-mers ( $16 \leq |Z_D| \leq 27$ ), the DBM is required.)

For any NA oligomer length our CEE analysis predicts that a cationic ligand is expected to bind more weakly at the ends than at the center of the anionic NA oligomer. (This analysis also predicts that if the ligand is an oligoanion, as in the binding of a short anionic dA oligomer to a longer anionic dU lattice, binding occurs preferentially at the ends of the dU lattice because the CEE causes the Coulombic contribution to the binding free energy to be least unfavorable at the lattice ends.) The DBM (and CCM, where applicable) explains the experimentally observed extreme sensitivity of the per-site  $K_{\text{obs}}$  of binding of a given oligocation ( $\text{L}^{Z+}$ ) to changes in NA length (26). The model confirms that these effects are a manifestation of significant Coulombic end effects present at typical experimental salt concentrations and provides expressions for quantitative predictions of  $K_{\text{obs}}$  and  $S_a K_{\text{obs}}$  as functions of  $|Z_D|$  and  $Z_L$  for single-stranded NA oligomers and oligocationic ligands of 6 or more charges each. Application of the DBM to analyze nonspecific binding of human polymerase  $\beta$  (pol  $\beta$ ) to polymeric or oligomeric ssDNA (27) shows that the behavior of both  $K_{\text{obs}}$  and  $SK_{\text{obs}}$  are identical to that expected for an  $\sim +4$  oligocation [for the 5 nucleotide (nt) binding mode] and for an  $\sim +10$  oligocation (for the 16 nt binding mode). In both modes, pol  $\beta$  is predicted to exhibit the same preference for the center compared to the end of the DNA oligomer as the corresponding  $Z_L$ -valent oligocation, indicating the utility of the DBM to analyze CEE in NA–protein interactions.

## BACKGROUND

Binding of an oligocationic ligand (L) with  $Z_L$  univalent charges to a oligoanionic lattice (D) with  $|Z_D|$  charges produces a complex (DL) with  $||Z_D| - Z_L|$  charges:



The effect of mean ionic activity ( $a_{\pm}$ ) of 1:1 salt on  $K_{\text{obs}}$  is related to the observed standard free energy change  $\Delta G_{\text{obs}}^{\circ}$  in the binding reaction by

$$S_a K_{\text{obs}} \equiv \frac{\partial \ln K_{\text{obs}}}{\partial \ln a_{\pm}} = -\frac{1}{RT} \left( \frac{\partial \Delta G_{\text{obs}}^{\circ}}{\partial \ln a_{\pm}} \right) = \Delta(|Z_j| + 2\Gamma_j) \equiv \Delta n_j \quad (2)$$

where  $n_j \equiv |Z_j| + 2\Gamma_j = -(RT)^{-1} \partial G_j^{\circ} / (\partial \ln a_{\pm})$  is the thermodynamic extent of salt ion association with species  $j$  ( $j = L, D, DL$ ) of chemical potential  $G_j^{\circ}$  (2, 3, 28). The derivative  $S_a K_{\text{obs}}$  is the experimentally accessible thermodynamic measure of ion release from the nucleic acid and oligocation upon binding (10);  $S_a K_{\text{obs}}$  is directly related to the difference in salt component–oligomer component preferential interaction coefficients ( $\Gamma_j$ ) for products and reactants (28).

We consider the limit of zero binding density where no more than one ligand is bound to each NA even if the NA lattice is long enough to support greater binding densities. We assume that a ligand binding site on a NA oligomer consists of  $Z_L$  contiguous charges (Figure 2A), so that the NA has  $Q + 1$  sites for the ligand to bind, and that binding interactions involving fewer than  $Z_L$  contiguous residues do not occur. The binding of a single  $Z_L$ -valent cationic ligand (L) to a lattice (D) with  $[Z_D]$  charged residues forms the complex  $DL_i$ , one of an ensemble of complexes (referred to collectively as distributed binding; see Figure 2B). The subscript  $i$  indicates that the bound ligand is positioned a distance of  $i$  charges from one end of the NA lattice (Figure 2A). Thus, the binding sites at the ends of the lattice have  $i = 0$  or  $i = Q$ . In its simplest form, this model describes ligands that have the same axial charge separation  $b$  as the nucleic acid [applicable to KWK<sub>6</sub> where  $b = 3.6$  Å (29), similar to the value of  $b$  determined for ssDNA,  $b = 3.4 \pm 0.2$  Å (11)]. This model may also be applied to flexible oligocations with  $Z_L$  charges but with larger mean axial charge separation than that of the NA lattice, if they undergo conformational changes upon binding to interact with  $Z_L$  consecutive charges on the nucleic acid. For example, the nonspecific DNA binding of 17-residue peptides with 4 cationic (K) residues interspaced with 3 or 4 alanines exhibits site sizes much closer to 4 than to 17 (30) as a result of  $\alpha$  helix formation coupled to binding (31). If the ligand has more charges than the NA oligomer, as in the case of KWK<sub>6</sub> (8 charges) and dT(pdT)<sub>6</sub> (6 charges), the subscript  $i$  in the designation of the complex  $DL_i$  indicates that the bound NA is positioned a distance of  $i$  charges from one end of the oligocation (Figure 2C). Such complexes are characterized by  $Q < 0$ . Analysis of this case requires some additional assumptions, which are explicitly presented in the Appendix.

For the process of forming a complex  $DL_i$  in the limit of zero binding density, the microscopic equilibrium binding constant,  $K_i$ , and corresponding free energy change,  $\Delta G_i^{\circ}$ , are both defined per site as

$$K_i \equiv e^{-(\Delta G_i^{\circ}/RT)} = [DL_i]/[D][L] \quad (3)$$

where  $[DL_i]$ ,  $[D]$ , and  $[L]$  are respectively the equilibrium molar concentrations of complexes, free NA, and free ligand

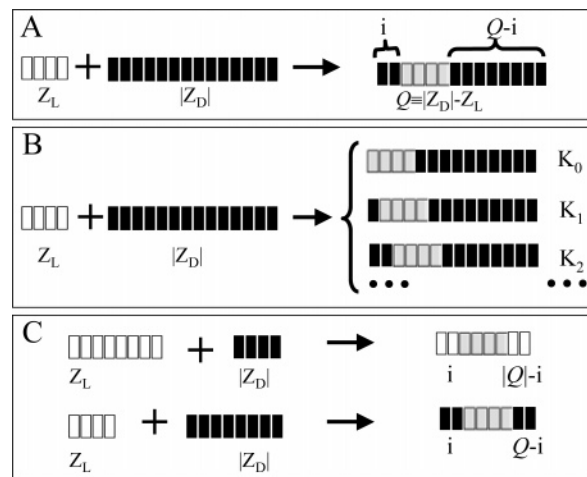


FIGURE 2: Complexes formed by distributed binding of oppositely charged oligoelectrolytes. In all panels a ligand with  $Z_L$  charges (white) binds to a lattice with  $[Z_D]$  charges (black) and occupies  $Z_L$  consecutive charges on the lattice (gray). (A) For  $[Z_D] - Z_L \equiv Q > 0$ , the complex has two charged flanking regions of size of  $i$  and  $Q - i$  charges. The flanking region of  $i$  charges is referred to as short if  $i < N_e$  and as long if  $i \geq N_e$ , where  $N_e$  is the length of the Coulombic end effect ( $N_e = 9.0 \pm 0.8$  phosphates for ssDNA lattice at 0.1–0.25 M salt; see Table 2). (B) Schematic representation of nonoverlapping members of the ensemble of complexes. The position of the bound ligand is described by the site number  $i$  from one (left) terminus of the lattice, which is equal to the number of unoccupied charged residues on the lattice; the corresponding binding constant is denoted  $K_i$ . Complexes at positions  $i$  and  $Q - i$  are identical, so  $K_i = K_{Q-i}$ . (C) Complexes with two short flanking regions for which  $Q < 0$  (top) and  $Q \geq 0$  (bottom). The position of the DNA oligomer on the ligand lattice is characterized by  $i$ , the number of unoccupied ligand charges in the complex. The number of sites for complexes with  $Q < 0$  is  $|Q| + 1$ .

in solution. In the distributed binding model (DBM), where  $\Delta G_i^{\circ}$  and  $K_i$  depend on  $i$ , the macroscopic binding constant is the average of all possible (nonoverlapping) binding positions:

$$K_{\text{obs}} = \sum_{i=0}^{|Q|} \frac{e^{-(\Delta G_i^{\circ}/RT)}}{|Q| + 1} \quad (\text{DBM}) \quad (4)$$

This expression provides the connection between the individual binding free energies for each position of the 1:1 complex and the position-averaged experimentally determined  $K_{\text{obs}}$  (26). With use of the absolute value of  $Q$ , eq 4 is applicable both to the usual situation where  $Z_L < [Z_D]$  and to the special case  $Z_L > [Z_D]$ . For all values of  $Z_L$  relative to  $[Z_D]$ ,  $|Q| + 1$  is the number of binding sites for the shorter lattice on the longer lattice, independently of which lattice (NA or ligand) is longer in terms of number of charges.

In the DBM,  $S_a K_{\text{obs}}$  (eq 2) is obtained from eq 4 for  $K_{\text{obs}}$ :

$$S_a K_{\text{obs}} = \frac{\sum_{i=0}^{|Q|} (S_a K_i) e^{-(\Delta G_i^{\circ}/RT)}}{\sum_{i=0}^{|Q|} e^{-(\Delta G_i^{\circ}/RT)}} \quad (\text{DBM}) \quad (5)$$

where  $S_a K_i \equiv [-(1/RT)] [\partial \Delta G_i^{\circ} / (\partial \ln a_{\pm})]$ . Thus, in the DBM,  $S_a K_{\text{obs}}$  is an ensemble average of the population-



weighted microscopic  $S_a K_i$  for each complex, where the distribution of complexes is determined by  $\Delta G_i^\circ$ . Since  $\Delta G_i^\circ$  and its derivative  $S_a K_i$  are thermodynamic quantities, they are determined by the difference in chemical potentials between final (product) and initial (reactant) values.

## RESULTS

*Analytical Description of the Coulombic End Effect on Oligomer Chemical Potential and Its Salt Derivative.* Binding studies using the peptide KWK<sub>6</sub> ( $Z_L = 8$ ) and short ss dT-mers (26) revealed large effects of dT-mer length on the site binding constant  $K_{\text{obs}}$  and on the amount of salt ion release  $\Delta n$  (as measured by  $S_a K_{\text{obs}}$ , eq 2). To interpret these results, we develop analytical expressions for the chemical potential  $G_j^\circ$  and the salt ion association  $n_j$  as a function of the number of charges  $|Z_j|$  on the NA ( $j = D$ ) or oligocation ( $j = L$ ). We consider nucleic acid oligomers and oligopeptides as linear lattices of charged residues (monomers). For each, the total lattice length ( $l$ ) is a linear function of the number of lattice charges ( $l = |Z|b + l_{\text{ci}}$ ), where  $b$  is the average axial spacing between nearest charged units. The  $l_{\text{ci}}$  ("ci" refers to "charge independent") term accounts for uncharged substituents on a lattice. For example, for a DNA oligomer with an uncharged terminal nucleoside,  $l_{\text{ci}}$  is the length of that nucleoside; for KWK<sub>6</sub>  $l_{\text{ci}}$  is the length of the tryptophan residue. Charge-independent contributions to the net ion accumulation and chemical potential are denoted by  $n_{\text{ci}}$  and  $G_{\text{ci}}^\circ$ ; those quantities are the values of  $n$  and  $G^\circ$  of an oligomer after deletion of all its charged residues. Any thermodynamic contributions to  $n$  and  $G^\circ$  which depend on the accessible surface area of the lattice (such as interactions with water) are well modeled as length dependent. Similarly, interactions which depend only on the lattice length (and not charge) can still be considered  $|Z|$  dependent via the linear relationship above between length and charge. Thus, we reduce all possible interactions for a lattice into  $|Z|$ -dependent and  $|Z|$ -independent terms.

For any homologous series of oligoelectrolytes (e.g., oligolysines or dT-mers) with sufficiently small numbers of charged residues  $|Z|$ , we assume that  $n$  (and also  $G^\circ$ ) can be represented as polynomial series in  $|Z|$  (see Appendix). Although the rigorous Taylor series is expected to converge slowly, we find [both from fitting experimental data from ref 26 to the DBM and from nonlinear Poisson–Boltzmann (NLPB) calculations of  $n(|Z|)$  for various NA oligomer models, details of which will be published elsewhere] that, for our purposes, truncation of the expansion at  $|Z|^2$  is an acceptable approximation:

$$n \cong n_{\text{ci}} + n_{\text{u},0}|Z| + \mathcal{J}_1|Z|^2 \quad (\text{small } |Z|) \quad (6)$$

Here  $n_{\text{ci}}$  is the charge-independent term as discussed above,  $n_{\text{u},0}$  is the salt ion accumulation per charge of a short oligomer in the limit as the number of charges approaches zero, and  $\mathcal{J}_1$  describes the nonlinear correction to  $n$  arising in part from CEE.

For polymeric DNA (where end effects are usually negligible), experimental studies (32) and theoretical calculations (33) using a cylindrical polyelectrolyte model find that the salt ion preferential interaction coefficient (Donnan coefficient)  $\Gamma$  and the ion accumulation  $n = |Z| + 2\Gamma \geq 0$  are directly proportional to the number of DNA charges,  $n$

$= n_{\text{u},\infty}|Z|$ , where  $n_{\text{u},\infty}$  is the per charge net ion accumulation for a polymer (similarly,  $G^\circ = G_{\text{u},\infty}^\circ|Z|$ ). For a homologous series of long oligoelectrolytes with  $|Z|$  charges, where end effects cannot be neglected, we find that the salt ion accumulation per oligomer molecule differs from that predicted for the same number of charged residues in the interior of an infinite polyelectrolyte by an end effect correction term designated  $2\gamma$ . The possible contributions of any charge- and length-independent factors to  $n$  for an oligomer (not accounted for by  $n_{\text{u},\infty}$ ) are designated  $n_{\text{ci}}$ , as in eq 6:

$$n \cong n_{\text{ci}} + n_{\text{u},\infty}|Z| - 2\gamma \quad (\text{large but not polymeric } |Z|) \quad (7)$$

The salt ion accumulation per charge of a polyion ( $n_{\text{u},\infty}$ , eq 7) is expected to be much larger than that characteristic of the low charge limit of a short oligomer ( $n_{\text{u},0}$ , eq 6). [Equations 6 and 7 are the two leading terms in series expansions (see Appendix) of the per charge salt ion accumulation  $n_{\text{u}} = (n - n_{\text{ci}})/|Z|$  in powers of  $|Z|$  (eq 6) and  $|Z|^{-1}$  (eq 7).]

