

Nearest Neighbor Thermodynamic Parameters for Internal G•A Mismatches in DNA

Hatim T. Allawi and John SantaLucia, Jr.*

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

Received October 8, 1997; Revised Manuscript Received December 3, 1997

ABSTRACT: Thermodynamics of 22 oligonucleotides with internal single G•A mismatches dissolved in 1 M NaCl were determined from absorbance versus temperature melting curves. These data, combined with five literature sequences, were used to derive nearest-neighbor thermodynamic parameters for seven linearly independent trimer sequences with internal G•A mismatches and Watson–Crick flanking base pairs. The G•A mismatch parameters predict ΔG_{37}° , ΔH° , ΔS° , and T_M with average deviations of 4.4%, 7.4%, 8.0%, and 1.5 °C, respectively. The nearest-neighbor parameters show that G•A mismatch stability is strongly context dependent, and ΔG_{37}° ranges from +1.16 kcal/mol for TGA/AAAT to –0.78 kcal/mol for GGC/CAG. In addition, one-dimensional ^1H NMR spectra show that the G•A pairing geometry is pH and context dependent.

Mismatches in DNA occur as a result of misincorporation of bases during DNA replication (1), recombination (2), and from mutagenic chemicals and ionizing radiation (3). In addition to Watson–Crick base pairs there are eight possible mispairs: A•A, A•C, C•C, C•T, G•G, G•A, G•T, and T•T. Repair of these mismatches requires their recognition by proofreading enzymes or by postreplication repair systems. Accurate knowledge of mismatch stability and thermodynamics will help in our understanding of these repair processes (4–9).

Accurate prediction of oligonucleotide hybridization thermodynamics to “matched” versus “mismatched” sites of a target DNA is helpful for several molecular biological techniques (4, 10–14) such as polymerase chain reaction (PCR) (15), sequencing by hybridization (16), gene diagnostics (17, 18), and specific probes for infectious agents (19). Thus, studies on structure and stability of mismatches in nucleic acids are necessary to minimize the chance of incorrect base pairing that can lead to amplification or detection of wrong sequences (13, 20, 21).

Previous studies have reported that repair of G•A mismatches is less efficient than all other mismatches in DNA (22). Thereafter, G•A mismatches have received considerable attention from the scientific community in terms of both their thermodynamics (4, 13, 14, 23) and structure (24–37). G•A mismatches, along with G•T and G•G mismatches, are among the most stable mismatches in DNA (4, 13, 14, 23, 38). To explain this unusual stability of G•A mismatches, several studies aimed to solve the structure of DNA oligonucleotides containing G•A mismatches. In X-ray structures, three geometries of single G•A mismatches have been observed (Figure 1): G(anti)•A(anti) (29), G(anti)•A(syn) (25, 26, 28), and G(syn)•A⁺(anti) (32, 33). In solution, it has been shown that single G•A mismatches usually adopt a G(anti)•A(anti) conformation at neutral pH (13, 24, 27, 34–36) and a G(syn)•A⁺(anti) conformation at pH ranges

of 4.0 to 5.5 (37). For tandem G•A mismatches in DNA, “sheared” G(anti)•A(anti) structure is often observed (13, 30, 39–42). In RNA, however, the structure and thermodynamics of tandem G•A mismatches were found to depend largely on neighboring Watson–Crick pairs (43–47).

Recently, we showed that a nearest-neighbor model is sufficient to accurately predict the thermodynamics of DNA duplexes composed of all Watson–Crick base pairs in addition to DNA duplexes containing internal G•T mismatches (38, 48). To expand our knowledge to include single G•A mismatches, we decided to test the applicability of the nearest-neighbor model to internal single G•A pairs and try to obtain nearest-neighbor thermodynamic parameters for G•A mismatches in DNA. Previous work on RNA suggested that thermodynamics of single G•A mismatches exhibit next-nearest-neighbor effects (49). Therefore, to test whether the nearest-neighbor model applies to single G•A mismatches in DNA, we obtained thermodynamic measurements of 22 oligonucleotides containing single internal G•A mismatches and combined them with five literature values to derive G•A mismatch nearest-neighbor parameters in 1 M NaCl, pH 7.0 buffer. To determine the effect of pH on the thermodynamics of G•A mismatches, the thermodynamics of several oligonucleotides containing G•A mismatches were determined at pH 5.0 and compared with those obtained at pH 7.0. Exchangeable one-dimensional proton NMR spectra at pH 7.0 and 5.0 were also used to evaluate the effects of pH on the structure of single G•A mismatches.

MATERIALS AND METHODS

DNA Synthesis and Purification. DNA oligonucleotides were supplied by Hitachi Chemical Research and were synthesized on solid support using standard phosphoramidite chemistry (50). 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, dissolved in 250 mL

* To whom correspondence should be addressed. Phone: (313) 577-0101. FAX: (313) 577-8822. E-mail: jsl@chem.wayne.edu.

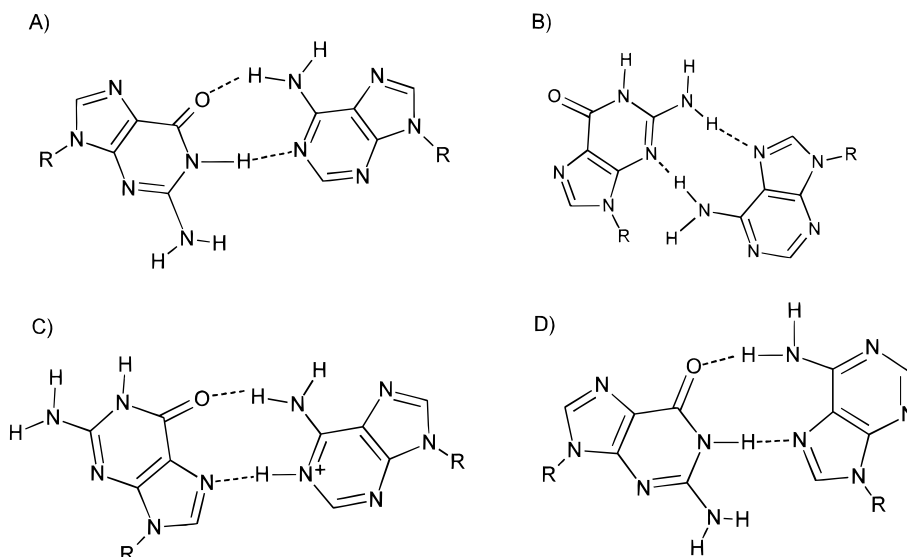


FIGURE 1: Four hydrogen-bonded structures of the G•A mismatch: (A) G(anti)•A(anti), (B) "sheared" G(anti)•A(anti), (C) G(syn)•A⁺(anti), and (D) G(anti)•A(syn).

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) (51). 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 (Perceptive Biosystems) and were greater than 95%.

Melting Curves. Absorbance versus temperature melting curves were carried out on an AVIV 14DS UV–vis spectrophotometer equipped with a five-cell rotor as described previously (48). For each duplex, 8–10 different concentrations were used to carry out melting curve experiments. The data were collected at 280 or 260 nm with a heating rate of 0.8 °C min⁻¹ in a 1 M NaCl, 20 mM sodium cacodylate, and 0.5 mM Na₂EDTA (disodium ethylenediaminetetraacetate), pH 7.0 or 5.0 buffer. Prior to the beginning of each melt, total strand concentrations (C_T) were determined using high-temperature (85 °C) absorbances at 260 nm and extinction coefficients were calculated from dinucleoside monophosphates and nucleotides (52).

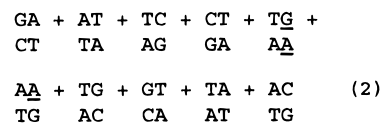
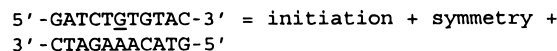
Data Analysis. Thermodynamic parameters for duplex formation were obtained from absorbance versus temperature profiles using the program MELTWIN v2.1 (53) by two methods: (i) averages of ΔH° and ΔS° from fits of individual melting curves as described (54) and (ii) plots of reciprocal melting temperature versus natural log of C_T according to the equation (55)

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

For self-complementary sequences, $N = 1$, and for nonself-complementary sequences, $N = 4$. Both methods assume a two-state model (i.e., duplex and random coil) and that the difference in heat capacity, ΔC_p° , of these states is zero (54, 56). For the two-state model to apply, agreement of the parameters derived using the two methods is a necessary, but not a sufficient criterion (38, 57).

