

Published in final edited form as:

Biochemistry. 2010 March 9; 49(9): 2018–2030. doi:10.1021/bi902027z.

# Contributions of the Histidine Side Chain and the N-terminal $\alpha$ -Amino Group to the Binding Thermodynamics of Oligopeptides to Nucleic Acids as a Function of pH

Jeff D. Ballin $^{\ddagger,*}$ , James P. Prevas $^{\ddagger}$ , Christina R. Ross $^{\ddagger}$ , Eric A. Toth $^{\ddagger}$ , Gerald M. Wilson $^{\ddagger}$ , and M. Thomas Record Jr. $^{\S,\dagger}$ 

<sup>‡</sup>Department of Biochemistry and Molecular Biology and Marlene and Stewart Greenebaum Cancer Center, University of Maryland School of Medicine, Baltimore, Maryland 21201

§Departments of Chemistry and Biochemistry, University of Wisconsin-Madison, Madison, WI 53706

## Abstract

Interactions of histidine with nucleic acid phosphates and histidine pK<sub>a</sub> shifts make important contributions to many protein-nucleic acid binding processes. To characterize these phenomena in simplified systems, we quantified binding of a histidine-containing model peptide HWKK (+NH<sub>3</sub>-His-Trp-Lys-Lys-NH<sub>2</sub>) and its lysine analog KWKK (+NH<sub>3</sub>-Lys-Trp-Lys-Lys-NH<sub>2</sub>) to a singlestranded RNA model, polyuridylate (polyU), by changes in tryptophan fluorescence as a function of salt concentration and pH. For both HWKK and KWKK, equilibrium binding constants, Kobs, and magnitudes of log-log salt derivatives  $SK_{obs} \equiv (\partial logK_{obs}/\partial log[Na^+])$ , decreased with increasing pH in the manner expected for a titration curve model in which deprotonation of the histidine and  $\alpha$ amino groups weakens binding and reduces its salt-dependence. Fully protonated HWKK and KWKK exhibit the same  $K_{obs}$  and  $SK_{obs}$  within uncertainty, and these  $SK_{obs}$  values are consistent with limiting-law polyelectrolyte theory for +4 cationic oligopeptides binding to single-stranded nucleic acids. The pH-dependence of HWKK binding to polyU provides no evidence for pK<sub>a</sub> shifts nor any requirement for histidine protonation, in stark contrast to the thermodynamics of coupled protonation often seen for these cationic residues in the context of native protein structure where histidine protonation satisfies specific interactions (e.g., salt-bridge formation) within highly complementary binding interfaces. The absence of pKa shifts in our studies indicates that additional Coulombic interactions across the nonspecific-binding interface between RNA and protonated histidine or the  $\alpha$ -amino group are not sufficient to promote proton uptake for these oligopeptides. We present our findings in the context of hydration models for specific versus nonspecific nucleic acid binding.

Nucleic acid binding proteins typically possess a higher density of positively charged amino acids in the DNA or RNA binding interface relative to elsewhere on the protein surface (1,2). In a comparison of 75 protein-nucleic acid crystal structures, DNA phosphates participated in 60% of all protein-DNA hydrogen bonds (including salt bridges), of which 41% were salt bridges with lysine or arginine (1). A recent survey of 45 protein-RNA co-structures found that 68% of hydrogen bond interactions to the phosphodiester backbone were through lysine and arginine residues, with donor-acceptor distances strongly clustered within 2.5 – 3.1 Å to make the most favorable (*i.e.*, dehydrated) hydrogen-bonds (3). The contributions of lysine (4-11)

<sup>\*</sup>Correspondence should be addressed to this author: Jeff D. Ballin, Department of Biochemistry and Molecular Biology, University of Maryland, 108 North Greene St., Baltimore MD 21201, U.S.A., Phone: (410) 706-8903, Fax: (410) 706-8297, jball003@umaryland.edu †This work was supported by funding from National Institutes of Health grants R01 CA102428 (to G.M.W) and partial support from the University of Wisconsin-Madison Graduate School.

and arginine (11-13) to peptide-nucleic acid binding have been extensively characterized and demonstrate that Coulombic interactions between cationic protein residues and the polyanionic nucleic acid backbone significantly stabilize the protein-nucleic acid complex and cause binding affinity to increase strongly with decreasing salt concentration.

Histidine can also be cationic and often makes important contacts in the protein-nucleic acid interface, many of which are essential or conserved across species. For example, direct phosphate-histidine interactions are observed for *E. coli* CAP *via* the minor groove adjacent to the site of DNA bending (14,15), in the NF-κB-DNA complex at the −1 position of the DNA recognition sequence (16) and for T7 DNA polymerase with the primer phosphate backbone (17). Figure 1 shows the conserved histidine-phosphodiester interactions in the co-crystal structures of rat glucocorticoid receptor (GR) with DNA (18) and of Vts1 (a homologue of Smaug in animals) with RNA (19). Mutational studies disrupting histidine-phosphate interactions impact binding and catalytic activity for T7 polymerase with DNA (20) and ribonuclease A with RNA (21). Overall, histidine is the seventh most prevalent natural amino acid at the protein-DNA interface (1). In amino acid distribution surveys of protein-DNA (1) and protein-RNA co-structures (22), histidine exhibits respectively the sixth and fifth highest ratio of binding interface *versus* surface distributions. Analysis of 129 protein-DNA structures revealed that 65% of direct histidine-DNA hydrogen bonds were to the phosphate backbone (23).

To our knowledge, no binding studies have been reported on histidine-nucleic acid model systems. Thermodynamic effects of histidine and other cationic residues on the thermal denaturation of DNA-oligopeptide conjugates have been reported (24), but data of this type only detect differences in interactions of the cationic residues with the native and denatured states of the nucleic acids. The frequency with which histidine is found in protein-nucleic acid interfaces suggests that histidine plays an important role in these binding processes. How does the behavior of histidine in a nonspecific ligand-nucleic acid complex compare to that seen for specific binding interactions such as those described above? What do these similarities and differences tell us? This study compares the nonspecific binding contributions of histidine versus lysine in model systems to address these questions.

The current work monitors tryptophan fluorescence as a function of pH and salt concentration to quantify the nonspecific binding of cationic peptides to single-stranded RNA in response to protonation changes of histidine and the  $\alpha$ -amino group. Since histidine and the  $\alpha$ -amino group are the only two basic amino acid residues which titrate near physiological pH, they are also the most likely to show any potential linkage effects between protonation and nucleic acid binding sometimes observed in native protein-nucleic acid binding processes (25-29). We examine whether protonation of histidine and the  $\alpha$ -amino group is coupled to or independent of nonspecific RNA binding and discuss coupled protonation events within the protein-nucleic acid binding interface in the context of recent hydration studies.

## **BACKGROUND**

# McGhee-von Hippel Analysis of the Oligopeptide-DNA Binding Isotherm

The nonspecific, primarily Coulombic interactions between polyanionic nucleic acids and a variety of oligocations and proteins, including polyamines and oligopeptides similar to those studied here (4-6,8,30-36) are well described by the non-cooperative McGhee-von Hippel (30) binding isotherm,

$$K_{\text{obs}} = \frac{v}{L_F (1 - nv)} \left( \frac{1 - (n - 1)v}{1 - nv} \right)^{n - 1}.$$
 (1)

This form of the McGhee-von Hippel equation assumes non-cooperative site binding of a linear ligand of arbitrary length to a linear, homogeneous infinite lattice with overlapping potential binding sites. In this case, the lattice is polyU and the ligand is the XWKK oligopeptide.  $L_F$  is the free ligand concentration. The site size (n) represents the number of adjacent phosphates (i.e.), lattice residues) occluded by the binding of a single oligopeptide to the RNA. The ligand binding density (v) represents the fraction of a ligand bound per lattice residue.

The McGhee-von Hippel isotherm and its finite lattice analog (37) have been extensively used to interpret fluorescence data (local accumulation of ligand) and equilibrium dialysis (global accumulation of ligand) for oligocation-nucleic acid interactions. Specific examples include fluorescence studies with oligolysines (4,6,34,35) and oligoarginines (13), and equilibrium dialysis studies with oligolysines (9,10), polyamines,  $Mg^{2+}$  (32), and  $Co(NH_3)_6^{3+}$  (31). Monte Carlo and Poisson-Boltzmann calculations modeling the nonspecific Coulombic accumulation of a divalent cation ( $Mg^{2+}$ ) in the vicinity of polynucleotides over a range of univalent salt concentration from 10-100 mM are well described by a McGhee-von Hippel isotherm with a site size between 2-3 (38). Although the McGhee-von Hippel isotherm was derived for site binding of a ligand and does not explicitly consider Coulombic or polyelectrolyte effects, its application to oligocation-polyanion data yield physically reasonable site sizes and binding constants. A comparison of the McGhee-von Hippel isotherm with alternative polyelectrolyte-based expressions for an oligocation-polyanion binding isotherm (39-41) is provided by Ni *et al.* (38).

Lohman and coworkers developed a quantitative fluorescence titration protocol that has been used extensively to obtain thermodynamic binding parameters of protein-nucleic acid interactions and is readily applied to oligopeptide-DNA/RNA binding (4-6,13,34,42,43). For ligands including E. coli SSB protein and various small oligopeptides, Lohman and coworkers found that the fractional fluorescence quenching ( $Q_{obs}/Q_{max}$ ), where  $Q_{max}$  is defined as the maximum quenching of the oligopeptide obtained at saturation, was equal to the fraction of ligand bound ( $L_B/L_T$ ),

$$\frac{Q_{obs}}{Q_{\text{max}}} = \frac{L_B}{L_T}.$$
 (2)

From eq.2, v and  $L_F$  of the McGhee-von Hippel equation (eq.1) can be expressed in terms of experimentally accessible variables,

$$v = \frac{L_B}{R_T} = \left(\frac{Q_{obs}}{Q_{max}}\right) \left(\frac{L_T}{R_T}\right),\tag{3}$$

$$L_F = \left(1 - \frac{Q_{obs}}{Q_{max}}\right)(L_T). \tag{4}$$

 $L_T$  and  $R_T$  are the total concentrations of ligand and RNA phosphate, and  $L_B$  is the concentration of bound ligand (4,6).

#### Effects of 1:1 Salt Concentration on Kobs

Binding interactions of nucleic acids with charged ligands are highly sensitive to salt concentration. At low to moderate salt concentrations ([NaCl] < 0.5 M), DNA and RNA binding

constants of oligocations and proteins typically exhibit a power-law dependence on the activity (concentration) of univalent salt which can be expressed as the log-log derivative

$$SK_{obs} \equiv \left(\frac{\partial log K_{obs}}{\partial log [Na^+]}\right) \cong \Delta\Gamma_+ + \Delta\Gamma_-,$$
 (5),

where the  $\Delta\Gamma$  terms are differences in preferential interaction coefficients for the interaction of the Na<sup>+</sup> cation ( $\Delta\Gamma_+$ ) and the Cl<sup>-</sup> anion ( $\Delta\Gamma_-$ ) with the complex vs. the uncomplexed reactants in the oligocation-nucleic acid binding process.  $SK_{obs}$  quantifies the thermodynamic consequences of the differences in preferential interactions of the monovalent salt ions (i.e., Na<sup>+</sup> and Cl<sup>-</sup>) with the complex and with the reactants (36,44,45). Below 0.25 M salt, the nature of the monovalent salt ions do not significantly impact  $SK_{obs}$  for  $Z_L < 10$  ligands binding to polyU (4,6), allowing comparisons of  $SK_{obs}$  across experimental systems even when obtained using different monovalent salts.

