Articles

Nearest-Neighbor Thermodynamics and NMR of DNA Sequences with Internal A·A, C·C, G·G, and T·T Mismatches[†]

Nicolas Peyret, P. Ananda Seneviratne, Hatim T. Allawi, and John SantaLucia, Jr.*

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

Received October 21, 1998

ABSTRACT: Thermodynamic measurements are reported for 51 DNA duplexes with A·A, C·C, G·G, and T·T single mismatches in all possible Watson−Crick contexts. These measurements were used to test the applicability of the nearest-neighbor model and to calculate the 16 unique nearest-neighbor parameters for the 4 single like with like base mismatches next to a Watson−Crick pair. The observed trend in stabilities of mismatches at 37 °C is G·G > T·T ≈ A·A > C·C. The observed stability trend for the closing Watson−Crick pair on the 5′ side of the mismatch is G·C ≥ C·G ≥ A·T ≥ T·A. The mismatch contribution to duplex stability ranges from −2.22 kcal/mol for G<u>G</u>C·G<u>G</u>C to +2.66 kcal/mol for A<u>C</u>T· A<u>C</u>T. The mismatch nearest-neighbor parameters predict the measured thermodynamics with average deviations of $\Delta G°_{37}$ = 3.3%, $\Delta H°$ = 7.4%, $\Delta S°$ = 8.1%, and T_M = 1.1 °C. The imino proton region of 1-D NMR spectra shows that G·G and T·T mismatches form hydrogen-bonded structures that vary depending on the Watson−Crick context. The data reported here combined with our previous work provide for the first time a complete set of thermodynamic parameters for molecular recognition of DNA by DNA with or without single internal mismatches. The results are useful for primer design and understanding the mechanism of triplet repeat diseases.

DNA mismatches occur in vivo due to misincorporation of bases during replication (*1*), heteroduplex formation during homologous recombination (*2*), mutagenic chemicals (*3, 4*), ionizing radiation (*5*), and spontaneous deamination (*6*). Knowledge of the thermodynamics of DNA mismatches will be useful for elucidating the mechanisms of polymerase fidelity and mismatch repair efficiency. Moreover, thermodynamic parameters for mismatch formation are important for DNA secondary structure prediction (see http://sun2.science.wayne.edu/~jslsun2 and http://mfold1.wustl.edu/~mfold/dna/form1.cgi). Recent work has shown that triplet repeat sequences form transiently stable hairpins that contain like with like base mismatches (*7−14*). The formation of these secondary structures can induce genome expansion or deletion during replication (*15, 16*) resulting in at least 11 different human diseases (*17−19*). Mismatch thermodynamics is also important for molecular biological techniques such as PCR (*20*), Southern blotting (*21*), single-stranded conformational polymorphism (SSCP) (*22−24*), sequencing by hybridization (*25, 26*), antigene targeting (*27*), Kunkel site-directed mutagenesis (*28*), and optimization of DNA chip arrays for diagnostics (*29*). These techniques require optimization of sequence, temperature,

and solution conditions to avoid detection or amplification of wrong sequences. Previous work from our laboratory has shown that a NN[1] model is valid to describe the thermodynamics of DNA structures involving canonical A·T and G·C base pairs (*30−32*) as well as G·T (*31*), G·A (*33*), C·T (*34*), and A·C (*35*) mismatches. We hypothesized that the nearest-neighbor model is also applicable to single A·A, C·C, G·G, and T·T mismatches. To test this hypothesis, thermodynamic measurements of 45 sequences combined with 6 from the literature (*36, 37*) were used to derive NN parameters for like with like base mismatches. 1-D NMR and CD studies were used to qualitatively probe the structures formed by the mismatches. These data combined with our previous results provide a complete thermodynamic database for DNA molecular recognition by DNA with or without single internal mismatches.

MATERIALS AND METHODS

DNA Synthesis and Purification. Oligonucleotides were graciously provided by Hitachi Chemical Research and were synthesized on solid support using standard phosphoramidite chemistry (*38*). Oligonucleotides were detached from the

---

[1] Abbreviations: Na$_2$EDTA, disodium ethylenediaminetetraacetate; eu, entropy unit; MES, 2-(4-morpholino)ethane sulfonate; NMR, nuclear magnetic resonance; NN, nearest-neighbor; SVD, singular value decomposition; TLC, thin-layer chromatography; UV, ultraviolet.

solid support, and the base blocking groups were removed by treatment with concentrated ammonia at 50 °C overnight. Each oligonucleotide sample was evaporated to dryness and dissolved in 250 μL of water. The solution was then purified on a Si500F thin-layer chromatography plate (Baker) by elution for 5 h with a mixture 1-propanol/ammonia/distilled deionized water in volumetric proportions of 55:35:10, respectively (*39*). Bands were visualized with a UV lamp, and the least mobile one was scrapped off and extracted three times with 3 mL of distilled deionized water. The oligonucleotides were further purified and desalted with Sep-pak C-18 cartridges (Waters). Then the oligonucleotides were aliquoted for UV melting, CD, and NMR experiments.

