

Nearest-Neighbor Thermodynamics of Internal A•C Mismatches in DNA: Sequence Dependence and pH Effects

Hatim T. Allawi and John SantaLucia, Jr.*

Department of Chemistry, Wayne State University, Detroit, Michigan 48202

Received February 17, 1998; Revised Manuscript Received April 21, 1998

ABSTRACT: Thermodynamics of 27 oligonucleotides with internal A•C mismatches at two different pHs were determined from UV absorbance versus temperature melting profiles. The data were combined with four literature values and used to derive nearest-neighbor parameters for all 16 trimer sequences with internal A•C mismatches at pH 7.0 and 5.0. The results indicate that the contribution of single A•C mismatches to duplex stability is strongly dependent on the solution pH and the nearest-neighbor context. On average, the protonation of an internal A•C mismatch stabilizes the duplex by 1.39 kcal/mol for ΔG°_{37} and 7.0 °C for the T_M . The nearest-neighbor parameters predict ΔG°_{37} , ΔH° , ΔS° , and T_M of oligonucleotides presented in this study with average deviations of 6.3%, 11.0%, 12.2%, and 1.8 °C, respectively, at pH 7.0 and 4.7%, 5.9%, 6.1%, and 1.3 °C, respectively, at pH 5.0. At pH 7.0, the contribution of single A•C mismatches to helix stability ranges from 2.25 kcal/mol for TCA/AAAT to 1.22 kcal/mol for GCG/CAC. At pH 5.0, however, the contribution of A⁺•C mismatches ranges from 1.09 kcal/mol for TCT/AAA to -0.43 kcal/mol for GCC/CAG. Implications of the results for replication fidelity and mismatch repair are discussed.

In addition to canonical Watson–Crick base pairs, there are eight possible single mismatches which occur in DNA with varying frequencies and stabilities, namely A•A, A•C, C•C, C•T, G•G, G•A, G•T, and T•T (1, 2). These mismatches occur as a result of errors made during DNA replication (3) and genetic recombination (4) and also from mutagenic chemicals, ionizing radiation, and spontaneous deamination (1, 5). Mismatches are also present in the secondary structures of several single-stranded DNA viruses (6–8). Efficiency of mismatch recognition and excision by postreplication repair enzymes has been shown to depend on the type of the mismatch and its flanking base pairs (9–13). Therefore, to understand the origins of various mismatch occurrences and to help in our understanding of their repair mechanisms, thermodynamics and structures of DNA mismatches need to be determined.

Several molecular biological techniques require accurate prediction of DNA hybridization to “matched” or “mismatched” sites such as the polymerase chain reaction (PCR) (14), mismatch amplification refractory mutation assay (ARMS) (15), Kunkel mutagenesis (16), antisense oligonucleotide probes (17–19), sequencing by hybridization (20), and oligonucleotide chip arrays as means for biochemical assays (21–24). In each of these techniques, knowledge of the context dependence of DNA mismatch thermodynamics will prove helpful in the design of better and more efficient experimental protocols such as choice of optimal sequence or temperature, pH, and salt conditions (25, 26). Furthermore, knowledge of thermodynamics of all possible base pairings in DNA is an important step toward the development

of single-stranded DNA secondary structure prediction algorithms (27, 28) (N. Peyert, H. T. Allawi, M. Zuker, and J. SantaLucia, Jr., unpublished experiments; see <http://sun2.science.wayne.edu/~jlsun2/servers/dna/form1.cgi>).

Adenine–cytosine mismatches have been shown to have pH-dependent thermodynamics and structure (12, 29–33). At neutral pH, an A•C mismatch can pair with one hydrogen bond involving the cytosine N3 and the adenine 6-amino proton (Figure 1a), or it may exist in low population in the rare enol tautomeric form with two hydrogen bonds (Figure 1b) (5, 29, 30, 33–35). At low pH, A•C mismatches can become protonated at the adenine N1 position and form a pair stabilized by two hydrogen bonds (Figure 1c) (12, 29–31). It has been shown that protonation of A•C mismatches can dramatically increase their stability (31, 36). However, the extent of pH stabilization effect on A•C mismatches in all contexts has not been tested.

We and others showed that the nearest-neighbor model is applicable to DNA duplexes with all canonical Watson–Crick base pairs (37). We also showed that the nearest-neighbor model is applicable to single G•T (25), G•A (38), and C•T mismatches (39) in DNA and derived nearest-neighbor parameters which make accurate predictions. To add to our thermodynamic database and to test the applicability of the nearest-neighbor model to A•C mismatches, we obtained thermodynamic measurements on 27 oligonucleotides with internal A•C mismatches and combined them with four previously reported literature values (36, 40, 41). Previously published Watson–Crick nearest-neighbor thermodynamics (25) and the A•C mismatch nearest-neighbor parameters presented here allow reliable prediction of duplex stability from sequence.

* To whom correspondence should be addressed.

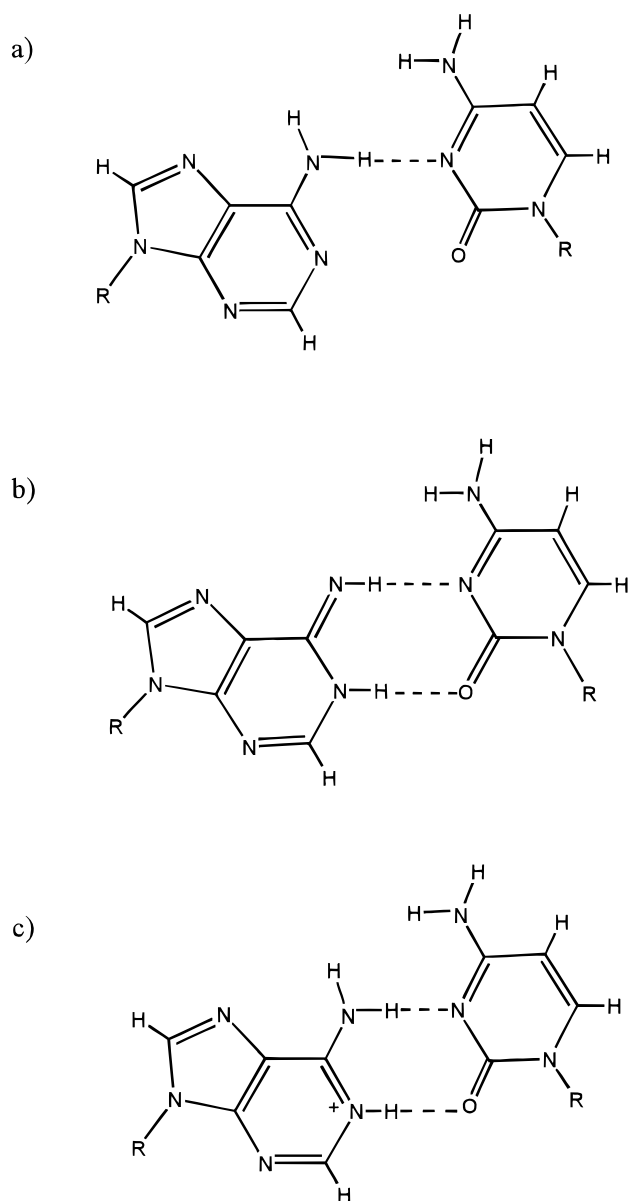


FIGURE 1: Three hydrogen-bonded structures of the A·C mispair at neutral pH (a and b) and at acidic pH (c).

MATERIALS AND METHODS

DNA Synthesis and Purification. DNA oligonucleotides were supplied by Hitachi Chemical Research or synthesized on Cruachem PS 250 DNA/RNA synthesizer using standard phosphoramidite chemistry (42). Upon completion of synthesis, oligonucleotides were incubated for 12 h in concentrated ammonia at 50 °C to remove solid support and blocking groups. Each sample was then evaporated to dryness, and the mixture was dissolved in 250 μ L of water and purified on a Si500F TLC plate (Baker) by eluting for 5–6 h with *n*-propanol/ammonia/water (55:35:10 by volume) (43). The least mobile band was visualized with a UV lamp, cut out, and eluted three times with 3 mL of distilled deionized water. The sample was then evaporated to dryness. Oligonucleotides were further purified and desalted using a Sep-pak C-18 cartridge (Waters). The DNA was eluted with 30% acetonitrile buffered with 10 mM ammonium bicarbonate, pH 7.0. Purities were checked by analytical C-8 HPLC and were greater than 95%.

