

# Simple Method for Determining Nucleobase $pK_a$ s by Indirect Labeling and Demonstration of a $pK_a$ of Neutrality in dsDNA

Ellen M. Moody,<sup>†</sup> Trevor, S. Brown,<sup>†,‡</sup> and Philip C. Bevilacqua<sup>\*,†,‡</sup>

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

and The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802

## Contents:

1. Preparation of DNA samples
2. NMR conditions
3. Fitting of the data
4. Chemical Shift versus pH data
5. Parameters from fits to additional pH titrations
6. Additional chemical shift versus pH data

### 1. Preparation of DNA samples

DNA oligonucleotides were synthesized, deblocked and desalted by the manufacturer (Integrated DNA Technologies). Oligonucleotides were dialyzed into sterile, distilled, deionized water as described.<sup>1</sup> The DNA sequences were DNA1 \*A: 5'-CG\*AGAATTC\*CG; DNA1 C\*: 5'-CGAGAATTCC\*CG; DNA2 A\*: CGCA\*GGACGCAGTCCCGCG; DNA2 C\*: 5'-CGCAGGACGCAGTCC\*GCG; DNA2 AG\*: 5'-CGCAG\*GACGCAGTCCCGCG and DNA2 A\* ctrl: 5'-CGCA\*GGACGCAGTCCTGCG, where \* indicates the position of the phosphorothioate substitution. Electrospray mass spectrometry confirmed the phosphorothioate substitution in DNA2 A\*. The secondary structures in Figure 2 were confirmed by *mfold 3.1*<sup>2</sup> with the smallest  $\Delta\Delta G_{37}^{\circ}$  equal to 3.3 kcal/mol, calculated as the difference between the optimal and first suboptimal predicted secondary structures. Self-complementary DNA1 was predicted to have a  $\Delta G_{37}^{\circ}$  of +1 kcal/mol in forming any intra-strand secondary structure, thus supporting duplex over hairpin formation. A single set of two <sup>31</sup>P peaks for all sequences at all pH values is consistent with a single secondary structure. All experiments were conducted without addition of a buffer and in the background of 100 mM KCl to maintain constant and physiologically-relevant ionic strength throughout the titrations, which involved addition of small amounts of KOH or HCl. NMR samples were 100 mM KCl and 5% D<sub>2</sub>O. DNA concentrations ranged from 640  $\mu$ M to 2.5 mM, and the oligonucleotides were renatured prior to the start of each experiment by heating to 90 °C for 3 minutes and cooling on bench top for 10 minutes. An internal coaxial tube containing 1% trimethyl phosphate (TMP) in 5% D<sub>2</sub>O was used as a reference and set to 0 ppm. The pH values were determined using a Accumet 3 mm micro combination electrode with a calomel reference or a Mettler Toledo 3 mm diameter AgCl reference electrode using a Corning 430 or Accumet Basic AB15 meter. The pH was determined both before and after the NMR experiment; values agreed within 0.05 pH units and were averaged for each point. All sequences were titrated from low to high pH by the addition of 0.1 M KOH, with the exception of DNA1 C\* and DNA2 C\*, which were titrated from high to low pH by adding 0.05 M HCl. The pH values were determined at 30 °C for all data in Table 1, except DNA1 C\*, which was determined at room temperature (~20 °C) and mathematically corrected to 31 °C. The pH values for NMR data at additional temperatures were determined at room temperature (~23 °C) for all sequences except DNA2 AG\*, which was determined at 33 °C.

### 2. NMR conditions

The NMR titrations in Table 1 were carried out at 31 °C for DNA1 C\*, as per Wang et al.,<sup>3</sup> and at 30 °C for all DNA2 sequences for comparison. The NMR spectra for additional temperatures given in Supporting Information were collected at 23 °C for DNA1, 19 °C for DNA2 A\* and 14 °C for DNA2 C\*, DNA2 AG\* and DNA2 A\* ctrl. The chemical shift of the R<sub>p</sub> and S<sub>p</sub> phosphorothioate resonances were followed as a function of pH. NMR data were collected on a Bruker AMX2-500 spectrometer using a 5 mm broadband probe with a <sup>31</sup>P frequency of 202.46 MHz and a 90° pulse of 20  $\mu$ sec. <sup>31</sup>P spectra were acquired with a spectral width of 23,809 Hz, a recycle delay of 3 s, and proton decoupling (WALTZ-16).<sup>4</sup> Data were processed using XWINNMR (Bruker). 32K data points were collected and the FIDs were zero-filled to 64K and apodized using 10 Hz line broadening.