Similar considerations apply for the dependence of  $G^\circ$  on the charge  $|Z|$ , since  $n$  is proportional to the derivative of  $G^\circ$  with respect to the logarithm of [salt] (eq 2):

$$G^\circ = G_{\text{ci}}^\circ + G_{\text{u},0}^\circ|Z| + \mathcal{J}_1|Z|^2 \quad (\text{small } |Z|) \quad (8)$$

$$G^\circ = G_{\text{ci}}^\circ + G_{\text{u},\infty}^\circ|Z| - 2g \quad (\text{large } |Z|) \quad (9)$$

Here, similarly to eq 6,  $G_{\text{ci}}^\circ$  is the charge independent term,  $G_{\text{u},0}^\circ$  and  $G_{\text{u},\infty}^\circ$  are the per charge chemical potentials of oligomer in the limit of zero and polymeric  $|Z|$ , respectively, and  $\mathcal{J}_1$  and  $g$  describe the corrections to  $G^\circ$  arising from CEE.

To quantify the ranges of  $|Z|$  which are small (eqs 6 and 8) or large (eqs 7 and 9), we introduce the length  $N_e$ , which will be seen to characterize the range of the CEE:  $N_e$  is the length (in charged residues) below which  $n$  ( $G^\circ$ ) is described more accurately by eq 6 (eq 8) than by eq 7 (eq 9). To match the behavior of  $n$  (and  $G^\circ$ ) as described by eq 6 (eq 8) and eq 7 (eq 9) at the length  $N_e$ , we require continuity of the functions ( $n$  and  $G^\circ$ ) and their first derivatives with respect to  $|Z|$  at  $|Z| = N_e$ , yielding (after simplification and expressing all other coefficients via  $\mathcal{J}_1$ ,  $\mathcal{J}_1$ , and  $N_e$ )

$$n_{\text{u},\infty} - n_{\text{u},0} = 2\mathcal{J}_1 N_e \quad (10)$$

$$2\gamma = \mathcal{J}_1 N_e^2 \quad (11)$$

$$G_{\text{u},\infty}^\circ - G_{\text{u},0}^\circ = 2\mathcal{J}_1 N_e \quad (12)$$

$$2g = \mathcal{J}_1 N_e^2 \quad (13)$$

Equations 10 and 11 relate the four coefficients characterizing salt ion association ( $n_{\text{u},\infty}$ ,  $n_{\text{u},0}$ ,  $\mathcal{J}_1$ ,  $\gamma$ ) to each other and to  $N_e$ ; eqs 12 and 13 relate the corresponding four chemical potential coefficients to each other and to  $N_e$ . Additional constraints on these coefficients are provided by the requirements that eqs 6 and 7 approximate the actual length dependence of  $n$  and that eqs 8 and 9 approximate the actual length dependence of  $G^\circ$ . These constraints are implemented here as a fitting algorithm minimizing the discrepancy

between the experimental observables  $K_{\text{obs}}$  and  $S_a K_{\text{obs}}$  and the analytical expressions developed below from eqs 6–9. The thermodynamic differences between short and long NA are then described simply by the length of the end effect,  $N_e$ , and by the coefficients,  $\mathcal{J}_1$  and  $\mathcal{J}_i$ , all of which are obtained from nonlinear least-squares analysis of experimental data.

*Analytical Expressions for Chemical Potential and Ion Accumulation of an Oligocation–Oligoanion Complex.* We model each complex formed by a nucleic acid oligomer and a shorter cationic ligand ( $|Z_D| > Z_L$ ) as a NA lattice where the ligand occupies a region of  $Z_L$  consecutive charged units on the NA oligomer, so that the complex in general has two anionic flanking regions. Analytical expressions for the chemical potential  $G_{\text{DL},i}^\circ$  and ion accumulation  $n_{\text{DL},i}$  of the complex are derived for the case where the ligand and NA are long enough that the two charged flanking regions of NA on either side of the bound ligand do not interact Coulombically with each other (see Appendix). NLPB calculations performed for cylindrical oligomer models of the ligand, ssDNA, and DNA–ligand complex at 0.15 M salt (Appendix, Figure 6B) indicate that this approximation is accurate for ligand and DNA lengths  $Z_L$ ,  $|Z_D| \geq 6$ . A flanking region is referred to as short if it has less than  $N_e$  charges and as long if it has  $N_e$  or more charges. As  $|Z_D|$  and  $Z_L$  are varied, the following cases are possible: (a)  $Q \equiv |Z_D| - Z_L < N_e$ , so that both flanking regions of all complexes in the ensemble are short; (b)  $N_e \leq Q < 2N_e$ , so that at least one flanking region of every complex is short; or (c)  $Q \geq 2N_e$ , so that at least one flanking region of every complex is long.

(a) *Distributed Binding of  $L^{Z+}$  to NA Oligomers Forming Short Flanking Regions ( $Q < N_e$ ).* Where the lattice is not much larger than the size of the ligand, all possible positions of the ligand in the complex result in two short flanking regions. These termini have lengths  $i$  and  $Q - i$ , both of which are smaller than  $N_e$ . We propose that the contributions from each short flanking region of a 1:1 complex to  $n_{\text{DL},i}$  and  $G_{\text{DL},i}^\circ$  obey eqs 6 and 8, such that each is a quadratic function of  $i$  or  $Q - i$ . Monte Carlo (MC) simulations using a cylindrical model for the complex of  $L^{8+}$  with a longer NA (22) support the assumption that each charged NA flanking region in the complex interacts independently with salt ions to approximately the same extent as the corresponding charged oligomer. All other contributions to  $G_{\text{DL},i}^\circ$  and  $n_{\text{DL},i}$  (described by the terms  $f_G, f_n$ ) are assumed to depend only on the valence of the ligand  $Z_L$  and not on the position of the ligand on the NA lattice:

$$G_{\text{DL},i}^\circ = [G_{\text{ci}}^\circ + G_{\text{u},0}^\circ i + \mathcal{J}_1 i^2] + [G_{\text{ci}}^\circ + G_{\text{u},0}^\circ (Q - i) + \mathcal{J}_1 (Q - i)^2] + f_G(Z_L) \quad (14)$$

(The expression for  $n_{\text{DL},i}$  is analogous to eq 14 for  $G_{\text{DL},i}^\circ$ , where each coefficient in  $G^\circ$  is replaced by the corresponding coefficient in  $n$ .)

Central binding ( $i = Q/2$ ) is of special significance; it is predicted to have the most favorable binding free energy because eq 14 has a minimum at  $i = Q/2$ . The chemical potential of the central complex (subscript “c”) is obtained from eq 14 by substitution of  $i = Q/2$ :

$$G_{\text{DL},c}^\circ = 2[G_{\text{ci}}^\circ + G_{\text{u},0}^\circ \frac{Q}{2} + \mathcal{J}_1 \left(\frac{Q}{2}\right)^2] + f_G(Z_L) \quad (15)$$

Equation 15 is exact for the case of even  $Q$ . For odd  $Q$ , there are two equivalent central binding sites [with  $i = (Q \pm 1)/2$ ], both of which have essentially the same binding free energy as that of the hypothetical center position at  $i = Q/2$ . [The difference in binding free energy between positions  $i = (Q \pm 1)/2$  and  $i = Q/2$  is  $\mathcal{J}_1/2$  and the difference in ion accumulation is  $\mathcal{J}_1/2$ , both of which are negligibly small compared to other terms in these equations.] Equation 15 is derived here for the case  $|Z_D| \geq Z_L$  ( $Q \geq 0$ ). The corresponding derivation for  $Q < 0$  is given in the Appendix.

(b) *Distributed Binding of  $L^{Z+}$  to NA Oligomers Forming One Long Flanking Region ( $N_e \leq Q < 2N_e$ ).* Where the lattice is somewhat larger than the ligand, binding of the ligand at a position  $i$  can produce one short and one long flanking region. This occurs for site numbers  $0 \leq i \leq Q - N_e$  (when  $i < N_e$  and  $Q - i \geq N_e$ ) and  $N_e \leq i \leq Q$  (when  $i \geq N_e$  and  $Q - i < N_e$ ). The short and long flanking regions are composed of  $Q/2 - |Q/2 - i|$  and  $Q/2 + |Q/2 - i|$  charges, respectively. We assume that their contributions to  $G_{\text{DL},i}^\circ$  are described by eq 9 for the long flanking region and by eq 8 for the short flanking region:

$$G_{\text{DL},i}^\circ = \left[ G_{\text{ci}}^\circ + G_{\text{u},0}^\circ \left( \frac{Q}{2} - \left| \frac{Q}{2} - i \right| \right) + \mathcal{J}_1 \left( \frac{Q}{2} - \left| \frac{Q}{2} - i \right| \right)^2 \right] + \left[ G_{\text{ci}}^\circ + G_{\text{u},0}^\circ \left( \frac{Q}{2} + \left| \frac{Q}{2} - i \right| \right) - 2g \right] + f_G(Z_L) \quad (16)$$

For site numbers  $Q - N_e < i < N_e$ , both flanking regions are short, and eq 14 applies. Correspondingly, central binding is described by eq 15.

(c) *Distributed Binding of  $L^{Z+}$  to NA Oligomers Forming Two Long Flanking Regions ( $Q \geq 2N_e$ ).* Where the lattice is much larger than the ligand, complexes can form where both regions of the lattice flanking a bound ligand are long. If binding at position  $i$  results in a complex with at least  $N_e$  charges in each flanking region ( $i \geq N_e$  and  $Q - i \geq N_e$ ), then both flanking regions are long, and their contributions to  $G_{\text{DL},i}^\circ$  are described by eq 9:

$$G_{\text{DL},i}^\circ = [G_{\text{ci}}^\circ + G_{\text{u},\infty}^\circ i - 2g] + [G_{\text{ci}}^\circ + G_{\text{u},\infty}^\circ (Q - i) - 2g] + f_G(Z_L) \quad (17)$$

The chemical potential of a complex determined by eq 17 does not depend on binding position  $i$ ; all complexes with two long flanking regions have the same chemical potential as the central complex:



$$G_{\text{DL},c}^\circ = G_{\text{DL},i}^\circ = 2[G_{\text{ci}}^\circ + G_{\text{u},\infty}^\circ \frac{Q}{2} - 2g] + f_G(Z_L) \quad (18)$$

If one flanking region is short ( $i < N_e$  or  $i > Q - N_e$ ), eq 16 applies to  $G_{\text{DL},i}^\circ$ .

Equations 14–18 express the chemical potential of central and distributed complexes ( $G_{\text{DL},c}^\circ$  and  $G_{\text{DL},i}^\circ$ ) in terms of  $G_{\text{ci}}^\circ$ ,  $G_{\text{u},0}^\circ$ ,  $G_{\text{u},\infty}^\circ$ ,  $\mathcal{J}_1$ , and  $g$ . The remaining term  $f_G$  (and  $f_n$  for  $n$ ) characterizes any contribution from the region of the complex occupied by the ligand (see below for determination of these terms).

*Analytical Expressions for the Free Energy Change and Its Salt Dependence for Binding  $L^{Z+}$  to a ssNA Oligomer.*

Table 1: Change in Free Energy ( $\Delta G^\circ$ ) and Its [Salt] Derivative ( $\Delta n$ ) of Binding of an Oligocationic Ligand to Oligoanionic NA for Distributed (A) and Central (B) Binding Models<sup>a</sup>

A. Distributed binding model				
				
$Q$	Binding position, $i$	$i$	$ Q  - i$	$\Delta G_i^\circ$ and $\Delta n_i (\equiv S_a K_i)^b$
$ Q  < N_e$	$0 \leq i \leq  Q $	short	short	$\Delta G_i^\circ = \Delta G_c^\circ + \frac{\mathcal{G}_1}{2}( Q  - 2i)^2$ (20)
	$Q - N_e < i < N_e$			$\Delta n_i = \Delta n_c + \frac{\mathcal{I}_1}{2}( Q  - 2i)^2$ (21)
$N_e \leq Q < 2N_e$	$0 \leq i \leq Q - N_e$	short	long	$\Delta G_i^\circ = \Delta G_\infty^\circ + \frac{\mathcal{G}_1}{4}(2N_e - Q +  2i - Q )^2$ (22)
	or $N_e \leq i \leq Q$	long	short	
$Q \geq 2N_e$	$0 \leq i < N_e$	short	long	$\Delta n_i = \Delta n_\infty + \frac{\mathcal{I}_1}{4}(2N_e - Q +  2i - Q )^2$ (23)
	or $Q - N_e < i \leq Q$	long	short	
	$N_e \leq i \leq Q - N_e$	long	long	$\Delta G_i^\circ = \Delta G_\infty^\circ$ (24)
				$\Delta n_i = \Delta n_\infty$ (25)
B. Central complex model				
				
$Q$	DNA charge	$ Q /2$	$ Q /2$	$\Delta G_c^\circ$ and $\Delta n_c (\equiv S_a K_c)$
	$6 \leq  Z_D  < N_e$	short	short	$\Delta G_c^\circ = \Delta G_\infty^\circ + \frac{\mathcal{G}_1}{2}(Q - 2N_e)^2 - \mathcal{G}_1( Z_D  - N_e)^2$ (26)
				$\Delta n_c = \Delta n_\infty + \frac{\mathcal{I}_1}{2}(Q - 2N_e)^2 - \mathcal{I}_1( Z_D  - N_e)^2$ (27)
$ Q  < 2N_e$	$ Z_D  \geq N_e$	short	short	$\Delta G_c^\circ = \Delta G_\infty^\circ + \frac{\mathcal{G}_1}{2}(2N_e - Q)^2$ (28)
				$\Delta n_c = \Delta n_\infty + \frac{\mathcal{I}_1}{2}(2N_e - Q)^2$ (29)
$Q \geq 2N_e$	$ Z_D  \geq N_e$	long	long	$\Delta G_c^\circ = \Delta G_\infty^\circ$ (30)
				$\Delta n_c = \Delta n_\infty$ (31)

<sup>a</sup>  $Z_L$  and  $|Z_D|$  should be larger than or equal to 6 charges for the assumption of noninteracting flanking regions to apply (see Appendix). <sup>b</sup>  $\Delta G_c^\circ$  and  $\Delta n_c$  in eqs 20 and 21 are substituted from part B of this table. <sup>c</sup> The expressions for the case of  $Q < 0$  are restricted to complexes allowing short flanking regions only,  $|Q| \leq N_e$ . The absolute value  $|Q|$  is used in this table and in the text where appropriate.