Melting Curves. The buffer for thermodynamic studies was 1.0 M NaCl, 20 mM sodium cacodylate, 0.5 mM Na₂EDTA (disodium ethylenediaminetetraacetate), adjusted to pH 7.0 or 5.0 with 1 M HCl. Absorbance versus temperature melting profiles were carried out at 280 and 260 nm on an AVIV 14DS UV–vis spectrophotometer from 0 to 85 or 90 °C as described previously (44). Melting curves were obtained with a constant heating rate of 0.8 °C/min. Extinction coefficients for oligonucleotide strands were calculated from dinucleoside monophosphates and nucleotides (45). Prior to the beginning of each melt, non-self-complementary strands were mixed in 1:1 concentration ratios and then annealed by raising the temperature to 85 °C for 5 min and slowly cooling to 0 °C. While at high temperature, absorbances were recorded at 260 nm for calculations of total oligonucleotide strand concentrations (C_T).

Data Analysis. Absorbance versus temperature melting curves were fit with sloping baselines to a two-state model using the program MELWIN 3.0 (46). Thermodynamic parameters for duplex formation were obtained by two methods: (1) enthalpy and entropy changes from fits of individual melting curves were averaged (47), and (2) plots of reciprocal melting temperature (T_M^{-1}) vs natural logarithm of the total strand concentration ($\ln C_T$) were fit to the equation (48):

$$T_M^{-1} = R/\Delta H^\circ \ln(C_T/N) + \Delta S^\circ/\Delta H^\circ \quad (1)$$

For self-complementary sequences, $N = 1$, and for non-self-complementary sequences, $N = 4$. For the two-state model to apply, agreement of the parameters obtained using the two different methods is a necessary, but not sufficient, condition (25, 49).

Design and Choice of Sequence. To get an accurate representation of all possible A·C mismatch contexts, sequences were designed such that all different 16 trimer sequences with an internal A·C mismatch are represented. Since the free energy changes reported in this work are at 37 °C (ΔG_{37}°), duplexes were designed to have melting temperatures between 30 and 55 °C to minimize extrapolation of errors to 37 °C. Furthermore, sequences were designed to minimize the potential of forming alternative competing secondary structures besides the desired duplex (e.g., hairpins or “slipped” duplexes) thus maximizing the likelihood of observing two-state transitions. In addition, duplexes were designed to have terminal G·C base pairs to minimize duplex “fraying”. Throughout this work dimer and trimer nearest-neighbor sequences are represented with a slash separating the two strands in antiparallel orientation and an underline indicating the position of A·C mismatches. For example, the sequence AA/TC means $5'AA^{3'/3'}TC^{3'}$ and AAC/TCG means $5'AAC^{3'/3'}TCG^{3'}$. In this study, the eight different A·C mismatch containing dimers occur with the following frequencies: AA/TC = 9, AC/TA = 10, CA/GC = 12, CC/GA = 9, GA/CC = 11, GC/CA = 8, TA/AC = 7, TC/AA = 12.

Determination of A·C Mismatch Contribution. To obtain the contribution of an A·C mismatch, $\Delta G_{37}^\circ(\text{mismatch})$, to duplex formation, the contributions from Watson–Crick pairs

and helix initiation are subtracted from the measured ΔG_{37}° . For example, the mismatch contribution to the GGACCGACG•CGTCAGTCC duplex is given by:

$$\begin{array}{r} \text{CC} + \text{CA} = 5' \text{-GGACCGACG-3}' - \text{initiation} - \text{GG} - \\ \text{GA} \quad \text{GC} \quad 3' \text{-CCTGACTGC-5}' \quad \text{CC} \\ \text{GA} - \text{AC} - \text{GA} - \text{AC} - \text{CG} \quad (2) \\ \text{CT} \quad \text{TG} \quad \text{CT} \quad \text{TG} \quad \text{GC} \end{array}$$

Using DNA Watson–Crick nearest-neighbor numbers (25) and the measured total free energy change (ΔG_{37}°) for the duplex above (see Table 2), eq 2 gives:

$$\begin{array}{r} \text{CC} + \text{CA} = -6.08 - (1.96) - (-1.84) - (-1.30) - \\ \text{GA} \quad \text{GC} \\ (-1.44) - (-1.30) - (-1.44) - (-2.17) = \\ 1.45 \text{ kcal/mol} = \Delta G_{37}^{\circ}(\text{mismatch}) \quad (3) \end{array}$$

Thus, the two nearest-neighbors CC/GA and CA/GC destabilize the free energy of the duplex by 1.45 kcal/mol. Similar calculations for ΔH° and ΔS° are carried out to obtain ΔH° -(mismatch) and ΔS° -(mismatch).

Error Analysis. Thermodynamic parameters obtained from averages of the fits of individual melting curves and from the T_M^{-1} vs $\ln C_T$ plots are equally reliable (25, 44, 50); thus their averages are considered to be the experimental values. The uncertainty in the experimental thermodynamics is assumed to be 4%, 8%, 8%, and 1 °C in ΔG_{37}° , ΔH° , ΔS° , and T_M , respectively. The small error in ΔG_{37}° (4%) compared to ΔH° and ΔS° is due the high correlation observed between ΔH° and ΔS° ($R^2 > 99\%$) (25, 47, 51). The errors in each A•C-containing dimer or trimer contribution were calculated by propagating the experimental uncertainties in the duplex formation and Watson–Crick nearest-neighbor dimer errors using standard error propagation methods (52) as described previously (25).

Unique versus Nonunique Representations of Internal A•C Mismatches. In DNA, internal mismatches are less stable than terminal mismatches (25, 38) (S. Varma, G. Jenkins, and J. SantaLucia, Jr., unpublished results). Therefore, different sets of nearest-neighbor parameters are needed to describe internal versus terminal mismatches (25). In this study, all A•C mismatches are located in the interior of the duplex, and thus, nearest-neighbor parameters derived from the data are applicable only to internal A•C mismatches. Imposing a restriction on the data set (i.e., all mismatches must be internal) reduces the number of nearest-neighbor parameters that can be uniquely derived from the data from eight to seven (25, 52). Instead of the usual format for nearest neighbors as dimer sequences, internal A•C mismatches can be represented as trimer sequences with the mismatch in the middle position. Of the 16 trimer sequences with an internal A•C mismatch, seven of them can be arbitrarily chosen as “linearly independent” and the other nine contexts can be derived from these seven within the limits if the nearest-neighbor model. For this study, we arbitrarily chose seven trimers with a closing 3' C•G pair as our linearly independent set, which can be used to predict any sequence. For instance, eq 2 can be rewritten as:

$$\begin{array}{r} 5' \text{-GGACCGACG-3}' = \text{initiation} + \text{GG} + \text{GA} + \text{AC} + \\ 3' \text{-CCTGACTGC-5}' \quad \text{CC} \quad \text{CT} \quad \text{TG} \\ \text{CCC} + \text{CAC} - \text{GCC} + \text{GA} + \text{AC} + \text{CG} \quad (4) \\ \text{GAG} \quad \text{GCG} \quad \text{CAG} \quad \text{CT} \quad \text{TG} \quad \text{GC} \end{array}$$

The trimer sequence GCC/CAG was subtracted to account for the extra CC/AG + AC/CG which are not found in the duplex but were added in by the trimer sequences CCC/GAG and CAC/GCG. Note that the linearly independent trimer sequences can also be derived by adding appropriate pairs of the eight A•C nearest-neighbor dimer terms.

Linear Regression Analysis of A•C Mismatch Nearest Neighbors. The ΔG_{37}° (mismatch), ΔH° (mismatch), and ΔS° (mismatch) calculated from all duplexes in this study formed a system of 28 equations (at pH 7.0) and 24 (at pH 5.0) equations with seven unknown A•C mismatch nearest-neighbor parameters. The least-squares solution to these equations was determined by performing multiple linear regression using singular value decomposition (SVD) analysis (53) using the program MATHEMATICA v3.0 (Wolfram research) as described (25, 44). The data in the SVD analysis were weighted by their errors (see above) (53). SVD calculations were performed for ΔG_{37}° , ΔH° , and ΔS° . To verify our results, the SVD solution obtained for entropy nearest-neighbor values was in agreement with those calculated from ΔG_{37}° and ΔH° .