UV Melting Curves. Absorbance versus temperature curves were measured at 260 or 280 nm with a heating rate of 0.8 °C min^{-1} using an AVIV 14DS UV−vis spectrophotometer with a five-cuvette thermoelectric controller as described previously (*30*). The buffer used for the melting curves was 1.0 M NaCl, 10 mM sodium cacodylate, and 0.5 mM NA$_2$EDTA, pH 7.0 or 4.9. For non-self-complementary duplexes the strands were mixed in a 1:1 molar ratio. Each duplex was melted at 8−10 different concentrations over an 80−100-fold range. Samples were annealed and degassed by raising the temperature to 85 °C for 5 min, and the absorbance of each sample (260 nm) was recorded to calculate total strand concentrations, C_T, using the single-strand extinction coefficients (*40*).

Calculation of Thermodynamic Parameters. Thermodynamic parameters were determined from melting curves using the program MELTWIN v3.0 (*41*). Thermodynamic parameters were calculated by two methods: (i) enthalpy and entropy changes from fits of individual melting curves at different concentrations were averaged (*42*), and (ii) plots of reciprocal melting temperatures (T_M^{-1} vs ln C_T) were fit to the equation (*43*):

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

where $N = 1$ for self-complementary oligonucleotides and $N = 4$ for non-self-complementary oligonucleotides. Both methods assume a two-state model and $\Delta C_p° = 0$ for the transition equilibrium (*42, 44*). For duplexes with agreement within 15% of the $\Delta H°$ values derived from the two methods, it was assumed that the two-state approximation is applicable (*45, 46*).

Design of Sequences. Oligonucleotides were designed to have melting temperatures between 30 and 60 °C and to minimize formation of undesired hairpin or slipped duplex conformations. The formation of such alternative structures would result in a non-two-state transition (*31*). Moreover, the sequences were chosen to represent uniformly the 16 single mismatch nearest-neighbor dimers as well as the 40 possible trimer contexts that contain a central X·X mismatch surrounded by Watson−Crick pairs.

Determination of Nearest-Neighbor Parameters. According to the nearest-neighbor model (*32, 43, 46−49*) thermodynamics for a given sequence can be decomposed into incremental contributions: helix initiation, helix symmetry, and nearest-neighbor interactions between base pairs. We recently showed that the nearest-neighbor model could be extended to include single G·T (*31*), G·A (*50*), C·T (*34*), and A·C mismatches (*35*). Similarly, the nearest-neighbor

model can be further extended to include single A·A, C·C, G·G, and T·T mismatches. For instance, the thermodynamic parameters $\Delta Y°_{total}$ (where $Y = G$, H, or S) for the sequence (GGAGTCTCC)$_2$ are decomposed as follows:

$$\Delta Y°_{total} = \Delta Y°_{initiation} + \Delta Y°_{sym} + 2\Delta Y°(GG/CC) +$$
$$2\Delta Y°(GA/CT) + 2\Delta Y°(AG/TC) + 2\Delta Y°(G\underline{T}/C\underline{T}) \quad (2)$$

The notation G$\underline{T}$/C$\underline{T}$ refers to a $^5{'}$GT$^{3'}$ dimer hydrogen bonded to a $^{3'}$CT$^{5'}$ dimer with the mismatch underlined. The mismatch contribution to duplex stability is given by rearranging eq 2:

$$2\Delta Y°(G\underline{T}/C\underline{T}) = \Delta Y°_{total} - \Delta Y°_{initiation} - \Delta Y°_{sym} -$$
$$2\Delta Y°(GG/CC) - 2\Delta Y°(GA/CT) - 2\Delta Y°(AG/TC) \quad (3)$$

Thus, the mismatch contribution is calculated by subtracting the initiation, symmetry, and Watson−Crick nearest-neighbor increments (*31*) from the total experimental value.