Design of Sequences. Sequences were designed to provide a uniform representation of all eight different G•A mismatch dimer nearest-neighbors. In addition, duplexes were selected to have a melting temperature (T_M) between 30 and 60 °C and to minimize the likelihood of forming alternative structures besides the desired duplex. Throughout this paper, nearest-neighbor base pairs are represented with a slash separating strands in antiparallel orientation and the mismatched residues are underlined (e.g. AG/TA means 5'AG3' paired with 3'TA5'). The eight G•A nearest-neighbor dimers represented in this study occur with the following frequencies: AA/TG = 7, AG/TA = 7, CA/GG = 13, CG/GA = 10, GA/CG = 11, GG/CA = 11, TA/AG = 8, and TG/AA = 11. In addition, all 16 Watson–Crick trimer contexts are represented at least once in the data set.

Determination of the G•A Mismatch Contribution to the Helix Stability. The total ΔG_{37}° obtained from UV melting curves for a duplex containing a G•A mismatch corresponds to strands going from random coil state to duplex state. According to the nearest-neighbor model, the total ΔG_{37}° can be described as the sum of ΔG_{37}° increments for helix initiation, helix symmetry, and nearest-neighbor interactions between base pairs (48, 56). Thus, knowing the contributions of helix initiation, symmetry, and Watson–Crick base pairs, which have been previously determined for DNA (38), allows the determination of the mismatch contribution to the total free energy of each duplex. For example, see eq 2. Thus,

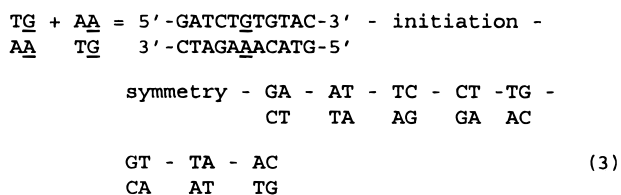


to obtain the contribution of the G•A mismatch to the duplex formation, $\Delta G_{37}^\circ(\text{mismatch})$, the contributions from Watson–Crick pairs, helix initiation, and symmetry (for self-complementary strands) are subtracted from the measured ΔG_{37}° . In the above example, eq 2 can be rearranged to obtain the contribution of the mismatch (eq 3). Using DNA

Table 1: Thermodynamics of Duplex Formation of Oligonucleotides with G·A 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)
Molecules with Two-State Transitions							
GGAC <u>ACT</u> CG	8.25 ± 0.65	53.8 ± 2.3	146.9 ± 5.2	52.6	8.39 ± 0.17	61.1 ± 1.7	169.9 ± 5.7
GGAC <u>AG</u> ACG	7.60 ± 0.58	50.5 ± 2.0	138.4 ± 4.6	49.3	7.70 ± 0.17	58.7 ± 1.2	164.4 ± 3.4
	(8.07 ± 0.91)	(64.9 ± 4.0)	(183.4 ± 9.9)	(48.6)	(8.05 ± 0.20)	(62.6 ± 3.6)	(175.9 ± 11.1)
GGAC <u>GCT</u> CG	7.46 ± 0.69	51.2 ± 2.4	141.0 ± 5.7	48.2	7.53 ± 0.14	58.2 ± 1.0	163.3 ± 3.2
	(7.63 ± 0.23)	(57.3 ± 0.9)	(160.3 ± 2.1)	(47.7)	(7.64 ± 0.04)	(56.1 ± 2.8)	(156.3 ± 9.3)
GGAC <u>GG</u> ACG	7.38 ± 0.51	51.6 ± 1.8	142.6 ± 4.1	47.6	7.38 ± 0.09	53.8 ± 1.0	149.8 ± 3.3
GGAG <u>G</u> CACG	8.86 ± 0.64	54.1 ± 2.2	145.9 ± 4.9	56.4	9.12 ± 0.24	61.1 ± 3.7	167.7 ± 5.8
CATG <u>A</u> AGCTAC	8.51 ± 0.75	75.2 ± 3.5	215.0 ± 8.9	49.2	8.41 ± 0.11	67.0 ± 1.0	188.8 ± 3.2
CATG <u>A</u> GGCTAC	9.25 ± 0.31	71.1 ± 1.3	199.5 ± 3.3	53.4	9.26 ± 0.06	71.4 ± 1.3	200.2 ± 3.9
CATG <u>T</u> A <u>ACT</u> AC	7.13 ± 0.58	56.3 ± 2.3	158.6 ± 5.5	45.2	7.14 ± 0.07	60.6 ± 3.7	172.3 ± 9.7
	(7.45 ± 0.31)	(61.3 ± 1.3)	(173.5 ± 3.2)	(46.5)	(7.54 ± 0.18)	(69.3 ± 5.7)	(199.2 ± 17.9)
GATCAATG <u>T</u> AC	8.08 ± 0.21	64.4 ± 0.9	181.7 ± 2.2	49.0	8.20 ± 0.10	73.1 ± 2.1	209.1 ± 6.4
GATCT <u>A</u> TG <u>T</u> AC	7.73 ± 0.42	62.8 ± 1.7	177.4 ± 4.2	47.5	7.78 ± 0.11	69.5 ± 3.0	199.0 ± 9.3
GATCTG <u>T</u> G <u>T</u> AC	6.87 ± 0.36	57.7 ± 1.5	163.7 ± 3.6	43.6	6.84 ± 0.10	65.9 ± 2.9	190.4 ± 8.9
CCGACTCTAG <u>C</u> G	10.06 ± 0.59	63.1 ± 2.1	170.9 ± 5.0	60.2	10.23 ± 0.16	66.4 ± 2.4	181.1 ± 7.4
CGA <u>G</u> CATG <u>A</u> T <u>C</u> G	8.80 ± 0.81	63.7 ± 3.4	177.1 ± 8.3	53.0	8.75 ± 0.24	60.7 ± 7.2	167.4 ± 9.3
	(9.30 ± 0.53)	(78.1 ± 2.6)	(221.7 ± 6.8)	(52.3)	(8.82 ± 0.15)	(65.0 ± 3.2)	(181.1 ± 10.1)
CGCAAGAGAC <u>G</u> CG	8.94 ± 0.81	47.2 ± 2.3	123.3 ± 4.8	60.1	9.19 ± 0.25	54.6 ± 2.9	146.5 ± 8.6
CGT <u>G</u> GAC <u>C</u> A <u>C</u> C	7.46 ± 0.80	53.4 ± 3.7	148.1 ± 8.7	47.7	7.47 ± 0.13	50.6 ± 2.9	138.9 ± 9.2
CTC <u>A</u> CATG <u>G</u> GAG	8.48 ± 0.74	64.7 ± 3.2	181.4 ± 7.9	51.0	8.25 ± 0.13	55.0 ± 6.2	150.8 ± 9.7
CTCGACG <u>T</u> AGAG	6.96 ± 0.36	76.6 ± 2.0	224.6 ± 5.4	42.3	6.93 ± 0.08	66.9 ± 3.1	193.3 ± 10.0
GAG <u>A</u> AC <u>T</u> G <u>C</u> AG	7.31 ± 0.31	59.6 ± 1.2	168.6 ± 3.0	45.8	7.35 ± 0.08	53.9 ± 1.5	150.1 ± 5.0
GAGGACCTACAG	7.81 ± 0.53	49.5 ± 1.7	134.3 ± 3.8	51.0	7.85 ± 0.03	47.5 ± 4.4	127.8 ± 8.9
GCA <u>A</u> CTCG <u>G</u> TAG	8.75 ± 0.48	59.5 ± 1.8	163.8 ± 4.1	53.9	8.82 ± 0.10	61.9 ± 3.0	171.2 ± 9.4
GCG <u>A</u> TCTCAG <u>C</u> C	9.25 ± 0.38	55.3 ± 1.3	148.4 ± 2.8	58.5	9.49 ± 0.18	60.5 ± 3.1	164.6 ± 9.3
GGC <u>A</u> GAG <u>A</u> AC <u>G</u> C	10.65 ± 0.73	74.5 ± 3.0	205.9 ± 7.3	59.2	10.57 ± 0.10	72.4 ± 1.6	199.4 ± 4.9
Molecule with Anomalous Two-State Transitions							
ATG <u>A</u> G <u>C</u> GCAT	6.45 ± 0.36	34.1 ± 1.0	89.2 ± 2.0	44.2	6.48 ± 0.10	36.3 ± 3.3	96.1 ± 0.5
Molecules with Non-Two-State Transitions							
GGAGAC <u>C</u> G	9.11 ± 0.59	55.9 ± 2.0	151.0 ± 4.6	57.3	9.49 ± 0.25	66.8 ± 2.5	184.7 ± 7.4
ATG <u>A</u> G <u>C</u> TAA <u>T</u>	4.57 ± 0.65	23.8 ± 1.4	62.1 ± 2.5	23.3	3.88 ± 0.14	34.7 ± 5.9	99.3 ± 19.0
CAT <u>G</u> TG <u>A</u> C <u>T</u> AC	7.05 ± 1.40	51.3 ± 5.2	142.6 ± 12.1	45.6	7.07 ± 0.18	60.8 ± 2.8	173.3 ± 8.9
CGT <u>G</u> T <u>C</u> GAA <u>A</u> C <u>G</u>	5.99 ± 0.46	48.1 ± 1.8	135.9 ± 4.2	39.0	5.98 ± 0.10	50.2 ± 5.4	142.5 ± 17.7