For the binding of a homologous series of cationic oligopeptides and polyamines to polyanionic DNA or RNA,  $logK_{obs}$  varies linearly as a function of  $[Na^+]$ 

$$\log K_{\text{obs}} = \log K_0 + (SK_{\text{obs}}) \log [Na^+], \tag{6}$$

where  $K_0$  is the extrapolated value of  $K_{\rm obs}$  at 1 M salt (4,8,32). Experimentally,  $SK_{\rm obs}$  is independent of  $[Na^+]$  and proportional to  $Z_L$ , the charge on the binding surface of the oligopeptide. The large magnitude of  $SK_{\rm obs}$  and the independence of  $SK_{\rm obs}$  versus salt concentration result primarily from the polyelectrolyte character of nucleic acids (35,46). Record *et al.* (8) predicted a thermodynamic limiting law (low [salt]) expression for  $SK_{\rm obs}$  for oligocation-polyanion binding

$$SK_{obs} \cong -Z_L \psi,$$
 (7)

where  $Z_L$  is the valence of the oligocation ligand and  $\psi$  represents the net thermodynamic extent of salt ion accumulation per phosphate of uncomplexed DNA (7,8,47,48). Experimentally,  $\psi \approx 0.88$  for double stranded DNA (47) and  $\psi \approx 0.74$  for single stranded nucleic acids (4,6). Equation 7 is a special case of the general expression for SK<sub>obs</sub> in eq. 5. Although derived as a limiting law, equation 7 is found to describe oligocation-polyanion binding data at higher salt concentration, possibly because of compensating salt concentration dependences of preferential interaction coefficients for the ligand and macromolecule (36,49).

Cylindrical Poisson-Boltzmann theory (50) and counterion condensation theory (8,51) predict that at low salt (limiting law) conditions,

$$\psi = 1 - \frac{1}{2\xi},\tag{8}$$

where the reduced axial charge density,  $\xi$ , is

$$\xi = \frac{e^2}{\varepsilon k T b}.\tag{9}$$

Here, e is the electronic charge,  $\epsilon$  is the bulk dielectric constant of water, k is the Boltzmann constant, T is the absolute temperature, and b is the axial charge spacing of the polyelectrolyte. In water at 25 °C,  $\xi = 7.14/b$ , where b is in units of Angstroms (Å). We use equations 6 - 9 as the basis for analyzing the effects of pH on  $K_{obs}$  and  $SK_{obs}$ , as described below.

# The Effects of pH on logKobs and SKobs

1) Titration curve model—The effects of pH on the interactions of pentalysine with double stranded DNA were shown to fit a "titration" curve model (7) where at fixed salt concentration, the dependence of  $K_{obs}$  on pH is determined by the pH-dependence of free ligand valence  $(Z_L)$ . In this model,  $Z_L$  does not increase upon binding to nucleic acids, *i.e.*, the pK<sub>a</sub> of the titratable groups on the oligopeptide are not affected by binding to DNA. For the simple case of independent titratable functional groups on the oligopeptide, the dependence of  $Z_L$  on  $[H^+]$  is given by

$$Z_{L} = Z_{L}^{\text{max}} - \sum_{i=1}^{g} \frac{1}{1 + k_{i,L} [H^{+}]},$$
(10)

where  $Z_L$  is the valence at a given pH,  $Z_L^{max}$  is the maximum valence at low pH ( $Z_L^{max}$ =4 in the present study),  $k_{i,L}$  is the equilibrium protonation constant of the i-th titratable group on the oligopeptide, and g is the number of groups titrating in the pH range of interest (6,7,47). Assuming that the two lysines at the C-terminal end of XWKK are always protonated at the pH and salt concentrations studied here,  $Z_L$  can be expressed as

$$Z_{L} = Z_{L}^{\text{max}} - \frac{1}{1 + k_{NH_{3}^{+}} [H^{+}]} - \frac{1}{1 + k_{X} [H^{+}]}, \tag{11}$$

where  $k_x$  is the protonation constant for the side chain of residue X (lysine or histidine) in

XWKK and  $k_{NH_3^+}$  is the protonation constant for the N-terminal  $\alpha$ -amino group. The uridylate moieties in polyU are also expected to deprotonate at sufficiently high pH, which in turn will reduce the axial charge spacing b. Taking the assumption of Mascotti and Lohman (6), we model the pH-dependence of b as

$$\frac{b_0}{b} = 1 + \frac{1}{1 + k_U [H^+]},\tag{12}$$

where  $b_0$  is the average charge spacing of polyU at neutral pH and  $k_U$  is the protonation constant of uridylate.

Provided that any salt concentration dependences of  $k_{NH_3^+}$ ,  $k_x$ , and  $k_U$  are negligible and that eq.7 for  $SK_{obs}$  is valid in the range of salt concentration of interest, the dependence of  $logK_{obs}$  on pH from eq. 6 - 9 is

$$\log K_{\text{obs}} - \log K_0 = -\psi \left( Z_L^{\text{max}} - \frac{1}{1 + k_{NH_3^+} [H^+]} - \frac{1}{1 + k_x [H^+]} \right) \log [Na^+],$$
(13)

where  $\psi$  is

$$\psi = 1 - \frac{b_0}{14.28} \left( 1 + \frac{1}{1 + k_U [H^+]} \right) \tag{14}$$

and K<sub>0</sub> is the extrapolated binding constant in the 1 M Na<sup>+</sup> reference state. From eq. 7 and 11,

$$SK_{obs} = -\psi \left( Z_L^{max} - \frac{1}{1 + k_{NH_3^+}[H^+]} - \frac{1}{1 + k_X[H^+]} \right)$$
(15)

with  $\psi$  as defined in eq. 14. In this model, any protonation state of the unbound oligopeptide can bind to RNA, with  $K_{obs}$  and the magnitude of  $SK_{obs}$  increasing with increasing protonation (i.e., greater  $Z_L$ ). At sufficiently low pH where all of the sites are protonated (i.e.,  $Z_L = Z_L^{max}$ ), both  $K_{obs}$  and  $SK_{obs}$  will become insensitive to pH.

**2)** Coupled Protonation Model—As an extreme alternative model, if protonation of the peptide were driven by complex formation so that the bound state of the peptide was the fully protonated state, then

$$SK_{obs} = -\psi(Z_L^{max}), \tag{16}$$

predicting that  $SK_{obs}$  would be to be independent of pH. However,  $logK_{obs}$  should still decrease with an increase in pH because of the increasing thermodynamic cost of protonating groups on the oligopeptide as the pH increases. This effect should give rise to a pH-dependence of  $K_0$  and hence of  $K_{obs}$ :

$$K_0 = K_0^0 \sum_{i=1}^g \frac{k_{i,L} [H^+]}{1 + k_{i,L} [H^+]},$$
(17)

where in the pH range of our experiments (5.2-8.7), g=1 or 2 for KWKK and g=2 for HWKK.  $K_0^0$  is the 1 M Na<sup>+</sup> extrapolated binding constant in the limiting case when all titratable sites on the ligand are protonated. As the pH increases,  $K_0$  decreases, thus reducing  $K_{obs}$ . Although the pH dependence of  $K_{obs}$  cannot be used to distinguish between the coupled protonation model and the titration curve model within an experimentally accessible range, the pH dependence of  $SK_{obs}$  can be so used. The titration model predicts that  $SK_{obs}$  should decrease in magnitude with increasing pH. In contrast, if protonation is coupled to binding,  $SK_{obs}$  is expected to be independent of pH.

# **EXPERIMENTAL PROCEDURES**

#### **Buffers and Reagents**

Reagents used were reagent grade, purchased from either Sigma Chemical Company (St. Louis, MO) or Fisher Scientific (Pittsburgh, PA). All solutions were prepared with 18 Mohm/cm deionized water. All buffers contained 0.2 mM Na<sub>2</sub>EDTA and were titrated to the indicated pH with concentrated HCl. The pH 5.2 buffer was 3 mM sodium acetate, the pH 6 buffer was sodium cacodylate, pH 7-8 buffers were 3 mM sodium HEPES, and the pH 8.7 buffer was 2.5 mM sodium borate. "High salt" and "low salt" solutions were prepared with 0.2 mM Na<sub>2</sub>EDTA and the specified buffer, with and without 1 M NaCl, respectively. Intermediate salt

concentrations at a given pH were obtained using linear combinations of the low salt and high salt buffers.

#### **Peptides**

Oligopeptides of the general form XWKK, where the  $\alpha$ -amino terminus is positively charged, the C-terminus is capped as an amide, and "X" is lysine or histidine, were synthesized, purified, lyophilized, and validated by matrix-assisted laser desorption ionization mass spectroscopy by the University of Maryland Biopolymer Core Facility. Stock peptide concentrations were determined spectrophotometrically from the 280 nm absorbance of the tryptophan in 6 M guanidinium chloride using  $\varepsilon_{280}^{\rm Trp} = 5690~{\rm M}^{-1}~{\rm cm}^{-1}$  (52). The extinction coefficients of the oligopeptides dissolved in the buffers described above (3 – 5 mM Na<sup>+</sup>) were then experimentally measured for use in determining the inner-filter corrections (see above) for the fluorescence titrations performed ( $\varepsilon_{292}^{\rm HWKK} = 3380 \pm 50~{\rm M}^{-1}~{\rm cm}^{-1}$ ),  $\varepsilon_{292}^{\rm KWKK} = 3140 \pm 110~{\rm M}^{-1}~{\rm cm}^{-1}$ ,  $\varepsilon_{350}^{\rm XWKK} < 5~{\rm M}^{-1}~{\rm cm}^{-1}$ ). The resulting inner-filter correction is small relative to the corresponding correction due to the RNA concentration (see below).

#### polyU single-stranded RNA

The potassium salt of polyuridylic acid (polyU) was manufactured (lot # 011805) by the Midland Certified Reagent Company (Midland, TX). After synthesis, phenol/chloroform/ isoamyl alcohol extraction and extensive dialysis against potassium chloride, the polyU was exhaustively dialyzed against deionized water to remove excess salt and then lyophilized by the manufacturer. Analytical polyacrylamide gel electrophoretic analysis indicated that the polyU substrates ranged between 40-200 bases in length. Studies of the transition between oligomeric *versus* polymeric binding for oligopeptides associating with single stranded nucleic acids found that a ligand with  $Z_L \le 4$  binds the central site of a nucleic acid with 22 or more charges with the same affinity as the central site of a polymeric nucleic acid (53,54). As such, XWKK is predicted to bind this distribution of polyU lengths equivalently in the low binding density limit reported by the McGhee-von Hippel isotherm (23). Extinction coefficients of the RNA were determined for use with the inner-filter corrections ( $\varepsilon_{292}^{RNA} = 160 \pm 40 \quad M^{-1} \quad cm^{-1}$ ,  $\varepsilon_{350}^{RNA} \sim 0 \quad M^{-1} \quad cm^{-1}$ ). The inner-filter correction (42) due to the RNA concentration in the cuvette can result in as much as a 17% change in fluorescence intensity.

## **Bioinformatics**

Sequence alignments were prepared using the CLUSTALW Version 3.2 software package available *via* the World Wide Web at the San Diego Supercomputing Center Workbench (http://www.workbench.sdsc.edu).