Number of Linearly Independent Parameters. In our previous studies of G·T, G·A, A·C, and C·T single mismatches, we showed that it is impossible to uniquely solve for eight dimer nearest neighbors from a data set of oligomers containing only single internal mismatches (*31*). Instead, within the limits of the nearest-neighbor model, only seven linearly independent trimers are sufficient to accurately predict internal mismatch thermodynamics. In the case of single like with like base mismatches (i.e., A·A, C·C, G·G, and T·T), however, symmetry allows for a unique solution of four internal nearest-neighbor dimers to be found. In particular, the dimer nearest neighbors can be uniquely solved from sequences that contain these trimers:

$$\frac{\text{T X}}{\text{A X}} = \frac{1}{2}\left(\frac{\text{T X A}}{\text{A X T}}\right) \qquad (4a)$$

$$\frac{\text{A X}}{\text{T X}} = \frac{1}{2}\left(\frac{\text{A X T}}{\text{T X A}}\right) \qquad (4b)$$

$$\frac{\text{C X}}{\text{G X}} = \frac{1}{2}\left(\frac{\text{C X G}}{\text{G X C}}\right) \qquad (4c)$$

$$\frac{\text{G X}}{\text{C X}} = \frac{1}{2}\left(\frac{\text{G X C}}{\text{C X G}}\right) \qquad (4d)$$

where X = A, C, G, or T. According to the nearest-neighbor model, any sequence with an internal X·X mismatch can be determined from linear combinations of eqs 4a−d. It should be noted, however, that even though it is possible to uniquely solve for the X·X dimer nearest-neighbor parameters from a set of oligonucleotides with only internal mismatches, these parameters cannot be used to accurately predict the thermodynamics of duplexes with terminal mismatches. As we found earlier (*31*), terminal mismatches always make favorable contributions to duplex stability, whereas single internal mismatches make favorable or unfavorable contributions depending on the mismatch type and context (see Discussion).

Regression Analysis. Since the number of sequences in this study is greater than the number of unknown mismatch

Table 1: Experimental and Predicted Thermodynamic Parameters of Duplex Formation of Oligonucleotides with A·A, C·C, G·G, and T·T Mismatches