Resampling Analysis of the Data. To assess the quality and confidence level of all experimental measurements obtained in this study and to point out any sequences that are either outliers in the fit or that substantially perturb the solution obtained by SVD, we performed a resampling analysis of the data (25). Since there are more measurements than unknowns, the solution obtained by performing SVD analysis is overdetermined. In this study, we performed 30 resampling trials for ΔG_{37}° , ΔH° , and ΔS° in which nine (for pH 7.0) and eight (for pH 5.0) sequences were randomly removed. In each resampling trial, the number of nonzero singular values was confirmed to be seven, and a solution was obtained. For each A•C mismatch nearest neighbor, the average solution of all 30 resampling trials was computed and standard deviations were determined. The averaged nearest-neighbor parameters calculated from all resampling trials were the same as the solution obtained with all sequences included. The standard deviation among nearest-neighbor solutions for all trials was considered to be the uncertainty in each nearest-neighbor term. The advantage of using resampling analysis to determine uncertainties in the nearest neighbors is that these uncertainties require no assumptions about the magnitude of the experimental errors (25, 54). The standard deviations for the solution obtained for all 30 resampling trials at both pH 7.0 and 5.0 are within round-off of the error obtained from propagation of errors in the variance–covariance matrix of the SVD analysis.

¹H NMR Spectroscopy. Oligomers were dissolved in 90% H₂O and 10% D₂O with 1 M NaCl, 10 mM disodium phosphate, and 0.1 mM Na₂EDTA at pH 7.0 or 5.0. Strand concentrations were between 0.2 and 1.0 mM. ¹H NMR spectra were recorded using a Varian Unity 500-MHz NMR spectrometer. One-dimensional exchangeable proton NMR spectra were recorded at 10 °C using the WATERGATE pulse sequence with “flip-back” pulse to suppress the water peak (55, 56). Spectra were recorded with the carrier placed

at the solvent frequency and with high-power and low-power pulse widths of 10 and 1850 μ s, sweep width of 12 kHz, gradient field strength of 10.0 G/cm, and duration of 1 ms; 512–1024 transients were collected for each spectrum. Data were multiplied by a 4.0-Hz line-broadening exponential function and Fourier transformed by a Silicon Graphics Indigo²Extreme computer with Varian VNMR software. No baseline correction or solvent subtraction was applied. 3-(Trimethylsilyl)propionic-2,2,3,3-*d*₄ acid (TSP) was used as the internal standard for chemical shift reference. 1D-NOE difference spectra were acquired as described above, but with selective decoupling of individual resonances during the 1-s recycle delay. Each resonance was decoupled with a power sufficient to saturate <80% of the signal intensity so that spillover artifacts would be minimized. The spectra were acquired in an interleaved fashion in blocks of 16 scans to minimize subtraction errors due to long-term instrument drift; 3200–6400 scans were collected for each FID.

RESULTS

Thermodynamic Data. Plots of T_M^{-1} vs $\ln C_T$ for all duplexes with two-state transition were linear (correlation coefficient > 99%) over the entire 80–100-fold range in concentration (see Supporting Information). Thermodynamic parameters at pH 7.0 and 5.0 derived from the fits of individual melting curves and from the T_M^{-1} vs $\ln C_T$ plots are listed in Table 1. The model used in this study assumes that the transition from duplex to random coil is two-state and ΔC_p° is zero. Typically, agreement within 15% of the parameters derived from the averages of the fits and the T_M^{-1} vs $\ln C_T$ is assumed to indicate a two-state transition (47, 57). However, this agreement does not necessarily rule out non-two-state behavior (25, 49), and caution is in order whenever a two-state approximation is invoked. In this study, since the seven unknown nearest-neighbor parameters are overdetermined, the validity of two-state approximation is confirmed by comparing thermodynamic contributions of trimer sequences in different duplexes. For duplexes with two-state thermodynamics, the data obtained from the averages of the fits and the T_M^{-1} vs $\ln C_T$ plots are equally reliable (25, 44, 50), and thus their averages are the experimental values listed in Table 2.

Molecules with Non-Two-State Transitions. At pH 7.0, three duplexes listed in Table 1 exhibited ΔH° differences between averages of the fits and T_M^{-1} vs $\ln C_T$ plots greater than 15%. At pH 5.0, however, four duplexes exhibited non-two-state transitions (Table 1). Note that some duplexes which had two-state transitions at pH 7.0 are not two-state at pH 5.0 and vice versa. At pH 7.0, A•C mismatches are destabilizing (see below), and thus it is possible that molecules with non-two-state behavior at pH 7.0 form alternative structures. There is also the possibility, since A•C mismatches are so unstable, that an intermediate internal loop is formed within the duplex. The pK_a of N1 of adenine in an A•C mismatch has been estimated to range between 7.5 and 6.6 (29, 58). Thus depending on the context, at pH 7.0 or 5.0 there is a competition between duplexes where the mismatch is protonated or not which may also result in non-two-state behavior.

Nearest-Neighbor Parameters of A•C Mismatches at pH 7.0. Table 3 lists parameters for seven uniquely determined

trimer sequences with A•C mismatches at pH 7.0 obtained using multiple linear regression analysis of the data (see Materials and Methods). The other nine possible unique trimer contexts are also listed in Table 3 and are linear combinations of the seven uniquely determined trimers. For example, thermodynamic parameters for the trimer sequence AAA/TCT can be obtained using linear combinations of the parameters of AAC/TCG, GAA/CCT, and GAC/CCG ($AAA/TCT = AAC/TCG + GAA/CCT - GAC/CCG$). Table 4 lists nonunique parameters for eight A•C dimer nearest-neighbor sequences. These parameters are the solutions obtained by performing SVD analysis of 27 equations fitted to eight unknowns. The singular matrix obtained in this eight-parameter fit was rank-deficient (rank of 7) indicating that the solution obtained is nonunique and that the data can be described using only seven parameters (see Table 3). However, linear combinations of the parameters in Table 4 can be used to derive the parameters in Table 3, but not vice versa unless an eighth parameter is given. The SVD fit to eight unknowns assumes this eighth parameter to be zero (25, 37). Nevertheless, both sets of parameters are equivalent keeping in mind that they both apply only to internal A•C mismatches. In fact, the parameters in Table 4 give predictions that are the same as those obtained using the parameters listed in Table 3.

Nearest-Neighbor Parameters of A•C Mismatches at pH 5.0. Tables 3 and 4 also list nearest-neighbor parameters for internal A•C mismatches obtained at pH 5.0 using SVD. On average, the pH 5.0 nearest-neighbor numbers are more stable than pH 7.0 parameters by 1.39 kcal/mol. Note that the stabilization effect of protonation of an A•C mismatch varies with context. For example, the largest protonation effect is observed for the trimer ACC/TAG (1.70 kcal/mol for ΔG°_{37}) and the smallest protonation effect is for CCT/GAA (1.05 kcal/mol for ΔG°_{37}).

¹H NMR of Internal A•C and A⁺•C Mismatches. To test the effect of pH change on the overall base pairing in DNA duplexes presented in this study, one-dimensional exchangeable ¹H NMR spectra for six duplexes containing internal A•C mismatches were acquired at pH 7.0 and 5.0. Figures 2 and 3 show representative imino 1D-NMR (9–15 ppm) region of two DNA duplexes with internal A•C mismatches at pH 7.0 and 5.0. Unfortunately, there are no imino protons in either the adenine or cytosine in A•C mismatches in the keto form (Figure 1a), and thus, exchangeable 1D-NMR does not give any resonances that indicate the formation of an A•C mispair. However, 1D-NMR gives a qualitative indication of the effect that internal A•C mismatches have on the entire duplex. Previous structural studies on A•C mismatches have suggested that at neutral pH internal A•C mispairs hydrogen-bond with only one hydrogen bond involving the cytosine N3 and the adenine 6-amino proton (Figure 1a). At lower pH, A⁺•C mismatches form two hydrogen bonds between the adenine NH(1)⁺ and NH₂(6) to the cytosine carbonyl O2 and N3 (12, 29, 30, 32, 33, 59). In this study, the imino proton NMR spectra do not provide direct evidence for hydrogen-bonded A•C mispairs (see Figures 2 and 3). However, all G•C and A•T imino proton pairs can be observed in the regions between 12 and 13 ppm and 13 and 15 ppm, respectively (see Figures 2 and 3). At pH 5.0, A•C mismatches neighboring G•C and A•T imino resonances become more resolved and sharper (Figures 2B and 3B)