For central binding of an oligocation to a NA lattice, both the strength and [salt] dependence of central binding must approach the polyelectrolyte limit as the length of the NA lattice increases. For example, Figure 6A illustrates the approach of  $SK_{\text{obs}}$  to its polyelectrolyte limit as  $|Z|$  increases, using NLPB calculations on a cylindrical oligomer model. Imposing this requirement (see Appendix) allows the functions  $f_G$  and  $f_n$  (eqs 14–18) to be determined. In particular, eq 47 predicts that, for central binding,  $K_c$  and  $|S_a K_c|$  will attain their maximum (polyelectrolyte) values once there are  $N_e$  or more charges in both flanking regions of the complex. This analysis provides expressions for  $f_G$  (cf. eq 14) and  $f_n$  in terms of model coefficients and polyelectrolyte values ( $\Delta G_\infty^\circ$  for  $f_G$ ,  $S_a K_\infty \equiv \Delta n_\infty$  for  $f_n$ ) for binding of the ligand to polymeric NA:

$$f_G = \Delta G_\infty^\circ - G_{\text{ci}}^\circ + G_L^\circ + G_{\text{u},\infty}^\circ Z_L + \mathcal{G}_1 N_e^2 \quad (19)$$

Substitution of eqs 14, 16, and 17 where  $f_G$  is given by eq 19 and the appropriate expression for  $G_D^\circ$  (eqs 8 and 9) into  $\Delta G_i^\circ = G_{\text{DL},i}^\circ - G_L^\circ - G_D^\circ$  provides expressions for the free energy change of forming a complex with ligand at position  $i$ . These expressions for  $\Delta G_i^\circ$  and  $\Delta G_c^\circ$  are given in Table 1 for the specified binding position ( $i$ ) and net numbers of charges on the complex ( $Q$ ) and on the nucleic acid ( $|Z_D|$ ).

Note that the quantities  $\Delta G_\infty^\circ$  and  $\Delta n_\infty (\equiv S_a K_\infty)$  in Table 1 are the limiting values of  $\Delta G_{\text{obs}}^\circ$  and  $S_a K_{\text{obs}}$  for binding a  $Z_L$ -valent oligocation to an infinite polyanionic lattice where end effects are negligible.

With an additional assumption discussed in Appendix, the equations of Table 1 for complexes with short flanking regions ( $Q < N_e$ ) can be extended to cases where  $|Z_D| < Z_L$  [such as KWK<sub>6</sub> binding to dT(pdT)<sub>6</sub>,  $|Q| < N_e$ ] (26). Thus, fitting data on the length dependence of  $K_{\text{obs}}$  and  $S_a K_{\text{obs}}$  yields five quantities ( $N_e$ ,  $\mathcal{G}_1$ ,  $\mathcal{I}_1$ ,  $\Delta G_\infty^\circ$ , and  $S_a K_\infty$ ) which in turn yield the quantities  $G_{\text{u},\infty}^\circ - G_{\text{u},0}^\circ$ ,  $g$ ,  $n_{\text{u},\infty} - n_{\text{u},0}$ , and  $\gamma$  (from eqs 10–13).

*Analysis of Experimental Data Using Distributed and Central Complex Binding Models: Determination of  $N_e$ , the Range of the Coulombic End Effect.* Experimental values of  $S_a K_{\text{obs}}$  and  $\log K_{\text{obs}}$  for binding  $L^{8+}$  to ss dT-mers (26) are fit globally vs  $|Z_D|$ , assuming that the end effect length  $N_e$  is a property of the NA lattice and hence is the same for analysis of both  $\log K_{\text{obs}}$  and  $S_a K_{\text{obs}}$ . Values of  $\log K_{\text{obs}}$  and  $S_a K_{\text{obs}}$  as a function of  $|Z_D|$  for  $|Z_D| = 6$ –15 at 0.1 and 0.2 M salt were fitted to the distributed binding model (DBM; eqs 4 and 5 and eqs 20–25 from Table 1A) using NONLIN (34) with 67% confidence intervals. Values of  $\Delta G_\infty^\circ$  and  $S_a K_\infty$  were fixed at the experimentally determined (26)



Table 2: Values of Globally Fitted Thermodynamic Coefficients of the Distributed Binding Model (Eqs 20–25) Applied to KWK<sub>6</sub> Binding to dT(pdT)<sub>|Z<sub>D</sub>|</sub> with |Z<sub>D</sub>| = 6–15 (26)

$N_e$	1:1 [salt] (M)	$\Delta G_{\infty}^{\circ}$ (kcal/mol) <sup>c</sup>	$\mathcal{E}_1$ (kcal/mol)	$G_{u,\infty}^{\circ} - G_{u,0}^{\circ}$ (kcal/mol)	$g^b$ (kcal/mol)
$9.0 \pm 0.8^d$	0.1	−8.77	$0.0193 \pm 0.003$	$0.35 \pm 0.07$	$0.78 \pm 0.16$
	0.2	−6.46	$0.0126 \pm 0.002$	$0.23 \pm 0.05$	$0.51 \pm 0.1$
		$S_a K_{\infty}$	$\mathcal{I}_1$	$n_{u,\infty} - n_{u,0}^e$	$\gamma^f$
0.1–0.25		−6.54	$0.019 \pm 0.003$	$0.34 \pm 0.07$	$0.77 \pm 0.16$

<sup>a</sup> Calculated from eq 12. <sup>b</sup> Calculated from eq 13. <sup>c</sup> Under the experimental conditions used ( $T = 298$  K),  $RT = 0.59$  kcal/mol. <sup>d</sup>  $N_e$  is rounded to the closest integer for easier comparison with |Z<sub>D</sub>|, Z<sub>L</sub>, and Q, all of which are also integer numbers. <sup>e</sup> Calculated from eq 10. <sup>f</sup> Calculated from eq 11.

polymeric values for each salt concentration; values of  $\mathcal{E}_1$ ,  $\mathcal{I}_1$ , and  $N_e$  were determined at each salt concentration from the fit and used to calculate the other coefficients (via eqs 10–13). Table 2 presents the results of these global fits and calculations for the coefficients  $\mathcal{E}_1$ ,  $G_{u,\infty}^{\circ} - G_{u,0}^{\circ}$ , and  $g$  at each salt concentration. Values of  $N_e$  at 0.1 M Na<sup>+</sup> ( $9.0 \pm 0.8$ ) and at 0.2 M Na<sup>+</sup> ( $9.2 \pm 0.8$ ) are the same within error. The coefficient  $\mathcal{I}_1$  changes only slightly in this range of [Na<sup>+</sup>] (from  $0.0191 \pm 0.003$  at 0.1 M Na<sup>+</sup> to  $0.0184 \pm 0.003$  at 0.2 M Na<sup>+</sup>). Therefore, Table 2 lists the average values of  $N_e$  and  $\mathcal{I}_1$  for this salt range and also lists values of  $n_{u,\infty} - n_{u,0}$  and  $\gamma$  calculated from them (from eqs 10 and 11).

We also tested individual fits to experimental data, when  $\log K_{\text{obs}}$  and  $S_a K_{\text{obs}}$  data are fit separately either to the central complex model (CCM; eqs 26, 28, and 30 for  $\log K_{\text{obs}}$ , eqs 27, 29, and 31 for  $S_a K_{\text{obs}}$ ) or to the DBM (eqs 20, 22, and 24 for  $\log K_{\text{obs}}$ , eqs 21, 23, and 25 for  $S_a K_{\text{obs}}$ ). Values of fitted quantities obtained from individual fits (not shown) differ only slightly from those obtained from global fits. Predictions of  $\Delta G_{\text{obs}}^{\circ}$  and  $S_a K_{\text{obs}}$  for 6–15-mers using global or individual fitted quantities differ by less than 3%; values obtained from the global fit (in Table 2) have a smaller uncertainty and are used for quantitative predictions.

As Table 2 reports, the length  $N_e$  characterizing the range of the CEE for ss dT-mers is  $9.0 \pm 0.8$  charged residues;  $N_e$  is not significantly [salt] dependent over the range of the binding data (0.1–0.25 M Na<sup>+</sup>). By contrast, the quantities  $\mathcal{E}_1$ ,  $G_{u,\infty}^{\circ} - G_{u,0}^{\circ}$ , and  $g$  all exhibit significant [salt] dependences between 0.1 and 0.2 M Na<sup>+</sup>, each decreasing 35% over this salt range. In the Discussion, we provide applications of these coefficients and the CEE length  $N_e$  to interpret Coulombic end effects on properties of charged oligomers in solution (e.g., extent of ion accumulation, parabolic vs trapezoidal profiles of surface counterion concentration) and on NA processes [parabolic vs trapezoidal profiles of free energies of binding LZ<sup>+</sup> to the NA oligomer, length dependence of  $ST_m = dT_m/(d \log [\text{salt}])$  for conformational transitions from the ds to ss state of NA oligomers].

Panels A and B of Figure 3 compare experimental values of  $\log K_{\text{obs}}$  and  $S_a K_{\text{obs}}$  with expressions for DBM (solid line, eqs 20–25) and CCM (dashed line, eqs 26–31) for the range |Z<sub>D</sub>| = 6–15 at 0.1 and 0.2 M salt. Ensemble averaging, including the weaker noncentral binding modes (eq 4), yields a smaller binding constant for DBM (solid lines) than that predicted by CCM only (dashed lines). Experimental values of  $\log K_{\text{obs}}$  agree well with predictions from the DBM (eqs 20–25) between |Z<sub>D</sub>| = 6–15. Values of  $\log K_{\text{obs}}$  and  $S_a K_{\text{obs}}$  predicted by CCM (eqs 26–31, dashed lines) are quite similar to the predictions of the DBM for |Z<sub>D</sub>| ≤ 11 and deviate slightly for |Z<sub>D</sub>| = 14, 15. The solid line in Figure

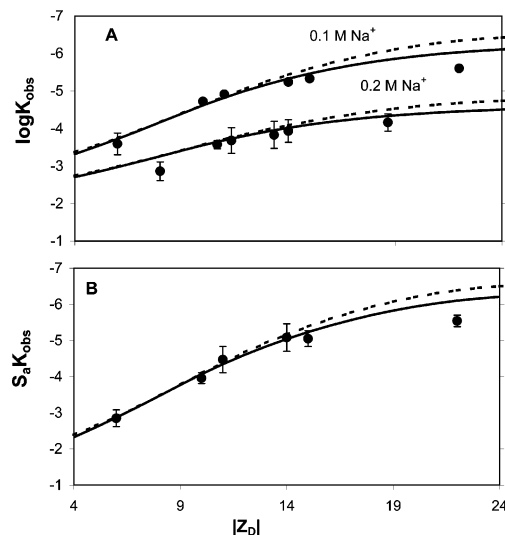


FIGURE 3: Analysis of binding constant  $K_{\text{obs}}$  and salt concentration dependence ( $S_a K_{\text{obs}}$ ) for L<sup>8+</sup>–ssDNA oligomer interactions using the distributed binding model (DBM; solid curves) and the central complex model (CCM; dashed curves). (A) Calculations of the logarithm of the ensemble average binding constant ( $\log K_{\text{obs}}$ ) at 0.1 M salt Na<sup>+</sup> (26) and at 0.2 M Na<sup>+</sup> as a function of DNA oligomer length (in charges) |Z<sub>D</sub>| using DBM (solid lines; eqs 4, 20, 22, and 24) and CCM (dashed lines; eqs 26, 28, and 30). (B) Analysis of the log–log derivative of  $K_{\text{obs}}$  with respect to salt activity ( $S_a K_{\text{obs}} = \partial \ln K_{\text{obs}} / \partial \ln a_{\pm}$ ; data from ref 26) as a function of |Z<sub>D</sub>| using DBM (solid lines; eqs 5, 21, 23, and 25) and CCM (dashed lines; eqs 27, 29, and 31).

3B is obtained from eq 5 with values of  $\Delta G_i^{\circ}$  either at 0.1 M Na<sup>+</sup> or at 0.2 M Na<sup>+</sup>.

Both DBM and CCM predictions of  $\log K_{\text{obs}}$  and  $S_a K_{\text{obs}}$  for dT(pdT)<sub>22</sub> overestimate the experimental values because only the binding of the first oligocation to this lattice was considered in this implementation of the DBM. However, dT(pdT)<sub>22</sub> is capable of binding two KWK<sub>6</sub> ligands, and the experimental data are analyzed to obtain an average binding constant. Binding of the second ligand to an available site at dT(pdT)<sub>22</sub> with one ligand bound is weaker than binding of the first ligand to site on the unoccupied lattice as a result of CEE introduced by the bound oligocationic ligand (see Appendix).