sequences[a]	ΔH° (kcal/mol)		ΔS° (eu)		ΔG°$_{37}$ (kcal/mol)		T$_M$ (°C)	
	experiment[b]	prediction	experiment[b]	prediction	experiment[b]	prediction	experiment[c]	prediction[c]
A·A mismatches								
CAAAAAAAG/[d]	−36.9	−41.9	−107.0	−123.8	−3.71	−3.47	21.3	21.5
CGATAATCG	−50.8	−42.4	−148.0	−120.8	−4.86	−4.96	32.1	31.9
GGAAATTCC	−51.5	−45.7	−151.4	−132.2	−4.59	−4.70	30.6	30.5
GGACAGTCC	−53.7	−50.7	−153.2	−144.4	−6.22	−5.94	40.2	38.6
GGAGACTCC	−51.6	−53.5	−145.7	−152.8	−6.38	−6.14	41.3	39.7
CATGAAGCTAC/	−65.2	−65.1	−185.4	−185.4	−7.70	−7.62	46.9	46.5
CATGTAACTAC/	−48.0	−54.8	−133.8	−155.5	−6.52	−6.62	42.5	42.4
GATCTATGTAC/	−59.3	−57.9	−170.6	−165.9	−6.42	−6.41	40.9	41.0
GGATGAATAGC/	−69.3	−61.9	−198.2	−174.9	−7.81	−7.63	46.9	47.1
GGATGAGTAGC/	−70.3	−68.6	−198.3	−194.0	−8.79	−8.45	51.4	50.1
CGCAAGAGACGG/	−66.3	−64.5	−186.6	−179.4	−8.42	−8.86	50.4	53.1
GGCAGAGAACGC/	−60.6	−65.6	−168.6	−183.3	−8.31	−8.77	51.1	52.3
GGA(CAG)₃AGG/[e]	−74.5	−73.2	−211.6	207.1	−8.87	−8.94	51.0	51.0
C·C mismatches								
CAAACAAAG/[d]	−55.3	−41.6	−170.0	−126.4	−2.57	−2.39	20.5	14.3
CGATCATCG	−36.6	−39.5	−104.7	−113.8	−4.14	−4.24	24.5	26.1
GGAACTTCC	−44.5	−48.1	−133.2	−144.4	−3.14	−3.26	20.3	22.2
GGACCGTCC	−53.6	−52.1	−155.4	−150.4	−5.35	−5.40	35.1	35.4
GGAGCCACG/	−48.9	−44.2	−138.9	−123.3	−5.86	−5.94	38.2	38.9
GGAGCCTCC	−44.1	−40.7	−125.9	−115.4	−5.09	−4.90	33.0	31.2
CATGTCACTAC/	−54.2	−51.9	−155.7	−148.4	−5.92	−5.90	38.4	38.3
GATCTCTGTAC/	−55.7	−57.6	−162.1	−168.5	−5.47	−5.33	35.9	35.2
GGATCCCTAGC/	−54.7	−62.0	−152.2	−176.4	−7.46	−7.25	47.5	45.1
GGATGCTTAGC/	−55.7	−60.9	−160.0	−175.3	−6.05	−6.49	39.1	41.2
GGATTCCTAGC/	−46.8	−54.0	−130.1	−152.8	−6.49	−6.60	42.5	42.4
GGATTCGTAGC/	−62.0	−59.7	−179.1	−170.4	−6.43	−6.85	40.8	43.2
GTAGCCTCATG/	−70.2	−67.0	−203.6	−194.5	−7.09	−6.63	43.4	41.5
G·G mismatches								
CAAAGAAAG/[d]	−53.5	−52.3	−158.0	−154.3	−4.50	−4.46	30.3	30.0
CGATGATCG	−48.4	−54.9	−138.8	−159.4	−5.34	−5.46	34.9	35.8
GGAAGTTCC	−52.2	−54.2	−148.9	−154.7	−6.00	−6.18	38.9	39.9
GGACGGTCC	−56.7	−58.7	−160.8	−166.8	−6.85	−7.02	43.6	44.3
GGAGGCTCC	−56.9	−59.8	−156.1	−164.8	−8.50	−8.70	53.2	53.5
CATGAGGCTAC/	−76.4	−73.4	−215.7	−207.8	−9.53	−8.90	53.5	51.3
CATGTGACTAC/	−57.0	−67.3	−160.8	−194.0	−7.18	−7.12	45.4	43.8
CCATCGCTACC/[f]	−78.7	−74.0	−221.0	−206.7	−10.17	−9.87	55.8	55.6
CCATTGCTACC/[f]	−75.1	−70.4	−212.3	−197.8	−9.30	−9.02	52.7	52.5
GATCTGTGTAC/	−64.4	−68.3	−183.4	−196.5	−7.49	−7.40	46.0	45.0
GCTAGGTATCC/	−69.6	−72.1	−194.6	−202.6	−9.24	−9.26	53.8	53.2
GCTATGTATCC/	−66.2	−69.8	−187.9	−199.3	−7.86	−8.01	47.6	47.7
T·T mismatches								
CAAATAAAG/[d]	−54.6	−50.6	−166.0	−152.1	−3.12	−3.40	23.1	23.6
CGAGTGTCC/	−55.9	−58.2	−158.3	−164.9	−6.83	−7.04	43.5	44.5
CGATTATCG	−48.6	−52.2	−140.6	−152.2	−4.95	−4.98	32.4	32.9
CGTCTGTCC/	−62.3	−61.6	−176.0	−173.6	−7.70	−7.77	47.4	47.9
CGTGTCTCC/	−60.3	−55.6	−172.0	−158.1	−6.90	−6.62	43.4	42.4
GGAATTTCC	−47.4	−53.3	−138.4	−157.3	−4.50	−4.54	29.5	30.5
GGACTGTCC	−59.4	−59.0	−168.5	−167.6	−7.08	−7.04	44.5	44.3
GGAGTCTCC	−51.1	−52.1	−146.7	−150.1	−5.62	−5.58	36.6	36.4
CATGATGCTAC/	−77.3	−73.1	−220.8	−209.8	−8.86	−8.09	50.3	47.6
CATGTTACTAC/	−61.4	−64.6	−175.5	−186.9	−6.99	−6.64	43.8	41.7
GATCTTTGTAC/	−77.7	−66.6	−227.7	−194.2	−7.09	−6.34	42.8	40.1
GGATGTATAGC/	−72.9	−66.1	−210.1	−189.3	−7.70	−7.36	45.9	45.1
CGCTAGAGTCGG/	−65.5	−72.2	−184.3	−204.1	−8.31	−8.86	50.0	51.3
GGCTGAGATCGC/	−79.7	−77.1	−226.0	−217.8	−9.64	−9.51	53.2	53.2