Table 1: Thermodynamics of Duplex Formation of Oligonucleotides with Internal A•C Mismatches^a

DNA duplex	1/T _M vs ln C _T parameters				curve fit parameters		
	−ΔG° ₃₇ (kcal/mol)	−ΔH° (kcal/mol)	−ΔS° (eu)	T _M (°C) ^b	−ΔG° ₃₇ (kcal/mol)	−ΔH° (kcal/mol)	−ΔS° (eu)
CGAG <u>C</u> GTCC	6.98 ± 0.01 <i>8.67 ± 0.44</i>	50.5 ± 1.2 <i>67.6 ± 1.9</i>	140.4 ± 3.7 <i>190.0 ± 4.6</i>	45.2 <i>51.4</i>	6.98 ± 0.14 <i>8.70 ± 0.07</i>	57.7 ± 1.4 <i>67.1 ± 1.3</i>	163.7 ± 4.2 <i>188.3 ± 4.3</i>
CGTCC <u>C</u> GTCC	6.59 ± 0.38 <i>7.80 ± 0.57</i>	55.7 ± 1.6 <i>60.4 ± 2.3</i>	158.4 ± 3.9 <i>169.6 ± 5.6</i>	42.2 <i>48.3</i>	6.57 ± 0.07 <i>7.73 ± 0.06</i>	59.9 ± 1.7 <i>61.7 ± 0.9</i>	171.9 ± 5.3 <i>174.0 ± 2.9</i>
CGTGC <u>C</u> TCC	*7.07 ± 0.32 <i>8.26 ± 0.55</i>	44.0 ± 1.0 <i>60.3 ± 2.1</i>	119.0 ± 2.2 <i>167.8 ± 5.1</i>	47.1 <i>50.9</i>	7.07 ± 0.34 <i>8.37 ± 0.11</i>	63.1 ± 3.3 <i>64.2 ± 2.3</i>	180.5 ± 9.5 <i>180.0 ± 7.2</i>
GGAC <u>C</u> CTCG	6.12 ± 0.53 <i>7.33 ± 0.84</i>	45.8 ± 1.8 <i>47.4 ± 2.6</i>	127.9 ± 4.2 <i>129.2 ± 5.7</i>	40.1 <i>48.2</i>	5.99 ± 0.11 <i>7.24 ± 0.04</i>	52.4 ± 1.9 <i>54.1 ± 2.5</i>	149.7 ± 6.0 <i>151.1 ± 8.1</i>
GGAC <u>C</u> GACG	6.12 ± 0.31 <i>6.42 ± 0.36</i>	48.3 ± 1.1 <i>61.0 ± 1.5</i>	135.9 ± 2.7 <i>172.8 ± 3.7</i>	39.9 <i>46.1</i>	6.05 ± 0.08 <i>7.39 ± 0.07</i>	52.4 ± 1.6 <i>56.7 ± 1.6</i>	149.4 ± 5.0 <i>159.0 ± 4.9</i>
GGAG <u>C</u> CACG	6.98 ± 0.19 <i>8.56 ± 0.50</i>	52.6 ± 0.7 <i>64.6 ± 2.0</i>	147.0 ± 1.6 <i>180.7 ± 4.9</i>	44.9 <i>51.5</i>	6.94 ± 0.10 <i>8.64 ± 0.12</i>	59.5 ± 1.7 <i>67.1 ± 3.6</i>	169.3 ± 5.1 <i>188.5 ± 11.3</i>
CATGA <u>A</u> GCTAC	7.25 ± 0.24 <i>8.50 ± 0.09</i>	65.0 ± 1.1 <i>72.7 ± 0.4</i>	186.2 ± 2.7 <i>207.0 ± 1.0</i>	44.7 <i>49.5</i>	7.24 ± 0.09 <i>8.51 ± 0.10</i>	70.7 ± 2.0 <i>77.7 ± 2.5</i>	204.6 ± 6.2 <i>223.1 ± 7.6</i>
CCACAC <u>C</u> AGAG	8.44 ± 0.24 <i>9.67 ± 0.27</i>	68.1 ± 1.0 <i>81.5 ± 1.3</i>	192.3 ± 2.5 <i>231.7 ± 3.4</i>	50.1 <i>53.0</i>	8.54 ± 0.09 <i>9.53 ± 0.05</i>	72.9 ± 1.9 <i>76.8 ± 1.5</i>	207.5 ± 5.8 <i>217.0 ± 4.7</i>
CCACAC <u>G</u> AGAG	7.86 ± 0.20 <i>9.07 ± 0.23</i>	63.9 ± 0.9 <i>72.9 ± 1.0</i>	180.7 ± 2.1 <i>205.8 ± 2.5</i>	48.0 <i>52.1</i>	7.93 ± 0.09 <i>9.13 ± 0.07</i>	67.7 ± 2.8 <i>74.7 ± 1.7</i>	192.8 ± 8.6 <i>211.4 ± 5.1</i>
GAGAG <u>C</u> ACACC	8.11 ± 0.67 <i>9.39 ± 0.30</i>	64.6 ± 2.8 <i>69.9 ± 1.3</i>	182.3 ± 6.8 <i>195.1 ± 3.1</i>	49.2 <i>54.4</i>	8.17 ± 0.12 <i>9.48 ± 0.10</i>	67.3 ± 2.9 <i>72.5 ± 2.2</i>	190.6 ± 9.0 <i>203.2 ± 6.6</i>
GATCA <u>A</u> TGTAC	6.62 ± 0.38 <i>7.47 ± 0.31</i>	61.7 ± 1.7 <i>66.8 ± 1.5</i>	177.5 ± 4.1 <i>191.3 ± 3.5</i>	41.8 <i>45.6</i>	6.51 ± 0.13 <i>7.49 ± 0.09</i>	71.0 ± 2.9 <i>71.0 ± 5.2</i>	208.0 ± 9.4 <i>204.6 ± 16.5</i>
GATCT <u>A</u> TGTAC	6.20 ± 0.55 <i>7.14 ± 0.36</i>	57.0 ± 2.3 <i>66.1 ± 1.6</i>	163.7 ± 5.6 <i>190.1 ± 4.0</i>	39.9 <i>44.0</i>	6.03 ± 0.09 <i>7.17 ± 0.04</i>	66.2 ± 4.4 <i>68.3 ± 2.7</i>	194.0 ± 13.9 <i>197.1 ± 8.7</i>
GTACA <u>A</u> AGATC	5.93 ± 0.17 <i>6.96 ± 0.24</i>	55.4 ± 0.7 <i>63.0 ± 1.0</i>	159.4 ± 1.7 <i>180.6 ± 2.6</i>	38.5 <i>43.4</i>	5.76 ± 0.09 <i>6.95 ± 0.06</i>	63.7 ± 3.9 <i>68.7 ± 3.7</i>	186.8 ± 12.5 <i>199.1 ± 11.8</i>
GTAGC <u>A</u> TCATG	7.75 ± 0.22 <i>9.68 ± 0.44</i>	65.3 ± 0.9 <i>81.2 ± 2.1</i>	185.6 ± 2.3 <i>230.6 ± 5.4</i>	47.2 <i>53.1</i>	7.79 ± 0.09 <i>10.07 ± 0.19</i>	70.9 ± 3.1 <i>93.9 ± 2.6</i>	203.5 ± 9.7 <i>270.3 ± 7.9</i>
GTAGT <u>A</u> ACATG	*6.42 ± 0.25 <i>7.42 ± 0.36</i>	53.2 ± 1.0 <i>65.5 ± 1.7</i>	150.7 ± 2.3 <i>187.1 ± 4.2</i>	41.4 <i>45.5</i>	6.25 ± 0.13 <i>7.46 ± 0.04</i>	67.9 ± 4.9 <i>66.2 ± 1.5</i>	198.7 ± 15.3 <i>189.4 ± 4.7</i>
GTAGT <u>C</u> ACATG	6.24 ± 0.84 <i>7.61 ± 0.21</i>	53.9 ± 3.3 <i>67.8 ± 1.0</i>	153.5 ± 8.0 <i>194.1 ± 2.4</i>	40.3 <i>46.1</i>	6.18 ± 0.10 <i>7.70 ± 0.09</i>	57.2 ± 2.4 <i>72.9 ± 4.8</i>	164.4 ± 7.8 <i>210.1 ± 15.2</i>
CCGACT <u>C</u> TAGCG	6.76 ± 0.41 <i>9.42 ± 0.56</i>	49.4 ± 1.5 <i>70.8 ± 2.4</i>	137.4 ± 3.4 <i>197.9 ± 5.9</i>	44.0 <i>54.3</i>	6.74 ± 0.10 <i>9.57 ± 0.13</i>	54.8 ± 2.4 <i>73.4 ± 2.5</i>	155.0 ± 7.4 <i>205.9 ± 7.6</i>
CGCA <u>A</u> AGAG <u>A</u> CGG	6.33 ± 0.52 <i>9.35 ± 0.24</i>	49.1 ± 1.9 <i>65.1 ± 1.0</i>	138.0 ± 4.3 <i>179.8 ± 2.3</i>	41.2 <i>55.5</i>	6.28 ± 0.09 <i>9.38 ± 0.10</i>	51.6 ± 1.3 <i>66.0 ± 3.5</i>	146.2 ± 4.3 <i>182.5 ± 10.9</i>
GCGCT <u>C</u> TCCGCC	7.00 ± 0.21 <i>9.72 ± 0.25</i>	47.0 ± 0.7 <i>72.8 ± 1.1</i>	129.1 ± 1.6 <i>203.3 ± 2.6</i>	46.0 <i>55.3</i>	6.96 ± 0.08 <i>9.63 ± 0.09</i>	54.1 ± 1.1 <i>69.5 ± 2.5</i>	152.0 ± 3.4 <i>193.1 ± 7.7</i>
GGCC <u>G</u> AGACCGC	7.15 ± 0.21 <i>9.99 ± 0.50</i>	49.3 ± 0.7 <i>75.3 ± 2.2</i>	136.0 ± 1.6 <i>210.4 ± 5.5</i>	46.4 <i>55.8</i>	7.19 ± 0.08 <i>9.98 ± 0.10</i>	54.3 ± 3.8 <i>74.6 ± 2.2</i>	151.9 ± 12.2 <i>208.5 ± 6.9</i>
CGTC <u>G</u> AGGACAACC	7.18 ± 0.46 <i>10.46 ± 0.24</i>	42.5 ± 1.4 <i>80.8 ± 1.1</i>	113.9 ± 2.9 <i>226.8 ± 2.8</i>	48.4 <i>56.5</i>	7.21 ± 0.05 <i>10.02 ± 0.03</i>	42.8 ± 2.1 <i>69.1 ± 3.3</i>	114.7 ± 6.8 <i>190.5 ± 10.7</i>
CGAC <u>C</u> ATATGATCG	6.21 ± 0.90 <i>*6.15 ± 0.31</i>	36.7 ± 2.7 <i>46.4 ± 1.1</i>	98.3 ± 5.7 <i>129.7 ± 2.6</i>	41.6 <i>40.2</i>	6.35 ± 0.14 <i>6.36 ± 0.25</i>	41.5 ± 9.9 <i>37.4 ± 7.6</i>	113.3 ± 20.0 <i>100.0 ± 25.2</i>
CGTCT <u>C</u> ATGAAACG	5.89 ± 0.37 <i>*8.67 ± 0.30</i>	64.4 ± 1.9 <i>92.6 ± 1.8</i>	188.7 ± 5.0 <i>270.6 ± 4.7</i>	38.0 <i>47.4</i>	5.95 ± 0.11 <i>8.16 ± 0.14</i>	58.2 ± 2.1 <i>71.9 ± 2.6</i>	168.6 ± 7.0 <i>205.5 ± 8.3</i>
CTCACATATG <u>C</u> GAG	*7.70 ± 0.26 <i>*10.57 ± 0.99</i>	83.9 ± 1.5 <i>89.6 ± 5.2</i>	245.7 ± 4.1 <i>254.8 ± 9.7</i>	44.7 <i>54.9</i>	7.53 ± 0.09 <i>9.99 ± 0.26</i>	70.3 ± 3.7 <i>75.0 ± 8.7</i>	202.2 ± 12.1 <i>209.6 ± 27.2</i>
CTCCACATGT <u>A</u> GAG	5.92 ± 0.41 <i>*8.07 ± 0.47</i>	61.9 ± 2.0 <i>85.3 ± 2.8</i>	180.5 ± 5.2 <i>249.0 ± 7.4</i>	38.2 <i>46.0</i>	6.02 ± 0.12 <i>7.82 ± 0.10</i>	56.2 ± 3.9 <i>66.1 ± 6.2</i>	161.9 ± 12.8 <i>187.9 ± 20.1</i>
GAG <u>A</u> AGCGGT <u>C</u> CAG	8.97 ± 0.77 <i>11.68 ± 0.24</i>	60.3 ± 2.8 <i>87.6 ± 1.1</i>	165.5 ± 6.7 <i>244.8 ± 2.8</i>	54.9 <i>59.8</i>	8.90 ± 0.10 <i>11.14 ± 0.10</i>	57.8 ± 1.7 <i>78.1 ± 1.7</i>	157.6 ± 5.1 <i>215.9 ± 5.1</i>
GCA <u>A</u> CTCCGG <u>T</u> AG	9.47 ± 0.29 <i>11.73 ± 0.46</i>	69.2 ± 1.2 <i>88.9 ± 2.2</i>	192.6 ± 2.8 <i>248.8 ± 5.6</i>	55.0 <i>59.7</i>	9.36 ± 0.07 <i>11.19 ± 0.11</i>	64.6 ± 3.0 <i>77.5 ± 2.1</i>	178.0 ± 9.5 <i>213.8 ± 6.6</i>