#### Fluorescence Quenching Studies of Oligopeptide Binding

Oligopeptide binding to polyU was monitored by tryptophan fluorescence quenching with a Cary Eclipse spectrofluorometer (Varian Instruments) equipped with a Peltier temperature controller maintaining solution at 25 °C. The excitation wavelength was 292 nm with a bandpass of 2.5 nm, and the emission wavelength was 350 nm with a bandpass of 10 nm. These wavelengths were chosen to minimize the inner-filter corrections and RNA absorbance (42). Solutions of the oligopeptides and RNA used in the fluorescent titrations were prepared from freezer stocks and diluted to identical buffer and salt concentration conditions. All titrations were performed in 1×1 cm square quartz cuvettes to allow for "crown" stir-bar mixing during titration and measurement. After addition of titrant (either with polyU during the "reverse titration" or with "high salt" buffer during "saltbacks") to the oligopeptide solution, samples

were incubated with stirring for at least one minute before measurement. Photobleaching of tryptophan in XWKK was minimized by illuminating the sample only during measurement. For comparison, unbound KWKK and HWKK exhibited less than 3% loss in fluorescence after an hour of continuous irradiation (data not shown). To correct for Raman light scattering from the water and any background fluorescent emission, the fluorescence of a buffer solution (no oligopeptide present) that was at the equivalent RNA and salt concentration in the titration was subtracted from the fluorescent signal of the sample. The extent of tryptophan fluorescence quenching is thus defined as

$$Q_{obs} \equiv \frac{F_0 - F_{obs}}{F_0},\tag{18}$$

where  $F_{obs}$  is the observed fluorescence intensity at a given point in the titration and  $F_0$  is the initial fluorescence intensity of the free ligand (*i.e.*, oligopeptide). All fluorescence intensities were corrected for background fluorescence, dilution, inner-filter contributions, and photobleaching effects as described previously (35,42).

#### **Reverse Titrations**

Isotherms of either HWKK or KWKK binding to polyU were generated by "reverse titration": 1-10~mM stock solutions of RNA were titrated into cuvettes containing  $1-8~\mu\text{M}$  of XWKK. Titrations continued until binding saturation was observed or a maximum concentration of 0.3 mM RNA phosphate was achieved, a limit dictated by the approximations used for inner-filter correction. Values of  $Q_{obs}$  (eq. 18) as a function of XWKK and polyU concentrations were then used to calculate binding affinity (see below).

HWKK deprotonates to a greater extent relative to KWKK within the range of pH considered, resulting in decreased binding affinity at moderate and high pH values. To maintain oligopeptide-RNA binding affinity within measurable limits yet allow for direct comparison of binding affinities across the full pH range, titrations were performed at the common salt concentrations of 24.4 mM Na<sup>+</sup> for KWKK and 9.8 mM Na<sup>+</sup> for HWKK (Table 1).

#### **Salt Back Titrations**

Salt concentration dependences of the oligopeptide-RNA binding equilibrium were determined in part via "saltback" titrations, where the pre-equilibrated XWKK-RNA complex was titrated with the "high salt" buffer with the appropriate pH (see *Buffers and Reagents* section above), monitoring the increase in fluorescence and therefore the decrease in quenching  $Q_{obs}$ , as the salt concentration of the solution increases. After all corrections were applied (*i.e.*, background fluorescence, dilution, inner-filter, photobleaching), more than 90% of the original fluorescence of the free oligopeptide is recovered, indicative of the reversibility of oligopeptide-RNA complex formation.  $SK_{obs}$  and  $logK_0$  values determined by saltback titrations were supported by at least three independent reverse titration determinations of  $K_{obs}$  at various salt concentrations, with each  $K_{obs}$  derived from at least three isotherms (see below and Figure 5).

#### Analysis of Oligopeptide-RNA Binding Isotherms

The binding constant  $K_{obs}$  was determined from reverse titration data by fitting to the non-cooperative McGhee-von Hippel (30) binding isotherm (eq. 1). The parameters v and  $L_F$  were determined via eq. 2-4.  $Q_{max}$  was calculated by global analysis of 3-7 isotherms for XWKK concentrations ranging between  $1-8~\mu M$  using the model-independent method of ligand binding density (LBD) analysis (42). The nonlinear least squares program NONLIN (55) was used to fit log $K_{obs}$  and n to eq.1 - 5 using a fixed  $Q_{max}$  obtained by LBD analysis.  $Q_{max}$ 

converged to the LBD-determined value when allowed to float for titrations where  $K_{\rm obs} \gtrsim 10^5~{\rm M}^{-1}$  and in global analysis of multiple reverse titrations across 4-6 salt concentrations (data not shown). The site size parameter, n, was typically not integral when allowed to float during nonlinear fitting. Any Coulombic or other source of negative or positive cooperativity that may exist has effectively been absorbed by n and these factors may be responsible for the nonintegral n. Nevertheless, no dependence of n on the values of v or  $L_F$  has been found in this or previous studies. Site size was indistinguishable from or near n=5 in global analysis of reverse titration data obtained at a common pH (Table SI). The  $n\approx 5$  site size was also determined using a model-independent approach (see *Results*). Unless otherwise indicated, binding constants and their salt dependence are reported using a site size fixed at n=5.

 $\log K_{obs}$  as a function of salt concentration was directly calculated for saltback data via eq. 1 using a constant  $Q_{obs}$  determined by LBD analysis and site size n=5. To facilitate comparison of HWKK versus KWKK binding to polyU in Figure 6,  $\log K_{obs}$  data for HWKK at 24.4 mM Na<sup>+</sup> was linearly interpolated via eq.6 and the SK<sub>obs</sub> and  $\log K_0$  parameters listed in Table 1. All  $K_{obs}$  data (Figure 5) were globally fit to eq.13 as a function of salt concentration and pH, with parameters listed in Table 2.

## Statistical comparison of fitting parameters and thermodynamic models

Several times in this study, we assess whether adding complexity to our models dramatically improves the description of the data. The F-test is a common means to determine the statistical significance of adding terms to nested models (*i.e.*, a series of models where fixing parameters to some value such as 0 or 1 generates a simpler expression from a more complex one). However, both the McGhee-von Hippel isotherm (eq.1) and the titration binding model (eq.13 - 15) are nonlinear equations where terms cannot be eliminated by setting a parameter to a constant value. The second-order Akaike Information Criterion (AIC<sub>c</sub>) (56,57) is a well accepted alternative to the F-test which does not have the nested-model restriction:

$$AIC_{c} = Nln\left(\frac{SS_{res}}{N}\right) + 2(p+1) + \frac{2(p+1)(p+2)}{N-p-2}$$
(19),

In eq.19, N is the number of data points,  $SS_{res}$  is the residual sum of squares of the fit, and p is the number of parameters in the model. The  $AIC_c$  is calculated for each model considered. For two models "A" versus "B", the better model has a smaller (or more negative)  $AIC_c$ . When  $AIC_c$  scores are close in value, the probability that one model is favored over the other is defined by

$$probability = \frac{e^{-\Delta AIC_c}}{1 + e^{-\Delta AIC_c}}$$
 (20),

where  $\Delta AIC_c = AIC_{c,B} - AIC_{c,A}$ . The evidence ratio (also called the relative likelihood) is the ratio of the probability (eq. 20) favoring model "A" *versus* the probability favoring "B." The evidence ratio can be calculated directly by  $e^{0.5\Delta AIC_c}$  since the  $\Delta AIC_c$  values for these two cases are related by a sign inversion. As an example, if "A" is favored over "B" with a  $\Delta AIC_c = 6.0$ , model "A" has a ~95% probability of being a better description of the data in comparison to "B" (with a ~5% chance). The evidence ratio of A/B states that "A" is 20-fold more likely than "B" to be the correct model (56). Model comparisons are reported using  $\Delta AIC_c$  or evidence ratios if  $\Delta AIC$  is small. We arbitrarily define that a model is convincingly preferred when  $\Delta AIC_c$  has a magnitude of 6.0 or greater (evidence ratio  $\geq$  20).

# **RESULTS**

The work presented here compares polyU binding by HWKK, which has not been studied before, with the previously characterized KWKK (4-6,13) to determine how the change from lysine to histidine affects binding affinity  $K_{obs}$  (eq. 1), its salt dependence  $SK_{obs}$ , and site size over a range of pH in which histidine and the  $\alpha$ -amino group deprotonate. Titrations under stoichiometric binding conditions were used to determine a model-independent site size of XWKK on polyU (Figure 2). The titrations were performed at low salt (3.4 mM Na<sup>+</sup>) and low pH (pH 5.2) to maximize binding affinity (log $K_{obs} \ge 7.3$ ). The site size is indicated by the intersection of the polyU concentration-dependent regime with the polyU concentration-independent plateau. The invariance of the intersection point across three titrations over a two-fold concentration range suggests that  $n \approx 5$  is a true site size for KWKK and HWKK binding to polyU at pH 5.2.

Figure 3 shows a typical series of reverse titration binding isotherms. As the initial oligopeptide concentration is increased, increasing amounts of polyU are required to achieve an equivalent level of binding saturation. The HWKK isotherms are well described by the noncooperative McGhee-von Hippel model, as has been shown for KWKK (4-6,13).  $Mg^{2+}$  (8) and model peptides (34,46) bind to nucleic acids with McGhee-von Hippel site sizes approximately equal to the net charge of the ligand. However, constraining the site size to n=4, the number of ligand charges neutralized in the XWKK-polyU complex, results in a statistically poorer fit relative to n=5 (Figure 3, dashed *versus* solid lines; Table SI). Site size converges to  $n\approx 5$  when allowed to float as a parameter (Table SI; Figure 3, solid lines), consistent with the model-independent determination of site size in Figure 2. At pH 6, increasing n from 4 to 5 increases  $K_{obs}$  35% and decreases  $\Delta G^o$  by about 3%; this site size effect on  $K_{obs}$  decreases at higher pH.

Calculation of binding affinity requires knowledge of how binding density v varies as a function of free oligopeptide concentration  $L_F$ .  $Q_{obs}$  can be directly related to v and  $L_F$  through  $Q_{max}$ (eq. 3 - 4). The upper asymptote, and thus  $Q_{max}$ , is easily defined when binding is saturable (e.g., at low pH and/or low [Na<sup>+</sup>]; cf. HWKK at pH 5.2 in Figure 4). However, with weaker binding, uncertainty in the estimation of  $Q_{max}$  from a single isotherm dramatically increases as the upper asymptote becomes more difficult to reach. Ligand binding density (LBD) analysis provides a model-independent means of determining  $Q_{max}$  and other thermodynamic parameters (42,58). Mascotti and Lohman applied LBD analysis to nucleic acid binding studies of a series of oligopeptides KWK<sub>x</sub> (x = 1 - 8), including KWKK. Figure S1 presents the results of LBD analysis of KWKK and HWKK, plotting  $Q_{obs}$  as a function of  $L_B/L_T$ . Because  $Q_{obs}$  is a linear function of the fraction of peptide bound  $(L_B/L_T)$ , eq. 3 shows that binding density vis also a linear function of  $Q_{obs}$ , and by extension,  $L_F$  can be calculated using eq. 4. Extrapolation of  $L_B/L_T$  to 1 allows calculation of  $Q_{max}$  without requiring XWKK binding saturation.  $Q_{max}$  was insensitive to salt concentration in the range considered (data not shown).  $Q_{max}$  was relatively constant over pH 5.2 – 7, decreasing at pH 8 for HWKK and pH 8.7 for KWKK (Table 1 and Figure S1). The pH dependence of XWKK  $Q_{max}$  is consistent with previous reports showing that  $Q_{max}$  was proportional to net oligopeptide charge for a series of oligolysines ( $Z_L \le 4$ ) binding to polyU (6), suggesting that tryptophan fluorescence quenching is modulated by net differences in adjacent charge, whether by deprotonation in the case of XWKK, or by the presence of additional lysines at the C-terminus, in the work of Mascotti and Lohman (6).