^a Listed in alphabetical order and by oligomer length. For each DNA duplex only the top strand is shown. Underlined residues indicate the position of a G·A mismatch. Molecules listed as two-state had ΔH° agreement within 15% by two different methods. Molecules listed as anomalous two-state showed two-state transition but had unusual thermodynamics (see text). Molecules listed as non-two-state had ΔH° disagreement of more than 15% by two different methods. Solutions are 1 M NaCl, 20 mM sodium cacodylate, 0.5 mM Na₂EDTA, pH 7.0. Values reported in parentheses 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. ΔS° values are in eu (entropy units). ^b Calculated for 10⁻⁴ M oligomer concentration for self-complementary sequences and 4 × 10⁻⁴ M for non-self-complementary sequences.



Watson-Crick nearest-neighbor numbers (38) and the measured total free energy change (ΔG°₃₇) (Table 2) for the duplex above, eq 3 simplifies to eq 4. Thus, the two nearest-

$$\begin{array}{l}
 \text{TG} + \text{AA} = -6.85 - (1.96) - (0) - (-1.30) - \\
 \text{AA} \quad \text{TG} \\
 \\
 (-0.88) - (-1.30) - (-1.28) - \\
 (-1.45) - (-1.44) - (-0.58) - \\
 (-1.44) \\
 \\
 = 0.86 \text{ kcal/mol} = \Delta G^{\circ}_{37}(\text{mismatch})
 \end{array} \quad (4)$$

neighbors TG/AA and AA/TG destabilize the free energy of the duplex by 0.86 kcal/mol. Similar calculations for ΔH° and ΔS° are carried out to calculate ΔH°(mismatch) and ΔS°(mismatch).

Another method to calculate the mismatch contribution is to measure the thermodynamics of a “core sequence”, subtracting its thermodynamics from the sequence containing the mismatch and adding back the nearest neighbor that is interrupted by the mismatch (58). The two methods are equally reliable (38) due to propagation of errors from both the experimental measurement and Watson-Crick nearest-neighbor parameters. In this study, we used the former method because measurement of the core sequence thermodynamics is not required.

Analysis of G·A mismatches in Terms of Linearly Independent Sequences. For mismatches in DNA, a mismatch located at the end of a duplex is usually more stabilizing than a mismatch located inside the duplex (38), and thus terminal and internal G·A mismatches should be characterized separately. For internal G·A mismatches, a result of making all sequences contain internal mismatches is that the eight single G·A mismatch containing dimer sequences cannot be uniquely determined (59), hence, internal G·A mismatches should be expressed in terms of seven linearly independent trimer sequences with the mismatch in the middle position (38). These seven linearly independent

Table 2: Experimental and Predicted Thermodynamics of Oligonucleotides with G•A Mismatches^a

DNA duplex	ref ^b	$-\Delta G_{37}^{\circ}$ (kcal/mol) ^c		$-\Delta H^{\circ}$ (kcal/mol) ^c		$-\Delta S^{\circ}$ (eu) ^c		T_M (°C) ^d	
		expt	pred	expt	pred	expt	pred	expt	pred
Molecules with Two-State Transitions									
CAAAAAAAG	4	3.92	3.89	39.9	45.3	116.0	133.0	23.9	26.3
CAAA <u>G</u> AAAG	4	4.22	4.33	52.6	47.7	156.0	139.7	28.6	28.7
GGAC <u>A</u> CTCG		8.32	7.86	57.4	51.2	158.4	139.6	52.0	51.1
GGAC <u>A</u> GACG		7.65	7.39	54.6	56.3	151.4	157.4	48.6	47.3
GGAC <u>G</u> CTCG		7.49	7.51	54.7	55.6	152.2	154.7	47.7	48.2
GGAC <u>G</u> GACG		7.38	7.39	52.7	56.3	146.2	157.4	47.4	47.3
GGAG <u>G</u> CACG		8.99	8.29	57.6	51.4	156.8	138.8	55.9	54.0
CATGAAGCTAC		8.46	8.41	71.1	70.0	201.9	198.2	49.6	50.2
CATGA <u>G</u> GCTAC		9.25	8.61	71.2	66.8	199.9	187.3	53.4	51.8
CATGTAACTAC		7.14	6.84	58.5	60.5	165.5	172.8	44.9	43.4
GATCAAATGTAC		8.14	7.72	68.7	65.7	195.4	186.5	48.5	47.6
GATCTATGTAC		7.75	7.27	66.1	63.7	188.2	181.8	47.1	45.2
GATCT <u>T</u> GTAC		6.85	6.83	61.8	61.3	177.1	175.1	43.0	43.8
CCATCGCTACC	68	9.49	8.79	74.5	67.7	209.4	189.6	53.7	52.5
CCATT <u>G</u> CTACC	68	8.18	7.86	67.1	60.4	189.8	169.0	49.0	49.3
CCGACTCTAGCG		10.15	10.56	64.7	66.0	176.0	178.8	60.0	61.7
CGAGCATGATCG		8.77	8.82	62.2	64.2	172.3	178.2	53.2	53.6
CGCAAATT <u>G</u> GCG	33	8.00	7.80	66.9	59.0	189.9	164.8	48.2	49.1
CGCAAAGAGACCG		9.06	10.32	50.9	60.4	134.9	161.2	59.1	63.3
CGT <u>G</u> GACCAACC		7.46	7.97	52.0	55.2	143.5	151.8	48.0	51.4
CTCACATG <u>G</u> GAG		8.36	7.56	59.9	56.4	166.1	157.4	51.6	47.8
CTC <u>G</u> ACGTAGAG		6.94	6.79	71.7	65.8	208.9	190.4	42.6	42.1
GAGAAACCTGCAG		7.33	6.93	56.7	51.5	159.3	143.0	46.3	46.1
GAG <u>G</u> ACCTACAG		7.83	8.11	48.5	53.9	131.0	148.0	51.4	51.0
GCAACTCGGTAG		8.78	9.10	60.7	59.9	167.5	163.6	53.7	56.2
GCGATCTCAGCC		9.37	10.50	57.9	68.2	156.5	185.7	58.1	61.2
GGC <u>A</u> GAGA <u>A</u> CGC		10.61	10.65	73.5	67.0	202.6	181.5	59.3	62.2
Molecules with Anomalous Two-State Transitions									
ATGAGCGCAT		6.46	4.72	35.2	45.8	92.7	132.2	44.1	31.2
ATG <u>A</u> GC <u>G</u> CAT	13	6.69	4.72	46.7	45.8	129.0	132.2	43.9	31.2
Molecules with Non-Two-State Transitions									
GGAGACACG		9.30	8.29	61.4	51.4	167.8	138.6	56.5	54.4
ATGAGCTAAT		4.23	1.78	29.3	27.8	80.7	84.2	22.4	-1.9
AT <u>G</u> AGCTAAT	13	3.43	1.78	36.0	27.8	105.0	84.2	18.8	-1.9
CATGTGACTAC		7.06	6.84	56.1	60.5	158.0	172.8	44.9	43.4
CGTGT <u>C</u> GAAACG		5.98	7.87	49.1	60.0	139.2	167.6	39.0	49.6