[a] Listed by mismatch type and by oligomer length. Experimental values are the averages of T_M^{-1} versus ln C_T and the curve fit parameters. Top strands are given in the 5′ to 3′ orientation. Underlined residues are mismatched. Sequences ending with a "/" are non-self-complementary. Experimental data for sequences without a literature reference are from this work. [b] Standard errors for experimental $\Delta G_{37}^°$, $\Delta H°$, and $\Delta S°$ are assumed to be 4%, 8%, and 8%, respectively. [c] Calculated for 10^{-4} M oligomer concentration for self-complementary sequences and 4×10^{-4} M for non-self-complementary sequences. [d] Aboul-ela et al. (36). [e] This sequence was not used to derive the mismatch nearest-neighbor parameters. [f] Arghavani et al. (37).

nearest-neighbor parameters, the problem is overdetermined (51 equations with 16 unknowns). This system of equations was solved by multiple linear regression using MATH-EMATICA version 2.1 (51). The data were arranged in a matrix form such that $\mathbf{G}_{MISM} = \mathbf{S} \cdot \mathbf{G}_{NN}$, where $\mathbf{G}_{MISM}$ is the column matrix containing the 51 experimental mismatch free energy increments (from equations analogous to eq 3) and $\mathbf{G}_{NN}$ is the column matrix containing the unknown 16 single symmetric mismatch dimer nearest neighbors. $\mathbf{S}$ is the stacking matrix with dimensions 51 × 16 which contains

the number of occurrences of the mismatch dimers in each sequence. Singular value decomposition (SVD) (*52*) was used to find the solution $\mathbf{G}_{NN}$ (*30, 31*). The column rank of the mismatch stacking matrix for our set of 51 oligonucleotides is 16, confirming that all 16 of the derived nearest-neighbor parameters are linearly independent (*31, 47, 53*). The same procedure was used to find $\Delta H°$ contributions. The nearest-neighbor $\Delta S°$ values were calculated from the equation:

$$\Delta S° = (\Delta H° - \Delta G°_{37})/310.15 \qquad (5)$$

The $\Delta S°$ values were also determined by linear regression and were within experimental error of those calculated using eq 5 (not shown).

Error Analysis. Errors in thermodynamic parameters were obtained using standard error propagation methods (*54*) and reflect the precision and reproducibility of the data (*41*) (see Supporting Information). Since instrument calibration and imposing the two-state approximation to fit the data can introduce systematic errors, conservative error estimates of 4%, 8%, and 8% in $\Delta G°_{37}$, $\Delta H°$, and $\Delta S°$ were assumed. These errors were propagated to the mismatch contribution used for SVD analysis as described previously (*31*). Resampling analysis of the data was performed to prove that no single measurement dramatically changed the solution of the multiple regression analysis and to verify that the assumed errors were approximately correct. Thirty resampling trials were created by randomly excluding one-third of the experimental equation data set in each trial. Resampling trials were only accepted if they contained at least one equation representative of every nearest neighbor. SVD was performed on each resampling trial. The nearest-neighbor parameters obtained from averaging the SVD results for the 30 resampling trials are the same as the nearest neighbors obtained by SVD analysis of the whole data set. Standard deviations were also calculated from the nearest-neighbor parameters obtained from these resampling trials. The resampling standard deviations are reported in Table 2 and do not depend on assumptions made about the magnitudes of experimental errors (*31*).

1H NMR Spectroscopy. Oligonucleotides were dissolved in 90% H_2O and 10% D_2O with 0.1 M NaCl, 10 mM disodium phosphate, and 0.1 mM Na_2EDTA at pH 7. Strand concentrations ranged from 0.2 to 1.0 mM. The internal standard for chemical shift reference was 3-(trimethylsilyl)-propionic-*2,2,3,3-d₄* acid. 1H NMR spectra were recorded at 10 °C using a Varian UNITY 500 MHz NMR spectrometer. One-dimensional exchangeable proton NMR spectra were recorded using the WATERGATE pulse sequence with "flip-back" pulse to suppress the water signal (*55, 56*). 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 μs, a sweep width of 12 kHz, and a gradient field strength of 10.0 G cm⁻¹ and duration of 1 ms. Data were multiplied by a 4.0 Hz line broadening exponential function and Fourier transformed on a Silicon Graphics Indigo[2] Extreme computer with Varian VNMR software. 1-D 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