*CEE Reduces End Binding of an Oligocation to an Oligoanion Lattice Relative to the Center.* For binding of the same oligocationic ligand (of valence Z<sub>L</sub> = 8) to oligoanionic lattices of different length |Z<sub>D</sub>| (5 ≤ |Z<sub>D</sub>| ≤ 40), Figure 4 compares free energy changes for central binding ( $\Delta G_c^{\circ}$ ; most favorable) and end binding ( $\Delta G_{i=0}^{\circ}$ ; least favorable) at 0.1 M salt. At Z<sub>L</sub> = |Z<sub>D</sub>| = 8 the difference between central and end binding vanishes because there is only one

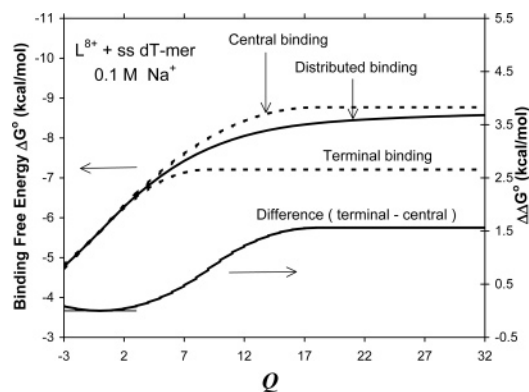


FIGURE 4: Plot of  $\Delta G_i^\circ$  for binding of  $L^{8+}$  to the lattice terminus,  $i = 0$  (lower dashed line), to the central site  $\Delta G_c^\circ$  (upper dashed line), and their difference  $\Delta\Delta G^\circ \equiv \Delta G_0^\circ - \Delta G_c^\circ$  (lower solid line) as a function of net charge of the complex  $Q$  at 0.1 M  $Na^+$ . Also shown is the ensemble average  $\Delta G_{obs}^\circ$  for the distributed binding model (upper solid line).

binding site for the ligand on the DNA lattice. The difference  $\Delta\Delta G^\circ = \Delta G_{i=0}^\circ - \Delta G_c^\circ$  is less than 0.04 kcal/mol for  $|Z_L| = 8$ ,  $|Z_D| = 10$  (so  $K_c$  exceeds  $K_0$  by less than 7%). For  $Z_L = 8$ ,  $|Z_D| = 16$ , the difference  $\Delta\Delta G^\circ$  is less than 0.6 kcal/mol, and the ratio of central to end binding constants is approximately 3. For these lengths ( $|Q| < N_e$ , e.g.,  $|Z_D| < 17$ ), the strength of central binding depends strongly on  $|Z_D|$ :  $\Delta G_c^\circ$  for central binding of  $L^{8+}$  to dT(pdT)<sub>14</sub> is more favorable (more negative) by 1.1 kcal/mol than for central binding to dT(pdT)<sub>10</sub>. The binding free energy for the end site ( $\Delta G_{i=0}^\circ$ ) also becomes more favorable in progressing from dT(pdT)<sub>10</sub> to dT(pdT)<sub>14</sub>, but the effect is smaller in magnitude (0.7 kcal/mol for end binding vs 1.1 kcal/mol for central binding), resulting in an increase in the magnitude of the difference  $\Delta\Delta G^\circ$ . For complexes with an intermediate net number of charges  $N_e \leq Q < 2N_e$ , the strong effect of  $|Z_D|$  on  $\Delta G_c^\circ$  observed for smaller  $Q$  is still present while the axial dependence of  $\Delta G_i^\circ$  becomes more apparent. Even though the free energy of end binding  $\Delta G_{i=0}^\circ$  does not decrease further when  $|Z_D| \geq Z_L + N_e$ ,  $\Delta G_c^\circ$  becomes 0.9 kcal/mol more favorable (more negative) as  $|Z_D|$  increases from 17 ( $Q = N_e$ ) to 26 ( $Q = 2N_e$ ), so that the difference between central binding and end binding ( $\Delta\Delta G^\circ$ ) becomes greater [ $\Delta\Delta G^\circ \sim 0.6$  kcal/mol for dT(pdT)<sub>14</sub> and  $\Delta\Delta G^\circ \sim 1.6$  kcal/mol for  $|Z_D| = 22$ ]. Therefore, the central binding constant for the 22-mer ( $K_c$ ) is predicted to be 11 times more favorable than the end binding constant ( $K_0$ ) at 0.1 M [salt]. For complexes with a large net number of charges ( $Q \geq 2N_e$ ),  $\Delta\Delta G^\circ$  is constant (at the value  $\Delta\Delta G^\circ = 2g = \mathcal{G}_1 N_e^2 = 1.56$  kcal/mol, as is evident from eq 22) for lengths beyond  $|Z_D| \geq Z_L + 2N_e$  (cf. Figure 4), and neither  $\Delta G_{i=0}^\circ$  nor  $\Delta G_c^\circ$  depends on  $|Z_D|$ . For binding of KWK<sub>6</sub> to such long NA oligomers, the binding constant for the eight phosphate sites at the ends of the oligomer ( $K_0$ ) is predicted to be only  $1/14$ th as large as the polymeric value  $K_\infty$  at 0.1 M [salt].

**The Axial Distribution of Complexes on a Lattice.** The predicted dependence of binding free energy  $\Delta G_i^\circ$  on binding position  $i$  for the DBM is plotted in Figure 5 for several oligomer lengths at 0.1 M [salt]. Equation 20 in Table 1A predicts a parabolic distribution of  $\Delta G_i^\circ$  for complexes with short flanking regions [ $|Q| < N_e$ , as is the case for oligomers in the range dT(pdT)<sub>6</sub> to dT(pdT)<sub>15</sub> in Figure 5].

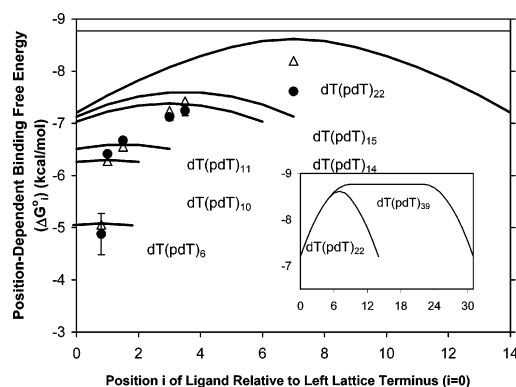


FIGURE 5: Plot of  $\Delta G_i^\circ$  (smooth curves) at 0.1 M  $Na^+$  vs site position  $i$  comparing DBM (triangles) and CCM (maxima of parabolas) with the observed per site experimental values of  $\Delta G_{obs}^\circ$  (circles) (26). The solid line at the top of the graph represents the free energy change for binding of  $L^{8+}$  to a polymeric DNA lattice ( $\Delta G_\infty^\circ$ ). The abscissa scale for the dT(pdT)<sub>6</sub> data and curves is shifted left 0.2 unit for clarity. The inset shows predicted  $\Delta G_i^\circ$  vs site position for a binding of  $L^{8+}$  to dT(pdT)<sub>22</sub> ( $|Z_D| = 22$ , parabolic curve) and to dT(pdT)<sub>39</sub> ( $|Z_D| = 39$ , trapezoidal curve).

For the case  $N_e \leq Q < 2N_e$  [applicable to dT(pdT)<sub>22</sub> in Figure 5], the parabolic distribution of  $\Delta G_i^\circ$  is described by eqs 20 and 22 (Table 1A). For large enough oligomeric lattices ( $Q \geq 2N_e$ ), the free energy of binding to interior sites  $N_e \leq i \leq Q - N_e$  is the same as that obtained for binding to the interior of the corresponding polymeric lattice (eq 24). This interior region corresponds to the plateau in the trapezoidal distribution of binding free energy with the axial position on the lattice [see the inset in Figure 5 for dT(pdT)<sub>39</sub>]. Equation 22 describes the free energy change for binding to  $N_e$  sites in each terminal region of a trapezoidal distribution. For each position  $i$ ,  $\Delta G_i^\circ$  initially decreases (becomes more favorable) as the lattice length  $|Z_D|$  increases; once  $|Z_D|$  exceeds  $Z_L + 2N_e$ , each  $\Delta G_i^\circ$  is a minimum independent of  $|Z_D|$ .

**Test of the Range of Applicability of the Central Complex Model (CCM).** For binding of  $L^{8+}$  to short NA oligomers, Figure 4 demonstrates the interplay between the greater stability of the central complex and the increase in number of noncentral sites with increasing NA length ( $|Z_D|$ ) on the ensemble-averaged  $\Delta G_{obs}^\circ$  for DBM (solid line). For complexes with a small net number of charges  $|Q| < N_e$ , complexes at all positions on the lattice make similar contributions to the overall binding affinity:  $\Delta G_i^\circ \cong \Delta G_c^\circ$  for all  $i$ . Therefore, for short NA oligomers with  $Z_L \leq |Z_D| < Z_L + N_e$ , the value of  $\Delta G_c^\circ$  ( $K_c$ ) for central binding provides a simple and accurate estimate of the experimentally observable (DBM ensemble-averaged) value of  $\Delta G_{obs}^\circ$  ( $K_{obs}$ ) for formation of 1:1 complexes, with a relative error not exceeding 3% in  $\ln K_{obs}$  (30% in  $K_{obs}$ ) at 0.1 M salt. For the complexes of intermediate length ( $N_e \leq Q < 2N_e$ ) the difference between  $K_c$  (from CCM) and  $K_{obs}$  (from DBM) is more significant. The maximum deviation between central and distributed binding free energies, observed for  $Q = 16$ , is 0.44 kcal/mol at 0.1 M salt, for which  $K_c$  is twice as large as  $K_{obs}$ . For  $Q \geq 2N_e$  (i.e.,  $|Z_D| \geq Z_L + 2N_e$ ), sites are added to the interior of the trapezoid (see inset in Figure 5), thereby weighting the distribution of possible complexes toward those in the polyelectrolyte-like interior of the long oligomer, and again the CCM well represents the observed binding affinity. With increasing  $Q$ ,  $K_{obs}$  approaches  $K_c$ ;  $K_{obs}/$



$K_c = 0.7$  for  $Q = 31$  ( $|Z_D| = 39$ ). As a result, the central binding constant  $K_c$  is a very good approximation to the experimentally determined site binding constant at both extremes of  $Q$ . At small  $Q$ , all sites are approximately equivalent because  $\Delta\Delta G^\circ$  is nearly zero. At very large  $Q$ , although there is a large  $\Delta\Delta G^\circ$ , the vast majority of possible sites are within the plateau region of the  $\Delta G_i^\circ$  distribution, and thus the average free energy of binding per site is well described by  $\Delta G_c^\circ = \Delta G_\infty^\circ$ .

## DISCUSSION

*Coulombic End Effect Causes Preference of Binding  $L^{Z+}$  to the Center of the NA Oligoanion and Strong Dependences of  $K_{obs}$  and  $S_a K_{obs}$  of  $L^{Z+}$  Binding on NA Lattice Length.* Theoretically predicted trapezoidal and parabolic axial profiles of surface counterion concentration (Figure 1) provide insight into understanding of Coulombic end effects in ligand DNA binding. The extreme ordering of salt ions (cation accumulation, coion exclusion) observed in the radial ion gradients near a NA surface represents an entropic cost to reduce intrinsically unfavorable interactions within the array of like charges on a NA molecule. The CEE results in smaller salt ion gradients and a smaller entropic cost of ordering salt ions in the terminal regions of any charged oligomer or polymer. Binding of a cationic ligand to anionic NA reduces the axial gradients in salt ion concentrations in the vicinity of the binding site due to charge neutralization (22). In the absence of Coulombic effects, a situation approached at high salt concentration, the oligocation would bind equally weakly to all sites on the nucleic acid. As the salt concentration is reduced, all site binding constants  $K_i$  for the cationic ligand at all positions on the NA lattice are predicted to increase. The increase in binding constant with decreasing [salt] is predicted to be much greater for interior binding than for binding to the terminal regions because the derivative  $(\partial \ln K_i)/(\partial \ln a_\pm) = -\Delta n_i$  is calculated to be larger in magnitude for central binding than for end binding (eqs 21, 23, and 25). [Given the steeper salt ion gradients in the interior than at the ends of the unbound NA, more salt ions are predicted to be released by central binding of the oligocationic ligand ( $L^{Z+}$ ) than in the case of end binding.] At any specified [salt] where Coulombic effects are important (e.g., 0.1 M), binding of a cationic ligand to the termini of any length NA and binding to any site on a sufficiently short NA oligomer are less favorable than binding to the polyelectrolyte interior of a long NA oligomer. (By contrast, nonspecific binding of an anionic ligand, such as a short ss dA oligomer to a longer ss dU oligomer, is predicted by the extension of our theory to occur preferentially to the ends of the oligo dU lattice.) For binding of a cationic ligand  $L^{Z+}$  to an oligoanionic NA lattice, the weaker binding modes reduce the observed (ensemble-averaged) binding constant below the value characteristic of binding the same oligocation to the corresponding polyanionic NA lattice and cause it to depend on NA lattice length.

*The Significance of  $N_e = 9.0 \pm 0.8$  Residues as a Characteristic of CEE Length for ssNA.* We obtain a CEE length  $N_e = 9.0 \pm 0.8$  residues for a ssNA at 0.1–0.25 M [salt]. In a series of NA lattices of varying length, the length  $N_e$  determines the transition from oligoelectrolyte to polyelectrolyte behavior in the interaction of the NA with salt

ions and in the binding of an oligocationic (or oligoanionic) ligand.

According to eqs 6–9, ssNA oligomers with less than nine charges are short oligomers at 0.1–0.25 M salt in the sense that the extent of salt ion accumulation is better described by eq 6 and the corresponding coefficients  $n_{u,0}$  and  $\mathcal{A}$ . Similarly, ssNA oligomers longer than 9 charges are long oligomers at 0.1–0.25 M salt in the sense that the extent of ion accumulation is better described by eq 7 and the corresponding coefficients  $n_{u,\infty}$  and  $\gamma$ . As the number of charges of a ssNA oligomer increases from  $|Z| = 1$  to  $|Z| = N_e - 1 = 8$ , eq 6 (and the estimate of  $n_{u,0}$  below) predicts a strong increase in the extent of salt ion accumulation *per oligoion (NA) charge* from  $n_u = 0.13 \pm 0.07$  to  $n_u = 0.26 \pm 0.08$ . (Of course, the total salt ion accumulation per NA molecule increases much more strongly as a result of the increase in the number of NA charges.) As the number of charges increases beyond  $N_e$ , salt ion accumulation is described by eq 7. Equation 7 predicts a more gradual increase in the extent of salt ion accumulation per NA charge from  $n_u = 0.28 \pm 0.08$  salt ions per oligomer charge at  $|Z| = N_e = 9$  to the polyelectrolyte value of  $n_{u,\infty} = 0.454 \pm 0.002$  as  $|Z| \rightarrow \infty$  [predicted from NLPB calculations for the cylindrical model of the ssNA oligomer (18) at 0.15 M salt], where  $n_u$  is within 10% of  $n_{u,\infty}$  only for  $|Z| \geq 34$ .