^a Listed in alphabetical order and by oligomer length. For each DNA duplex only the top strand is shown. Underlined residues indicate the position of a G•A mismatch. Experimental values are the averages of the T_M^{-1} vs $\ln C_T$ and the curve fit parameters given in Table 1. ^b Sequences without a literature reference are from Table 1 of this work. ^c Standard errors for experimental ΔG_{37}° , ΔH° , and ΔS° are assumed to be 4%, 8%, and 8%, respectively. ^d Calculated for 10^{-4} M oligomer concentration for self-complementary sequences and 4×10^{-4} M for non-self-complementary sequences.

sequences are linear combinations of the eight G•A nearest-neighbor dimers. A set of linearly independent trimer sequences is built by adding one Watson-Crick pair to the left (3'-end) of each G•A nearest-neighbor dimer. The choice of which base pair to place at the left is arbitrary, but adding the same pair (in this case C-G) to the end of all dimers simplifies the analysis in terms of linearly independent sequences (38). For instance, eq 2 can be rewritten as eq 5.

$$\begin{aligned}
 &5' \text{-GATCTGTGTAC-3}' = \text{initiation} + \text{symmetry} + \\
 &3' \text{-CTAGAAACATG-5}' \\
 &\quad \text{GA} + \text{AT} + \text{TC} + \text{CT} + \text{TGC} + \\
 &\quad \text{CT} \quad \text{TA} \quad \text{AG} \quad \text{GA} \quad \text{AAG} \\
 &\quad \text{AAC} - \text{GGC} + \text{TG} + \text{GT} + \text{TA} + \\
 &\quad \text{TGG} \quad \text{CAG} \quad \text{AC} \quad \text{CA} \quad \text{AT} \\
 &\quad \text{AC} \\
 &\quad \text{TG}
 \end{aligned} \tag{5}$$

The trimer sequence GGC/CAG is subtracted to account for the extra GC/AG and GG/CA neighbors that are added in by the trimer sequences TGC/AAG and AAC/TGG (i.e., $\text{TGT/AAA} = \text{TGC/AAG} + \text{GGT/CAA} - \text{GGC/CAG}$).

Regression Analysis. The free energy, enthalpy, and entropy contributions of the seven linearly independent G•A mismatch-containing trimers were determined by multiple linear regression analysis using the program MATHEMATICA v2.1 (Wolfram research). Thermodynamic parameters derived from the averages of the fits and T_M^{-1} vs $\ln C_T$ plots are equally reliable when the ΔH° parameters agree within 15% (48, 60), and thus the averages of these parameters (Table 2) were used to construct a list of 27 equations analogous to eq 4 with seven unknowns. These equations were cast into matrix form and used as input for linear regression by singular value decomposition (SVD) as described previously (38, 48). In our SVD analysis, to obtain reliable solutions that represent the uncertainties in each measurement, we used an error weighted fit of the data (38) (see below).

Error Analysis. The errors reported in Table 1 for averages of the fits and T_M^{-1} vs $\ln C_T$ plots were obtained using standard methods (53) and reflect the reproducibility and precision of the data. The ΔG_{37}° , ΔH° , ΔS° , and T_M parameters reported in Table 3 are the averages of the fits

Table 3: Thermodynamic Parameters for 16 Unique Trimer Sequences with Internal G·A Mismatches in 1 M NaCl^a

propagation sequence	ΔH° (kcal/mol)	ΔS° (eu)	ΔG°_{37} (kcal/mol)
Seven Linearly Independent Trimers			
AAC/TGG	-0.1 ± 1.5	0.9 ± 3.2	-0.38 ± 0.11
CAC/GGG	-0.2 ± 1.1	0.9 ± 3.2	-0.49 ± 0.12
GAA/CGT	2.5 ± 1.7	6.5 ± 3.5	0.49 ± 0.09
GAC/CGG	-0.1 ± 1.7	2.2 ± 2.9	-0.78 ± 0.11
GAG/CGC	-4.6 ± 1.6	-14.4 ± 3.0	-0.14 ± 0.10
GAT/CGA	-1.2 ± 2.0	-3.1 ± 3.8	-0.24 ± 0.14
TAC/AGG	1.1 ± 1.9	3.9 ± 2.9	-0.10 ± 0.10
The Nine Other Trimer Contexts ^b			
AAA/TGT	2.4 ± 2.3	4.9 ± 3.9	0.88 ± 0.13
AAG/TGC	-4.6 ± 2.4	-15.6 ± 3.3	0.25 ± 0.14
AAT/TGA	-1.3 ± 2.0	-4.7 ± 3.0	0.16 ± 0.11
CAA/GGT	2.3 ± 1.6	4.9 ± 2.8	0.77 ± 0.09
CAG/GGC	-4.7 ± 2.1	-15.6 ± 2.9	0.14 ± 0.12
CAT/GGA	-1.4 ± 1.8	-4.7 ± 2.5	0.05 ± 0.09
TAA/AGT	3.7 ± 2.1	8.2 ± 3.1	1.16 ± 0.11
TAG/AGC	-3.3 ± 2.2	-12.3 ± 3.0	0.53 ± 0.12
TAT/AGA	0.0 ± 1.7	-1.4 ± 2.7	0.44 ± 0.10

^a Errors are resampling standard deviations (see text). ^b These nine other contexts can be derived from linear combinations of the seven linearly independent trimer sequences (see text, eq 5).

and T_M^{-1} vs $\ln C_T$ plots, and their accuracies are estimated to be within 4%, 8%, 8%, and 1 °C, respectively (38). The small errors observed for ΔG°_{37} and T_M are due to the high degree of correlation between ΔH° and ΔS° ($R^2 > 99\%$) (38, 54). The errors in the experimental measurements and Watson–Crick nearest-neighbors were propagated to the G·A nearest-neighbor parameters in the variance–covariance matrix in the SVD analysis (48, 61, 62). For example, to estimate the error in the free energy for the nearest neighbors in eq 2:

$$(\sigma_{\Delta G^\circ_{37}(\text{mismatch})})^2 = (\sigma_{\Delta G^\circ_{37}(\text{measured})})^2 + \sum_{NN} (\sigma_{\Delta G^\circ_{37}})^2 \quad (6)$$

where, $\sigma_{\Delta G^\circ_{37}(\text{mismatch})}$ is the propagated error associated with the $\Delta G^\circ_{37}(\text{mismatch})$, $\sigma_{\Delta G^\circ_{37}(\text{measured})}$ is the uncertainty in the measured free energy for the duplex (~4%), and $\sum_{NN}(\sigma_{\Delta G^\circ_{37}})$ is the squared sum of errors for the Watson–Crick nearest-neighbors present in the duplex (38). The error from the initiation parameter essentially does not contribute because it covaries with the nearest-neighbors (63). For example, the error associated with the mismatch ΔG°_{37} (eqs 3 and 4) can be calculated as follows:

$$\begin{aligned} (\sigma_{\Delta G^\circ_{37}(\text{mismatch})})^2 &= (0.27)^2 + (0.03)^2 + (0.04)^2 + \\ &(0.03)^2 + (0.03)^2 + (0.06)^2 + (0.04)^2 + (0.06)^2 + (0.04)^2 \\ \sigma_{\Delta G^\circ_{37}(\text{mismatch})} &= 0.30 \text{ kcal/mol} \quad (7) \end{aligned}$$

Therefore, the measured value for the nearest neighbors TG/AA and AA/TG is 0.86 ± 0.30 kcal/mol. The propagated errors in the G·A nearest-neighbor parameters have been independently confirmed using standard deviations obtained by performing resampling analysis of the data (38, 64).