Using  $Q_{max}$  determined by LBD analysis and eq. 2 - 4 to calculate v and  $L_F$ , we assessed the pH sensitivity of XWKK-polyU binding for HWKK relative to KWKK (Figure 4). KWKK binding is only moderately affected at low pH, with marked decreases at pH 8 and 8.7. In contrast,  $\log K_{obs}$  for HWKK is a strong function of pH and is weaker than KWKK at neutral pH even at a relatively lower salt concentration which favors tighter binding (9.8 mM Na<sup>+</sup>

*versus* 24.4 mM Na<sup>+</sup>). With the exception of HWKK at pH 7.0 which has a fitted  $n = 4.36 \pm 0.36$ , the McGhee-von Hippel site size is invariant at n = 5.0 within error across the full 5.2 - 8.7 pH range for both HWKK and KWKK (Table SI).

The linear dependence of  $Q_{obs}$  on binding density and the apparent invariance of site size across salt concentration and pH permits simple calculation of K<sub>obs</sub> from eq. 1, 3, and 4 for a given  $Q_{obs}$ . In cases where NaCl titration of preformed XWKK-polyU complexes results in full recovery of the initial tryptophan fluorescence, "saltback" titrations provide a much more direct route to calculating the salt dependence of logK<sub>obs</sub> via eq. 6. Figure 5 presents the logK<sub>obs</sub> data as a function of salt concentration and pH for KWKK and HWKK. Black symbols represent reverse titration data measured and analyzed across 3 – 7 oligopeptide concentrations, while red symbols indicate logK<sub>obs</sub> values determined via the saltback methodology (42). The colinearity of logK<sub>obs</sub> vs. log[Na<sup>+</sup>] data observed by reverse titration versus saltback approaches suggests that saltback measurements accurately reflect the salt dependence of polyU-binding affinity for KWKK and HWKK. In light of this observation, the reverse titration and saltback data were analyzed as a whole. Black lines in Figure 5 represent linear least squares fits of logK<sub>obs</sub> vs. log[Na<sup>+</sup>] for each peptide and pH. Green lines indicate the predicted salt dependence of binding affinity using eq.13 and parameters determined by global nonlinear least-squares analysis of the combined  $log K_{obs}$  data from reverse titrations and saltbacks across the full pH range (Figure 5). The SK<sub>obs</sub> of HWKK-polyU shows a steep magnitude decrease as pH increases (Table 1), reflected in more shallow slopes in the salt dependence of logK<sub>obs</sub> in Figure 5B. KWKK in Figure 5A also exhibits a decreased SK<sub>obs</sub> magnitude at high pH, although to a lesser extent than HWKK. The small magnitude of logK<sub>0</sub> (the residual binding affinity at 1 M Na<sup>+</sup> where polyelectrolyte contributions are considered to be negligible) indicates that the polyelectrolyte effect accounts for the majority of the  $\Delta G^{0}$  of the binding interaction for XWKK (Table 1). This is consistent with previous oligocation-nucleic acid binding studies (4-8,11,32,35,46,53).

XWKK binding affinity K<sub>obs</sub> and its salt dependence SK<sub>obs</sub> are presented in Figure 6. To facilitate comparison to the KWKK logKobs data, HWKK logKobs values were interpolated to 24.4 mM Na<sup>+</sup> by eq. 6 using the SK<sub>obs</sub> and logK<sub>0</sub> parameters reported in Table 1. The same global analysis which generated the green logKobs vs log[Na+] fits in Figure 5 are presented as a function of pH in Figure 6. At low pH, K<sub>obs</sub> of HWKK approaches K<sub>obs</sub> of KWKK, indicating that the nucleic acid binding free energy contribution of protonated histidine is similar to that of lysine. The titration model (eq. 13 - 15) is overwhelmingly favored over the proton-uptake model (eq. 16 - 17) to describe the pH dependence of polyU-binding for KWKK (evidence ratio =  $2 \times 10^{19}$ ) and HWKK (evidence ratio =  $2 \times 10^{39}$ ). In other words, the titration model which accounts for changes in the oligopeptide  $Z_L$  charge (eq.10), coupled to  $SK_{obs}$ through  $\psi$  (eq. 7) and thus to  $\log K_{obs}$  by eq. 6, describes all of the observed binding phenomena for both HWKK and KWKK across changes in salt concentration and pH. The net thermodynamic ion accumulation around each polyU phosphate was equivalent for KWKK and HWKK with  $\psi$  KWKK = 0.77  $\pm$  0.01 and  $\psi$  HWKK = 0.76  $\pm$  0.03 (Table 2), consistent with literature values for KWKK ( $\psi = 0.78 \pm 0.05$ ) and its KWK<sub>x</sub> variants (composite  $\psi = 0.74 \pm 0.05$ ) 0.04) at pH 6 (4,6).

With increasing pH, HWKK-polyU binding exhibits two observable deprotonation events with pKa values of  $5.75 \pm 0.18$  and  $7.86 \pm 0.13$  for the histidine imidazole and the  $\alpha$ -amino group, respectively. The fitted  $\alpha$ -amino pKa for KWKK was  $7.57 \pm 0.08$  in agreement with previous results (6). Accounting for potential lysine  $\epsilon$ -NH3<sup>+</sup> deprotonation at pH > 8 by adding a second titration term to the fixed- $\psi$  model provides a minor improvement to the fit as judged by comparison of the second-order Akaike Information Criteria (AICc) with a only 78% probability favoring the two pKa fixed- $\psi$  model (evidence ratio = 3.5). Additional model complexity compensating for uracil deprotonation through eq.14 was not necessary for HWKK

(99% probability favoring the fixed- $\psi$  model; evidence ratio = 79.1) or the two-pK<sub>a</sub> model for KWKK (37% favoring the 2-pK<sub>a</sub> titrating- $\psi$  model with  $\Delta$ AIC<sub>c</sub>= 1.05; evidence ratio = 1.7). However, fitting KWKK to the single-pK<sub>a</sub> model was dramatically improved by using eq.14 in lieu of a fixed- $\psi$  model ( $\Delta$ AIC<sub>c</sub>= 38.4). Including the potential deprotonation of uracil or of lysine by increasing g=1 to g=2 in eq.10 contributes another KWKK-polyU titration event with a pK<sub>a</sub>  $\geq$  9.6 (Figure S2). Inspection of Figure S2 suggests that the KWKK preference for an additional alkaline pK<sub>a</sub> is dictated predominantly by the pH 8.7 data. Indeed, all other KWKK-polyU fitted parameters ( $\alpha$ -amino pK<sub>a</sub>, logK<sub>0</sub>, and  $\psi$  fitted directly or through eq.14 using the fitted  $b_0$ ) were equivalent within error across the four model variants. Taking all of these factors in consideration, we chose the fixed- $\psi$  mode using a single titratable group for KWKK and two titratable groups for HWKK to describe XWKK binding data in Figures 5 - 6 and Table 1.

# **DISCUSSION**

# Protonation of histidine and N-terminal $\alpha$ -amino residues on HWKK and KWKK is not required for binding to polyU

This study determines whether nonspecific nucleic acid-binding exhibits pK<sub>a</sub> shifts of titratable amino acid moieties in histidine- and lysine-containing peptides. Three observations indicate that both KWKK and HWKK bind to polyU without pK<sub>a</sub> shifts and coupled uptake of protons. First and foremost, the pK<sub>a</sub> values found here for histidine, lysine, and the α-amino group are consistent with pK<sub>a</sub> measurements for free oligopeptides. Second, the magnitude of SK<sub>obs</sub> decreases with increasing pH (Figures 5 - 6), indicating that the number of positive charges and therefore the protonation state of HWKK and KWKK in the complex with polyU decreases with increasing pH, consistent with the titration model (eq.15) and not the coupled protonation model (eq.16). Finally, logK<sub>0</sub>, which is small in magnitude and does not vary significantly with increasing pH, does not reflect the decrease in  $K_0$  predicted by the coupled protonation model (eq.17) to account for the enthalpic cost of stabilizing the peptide-polyU complex by proton uptake. Our demonstration that histidine and α-amino group protonation is intrinsically uncoupled from nonspecific nucleic acid binding indicates that the histidine pK<sub>a</sub> shifts observed for many specific protein-nucleic acid recognition processes require additional noncovalent interactions to stabilize the protonated state which are not available in generalized nonspecific binding interfaces.

# Histidine-polyU binding thermodynamics are well described by limiting law polyelectrolyte theory

Polyelectrolyte theory developed in the limit of low monovalent salt concentration and nearzero macromolecule binding density accurately describes XWKK-polyU binding between pH 5.2 - 8.7 and 5.4 - 200 mM Na<sup>+</sup>. The non-electrostatic binding constant K<sub>0</sub> at 1 M Na<sup>+</sup> was indistinguishable from zero for both HWKK and KWKK as expected for a Coulombic interaction entropically driven by salt ion release, consistent with previous model studies (4-6,32,36,44). The contributions of protonated histidine and protonated lysine to binding free energy are indistinguishable. From the perspective of the nucleic acid, the polyU per-phosphate ion accumulation is equivalent within error for  $Z_L = 4$  oligopeptides HWKK ( $\psi = 0.76 \pm 0.03$ ), KWKK ( $\psi = 0.77 \pm 0.01$ ) and the arginine-containing oligopeptide RWRR ( $\psi = 0.80 \pm 0.05$ ) (13). If  $\psi$  is expressed as a function of the polyU axial charge spacing through eq.14, values of  $b_0 = 3.2 \pm 0.2$  Å and  $3.3 \pm 0.3$  Å obtained from the salt dependence of binding for KWKK and HWKK respectively (Figure S2), are consistent with a previous report of  $b_0 = 3.2 \pm 0.6$  Å for single-stranded polyU and poly(rA) obtained by analysis of nucleic acid folding-unfolding processes (59). The reproducibility of  $\psi$  and  $b_0$  determined for differing processes (binding versus melting) and with ligands with varying amino acid composition (lysine, arginine, or histidine) indicate that  $\psi$  and  $b_0$  are thermodynamic parameters entirely dictated by the

polyelectrolyte character of polyU. With  $\log K_0 \approx 0$  and  $\psi \approx 0.78$ , HWKK-polyU binding and its dependence on salt concentration and pH are fully explained by changes in the protonation state of histidine and the N-terminal  $\alpha$ -amino group.

#### Protonation states of amino acid moieties are tunable

Statistically validated differences in the  $\alpha$ -amino pK<sub>a</sub> for HWKK versus KWKK indicate that replacing the neighboring lysine moiety with histidine promotes a pK<sub>a</sub> shift (Table 2). Fitting HWKK data with a pK<sub>a</sub> held constant using the pK<sub>a</sub> = 7.57 from KWKK is strongly disfavored ( $\Delta$ AIC<sub>c</sub> = 8.6; evidence ratio = 75.4), while the converse constraint of fitting KWKK data with the HWKK pK<sub>a</sub> = 7.81 results in an even poorer fit ( $\Delta$ AIC<sub>c</sub> = 20.0; evidence ratio = 1.3 × 10<sup>5</sup>). The decreased acidity of the  $\alpha$ -amino group in HWKK relative to KWKK likely reflects the reduced free energy cost of adjacent charge-charge repulsion due to prior histidine deprotonation.