In the binding of  $L^{Z+}$  to ssNA,  $N_e$  determines the transition from a parabolic to a trapezoidal profile of binding free energy  $\Delta G_i^\circ$  vs axial position of the  $L^{Z+}$  ligand on the NA lattice, which occurs at the DNA oligomer length  $|Z_D| = Z_L + 2N_e$ . At this NA length, the binding constant of the central complex approaches the value characteristic of the interior of a polyanionic NA, because the central region of the trapezoidal profile has the binding strength and [salt] dependence characteristic of the corresponding polyelectrolyte. For oligomers longer than  $Z_L + 2N_e$ ,  $N_e$  binding sites at each end of the oligomer exhibit weaker binding strength and smaller magnitude of [salt] dependence of binding than the values characteristic of interior sites. *In parallel to the sharp decrease in salt ion accumulation per charge for an oligomer when its length is reduced to less than  $N_e$ , sharp reductions in  $\log K_{obs}$  and  $|S_a K_{obs}|$  of  $L^{Z+}$ –ssNA binding are observed when the NA length is reduced to less than  $Z_L + 2N_e$ .*

As stated above, the predicted regimes of trapezoidal and parabolic profiles of binding free energies  $\Delta G_i^\circ$  as a function of lattice length  $|Z_D|$  are fundamentally related to the trapezoidal and parabolic profiles of surface cation concentration on the same lattice. Considering a univalent cation as a ligand with charge  $Z_L = 1$ , our model predicts parabolic and trapezoidal profiles of surface concentration and a change over from parabolic to trapezoidal as the DNA oligomer length exceeds  $|Z_D| = 2N_e + 1$ . (This prediction has been tested with the cylindrical NLPB model, Figure 1. Although this NLPB model somewhat overestimates the experimental value of  $\mathcal{A}$  and underestimates  $n_{u,0}$  and  $N_e$ , it correctly predicts the transition from a parabolic to a trapezoidal axial distribution of the surface cation concentration (as shown in Figure 1) between  $|Z_D| = 8$  and  $|Z_D| = 12$ .)

The cylindrical NLPB model of DNA oligomers (18) at 0.15 M univalent salt predicts for ssDNA that  $\gamma = 0.60 \pm 0.02$ ,  $n_{u,\infty} = 0.454 \pm 0.02$  and for dsDNA that  $\gamma = 1.42 \pm 0.02$ ,  $n_{u,\infty} = 0.612 \pm 0.02$ . These values were previously

used to analyze the [salt] derivative of the transition temperature [ $ST_m = dT_m/(d \log [\text{salt}])$ ] for denaturation of hairpin and two-stranded helices (18). Experimentally observed  $ST_m$  exhibit the same signature of the CEE as observed in oligocation binding; transitions involving short oligomer strands exhibit a strong dependence of  $ST_m$  on  $|Z_D|$ . This occurs when a ssNA oligomer is shorter than  $N_e = 9.0 \pm 0.8$  phosphate charges and a dsNA oligomer is shorter than  $N_e = 12 \pm 1$  phosphate charges. (The latter estimation of the characteristic length of CEE for ds NA is obtained from eq 11 using  $\gamma = 1.42 \pm 0.02$  and  $\mathcal{I}_1 = 0.019 \pm 0.003$ .) For a short oligoelectrolyte hairpin helix with  $|Z_D| = 12$  in ds form, the salt dependence of the melting temperature ( $ST_m$ ) is only one-quarter of that observed for the corresponding polymeric NA. The CEE exerts a smaller effect on the [salt] dependence of stability of short two-stranded helices because the total extent of salt ion accumulation by the two denatured strands (with a total of four ends exhibiting the CEE and with each strand containing only half of the number of charges of the ds form) is significantly less than that of the one denatured strand [which has the same number of ends (two) and charges as the ds form] obtained by melting a hairpin helix. As a consequence, the amount of ion release  $\Delta n$  upon melting an oligomeric ( $> 18$  charges) two-stranded helix is much larger than for a hairpin helix of same number of charges  $|Z|$  in ds form and in some cases is coincidentally close to the amount of ion release from melting corresponding polymeric NA. For shorter two-stranded helices ( $\leq 18$  charges), where the single strands formed in the transition are shorter than  $N_e = 9$ , the CEE causes a significant reduction in  $ST_m$ . For example,  $ST_m$  is reduced from 16 to 11 °C when the number of charges in the two-stranded helical form is reduced from 18 to 10.

*Other Measures of Coulombic End Effect at 0.1–0.25 M Salt.* The ion accumulation  $n_u$  per residue of an oligomer or polyelectrolyte is the derivative of the per residue chemical potential  $G_u^\circ$  with respect to the logarithm of mean ionic activity of the salt (eq 2). The same relation approximately holds for each pair of model coefficients

$$n_\alpha \cong - \frac{1}{RT} \left( \frac{\partial G_\alpha^\circ}{\partial \ln a_\pm} \right) \quad (32)$$

where  $n_\alpha = \{n_{ci}, n_{u,0}, n_{u,\infty}, \mathcal{I}_1, \gamma\}$  and the corresponding  $G_\alpha^\circ = \{G_{ci}^\circ, G_{u,0}^\circ, G_{u,\infty}^\circ, \mathcal{I}_1, g\}$ . Coefficients  $n_\alpha$ , defined in the expressions (eqs 6 and 7) relating salt ion accumulation to oligocation valence (or number of charges), are approximately constant between 0.1 and 0.25 M [salt]. Coefficients  $G_\alpha^\circ$ , defined in the expressions (eqs 8 and 9) relating the chemical potential of an oligocation to its valence, are [salt] dependent because of eq 32. Given a value at one [salt] (e.g., 0.1 or 0.2 M from Table 2), a value at another [salt] is found by integrating eq 32:

$$G_\alpha^\circ(a_\pm) = G_\alpha^\circ(a_{\pm,\text{ref}}) - RT \int_{a_{\pm,\text{ref}}}^{a_\pm} n_\alpha d(\ln a_\pm) \cong G_\alpha^\circ(a_{\pm,\text{ref}}) - RT n_\alpha \ln(a_\pm/a_{\pm,\text{ref}}) \quad (33)$$

In eq 33,  $a_{\pm,\text{ref}}$  is the reference salt activity; in the present context either  $a_{\pm,\text{ref}} = 0.083$  M ( $[\text{Na}^+] = 0.1$  M) or  $a_{\pm,\text{ref}} = 0.15$  M ( $[\text{Na}^+] = 0.2$  M) can be used. Equation 33 is accurate over the range 0.1–0.25 M salt ( $0.083 \text{ M} \leq a_\pm \leq 0.19 \text{ M}$ )

where  $\ln K_{\text{obs}}$  is linear with respect to  $\ln a_\pm$  (26). Values of coefficients  $G_\alpha^\circ$  predicted at 0.2 M by eq 33 with 0.1 M as a reference point agree well with the values at 0.2 M [salt] in Table 2.

*Comparison of Model Coefficients with Theoretical Predictions.* Determination of  $\gamma$  for ssDNA from the  $S_0 K_{\text{obs}}$  vs  $|Z_D|$  data (cf. Figure 3B) predicts a value of  $0.77 \pm 0.16$ , and theoretical calculations of  $\gamma$  with the NLPB cylindrical model of the ssDNA oligomer yield a value of  $0.60 \pm 0.02$  (18). The good agreement between the two values of  $\gamma$  supports the thermodynamic picture that Coulombic interactions and ion release (or uptake) are the predominant source of the [salt] dependence of the nucleic acid processes in this [salt] range. The same set of model coefficients ( $N_e, n_{u,0}, n_{u,\infty}, \mathcal{I}_1, \gamma$ ) describes salt ion accumulation in the vicinity of a NA oligomer and the salt dependence of helix melting and oligocation binding, indicating that the CEE model provides a unified quantitative description of [salt] effects in processes involving nucleic acid oligomers and/or involving oligomeric charged ligands.

Since  $n_{u,\infty}$  is accessible via both experiment and theoretical calculations (18), we can estimate the extent of salt ion accumulation by an isolated charge on a short NA oligomer ( $n_{u,0}$ ) from the difference ( $n_{u,\infty} - n_{u,0}$ ) of Table 2. The NLPB calculations for the cylindrical model of the ssDNA oligomer (11, 36) show that  $n_{u,\infty} = 0.454 \pm 0.002$  at 0.15 M univalent salt. Therefore,  $n_{u,0} = n_{u,\infty}^{\text{NLPB}} - (n_{u,\infty} - n_{u,0})^{\text{exp}}$  or, equivalently,  $n_{u,0} = 0.11 \pm 0.07$  salt ion accumulated by a charge in an oligomer in the limit as oligomer charge approaches zero. Salt ion accumulation from Coulombic interactions with a weakly charged species is much smaller than salt ion accumulation in the vicinity of a phosphate within the interior region of a polymeric single-stranded nucleic acid ( $0.11 \pm 0.07$  vs  $0.454 \pm 0.002$  at 0.15 M 1:1 salt). Our preliminary NLPB calculations of  $n_{u,0}$  from the all-atom model of short ( $|Z_D| < 9$ ) ssDNA oligomers are consistent with this value and provide means of an alternative estimation of the model coefficients.

When  $|Z| = 1$  is substituted into eq 6, it yields  $n - n_{ci} = n_{u,0} + \mathcal{I}_1 \cong 0.13 \pm 0.07$  salt ions accumulated because of Coulombic interactions by an individual (isolated) nucleotide at 0.1–0.25 M salt. We predict that this limiting value would also be obtained as the limiting case for analysis of binding of  $L^{Z+}$  to a series of short multistranded NA oligomers and of melting transitions of multistranded NA oligomers.

*A Coulombic End Effect on Protein–DNA Binding.* Do protein–nucleic acid interactions exhibit Coulombic end effects (CEE) comparable to those observed for interactions of  $L^{Z+}$ ? To what extent does the protein behave like an oligocation and contribute only a minor part to  $SK_{\text{obs}}$  when binding to polyelectrolyte DNA at a typical salt concentration? In the absence of higher valent salt cations, the value of  $SK_{\text{obs}}$  describing the effects of 1:1 salt concentration on a protein–nucleic acid interaction is typically independent of salt concentration over the observable range. By analogy with the binding of a sufficiently short oligocation to ss or ds polyanionic nucleic acids, this  $SK_{\text{obs}}$  can be used to define an effective charge on the binding surface of the protein  $Z_{L,\text{eff}}$  assuming for ssDNA that

$$SK_{\text{obs},\infty}^{\text{ss}} \cong -0.7Z_{L,\text{eff}} \quad (34)$$

Table 3: Salt Effects on a Binding Human  $\beta$  Polymerase to ssDNA<sup>a</sup>

binding mode (nucleotide)	DNA	$K_{\text{obs}}$ (M <sup>-1</sup> )	$SK_{\text{obs}}$	$Z_{\text{L,eff}}$
5	polymeric	$(1.6 \pm 0.6) \times 10^5$	$-2.9 \pm 0.6$	$4.1 \pm 1.1^b$
	16-mer	$(2 \pm 1) \times 10^4$	$-2.1 \pm 0.5$	$4.6 \pm 3.8^c$
16	polymeric	$(1 \pm 0.5) \times 10^6$	$-6 \pm 1$	$8.6 \pm 1.5^b$
	16-mer	$(7 \pm 3) \times 10^5$	$-3.9 \pm 0.7$	$11.8 \pm 3.2^c$

<sup>a</sup> Values of  $K_{\text{obs}}$  are at 100 mM Na<sup>+</sup> and 1 mM Mg<sup>2+</sup>; values of  $SK_{\text{obs}}$  are at 50–150 mM Na<sup>+</sup> and 1 mM Mg<sup>2+</sup>. <sup>b</sup> Determined from  $SK_{\text{obs}}$  using eq 34. <sup>c</sup> Determined from  $(SK_{\text{obs},\infty} - SK_{\text{obs},16})$  using the distributed binding model and CEE analysis (see text).

and  $SK_{\text{obs},\infty}^{\text{ds}} \cong -0.9Z_{\text{L,eff}}$  for dsDNA (6–9). (Here the subscript “ $\infty$ ” refers to binding of the protein to polymeric DNA.) For a binding surface of a protein of effective charge  $Z_{\text{L,eff}}$ , are the CEE on its interaction with oligoanionic DNA or RNA the same as for an oligocation of this valence? Data are currently available for only one well-characterized system: human DNA polymerase  $\beta$  (pol  $\beta$ ) binding to ssDNA (27). This enzyme binds to ssDNA in two modes characterized by occluded site sizes on ssDNA of 5 and 16 nucleotides (nt). Equilibrium constants for binding of pol  $\beta$  to both polymeric ssDNA [poly(dεA)] and moderate length oligomeric ssDNA [dεA(pεA)<sub>15</sub>] have been obtained as a function of [salt] (27). Binding constants at 100 mM Na<sup>+</sup> and 1 mM Mg<sup>2+</sup> and the [salt] dependence of binding ( $SK_{\text{obs}}$ ) at 0.05–0.15 M Na<sup>+</sup> and 1 mM Mg<sup>2+</sup> are given in Table 3. Both  $K_{\text{obs}}$  and  $SK_{\text{obs}}$  are significantly smaller in magnitude for binding to the 16-mer oligoanion than to polyanionic DNA. Specifically,  $K_{\text{obs}}$  for pol  $\beta$  binding to the 16-mer in the 5 nt binding mode is 8 times weaker than  $K_{\text{obs}}$  for binding polyanionic ssDNA. For the 16 nt mode,  $K_{\text{obs}}$  for binding the 16-mer is 10 times smaller than  $K_{\text{obs}}$  for binding polyanionic ssDNA at low [salt] (0.05 and 0.075 M Na<sup>+</sup>). (However,  $K_{\text{obs}}$  for the 16 nt mode is similar at higher [salt] (0.10 and 0.15 M Na<sup>+</sup>); the lack of an observed CEE on  $K_{\text{obs}}$  at 0.1–0.15 M salt is unexpected.) The magnitude of the [salt] dependence ( $|SK_{\text{obs}}|$ ) is reduced 30–35% from polymer to oligomer in both modes at any [salt] in this range. Is this a CEE or is it another end effect, such as the loss of contacts from truncation of the lattice? (This possibility is unlikely for the 5 nt mode, though conceivable for the 16 nt mode.)

