

---

# Linkage between proton binding and folding in RNA: A thermodynamic framework and its experimental application for investigating $pK_a$ shifting

---

ELLEN M. MOODY, JULIETTE T. J. LECOMTE, and PHILIP C. BEVILACQUA

Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

## ABSTRACT

Perturbation of  $pK_a$  values can change the favored protonation states of the nucleobases at biological pH and thereby modulate the function of RNA and DNA molecules. In an effort to understand the driving forces for  $pK_a$  shifting specific to nucleic acids, we developed a thermodynamic framework that relates proton binding to the nucleobases and the helix–coil transition. Key features that emerge from the treatment are a comprehensive description of all the actions of proton binding on RNA folding; acid and alkaline denaturation of the helix and  $pK_a$  shifting in the folded state. Practical experimental approaches for measuring  $pK_a$ s from thermal denaturation experiments are developed. Microscopic  $pK_a$  values (where  $k_a$  is the acid dissociation constant) for the unfolded state were determined directly by experiments on unstructured oligonucleotides, which led to a macroscopic  $pK_a$  for the ensemble of unfolded states shifted toward neutrality. The formalism was then applied to pH-dependent UV melting data for model DNA oligonucleotides. Folded-state  $pK_a$  values were in good agreement with the outcome of pH titrations, and the acid and alkaline denaturation regions were well described. The formalism developed here is similar to that of Draper and coworkers for  $Mg^{2+}$  binding to RNA, except that the unfolded state is described explicitly owing to the presence of specific proton-binding sites on the bases. A principal conclusion is that it should be possible to attain large  $pK_a$  shifts by designing RNA molecules that fold cooperatively.

**Keywords:**  $pK_a$  shifting; RNA folding; RNA catalysis

## INTRODUCTION

Under physiological conditions, nucleic acids are poly-anions and therefore associate with and are stabilized by the binding of cations. Binding of mono- and divalent metal ions has been studied in depth (Record et al. 1976; Manning 1978; Misra et al. 2003). Less is known, however, about the binding of protons. Nucleobases in the unfolded state have  $pK_a$  values of  $\leq 4.2$  and  $\geq 9.3$  (Saenger 1984), and shift even further from neutrality in typical Watson-Crick base pairs (Legault and Pardi 1997). Numerous folded-state  $pK_a$ s perturbed toward neutrality have been reported as well. For instance,  $A^+ \cdot C$  mismatches with  $pK_a$ s of  $\approx 6.5$  include the lead-dependent RNA enzyme, or leadzyme (Legault and Pardi 1997); the U6 loop of the spliceosome (Huppler et al. 2002); and a DNA duplex (Wang et al. 1991). Also, the A

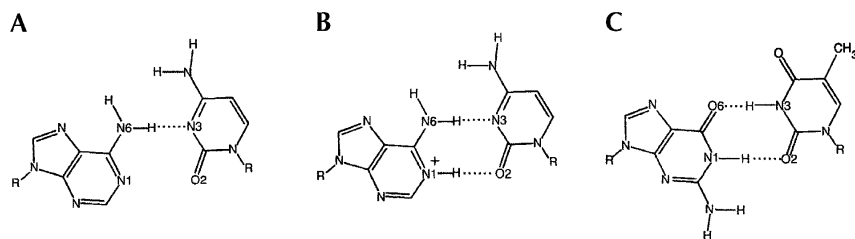
loop of 23S ribosomal RNA has a cytosine with a  $pK_a$  of 6.4 (Blanchard and Puglisi 2001), and an  $A^+ \cdot C$  mismatch in a DNA hairpin has been recently shown to have a  $pK_a$  of 7.1 at temperatures near 15°C (Moody et al. 2004a). In all these cases, it is likely that linkage of proton binding and folding is responsible for  $pK_a$  shifting. Other physical means of shifting  $pK_a$ s are also possible; these include charge–charge interactions (Nakano et al. 2000; Huppler et al. 2002) and direct coordination of metal ions to the bases (Roitzsch and Lippert 2004).

$pK_a$ s shifted toward neutrality have been separated into two classes on the basis of structural and functional features (Narlikar and Herschlag 1997; Bevilacqua et al. 2004). Class I sites are defined as those for which the loaded proton is involved in hydrogen bonding, as in the case of the N1 of A in an  $A^+ \cdot C$  mismatch (Fig. 1), whereas Class II sites are those for which the proton is not sequestered (Bevilacqua et al. 2004). Protonated nucleobases have the potential to play functional roles in both catalysis and folding. Class II sites have been implicated in proton transfer in the HDV and hairpin ribozymes (Perrotta et al. 1999; Nakano et al. 2000;

---

**Reprint requests to:** Philip C. Bevilacqua, Department of Chemistry, The Pennsylvania State University, University Park, PA 16802, USA; e-mail: pcb@chem.psu.edu; fax: (814) 863-8403.

Article and publication are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.7177505>.



**FIGURE 1.**  $A^+ \cdot C$  and  $G \cdot U/T$  mismatches adopt wobble structures. R represents the ribose sugar. (A) Unprotonated  $A \cdot C$ . This pairing is unlikely to form because it has only a single hydrogen bond. (B)  $A^+ \cdot C$  wobble pair involving a protonated N1 with two hydrogen bonds. This pairing is isosteric to (C) the  $G \cdot T$  wobble pair.

Bevilacqua 2003), analogous to the mechanism of certain protein enzymes such as RNase A (Thompson et al. 1994). The ideal  $pK_a$  for these residues is near biological pH, to provide a histidine-like property (Bevilacqua 2003). Although proton transfer is unlikely for Class I sites, these sites avail themselves to electrostatic catalysis through oxyanion hole formation. Free energy calculations on serine protease active sites indicate that the primary mechanism for rate acceleration is electrostatic stabilization in the transition state (Warshel et al. 1989). In a parallel behavior, a cationic Class I  $A^+ \cdot C$  mismatch was recently proposed to serve as an oxyanion hole in the active site of the ribosome (Bevilacqua et al. 2004). To be optimal, the  $pK_a$  of an oxyanion hole should be shifted 1–2 units above biological pH, to provide a lysine-like property, which would afford >90% of the molecules in the functional protonated state at neutrality. Such  $pK_a$ s would be especially valuable because the free nucleobases lack positive functionality at biological pH. One goal of the present study is to investigate the feasibility of such perturbations.

In addition to catalytic roles, nucleobases with shifted  $pK_a$  values may assist in the folding of certain RNA structures. Cytosine protonation drives pseudoknot formation via a triplet (Nixon and Giedroc 2000) and ribozyme folding via a quartet (Ferre-D'Amare and Doudna 2000); both of these systems have critical biological functions, programmed –1 mRNA ribosomal frameshifting and hepatitis D virus rolling circle replication, respectively.

There is a rich literature describing effects of ligand binding on protein stability (e.g., Acampora and Hermans 1967; Hermans and Acampora 1967; Tanford 1968, 1970; Schellman 1975; Wyman and Gill 1990; Freire 1998; Luque et al. 2002). Misra and Draper (2002) showed that binding of  $Mg^{2+}$  ions is linked to RNA folding. In the present study, we investigate the linkage between  $H^+$  binding and RNA and DNA folding, drawing on these earlier formalisms. We focus on Class I  $A^+ \cdot C$  mismatches (Fig. 1) since these are common and are amenable to simple design and experimentation. In these mismatches, the N1 of A is protonated and donates a hydrogen bond to the O2 of C, while the N6 of A donates a hydrogen bond to the N3 of C (Fig. 1B). As with the isosteric  $G \cdot U/T$  mismatch (Fig. 1C), the wobble

geometry of an  $A^+ \cdot C$  mismatch causes local fluctuations in B-form DNA, but does not affect global conformation (Hunter et al. 1986).

Herein we establish a thermodynamic framework for proton binding to RNA and DNA that involves linkage to folding. Our motivation for this work is twofold: (1) to provide a formalism for fitting pH-dependent denaturation data of RNA and DNA that properly identifies microscopic and macroscopic folded and unfolded state  $pK_a$  values;

and (2) to elucidate molecular and thermodynamic driving forces for  $pK_a$  shifting toward neutrality in nucleic acids.

## RESULTS

### Background: comparison of $H^+$ and $Mg^{2+}$ binding to nucleic acids

Our general approach for relating  $H^+$  binding to RNA folding parallels that for  $H^+$  binding to proteins (Tanford 1968, 1970; Schellman 1975; Freire 1998; Luque et al. 2002) and  $Mg^{2+}$  binding to RNA (Misra et al. 2003). In brief, folding equilibria between U and F, as well as  $U^+$  and  $F^+$ , are coupled to protonation equilibria involving  $U^+$  and  $F^+$ .<sup>1</sup> One major difference between  $Mg^{2+}$  and  $H^+$  binding to nucleic acids is that U and F generally contain specific sites for  $H^+$  but not for  $Mg^{2+}$ . Misra and Draper (1998) divide the modes of ion binding to RNA into diffuse and site binding, with site binding subdivided into outer or inner sphere depending on the degree of ion dehydration. Diffuse binding is nonspecific and cannot be described by mass action (Wyman and Gill 1990). Proton binding, on the other hand, is site-specific and obeys the law of mass action.

### Developing the thermodynamic framework: macroscopic and microscopic states

The equilibrium between the two macroscopic states, helix and coil, is described by an apparent equilibrium constant,  $K_u^{obs}$ .<sup>2</sup> This constant is given by the ratio of the ensemble of all ligand-free and ligand-bound<sup>3</sup>coil species, {U}, to all such helix species, {F},

<sup>1</sup>Table 1 provides a list of abbreviations, thermodynamic constants, and notation for equations.

<sup>2</sup>Here, we use the convention of following an unfolding reaction.

<sup>3</sup>“Ligand-free” refers to the bases in their standard, neutral forms. “Ligand-bound” refers to protonation of A and C at the imino nitrogen, which leads to cationic bases. Later, we also consider deprotonation of the imino protons of G and U/T, which leads to anionic bases. The imino nitrogens have  $pK_a$ s closest to neutrality and thus make the most significant contribution to the helix–coil equilibrium.

**TABLE 1.** Abbreviations, thermodynamic constants, and notation for equations

Variable	Equation(s) <sup>a</sup>	Definition <sup>b</sup>
$a_{H^+}$		Ligand activity, measured experimentally through pH. Defined as $pH = -\log a_{H^+}$
U, F		Unfolded and folded states in their standard neutral forms
{U}, {F}	1	{U} is the ensemble of all ligand-bound and ligand-free coil states; {F} represents all such helical species
$K_u^{obs}$	1	Observed equilibrium constant for the unfolding reaction. Defined as $K_u^{obs} = \{U\}/\{F\}$ . Dependent on pH
$n_u, n_f$	1	Number of proton-binding sites in the unfolded and folded states, respectively. Independent of pH
$U_a^+$	2	Same as $UH_a^+$ ; others follow this shorthand
$k_{a,w}, k_{a,f}$	4	Microscopic acid dissociation constant for a nucleobase in the unfolded and folded states, respectively. Independent of pH
$K_u^{ref}$	8	Equilibrium constant for the reference unfolding reaction. Defined as $K_u^{ref} = [U]/[F]$ . Independent of pH
$\Sigma_u, \Sigma_f$	8	Binding polynomials for the unfolded and folded states, respectively. Dependent on pH
$\Delta H_u^{ref}$	12	Enthalpy change for the reference unfolding reaction, U-to-F. Independent of pH
$T_M^{ref}$	12	Melting temperature for the reference unfolding reaction, U-to-F. Independent of pH
$T_M^{obs}$	13	Observed melting temperature. Defined as the temperature at which $K_u^{obs} = 1$ . Dependent on pH
$K_{a,u}, K_{a,f}$	Various figures	Macroscopic (observed) acid dissociation constant for an oligonucleotide in the unfolded and folded states, respectively. Independent of pH
$\Delta G_f^{Hsat}$		Folding free energy difference between unfolded and folded states in which the site with the shifted $pK_a$ is saturated with proton and no other protons are bound. Independent of pH
$\Delta G_f^{obs}$	14	Observed free energy change for folding. Defined as $\Delta G_f^{obs} = -RT \ln K_f^{obs}$ . Dependent on pH
$\Delta G_f^{ref}$	14	Free energy change for the reference folding reactions. Defined as $\Delta G_f^{ref} = -RT \ln K_f^{ref}$ . Independent of pH
$\Delta G_{H^+}^u, \Delta G_{H^+}^f$	16	Contribution of proton activity to observed free energy. Defined as $\Delta G_{H^+}^u = -RT \ln \Sigma_u$ ; $\Delta G_{H^+}^f = -RT \ln \Sigma_f$ . Dependent on pH
$\Delta \Delta G_{H^+}, \Delta \Delta G_f^{obs}$	16	Total free energy contribution for proton binding to folding. Dependent on pH
$\Delta G_x$	24	Intrinsic free energy of an $A^+ \cdot C$ mismatch. Independent of pH
$A_u, A_p$	25	Absorbance of unprotonated and protonated forms of an oligonucleotide, respectively. Independent of pH
$n$	25	Hill constant. Values >1 reflect multiple ionizations

<sup>a</sup>“Equation” denotes the first equation in which the term is found.