\* To whom correspondence should be addressed. E-mail: pcb@chem.psu.edu

<sup>†</sup> The Department of Chemistry

<sup>‡</sup> The Huck Institutes of the Life Sciences

### 3. Fitting of the data

Chemical shift ( $\delta$ ) versus pH data were fit to equation (1), which is the Henderson-Hasselbalch equation adapted for fast chemical exchange on the NMR time scale, to obtain the  $pK_a$  value,  $n$ , and  $\Delta\delta$ ,

$$\delta = \delta_p + \frac{\Delta\delta}{1 + 10^{n(pK - pH)}}, \quad (1)$$

where  $\Delta\delta = \delta_u - \delta_p$ ,  $\delta_u$  and  $\delta_p$  are the chemical shift values of the unprotonated and protonated states (in ppm), 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, or independent sites. The values of  $pK_a$ ,  $n$ , and  $\Delta\delta$  from fits of data collected at 30 °C and 31 °C are provided in Table 1. As a control, the slope of the linear fit of DNA2 A\* ctrl was used as a baseline correction to eq 1 in fitting peak 1 of DNA2 C\*; at 14 °C this was done to test the possible effect on the resultant  $pK_a$ s. The curve fits yielded  $pK_a$  values of  $7.08 \pm 0.03$  without a correction,  $6.99 \pm 0.02$  with a lower baseline correction,  $7.18 \pm 0.03$  with an upper baseline correction, and  $7.08 \pm 0.02$  with a correction to both baselines. This demonstrates that sloping baselines as seen in DNA2 A\* ctrl do not affect the calculated  $pK_a$  value by more than  $\pm 0.1$ .

### 4. Chemical Shift versus pH data

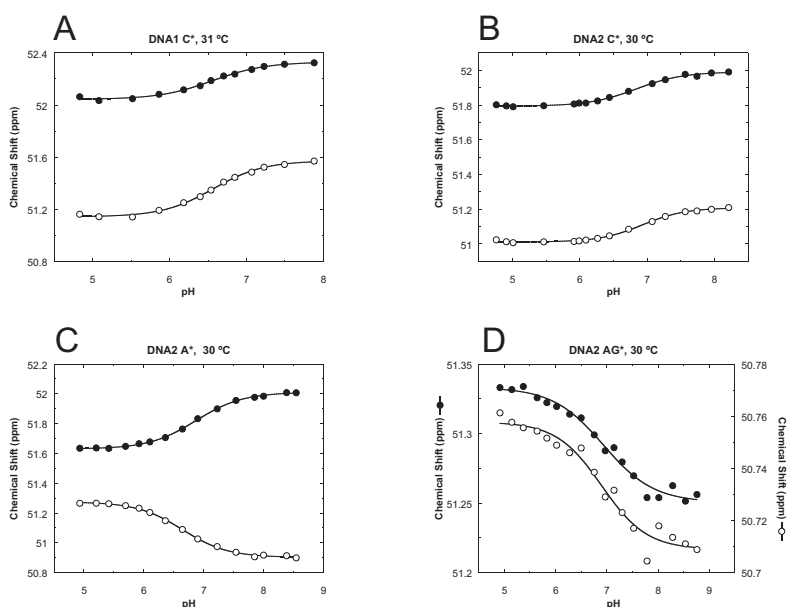


Figure S1: Titration curves for data in Table 1. A) DNA1 C\* at 31 °C, B) DNA2 C\* at 30 °C, C) DNA2 A\* at 30 °C, and D) DNA2 AG\* at 30 °C. Both Peak 1 (the most downfield shifted peak; filled circles) and Peak 2 (open circles) are shown. The reported chemical shifts reported are relative to TMP (0 ppm). Data are fit with equation 1 and  $pK_a$  and  $n$  values for panels A-D are provided in Table 1.

## 5. Parameters from fits to additional pH titrations

Sequence	Peak <sup>a</sup>	$\Delta\delta$ (ppm) <sup>b,c</sup>	n <sup>c</sup>	$pK_a^c$
DNA1 C*	1	0.45±0.01	0.89±0.06	6.64±0.03
	2	0.6±0.02	1.01±0.08	6.63±0.03
DNA2 C*	1	0.306±0.007	0.98±0.07	7.08±0.03
	2	0.243±0.005	1.26±0.09	7.00±0.03
DNA2 A*	1	0.367±0.006	1.05±0.05	7.17±0.02
	2	-0.404±0.007	0.92±0.04	7.00±0.02
DNA2 AG*	1	-0.127±0.002	0.94±0.05	6.94±0.02
	2	-0.035±0.003	0.9±0.3	7.2±0.1

Table S1: <sup>a</sup>Peak 1 is the more downfield shifted phosphorothioate peak. <sup>b</sup>Calculated as the difference between high and low pH. <sup>c</sup>Values of  $\Delta\delta$ , n, and  $pK_a$  were obtained from non-linear curve fitting to eq 1. Temperature values are between 14 and 23 °C and are provided in Section 2. The average  $\Delta pK_a$  at these lower temperatures (0.42) is similar to that at 30 °C (0.29).