Table 2: Linearly Independent Nearest-Neighbor Thermodynamic Parameters for A·A, C·C, G·G, and T·T Mismatch Formation in 1 M NaCl, pH 7[a]

propagation sequence	$\Delta H°$ (kcal/mol)	$\Delta H°$ (eu)	$\Delta G°_{37}$ (kcal/mol)
	A·A mismatches		
A<u>A</u>/T<u>A</u>	1.2 ± 2.5	1.7 ± 8.0	0.61 ± 0.13
C<u>A</u>/G<u>A</u>	−0.9 ± 2.3	−4.2 ± 7.3	0.43 ± 0.17
G<u>A</u>/C<u>A</u>	−2.9 ± 4.1	−9.8 ± 13.1	0.17 ± 0.23
T<u>A</u>/A<u>A</u>	4.7 ± 2.4	12.9 ± 7.7	0.69 ± 0.05
	C·C mismatches		
A<u>C</u>/T<u>C</u>	0.0 ± 2.1	−4.4 ± 6.5	1.33 ± 0.09
C<u>C</u>/G<u>C</u>	−1.5 ± 1.1	−7.2 ± 4.7	0.70 ± 0.50
G<u>C</u>/C<u>C</u>	3.6 ± 3.2	8.9 ± 9.8	0.79 ± 0.09
T<u>C</u>/A<u>C</u>	6.1 ± 1.1	16.4 ± 3.5	1.05 ± 0.06
	G·G mismatches		
A<u>G</u>/T<u>G</u>	−3.1 ± 1.3	−9.5 ± 4.0	−0.13 ± 0.09
C<u>G</u>/G<u>G</u>	−4.9 ± 1.1	−15.3 ± 3.3	−0.11 ± 0.11
G<u>G</u>/C<u>G</u>	−6.0 ± 2.5	−15.8 ± 7.8	−1.11 ± 0.13
T<u>G</u>/A<u>G</u>	1.6 ± 0.8	3.6 ± 2.5	0.44 ± 0.14
	T·T mismatches		
A<u>T</u>/T<u>T</u>	−2.7 ± 4.1	−10.8 ± 13.1	0.69 ± 0.23
C<u>T</u>/G<u>T</u>	−5.0 ± 1.4	−15.8 ± 13.9	−0.12 ± 0.23
G<u>T</u>/C<u>T</u>	−2.2 ± 1.1	−8.4 ± 3.2	0.45 ± 0.05
T<u>T</u>/A<u>T</u>	0.2 ± 1.8	−1.5 ± 5.8	0.68 ± 0.08

[a] Extra significant figures are given to allow accurate calculation of the T_M and $\Delta G°_{37}$. Underlined residues are mismatched. Errors are resampling standard deviations (see text). These parameters should not be used to calculate the stability of oligonucleotides with mismatches in the terminal or penultimate positions.

acquired in an interleaved fashion in blocks of 16 scans to minimize subtraction errors due to long-term instrument drift. A total of 3200−6400 scans were collected for each FID.

RESULTS

Thermodynamic Data. Plots of T_M^{-1} vs ln C_T were linear over the entire 80−100-fold range in concentration with correlation coefficients ≥0.98 (see Supporting Information). Table S-1 shows the thermodynamic parameters derived from fits of melting curves and T_M^{-1} vs ln C_T plots (see Supporting Information). The agreement in $\Delta H°$ values of the two methods is within 15% for all the sequences used to derive the nearest-neighbor parameters. Therefore, the two-state approximation appears to be valid for these sequences. Since the data from both methods are equally reliable, the average parameters (Table 1) were used in linear regression analysis to derive nearest-neighbor parameters. The fact that none of the sequences were outliers in the SVD fit of nearest-neighbor parameters provides additional confidence in the reliability of the two-state approximation. Table 2 lists the 16 linearly independent nearest-neighbor parameters for like with like base mismatches which were derived from the experimental data in Table 1 using SVD. Table 1 also presents the predictions made by the mismatch nearest-neighbor parameters (Table 2) and the Watson−Crick nearest-neighbor parameters (*31*). The measured thermodynamics of the 51 sequences are predicted with average deviations of $\Delta G°_{37}$ = 3.3%, $\Delta H°$ = 7.4%, $\Delta S°$ = 8.1%, and T_M = 1.1 °C. Self-complementary sequences and non-self-complementary sequences are predicted with the same accuracy, demonstrating the consistency of our approach.