^aListed in alphabetical order and by oligomer length. For each DNA duplex, only the top stand is shown. Underlined residues indicate the position of an A•C mismatch. Molecules that exhibited an agreement in ΔH° with 15% by two different methods are considered to have two-state transitions. Molecules that exhibited a disagreement of ΔH° of more than 15% are considered to be non-two-state and are marked with an asterisk. Solutions are 1 M NaCl, 20 mM sodium cacodylate, 0.5 mM Na₂EDTA, pH 7.0. Values reported in italics were obtained in the same solution conditions as above except at pH 5.0. Errors are standard deviations from the regression analysis of the melting data. Extra significant figures are given to allow accurate calculation of ΔG°₃₇ and T_M. ^bCalculated for 10^{−4} M oligomer concentration for self-complementary sequences and 4 × 10^{−4} M for non-self-complementary sequences.

suggesting that the protonation perturbs ring current shifts. Since protonation of A•C mismatches increases duplex stability (Tables 1 and 2), it is expected that this increase in stability may be reflected in the strength of hydrogen bonds of Watson–Crick base pairs adjacent to the A•C mismatch, and hence reduced exchange broadening is expected.

DISCUSSION

Applicability of the Nearest-Neighbor Model to Internal A•C and A⁺•C Mismatches in DNA. Table 2 compares experimental results of A•C mismatch-containing duplexes with those predicted using the parameters listed in Tables 3 and 4 in conjunction with Watson–Crick nearest-neighbor

Table 4: Nearest-Neighbor Thermodynamics of A•C Mismatches in 1 M NaCl^a

dimer sequence	ΔH° (kcal/mol)		ΔS° (eu)		ΔG°_{37} (kcal/mol)	
	pH 7.0	pH 5.0	pH 7.0	pH 5.0	pH 7.0	pH 5.0
AA/TC	2.3	-0.8	4.6	-3.8	0.88	0.39
A \bar{C} /TA	5.3	-6.3	14.6	-20.2	0.77	-0.02
CA/GC	1.9	-4.2	3.7	-13.6	0.75	0.02
C \bar{C} /GA	0.6	-1.3	-0.6	-4.9	0.79	0.23
GA/CC	5.2	-3.3	14.2	-10.3	0.81	-0.10
G \bar{C} /CA	-0.7	-4.9	-3.8	-14.7	0.47	-0.33
T \bar{A} /AC	3.4	-2.1	8.0	-7.6	0.92	0.26
T \bar{C} /AA	7.6	2.2	20.2	4.8	1.33	0.70

^a These parameters are a linear least-squares fit of the data for a singular matrix with a rank of 7. These parameters make predictions that are the same as those made by the parameters listed in Table 3. Linear combinations of the parameters in this table give the parameters in Table 3. These parameters are not applicable to terminal or penultimate A•C mismatches (25).