<sup>b</sup>For clarity, definitions indicate whether the value is dependent on pH.

$$\{F\} \xrightleftharpoons{K_u^{obs}} \{U\} \quad K_u^{obs} = \frac{\{U\}}{\{F\}} = \frac{\sum_{i=0}^{n_u} [U_i^{i+}]}{\sum_{j=0}^{n_f} [F_j^{j+}]} \quad (1)$$

where  $n_u$  and  $n_f$  are the numbers of proton-binding sites in the coil and helical states, respectively. This binding model is the basis for all simulations in this work. For the simple DNA oligonucleotides studied herein, the change in absorbance at 280 nm is dominated by the {F}-to-{U} transition ( $\Delta \epsilon \approx 25\%$ ), with processes such as the protonation of F accounting only for a small portion of the effect ( $\Delta \epsilon < 5\%$ , as seen in Fig. 7B, below). Thus,  $K_u^{obs}$  can be obtained experimentally with a two-state analysis of absorbance data.

It is convenient to parse  $K_u^{obs}$  into  $K_u^{ref}$  and  $\Sigma_u$  and  $\Sigma_f$  terms, where  $K_u^{ref}$  is a two-state equilibrium constant for unfolding with the bases in their standard neutral forms, and  $\Sigma_u$  and  $\Sigma_f$  are binding polynomials that enumerate all species for the unfolded and folded states (Schellman 1975; Graziano et al. 1999; Laing et al. 1994; Sharp and Englander 1994). The relationship among these four variables can be illustrated by considering a hypothetical case involving a single proton-binding site in the folded state,  $a$ , and two

proton-binding sites in the unfolded state,  $a$  and  $b$ . Here, we impose distinct proton affinities for site  $a$  in the folded and unfolded states. This model also posits that protonation at site  $b$  is incompatible with folding and corresponds to the limit of strong coupling. This is a reasonable situation since protonation would result in the loss of a base pair, and each base pair typically contributes at least a factor of 10 to the observed equilibrium constant in RNA and DNA folding (SantaLucia 1998; Mathews et al. 1999). According to equation 1 this gives

$$K_u^{obs} = \frac{[U] + [U_a^+] + [U_b^+] + [U_{ab}^{++}]}{[F] + [F_a^+]} \quad (2)$$

Using [U] and [F] as reference states for {U} and {F}, respectively,

$$K_u^{obs} = \frac{[U] \left( 1 + \frac{[U_a^+]}{[U]} + \frac{[U_b^+]}{[U]} + \frac{[U_{ab}^{++}]}{[U]} \right)}{[F] \left( 1 + \frac{[F_a^+]}{[F]} \right)} \quad (3)$$

It is first assumed that the  $a$  and  $b$  sites have identical proton affinities in the unfolded state and are independent

of each other (e.g., two distant adenines). Let  $k_{a,u}$  and  $k_{a,f}$  denote unfolded- and folded-state microscopic acid dissociation constants, respectively,

$$k_{a,u} = \frac{[U][H^+]}{[U_a^+]} = \frac{[U][H^+]}{[U_b^+]} = \frac{[U_a^+][H^+]}{[U_{ab}^{++}]} = \frac{[U_b^+][H^+]}{[U_{ab}^{++}]}, \quad (4)$$

$$k_{a,f} = \frac{[F][H^+]}{[F_a^+]}. \quad (5)$$

By substitution

$$K_u^{\text{obs}} = \frac{[U](1 + [H^+]/k_{a,u})^2}{[F](1 + [H^+]/k_{a,f})} \quad (6)$$

In general, for  $n_u$  identical and independent sites in U and  $n_f$  identical and independent sites in F,  $K_u^{\text{obs}}$  is

$$K_u^{\text{obs}} = \frac{[U](1 + 10^{\text{p}K_{a,u} - \text{pH}})^{n_u}}{[F](1 + 10^{\text{p}K_{a,f} - \text{pH}})^{n_f}} \quad (7)$$

This expression gives the position of the observed equilibrium as a function of  $a_{\text{H}^+}$ , ligand activity, which is related to pH through  $\text{pH} = -\log a_{\text{H}^+}$  (Westcott 1978). By substitution

$$K_u^{\text{obs}} = K_u^{\text{ref}} \frac{\sum_u}{\sum_f} \quad (8)$$

where  $K_u^{\text{ref}}$  is the equilibrium constant  $[U]/[F]$  for the reference reaction, and the influence of proton binding on stability is described with the ratio of the binding polynomials,  $\sum_u$  and  $\sum_f$ . The binding polynomials are functions of pH and temperature and are given by

$$\sum_u = (1 + 10^{\text{p}K_{a,u} - \text{pH}})^{n_u}, \quad (9)$$

$$\sum_f = (1 + 10^{\text{p}K_{a,f} - \text{pH}})^{n_f}. \quad (10)$$

Before proceeding, we inspect the validity of assuming independent sites, that is, identical  $k_{a,u}$  values for all sites of a particular type (e.g., adenine). In general, the strength of the interaction ( $\Delta G_1$ ) between two functionally identical sites (e.g., two primary amines or two carboxylic acids) depends on the distance  $d$  between the ionizable groups. At the limit of large  $d$ ,  $\Delta G_1$  tends to zero, and the difference in the macroscopic  $\text{p}K_a$  ( $\Delta\text{p}K_a = \text{p}K_{a,1} - \text{p}K_{a,2}$ ) reaches the statistical value of  $\log_{10} 4 = 0.602$  (Jencks and Regenstein 1970; Albert and Serjeant 1984). As  $d$  decreases,  $\Delta\text{p}K_a$  deviates from this limit, and the difference ( $\Delta\text{p}K_a - 0.602$ ) corresponds to an interaction energy  $\Delta G_1$  of 1.42 ( $\Delta\text{p}K_a - 0.602$ ) expressed in kilocalories per mole and calculated at 37°C. The distance between two adjacent bases in a random coil state for RNA was estimated by considering model systems. In the leadzyme, the distance between the N3 of C11 and N1 of A12 in a helical region is 5.15 Å, whereas the distance between the N1s of A16 and A17 in a loop region is 3.99 Å (Hoogstraten et al. 1998). In an un-

folded state, the distance between protonated bases should be even greater owing to increased conformational freedom and electrostatic repulsion. Symmetric diamines can be invoked as a relevant parallel to the nucleobases since they become dications upon protonation. In 1,3-diaminopropane the  $\Delta\text{p}K_a$  is 1.67 in water at 25°C (Albert and Serjeant 1984), which gives a  $\Delta G_1$  of 1.5 kcal/mol at 37°C. The distance between the protonated heteroatoms in fully extended 1,3-diaminopropane is 4.93 Å. Based on comparison,  $\Delta G_1$  for adjacent protonations in the random coil state of a nucleic acid is predicted to be  $\leq 1.5$  kcal/mol, and this value should decrease by  $\approx 0.4$  kcal/mol at 100 mM ionic strength (Martel and Smith 1974). As such, we ignore the small amount of anticooperativity between adjacent sites; however, negative cooperativity could be treated on a case-by-case basis by explicitly using the  $\text{p}K_a$ s in the appropriate binding polynomial.

### Developing the thermodynamic framework: dependence of $T_M$ on pH

In order to relate proton binding to folding, the effect of pH on the helix-coil equilibrium constant needs to be determined. The melting temperature,  $T_M$ , is a convenient and robust experimental parameter. Its dependencies on ionic strength (Record et al. 1976; Laing et al. 1994) and pH (Nixon and Giedroc 2000) are routinely used as characterization tools for RNA and DNA sequences. Although we prefer to treat the effect of pH on  $\Delta G$  for reasons explained below, a brief description of the dependence of  $T_M$  on pH is provided here. This description includes practical equations for determining  $\text{p}K_{a,f}$  from such plots if desired.

We begin by taking the natural log of equation 8 in order to separate terms:

$$\ln K_u^{\text{obs}} = \ln K_u^{\text{ref}} + \ln(\sum_u/\sum_f). \quad (11)$$

The melting temperature observed in a thermal denaturation experiment,  $T_M^{\text{obs}}$ , at a given pH is defined as the temperature at which  $\{U\} = \{F\}$  and, therefore,  $K_u^{\text{obs}} = 1$ ; this condition corresponds to  $\Delta G_u^{\text{obs}} = 0$  and denotes constant observed free energy conditions. The temperature dependence of  $K_u^{\text{ref}}$  is obtained from the van't Hoff relationship:

$$\ln K_u^{\text{ref}}(T) = \frac{\Delta H_u^{\text{ref}}}{R} \left( \frac{1}{T_M^{\text{ref}}} - \frac{1}{T} \right) \quad (12)$$

where  $T_M^{\text{ref}}$  and  $\Delta H_u^{\text{ref}}$  are defined for the ligand-free F-to-U unfolding reaction, and  $R$  is the gas constant. Equation 12 assumes that  $\Delta H_u^{\text{ref}}$  is independent of temperature, that is,  $\Delta C_p^{\text{ref}} = 0$ . Substituting equation 12 into 11 at  $T = T_M^{\text{obs}}$  gives the dependence of  $T_M^{\text{obs}}$  on pH:

$$\frac{1}{T_M^{\text{obs}}} = \frac{1}{T_M^{\text{ref}}} + \frac{R}{\Delta H_u^{\text{ref}}} \ln \left( \frac{\sum_u|_{T=T_M^{\text{obs}}}}{\sum_f|_{T=T_M^{\text{obs}}}} \right) \quad (13)$$

where  $\Sigma_u$  and  $\Sigma_f$  are evaluated at  $T_M^{\text{obs}}$  of a particular thermal denaturation experiment. Thus, a  $T_M$  analysis necessitates accounting for the temperature dependence of the binding polynomials, which potentially complicates application of equation 13 to curve fitting. For  $\Sigma_\beta$  proton binding will be associated with completion of secondary structural folding within {F}, which has a large negative enthalpy, whereas for  $\Sigma_u$ , the small enthalpy change associated with proton binding to the bases, which is manifested in the temperature dependence of the microscopic  $pK_a$ s (Izatt et al. 1971), may be amplified by the large number of such protonations. In addition, the assumption that  $\Delta C_p^{\text{ref}} = 0$  can be problematic for certain tertiary and secondary structures (Takach et al. 2004).

### Developing the thermodynamic framework: dependence of $\Delta G^{\text{obs}}$ on pH

As an alternative to  $T_M$  analysis, the dependence of observed free energy,  $\Delta G^{\text{obs}}$ , on pH can be analyzed at constant temperature. An expression for the observed folding free energy,  $\Delta G_f^{\text{obs}}$ , can be defined from equation 11:

$$\Delta G_f^{\text{obs}} = \Delta G_f^{\text{ref}} - RT \ln(\Sigma_f / \Sigma_u), \quad (14)$$

where  $\Delta G_f^{\text{ref}}$  is defined for the ligand-free U-to-F folding reaction.<sup>4</sup> Moreover, the net influence of proton binding on experimental folding free energy can be defined as

$$\Delta \Delta G_f^{\text{obs}} = -RT \ln(\Sigma_f / \Sigma_u) \quad (15)$$

where  $\Delta \Delta G_f^{\text{obs}} = \Delta G_f^{\text{obs}} - \Delta G_f^{\text{ref}}$ . In the next section, equation 15 and the appropriate binding polynomials are used to simulate the dependence of  $\Delta \Delta G_f^{\text{obs}}$  on pH.