Resampling Analysis of the Data. Previously we showed that a resampling analysis of the data is an effective technique for pointing out any measurements that dramatically affect the solution from the regression analysis (38). This resampling technique is applicable when there are more measure-

ments than unknowns (i.e., when the solution is overdetermined). In our case, we have 27 equations with seven unknowns. We performed 30 resampling trials. In each trial, we removed seven randomly selected sequences from our data and performed SVD analysis on the remaining 23 sequences to calculate the seven unknowns. Each time, the rank of the matrix was confirmed to be seven and a solution was obtained. This resampling analysis was performed for ΔG°_{37} , ΔH° , and ΔS° . The nearest-neighbor parameters from the 30 trials were averaged, and standard deviations were calculated. The nearest-neighbors from the averaged resampling trials are the same as those obtained from SVD analysis using 27 equations. The resampling standard deviations, reported in Table 3, have the advantage that they are independent of any assumptions made about the experimental uncertainties (64).

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 or 5. Duplex 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 (65, 66). Spectra were recorded with the carrier placed at the solvent frequency and with high-power and low-power pulse widths of 8.8 and 1700 μ s, sweep width of 12 kHz, gradient field strength of 10.0 G/cm, and duration of 1 ms. In all, 512–1024 transients were collected for each spectrum. Data were multiplied by a 2.0 Hz line-broadening exponential function and Fourier transformed by a Silicon Graphics Indigo²Extreme computer with Varian VNMR and Felix-95.0 (Biosym/MSI) 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. One-dimensional nuclear Overhauser enhancement (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. In all, 3200–6400 scans were collected for each FID.

RESULTS

Thermodynamic Data. Plots of T_M^{-1} vs $\ln C_T$ were linear (correlation coefficient >0.98) over the entire 80–100-fold range in concentrations (not shown). Thermodynamic parameters of helix formation derived from these plots and averages of fits of the melting curves are listed in Table 1. The model used to derive all thermodynamic parameters in this study assumes that the transition from helix to random coil is two-state. Typically, an agreement within 10–15% for ΔH° from the averages of curve fits and T_M^{-1} vs $\ln C_T$ plot is considered to indicate a two-state transition, but caution is in order (see below) (38, 54, 56). Duplexes listed in Table 1 to have two-state transitions showed monophasic and concentration dependent melting curves and had differences in ΔH° less than 15%.

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

propagation sequence	ΔH° (kcal/mol)	ΔS° (eu)	ΔG°_{37} (kcal/mol)
AA/TG	-0.6	-2.3	0.14
AG/TA	-0.7	-2.3	0.02
CA/GG	-0.7	-2.3	0.03
CG/GA	-4.0	-13.2	0.11
GA/CG	-0.6	-1.0	-0.25
GG/CA	0.5	3.2	-0.52
TA/AG	0.7	0.7	0.42
TG/AA	3.0	7.4	0.74

^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 within roundoff error of the parameters listed in Table 3. Linear combinations of the parameters in this table give the parameters in Table 3. Trends in these parameters should not be considered physically relevant (see text).

Linear Regression Analysis of G•A Mismatches in Terms of Seven Linearly Independent Sequences. Table 3 lists parameters for seven uniquely determined trimer sequences with G•A single mismatches obtained using a multiple linear regression analysis of the data (see Materials and Methods). The other nine possible trimer contexts are also listed in Table 3 and are linear combinations of the seven uniquely determined trimers. The errors listed in Table 3 are the standard deviations from resampling analysis of the data (see Material and Methods). The resampling errors are within roundoff of the errors obtained by propagating experimental errors in the SVD analysis. The parameters in Table 3 along with Watson–Crick parameters (38) predict the thermodynamics of all 27 sequences listed in Table 2 as two-state transitions with an average deviation for ΔG°_{37} , ΔH° , ΔS° , and T_M of 4.4%, 7.4%, 8.0%, and 1.5 °C, respectively. Previously, we and others (38, 48, 56, 60) showed that this level of agreement between experiment and prediction is indicative of the applicability of the nearest-neighbor model.

Linear Regression Analysis of G•A Mismatches in Terms of Eight Nonunique Dimers. Table 4 lists nearest-neighbor parameters for dimer sequences containing G•A mismatches. These parameters are an alternative presentation of the data in Table 3 and were obtained using SVD analysis of 27 equations fitted to eight unknowns. The singular matrix obtained using this method is rank deficient (rank is seven), indicating that the fit can be represented in terms of seven unknowns rather than eight and, hence, the eight parameters are nonunique (38, 62). The nonuniqueness of the parameters in Table 4 is manifested in the fact that it is possible to obtain the parameters in Table 3 using those listed in Table 4 but not vice versa unless an eighth parameter is given. The SVD analysis in terms of eight dimer sequences assumes this eighth parameter to be zero (38). The parameters in Table 4 make predictions that are the same as those obtained using the parameters listed Table 3.

Molecules with Non-Two-State Thermodynamics. Tables 1 and 2 list four sequences that have non-two-state thermodynamics. It is likely that these sequences form a number of structures in addition to the desired duplex. For three of these sequences, non-two-state behavior is evident from the large differences in ΔH° obtained from T_M^{-1} vs $\ln C_T$ and the averages of the fits (>15%). However, one sequence CGTGTCGAAACG/GCAAAGCTGTGC, listed as non-two-

state, has a good agreement (within 4%) of enthalpies derived by the two methods, but a biphasic transition is observed, indicating non-two-state behavior.

Molecules with Anomalous Two-State Thermodynamics. Thermodynamics of the two sequences, (ATGAGCGCAT)₂ (SEQ-1), and (ATGAGCTAAT)₂ (SEQ-2), were first reported by Li and co-workers (13) (see Table 2). When these two sequences were included in the regression analysis, the thermodynamic parameters obtained were within roundoff error of those listed in Tables 3; however, the experimental and predicted thermodynamics of the two sequences did not agree (differences in the experimental and predicted T_M were 7.5 °C for SEQ-1 and 21.2 °C for SEQ-2). To investigate the origin of this poor agreement, we synthesized and melted these two duplexes and obtained thermodynamics that are in good agreement with those reported by Li et al. (13) (see Table 2), indicating that the origin of disagreement between prediction and experiment is not due to instrumental calibration. However, the sequence SEQ-1 turned out to have a difference in ΔH° obtained from T_M^{-1} vs $\ln C_T$ and averages of the fits of 37% (see Table 1), indicating a non-two-state behavior. Nonetheless, the sequence SEQ-2 appeared to have a good ΔH° agreement (~6%) and did not exhibit a multiphasic transition. This ΔH° agreement of the two methods does not rule out the presence of alternative structures. Since there is no evidence that would support the presence of competing structures, this sequence is listed to have anomalous two-state thermodynamics. At the present, the origin of this anomalous behavior is unknown.

Thermodynamics of G•A Mismatches at pH 5.0. The thermodynamics of four sequences listed in Table 1 were obtained at pH 5.0. These four sequences were selected to represent G•A mismatches in different Watson–Crick contexts. The thermodynamics obtained for these four sequences at pH 5.0 and 7.0 are within experimental error. Thus it appears that the thermodynamics of single G•A mismatches do not exhibit a strong pH dependence. However, the G•A mismatch structure does appear to be pH and context dependent (see below).