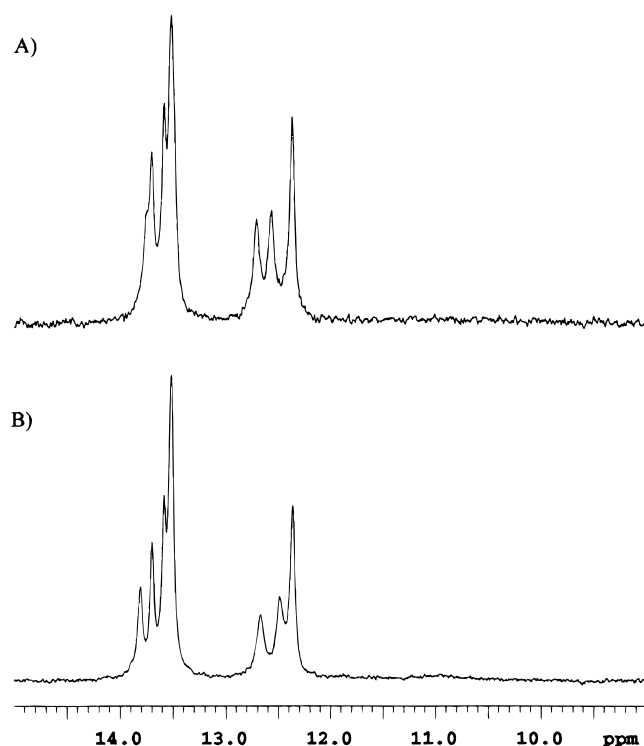


FIGURE 2: 500-MHz ¹H NMR spectra of the exchangeable imino region (9–15 ppm) in 1 M NaCl, 10 mM disodium phosphate, and 0.1 mM Na₂EDTA at 10 °C in 90% H₂O/10% D₂O of GATCAATGTAC•GTACACTGTAC at (a) pH 7.0 and (b) pH 5.0.

respectively (25, 38). On the other hand, for internal C•T mismatches, agreement between experimental and predicted thermodynamics using a nearest-neighbor model is slightly higher than for G•T and G•A mismatches (38) (average deviations for ΔG°_{37} , ΔH° , ΔS° , and T_M of 6.4%, 9.9%, 10.6%, and 1.9 °C, respectively). In this study, we find that at pH 7.0, the A•C mismatch parameters listed in Tables 3 and 4 predict the thermodynamics of all 28 duplexes with two-state transitions listed in Table 2 with average deviations for ΔG°_{37} , ΔH° , ΔS° , and T_M of 6.3%, 11.0%, 12.2%, and 1.8 °C, respectively. Apparently, non-nearest-neighbor interactions are more apparent in weak mismatches than in stable mismatches. At pH 5.0, however, the A⁺•C mismatch parameters listed in Tables 3 and 4 predict all 24 two-state molecules in this study with average deviations for ΔG°_{37} , ΔH° , ΔS° , and T_M of 4.7%, 5.9%, 6.1%, and 1.3 °C,

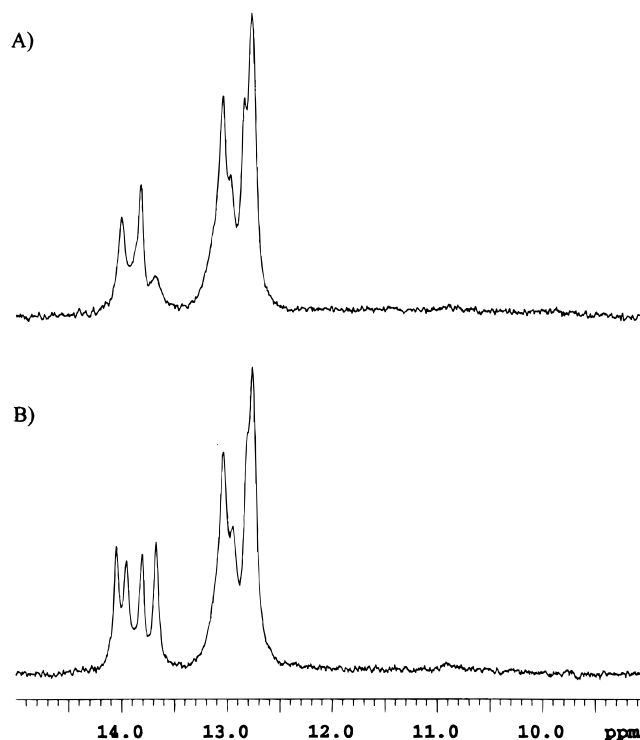


FIGURE 3: 500-MHz ¹H NMR spectra of the exchangeable imino region (9–15 ppm) in 1 M NaCl, 10 mM disodium phosphate, and 0.1 mM Na₂EDTA at 10 °C in 90% H₂O/10% D₂O of GCAACTCCGGCTAG•CTAACCGGAGCTGC at (a) pH 7.0 and (b) pH 5.0.

respectively. Consistent with our previous studies on G•T and G•A mismatches, the quality of predictions made by the parameters listed in Tables 3 and 4 indicates that the nearest-neighbor model is a good approximation for prediction of the thermodynamics of oligonucleotides with either internal A•C or A⁺•C mismatches.

A•C versus A⁺•C Mismatch Nearest Neighbors. At pH 7.0, the average free energy contribution (ΔG°_{37}) of trimer sequences with an internal A•C mismatch (Table 3) is 1.68 kcal/mol, and they vary over a range of 1.03 kcal/mol. The most stable A•C mismatch trimer sequence is GCG/CAC (1.22 kcal/mol). The least stable A•C mismatch trimer, however, is TCA/AAT (2.25 kcal/mol). At pH 5.0, on the other hand, the average contribution of trimer sequences with an A⁺•C mismatch (Table 3) is 0.29 kcal/mol, and they vary over a range of 1.52 kcal/mol. The most stable A⁺•C mismatch trimer sequence is GCC/CAG (-0.43 kcal/mol). The least stable A⁺•C mismatch trimer is TCT/AAA (1.09 kcal/mol). A plot of ΔG°_{37} of A•C versus A⁺•C mismatches reveals that the data can be fit to a line with a correlation coefficient (R^2) of 0.84 (Figure 4). This suggests that A•C and A⁺•C mismatches stack slightly different depending on the neighboring pairs (see below).

Trends in A•C and A⁺•C Mismatch Nearest-Neighbor Parameters. At pH 7.0, when the adenine of an A•C mismatch is on the top strand, the general trend observed for the 5' Watson-Crick neighboring base (with decreasing order of stabilities) is C \approx G \approx A \approx T (Table 4). However, when the cytosine of the A•C mismatch is on the top strand, the trend for the 5' neighboring base becomes G > A \approx C > T. These data suggest adenine stacking with a 5' base pair plays a dominant role in the sequence dependence of

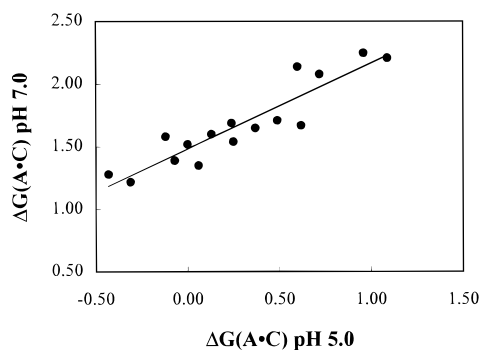


FIGURE 4: Plot of ΔG°_{37} of 16 unique A·C trimer sequences at pH 7.0 vs pH 5.0. The least-squares regression line follows the equation $Y = 0.687X + 1.48$ ($R^2 = 0.84$).

A·C mismatch formation. In addition, since G·C pairs are inherently more stable than A·T pairs (G·C have three hydrogen bonds whereas A·T have two), it is not surprising that 5' G·C or C·G stacked on top of A·C mismatches are more stable than 5' neighboring A·T or T·A. For A⁺·C mismatches, however, the trend for a 5' Watson–Crick neighboring pair when the protonated adenine of the A⁺·C is on the top strand (with decreasing stabilities) is $G > C > T \approx A$. When cytosine of the C·A⁺ is on the top stand, the trend for a 5' Watson–Crick pair becomes $G > A > C > T$. The thermodynamic trends at pH 7 and 5 are slightly different. This suggests that protonation of A·C mismatches results in slightly different stacking geometries and, thus, different stability trends.