The influence of proton binding on observed free energy can be divided into effects on the folded state,  $\Delta G_H^f = -RT \ln \Sigma_\beta$  and unfolded state,  $\Delta G_H^u = -RT \ln \Sigma_u$  (Schellman 1975; Bukhman and Draper 1997; Misra et al. 2003). This leads to the following expression relating  $\Delta \Delta G_f^{\text{obs}}$  and  $\Delta \Delta G_H$ :

$$\Delta \Delta G_f^{\text{obs}} = \Delta \Delta G_H \quad (16)$$

where  $\Delta \Delta G_H = \Delta G_H^f - \Delta G_H^u$ . This simply expresses that to determine the influence of folding on proton binding, which is ultimately reflected in shifted  $pK_a$  values, is to determine the influence of proton binding on experimental folding free energy.

The ligand and temperature dependencies of  $\Delta G_f^{\text{obs}}$  can be defined by taking the differential of  $\Delta G_f^{\text{obs}}$ :

$$d(\Delta G_f^{\text{obs}}) = \left[ \frac{\partial \Delta G_f^{\text{ref}}}{\partial(1/T)} + RT^2 \ln \frac{\Sigma_f}{\Sigma_u} \right]$$

$$\begin{aligned} & - RT \left( \frac{\partial \ln \Sigma_f}{\partial(1/T)} - \frac{\partial \ln \Sigma_u}{\partial(1/T)} \right) d(1/T) \\ & - RT \left( \frac{\partial \ln \Sigma_f}{\partial \ln [H^+]} - \frac{\partial \ln \Sigma_u}{\partial \ln [H^+]} \right) d \ln [H^+] \quad (17) \end{aligned}$$

If  $T$  is held constant, the ligand dependence of  $\Delta G_f^{\text{obs}}$  can be isolated:

$$\frac{d\Delta G_f^{\text{obs}}}{d \ln [H^+]} = -RT \left( \frac{\partial \ln \Sigma_f}{\partial \ln [H^+]} - \frac{\partial \ln \Sigma_u}{\partial \ln [H^+]} \right). \quad (18)$$

However, since  $\Delta \Delta G_f^{\text{obs}} = \Delta G_f^{\text{obs}} - \Delta G_f^{\text{ref}}$  and  $\Delta G_f^{\text{ref}}$  is not a function of pH,

$$\frac{d\Delta \Delta G_f^{\text{obs}}}{d \text{pH}} = -RT \left( \frac{\partial \ln \Sigma_f}{\partial \text{pH}} - \frac{\partial \ln \Sigma_u}{\partial \text{pH}} \right). \quad (19)$$

For the case of independent sites described above, the derivatives of equations 10 and 9 are

$$\frac{\partial \ln \Sigma_f}{\partial \text{pH}} = \frac{-2.303 n_f [H^+] / k_{a,f}}{1 + [H^+] / k_{a,f}}, \quad (20)$$

$$\frac{\partial \ln \Sigma_u}{\partial \text{pH}} = \frac{-2.303 n_u [H^+] / k_{a,u}}{1 + [H^+] / k_{a,u}}. \quad (21)$$

For each simulated binding model presented below, analytical expressions for  $\partial \ln \Sigma_i / \partial \text{pH}$  are provided, which upon substitution into equation 19 afford the slope of  $\Delta \Delta G_f^{\text{obs}}$  versus pH plots. It is of interest to note that over pH values where  $k_{a,f} < [H^+] < k_{a,u}$ , the slope of a  $\Delta \Delta G_f^{\text{obs}}$  versus pH plot approaches a constant whose value reflects the number of proton-binding sites in the folded state,  $n_f$ . Somewhat counterintuitively, saturation of the folded-state binding sites does not preclude further stabilization upon further lowering of the pH in this regime. As described below, the folded-state  $pK_a$  corresponds to the pH of the inflection point in a first derivative plot, but not in the free energy versus pH plot.

The partial derivatives of equation 19 are related to the proton-binding densities  $\nu_f$  and  $\nu_u$  (Wyman 1965; Schellman 1975):

$$\frac{d\Delta \Delta G_f^{\text{obs}}}{d \text{pH}} = +2.303 RT (\nu_f - \nu_u) \quad (22)$$

When the pH is low enough to saturate the proton-binding sites in both F and U, the slope of a  $\Delta \Delta G_{f,37}^{\text{obs}}$  versus pH plot approaches a constant whose value reflects the difference in the number of proton-binding sites,  $\Delta n$ , between the two states:

$$\frac{d\Delta \Delta G_{f,37}^{\text{obs}}}{d \text{pH}} = 1.42 \Delta n \quad (23)$$

<sup>4</sup>This is the convention for describing the stability of oligonucleotides.

### Developing the thermodynamic framework: microscopic $pK_a$ s for the unfolded state

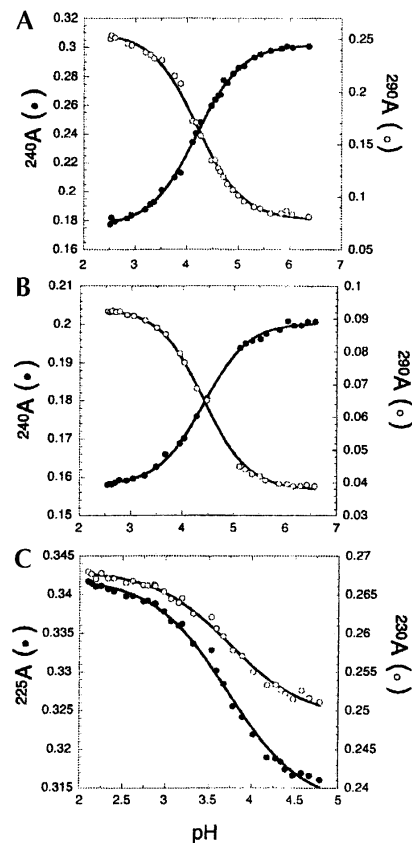
Binding polynomials for the unfolded state require the corresponding microscopic  $pK_a$  values as per equation 9. Values of  $pK_a$  for the bases free in solution are well known; however, we wanted to determine whether the anionic phosphate backbone of a longer, unstructured oligonucleotide influences  $pK_{a,u}$ . Three RNA oligonucleotides, r(5'-UC), r(5'-UUCUU), and r(5'-UUAUU), were chosen for determining the influence of the phosphate backbone.<sup>5</sup> These oligonucleotides were designed to have only one base that ionizes between  $\approx$ pH 2.5 and 6.5, are not self-complementary, and should be unstructured. Uracil was chosen as the nontitrating base since it is a poor stacker (Richards et al. 1963; Turner and Bevilacqua 1993); in contrast, guanine self-associates (Kim et al. 1991) and has an N7  $pK_a$  that would interfere with the titration (Izatt et al. 1971; Saenger 1984).

The microscopic  $pK_a$ s of the three oligonucleotides were determined by pH titrations monitored by UV absorbance (Fig. 2). An ionic strength of 100 mM KCl was chosen, similar to titrations and melts on folded oligonucleotides. Each titration was carried out at two wavelengths where  $\Delta\epsilon$  was large, and good fits to a Henderson-Hasselbalch equation were obtained. Observation of similar  $pK_a$  and Hill coefficient ( $n$ ) values (see equation 25 in Materials and Methods) at both wavelengths was consistent with two-state processes. In addition,  $n$  values near unity were found as expected (Saenger 1984). Titration of the dinucleotide, r(UC), gave a  $pK_a$  of 4.20 (Fig. 2A), similar to the literature value for cytidine (Saenger 1984). A base in an RNA oligonucleotide is embedded in negative potential, which might lead to a favorable interaction upon base protonation and thereby raise the  $pK_a$ . To test this possibility, titrations were carried out with a pentamer, r(UUCUU). Indeed, titration of r(UUCUU) led to a  $pK_a$  of 4.41 (Fig. 2B),  $\approx$ 0.2 units higher than the dinucleotide. Likewise, titration of r(UUAUU) gave a  $pK_a$  of 3.70 (Fig. 2C),  $\approx$ 0.2 higher than the typical value for adenosine (Izatt et al. 1971). On the basis of these experiments, we used  $pK_{a,u}$  values of 3.7 for A, 4.4 for C, and 9.4 for G and U in the simulations; the latter two values are also  $\approx$ 0.2 units higher than the typical values for the nucleosides (Izatt et al. 1971) and are based on the repulsion expected between the negative charge of the backbone and the deprotonated base.

### Developing the thermodynamic framework: simulations of $\Delta\Delta G^{\text{obs}}$ versus pH in the simplest cases

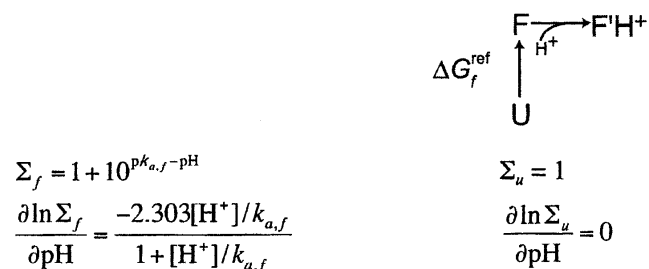
Next, we present a series of models that link proton binding and folding. For each case, the associated binding poly-

<sup>5</sup>RNA oligonucleotides were chosen for unfolded-state studies. Since we are concerned with the effect of the phosphate backbone, the general trends should be applicable to DNA. Also, removal of the 2'-hydroxyl has no observable effect on the  $pK_a$  values of adenosine and cytidine within experimental error (Izatt et al. 1971).

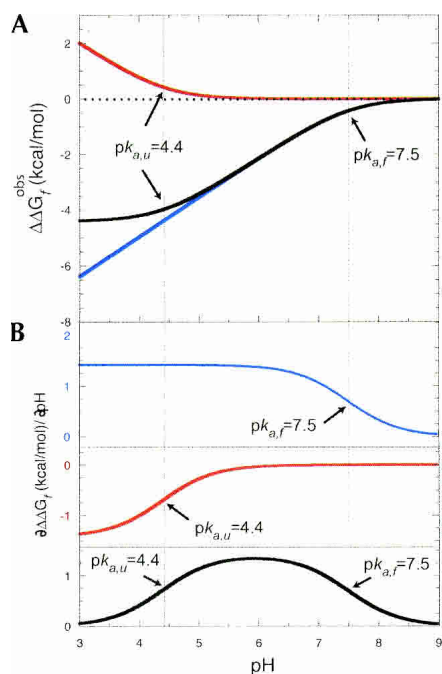


**FIGURE 2.** Determination of unfolded-state  $pK_a$  values. Here 100 mM KCl was added to all samples prior to the titrations, and data were fit to equation 25. (A) Titration of r(UC) with KOH monitored at 240 nm (●) and 290 nm (○), which provided an average  $pK_{a,u}$  of  $4.20 \pm 0.01$  and  $n$  of 1.08. (B) Titration of r(UUCUU) with KOH monitored at 240 nm (●) and 290 nm (○), which provided an average  $pK_{a,u}$  of  $4.41 \pm 0.02$  and  $n$  of 1.01. (C) Titration of r(UUAUU) with HCl monitored at 225 nm (●) and 230 nm (○), which provide an average  $pK_a$  of  $3.70 \pm 0.04$  and  $n$  of 1.20.

mials and their partial derivatives are provided. All models have only two macroscopic states, {U} and {F}, but variable numbers of proton-bound or proton-depleted microscopic states. Scheme 1 depicts the simplest model having linkage between proton binding and folding. This model involves a single protonation of the folded state (e.g., a site is created



**SCHEME 1.** Linkage among unfolded, folded, and folded protonated states.



**FIGURE 3.** Contributions of proton activity to stability for binding of 1 proton to F, U, or both. (A) Simulation of  $\Delta\Delta G_f^{\text{obs}}$  versus pH comparing the models in Schemes 1 (blue), 2 (red), and 3 (black). Simulations are according to equation 15, binding polynomials from Schemes 1–3, and  $pK_{a,f}$  and  $pK_{a,u}$  of 7.5 and 4.4, respectively. The black line is the summation of  $\Delta G_H^f$  (blue) and  $-\Delta G_H^u$  (red). (B) Plot of first derivative of data in panel A versus pH. Panels share the same X-axis to facilitate comparison of flex points (A) and inflection points (B).  $pK_a$  values are noted in the figure.

in F that disappears in U) and has three microscopic states.  $\Sigma_f$  is given by equation 10;  $\Sigma_u$ , which is given by equation 9, has a value of unity in this case. Scheme 1 is useful since it illustrates one of the limiting behaviors of the binding model. Figure 3 illustrates the dependence on pH of  $\Delta\Delta G_f^{\text{obs}}$  and its first derivative with respect to pH. As shown in Figure 3A (blue), proton binding begins to contribute noticeably to free energy near the input  $pK_a$  value of 7.5.<sup>6</sup> In addition, the proton activity contribution to free energy does not level off at pH values below the  $pK_a$ ; instead, lower pH favors folding indefinitely, with the  $pK_a$  found graphically near the flex point<sup>7</sup> of the  $\Delta\Delta G_f^{\text{obs}}$  versus pH curve. The  $pK_a$  can also be seen as the inflection point in the  $\partial\Delta\Delta G_f^{\text{obs}}/\partial\text{pH}$  versus pH plot (Fig. 3B, blue), as expected from equations 20 and 21 and recalling that pH is a logarithmic axis. Behavior for this system parallels that for site binding of  $\text{Mg}^{2+}$  to RNA (Laing et al. 1994; Nakano et al. 2003). The value of the first derivative at low pH is +1.42, as expected from equation 23.