Analogous to our findings that the α-amino group pKa depends on adjacent sequence in XWKK, numerous reports show that histidine protonation equilibria are sensitive to local environment changes (28,60-63). The highly tunable nature of histidine protonation in native proteins is exemplified by sperm whale myoglobin which possesses seven histidines with pK<sub>a</sub> values ranging between 5.13 – 8.08 (60). Histidine pK<sub>a</sub> variations for a series of model peptides based on sequences from sperm whale myoglobin were assessed at 0.02 M NaCl and 25 °C by NMR spectroscopy (60). Flanking lysines in Phe-Lys-His-Leu-Lys (FKHLK) acidify the adjacent histidine by  $\Delta pK_a = -0.52$  relative to Gly-His-Gly (GHG) (Hisp $K_a = 6.16 \pm 0.02$ versus  $6.68 \pm 0.02$ , respectively). Proline isomerization from cis to trans in Lys-Ser-His-Pro-Glu (KSHPE) suppresses histidine deprotonation with a  $\Delta pK_a = +0.46$  from  $6.23 \pm 0.02$  to 6.69 $\pm$  0.02, likely due to the difference in relative orientation of glutamate and histidine residues. These findings demonstrate that simple electrostatic differences in the vicinity of histidine are sufficient to modulate its protonation pH dependence in either direction. However, the histidines in FKHLK and KSHPE exhibit further pK<sub>a</sub> perturbations in the context of native protein structure. FKHLK and KSHPE are the respective oligopeptide models for His36 and His48 in sperm whale myoglobin (60). His36, only observed in the trans conformation in the wild-type protein, has a pK<sub>a</sub> = 7.67  $\pm$  0.02, which is  $\Delta$ pK<sub>a</sub> = +0.98 more basic than its corresponding model peptide, trans-KSHPE. Similarly, His48 is more acidic in sperm whale myoglobin (pK<sub>a</sub> =  $5.42 \pm 0.02$ ), reflecting a  $\Delta$ pK<sub>a</sub> = -0.74 relative to FKHLK. The potential to tune the histidine protonation equilibrium constant across three orders of magnitude suggests that the prevalence of histidine at nucleic acid-binding interfaces may serve as an exquisite means to regulate protein-nucleic acid processes by providing important noncovalent interactions through coupled protonation.

Many nucleic acid-binding proteins exhibit histidine protonation coupled to binding. The pK<sub>a</sub> of highly conserved His318 in the DNA-binding interface of human papillomavirus type 16 E2 protein (E2C) increases upon DNA binding (from 6.7 to 7.8;  $\Delta$ pK<sub>a</sub> = +1.1) in a mechanism proposed to be critical for the transition from nonspecific to specific DNA recognition (28). The pK<sub>a</sub> of His451 in rat glucocorticoid receptor increases from 5.9 to 7.9 after binding the phosphodiester backbone of its DNA recognition element (25). His203 of the origin-binding domain of SV40 T antigen (T-ag-obd) exhibits a pK<sub>a</sub> shift from ~5 to >8 when it binds specifically to the middle phosphate of the GAGGC pentanucleotide (27). His49 of the thermophile *Pyrococcus woesei* TATA-binding protein (*Pw*TBP) has a pK<sub>a</sub> of 6.2 in the unbound state but increases to pK<sub>a</sub> = 7.2 when *Pw*TBP complexes with its recognition sequence in which His49 interacts directly with a DNA phosphate (29). In each of the structural studies described above, proton uptake is coupled to specific binding to generate salt-bridges or other interactions between the protonated histidine and complementary surfaces within the binding interface (25,27-29). In cases where nonspecific binding was considered, histidine did not show

these interactions (18,27,28). For example, His451 which exhibits a  $\Delta pK_a = +2$  in the specific complex did not interact with a non-cognate element (18). The coupled protonation of His203 with specific binding of T-ag-obd to its recognition sequence was not observed for the binding a non-cognate sequence (27). In E2C, His322 and His326 reside in a flexible and fully solventexposed loop domain known to participate in long-range nonspecific electrostatic interactions with cognate DNA (64). However, the His322 and His326 pK<sub>a</sub> values (6.3 and 6.2, respectively) do not change upon specific binding, whereas the nearby His318 protonates during the same process to make a hypothesized water-mediated contact to a DNA phosphate (28). Together, the unbound proteins highlighted above (25,27-29) have comparable histidine p $K_a$  values (average p $K_a = 6.0 \pm 0.7$ ) to those seen for nonspecific binding histidines in E2C (28) and to the pK<sub>a</sub> =  $5.75 \pm 0.18$  for HWKK nonspecific binding to single-stranded RNA seen here. We note that all of these pK<sub>a</sub> values are nearly indistinguishable from that of free histidine  $(pK_a = 6.04 (65))$ . In addition to the proteins above which show coupled protonation upon specific binding, lac repressor (lacR) exhibits cooperative uptake of 2 protons during nonspecific binding between pH 7.7 – 8.4 (26), even though specific binding of lacR with lac operator DNA is well-described by the titration model for a single titratable group with no evidence of coupled protonation between pH 7.1 – 8.4 (66). deHaseth et al. did not identify which residues were protonated, but Kalidomos et al. observed that lacR His29 makes electrostatic contacts with a DNA phosphate in both specific and nonspecific complexes (67).

The striking  $pK_a$  shifts described above are likely independent of large changes in protein structure. For example, E2C does not significantly change its conformation upon specific binding of its recognition element (68). Several structural comparisons of protein-DNA complexes report that the distribution of protein residues interacting with DNA are overall quite similar for a given protein in specific versus nonspecific complexes, although the DNA targets of these interactions typically change (18,67,69-72). For ~80% of the 44 specific DNA-binding proteins considered in rigid-body docking simulations, the sites involved in specific DNA-binding (72). These combined findings suggest that protonation events coupled to specific binding are principally due to many small changes in the local environment rather than large domain rearrangements within the protein.

Structural and thermodynamic studies provide evidence that the interfaces in specifically bound protein-DNA complexes are highly dehydrated, with specific interactions between protein and nucleic acid functional groups replacing the interactions of these functional groups otherwise provided by water (1,3,73). Analyses (1,3) of structures of protein-nucleic acid complexes illustrate that a large fraction of the hydrogen bonds spanning the binding interfaces of cognate complexes have donor-acceptor distances that are too short to accommodate bridging waters (< 3.5 Å). For example, Janin and coworkers reported that 98% of protein-nucleic acid donoracceptor distances were shorter than 3.4 Å and 67% were shorter than 3.0 Å (3). Recently, the large stabilizing effect of glycine betaine on the equilibrium constant for specific binding of lac repressor to lac operator DNA has been quantitatively interpreted in terms of the loss of more than 1200 waters of hydration (and some glycine betaine) from almost 7000 Å<sup>2</sup> of repressor and operator surface buried in this interaction (74). (Burial of 630 Å<sup>2</sup> of DNA phosphate surface area, with release of approximately 170 waters of hydration from which glycine betaine was completely excluded in the uncomplexed DNA, was shown to contribute the majority of the stabilizing effect of glycine betaine.) More generally, effects of glycine betaine, urea and Hofmeister salts on protein-DNA interactions, on protein folding and other protein processes are explained in terms of the loss of approximately 2 layers of water from all surfaces buried in the interface (74-77). Conversely, interfaces of nonspecific complexes of oligocations like KWKK and HWKK with nucleic acids probably are fully hydrated, based on several lines of evidence. First, experimentally-determined Kobs and SKobs values for oligocation-nucleic acid binding processes were accurately predicted by Poisson-Boltzmann

calculations which only account for Coulombic interactions and not changes in surface hydration (78). Second, oligopeptide-polyU binding constants were essentially indistinguishable as a function of anion composition at constant salt concentration (4). If significant dehydration occurred in these interactions, specific effects of different Hofmeister anions on the binding constant would be expected (77). Finally, NMR studies (79,80) demonstrated that polyamines bound to DNA are highly mobile, which is consistent with a hydrated, Coulombic, "loose" polyamine-DNA complex. Therefore, XWKK-RNA complexes most likely have hydrated binding interfaces which are exclusively stabilized by long-range Coulombic interactions. Our studies show that an addition of 1 – 2 positive charges on XWKK to increase Coulombic interactions across the hydrated interface is not sufficiently favorable to promote protonation of histidine or the α-amino group at a pH above their respective typical pKa values. However, in the dehydrated interfaces of specifically bound complexes, residues interact directly instead of through water. Protein-nucleic acid binding processes exhibiting coupled histidine protonation require that histidine interact via the proton that makes it positively charged to satisfy the salt-bridges or other necessary hydrogen bonds with the nucleic acid. In these cases, the free energy penalty of protonating histidine above its nominal pK<sub>a</sub> is much less than the consequences of not fulfilling the interactions it provides. These findings provide a rationale explaining why the presence of histidine coupled protonation exhibited by some nucleic acid-binding proteins depends on the nucleic acid sequences to which they bind (18, 26-28).

# Site size of XWKK binding to polyU is invariant with pH and larger than predicted by charge neutralization alone

Nonspecific binding of many cationic ligands to nucleic acids exhibits a site size approximately equal to the number of charges neutralized in the ligand-nucleic acid complex. Examples include inorganic salts such as  $\mathrm{Mg^{2+}}$  (6,47) and  $\mathrm{Co(NH_3)_6^{3+}}$  (31), polyamines (32), and oligopeptides of varying lengths (4-6,13,34,35,46). A small ligand such as  $\mathrm{Mg^{2+}}$  binding to polyU or poly(rA)-polyU has a site size  $n=2.0\pm0.1$  and  $n=2.2\pm0.1$ , respectively (47), whereas KWK<sub>6</sub> bound to homopolymeric dT ssDNA with  $n\approx8$  under stoichiometric binding conditions (35). The site sizes of KWKK and KWKK were larger than expected if a 1:1 correlation of charge to lattice occupation size is assumed. For both HWKK and KWKK, the site size was about 5 nucleotides both by nonlinear least squares analysis of binding isotherms and via independent measurement of stoichiometric binding at low pH. In addition, we saw no obvious dependence on site size with protonation state, even for HWKK whose charge ( $Z_L$ ) decreases from ~4 to ~2 over pH 5.2 – 8.7 (Table S1).

Flexibility of the oligopeptide and/or nucleic acid may play a role in explaining these findings. Unstructured oligopeptides of intermediate length appear to have slightly larger site sizes when binding double stranded DNA (dsDNA) than would be expected from ligand charge alone. Spermine (a polyamine with  $Z_L = 4$ ) has an average site size of  $n = 4.9 \pm 0.3$  phosphates at pH 6.5 between 71 – 132 mM Na<sup>+</sup> (32) and WK<sub>4</sub> ( $Z_L = 4$ ) binds at 6.4 mM Na<sup>+</sup> pH 7.0 with  $n = 4.6 \pm 0.5$  (34). Larger ligands of the form  $\epsilon$ -DNP-Lys-(Lys)<sub>x</sub>, where x indicates the number of unmodified lysines and the total charge  $Z_L$ , bind poly(rA)-polyU with n = 6.6 when  $Z_L = 5$  and with n = 7.1 when  $Z_L = 6$  (9,30). These results suggest that ligand conformational constraints may require a larger nucleic acid site size than expected based on charge neutralization alone. The converse of this phenomenon, where ligand flexibility permits interstitial residue compaction upon nucleic acid binding, was observed for a series of 17-residue  $Z_L = 4$  oligopeptides which each bound dsDNA with a comparable site size of  $n = 4.0 \pm 0.7$  and most exhibited a concomitant conformational transition from random-coil to  $\alpha$ -helix by circular dichroism (34). The pH-invariant site size  $n \approx 5$  for XWKK may be due to inherent inflexibility which prohibits decreases in site size as  $Z_L$  decreases.