## 6. Additional Chemical Shift versus pH data

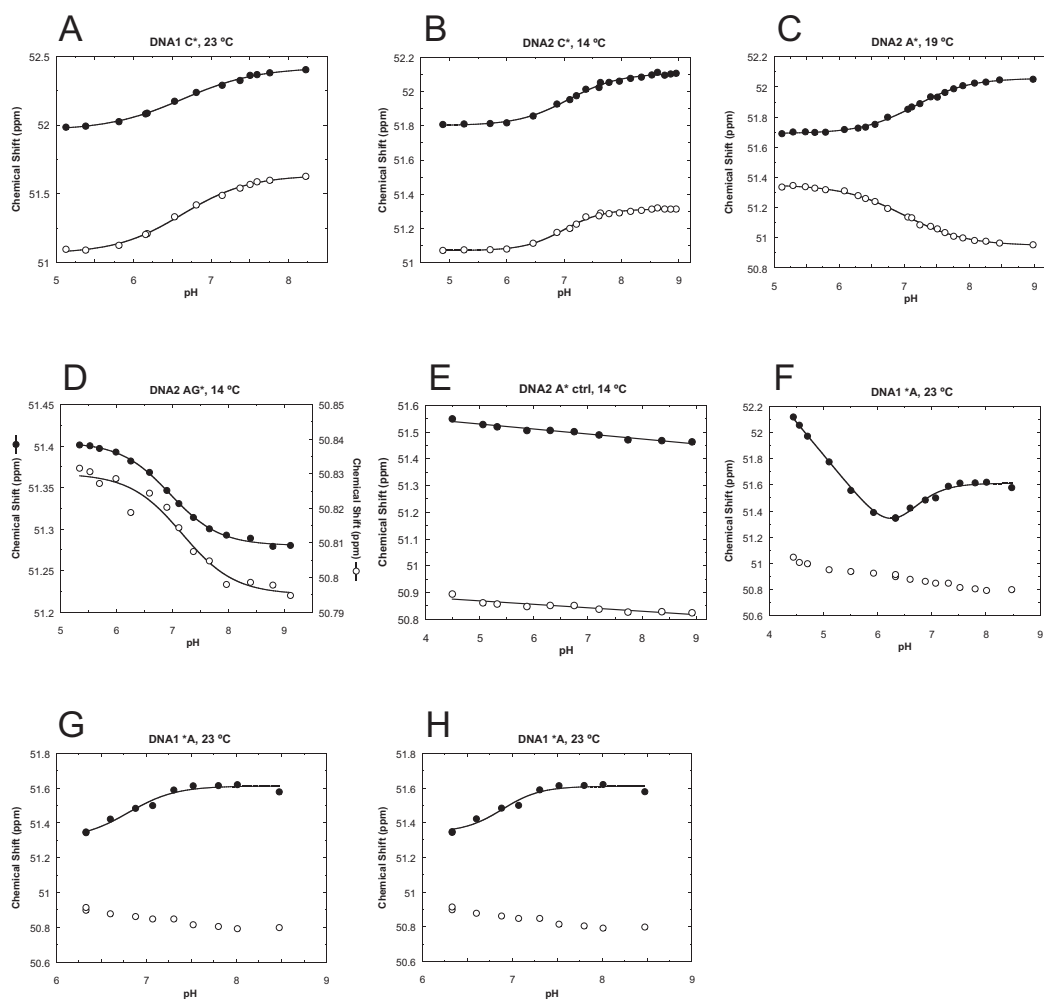


Figure S2: Additional titration curves. A) DNA1 C\* at 23 °C, B) DNA2 C\* at 14 °C, C) DNA2 A\* at 19 °C, D) DNA2 AG\* at 14 °C, E) DNA2 A\* ctrl at 14 °C, and F, G, H) DNA1 \*A at 23 °C. Both peak 1 (the most downfield shifted peak; filled circles) and peak 2 (open circles) are shown. The reported chemical shifts are relative to TMP (0 ppm). In panels A – D, the data are fit with equation 1. Panel E data are fit with a straight line. Panel F data are fit with equation 1 using the slope of the data below pH 5.93 to represent the dependence of  $\delta_p$  on pH. Panel G and H are DNA1 \*A data  $\geq$  pH 6.33 fit with equation 1 with a lower baseline that is either variable or fixed at the chemical shift value observed at pH 6.33 (51.344 ppm), respectively. In these two panels, only peak 1 is fit, but peak 2 data are shown for comparison.  $pK_a$  and  $n$  values for panels A-D are provided in Table S1.  $pK_a$  and  $n$  values for panels F, G, and H are  $6.46 \pm 0.03$ ,  $6.8 \pm 0.1$ , and  $6.89 \pm 0.06$ , respectively, and  $1.5 \pm 0.1$ ,  $1.7 \pm 0.7$ , and  $2.0 \pm 0.4$ , respectively. Because of the uncertainty in DNA1 \*A  $n$  values, caused by the low pH baseline, this  $pK_a$  data is treated only qualitatively.

---

## References

- (1) Moody, E. M.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2003**, 125, 16285-16293.
- (2) (a) Zuker, M. *Nucleic Acid. Res.* **2003**, 31, 3406-3415 (b) SantaLucia, J., Jr. *Proc. Natl. Acad. Sci. USA* **1998**, 95, 1460-1465.
- (3) Wang, C.; Gao, H.; Gaffney, B. L.; Jones, R. A. *J. Am. Chem. Soc.* **1991**, 113, 5486-5488.
- (4) Shaka, A. J.; Keeler, J.; Freeman, R. *J. Magn. Reson.* **1983**, 52, 313-340.