<sup>6</sup>This value was chosen because  $pK_a$ s near neutrality have been observed experimentally for oligonucleotides.

<sup>7</sup>The “flex point” is the region between the two linear pieces of the plot; it is defined more precisely in the first derivative plot, where it corresponds to an inflection point.

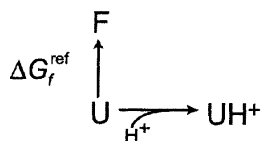
Scheme 2 depicts a model that involves a single protonation of the unfolded state. Values of  $\Sigma_f$  and  $\Sigma_u$  are exchanged from Scheme 1. In this case protonating a base with a low  $pK_a$  value, for example, the base-pairing face of A or C, disfavors folding since standard hydrogen bonding is no longer possible. As shown in Figure 3A (red), proton binding begins to contribute noticeably to free energy near the input  $pK_a$  value of 4.4. The other properties of the  $\Delta\Delta G_f^{\text{obs}}$  function and its first derivative mirror those described above, except that proton binding disfavors folding (Fig. 3B).

Scheme 3 depicts a model involving single protonations of both the folded and unfolded states, and has four microscopic states; this model begins to approach a realistic scenario for simple RNA and DNA oligonucleotides. Note that  $\Sigma_f$  and  $\Sigma_u$  are the same as in Schemes 1 and 2, respectively. The resulting curve has two level regions in the pH extremes and has a positive slope between the  $pK_a$  values of 4.4 for the unfolded state and 7.5 for the folded state (Fig. 3A, black). Although for certain  $pK_a$  values the  $\Delta\Delta G_f^{\text{obs}}$  versus pH plot may appear as a standard binding curve having a single  $pK_a$  near the inflection point, this appearance is misleading. In some instances, it is possible to fit the data adequately with a binding equation (Nixon and Giedroc 2000; Nixon et al. 2002); however, the fitted  $pK_a$  so obtained would correspond to the midpoint of the curve and thus underestimate the true  $pK_{a,f}$  at the flex point. For example, underestimates of 0.3 and 1.3 units were found for input  $pK_{a,f}$  values of 6.0 and 8.5, respectively (Moody 2004). In other words, the  $\Delta\Delta G_f^{\text{obs}}$  versus pH plot should not be confused with a site occupancy plot. The latter identifies clearly the pH at which the populations of protonated and unprotonated species are equal, whereas the former does not.

The free energy contribution of proton activity to folding levels off at low pH values; this is due to compensating effects of proton binding to both F and U (Fig. 3A). The first derivative of  $\Delta\Delta G_f^{\text{obs}}$  with respect to pH shows two inflection points, which give the two  $pK_a$  values (Fig. 3B). This is obtained from equation 22, which illustrates that  $d\Delta\Delta G_f^{\text{obs}}/d\text{pH}$  is related to site occupancy, and therefore represents a standard binding curve (Wyman and Gill 1990). The derivative approaches a value of +1.4 kcal/mol at pH values intermediate between the two  $pK_a$ s (Fig. 3B, black), but has a final value of 0 at very low pH values, as expected from equation 23.

For Scheme 3A, the thermodynamic box is completed with  $\Delta G_f^{\text{Hsat}}$ , which is the free energy for folding from ligand-saturated U to ligand-saturated F. It is useful to break this free energy term into the free energy for the ligand-free U-to-F folding reaction,  $\Delta G_f^{\text{ref}}$ , and a bonus free energy for the additional interactions possible upon proton saturation in U,  $\Delta G_x$ .<sup>8</sup> Schemes 3B and 3C provide this alternative

<sup>8</sup> $\Delta G_x$  is also referred to as the “intrinsic energy” of the  $A^+ \cdot C$  base pair



$$\begin{aligned}
 \Sigma_f &= 1 \\
 \frac{\partial \ln \Sigma_f}{\partial \text{pH}} &= 0
 \end{aligned}$$

$$\begin{aligned}
 \Sigma_u &= 1 + 10^{\text{p}K_{a,u} - \text{pH}} \\
 \frac{\partial \ln \Sigma_u}{\partial \text{pH}} &= \frac{-2.303[\text{H}^+]/k_{a,u}}{1 + [\text{H}^+]/k_{a,u}}
 \end{aligned}$$

**SCHEME 2.** Linkage among unfolded, unfolded protonated, and folded states.

view. In Scheme 3B, the upper edge of Scheme 3A is broken into two coupled steps. The first involves protonation without additional interactions and is denoted by making  $\text{FH}^+$  with unperturbed  $\text{p}K_{a,u}$ ; the second accounts for the formation of the additional interactions and is denoted  $\text{F}'\text{H}^+$  with a free energy  $\Delta G_x$ . Equating the upper edges of Schemes 3A and 3B leads to an expression for the origin of  $\text{p}K_a$  shifting in terms of RNA folding:

$$\text{p}K_{a,f,37} = \text{p}K_{a,u,37} - \Delta G_{x,37}/1.42 \quad (24)$$

where  $\Delta G_{x,37}$  is in units of kilocalories per mole. Because species  $\text{FH}^+$  is the product of an unfavorable reaction (protonation) and the reactant for a favorable reaction (formation of a wobble pair), it should not be populated significantly at any pH, making Scheme 3A a good approximation of Schemes 3B and 3C. Scheme 3C provides a molecular picture of this process, in which the additional interactions are in the form of stacking, hydrogen bonding, and enhanced electrostatics due to higher electrostatic potential. The relationship between interactions and  $\Delta G_x$  is considered in the Discussion.

### Developing the thermodynamic framework: simulations of $\Delta\Delta G^{\text{obs}}$ versus pH in practical cases

The models in Scheme 3 are oversimplified for the helix-coil process<sup>9</sup> because when a typical oligonucleotide is unfolded, a proton has access to all As and Cs in the sequence. Thus, it is necessary to consider multiple protonations in the unfolded state. Scheme 4 depicts such a model, with a single protonation in the folded state and multiple protonations in the unfolded state. The folded-state binding polynomial remains unchanged from Schemes 1 and 3; the new unfolded-state binding polynomial, however, contains terms for protonation of all combinations of As and Cs.

based on Jencks's use of the term in describing the contribution of an interaction to stability in the absence of associated energetic penalties (Jencks 1975).

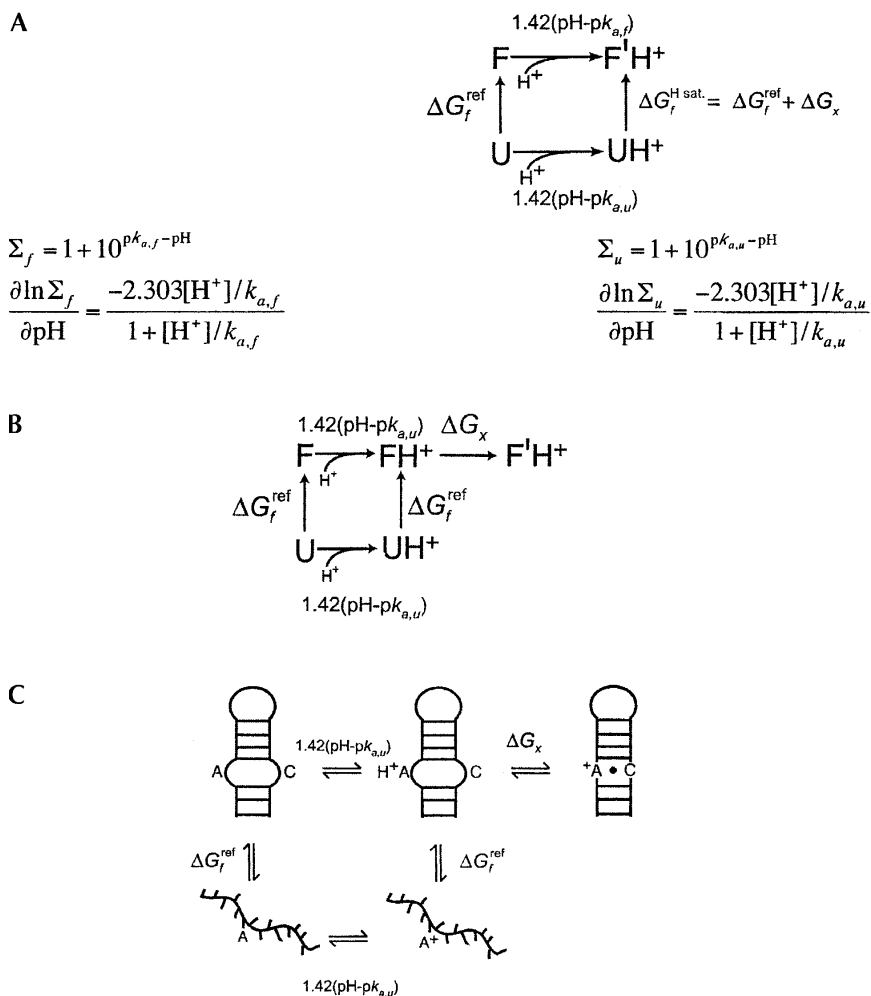
<sup>9</sup>Scheme 3 may, however, be a good model for processes in which the reference state is not completely unfolded, and consequently only a single reference-state protonation is possible.

Note that only ionizations involving imino nitrogen are considered (see footnote 3), and  $\Sigma_u$  assumes an absence of interaction between protonated sites, as described above. Terms involving deprotonation of G and U imino protons are considered below.

The plot in Figure 4A shows the influences of proton binding on the free energies of individual macroscopic states ( $\Delta G_H^f$  and  $-\Delta G_H^u$ ) and their sum ( $\Delta\Delta G_f^{\text{obs}}$ ). For this illustration, we used a model oligonucleotide with four Cs and two As, which mimics a hairpin with six Watson-Crick base pairs in its stem. The resulting curve has a level region at high pH, a positive slope from  $\sim\text{pH}$  5 to 7.5, and a negative slope at pH values below 5. The origin of this behavior can be understood in terms of the component curves,  $\Delta G_H^f$  (blue) and  $-\Delta G_H^u$  (red) (Fig. 4A). Between pH 7.5 and  $\approx 5.2$ , proton binding to F dominates the summation curve. Below pH 5, however, the presence of the many proton-binding sites in U gives rise to a  $-\Delta G_H^u$  curve that has a steep dependence on pH and a macroscopic  $\text{p}K_{a,u}$  shifted toward neutrality (see below); the  $-\Delta G_H^u$  curve dominates the overall dependence, which leads to the negative slope at low pH values, characteristic of the acid denaturation behavior of oligonucleotides. Overall, pH both stabilizes and destabilizes folding depending on the pH values (Fig. 4A). As described below, this is consistent with experimental observations.