To determine the dependence of nucleic acid binding as a function of peptide charge, nucleic acid base composition and salt effects, Lohman  $et\ al.$  estimated site size either by linear extrapolation of  $v/L_F$  to  $v/L_F=0$  on a Scatchard plot (4) or assumed that site size was proportional to oligopeptide charge (5,6,13). Our results suggest that this assumption may not be justified for the shorter oligopeptides they considered. However, because the McGhee-von Hippel isotherm determines  $K_{obs}$  in the zero-binding density limit (i.e., as if a single ligand were binding at the center of an infinitely long lattice), site size effects are expected to be fairly inconsequential. Indeed, Mascotti and Lohman reported that increasing the site size ~20% resulted in only a 4% perturbation to  $log K_{obs}$  (6). Given the comparative approach of their studies and the relatively insensitive nature of the McGhee-von Hippel equilibrium binding constant to n, the discrepancy in site size does not impinge the major findings of these previous studies. Nevertheless, the larger site size has potential implications for complex formation with finite lattices at higher binding densities where ligands have to compete for overlapping sites.

In conclusion, our nonspecific binding studies of HWKK and KWKK to polyU reveal that protonated histidine and lysine have equivalent effects on the stability of these nucleic acid complexes and that reduction in the binding constant with increasing salt concentration is welldescribed by limiting-law polyelectrolyte theory. Both tetrapeptides, with a maximum of 4 positive charges, have a site size of ~5 nucleotides and do not vary significantly with pH or salt concentration. Binding constants for nonspecific interactions of HWKK and KWKK to polyU increase with protonation of histidine and the  $\alpha$ -amino group, but pK<sub>a</sub> shifts coupled to binding are not observed and there is no requirement for protonation. This contrasts with welldocumented examples of histidine pKa shifts and proton uptake in the binding of native proteins to nucleic acids. We propose that this disparity in histidine protonation behavior derives from differences in interface hydration. The interface in the nonspecific, Coulombically-driven complexes of HWKK and KWKK with polyU is almost certainly fully hydrated, in contrast to the largely dehydrated, complementary interfaces of many native protein-nucleic acid complexes. We find that the additional Coulombic stabilization of a nonspecific HWKK-polyU complex provided by protonation is insufficient to cause a histidine pKa shift and proton uptake upon binding. Conversely, the pKa shifts and coupled protonation observed in several proteinnucleic acid binding processes (presumably) reflect the requirement for a salt-bridge or other specific interaction with the protonated histidine in the significantly dehydrated binding interface.

# **Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

#### Abbreviations used

ss single-stranded

dsDNA double stranded DNA
GR glucocorticoid receptor

PNAI protein-nucleic acid interactions

polyU polyuridylic acid

AICc second-order Akaike Information Criterion

EDTA ethylenediaminetetraacetic acid

HEPES 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

LBD ligand binding density

SAM sterile alpha motif

lacR lac repressor

PwTBP Pyrococcus woesei TATA-binding protein
 E2C human papillomavirus type 16 E2 protein
 T-ag-obd origin-binding domain of SV40 T antigen.

## References

 Nadassy K, Wodak SJ, Janin J. Structural features of protein–nucleic acid recognition sites. Biochemistry 1999;38:1999–2017. [PubMed: 10026283]

- 2. Honig B, Nicholls A. Classical electrostatics in biology and chemistry. Science 1995;268:1144–1149. [PubMed: 7761829]
- Allers J, Shamoo Y. Structure-based analysis of protein-RNA interactions using the program ENTANGLE. J. Mol. Biol 2001;311:75–86. [PubMed: 11469858]
- Mascotti DP, Lohman TM. Thermodynamic extent of counterion release upon binding oligolysines to single-stranded nucleic acids. Proc. Natl. Acad. Sci. USA 1990;87:3146.
- Mascotti DP, Lohman TM. Thermodynamics of single-stranded RNA and DNA interactions with oligolysines containing tryptophan. Effects of base composition. Biochemistry 1993;32:10568–10579.
   [PubMed: 7691177]
- Mascotti DP, Lohman TM. Thermodynamics of single-stranded RNA binding to oligolysines containing tryptophan. Biochemistry 1992;31:8932–8946. [PubMed: 1382582]
- Lohman TM, DeHaseth PL, Record MT Jr. Pentalysine-deoxyribonucleic acid interactions: a model for the general effects of ion concentrations on the interactions of proteins with nucleic acids. Biochemistry 1980;19:3522–3530. [PubMed: 7407056]
- Record J, Lohman TM, Haseth P. d. Ion effects on ligand-nucleic acid interactions. J. Mol. Biol 1976;107:145–158. [PubMed: 1003464]
- Latt SA, Sober HA. Protein-nucleic acid interactions. II. Oligopeptide-polyribonucleotide binding studies. Biochemistry 1967;6:3293–3306. [PubMed: 6056990]
- 10. Latt SA, Sober HA. Protein-nucleic acid interactions. III. Cation effect on binding strength and specificity. Biochemistry 1967;6:3307–3314. [PubMed: 4861938]
- 11. Helene C, Maurizot JC. Interactions of oligopeptides with nucleic acids. CRC Crit. Rev. Biochem 1981;10:213–258. [PubMed: 6260429]
- 12. Helene C, Lancelot G. Interactions between functional groups in protein-nucleic acids associations. Prog. Biophys. Mol. Biol 1982;39:1–68. [PubMed: 6175011]
- 13. Mascotti DP, Lohman TM. Thermodynamics of oligoarginines binding to RNA and DNA. Biochemistry 1997;36:7272–7279. [PubMed: 9188729]
- 14. Schultz SC, Shields GC, Steitz TA. Crystal structure of a CAP-DNA complex the DNA is bent by 90°. Science 1991;253:1001–1007. [PubMed: 1653449]
- Parkinson G, Wilson C, Gunasekera A, Ebright YW, Ebright RE, Berman HM. Structure of the CAP-DNA complex at 2.5 Å resolution: A complete picture of the protein-DNA interface. J. Mol. Biol 1996;260:395–408. [PubMed: 8757802]
- 16. Muller CW, Rey FA, Sodeoka M, Verdine GL, Harrison SC. Structure of the NF-κB p50 homodimer bound to DNA. Nature 1995;373:311–317. [PubMed: 7830764]
- 17. Doublie S, Tabor S, Long AM, Richardson CC, Ellenberger T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. Nature 1998;391:251–258. [PubMed: 9440688]
- 18. Luisi BF, Xu WX, Otwinowski Z, Freedman LP, Yamamoto KR, Sigler PB. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. Nature 1991;352:497–505. [PubMed: 1865905]
- 19. Aviv T, Lin Z, Ben-Ari G, Smibert CA, Sicheri F. Sequence-specific recognition of RNA hairpins by the SAM domain of Vts1p. Nat Struct Mol Biol 2006;13:168–176. [PubMed: 16429151]

 Astatke M, Grindley NDF, Joyce CM. Deoxynucleoside triphosphate and pyrophosphate bindingsites in the catalytically competent ternary complex for the polymerase reaction catalyzed by DNA-Polymerase-I (Klenow fragment). J. Biol. Chem 1995;270:1945–1954. [PubMed: 7829532]

- Park C, Schultz LW, Raines RT. Contribution of the active site histidine residues of Ribonuclease A to nucleic acid binding. Biochemistry 2001;40:4949

  –4956. [PubMed: 11305910]
- Kim OTP, Yura K, Go N. Amino acid residue doublet propensity in the protein-RNA interface and its application to RNA interface prediction. Nucl. Acids Res 2006;34:6450–6460. [PubMed: 17130160]
- Luscombe NM, Laskowski RA, Thornton JM. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucl. Acids Res 2001;29:2860–2874.
   [PubMed: 11433033]
- 24. Harrison JG, Balasubramanian S. Synthesis and hybridization analysis of a small library of peptideoligonucleotide conjugates. Nucl. Acids Res 1998;26:3136–3145. [PubMed: 9628910]
- 25. Lundback T, van den Berg S, Hard T. Sequence-specific DNA binding by the glucocorticoid receptor DNA-binding domain is linked to a salt-dependent histidine protonation. Biochemistry 2000;39:8909–8916. [PubMed: 10913303]
- DeHaseth PL, Lohman TM, Record MT. Nonspecific interaction of lac repressor with DNA: an association reaction driven by counterion release. Biochemistry 1977;16:4783–4790. [PubMed: 911789]
- 27. Bradshaw EM, Sanford DG, Luo X, Sudmeier JL, Gurard-Levin ZA, Bullock PA, Bachovchin WW. T antigen origin-binding domain of Simian Virus 40: determinants of specific DNA binding. Biochemistry 2004;43:6928–6936. [PubMed: 15170330]
- 28. Eliseo T, Sánchez IE, Nadra AD, Dellarole M, Paci M, de Prat Gay G, Cicero DO. Indirect DNA readout on the protein side: coupling between histidine protonation, global structural cooperativity, dynamics, and DNA binding of the human papillomavirus type 16 E2C domain. J. Mol. Biol 2009;388:327–344. [PubMed: 19285507]
- 29. Bergqvist S, Williams MA, O'Brien R, Ladbury JE. Heat capacity effects of water molecules and ions at a protein-DNA interface. J. Mol. Biol 2004;336:829–842. [PubMed: 15095863]
- 30. McGhee JD, von Hippel PH. Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. J. Mol. Biol 1974;86:469–489. [PubMed: 4416620]
- 31. Plum GE, Bloomfield VA. Equilibrium dialysis study of binding of hexammine cobalt(III) to DNA. Biopolymers 1988;27:1045–1051. [PubMed: 3401555]
- 32. Braunlin WH, Strick TJ, Record MT Jr. Equilibrium dialysis studies of polyamine binding to DNA. Biopolymers 1982;21:1301–1314. [PubMed: 7115891]
- 33. Bujalowski W, Lohman TM. Negative co-operativity in *Escherichia coli* single strand binding protein-oligonucleotide interactions: I. Evidence and a quantitative model. J. Mol. Biol 1989;207:249–268. [PubMed: 2661832]
- 34. Padmanabhan S, Zhang W, Capp MW, Anderson CF, Record MT Jr. Binding of cationic (+4) alanine-and glycine-containing oligopeptides to double-stranded DNA: thermodynamic analysis of effects of Coulombic interactions and α-helix induction. Biochemistry 1997;36:5193–5206. [PubMed: 9136881]
- 35. Zhang W, Bond JP, Anderson CF, Lohman TM, Record MT. Large electrostatic differences in the binding thermodynamics of a cationic peptide to oligomeric and polymeric DNA. Proc. Natl. Acad. Sci. USA 1996;93:2511–2516. [PubMed: 8637905]
- 36. Record J, Zhang W, Anderson CF. 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 1998;51:281–353. [PubMed: 9615173]
- 37. Epstein IR. Cooperative and non-cooperative binding of large ligands to a finite one-dimensional lattice. A model for ligand-oligonucleotide interactions. Biophys. Chem 1978;8:327–339. [PubMed: 728537]
- 38. Ni HH, Anderson CF, Record MT. Quantifying the thermodynamic consequences of cation (M<sup>2+</sup>, M<sup>+</sup>) accumulation and anion (X<sup>-</sup>) exclusion in mixed salt solutions of polyanionic DNA using Monte