Calculation of effective charges on the pol  $\beta$  binding sites from eq 34 and the [salt] dependences of binding pol  $\beta$  to polymeric DNA (Table 3) yields  $Z_{\text{L,eff}} = 4$  for the 5 nt binding mode and  $Z_{\text{L,eff}} = 9$  for the 16 nt binding mode. (At 0.05–0.15 M NaCl, the presence of 1 mM Mg<sup>2+</sup> is predicted by cylindrical NLPB model to reduce the magnitude of  $SK_{\text{obs},\infty}$  by no more than experimental uncertainty,  $\sim 10\%$ , and therefore does not significantly affect these estimates of  $Z_{\text{L,eff}}$ .) An independent estimate of  $Z_{\text{L,eff}}$  can be obtained from the difference  $SK_{\text{obs},16} - SK_{\text{obs},\infty}$ , where the subscript “16” refers to binding of the protein to a DNA 16-mer. Analysis of the differences  $SK_{\text{obs},\infty} - SK_{\text{obs},16}$  using the DBM and coefficients from Table 2 (again neglecting the expected small effect of 1 mM Mg<sup>2+</sup> on  $S_aK_{\text{obs}}$ ) yields values for  $Z_{\text{L,eff}}$  of  $\sim 5$  (with a large uncertainty) for the 5 nt mode and  $\sim 12$  for the 16 nt mode. The two predictions for  $Z_{\text{L,eff}}$  are in agreement with each other within uncertainty, yielding the conclusion that pol  $\beta$  binds to ssDNA in the 5 nt mode as an oligocationic ligand with a charge approximately equal

to its site size ( $5 \pm 1$  nt,  $\sim 4 \pm 1$  charges). In the 16 nt mode, pol  $\beta$  binds to ssDNA as an oligocationic ligand with  $\sim 10 \pm 2$  charges.

This picture is also consistent with structural data for the only available  $\beta$  polymerase–DNA complex (37) and with the proposed model of binding of pol  $\beta$  to ssDNA (27). In the structure of a complex of pol  $\beta$  with 16 bp dsDNA (37), the number of protein charges that are in the vicinity of DNA phosphate charges [within 8 Å, calculated by CCP4 Suite (38)] is 8 positive and 3 negative charges for the 5′ to 3′ DNA strand (referred to as chain T in ref 37) and 10 positive and 4 negative charges for the 3′ to 5′ gapped DNA strand [referred to as chains P and D (37)]. The 3 phosphates of chain D of dsDNA are in contact with 6 positive and 2 negative charges from the small 8 kDa domain of the protein, which has higher affinity for binding to ssDNA than the other 31 kDa domain of pol  $\beta$  (27). The observed contacts between phosphates of chain D and the 8 kDa domain of pol  $\beta$  (net +4 positive charge on its binding site) agree with the proposal (27) that only the 8 kDa domain of pol  $\beta$  interacts with ssDNA in the 5 nt mode, because  $K_{\text{obs}}$  and  $SK_{\text{obs}}$  for this mode are consistent with binding of an oligocationic ligand with  $Z_{\text{L}} = 4$ . The effective charge on the binding surface of the protein deduced from  $SK_{\text{obs}}$  for the 16 nt mode ( $Z_{\text{L,eff}} = 10 \pm 2$ ) is the same within uncertainty as the net charge on both the 8 and 31 kDa domains of the protein within 8 Å of the DNA in the above structural analysis (net +11 charge; 18 positive, 7 negative). This is consistent with the structural proposal that both 8 and 31 kDa domains interact with the ssDNA in the 16 nt binding mode (27). The structural analysis (37) also suggests that 16 bp dsDNA is long enough to have all contacts with charges on the protein surface of pol  $\beta$  in both 5 and 16 nt binding modes. Therefore, the most likely explanation of the observed reductions in  $K_{\text{obs}}$  and  $|SK_{\text{obs}}|$  (Table 3) is the CEE.

The above analysis, in terms of an effective charge on the interface of the protein ( $Z_{\text{L,eff}}$ ), is of course a very approximate method to treat CEE. Proteins differ greatly from oligocations; even if the binding interface is oligocationic, it is surrounded by the remainder of the heterogeneously charged protein exterior. Charges on the protein in the vicinity of (but not in) the interface may have a significant effect on  $K_{\text{obs}}$  and  $SK_{\text{obs}}$ . For example, individual changes in cationic and anionic residues at distances of 3–11 Å from the interface of a peptide–RNA complex (with 4 positive charges in the interface with RNA) were shown by experiment and NLPB calculations to modulate  $SK_{\text{obs}}$  by up to 50% of the amount expected from the same change in the number of charges in the interface (40). The numerous anionic residues flanking the DNA binding interface of IHF appear to reduce  $|SK_{\text{obs}}|$  to less than half the value predicted for binding of an oligocation with the same number of positive charges as in the interface with dsDNA (39).

Recently obtained structures of complexes of a disulfide cross-linked dimer of the lac repressor DNA-binding domain with specific (22 bp,  $|Z_{\text{D}}| = 40$ ) and nonspecific (18 bp,  $|Z_{\text{D}}| = 34$ ) DNA sequences show that DNA–protein charge–charge interactions are localized at the 3–5 phosphates at each terminus of DNA oligomers (41). Our model predicts that the CEE will be manifested as a significant reduction in  $|S_aK_{\text{obs}}|$  (by up to  $\sim 1.6$  units at 0.1–0.3 M 1:1 [salt]) for binding to these DNA oligomers as compared to  $|S_aK_{\text{obs}}|$  for



binding to the same site embedded in polymeric DNA. Experimental and theoretical studies of the Coulombic end effect in this system are in progress.

## APPENDIX

*Approximation of Length Dependence of  $G^\circ$  and  $n$  by Truncated Taylor Expansions in  $|Z| \rightarrow 0$  and  $|Z| \rightarrow \infty$  Limits.* Two fundamental quantities describe the thermodynamic behavior of ligand–lattice interactions studied here: excess chemical potential ( $G^\circ$ ) and the net salt ion accumulation [ $n \equiv -(RT)^{-1} \partial G^\circ / (\partial \ln a_{\pm})$ ]. To obtain the approximate functional dependences of  $G^\circ$  on the number of charges  $|Z|$  of a charged linear molecule, it is useful to work with the per residue Coulombic quantity  $G_u^\circ \equiv (G^\circ - G_{ci}^\circ)/|Z|$ , where  $G_{ci}^\circ$  represents any charge-independent (ci) and length-independent contribution to the  $G^\circ$  of the species and is the same for all lattice lengths,  $G_{ci}^\circ = G^\circ|_{|Z|=0}$ . For very long polyions,  $G_u^\circ$  is expected to be independent of  $|Z|$  because CEE is insignificant for average (per charge) properties. For large oligomers, where CEE becomes significant, expansion in inverse powers of  $|Z|$  yields

$$G_u^\circ = \tilde{G}_{u,\infty}^\circ - 2\tilde{g} \frac{1}{|Z|} + \sum_{i=2}^{\infty} \frac{1}{i!} \left( \frac{\partial^i G_u^\circ}{\partial (1/|Z|)^i} \right)_{1/|Z|=0} \frac{1}{|Z|^i} \quad (35)$$

where  $\tilde{G}_{u,\infty}^\circ = G_u^\circ|_{1/|Z|=0}$  and  $2\tilde{g} = -[\partial G_u^\circ / \partial (1/|Z|)]|_{1/|Z|=0}$ .

At small  $|Z|$ , the Taylor expansion yields

$$G_u^\circ = \tilde{G}_{u,0}^\circ + \tilde{\mathcal{J}}_1 |Z| + \sum_{i=2}^{\infty} \frac{1}{i!} \left( \frac{\partial^i G_u^\circ}{\partial |Z|^i} \right)_{|Z|=0} |Z|^i \quad (36)$$

where  $\tilde{G}_{u,0}^\circ = G_u^\circ|_{|Z|=0}$  and  $\tilde{\mathcal{J}}_1 = [\partial G_u^\circ / \partial |Z|]|_{|Z|=0}$ .

The Taylor expansions in general are slowly convergent, and truncations as in the text would be expected to be valid over limited ranges of  $1/|Z|$  and  $|Z|$ , respectively. Nevertheless, we found that truncating each expansion at two terms and matching them at some intermediate length to obtain the corresponding coefficients (which differ from those of original Taylor series) provide a very good approximation to the overall length dependence of  $G^\circ(|Z|)$ :

$$\text{short: } G_u^\circ \cong \tilde{G}_{u,0}^\circ + \tilde{\mathcal{J}}_1 |Z| \quad (37)$$

$$\text{long: } G_u^\circ \cong \tilde{G}_{u,\infty}^\circ - 2\tilde{g}/|Z| \quad (38)$$

Analysis of the  $|Z|$  dependence of the per charge salt ion association quantity  $n_u = (n - n_{ci})/|Z|$  proceeds similarly to give

$$\text{short: } n_u \cong \tilde{n}_{u,0} + \tilde{\mathcal{J}}_1 |Z| \quad (39)$$

$$\text{long: } n_u \cong \tilde{n}_{u,\infty} - 2\tilde{\gamma}/|Z| \quad (40)$$

Equations 37–40 can be rearranged to the equations presented in the text (eqs 6–9):

$$n = n_{ci} + n_{u,0}|Z| + \mathcal{J}_1 |Z|^2, \quad |Z| < N_e \quad (41)$$

$$n = n_{ci} + n_{u,\infty}|Z| - 2\gamma, \quad |Z| \geq N_e \quad (42)$$

$$G^\circ = G_{ci}^\circ + G_{u,0}^\circ |Z| + \mathcal{J}_1 |Z|^2, \quad |Z| < N_e \quad (43)$$

$$G^\circ = G_{ci}^\circ + G_{u,\infty}^\circ |Z| - 2g, \quad |Z| \geq N_e \quad (44)$$

Whereas  $\tilde{G}_{u,0}^\circ$ ,  $\tilde{\mathcal{J}}_1$ ,  $\tilde{G}_{u,\infty}^\circ$  ( $\tilde{n}_{u,0}$ ,  $\tilde{\mathcal{J}}_1$ ,  $\tilde{n}_{u,\infty}$ ), and  $2\tilde{g}$  ( $2\tilde{\gamma}$ ) represent true expansion coefficients which describe the asymptotic behavior of  $G^\circ$  and  $n$  as  $|Z| \rightarrow 0$  or  $|Z| \rightarrow \infty$ , respectively,  $G_{u,0}^\circ$ ,  $\mathcal{J}_1$ ,  $G_{u,\infty}^\circ$  ( $n_{u,0}$ ,  $\mathcal{J}_1$ ,  $n_{u,\infty}$ ) and  $2g$  ( $2\gamma$ ) are approximations of these coefficients. The set of coefficients differ from the Taylor coefficients as a result of the Taylor series truncation error, the smooth continuity constraints (eqs 10–13), and of fitting equations of Table 1 (derived from eqs 6–9) to the actual  $\Delta G^\circ$  and  $\Delta n$ . Consequently, eqs 6 and 7 are approximations of the true functional dependence of  $n$  on  $|Z|$ . The number of terms in eqs 6–9 is the minimum necessary to account for the deviation from the polymeric limit (via  $\gamma$  and  $g$ ) and yet still allow continuous first derivatives of  $G^\circ(|Z|)$  and  $n(|Z|)$  across the full range of  $|Z|$  values.

In general,  $G^\circ$  and  $n$  depend on the number of oligomer charges  $|Z|$ , the monomer sequence, [salt], and salt composition, etc. The expansions (eqs 35 and 36) consider only the length dependence. Consequently, the expansion coefficients will be functions of all other significant variables (DNA composition, structure, and solution conditions).

*Determination of  $f_G$  ( $f_n$ ) from the Assumption of Noninteracting Flanking Regions in the Limit of Large Ligand.* With the assumptions specified in the Results section, the values of the chemical potential  $G_{DL}^\circ$  and  $n_{DL}$  of the complex are modeled as polynomial functions of  $|Z_D|$  and  $Z_L$ . In principle, one can build Taylor expansions of  $G_{DL}^\circ$  and  $n_{DL}$  similar to eqs 35 and 36 considering various ranges of  $|Z_D|$  and  $Z_L$ . But with two independent variables ( $|Z_D|$  and  $Z_L$ ) the large number of permutations makes the derivation lengthy. We use an equivalent and more intuitive approach relating complex chemical potential  $G_{DL}^\circ$  (or  $n_{DL}$ ) to chemical potential (or  $n$ ) of charged flanking regions. We assume that the chemical potential of the central complex is equal to the sum of chemical potentials of the two charged ends (each described by eq 6 for a  $Q/2$  charged oligomer) and of the neutralized part of the complex [represented here by the term  $f_{G,n}(Z_L)$ , which is dependent only on  $Z_L$ ]. For a central complex formed between a large enough oligocation and a somewhat larger lattice so that the flanking regions are short and noninteracting

$$G_{DL,c}^\circ = 2 \left[ G_{ci}^\circ + G_{u,0}^\circ \frac{Q}{2} + \mathcal{J}_1 \left( \frac{Q}{2} \right)^2 \right] + f_G(Z_L) \quad (45)$$

The assumption is tested with NLPB calculations (Figure 6B), which indicate that the contribution of interactions between the two flanking regions to the  $SK_{obs} = \Delta n^{NLPB}$  of the central complex at 0.10–0.25 M salt is less than 5% for  $Z_L \geq 6$ . (The approximated values of  $\Delta n$  differ by less than 2% at  $Z_L = 8$ , 5% at  $Z_L = 6$ , and 9% at  $Z_L = 4$  from the exact values.)