Similar to Scheme 3, the first derivative plot for Scheme 4 displays two apparent inflection points, or macroscopic  $\text{p}K_a$ s (Figs. 3B, 4B). As with Scheme 3, the higher macroscopic  $\text{p}K_a$  corresponds to the microscopic  $\text{p}K_a$  for F. However, the lower macroscopic  $\text{p}K_a$  does not correspond to a single  $\text{p}K_a$ , but instead reflects behavior for the ensemble of microscopic unfolded states. Because this macroscopic  $\text{p}K_a$  does not correspond to any individual protonation, its effect on any functional properties of the molecule will be case-dependent. For instance, if functional properties of F, such as structure or catalysis, are being observed, then  $\Delta G_f^{\text{obs}}$  values above  $\approx -1.5$  kcal/mol (corresponding to more than  $\approx 10\%$  {U}) should be observable; however, the  $\Delta\Delta G_f^{\text{obs}}$ , and therefore pH, needed to promote such denaturation will be dependent on the value of  $\Delta G_f^{\text{ref}}$  for the system of interest. If  $\Delta G_f^{\text{ref}}$  is large and negative, then a lower pH will be needed to promote observable denaturation than if  $\Delta G_f^{\text{ref}}$  is more modest. The ability of multiple protonations to provide an apparent  $\text{p}K_a$  shifted toward neutrality was reported by Knitt and Herschlag (1996). They found that deprotonation of any of 19 G and U residues with microscopic  $\text{p}K_a$ s of 9.4 interfered with catalysis, which gave rise to an apparent  $\text{p}K_a$  of 7.6. Lastly, it can be noted that the value of the derivative approaches  $-7.1$  kcal/mol at very low pH values (Fig. 4B), as expected from equation 23 and  $\Delta n = -5$ .

Next, we extend Scheme 4 and consider several cases: behavior when F does not have an ionizable group (i.e.,  $\Sigma_f = 1$ ), the effect of oligonucleotide length on proton-de-



**SCHEME 3.** (A) Linkage among unfolded, unfolded protonated, folded, and folded protonated states. (B) Expansion of Scheme 3A showing five microscopic states: unfolded, unfolded protonated, folded, folded protonated without  $\text{A}^+ \cdot \text{C}$  mismatch formed and folded protonated with  $\text{A}^+ \cdot \text{C}$  mismatch formed. (C) Possible base-pairing status of the five microscopic states shown in Scheme 3B.

pendent destabilization, and the effect of deprotonation of G and U residues at higher pH. Using a  $\Sigma_f$  of unity and a  $\Sigma_u$  that accounts for protonations of As and Cs but not deprotonation of Gs and Us, gives rise to  $\Delta\Delta G_f^{\text{obs}}$  curves equivalent to  $\Delta G_H^u$  (Fig. 5A). As oligonucleotide length increases, the onset of destabilization occurs at pH values closer to neutrality, which could be reflected in apparent  $pK_a$  values closer to neutrality. However, experimental realization of this effect, in terms of  $\Delta G_f^{\text{obs}}$ , depends on the associated  $\Delta G_f^{\text{ref}}$ , which typically increases with oligonucleotide length as well. Of course, this model will be oversimplified in some cases since it assumes that any individual protonation prevents folding completely; nevertheless, it illustrates the trend that increased length can give rise to apparent  $pK_a$  values more shifted toward neutrality. The effect of oligonucleotide length is also illustrated for a case with a single protonation in F (Fig. 5B), where it can be seen that the onset of destabilization has the potential to distort the

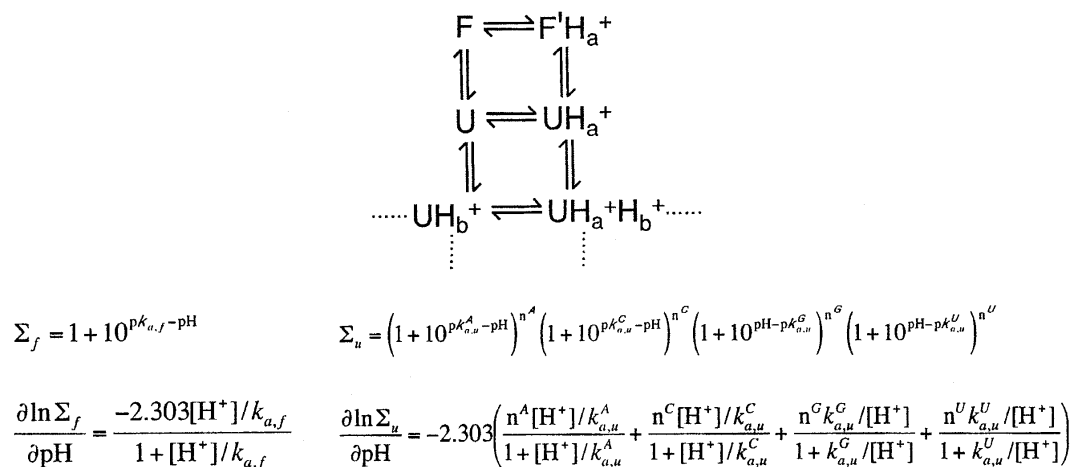
$\Delta\Delta G_f^{\text{obs}}$  versus pH profile near  $pK_{a,f}$  and thus interfere with determination of its value.

Introduction of terms for deprotonation of Gs and Us was considered next (Scheme 4;  $\Sigma_u$ ), and gave rise to bell-shaped  $\Delta\Delta G_f^{\text{obs}}$  versus pH curves, with proton-dependent destabilization observed in both the low and high pH regimes (Fig. 5C). This behavior corresponds to the well-known acid and alkaline denaturation properties of simple oligonucleotides. Lastly, a series of plots allowing a single protonation in F with all protonations and deprotonations in U is provided in Figure 5D. This curve shows the diverse and complex nature of pH effects on apparent stability.

### Applying the thermodynamic framework: experimental determination of folded state $pK_a$ s

Next, we applied the thermodynamic framework to experimental data on a DNA hairpin with an  $\text{A}^+ \cdot \text{C}$  mismatch in its stem, DNA(AC) (Fig. 6A). A DNA hairpin was chosen because it should be relatively resistant to degradation during thermal melting and titration, and the principles discerned, which are based on stacking and hydrogen bonding, should be extendable to RNA. In addition, a  $pK_a$  for DNA(AC) was recently measured by NMR (Moody et al. 2004a). Through melting experiments,  $\Delta G_{f,37}^{\text{obs}}$  was determined as a function of pH, and the data

were fit to equation 14 using binding polynomials derived from Scheme 4. Melts were conducted at a series of pH values, yielding the representative profiles provided in Figure 6B. Absorbance–temperature plots were fit to a standard two-state model using the two-macroscopic-state equilibrium constant defined in equation 1. A plot of  $\Delta G_{f,37}^{\text{obs}}$  versus pH (4.5–8.5) shows the same general shape as observed for the simulations shown in Figures 4A and 5B.  $\Delta G_{f,37}^{\text{obs}}$  is relatively independent of pH from 7 to 8.5, shows a positive slope from pH  $\approx 5.2$  to 7, and has a negative slope at pH values below 5.2. Similar experiments for DNA(GC), in which the  $\text{A}^+ \cdot \text{C}$  mismatch was replaced with a GC base pair, revealed only the acid-denaturation behavior characteristic of the simulations in Figure 5A, with a  $\Delta G_{f,37}^{\text{ref}}$  of  $-8.0$  kcal/mol (data not shown). The plot for DNA(AC) of  $\Delta G_{f,37}^{\text{obs}}$  versus pH was fit to equation 14 using a  $\Sigma_f$  with one protonation and a  $\Sigma_u$  with three As and eight Cs (Scheme 4), because residues in the stable loop of DNA(AC) are ex-



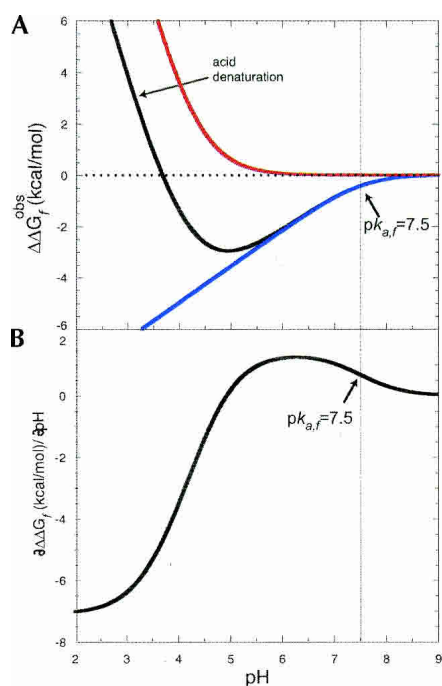
**SCHEME 4.** Linkage among a full ensemble of microscopic states: unfolded, all unfolded protonated, all unfolded deprotonated, folded, and folded protonated states.

pected to be important contributors to stability (Fig. 6A; Hirao et al. 1994). Deprotonations of Gs and Us were not included in  $\Sigma_u$  because their effect only becomes important above  $pH \approx 8.5$  for shorter oligonucleotides (Fig. 5C,D). The fit therefore had only two adjustable parameters,  $\Delta G_{f,37}^{ref}$  and

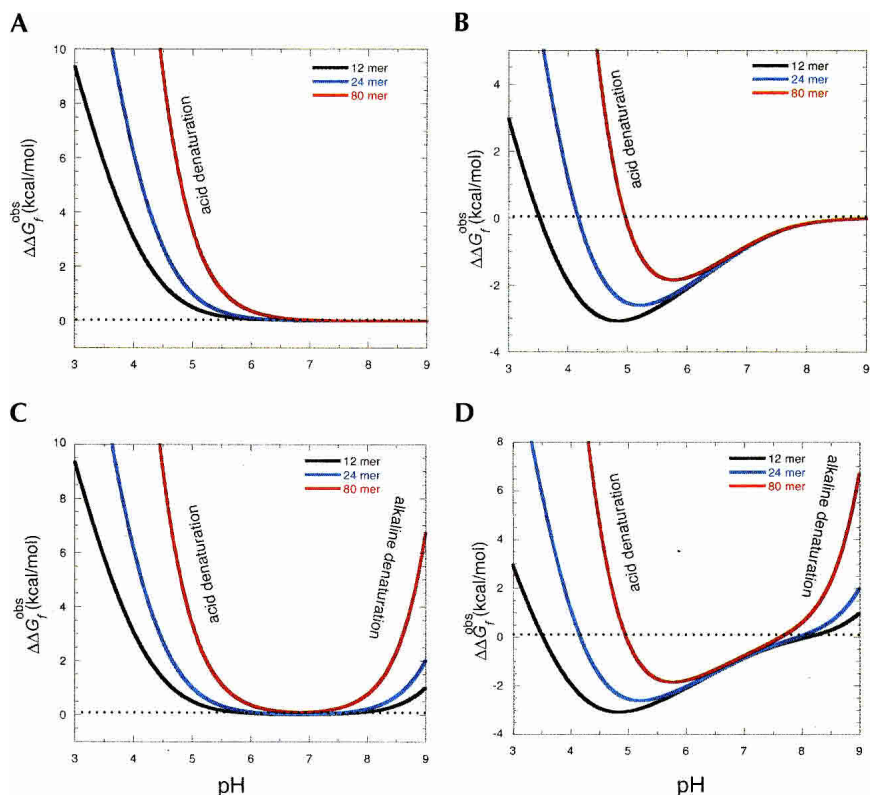
$pk_{a,f}$ , and resulted in values of  $-3.6 \pm 0.2$  kcal/mol and  $6.7 \pm 0.1$ , respectively (Fig. 6C). Besides providing a qualitatively good fit to the shape of the  $\Delta G_{f,37}^{obs}$  data versus pH, the accuracy of the fit can be assessed by comparison of the  $\Delta G_{f,37}^{ref}$  and  $pk_{a,f}$  parameters to expected or independently determined values.

The values of  $\Delta G_{f,37}^{ref}$  for DNA(AC) and DNA(GC) can be estimated by free energy minimization using mfold v3.1 (Zuker 2003) and DNA nearest-neighbor parameters adjusted for 100 mM monovalent salt (Allawi and SantaLucia 1998; SantaLucia 1998; Peyret 2000). Using this approach,  $\Delta G_{f,37}^{ref}$  values of  $-4.1$  and  $-9.7$  kcal/mol were predicted for DNA(AC) and DNA(GC), respectively, which are in reasonable agreement with the measured values of  $-3.6$  kcal/mol and  $-8.0$  kcal/mol.