NMR Spectra of G•A Mismatches. To determine the pairing geometry for G•A mismatches in this study, we acquired 1D exchangeable proton NMR spectra of six DNA duplexes representing different possibilities for closing Watson–Crick pairs at pH 7.0 and 5.0. For G•A pairs, it has been reported that pairing geometry is pH-dependent with a transition pK_a around 6.0 (31, 37). Figures 2 and 3 show the imino region (9–15 ppm) of six DNA duplexes containing G•A mismatches in different contexts at pH 7.0 and 5.0, respectively. Resonances were assigned either by using 1D-NOE and temperature dependent broadening of terminal base pairs or by assuming that guanine and thymine imino protons in canonical Watson–Crick G•C and A•T pairs resonate between 12 and 13 and 13–15 ppm, respectively. The imino proton NMR data at pH 7.0 and 5.0 correspond to the number of Watson–Crick pairs present in each duplex. At pH 7.0 (Figure 2), a broad resonance at 9.7 ppm is observed for three of the sequences, and no resonance is observed for the other three sequences, indicating chemical exchange with solvent. On the basis of previous observations made on single G•A mismatches at neutral pH (31, 37), we tentatively assign this resonance as the adenine amino proton (A-NH₂)

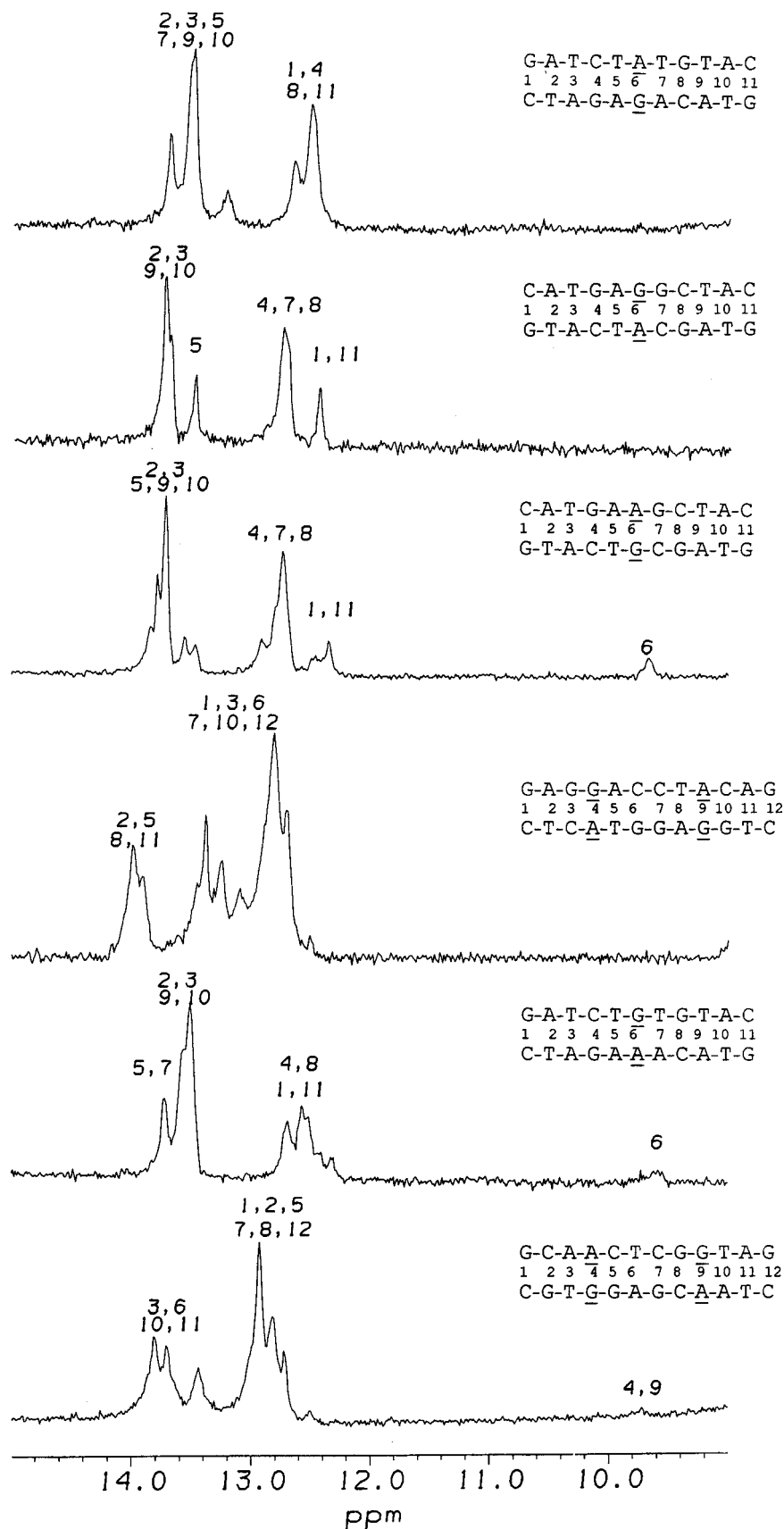


FIGURE 2: 500 MHz ^1H NMR spectra of the exchangeable imino region (9–15 ppm) at pH 7.0 at 10 °C in 1 M NaCl, 10 mM disodium phosphate, and 0.1 mM Na_2EDTA in 90% $\text{H}_2\text{O}/10\%$ D_2O . Duplexes are shown above the spectrum to which they correspond. Tentative assignments are shown above each spectrum.

in a G(anti)·A(anti) or G(syn)·A⁺(anti) pairing geometries. The imino proton of guanine in a G(anti)·A(anti) pair was

reported to resonate at 11.7 or 13.6 ppm depending on the mismatch context (31, 37, 67). At pH 7.0 (Figure 2), the