Given eq 45 for a complex with short flanking regions, the change in free energy with complexation is

$$\Delta G_c^\circ \equiv G_{DL,c}^\circ - G_D^\circ - G_L^\circ = 2G_{ci}^\circ + G_{u,0}^\circ Q + \frac{1}{2} \mathcal{J}_1 Q^2 + f_G - G_D^\circ - G_L^\circ \quad (46)$$

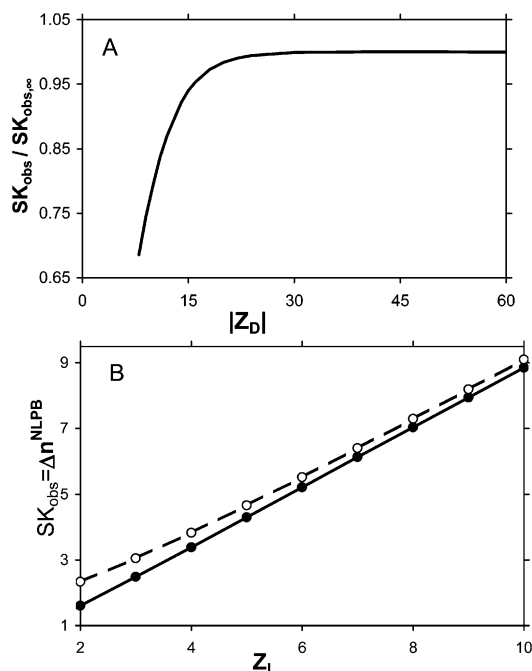


FIGURE 6: NLPB prediction of the log–log salt derivative of the binding constant  $S_a K_{\text{obs}}$  for (A) central binding of  $L^{8+}$  to variable ssDNA length (in charges)  $|Z_D|$  and (B) central binding of  $L^{8+}$  to a ssDNA 60-mer ( $|Z_D| = 60$ ) for  $2 \leq Z_L \leq 10$  (filled symbols, solid line). Open symbols and dashed line show the approximation when  $n_{DL}$  of the central complex is calculated as a sum of  $n$  of two oligomers of  $(60 - Z_L)/2$  charges each. Details of the NLPB calculation are given in ref 18; the central complex is modeled as a cylinder where the charges at the center of the cylinder axis are turned off over the length of the bound ligand ( $Z_L b$ ).

where  $G_L^\circ$  is the ligand chemical potential and  $G_D^\circ$  is the NA chemical potential determined by eq 8 or eq 9. Analogous equations apply to obtain  $\Delta n_c$ .

To determine  $f_G$  and  $f_n$ , we consider the situation where the ligand is large so flanking regions do not interact ( $Z_L \geq 6$ ), and the NA lattice is always sufficiently longer such that there are less than  $N_e$  charges on either side of the bound ligand (i.e.,  $Z_L \gg N_e$  and  $0 \leq Q < 2N_e$ ). We require that with fixed  $Z_L$  and varying  $|Z_D|$ ,  $\Delta G_c^\circ$  and  $\Delta n_c$  for central binding initially decrease and eventually plateau (i.e.,  $|\Delta G_c^\circ|$  and  $|\Delta n_c|$  increase to plateaus) where  $\Delta G_c^\circ = \Delta G_\infty^\circ$  and  $\Delta n_c = \Delta n_\infty$ , as  $|Z_D|$  increases. (Figure 6A provides evidence for such behavior of  $SK_{\text{obs}}$  calculated from the cylindrical NLPB model of ligand–ssDNA binding; details of calculations will be published elsewhere.) We require that  $\Delta G_c^\circ$  and  $\Delta n_c$  have continuous first derivatives at the value of  $|Z_D|$  where their minima are attained.

The first condition on the minimum of  $\Delta G_c^\circ$  [substituting  $G_D^\circ = G_{ci}^\circ + G_{u,\infty}^\circ |Z_D| - 2g$  within this range of  $|Z_D|$  into eq 46 and utilizing the linkage equation (eq 13)] requires

$$\frac{\partial \Delta G_c^\circ}{\partial |Z_D|} = \mathcal{G}_1(|Z_D| - Z_L) + G_{u,0}^\circ - G_{u,\infty}^\circ = 0 \quad (47)$$

which yields

$$|Z_D| = Z_L + 2N_e \quad (48)$$

Equating  $\Delta G_c^\circ$  from eq 46 to  $\Delta G_\infty^\circ$  at  $|Z_D| = Z_L + 2N_e$  ( $Q = 2N_e$ ) we obtain

$$\Delta G_\infty^\circ = 2G_{ci}^\circ + 2G_{u,0}^\circ N_e + 2\mathcal{G}_1 N_e^2 + f_G - G_{ci}^\circ - G_{u,\infty}^\circ (Z_L + 2N_e) + 2g - G_L^\circ \quad (49)$$

$\Delta G_\infty^\circ$  is the limit of  $\Delta G_c^\circ$  at fixed  $Z_L$  and  $|Z_D| \rightarrow \infty$  and is a function of  $Z_L$  only. The value of this function is known to us experimentally for oligolysine with  $Z_L = 8$ . We use eq 49 to express  $f_G$  via  $\Delta G_\infty^\circ$ ,  $G_L^\circ$ , and model coefficients

$$f_G(Z_L) = \Delta G_\infty^\circ - G_{ci}^\circ + G_L^\circ + G_{u,\infty}^\circ Z_L + \mathcal{G}_1 N_e^2 \quad (50)$$

after simplification using eqs 12 and 13. Substitution of eq 50 into eq 46, along with the appropriate expression for  $G_D^\circ$ , provides the final expression (eq 20 in Table 1).

*Complexes Where the Peptide Has More Charges Than the Nucleic Acid ( $Z_L > |Z_D|$ ).* Equations in Table 1 are derived for  $|Z_D| \geq Z_L$  and expressed through NA lattice coefficients only ( $\mathcal{G}_1$ ,  $\mathcal{A}_1$ , and  $N_e$ ). Corresponding coefficients for the ligand in general may be different from NA coefficients. They are implicitly included in  $\Delta G_\infty^\circ$  and  $S_a K_\infty$ , which are measured experimentally (24–26).

The case of the shortest DNA oligomer, dT(pdT)<sub>6</sub>, requires special treatment. Simply following the derivation in the text produces expressions similar to eqs 20–31, where  $|Z_D|$  and  $Z_L$  are interchanged and  $\Delta G_\infty^\circ(Z_L)$  [ $S_a K_\infty(Z_L)$ ] are replaced with  $\Delta G_\infty^\circ(|Z_D|)$  [ $S_a K_\infty(|Z_D|)$ ], which are the polyelectrolyte values for binding dT(pdT)<sub>6</sub> to polylysine. Such information is not available to us, and we need to relate  $\Delta G^\circ$  for binding of KWK<sub>6</sub><sup>8+</sup> to dT(pdT)<sub>6</sub><sup>6−</sup> to the same set of model coefficients as binding free energies for other DNA oligomer lengths. We do this with the additional assumption that complexes with short flanking regions ( $|Q| < N_e$ ) have approximately the same  $G^\circ$  and  $n$  independent of which lattice, DNA or oligolysine, is longer and that  $G^\circ$  and  $n$  are determined by the number of charges  $|Q|$  only. For example,  $G^\circ$  (and  $n$ ) for the complex of KWK<sub>6</sub><sup>8+</sup> and dT(pdT)<sub>6</sub><sup>6−</sup> ( $Q = -2$ ) is assumed to be approximately equal to  $G^\circ$  (and  $n$ ) for the complex of KWK<sub>4</sub><sup>6+</sup> and dT(pdT)<sub>8</sub><sup>8−</sup> ( $Q = 2$ ). We assume that the shortest lattice occupies the same number of charges as its valence on the long lattice. In this case, the two complexes differ only by sign (and not by number) of charges in flanking regions (Figure 2C). The above assumption that the contribution of short flanking regions to complex chemical potential strongly depends on the number of charges, but not on the particular details of charge arrangement on the lattice, yields

$$G_{LD,c,+6,-8}^\circ = 2 \left[ G_{ci}^\circ + G_{u,0}^\circ \frac{|Z_D| + Z_L}{2} + \mathcal{G}_1 \left( \frac{|Z_D| - Z_L}{2} \right)^2 \right] + f_G(Z_L) - 2G_{u,0}^\circ Z_L \cong G_{LD,c,-6,+8}^\circ = 2 \left[ \hat{G}_{ci}^\circ + \hat{G}_{u,0}^\circ \frac{|Z_D| + Z_L}{2} + \hat{\mathcal{G}}_1 \left( \frac{Z_L - |Z_D|}{2} \right)^2 \right] + \hat{f}_G(|Z_D|) - 2\hat{G}_{u,0}^\circ |Z_D|$$

Here  $\hat{G}_{ci}^\circ$ ,  $\hat{G}_{u,0}^\circ$ , and  $\hat{\mathcal{G}}_1$  are coefficients for the ligand. If we assume that they are approximately equal to DNA coefficients, the two bracketed terms in the preceding equation are equal. Then  $f_G(Z_L) - 2G_{u,0}^\circ Z_L$  and  $\hat{f}_G(|Z_D|) - 2\hat{G}_{u,0}^\circ |Z_D|$  are constants (they have to be equal to each other and therefore cannot depend on different length variables). The

expression for the complex is symmetric with respect to  $Z_L$  and  $|Z_D|$ , and therefore eqs 26 and 27 for complexes with short flanking regions are valid as written for the case of  $|Z_D| = 6$  and  $Z_L = 8$ . The assumption of equal model coefficients for NA and ligand lattices is tested by NLPB calculations with the cylindrical model of the DNA oligomer, showing that model coefficients in the limit of small  $|Z|$  (e.g.,  $n_{u,0}$  and  $\mathcal{N}$ ) are not sensitive to the structural details of the lattice (to be published).

**Contribution of Multiple Binding to  $K_{obs}$  and  $S_a K_{obs}$  for  $KWK_6-dT(pdT)_{|Z_D|}$  Interactions.** For DNA oligomers longer than  $2Z_L$  charges, more than one cationic ligand can bind to the lattice. Coulombic end effects provide a second source of anticooperativity in addition to the ligand size effect (42) on  $K_{obs}$  for binding of multiple ligands to such a lattice. CEE reduces the zero binding density binding constant for an oligomer from that characteristic for binding of the ligand to a polymeric lattice where CEE is absent. Moreover, binding of the first ligand creates two flanking regions of anionic lattice of smaller length than the original oligomer, both of which exhibit either trapezoidal or parabolic profiles of cation concentration (22) and binding free energy depending on their length. Thus, for a long oligomer allowing multiple bound ligands, binding of each subsequent ligand becomes increasingly less favorable as the flanking regions, on average, become shorter and change from a trapezoidal to a parabolic distribution of surface counterion concentration and binding free energy.

$dT(pdT)_{22}$  represents an intermediate case between a long and short lattice. From the analysis of  $KWK_6-dT(pdT)_{|Z_D|}$  binding data with  $|Z_D| = 6-15$ ,  $dT(pdT)_{22}$  is predicted to be a long oligomer ( $|Z_D| \geq N_e$ ) with parabolic distribution of  $\Delta G_i^\circ$  for ligands with  $Z_L = 8$  ( $N_e < Q < 2N_e$ ).  $dT(pdT)_{22}$  can have up to two ligands simultaneously bound if overhanging modes are ignored. If the DBM (eqs 20, 22, and 28, Table 1) is applied to the first binding event,  $\log K_1$  is predicted to be  $6.0 \pm 0.1$ . The binding constant for the second binding event is related to  $K_{i,j}$ , the binding constants for two ligands bound at the  $i$ th and  $(i + j + Z_L)$ th positions on the lattice as  $K_2 = (K_1)^{-1} \sum_{i=0}^{|Z_D|-2Z_L} (\sum_{j=0}^{|Z_D|-2Z_L-i} K_{i,j})$ . Development of model expressions for  $K_{i,j}$  and multiple binding is outside of the scope of this paper. An estimation of  $K_2$  can be obtained by assuming a model in which the second ligand interacts with an unoccupied oligomer lattice with only 8–14 charges. The upper limit for binding to a site with 8–14 consequent unoccupied monomers is obtained from the DBM for a 14-mer, yielding  $\log K_2 \leq 5.2 \pm 0.1$ . The observed binding constant is bounded between two model predictions for  $K_1$  and  $K_2$  showing that finite lattice isotherm with equivalent sites (44) (without end effects) yields an average of two binding constants. This is also consistent with the observations from experimental data (26). The titration curves are statistically better described (26) when two different microscopic binding constants are used in the binding isotherm:

$$\frac{\nu}{[L]} = \frac{\Omega_1 K_1 [L] + 2\Omega_2 K_1 K_2 [L]^2}{1 + \Omega_1 K_1 [L] + \Omega_2 K_1 K_2 [L]^2} \quad (51)$$

where  $\Omega_1 = |Z_D| - Z_L + 1$  and  $\Omega_2 = (|Z_D| - 2Z_L + 2)(|Z_D| - 2Z_L + 1)/2$ . For  $L^{8+}$  and  $dT(pdT)_{22}$  under conditions

investigated (0.1 M salt) the first microscopic binding constant,  $K_1$  ( $\log K_1 = 5.73 \pm 0.01$ ), is approximately a factor of 2 larger than the second,  $K_2$  ( $\log K_2 = 5.43 \pm 0.02$ ). Thus the binding constant determined from the finite lattice model without end effects ( $K_{obs} = 5.6 \pm 0.04$ ) is bounded above by  $K_1$  and below by  $K_2$ . These results support model prediction for CEE in multiple bound ligands. In particular, that after binding the first ligand to such a lattice, subsequent ligands bind more weakly; each subsequent binding event can be modeled as an interaction with a shorter lattice of a length equal to the number of contiguous noncomplexed charges.