Circular Dichroism. Circular dichroism spectra for duplexes in 1 M Na⁺ at pH = 7 acquired for the four types of
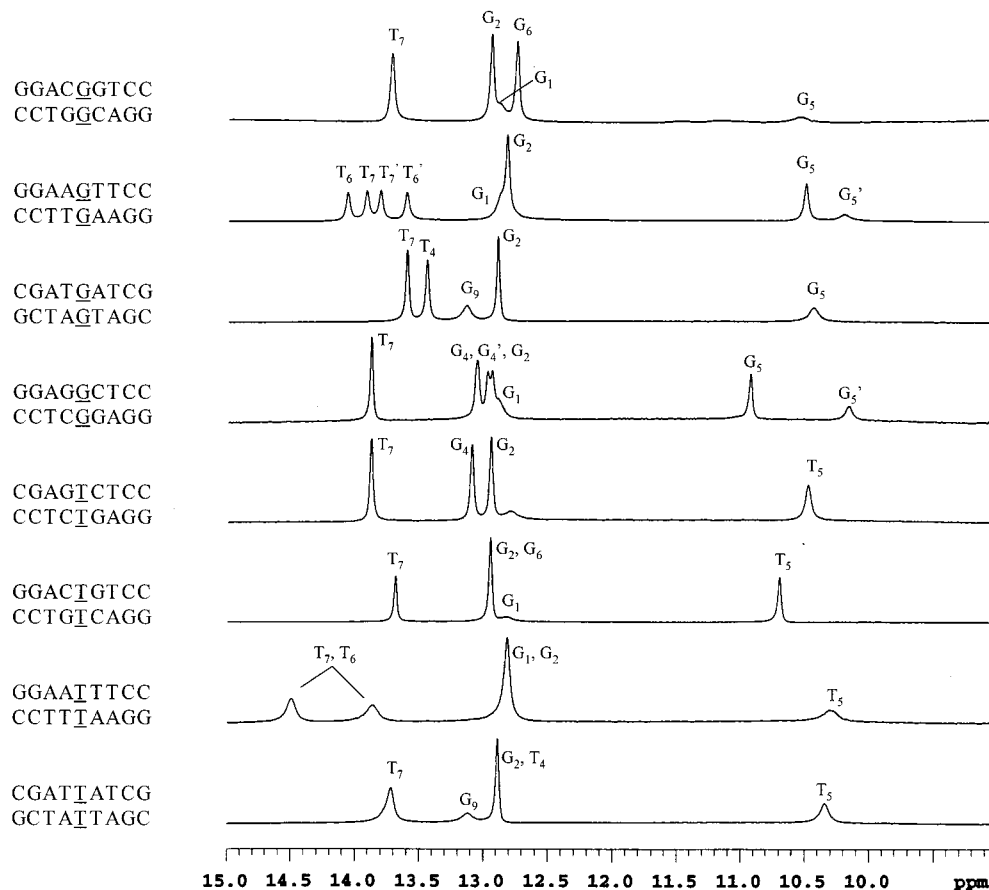
FIGURE 1: 500 MHz NMR spectra of the exchangeable imino region (9−15 ppm) at 10 °C in 0.1 M NaCl, 10 mM disodium phosphate, and 0.1 mM Na$_2$EDTA at pH 7.0 in 90% H$_2$O/10% D$_2$O of self-complementary sequences with G·G and T·T mismatches.

mismatches show shapes characteristic of B-helical form, with minor variations presumably due to the different mismatch structures (data not shown).

NMR Spectroscopy. Figure 1 shows the exchangeable imino region (9−15 ppm) of the 1-D ^{1}H NMR of eight sequences containing G·G and T·T mismatches. 1-D NOE difference spectroscopy was used to assign peaks (see Supporting Information). The mismatch imino proton chemical shift and line width are dependent on sequence context (see Discussion).

DISCUSSION

Validity of the Nearest-Neighbor Model for Internal Like with Like Base Mismatches. The nearest-neighbor parameters predict the thermodynamics of all sequences within an average deviation of $\Delta G°_{37} = 3.3\%$, $\Delta H° = 7.4\%$, $\Delta S° = 8.1\%$, and $T_M = 1.1$ °C. These average deviations are comparable to the ones observed for the prediction of Watson−Crick base pairs (*31*) as well as G·T (*31*), G·A (*50*), C·T (*34*), and A·C (*35*) internal mismatches. For G·T mismatches, for instance, average deviations were $\Delta G°_{37} = 5.1\%$, $\Delta H° = 7.5\%$, $\Delta S° = 8.0\%$, and $T_M = 1.4$ °C (*31*). Consequently, it can be concluded that the nearest-neighbor model is a good approximation for both Watson−Crick base pairs and all single mismatches.