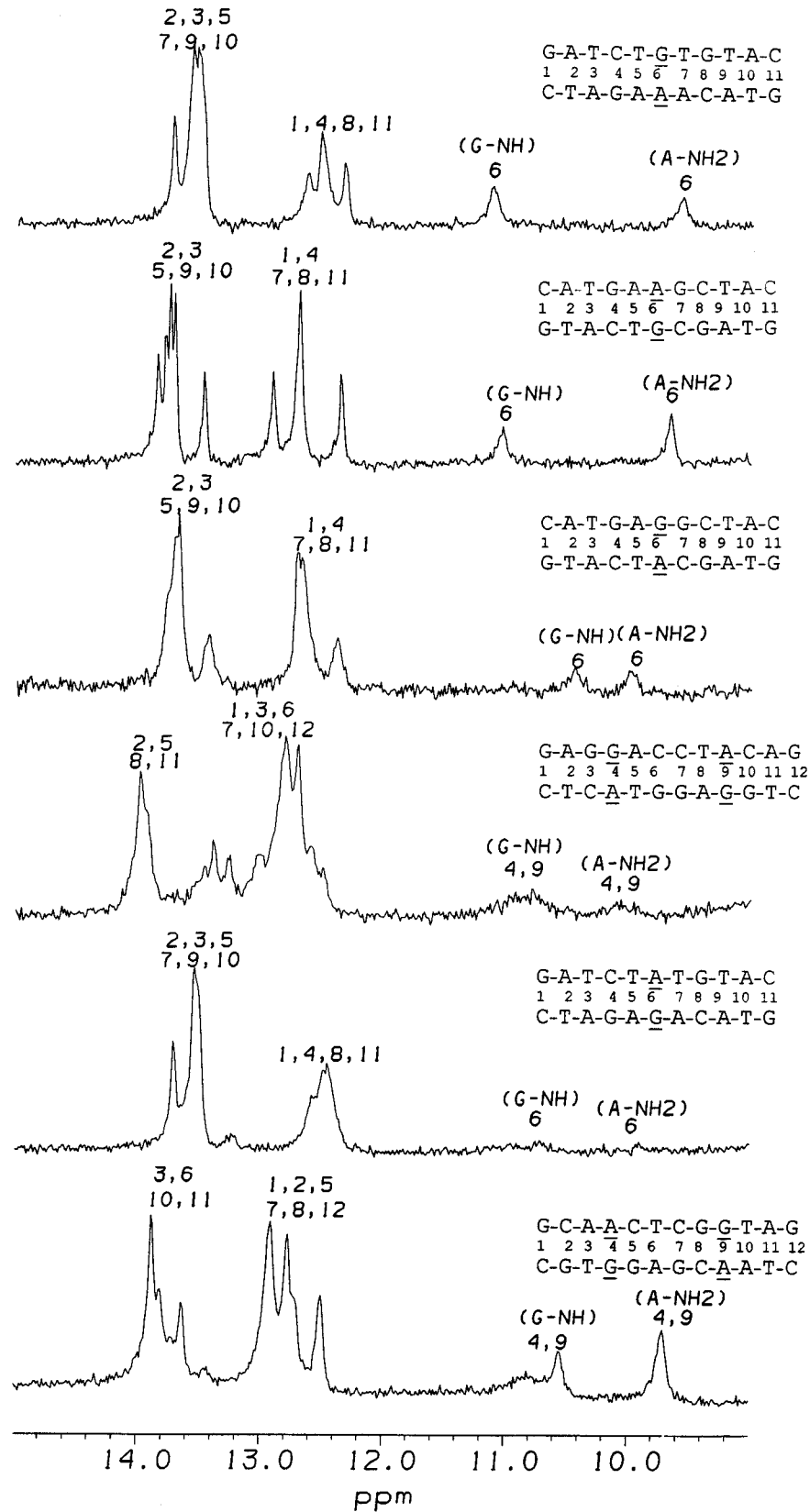


FIGURE 3: 500 MHz ^1H NMR spectra of the exchangeable imino region (9–15 ppm) at pH 5.0 at 10 °C in 1 M NaCl, 10 mM disodium phosphate, and 0.1 mM Na_2EDTA in 90% $\text{H}_2\text{O}/10\%$ D_2O . Duplexes are listed above the spectrum to which they correspond. Tentative assignments are shown above each spectrum.

region around 13.6 ppm is crowded and no resonances are observed near 11.7 ppm so the G·A mismatch guanine imino resonances are not definitively assigned. Upon lowering the pH to 5.0 an extra resonance around 11.0 ppm appears

(Figure 3). The appearance of this resonance indicates a structural change within the G·A pairing geometry which resulted in the involvement of an imino or amino proton that is paired weakly or not at all at neutral pH. Previous work

suggests that this resonance is the imino proton of the guanine (G-NH) in G(syn)•A⁺(anti) pair (31, 37).

DISCUSSION

Applicability of the Nearest-Neighbor Model to Single G•A Mismatches. Table 2 compares the experimental results of 27 G•A mismatch containing duplexes with those predicted using the parameters listed in Table 4 in conjunction with Watson–Crick nearest-neighbors (38). We have previously shown that the nearest-neighbor model (38) can predict canonical Watson–Crick duplexes with average deviations for ΔG°_{37} , ΔH° , ΔS° , and T_M of 4%, 7%, 8%, and 2 °C, respectively. The G•A mismatch parameters listed in Table 4 predict the thermodynamics of all 27 sequences with two-state transitions with a standard deviation of 4.0%, 7.4%, 8.0%, and 1.5 °C for ΔG°_{37} , ΔH° , ΔS° , and T_M , respectively. This indicates that the nearest-neighbor model applies to oligonucleotides with single G•A mismatches equally as well as to DNA duplexes with only Watson–Crick pairs.

Errors in G•A Mismatch Nearest-Neighbors and Sequence Trends. Table 3 lists error estimates for 16 unique trimer sequences with internal single G•A mismatches obtained using standard deviations of resampling analysis (see above). When considering errors, what matters is determining whether statistically significant distinctions can be made when comparing various measurements. The ΔG°_{37} contribution of a single G•A mismatch to duplex stability is dependent on the neighboring base pairs and ranges from +1.16 kcal/mol (for the context TGA/AAT) to -0.78 kcal/mol (for the context GGC/CAG) (Table 3). The average error in the ΔG°_{37} contribution of single G•A mismatches is 0.11 kcal/mol. Since the average error is 5% of the range, it is clear that a statistically significant sequence trend for G•A mismatches is observed. The general trend for the nucleotide at the 5' side of the mismatch in order of decreasing stability is G > C > A > T. Note that this trend is the same whether the adenine or the guanine of the G•A pair are in the top strand. Similar observations are made for ΔH° and ΔS° nearest-neighbor parameters (Table 3); but the errors for G•A mismatches are about 15% of the range, so that it is more difficult to deduce sequence trends. However, the ΔH° and ΔS° parameters are still useful for predicting the duplex T_M (average deviation of experimental vs predicted T_M is 1.5 °C), since errors in ΔH° and ΔS° are greater than 99% correlated (38). The errors are also small enough to distinguish the G•A mismatch contribution from that of other mismatches and Watson–Crick pairs (see below).

Comparison of Thermodynamics of G•A Mismatches with Watson–Crick Base Pairs. G•A trimer sequences cover a range of 1.94 kcal/mol, while the corresponding trimer sequences with G•C and T•A trimer sequences range over 1.83 and 1.43 kcal/mol, respectively. The most stable trimer sequences containing a central G•C or T•A pair (GCG/CGC and CAC/GTG) are -4.41 and -2.89 kcal/mol (38), respectively, and the most stable trimer sequence with a G•A mismatch (GAC/CGG) is -0.78 kcal/mol. The least stable trimer sequences with a central G•C or a T•A pair (TCT/AGA and TAT/ATA) are -2.58 and -1.46 kcal/mol, respectively, and the least stable G•A mismatch trimer (TAA/AGT) is 1.16 kcal/mol. This comparison indicates that the most stable G•A, G•C, and T•A trimers always have a

5' G•C closing base pair and the least stable trimer have a 5' T•A closing base pair.

Comparison of G•A Mismatch Thermodynamics with G•T DNA Mismatches. We recently showed that the nearest-neighbor model is applicable to internal G•T mismatches in DNA (38). In our study of G•T mismatches (38), the most stable trimer sequence containing a G•T mismatch was -1.05 kcal/mol for CGC/GTG and the least stable was +1.05 kcal/mol for AGA/TTT. In this work, we find that the most stable G•A trimer sequence is -0.79 kcal/mol for GGC/CAG and the least stable is +1.16 kcal/mol for TGA/AAT. On average, when the closing Watson–Crick pair on the 5' side of the mismatch is an A•T or a G•C pair, G•A mismatches are more stable than G•T mismatches by about 0.40 and 0.30 kcal/mol, respectively. When the 5' closing pair is a T•A or a C•G, then G•T mismatches are more stable than G•A mismatches by 0.54 and 0.75 kcal/mol, respectively. Evidently, the different hydrogen-bonding and stacking in G•T and G•A mismatches results in different thermodynamic trends.

Comparison of Single G•A Mismatch Thermodynamics in DNA and RNA. Single G•A mismatch thermodynamics in RNA have previously been investigated (49). The authors reported data for three duplexes with single G•A mismatches. Two of the sequences had the same nearest-neighbor composition but showed different thermodynamics, and on this basis the authors concluded that next-nearest-neighbor effects were evident. However, one of the sequences (r(CGCAGCGGCG)₂) has the potential to form a competing hairpin structure similar to that observed for a G•T mismatch sequence studied previously in our laboratory (38). Interestingly, we find that the nearest-neighbor model applies to single G•A mismatches in DNA and next-nearest-neighbor effects are negligible. More measurements are required to deduce if a nearest-neighbor model is applicable to G•A mismatches in RNA.

NMR Spectroscopy and Pairing Geometries in Single G•A Mismatches. Previous studies on single G•A mismatches (31, 37) have shown that the geometry of the pairing in the G•A mismatch is pH dependent. Our data indicate that at neutral pH either the G(anti)•A(anti) (Figure 1A) or the unprotonated G(syn)•A(anti) (Figure 1C) is formed or is interconverting. At pH 5.0, our data are most consistent with the formation of the protonated G(syn)•A⁺(anti) structure (Figure 1C). Our data are consistent with previous data that the pK_a for protonation of G•A mismatches is about 6.0 (37). This suggests that at pH 7.0 about 90% of G•A mismatches are in the G(anti)•A(anti) conformation and at pH 5.0 90% are in the G(syn)•A⁺(anti). Interestingly, while the thermodynamics of single G•A mismatches seem to be pH independent (see above), we observe that G•A mismatches in the present study adopt different conformations at pH 7.0 and 5.0. This seems to indicate that the overall stacking and hydrogen bonding energies of G(anti)•A(anti) at pH 7.0 are similar to G(syn)•A⁺(anti) at pH 5.0. Thus, structure appears to be more sensitive to pH than thermodynamics.