Comparison of Thermodynamics of A·C Mismatches and Watson–Crick Pairs. The most stable Watson–Crick trimers with a central G·C or T·A pair (GCG/CGC and CAC/GTG) are -4.41 and -2.89 kcal/mol (25), respectively, and the most stable A·C or A⁺·C trimer sequences are GCC/CAG and GCG/CAC (Table 3). In addition, the least stable Watson–Crick trimers with a central G·C or T·A pair are TCT/AGA and TAT/ATA (ΔG°_{37} of -2.58 and -1.46 kcal/mol, respectively), and the least stable A·C or A⁺·C trimer sequence is TCT/AAA (ΔG°_{37} of 2.21 kcal/mol at pH 7.0 and 1.09 kcal/mol at pH 5.0). While in these two instances the most and least stable A·C mismatch contexts are similar to the corresponding Watson–Crick trimer sequence with a central G·C pair, however, this is not the case for all 16 trimer A·C mismatch sequences, and no obvious correlation is observed (not shown).

Comparison of Thermodynamics of A·C Mismatches with G·T Mismatches. X-ray crystallography studies of DNA duplexes with A·C mismatches suggest that an A⁺·C mismatch adopts a wobble structure similar to the G·T mispair (Figure 1c) (34). It is thus reasonable to predict that A⁺·C and G·T mismatches have similar thermodynamics or that there might be a correlation between the sequence dependence of their thermodynamics. Recently, we determined thermodynamic parameters for all 16 unique trimer sequences with single-internal G·T mismatches in DNA (25). Results from that study indicated that free energy contributions of G·T mismatches range between 1.05 kcal/mol for AGA/TTT and -1.05 kcal/mol for CGC/GTG. Furthermore, the average ΔG°_{37} for trimer sequences with internal G·T mismatches was found to be 0.10 kcal/mol. In this study, A⁺·C mismatch trimer sequences range in stability from 1.09 kcal/mol for TCT/AAA to -0.43 kcal/mol GCC/CAG, and

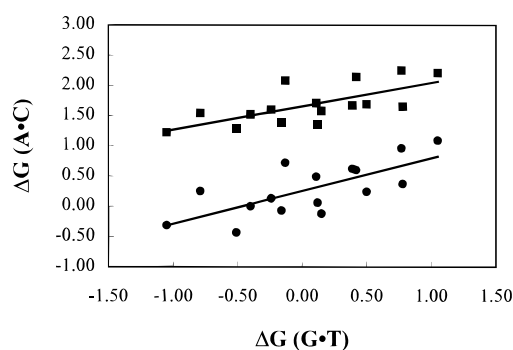


FIGURE 5: Plot of ΔG°_{37} of 16 unique A·C trimer sequences at pH 7.0 (■) and pH 5.0 (●) vs ΔG°_{37} of 16 unique G·T trimer sequences (25). The data at pH 7.0 can be fit to a straight line ($R^2 = 0.48$) using the equation $Y = 0.389X + 1.65$ and at pH 5.0 ($R^2 = 0.52$) using the equation $Y = 0.542X + 0.25$.

the average contribution for all 16 unique trimer sequences is 0.29 kcal/mol. Overall, the most and least stable trimer sequences for G·T, A·C, and A⁺·C have the same Watson–Crick sequence context suggesting that a thermodynamic correlation between G·T mismatches and A·C or A⁺·C mismatches may exist. However, a plot of all 16 unique trimer free energies of A⁺·C or A·C versus the 16 unique trimer free energies of G·T mismatches reveals that only a weak correlation is observed (Figure 5). These data suggest that A·C, A⁺·C, and G·T mismatches stack differently in subtle ways.

Comparison of Thermodynamics of A·C Mismatches with G·A and C·T Mismatches. Previous studies from our lab (38, 39) have shown that average free energy (ΔG°_{37}) contributions of 16 unique trimer sequences with single-internal G·A and C·T mismatches are 0.17 and 1.41 kcal/mol, respectively. The most stable trimer sequence with G·A mismatches is -0.78 kcal/mol for GGC/CAG, and the most stable trimer context for C·T mismatches is 1.02 kcal/mol for GCG/CTC. For A·C and A⁺·C mismatches, the most stable contexts are GCC/CAG and GCG/CAC (ΔG°_{37} of 1.22 for A·C mismatches and -0.43 kcal/mol for A⁺·C mismatches). The least stable trimer context for G·A mismatches is TGA/AAT (ΔG°_{37} of 1.16 kcal/mol), and the least stable trimer context for C·T mismatches is 1.95 kcal/mol for TCC/ATG. The least stable contexts for A·C and A⁺·C mismatches are TCA/AAT (ΔG°_{37} of 2.25 kcal/mol) and TCT/AAA (ΔG°_{37} of 1.09 kcal/mol), respectively. These comparisons lead us to conclude that, in general, for the most stable context a 5' G·C pair is adjacent to the mismatch and for the least stable context a 5' T·A pair is adjacent to the mismatch. Our results indicate that, on average, the trend of stability for internal mismatches in DNA (with decreasing order of stability) is $G \cdot T \geq G \cdot A > A^+ \cdot C > C \cdot T > A \cdot C$ and that the more stable mismatches (G·T, G·A, and A⁺·C) show a larger context dependence than unstable mismatches (A·C and T·C).

Comparison of Mismatch Thermodynamics, Replication Fidelity, and Mismatch Repair. Previous investigators have suggested that DNA replication fidelity may or may not be governed by the thermodynamics of base pairing (1, 60–62). Others have also suggested that replication fidelity is influenced by steric interactions at the site of catalysis (63) or due to nearest-neighbor effects (3). Base insertion experiments of Goodman and co-workers (61, 64) with

Escherichia coli pol III indicate that G•T, G•A, and A•C mismatches occur more often than other mismatches (e.g., A•A, C•C, C•T, G•G, and T•T). G•T and G•A mismatches are generally thermodynamically stabilizing and occur often (61), while this work shows that A•C mismatches can be stabilizing or destabilizing depending on the pH and context. Since A•C mismatches are often incorporated, it is possible that A•C is protonated at the moment of insertion during polymerization. Our thermodynamic data for these mismatches are in general agreement with the pol III misinsertion trend with more thermodynamically stabilizing mismatches (G•T, G•A, and A⁺•C) occurring more often than destabilizing mismatches (A•A, C•C, and C•T). However, thermodynamic data for G•G mismatches (Peyret, Seneviratne, Allawi, and SantaLucia, unpublished) do not agree with this trend since G•G mismatches are stabilizing yet are not efficiently incorporated by DNA polymerases. This could be due to unfavorable steric effects of a purine–purine mismatch in the polymerase active site. It is also plausible that thermodynamics plays an important role regulating the efficiency of mismatch repair enzymes. According to previous studies on mismatch repair (9, 10, 13, 65–69), G•T, A•C, A•A, G•G, and T•T mismatches are well-repaired in *Streptococcus pneumoniae* and *E. coli*, while G•A, C•T, and C•C mismatches are poorly repaired (10, 65). In this case, A•A and G•A mismatch repair efficiencies appear not to follow the thermodynamic trend. This suggests that kinetic or steric factors may play important roles in addition to thermodynamics in mismatch repair efficiency.

ACKNOWLEDGMENT

We thank Hitachi Chemical Research for DNA synthesis and Christine Chow for stimulating discussions and help with DNA synthesis.

SUPPORTING INFORMATION AVAILABLE

Representative plots of two normalized melting curves at pH 7.0 and 5.0 for the duplex GAGAGCACACC•GGTGTACTCTC and two T_M^{-1} vs $\ln C_T$ plots at pH 7.0 and 5.0 for four duplexes (3 pages). Ordering information is given on any current masthead page.