- Carlo and Poisson-Boltzmann calculations of ion-polyion preferential interaction coefficients. J. Phys. Chem. B 1999;103:3489–3504.
- Rouzina I, Bloomfield VA. Competitive electrostatic binding of charged ligands to polyelectrolytes: Practical approach using the non-linear Poisson-Boltzmann equation. Biophys. Chem 1997;64:139–155. [PubMed: 17029833]
- Rouzina I, Bloomfield VA. Macroion attraction due to electrostatic correlation between screening counterions .1. Mobile surface-adsorbed ions and diffuse ion cloud. J. Phys. Chem 1996;100:9977– 9989
- 41. Ray J, Manning GS. Theory of delocalized ionic binding to polynucleotides structural and excluded-volume effects. Biopolymers 1992;32:541–549. [PubMed: 1515545]
- 42. Lohman TM, Mascotti DP. Nonspecific ligand-DNA equilibrium binding parameters determined by fluorescence methods. Methods Enzymol 1992;212:424–458. [PubMed: 1518458]
- 43. Overman LB, Bujalowski W, Lohman TM. Equilibrium binding of *Escherichia coli* Single-Strand Binding Protein to single-stranded nucleic-acids in the (SSB)<sub>65</sub> binding mode cation and anion effects and polynucleotide specificity. Biochemistry 1988;27:456–471. [PubMed: 3280021]
- Anderson CF, Record MT. Salt nucleic-acid interactions. Ann. Rev. of Phys. Chem 1995;46:657–700. [PubMed: 7495482]
- 45. Anderson CF, Record MT. Salt dependence of oligoion polyion binding a thermodynamic description based on preferential interaction coefficients. J. Phys. Chem 1993;97:7116–7126.
- 46. Zhang WT, Ni HH, Capp MW, Anderson CF, Lohman TM, Record MT. The importance of Coulombic end effects: experimental characterization of the effects of oligonucleotide flanking charges on the strength and salt dependence of oligocation (L<sup>8+</sup>) binding to single-stranded DNA oligomers. Biophys. J 1999;76:1008–1017. [PubMed: 9916032]
- 47. Record MT, Anderson CF, Lohman TM. Thermodynamic analysis of ion effects on 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 1978;11:103–178. [PubMed: 353875]
- 48. Record MT, Ha JH, Fisher MA. Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA. Methods Enzymol 1991;208:291–343. [PubMed: 1779839]
- 49. Olmsted MC, Bond JP, Anderson CF, Record MT. Grand canonical Monte Carlo molecular and thermodynamic predictions of ion effects on binding of an oligocation (L<sup>8+</sup>) to the center of DNA oligomers. Biophys. J 1995;68:634–647. [PubMed: 7696515]
- Anderson CF, Record MT. The relationship between the Poisson-Boltzmann model and the condensation hypothesis: an analysis based on the low salt form of the Donnan coefficient. Biophys. Chem 1980;11:353–360. [PubMed: 16997254]
- Manning GS. Limiting laws and counterion condensation in polyelectrolyte solutions I. Colligative properties. J. Chem. Phys 1969;51:924–933.
- 52. Edelhoch H. Spectroscopic determination of tryptophan and tyrosine in proteins. Biochemistry 1967;6:1948–1954. [PubMed: 6049437]
- 53. Ballin JD, Shkel IA, Record MT Jr. 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. Nucl. Acids Res 2004;32:3271–3281. [PubMed: 15205469]
- 54. Shkel IA, Ballin JD, Record MT Jr. 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. Biochemistry 2006;45:8411–8426. [PubMed: 16819840]
- 55. Johnson ML, Frasier SG. Nonlinear least-squares analysis. Methods Enzymol 1985;117:301–342.
- Motulsky, H.; Christopoulos, A. Fitting Models to Biological Data Using Linear and Nonlinear Regression. A Practical Guide to Curve Fitting. GraphPad Software, Inc.; San Diego, CA: 2003.
- 57. Hurvich CM, Tsai CL. Regression and time series model selection in small samples. Biometrika 1989;76:297–307.
- 58. Bujalowski W, Lohman TM. A general method of analysis of ligand-macromolecule equilibria using a spectroscopic signal from the ligand to monitor binding. Application to *Escherichia coli* single-

- strand binding protein-nucleic acid interactions. Biochemistry 1987;26:3099–3106. [PubMed: 3300771]
- Bond JP, Anderson CF, Record MT Jr. Conformational transitions of duplex and triplex nucleic acid helices: thermodynamic analysis of effects of salt concentration on stability using preferential interaction coefficients. Biophys. J 1994;67:825–836. [PubMed: 7948695]
- Kao YH, Fitch CA, Bhattacharya S, Sarkisian CJ, Lecomte JTJ, Garcia-Moreno E,B. Salt effects on ionization equilibria of histidines in myoglobin. Biophys. J 2000;79:1637–1654. [PubMed: 10969024]
- 61. Lee KK, Fitch CA, Lecomte JTJ, Garcia-Moreno E,B. Electrostatic effects in highly charged proteins: salt sensitivity of pK<sub>a</sub> values of histidines in Staphylococcal nuclease. Biochemistry 2002;41:5656–5667. [PubMed: 11969427]
- 62. Shire S, Hanania G, Gurd F. Electrostatic effects in myoglobin. Application of the modified Tanford-Kirkwood theory to myoglobins from horse, California grey whale, harbor seal, and California sea lion. Biochemistry 1975;14:1352–1358. [PubMed: 235950]
- 63. Shire S, Hanania G, Gurd F. Electrostatic effects in myoglobin. Hydrogen ion equilibria in sperm whale ferrimyoglobin. Biochemistry 1974;13:2967–2974. [PubMed: 4407621]
- 64. Ferreiro DU, Dellarole M, Nadra AD, de Prat-Gay G. Free energy contributions to direct readout of a DNA sequence. J. Biol. Chem 2005;280:32480–32484. [PubMed: 16000299]
- 65. Dawson, RMC.; Elliott, DC.; Elliot, WH.; Jones, KM. Data for Biochemical Research. Oxford University Press; New York: 1995.
- 66. Barkley MD, Lewis PA, Sullivan GE. Ion effects on the lac repressor-operator equilibrium. Biochemistry 1981;20:3842–3851. [PubMed: 7272280]
- 67. Kalodimos CG, Biris N, Bonvin AMJJ, Levandoski MM, Guennuegues M, Boelens R, Kaptein R. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. Science 2004;305:386–389. [PubMed: 15256668]
- 68. Cicero DO, Nadra AD, Eliseo T, Dellarole M, Paci M, de Prat-Gay G. Structural and thermodynamic basis for the enhanced transcriptional control by the human papillomavirus strain-16 E2 protein. Biochemistry 2006;45:6551–6560. [PubMed: 16716065]
- 69. Winkler FK, Banner DW, Oefner C, Tsernoglou D, Brown RS, Heathman SP, Bryan RK, Martin PD, Petratos K, Wilson KS. The crystal structure of *Eco*RV endonuclease and of its complexes with cognate and non-cognate DNA fragments. EMBO J 1993;12:1781–1795. [PubMed: 8491171]
- 70. Gewirth DT, Sigler PB. The basis for half-site specificity explored through a non-cognate steroid receptor-DNA complex. Nat Struct Mol Biol 1995;2:386–394.
- 71. Ferreiro DU, Sanchez IE, de Prat Gay G. Transition state for protein–DNA recognition. Proc. Natl. Acad. Sci. USA 2008;105:10797–10802. [PubMed: 18664582]
- Gao M, Skolnick J. From nonspecific DNA-protein encounter complexes to the prediction of DNA-protein interactions. PLoS Comput Biol 2009;5:e1000341. [PubMed: 19343221]
- 73. Woda J, Schneider B, Patel K, Mistry K, Berman HM. An analysis of the relationship between hydration and protein-DNA interactions. Biophys. J 1998;75:2170–2177. [PubMed: 9788911]
- 74. Capp MW, Pegram LM, Saecker RM, Kratz M, Riccardi D, Wendorff T, Cannon JG, Record MT. Interactions of the osmolyte glycine betaine with molecular surfaces in water: thermodynamics, structural interpretation, and prediction of m-values. Biochemistry 2009;48:10372–10379. [PubMed: 19757837]
- 75. Meulen KAV, Saecker RM, Record MT. Formation of a wrapped DNA-protein interface: experimental characterization and analysis of the large contributions of ions and water to the thermodynamics of binding IHF to H' DNA. J. Mol. Biol 2008;377:9–27. [PubMed: 18237740]
- 76. Kontur WS, Saecker RM, Davis CA, Capp MW, Record MT. Solute probes of conformational changes in open complex (RP $_{\rm O}$ ) formation by *Escherichia coli* RNA polymerase at the  $\lambda P_{\rm R}$  promoter: evidence for unmasking of the active site in the isomerization step and for large-scale coupled folding in the subsequent conversion to RP $_{\rm O}$ . Biochemistry 2006;45:2161–2177. [PubMed: 16475805]
- 77. Pegram LM, Record MT. Thermodynamic origin of Hofmeister ion effects. J. Phys. Chem. B 2008;112:9428–9436. [PubMed: 18630860]
- 78. Stigter D, Dill KA. Binding of ionic ligands to polyelectrolytes. Biophys. J 1996;71:2064–2074. [PubMed: 8889181]

79. Srivenugopal K, Wemmer D, Morris D. Aggregation of DNA by analogs of spermidine; enzymatic and structural studies. Nucl. Acids Res 1987;15:2563–2580. [PubMed: 3031604]

80. Wemmer DE, Srivenugopal KS, Reid BR, Morris DR. Nuclear magnetic resonance studies of polyamine binding to a defined DNA sequence. J. Mol. Biol 1985;185:457–459. [PubMed: 4057253]

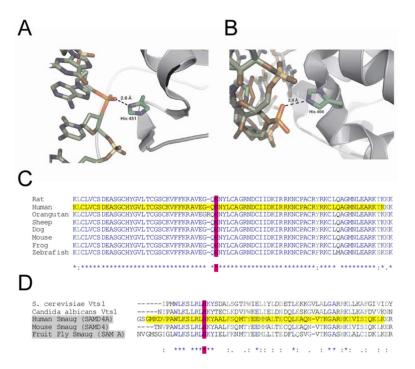


Figure 1.

Representative crystal structures of conserved histidine-phosphodiester interactions. (A) The co-crystal of the rat glucocorticoid receptor DNA binding domain complexed with a glucocorticoid recognition element (pdb:1R4R) (18). (B) A co-crystal of the SAM domain of the Saccharomyces cerevisiae post-transcriptional regulator Vts1p recognizing an RNA hairpin SAM recognition element (pdb:2F8K) (19). (C) CLUSTALW analysis of a region within the glucocorticoid receptor DNA binding domain, highlighted in yellow. Database sequences were extracted from the following accession numbers: rat (*Rattus norvegicus*), NP\_036708; human (Homo sapiens), CAJ65924; orangutan (Pongo abelii), NP\_001126305; sheep (Ovis aries), NP\_001107658; dog (Canis familiaris), ABA40754; mouse (Mus musculus), NP\_032199; frog, NP\_001081531; zebrafish (Danio rerio), NP\_032199. (D) CLUSTALW analysis of the SAM RNA binding domain. Species highlighted in grey possess Smaug, a homologue of Vts1, which contains a conserved SAM domain highlighted yellow for the human sequence. Accession numbers used were Saccharomyces cerevisiae, NP\_015004; Candida albicans, Q5AI80; human (Homo sapiens), NP\_056404; mouse (Mus musculus), NP\_083242; fruit fly (Drosophila melanogaster), NP\_523987. In (C) and (D), the respective conserved histidines shown in panels (A) and (B) are highlighted in pink. The symbols below the sequence data in (C) and (D) denote CLUSTALW-defined sequence conservation: "\*" indicates a fully conserved residue; ":", strongly conserved; ".", weakly conserved.