ACKNOWLEDGMENT

We thank David Hyndman (Advanced Gene Computing Technologies) for stimulating conversations and Meiko Ogura (Hitachi Chemical Research) for oligonucleotide

synthesis. We also thank Shikha Varma and Nicolas Peyret for critical reading of the manuscript.

REFERENCES

- 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) *Aldrichimica Acta* 28, 15–20.
- Aboul-ela, F., Koh, D., Tinoco, I., Jr., and Martin, F. H. (1985) *Nucleic Acids Res.* 13, 4811–4824.
- Werntges, H., Steger, G., Riesner, D., and Fritz, H. J. (1986) *Nucleic Acids Res.* 14, 3773–3790.
- 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.
- Mendelman, L. V., Boosalis, M. S., Petruska, J., and Goodman, M. F. (1989) *J. Biol. Chem.* 264, 14415–14423.
- Johnson, K. (1993) *Annu. Rev. Biochem.* 62, 685–713.
- Brown, T., Hunter, W. N., and Leonard, G. A. (1993) *Chem. Br.* 6, 484–488.
- Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T., and Itakura, K. (1979) *Nucleic Acids Res.* 6, 3543–3557.
- Kawase, Y., Iwai, S., Inoue, H., Miura, K., and Ohtsuka, E. (1986) *Nucleic Acids Res.* 14, 7727–7736.
- Ikuta, S., Takagi, K., Wallace, R. B., and Itakura, K. (1987) *Nucleic Acids Res.* 15, 797–811.
- Li, Y., Zon, G., and Wilson, W. D. (1991) *Biochemistry* 30, 7566–7572.
- Li, Y., and Agrawal, S. (1995) *Biochemistry* 34, 10056–10062.
- 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.
- 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.
- 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.
- Chang, E. H., Yu, Z., Shinozuki, K., Zon, G., Wilson, W. D., and Sterkowska, A. (1989) *AntiCancer Drug Des.* 4, 221.
- Steger, G. (1994) *Nucleic Acids Res.* 22, 2760–2768.
- Fersht, A. R., Knill-Jones, J. W., and Tsui, W. C. (1982) *J. Mol. Biol.* 156, 37–51.
- 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.
- Patel, D. J., Kozlowski, S. A., Ikuta, S., and Itakura, K. (1984) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* 43, 2663–2670.
- Brown, T., Hunter, W. N., Kneale, G., and Kennard, O. (1986) *Proc. Natl. Acad. Sci. U.S.A.* 83, 2402–2406.
- Webster, G. D., Sanderson, M. R., Skelly, J. V., Swann, P. F., Li, B. F., and Tickel, I. J. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 6693–6697.
- Kan, L.-S., Chandrasegaran, S., Pulford, S. M., and Miller, P. S. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80, 4263–4265.
- Hunter, W. N., Brown, T., and Kennard, O. (1986) *J. Biomol. Struct. Dyn.* 4, 173–191.
- Prive, G. G., Heinemann, U., Kan, L.-S., Chandrasegaran, S., and Dickerson, R. E. (1987) *Science* 238, 498–504.
- Cheng, J. W., Chou, S. H., and Reid, B. R. (1992) *J. Mol. Biol.* 228, 1037–1041.
- Gao, X., and Patel, D. J. (1988) *J. Am. Chem. Soc.* 110, 5178–5182.
- Brown, T., Leonard, G. A., Booth, E. D., and Chambers, J. (1989) *J. Mol. Biol.* 207, 455–457.
- Leonard, G. A., Booth, E. D., and Brown, T. (1990) *Nucleic Acids Res.* 18, 5617–5623.
- Maskos, K., Gunn, B. M., LeBlanc, D. A., and Morden, K. M. (1993) *Biochemistry* 32, 3583–3595.
- Nikonowicz, E. P., and Gorenstein, D. G. (1990) *Biochemistry* 29, 8845–8858.
- Nikonowicz, E. P., Meadows, R. P., Fagan, P., and Gornstein, D. G. (1991) *Biochemistry* 30, 1323–1334.
- Carbonnaux, C., van der Marel, G. A., van Boom, J. H., Guschlbauer, W., and Fazakerley, G. V. (1991) *Biochemistry* 30, 5449–5458.
- Allawi, H. T., and SantaLucia, J., Jr. (1997) *Biochemistry* 36, 10581–10594.
- Chou, S.-H., Zhu, L., and Reid, B. (1997) *J. Mol. Biol.* 267, 1055–1067.
- Green, K. L., Jones, R. L., Li, Y., Robinson, D. L., Wang, A. H.-J., Zon, G., and Wilson, W. D. (1994) *Biochemistry* 33, 1053–1062.
- Chou, S.-H., Cheng, J.-W., Fedoroff, O. Y., Chuprina, V. P., and Reid, B. R. (1992) *J. Am. Chem. Soc.* 114, 3114–3115.
- Chou, S.-H., Cheng, J.-W., and Reid, B. R. (1992) *J. Mol. Biol.* 228, 138–155.
- SantaLucia, J. J., Kierzek, R., and Turner, D. H. (1990) *Biochemistry* 29, 8813–8819.
- SantaLucia, J., Jr., Kierzek, R., and Turner, D. (1991) *J. Am. Chem. Soc.* 113, 4313–4322.
- Wu, M., SantaLucia, J., Jr., and Turner, D. H. (1997) *Biochemistry* 36, 4449–4460.
- Wu, M., and Turner, D. H. (1996) *Biochemistry* 35, 9677–9689.
- SantaLucia, J. J., and Turner, D. H. (1993) *Biochemistry* 32, 12612–12613.
- SantaLucia, J., Jr., Allawi, H. T., and Seneviratne, P. A. (1996) *Biochemistry* 35, 3555–3562.
- Morse, S. E., and Draper, D. E. (1995) *Nucleic Acids Res.* 23, 302–306.
- 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.
- Richards, E. G. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.), CRC Press, Cleveland, OH.
- McDowell, J. A., and Turner, D. H. (1996) *Biochemistry* 35, 14077–14089.
- Petersheim, M., and Turner, D. H. (1983) *Biochemistry* 22, 256–263.
- Borer, P. N., Dengler, B., Tinoco, I., Jr., and Uhlenbeck, O. C. (1974) *J. Mol. Biol.* 86, 843–853.
- 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.
- Marky, L. A., and Breslauer, K. J. (1987) *Biopolymers* 26, 1601–1620.
- Wu, M., McDowell, J. A., and Turner, D. H. (1995) *Biochemistry* 34, 3204–3211.
- Gray, D. M. (1997) *Biopolymers* 42, 783–793.
- Sugimoto, N., Nakano, S., Yoneyama, M., and Honda, K. (1996) *Nucleic Acids Res.* 24, 4501–4505.
- Bevington, P. R. (1969) *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York.
- Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1989) *Numerical Recipes*, Cambridge University Press, New York.
- SantaLucia, J., Jr. and Turner, D. H. (submitted for publication).
- Efron, B., and Tibshirani, R. (1993) *An Introduction to the Bootstrap*, Chapman & Hall, London.
- Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR* 2, 661–665.
- Lippens, G., Dhalluin, C., and Wieruszkeski, J.-M. (1995) *J. Biomol. NMR* 5, 327–331.
- Patel, D. J., Kozlowski, S. A., Ikuta, S., and Itakura, K. (1984) *Biochemistry* 23, 3207–3217.
- Arghavani, M., SantaLucia, J., Jr., and Romano L. (submitted for publication).