REFERENCES

- Johnson, K. A. (1993) *Annu. Rev. Biochem.* 62, 685–713.
- Gaffney, B. L., and Jones, R. A. (1989) *Biochemistry* 28, 5881–5889.
- Goodman, M. F., Creighton, S., Bloom, L. B., and Petruska, J. (1993) *Crit. Rev. Biochem. Mol. Biol.* 28, 83–126.
- Bhattacharyya, A. (1989) *J. Mol. Biol.* 209, 583–597.
- Brown, T. (1995) *Aldrichim. Acta* 28, 15–20.
- Ikoku, A. S., and Hearst, J. E. (1981) *J. Mol. Biol.* 151, 245–259.
- Morozov, S. Y., Chernov, B. K., Merits, A., and Blinov, V. M. (1994) *J. Biomol. Struct. Dyn.* 11, 837–847.
- Ozawa, K., Kurtzman, G., and Young, N. (1986) *Science* 233, 883–886.
- Kramer, B., Kramer, W., and Fritz, H.-J. (1984) *Cell* 38, 879–887.
- Modrich, P., and Lahue, R. (1996) *Annu. Rev. Biochem.* 65, 101–133.
- Radman, M., and Wagner, R. (1986) *Annu. Rev. Genet.* 20, 523–538.
- Gao, X., and Patel, D. J. (1987) *J. Biol. Chem.* 262, 16973–16984.
- Dohet, C., Wagner, R., and Radman, M. (1985) *Proc. Natl. Acad. Sci. U.S.A.* 82, 503–505.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S., Higuchi, R. H., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) *Science* 239, 487–494.
- Cebula, T. A., Payne, W. L., and Feng, P. (1995) *J. Clin. Microbiol.* 33, 248–250.
- Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol.* 154, 367–382.
- Freier, S. M. (1993) in *Antisense Research and Applications* (Crooke, S. T., and Lebleu, B., Eds.) pp 67–82, CRC Press, Boca Raton, FL.
- Zon, G. (1989) in *Oligonucleotides* (Cohen, J. S., Ed.) pp 233–249, CRC Press, Boca Raton, FL.
- Symons, R. H. (1989) *Nucleic Acids Probes*, CRC Press, Boca Raton, FL.
- Parinov, S., Barsky, V., Yershov, G., Kirillov, E., Timofeev, E., Belgovskiy, A., and Mirzabekov, A. (1996) *Nucleic Acids Res.* 24, 2998–3004.
- Fodor, S. P. A., Rava, R. P., Huang, X. C., Pease, A. C., Holmes, C. P., and Adams, C. L. (1993) *Nature* 364, 555–556.
- Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S., and Fodor, S. P. (1996) *Science* 274, 610–614.
- Pease, A. C., Solas, D., Sullivan, E. J., Cronin, M. T., Holmes, C. P., and Fodor, S. P. A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* 91, 5022–5026.
- Sapolsky, R. J., and Lipshutz, R. J. (1996) *Genomics* 33, 445–456.
- Allawi, H. T., and SantaLucia, J., Jr. (1997) *Biochemistry* 36, 10581–10594.
- Steger, G. (1994) *Nucleic Acids Res.* 22, 2760–2768.
- Zuker, M. (1989) *Science* 244, 48–52.
- Zuker, M., Jaeger, J. A., and Turner, D. H. (1991) *Nucleic Acids Res.* 19, 2707–2714.
- Boulard, Y., Cognet, J. A. H., Gabarro-Arpa, J., Bret, M. L., Sowers, L. C., and Fazakerley, G. V. (1992) *Nucleic Acids Res.* 20, 1933–1941.
- Boulard, Y., Cognet, J. A. H., Gabarro-Arpa, J., Bret, M. L., Carbonnaux, C., and Fazakerley, G. V. (1995) *J. Mol. Biol.* 246, 194–208.
- Brown, T., Leonard, G. A., and Booth, E. D. (1990) *J. Mol. Biol.* 212, 437–440.
- Patel, D. J., Kozlowski, S. A., Ikuta, S., and Itakura, K. (1984) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* 43, 2663–2670.
- Hunter, W. N., Brown, T., and Kennard, O. (1987) *Nucleic Acids Res.* 15, 6589–6606.
- Hunter, W. N., Brown, T., Anand, N. N., and Kennard, O. (1986) *Nature* 320, 552–555.
- Brown, T., Hunter, W. N., and Leonard, G. A. (1993) *Chem. Brit.* 6, 484–488.
- Leonard, G. A., Booth, E. D., and Brown, T. (1990) *Nucleic Acids Res.* 18, 5617–5623.
- SantaLucia, J., Jr. (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95, 1460–1465.
- Allawi, H. T., and SantaLucia, J., Jr. (1998) *Biochemistry* 37, 2170–2179.
- Allawi, H. T., and SantaLucia, J., Jr. (1998) *Nucleic Acids Res.* (in press).
- Aboul-ela, F., Koh, D., Tinoco, I., Jr., and Martin, F. H. (1985) *Nucleic Acids Res.* 13, 4811–4824.
- Tibanyenda, N., De Bruin, S. H., Haasnoot, C. A. G., van der Marel, G. A., van Boom, J. H., and Hilbers, C. W. (1984) *Eur. J. Biochem.* 139, 19–27.
- Brown, T., and Brown, D. J. S. (1991) in *Oligonucleotides and Analogues* (Eckstein, F., Ed.) pp 1–24, IRL Press, New York.
- Chou, S.-H., Flynn, P., and Reid, B. (1989) *Biochemistry* 28, 2422–2435.
- SantaLucia, J., Jr., Allawi, H. T., and Seneviratne, P. A. (1996) *Biochemistry* 35, 3555–3562.

45. Richards, E. G. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) CRC Press, Cleveland, OH.
46. McDowell, J. A., and Turner, D. H. (1996) *Biochemistry* 35, 14077–14089.
47. Petersheim, M., and Turner, D. H. (1983) *Biochemistry* 22, 256–263.
48. Borer, P. N., Dengler, B., Tinoco, I., Jr., and Uhlenbeck, O. C. (1974) *J. Mol. Biol.* 86, 843–853.
49. Marky, L. A., and Breslauer, K. J. (1987) *Biopolymers* 26, 1601–1620.
50. Sugimoto, N., Nakano, S., Yoneyama, M., and Honda, K. (1996) *Nucleic Acids Res.* 24, 4501–4505.
51. SantaLucia, J., Jr., Kierzek, R., and Turner, D. (1991) *J. Am. Chem. Soc.* 113, 4313–4322.
52. Gray, D. M. (1997) *Biopolymers* 42, 783–793.
53. Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1989) *Numerical Recipes*, Cambridge University Press, New York.
54. Efron, B., and Tibshirani, R. (1993) *An Introduction to the Bootstrap*, Chapman & Hall, London.
55. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR* 2, 661–665.
56. Lippens, G., Dhalluin, C., and Wieruszski, J.-M. (1995) *J. Biomol. NMR* 5, 327–331.
57. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., and Turner, D. H. (1986) *Proc. Natl. Acad. Sci. U.S.A.* 83, 9373–9377.
58. Wang, C., Gao, H., Gaffney, B. L., and Jones, R. A. (1991) *J. Am. Chem. Soc.* 113, 5486–5488.
59. Patel, D. J., Kozlowski, S. A., Ikuta, S., and Itakura, K. (1984) *Biochemistry* 23, 3218–3226.
60. Mendelman, L. V., Boosalis, M. S., Petruska, J., and Goodman, M. F. (1989) *J. Biol. Chem.* 264, 14415–14423.
61. Echols, H., and Goodman, M. F. (1991) *Annu. Rev. Biochem.* 60, 477–511.
62. Petruska, J., Goodman, M. F., Boosalis, M. S., Sowers, L. C., Cheong, C., and Tinoco, I., Jr. (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85, 6252–6256.
63. Wang, S., Friedman, A. E., and Kool, E. T. (1995) *Biochemistry* 34, 9774–9784.
64. Sloane, D. L., Goodman, M. F., and Echols, H. (1998) *Nucleic Acids Res.* 16, 6465–6475.
65. Modrich, P. (1987) *Annu. Rev. Biochem.* 56, 435–466.
66. Lu, A.-L., Clark, S., and Modrich, P. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80, 4639–4643.
67. Wagner, R., Dohet, C., Jones, M., Doutriaux, M.-P., Hutchinson, F., and Radman, M. (1984) *Cold Spring Harbor Symp. Quant. Biol.* 49, 611–615.
68. Lacks, S. A., Dunn, J., and Greenberg, B. (1982) *Cell* 31, 327–336.
69. Claverys, J.-P., Mejean, V., Gasc, A.-M., and Sicard, A. M. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80, 5956–5960.

BI9803729