For further comparison, the value of  $pk_{a,f}$  was determined independently using pH titrations with NMR or UV detection and fitting to a two-microscopic-state model. We recently reported the use of phosphorothioates as indirect labels for determining  $pk_{a,f}$  values in dsDNA (Moody et al. 2004a). Using the same oligonucleotide as in Figure 6, we determined an average  $pk_{a,f}$  of 6.85 at 30°C and 7.06 at 14°–19°C, both with Hill coefficients of  $\approx 1$ . We also determined a  $pk_{a,f}$  for this oligonucleotide by pH titration with UV detection at 23°C, fitting the data according to equation 25 (Fig. 7). The  $\approx 4\%$  hypochromicity upon  $A^+ \cdot C$  wobble pair formation was monitored at three wavelengths between 265 and 285 nm. The value of  $pk_{a,f}$  was determined to be 7.0–7.1 with a Hill coefficient near 1, independent of wavelength (Fig. 7). Control experiments on an oligonucleotide with a G in place of the A of the  $A^+ \cdot C$  mismatch in DNA(GC) did not show the sigmoidal hypochromicity behavior at these wavelengths, as expected; an observed increase in absorbance near pH 9 is likely due to the onset of alkaline denaturation (Fig. 7). In summary, pH titration experiments revealed a  $pk_{a,f}$  for DNA(AC) between 6.85 and



**FIGURE 4.** Contributions of proton activity to stability for binding of 1 proton to F and 6 protons to U. (A) Simulation of  $\Delta\Delta G_f^{obs}$  versus pH for the model in Scheme 4. A hairpin with two As and four Cs was assumed. Simulations are according to equation 15, binding polynomials from Scheme 4,  $pk_{a,f}$  of 7.5, and  $pk_{a,u,s}$  of 3.7 and 4.4 for As and Cs, respectively. The black line is the summation of  $\Delta G_H^f$  (blue) and  $-\Delta G_H^u$  (in red). (B) Plot of the first derivative of the data in panel A versus pH. Panels share the same X-axis to facilitate comparison of flex points (A) and inflection points (B).  $pk_a$  values are noted in the figure.



**FIGURE 5.** Contribution of proton activity to stability for binding of 0–1 proton to F and multiple protons to U. Simulations are of  $\Delta\Delta G_f^{\text{obs}}$  versus pH for the model in Scheme 4. Each simulation considers a hairpin 12-mer (black), 24-mer (blue), and 80-mer (red) in which protonation of a single nucleotide precludes the entire oligonucleotide from folding. Equal representation of the four bases is assumed. Simulations are according to equation 15, binding polynomials from Scheme 4,  $pK_{a,f}$  of 7.5, and  $pK_{a,u}$ s of 3.7, 4.4, 9.4, and 9.4 for As, Cs, Gs, and Us, respectively. (A) Simulations with no protonations in the folded state (i.e.,  $\Sigma_f = 1$ ; e.g., canonical Watson-Crick base pairs only) and multiple A and C protonations in the unfolded state. (B) Simulations with a single protonation in the folded state and multiple A and C protonations in the unfolded state. (C) Simulations with no protonations in the folded state and multiple A, C, G, and U protonations in the unfolded state. (D) Simulations with a single protonation in the folded state and multiple A, C, G, and U protonations in the unfolded state.

7.1 (Fig. 7), with the smaller values near the higher temperature of 30°C, indicative of a negative enthalpy upon protonation. These values are in good agreement with the  $pK_{a,f}$  of 6.7 determined at 37°C from pH-dependent melts (Fig. 6C).

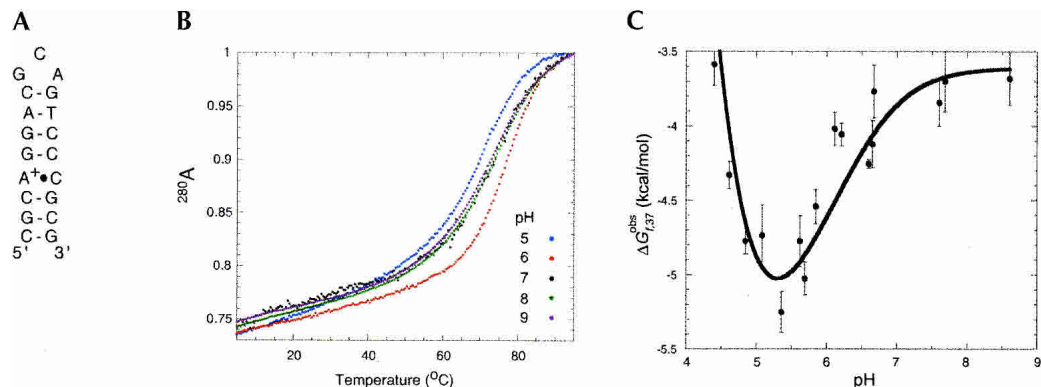
## DISCUSSION

The thermodynamic framework presented herein involves linkage between proton binding and RNA and DNA folding. The equations derived from the framework take into consideration protonations of both the folded and unfolded states, and provide appropriate equations to allow folded-state  $pK_a$  values to be determined from melting experiments. The formalism is similar to that for proton binding to proteins (Tanford 1968, 1970; Schellman 1975; Freire 1998; Luque et al. 2002), and gives expected dependencies for nucleic acid folding upon proton binding to U and F.

Although monitoring observed free energy as a function of pH appears to provide a reasonable  $pK_{a,f}$  for model oligonucleotides, the value of this approach may be more in the insight it affords into the driving forces for  $pK_a$  shifting and less in its experimental feasibility for obtaining  $pK_{a,f}$  values. For instance, obtaining plots of  $\Delta G_{f,37}^{\text{obs}}$  versus pH requires the average of several complete melting profiles, which is more labor-intensive than performing pH titrations. Moreover, the resulting curves have a complex shape, necessitating that melts be acquired at a large number of pH values. An additional problem is that buffers have temperature-dependent  $pK_a$ s; this causes the actual pH of the solution to change during the course of a given melt (see Materials and Methods).

In general, isothermal pH titrations with NMR or UV detection appear to provide a better alternative for obtaining a  $pK_{a,f}$  value. Isothermal titrations have the added benefits of a simpler functional dependence on pH and a much faster acquisition of data. Nevertheless, the framework developed herein for fitting  $\Delta G_{f,37}^{\text{obs}}$  versus pH plots provides insight into the role of RNA and DNA folding in shifting  $pK_a$ s. It is particularly instructive to consider the intrinsic contribution of a protonated  $A^+ \cdot C$  mismatch to stability (see footnote 8).

In Scheme 3B, the intrinsic energetic contribution from folding of the  $A^+ \cdot C$  wobble pair is isolated in the  $\Delta G_x$  term, where  $\Delta pK_{a,37} = -\Delta G_{x,37}/1.42$  (equation 24). For DNA(AC),  $pK_a$  shifting from a  $pK_{a,u,37}$  of 3.7 to a  $pK_{a,f,37}$  of 6.7 leads to a surprisingly large  $\Delta G_{x,37}$  of  $-4.25$  kcal/mol. Assuming that the A and C are partially stacked in the unprotonated state, this free energy reflects completion of stacking, hydrogen bonding, and electrostatics interactions. SantaLucia and coworkers calculated nearest-neighbor parameters for the formation of  $G \cdot T$  mismatches (Allawi and SantaLucia 1997), which are isosteric with  $A^+ \cdot C$  wobble pairs (Fig. 1). Free energy terms for formation of  $G \cdot T$  mismatches next to Watson-Crick base pairs are modest, ranging from  $+0.74$  to  $-0.59$  kcal/mol (Allawi and SantaLucia 1997). These values reflect formation of stacks and two hydrogen bonds. Upon protonation, the  $A^+ \cdot C$  mismatch also forms two hydrogen bonds and probably has a similar, or perhaps even lesser, gain in stacking. Hydrogen bonding should be strengthened in this base pair because



**FIGURE 6.** Dependence of observed free energy on pH for DNA(AC). Melts were conducted in a background of 100 mM KCl. (A) Secondary structure of DNA(AC) in the folded protonated form. (B) Representative melting curves. Normalized absorbance at 280 nm was plotted versus temperature at various pH values. Note that lowering the pH increases stability until the pH is near 5. (C) Plot of  $\Delta G_{f,37}^{obs}$  versus pH.  $\Delta G_{f,37}^{obs}$  values were determined from fitting the data in B to a two-state model. All data points are the average of three or more melts. Error bars are the standard deviations of the  $\Delta G_{f,37}^{obs}$  values from individual melts. Data were fit to equation 14 using a  $\Sigma_f$  with one protonation and a  $\Sigma_u$  with three As and eight Cs (Scheme 4), as dictated by the sequence of DNA(AC). The  $pK_{a,s}$  for A and C were fixed at 3.7 and 4.4, respectively. Values for  $pK_f$  and  $\Delta G_{f,37}^{ref}$ , which were the only two adjustable parameters, were  $6.7 \pm 0.1$  and  $-3.6 \pm 0.2$  kcal/mol. The pH values were corrected from room temperature pH values to values at the  $T_M$  using  $\Delta pK/^\circ C$  values for the respective buffers (Gueffroy 1985).

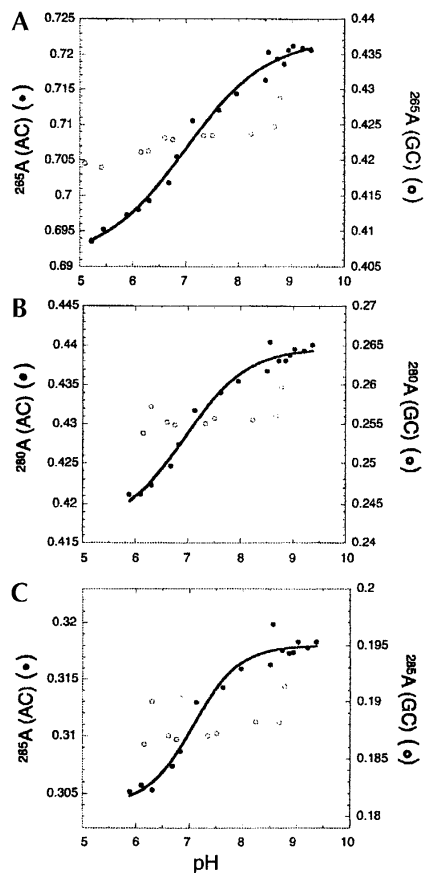
the donor group is a cation (Jeffrey and Saenger 1991). We conclude that favorable electrostatic interaction between the cationic adenine and the phosphate backbone accounts for the other portion of the  $-4.25$  kcal/mol free energy bonus.

Nucleic acids have a high charge density in the folded state, making electrostatics important to folding (Chin et al. 1999; Misra and Draper 1999). Curiously, the phosphate backbone did not strongly perturb the  $pK_a$  in the unfolded state in 100 mM ionic strength, with changes of only  $\approx +0.2$  units (Fig. 2). However, the axial charge density of DNA increases  $\approx 2.5$  times upon duplex formation (Manning 1978), which provides a significantly more favorable environment for the positive charge. Also, an internal positive charge in a helix reduces the number of counterions needed for neutralization, and may promote a release of local counterions upon folding, although this is likely to be dependent on oligonucleotide length. In the future, it will be important to develop formalisms that account for the linkage of folding to both protons and metal cations. Intriguingly, these calculations suggest that if in the unfolded state a base analog were cationic at neutrality, the analog would make a very large contribution to stability.

SantaLucia and coworkers carried out studies of the energetic contribution of  $A^+ \cdot C$  mismatches to duplex stability, compiling observed nearest-neighbor free energy terms at both pH 5.0 and 7.0 (Allawi and SantaLucia 1998). They found that the average contribution of a trimer sequence, which is also equivalent to two AC-containing nearest-neighbor terms, was  $+0.29$  kcal/mol at pH 5 and  $+1.68$  kcal/mol at pH 7.0. The difference between their small, destabilizing observed free energies and our large, stabilizing intrinsic free energy can be accounted for mostly in terms of thermodynamic penalties for both unloading “incorrect” and loading “correct” protons in the unfolded

state. In particular, although the macroscopic folded state consists primarily of the  $F'H^+$  species at pH 5.0, the macroscopic unfolded state consists of a complex ensemble of protonated species. Since C has a substantially higher  $pK_{a,u}$  than A (4.4 vs. 3.7), protonation of C residues will contribute to  $\Delta G_H^u$  even at pH 5.0; such protonated Cs are among the “incorrect” protons that must be unloaded before folding. SantaLucia and coworkers used  $\approx 12$ -bp model oligonucleotides in their studies, making the 24-mer simulation in Figure 5 relevant. This simulation shows that  $\Delta G_H^u$  contributes  $\approx +1.0$  kcal/mol to folding at pH 5.0. Also, there is an additional penalty for loading the “correct” proton onto the A of the  $A^+ \cdot C$  mismatch at pH 5.0 of  $\approx 1.85$  kcal/mol [ $= 1.42 \times (5.0 - 3.7)$ ] (Scheme 3B). Together, these two terms account for much of the difference between  $\Delta G_x$  and observed nearest-neighbor free energy for the  $A^+ \cdot C$  mismatch at pH 5.0.