**Introduction of CEE in the Binding Isotherm.** Our results allow a test of the finite lattice model (44) and 1:1 binding model in applications to binding of a cationic ligand to DNA oligomers. Short complexes with  $|Q| < N_e$  do not exhibit a large difference in binding constant between the end and center of the oligomer. Thus each  $K_i$  is well approximated by  $K_c$ . If, in addition, the NA lattice allows only a single bound ligand,  $Q < Z_L$ , then fitting of experimental data with the 1:1 binding model yields a good estimation of  $K_{obs} \cong K_c$ , with error not exceeding 30% for  $L^{8+}$  at 0.1–0.25 M salt. Thus, for DNA–ligand complexes, where  $Q < \min\{N_e, Z_L\}$ , inclusion of CEE into the 1:1 binding isotherm is sufficient at the level of application of central or distributed (eq 4 and eqs 20–25) model expressions for the interpretation of  $K_{obs}$ , obtained by the 1:1 isotherm.

Intermediate length complexes ( $\min\{N_e, Z_L\} < Q \leq 2N_e$ ) exhibit a more significant difference between  $K_c$  and ensemble-averaged  $K_{obs}$  calculated from DBM. If an accuracy better than 50% is required for predicting a binding constant, one should use the DBM to interpret results obtained from experimental data by fitting the 1:1 binding isotherm. More complexity is introduced when multiple binding is present ( $Q \geq Z_L$ ). As discussed above, the use of the finite lattice model without end effects (44) to fit experimental data for  $dT(pdT)_{22}$  does not yield a low-density limit of binding constant, but an average of binding constants for complexes with a single ligand and a pair of bound ligands. The better fit to experimental data of a modified isotherm (eq 51) with distinct binding constants for binding of a single ligand and for a pair of ligands allows a more detailed interpretation of experimental data than the finite lattice model with a single binding constant. However, additional theoretical work is required to interpret  $K_2$  (eq 51) in terms of model coefficients of Table 2 and dissect effects of multiple binding modes on  $K_{obs}$ .

Long complexes with  $Q \geq 2N_e$  exhibit only small deviations of  $\log K_{obs}$  from the polyelectrolyte interior value. As in the case of multiple binding to intermediate length complexes, further development of the model to include multiple binding is necessary.

**Comparison of the Distributed Binding Model to Previous Models of CEE.** The distributed model replaces the two-state model of CEE suggested in ref 25, where distribution of binding free energy along the DNA oligomer was modeled as a polymeric interior and  $N_E$  binding sites on both DNA termini with binding free energy much smaller than that for the polymeric interior. The Coulombic end effect length  $N_e$ , defined here, represents the region on each end of the lattice where binding free energy is different from the polymeric interior. The determination of  $N_E$  in the two-state model of



CEE (25) approximated the free energy of binding to the end of the oligomer (which is not an experimentally accessible value) by the free energy of binding to a short oligomer. The current model provides the entire distribution of binding constants along the oligomer with a set of coefficients that are obtainable by fitting experimental data.

Recently (26), the CEE length  $N_e$  was estimated as  $11 \pm 4 + \sum_i K_i/K_\infty$  from long oligomer  $|Z_D| \geq 39$  experimental data. For such oligomers,  $\sum_i K_i/K_\infty$  (the ratio of the sum of binding constants over one terminal region to a single site polymeric binding constant) does not depend on  $|Z_D|$ . The large error in this estimation originates from the large relative error in  $\log K_{\text{obs}} - \log K_{\text{obs},\infty}$  for long oligomers. The unknown contribution to  $\log K_{\text{obs}} - \log K_{\text{obs},\infty}$  from multiple binding also produces uncertainty in  $N_e$ . Data for short oligomers ( $|Z_D| \leq 15$ ) of the type presented in ref 26 and analyzed here and in ref 26 are free of these complications and provide a more accurate determination of the CEE length  $N_e = 9.0 \pm 0.8$ .

## ACKNOWLEDGMENT

We thank Prof. J. C. Weisshaar for multiple discussions of the manuscript and Dr. Ruth Saecker for helpful discussions of DNA–protein interactions.

## REFERENCES

- Latt, S. A., and Sober, H. A. (1967) Protein-nucleic acid interactions. II. Oligopeptide-polyribonucleotide binding studies, *Biochemistry* 6, 3293–3306.
- Record, M. T., Jr., Lohman, T. M., and DeHaseth, P. (1976) Ion effects on ligand-nucleic acid interactions, *J. Mol. Biol.* 107, 145–158.
- Record, M. T., Jr., Anderson, C. F., and Lohman, T. M. (1978) Thermodynamic analysis of ion effects in the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity, *Q. Rev. Biophys.* 11, 103–178.
- Lohman, T. M., deHaseth, P. L., and Record, M. T., Jr. (1980) Pentalsine-deoxyribonucleic acid interactions: A model for the general effects of ion concentrations on the interactions of proteins with nucleic acids, *Biochemistry* 19, 3522–3530.
- Plum, G. E., and Bloomfield, V. A. (1988) Equilibrium dialysis study of binding of hexamine cobalt (III) to DNA, *Biopolymers* 27, 1045–1051.
- Mascotti, D. P., and Lohman, T. M. (1990) Thermodynamic extent of counterion release upon binding oligolysines to single-stranded nucleic acids, *Proc. Natl. Acad. Sci. U.S.A.* 87, 3142–3146.
- Mascotti, D. P., and Lohman, T. M. (1992) Thermodynamics of single-stranded RNA binding to oligolysines containing tryptophan, *Biochemistry* 31, 8932–8946.
- Mascotti, D. P., and Lohman, T. M. (1993) Thermodynamics of single-stranded RNA and DNA interactions with oligolysines containing tryptophan. Effects of base composition, *Biochemistry* 32, 10568–10579.
- Mascotti, D. P., and Lohman, T. M. (1997) Thermodynamics of oligoarginines binding to RNA and DNA, *Biochemistry* 36, 7272–7279.
- Record, M. T., Jr., Zhang, W., and Anderson, C. F. (1998) Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: A practical guide to recognizing and interpreting polyelectrolyte effects, Hofmeister effects, and osmotic effects of salts, *Adv. Protein Chem.* 51, 281–353.
- Bond, J. P., Anderson, C. F., and Record, M. T., Jr. (1994) Conformational transitions of duplex and triplex nucleic acid helices: Thermodynamic analysis of effects of salt concentration on stability using preferential interaction coefficients, *Biophys. J.* 67, 825–836.
- Bloomfield, V. A., Crothers, D. M., and Tinoco, I., Jr. (2000) *Nucleic Acids, Structure, Properties, and Functions*, University Science Books, Sausalito, CA.
- Manning, G. S. (1969) Limiting laws and counterion condensation in polyelectrolyte solutions I. Colligative properties, *J. Chem. Phys.* 51, 924–938.
- Manning, G. S., and Ray, J. (1998) Counterion condensation revisited, *J. Biomol. Struct. Dyn.* 16, 461–476.
- Stigter, D., and Dill, K. A. (1996) Binding of ionic ligands to polyelectrolytes, *Biophys. J.* 71, 2064–2074.
- Misra, V. K., and Draper, D. E. (1999) The interpretation of  $\text{Mg}^{2+}$  binding isotherms for nucleic acids using Poisson–Boltzmann theory, *J. Mol. Biol.* 294, 1135–1147.
- Zacharias, M., Luty, B. A., Davis, M. E., and McCammon, J. A. (1992) Poisson–Boltzmann analysis of the lambda-repressor-operator interaction, *Biophys. J.* 63, 1280–1285.
- Shkel, I. A., and Record, M. T., Jr. (2004) Effect of the number of nucleic acid oligomer charges on the salt dependence of stability ( $\Delta G_{37}^\circ$ ) and melting temperature ( $T_m$ ): NLPB analysis of experimental data, *Biochemistry* 43, 7090–7101.
- Record, M. T., Jr., and Lohman, T. M. (1978) A semiempirical extension of polyelectrolyte theory to the treatment of oligoelectrolytes: Application to oligonucleotide helix-coil transitions, *Biopolymers* 17, 159–166.
- Olmsted, M. C., Anderson, C. F., and Record, M. T., Jr. (1989) Monte Carlo description of oligoelectrolyte properties of DNA oligomers: Range of the end effect and the approach of molecular and thermodynamic properties to the polyelectrolyte limits, *Proc. Natl. Acad. Sci. U.S.A.* 86, 7766–7770.
- Olmsted, M. C., Anderson, C. F., and Record, M. T., Jr. (1991) Importance of oligoelectrolyte end effects for the thermodynamics of conformational transitions of nucleic acid oligomers: A grand canonical Monte Carlo analysis, *Biopolymers* 31, 1593–1604.
- Olmsted, M. C., Bond, J. P., Anderson, C. F., and Record, M. T., Jr. (1995) Grand canonical Monte Carlo molecular and thermodynamic predictions of ion effects on binding of an oligocation ( $\text{L}^{8+}$ ) to the center of DNA oligomers, *Biophys. J.* 68, 634–647.
- Allison, S. A. (1994) End effects in electrostatic potentials of cylinders: Models for DNA fragments, *J. Phys. Chem.* 98, 12091–12096.
- Zhang, W., Bond, J. P., Anderson, C. F., Lohman, T. M., and Record, M. T., Jr. (1996) Large electrostatic differences in the binding thermodynamics of a cationic peptide to oligomeric and polymeric DNA, *Proc. Natl. Acad. Sci. U.S.A.* 93, 2511–2516.
- Zhang, W., Ni, H., Capp, M. W., Anderson, C. F., Lohman, T. M., and Record, M. T., 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 ( $\text{L}^{8+}$ ) binding to single-stranded DNA oligomers, *Biophys. J.* 76, 1008–1017.
- Ballin, J. D., Shkel, I. A., and Record, M. T., Jr. (2004) Interactions of the KWK<sub>6</sub> cationic peptide with short nucleic acid oligomers: Demonstration of large Coulombic end effects on binding at 0.1–0.2 M salt, *Nucleic Acids Res.* 32, 3271–3281.
- Rajendran, S., Jezewska, M. J., and Bujalowski, W. (1998) Human DNA polymerase  $\beta$  recognizes single-stranded DNA using two different binding modes, *J. Biol. Chem.* 273, 31021–31031.
- Anderson, C. F., and Record, M. T., Jr. (1993) Salt dependence of oligoion-polyion binding: A thermodynamic description based on preferential interaction coefficients, *J. Phys. Chem.* 97, 7116–7126.
- Wilcoxon, J. P., and Schurr, J. M. (1983) Electrophoretic light scattering studies of poly(L-lysine) in the ordinary and extraordinary phase. Effects of salt, molecular weight, and polyion concentration, *J. Chem. Phys.* 78, 3354–3364.
- Padmanabhan, S., Zhang, W., Capp, M. W., Anderson, C. F., and Record, M. T., Jr. (1997) Binding of cationic (+4) alanine- and glycine-containing oligopeptides to double-stranded DNA: Thermodynamic analysis of effects of Coulombic interactions and  $\alpha$ -helix induction, *Biochemistry* 36, 5193–5206.
- Johnson, N. P., Lindstrom, J., Baase, W. A., and von Hippel, P. H. (1994) Double-stranded DNA templates can induce  $\alpha$ -helical conformation in peptides containing lysine and alanine: Functional implications for leucine zipper and helix-loop-helix transcription factors, *Proc. Natl. Acad. Sci. U.S.A.* 91, 4840–4844.
- Strauss, U. P., Helfgott, C., and Pink, H. (1967) Interactions of polyelectrolytes with simple electrolytes. II. Donnan equilibria obtained with DNA in solutions of 1-1 electrolytes, *J. Phys. Chem.* 71, 2550–2556.
- Gross, L. M., and Strauss, U. P. (1966) Interactions of polyelectrolytes with simple electrolytes. I. Theory of electrostatic potential

- and Donnan equilibrium for a cylindrical rod model: the effect of site-binding, in *Chemical Physics of Ionic Solutions* (Conway, B. E., and Barradas, R. G., Eds.) pp 361–389, John Wiley and Sons, New York.
34. Johnson, M. L., and Frasier, S. G. (1985) Nonlinear least-squares analysis, *Methods Enzymol.* **117**, 301–342.
  35. Privalov, P. L., Ptitsyn, O. B., and Birshtein, T. M. (1969) Determination of stability of the DNA double helix in an aqueous medium, *Biopolymers* **8**, 559–571.
  36. Shkel, I. A., Tsodikov, O. V., and Record, M. T., Jr. (2000) Complete asymptotic solution of cylindrical and spherical Poisson–Boltzmann equations at experimental salt concentrations, *J. Phys. Chem. B* **104**, 5161–5170.
  37. Sawaya, M. R., Prasad, R., Wilson, S. H., Kraut, J., and Pelletier, H. (1997) Crystal structures of human DNA polymerase  $\beta$  complexed with gapped and nicked DNA: Evidence for an induced fit mechanism, *Biochemistry* **36**, 11205–11215.
  38. Bailey S. (1994) The CCP4 Suite—programs for protein crystallography, *Acta Crystallogr. D* **50**, 760–763.
  39. Holbrook, J. A., Capp, M. W., Saecker, R. M., and Record, M. T., Jr. (1999) Enthalpy and heat capacity changes for formation of an oligomeric DNA duplex: Interpretation in terms of coupled processes of formation and association of single-stranded helices, *Biochemistry* **38**, 8409–8422.
  40. Garcia-Garcia, C., and Draper, D. E. (2003) Electrostatic interactions in a peptide-RNA complex, *J. Mol. Biol.* **331**, 75–88.
  41. Kalodimos, C. G., Biris, N., Bonvin, A. M. J. J., Levandovski, M. M., Guennegues, M., Boelens, R., Kaptein, R. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes, *Science* **305**, 386–389.
  42. McGhee, J. D., and von Hippel, P. H. (1974) Theoretical aspects of DNA-protein interactions: Co-operative and non-cooperative binding of large ligands to a one-dimensional homogeneous lattice, *J. Mol. Biol.* **86**, 469–489.
  43. Shkel, I. A., Tsodikov, O. V., and Record, M. T., Jr. (2002) Asymptotic solution of the cylindrical nonlinear Poisson–Boltzmann equation at low salt concentration: Analytic expressions for surface potential and preferential interaction coefficient, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2597–2602.
  44. 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.

BI0520434