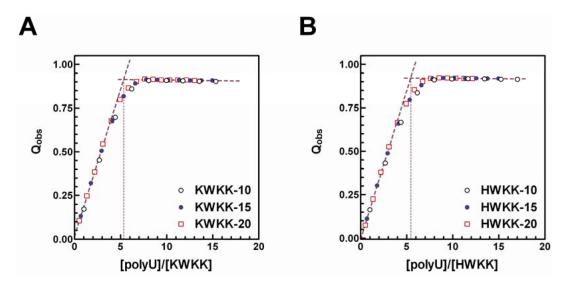


Figure 2. Determination of XWKK site size on polyU by stoichiometric binding. Titrations of (A) KWKK or (B) HWKK as a function of initial oligopeptide concentration (see legend above) at 3.4 mM Na<sup>+</sup> and 25 °C (*i.e.*, high affinity conditions). Intersection of the linear least square fits of the two regimes (dashed purple lines) is indicated by a light purple dotted line for clarity. The intersections for indicate a site size  $n = 5.3 \pm 0.1$  for KWKK and  $n = 5.4 \pm 0.1$  for HWKK.

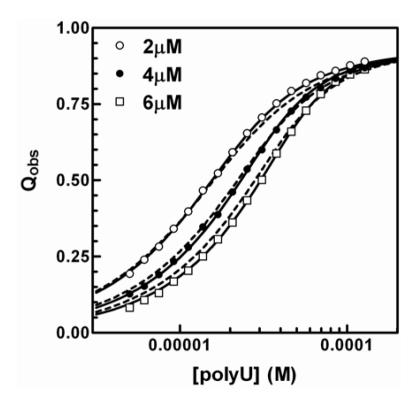
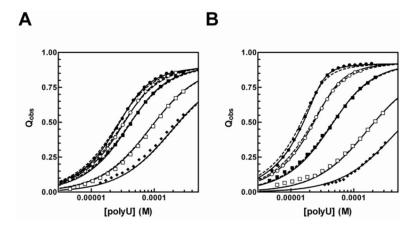


Figure 3. Titrations of the oligopeptide HWKK with polyU. Titrations at three HWKK concentrations  $(2 \,\mu\text{M}\,(\odot), 4 \,\mu\text{M}\,(\bullet))$ , and  $6 \,\mu\text{M}\,(\Box))$  observed as function of tryptophan fluorescence quenching (eq. 14) at 9.8 mM Na<sup>+</sup>, pH 6.0, and 25 °C. The concentration of polyU is given in moles of nucleotides per liter. Solid and dashed lines are the nonlinear least squares global fits across all three titrations using eq. 1, 3, and 4, with  $Q_{max} = 0.93$  as determined by ligand binding density analysis (35). The dashed line reports the McGhee-von Hippel isotherm with  $\log K_{obs} = 5.11 \pm 0.02$  with a site size fixed at n = 4. The solid line isotherm shows the fit with the site size allowed to float ( $\log K_{obs} = 5.22 \pm 0.02$ ,  $n = 4.96 \pm 0.11$ ). If all parameters are allowed to float,  $\log K_{obs} = 5.23 \pm 0.03$ ,  $Q_{max} = 0.929 \pm 0.007$ , and  $n = 4.97 \pm 0.13$  (data not shown in plot above).



**Figure 4.** XWKK binding isotherms as a function of pH. Titrations of (A) KWKK at 24.4 mM Na<sup>+</sup> or (B) HWKK at 9.8 mM Na<sup>+</sup> as a function of polyU. Data from titrations performed at 25 °C with 4  $\mu$ M oligopeptide are presented for each pH with the following symbols: (•), pH 5.2; (o), pH 6.0; (•), pH 7.0; (□), pH 8.0; (•), pH 8.7. Curves represent fits determined by global nonlinear least squares analysis across all oligopeptide concentrations (2 – 6  $\mu$ M) at the specified pH and salt concentration (*cf.* Figure 3). Lines represent the best global fits to the data when the site size is constrained to n = 4 (dashed lines) or n = 5 (solid lines; Table 1).

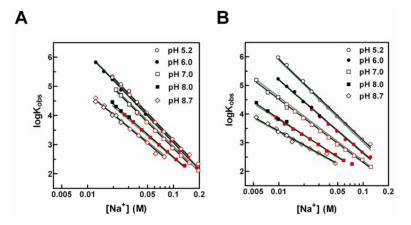
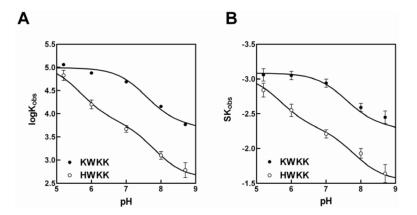


Figure 5. Dependence of  $\log K_{obs}$  on  $[Na^+]$  and pH.  $\log K_{obs}$ , determined directly by global analysis of multiple reverse titrations (black symbols) or via calculation using eq. 1, 2 and 4 and saltback data (red symbols) are plotted as a function of pH and salt concentration. Black solid lines represent linear least squares fits of combined reverse titration and saltback data (reported in Table 1). Green solid lines indicate the predicted dependence of  $\log K_{obs}$  on  $[Na^+]$  and pH using eq. 13 - 14 and parameters found by global nonlinear least squares analysis of all  $\log K_{obs}$  vs  $[Na^+]$  data (Table 2).



**Figure 6.** Model description of XWKK-polyU logK<sub>obs</sub> and SK<sub>obs</sub> as a function of pH. (A) LogK<sub>obs</sub> for XWKK binding to polyU at 25 °C and 24.4 mM Na<sup>+</sup> are plotted as a function of pH. KWKK logK<sub>obs</sub> was directly determined from reverse titration data ( $\bullet$ ), whereas HWKK logK<sub>obs</sub> ( $\circ$ ) was interpolated *via* eq. 6 using the parameters in Table 1. Curves represent the global nonlinear least squares fit of the combined logK<sub>obs</sub> vs [salt] and pH data in Figure 6 to eq. 9 (Table 2). (B) SK<sub>obs</sub>, as reported in Table 1, is plotted as a function of pH for KWKK and HWKK. Curves represent the fits to eq. 13 - 14 using the parameters in Table 2.

 $\label{eq:Table 1} \textbf{Table 1}$  pH dependence of McGhee-von Hippel log K  $_{obs}$  and SK  $_{obs}$  as a function of amino acid composition

pН	$\log { m K_{obs}}^a$	$\mathrm{SK}_{\mathrm{obs}}{}^{b}$	logK <sub>0</sub> <sup>b</sup> (1 M Na <sup>+</sup> )	$Q_{max}^{c}$ (%)
	<b>KWKK</b> (24.4 mM Na <sup>+</sup> )			
5.2	$5.06 \pm 0.02$	$-3.06 \pm 0.09$	$0.10 \pm 0.12$	$0.91 \pm 0.01$
6.0	$4.88 \pm 0.01$	$-3.05 \pm 0.06$	$0.01 \pm 0.09$	$0.91 \pm 0.01$
7.0	$4.69 \pm 0.02$	$-2.94 \pm 0.06$	$-0.06 \pm 0.08$	$0.91 \pm 0.02$
8.0	$4.16\pm0.01$	$-2.59 \pm 0.06$	$-0.01 \pm 0.09$	$0.91 \pm 0.02$
8.7	$3.77 \pm 0.02$	$-2.45\pm0.11$	$-0.13 \pm 0.17$	$0.87 \pm 0.01^{\textit{d}}$
	<b>HWKK</b> (9.8 mM Na <sup>+</sup> )			
5.2	$5.98 \pm 0.02$	$-2.84\pm0.10$	$0.25 \pm 0.15$	$0.92 \pm 0.02$
6.0	$5.23 \pm 0.02$	$-2.55 \pm 0.09$	$0.09 \pm 0.12$	$0.93 \pm 0.01$
7.0	$4.59 \pm 0.02$	$-2.21 \pm 0.06$	$0.11 \pm 0.11$	$0.93 \pm 0.01$
8.0	$3.83 \pm 0.02$	$-1.93 \pm 0.07$	$-0.01 \pm 0.12$	$0.88 \pm 0.01^{\textit{e}}$
8.7	$3.38 \pm 0.02$	$-1.64 \pm 0.13$	$0.14 \pm 0.24$	$0.81 \pm 0.02^e$

<sup>&</sup>lt;sup>a</sup>Determined by global nonlinear least squares analysis of data collected at 25 °C across multiple oligopeptide concentrations at the specified salt concentration using the noncooperative McGhee-von Hippel equation (eq. 1) with the indicated  $Q_{max}$  and n = 5. Errors represent 95% confidence intervals.

 $<sup>^</sup>b$ SK<sub>Obs</sub> and logK<sub>O</sub> values were calculated by nonlinear least squares analysis of aggregate logK<sub>Obs</sub> versus [Na<sup>+</sup>] data from reverse titrations and saltbacks. Saltback logK<sub>Obs</sub> data were obtained using the indicated  $Q_{max}$  and n = 5 via eq. 1 – 4.

<sup>&</sup>lt;sup>c</sup>Calculated via model-independent ligand binding density analysis (42) of reverse titration data at the specified salt concentration.

dWhen binding was weaker than  $K_{\text{Obs}} \lesssim 10^4 \,\text{M}^{-1}$ ,  $Q_{max}$  was found by ligand binding density analysis using data collected at the same pH but lower salt concentration to improve the accuracy of the determination (*cf.* Figure S1). At high pH,  $Q_{max}$  is reported at 12.4 mM Na<sup>+</sup> for KWKK

<sup>&</sup>lt;sup>e</sup>When binding was weaker than  $K_{Obs} \lesssim 10^4 \,\mathrm{M}^{-1}$ ,  $Q_{max}$  was found by ligand binding density analysis using data collected at the same pH but lower salt concentration to improve the accuracy of the determination (*cf.* Figure S1). At high pH,  $Q_{max}$  is reported at 5.4 mM Na<sup>+</sup> for HWKK.

 $\label{eq:Table 2} \textbf{Global fit parameters describing the pH dependence of } \log K_{obs} \ \textit{versus} \ \text{amino acid composition}^{\textit{a}}$ 

			$log K_0$	$\mathbf{pK}_{\mathbf{a}}$	
pe	peptide	Ψ	(1 M Na <sup>+</sup> )	Histidine	$\alpha N{H_3}^+$
K	WKK	$0.77 \pm 0.01$	$0.02 \pm 0.07$		$7.57 \pm 0.08$
Н	WKK	$0.76 \pm 0.03$	$0.12 \pm 0.10$	$5.75 \pm 0.18$	$7.86 \pm 0.13$

 $<sup>^</sup>a$ Values reported are the globally fitted values to eq. 13 for all reverse titration and salt back data in Figure 5 (N  $\approx$  70 – 71 data points each). Errors represent 95% confidence intervals for the global fit.