One of the goals of the present study was to explore the feasibility of large  $pK_a$  perturbations in RNA. The work presented herein suggests that the major contributor to  $pK_a$  shifting in the helix-coil transition for DNA and RNA appears to be the additional folding conferred by binding of a proton. It follows that  $pK_a$  shifting might be maximized by increasing the cooperativity of structure formation upon protonation. Folding of secondary structures has been shown to be cooperative in the proper context (Moody and Bevilacqua 2003, 2004; Moody et al. 2004b), and should carry over to tertiary structures, suggesting that large  $pK_a$  shifts may be feasible. Highly cooperative structure formation might lead to  $pK_a$  shifting into the basic regime, providing cationic base pairs at neutrality with possible lysine-like function. The thermodynamic framework developed herein should be applicable to folding of tertiary structure motifs as well. The major obstacle in accomplishing this



**FIGURE 7.** pH titrations with UV detection of DNA(AC) (●) and DNA(GC) (○). Here 100 mM KCl was added to all samples prior to the titrations, and data were fit to equation 25. The ranges for the right-hand Y-axes were set the same as for the left Y-axes. (A) Monitoring at 265 nm;  $pK_f$  from fit is  $7.0 \pm 0.1$  and  $n$  is  $0.6 \pm 0.1$ . (B) Monitoring at 280 nm;  $pK_f$  from fit is  $7.0 \pm 0.2$  and  $n$  is  $0.7 \pm 0.2$ . (C) Monitoring at 285 nm;  $pK_f$  from fit is  $7.1 \pm 0.1$  and  $n$  is  $1.0 \pm 0.3$ .

may prove to be isolating the folding transition of interest, as well as determining the number of critical ionizable bases in the reference state so that a suitable  $\Sigma_u$  can be developed.

## MATERIALS AND METHODS

### Preparation of DNA and RNA

DNA oligonucleotides were prepared by solid-phase synthesis and deblocked and desalted by the manufacturer (IDT). Oligonucleotides were then dialyzed against deionized water for 3–4 h using a GIBCO BRL microdialysis apparatus with a flow rate of ~1 L/h and a SpectraPor membrane with a molecular weight cutoff of 1000. The sequence of DNA(AC), which forms a hairpin with an internal  $A^+ \cdot C$  mismatch (underlined), is  $d(5'-CGCAGGACGCAGTCCC\underline{GC}G)$ . This sequence was chosen to have at least three GC base pairs on either side of the AC to facilitate retention of base-pairing upon adenine deprotonation and to also help resist alkaline denaturation of the stem at higher pH values. The sequence of the control oligonucleotide DNA(GC) involves changing the under-

lined A to a G. RNA oligonucleotides were also prepared by solid-phase synthesis and were deblocked and desalted according to the manufacturer's instructions (Dharmacon) and stored in water. The sequences for model oligonucleotides for unfolded-state microscopic pks were  $r(5'-UA)$ ,  $r(5'-UC)$ , and  $r(5'-UUCUU)$ .

### Thermal denaturation experiments

All experiments were conducted in a background of 100 mM KCl to maintain constant and physiologically relevant ionic strength (Feig and Uhlenbeck 1999) and to overwhelm small salt differences generated during titrations or preparing of buffers at different pH values. In order to cover the necessary pH range for melting experiments, a series of different buffers was used: 25 mM MES (pH 5.0–6.0), 25 mM MOPS (pH 6.0–7.0), 25 mM HEPES (pH 6.45–8.5), and 25 mM HEPPS (pH 7.0–9.0), where MES, MOPS, HEPES, and HEPPS have  $pK_a$  values at 20°C of 6.15, 7.20, 7.55, and 8.00, respectively (Gueffroy 1985). Representative melting experiments at a given pH were repeated using different buffers. In particular, melts at pH 6.0 were carried out in MOPS or MES, and melts at pH 7.0 were carried out in MOPS, HEPPS, or HEPES; resulting thermodynamic parameters were independent of the identity of the buffer, suggesting that the buffer is not directly involved in the equilibrium. These four buffers were chosen in part because their pKs have a shallow dependence on temperature:  $\Delta pK/^\circ C$  is  $-0.011$ ,  $-0.006$ ,  $-0.014$ , and  $-0.007$  for MES, MOPS, HEPES, and HEPPS, respectively (Gueffroy 1985). Despite these precautions, changes in pH during the actual thermal denaturation experiments, which typically covered a range of 60°–80°C, were unavoidable. For  $\Delta G_f^{obs}$  versus pH plots, the pH of a given melt was assigned as the pH at the temperature of the transition midpoint, which was calculated using the appropriate  $\Delta pK/^\circ C$  factor; this choice was made since solution conditions near the transition midpoint most strongly influence the values of  $\Delta H$  and  $\Delta S$ . As an example, a melting profile with a  $T_M$  of 63°C in the presence of MOPS (pH 6.00 at 25°C) was adjusted to pH 5.77 for  $\Delta G_f^{obs}$  versus pH curve fitting.

Thermal denaturation with UV detection was carried out similarly to that previously described (Moody and Bevilacqua 2003). Briefly, melting experiments were performed using a Gilford Response II spectrophotometer at 280 nm with a heating rate of 0.5°C/min. Melts were in 1-, 5-, or 10-mm-pathlength microcuvettes, at a minimum of three different strand concentrations ranging from 1 to 80  $\mu M$ . Melting profiles were independent of strand concentration, consistent with the hairpin conformation; if duplexes were forming, calculations using a nearest-neighbor model and an ionic strength of 100 mM (Zuker 2003) predicted a  $T_M$  change of  $\sim 10^\circ C$  over a 20-fold change in strand concentration, which is well outside the error limits of the experiments ( $\pm 1^\circ C$ ). At the start of each experiment, the hairpin was renatured by heating to 90°C in melting buffer. Data for forward and reverse melts were 90% or more similar, consistent with reversibility of the melting transition. Concentrations of RNA and DNA were calculated using absorbance values at 90°C and extinction coefficients from a nearest-neighbor analysis (Borer 1975; Richards 1975). Data were the average of three or more melts. Thermodynamic parameters were obtained by nonlinear least squares fitting to a two-state model using a set of parametric equations defined in Kaleidagraph v3.5 (Synergy software) (Nakano et al. 2003). Linear

baselines with nonzero slopes were included in the fitting of the signal.

### UV-absorbance detected pH titrations

For DNA and RNA hairpin oligonucleotides,  $pK_a$  values were directly determined by a UV-absorbance-detected pH titration. Depending on the case, oligonucleotides were designed to remain either fully folded or fully unfolded throughout the titration, and so afford a two-microscopic-state system. Desalted oligonucleotides were dried in vacuo and resuspended in distilled-deionized water from a Barnstead NANOpure Diamond unit, which minimized absorbance in the far-UV region of the spectrum. As in the case of the melting experiments, 100 mM KCl was added to overwhelm small changes in salt concentration generated during the pH titration. In these experiments, buffer was deliberately omitted to allow the  $pK_a$  value for the oligonucleotide to be directly observed. Concentrations of oligonucleotides were 10–35  $\mu\text{M}$  and a total volume of  $\approx 3.5$  mL in a 1-cm-pathlength cell was used. UV absorbance was monitored at room temperature on a Beckman Coulter DU 650 spectrophotometer, and pH was monitored using a Corning 430 pH meter with a Mettler Toledo 3-mm-diameter AgCl reference electrode. The spectrophotometer was zeroed on 100 mM KCl in distilled-deionized water at the beginning of the experiment. Approximately 1  $\mu\text{L}$  of a dilute HCl stock solution (pH  $\approx 1.0$ ) or KOH stock solution (pH  $\approx 12$ ) was added to the sample, followed by mixing, a pH determination, and a UV scan from 220 to 320 nm. To minimize scatter in the data, the cuvette remained in the spectrophotometer for the duration of the experiment, with mixing of the sample performed manually with a Pipetman set to 1 mL. The pH electrode was allowed to equilibrate in the sample for  $\approx 1$  min before each reading. This time was sufficient to remove any noticeable drift in the reading. Throughout the titration, the change in the total volume was  $<1\%$ , and therefore no corrections for volume were made.

Absorbance ( $A$ ) at a particular wavelength versus pH data was fit to equation 25 to obtain the  $pK_a$  and Hill coefficient:

$$A = A_p + (A_U - A_p) \frac{1}{1 + 10^{n(pk_a - pH)}} \quad (25)$$

where  $A_U$  and  $A_p$  are the absorbance values of the unprotonated and protonated states, respectively, and  $n$  is a Hill constant reflecting the number of proton-binding sites. This equation assumes infinite cooperativity between multiple sites, if present. A value of  $n$  near unity is consistent with one proton-binding site, whereas a value significantly greater than unity would be consistent with the closest higher integer being the number of sites. Several different wavelengths were chosen for fitting titration data, and were determined on the basis of good signal to noise and locations near isosbestic points for any neighboring transitions. Particular wavelengths that gave the best signal depended on whether the oligonucleotide was folded or unfolded throughout the titration, as well as on the identity of the ionizing base, adenine or cytosine. Titrations of r(UC) and r(UUCUU) were monitored at 240 and 290 nm, where upon protonation the absorbance of a C has the greatest change; titrations of r(UUAUU) were monitored at 225 and 230 nm, where upon protonation the absorbance of an A has the greatest change; and titrations of folded oligonucleotides were monitored at 265, 280, or 285 nm to observe

changes in hypochromicity caused by local folding events dependent on adenine protonation. In all cases, baselines were taken as horizontal (equation 25), since there was no evidence against this simple behavior. This assumption has been applied in numerous other  $pK_a$  determinations in oligonucleotides (Wang et al. 1991; Legault and Pardi 1997; Ravindranathan et al. 2000).

### Simulation of data and curve fitting

Simulations and nonlinear curve fitting were performed using Kaleidagraph v3.5 (Synergy Software). Simulations were developed from equation 15 and the appropriate binding polynomials in the schemes. Experimental data ( $\Delta G_{f,37}^{\text{obs}}$  vs. pH plots) were fit to equation 14 using the appropriate binding polynomials from Scheme 4; the only adjustable parameters were  $pK_{a,f}$  and  $\Delta G_f^{\text{ref}}$ . In general, equations used in curve fitting were expressed in terms of pH and  $pK_a$  to allow plotting versus pH and direct accessing of  $pK_a$ , whereas equations for first derivatives were expressed in terms of  $\text{H}^+$  concentration and  $K_a$  to suggest the hyperbolic nature of the functions.

### ACKNOWLEDGMENTS

We thank Doug Turner and members of the Bevilacqua lab for comments on the manuscript. This work was supported by a Camille Dreyfus Teacher-Scholar Award (P.C.B.), a fellowship from the Alfred P. Sloan Foundation (P.C.B.), NSF Grant MCB-9984129 (P.C.B.), NIH grants GM58709 (P.C.B.) and GM 54217 (J.T.J.L.), and a Paul Berg fellowship (E.M.M.).

Received September 9, 2004; accepted November 17, 2004.

### REFERENCES

- Acampora, G. and Hermans Jr., J. 1967. Reversible denaturation of sperm whale myoglobin. I. Dependence on temperature, pH, and composition. *J. Am. Chem. Soc.* **89**: 1543–1547.
- Albert, A. and Serjeant, E.P. 1984. *The determination of ionization constants*, 3d ed. Chapman and Hall, London.
- Allawi, H.T. and SantaLucia Jr., J. 1997. Thermodynamics and NMR of internal G·T mismatches in DNA. *Biochemistry* **36**: 10581–10594.
- . 1998. Nearest-neighbor thermodynamics of internal A·C mismatches in DNA: Sequence dependence and pH effects. *Biochemistry* **37**: 9435–9444.
- Bevilacqua, P.C. 2003. Mechanistic considerations for general acid-base catalysis by RNA: Revisiting the mechanism of the hairpin ribozyme. *Biochemistry* **42**: 2259–2265.
- Bevilacqua, P.C., Brown, T.S., Nakano, S., and Yajima, R. 2004. Catalytic roles for proton transfer and protonation in ribozymes. *Bio-polymers* **73**: 90–109.
- Blanchard, S.C. and Puglisi, J.D. 2001. Solution structure of the A loop of 23S ribosomal RNA. *Proc. Natl. Acad. Sci.* **98**: 3720–3725.
- Borer, P.N. 1975. In *Handbook of biochemistry and molecular biology: Nucleic acids*, 3d ed. (ed. G.D. Fasman), Vol. I, p. 597. CRC Press, Cleveland, OH.
- Bukhman, Y.V. and Draper, D.E. 1997. Affinities and selectivities of divalent cation binding sites within an RNA tertiary structure. *J. Mol. Biol.* **273**: 1020–1031.
- Chin, K., Sharp, K.A., Honig, B., and Pyle, A.M. 1999. Calculating the electrostatic properties of RNA provides new insights into molecular interactions and functions. *Nat. Struct. Biol.* **6**: 1055–1061.

- Feig, A.L. and Uhlenbeck, O. 1999. The role of metal ions in RNA biochemistry. In *The RNA world*, 2d ed. (eds. R.F. Gesteland et al.), pp. 287–319. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Ferre-D'Amare, A.R. and Doudna, J.A. 2000. Crystallization and structure determination of a hepatitis delta virus ribozyme: Use of the RNA-binding protein U1A as a crystallization module. *J. Mol. Biol.* **295**: 541–556.
- Freire, E. 1998. Statistical thermodynamic linkage between conformational and binding equilibria. *Adv. Protein Chem.* **51**: 255–279.
- Graziano, G., Catanzano, F., and Nappa, M. 1999. Linkage of proton binding to the thermal unfolding of Sso7d from the hyperthermophilic archaeobacterium *Sulfolobus solfataricus*. *Int. J. Biol. Macromol.* **26**: 45–53.
- Gueffroy, D.E. 1985. *A guide for the preparation and use of buffers in biological systems*. Behring Diagnostics.
- Hermans Jr., J. and Acampora, G. 1967. Reversible denaturation of sperm whale myoglobin. II. Thermodynamic analysis. *J. Am. Chem. Soc.* **89**: 1547–1552.
- Hirao, I., Kawai, G., Yoshizawa, S., Nishimura, Y., Ishido, Y., Watanabe, K., and Miura, K. 1994. Most compact hairpin-turn structure exerted by a short DNA fragment, d(GCGAAGC) in solution: An extraordinarily stable structure resistant to nucleases and heat. *Nucleic Acids Res.* **22**: 576–582.
- Hoogstraten, C.G., Legault, P., and Pardi, A. 1998. NMR solution structure of the lead-dependent ribozyme: Evidence for dynamics in RNA catalysis. *J. Mol. Biol.* **284**: 337–350.
- Hunter, W.N., Brown, T., Anand, N.N., and Kennard, O. 1986. Structure of an adenine–cytosine base pair in DNA and its implications for mismatch repair. *Nature* **320**: 552–555.
- Huppler, A., Nikstad, L.J., Allmann, A.M., Brow, D.A., and Butcher, S.E. 2002. Metal binding and base ionization in the U6 RNA intramolecular stem–loop structure. *Nat. Struct. Biol.* **9**: 431–435.
- Izatt, R.M., Christensen, J.J., and Rytting, J.H. 1971. Sites and thermodynamic quantities associated with proton and metal ion interaction with ribonucleic acid, deoxyribonucleic acid, and their constituent bases, nucleosides, and nucleotides. *Chem. Rev.* **71**: 439–481.
- Jeffrey, G.A. and Saenger, W. 1991. *Hydrogen bonding in biological structures*. Springer-Verlag, Berlin.
- Jencks, W.P. 1975. Binding energy, specificity, and enzymic catalysis: The circe effect. *Adv. Enzymol.* **43**: 219–410.
- Jencks, W.P. and Regenstein, J. 1970. Ionization constants of acids and bases. *Handbook of biochemistry: Selected data for molecular biology*, 2d ed. (ed. H.A. Sober), pp. J150–J189. Cleveland Rubber Co.
- Kim, J., Cheong, C., and Moore, P.B. 1991. Tetramerization of an RNA oligonucleotide containing a GGGG sequence. *Nature* **351**: 331–332.
- Knitt, D.S. and Herschlag, D. 1996. pH dependencies of the *Tetrahymena* ribozyme reveal an unconventional origin of an apparent  $pK_a$ . *Biochemistry* **35**: 1560–1570.
- Laing, L.G., Gluick, T.C., and Draper, D.E. 1994. Stabilization of RNA structure by Mg ions. Specific and non-specific effects. *J. Mol. Biol.* **237**: 577–587.
- Legault, P. and Pardi, A. 1997. Unusual dynamics and  $pK_a$  shift at the active site of a lead-dependent ribozyme. *J. Am. Chem. Soc.* **119**: 6621–6628.
- Luque, I., Leavitt, S.A., and Freire, E. 2002. The linkage between protein folding and functional cooperativity: Two sides of the same coin? *Annu. Rev. Biophys. Biomol. Struct.* **31**: 235–256.
- Manning, G.S. 1978. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* **11**: 179–246.
- Martel, A.E. and Smith, R.M. 1974. Critical stability constants. In *Critical stability constants*, Vol. 2, p. 36. Plenum Press, New York.
- Mathews, D.H., Sabina, J., Zuker, M., and Turner, D.H. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**: 911–940.
- Misra, V.K. and Draper, D.E. 1998. On the role of magnesium ions in RNA stability. *Biopolymers* **48**: 113–135.
- . 1999. The interpretation of  $Mg^{2+}$  binding isotherms for nucleic acids using Poisson-Boltzmann theory. *J. Mol. Biol.* **294**: 1135–1147.
- . 2002. The linkage between magnesium binding and RNA folding. *J. Mol. Biol.* **317**: 507–521.
- Misra, V.K., Shiman, R., and Draper, D.E. 2003. A thermodynamic framework for the magnesium-dependent folding of RNA. *Biopolymers* **69**: 118–136.
- Moody, E.M. 2004. “Stability in nucleic acid hairpins. (1) Molecular determinants of cooperativity and (2) linkage between proton binding and folding.” Ph.D. thesis, The Pennsylvania State University, University Park, PA.
- Moody, E.M. and Bevilacqua, P.C. 2003. Folding of a stable DNA motif involves a highly cooperative network of interactions. *J. Am. Chem. Soc.* **125**: 16285–16293.
- . 2004. Structural and energetic consequences of expanding a highly cooperative stable DNA hairpin loop. *J. Am. Chem. Soc.* **126**: 9570–9577.
- Moody, E.M., Brown, T.S., and Bevilacqua, P.C. 2004a. Simple method for determining nucleobase  $pK_a$  values by indirect labeling and demonstration of a  $pK_a$  of neutrality in dsDNA. *J. Am. Chem. Soc.* **126**: 10200–10201.
- Moody, E.M., Feerrar, J.C., and Bevilacqua, P.C. 2004b. Evidence that folding of an RNA tetraloop hairpin is less cooperative than its DNA counterpart. *Biochemistry* **43**: 7992–7998.
- Nakano, S., Chadalavada, D.M., and Bevilacqua, P.C. 2000. General acid–base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science* **287**: 1493–1497.
- Nakano, S., Cerrone, A.L., and Bevilacqua, P.C. 2003. Mechanistic characterization of the HDV genomic ribozyme: Classifying the catalytic and structural metal ion sites within a multichannel reaction mechanism. *Biochemistry* **42**: 2982–2994.
- Narlikar, G.J. and Herschlag, D. 1997. Mechanistic aspects of enzymatic catalysis: Lessons from comparison of RNA and protein enzymes. *Annu. Rev. Biochem.* **66**: 19–59.
- Nixon, P.L. and Giedroc, D.P. 2000. Energetics of a strongly pH dependent RNA tertiary structure in a frameshifting pseudoknot. *J. Mol. Biol.* **296**: 659–671.
- Nixon, P.L., Cornish, P.V., Suram, S.V., and Giedroc, D.P. 2002. Thermodynamic analysis of conserved loop–stem interactions in P1–P2 frameshifting RNA pseudoknots from plant Luteoviridae. *Biochemistry* **41**: 10665–10674.
- Perrotta, A.T., Shih, I., and Been, M.D. 1999. Imidazole rescue of a cytosine mutation in a self-cleaving ribozyme. *Science* **286**: 123–126.
- Peyret, N. 2000. “Prediction of nucleic acid hybridization: Parameters and algorithms.” Ph.D. thesis, Wayne State University, Detroit, MI.
- Ravindranathan, S., Butcher, S.E., and Feigon, J. 2000. Adenine protonation in domain B of the hairpin ribozyme. *Biochemistry* **39**: 16026–16032.
- Record Jr., M.T., Lohman, T.M., and deHaseth, P. 1976. Ion effects on ligand–nucleic acid interactions. *J. Mol. Biol.* **107**: 145–158.
- Richards, E.G. 1975. In *Handbook of biochemistry and molecular biology: Nucleic acids*, 3d ed. (ed. G.D. Fasman), Vol. 1, p. 597. CRC Press, Cleveland, OH.
- Richards, E.G., Flessel, C.P., and Fresco, J.R. 1963. Polynucleotides. VI. Molecular properties and conformation of polyribouridylic acid. *Biopolymers* **1**: 431–446.
- Roitzsch, M. and Lippert, B. 2004. Metal coordination and imine–amine hydrogen bonding as the source of strongly shifted adenine  $pK_a$  values. *J. Am. Chem. Soc.* **126**: 2421–2424.
- Saenger, W. 1984. *Principles of nucleic acid structure* (ed. C.R. Cantor). Springer-Verlag, New York.
- SantaLucia Jr., J. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci.* **95**: 1460–1465.
- Schellman, J.A. 1975. Macromolecular binding. *Biopolymers* **14**: 999–1018.

- Sharp, K.A. and Englander, S.W. 1994. How much is a stabilizing bond worth? *Trends Biochem. Sci.* **19**: 526–529.
- Takach, J.C., Mikulecky, P.J., and Feig, A.L. 2004. Salt-dependent heat capacity changes for RNA duplex formation. *J. Am. Chem. Soc.* **126**: 6530–6531.
- Tanford, C. 1968. Protein denaturation. *Adv. Protein Chem.* **23**: 121–282.
- . 1970. Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv. Protein Chem.* **24**: 1–95.
- Thompson, J.E. and Raines, R.T. 1994. Value of general acid-base catalysis to ribonuclease A. *J. Am. Chem. Soc.* **116**: 5467–5468.
- Turner, D.H. and Bevilacqua, P.C. 1993. Thermodynamic considerations for evolution by RNA. In *The RNA world* (eds. R.F. Gesteland and J.F. Atkins), pp. 447–464. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Wang, C., Gao, H., Gaffney, B.L., and Jones, R.A. 1991. Nitrogen-15-labeled oligodeoxynucleotides. 2. Protonation of the adenine N1 in the A-C and A-G mispairs of the duplexes {d[CG(<sup>15</sup>N<sup>1</sup>)AGAA TTCCCG]}<sub>2</sub> and {d[CGGGAATTC(<sup>15</sup>N<sup>1</sup>)ACG]}<sub>2</sub>. *J. Am. Chem. Soc.* **113**: 5486–5488.
- Warshel, A., Naray-Szabo, G., Sussman, F., and Hwang, J.K. 1989. How do serine proteases really work? *Biochemistry* **28**: 3629–3637.
- Westcott, C.C. 1978. *pH measurements*. Academic Press, New York.
- Wyman, J. 1965. The binding potential, a neglected linkage concept. *J. Mol. Biol.* **11**: 631–644.
- Wyman, J. and Gill, S.J. 1990. *Binding and linkage: Functional chemistry of biological macromolecules*. University Science Books, Mill Valley, CA.
- Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**: 3406–3415.