Mismatches in Penultimate and Penpenultimate Positions. Since a duplex folds into its thermodynamically lowest energy structure, mismatches near the terminus of a duplex (particularly the penultimate and penpenultimate positions)

often behave as terminal mismatches instead of internal mismatches. Consider the following self-complementary duplex structures:

```
A A G C G C A T        A A G C G C A T
| × | | | | × |   vs    A A       A A
T A C G C G A A        T A C G C G A A
```
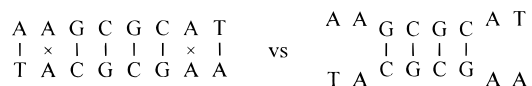
Our data predict that the structure on the right, without terminal A·T hydrogen bonding, is more stable by approximately 4 kcal/mol (S. Varma, G. Jenkins, and J. SantaLucia, Jr., unpublished results). This effect is presumably due to unfavorable steric interactions that occur when destabilizing mismatches are placed in the interior of a duplex (*31*). The data presented here (Table 1) as well as our previous work on other mismatches (*31, 34, 35, 50*) indicate that mismatches more than three base pairs from the end show no position dependence and are well predicted with the nearest-neighbor parameters reported here.

Trends in Nearest-Neighbor Parameters. The observed order of single mismatch stability at 37 °C is G·G > T·T ≈ A·A > C·C. However, a large context dependence is observed. This context dependence is related to the different stacking and H-bonding interactions formed by mismatches in different contexts (Figure 2). Various studies have investigated the context dependence of G·G mismatch structures. Cognet et al. (*57*) and Casati et al. (*58*) found that G·G mismatches in the contexts AGG and GGA form *syn-anti* conformations stabilized by two hydrogen bonds. Lane and Peck (*59*) also found that G·G mismatches in CGG,
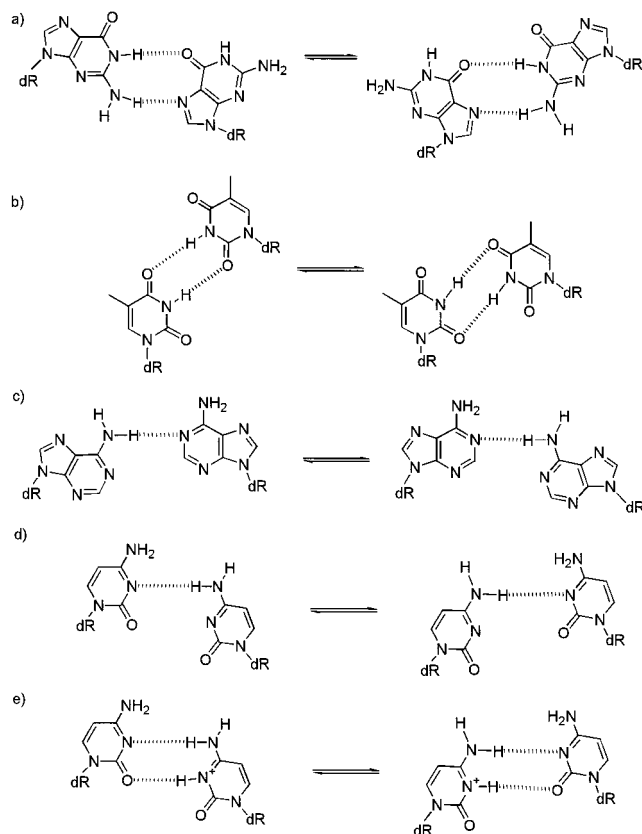
FIGURE 2: Structures of like with like base mismatches as described in the literature (a) G·G (*57–59, 62*) (b) T·T (*63, 64, 66*), (c) A·A (*63–65*), (d) C·C (*67*), and (e) protonated C·C (*67*).

TGA, and CGA contexts are in *syn-anti* conformation, but they also observed a rapid exchange between the *syn-anti* and *anti-syn* conformations of the mismatches. Borden et al. (*60*) studied a G·G mismatch in the CGA context and concluded that the main conformation was *anti-anti* with no strong hydrogen bonding. Faibis et al. (*61*) also studied the CGA context, but they suggested an *anti-anti* conformation stabilized by one hydrogen bond. A crystal structure of the sequence studied by Borden et al. (1992) indicated that the G·G mismatch in the CGA context forms a *syn-anti* conformation stabilized by two hydrogen bonds (*62*). Gervais et al. (*63*) studied A·A mismatches in the CAA context and found base pairing between adenines in the *anti* conformation stabilized by one hydrogen bond. An NMR study of A·A in the CAG context showed that only a quickly exchanging hydrogen-bonded structure could exist (*64*). Another NMR study showed that an A·A mismatch in the AAA context was also in *anti-anti* conformation and probably stabilized by one hydrogen bond (*65*). T·T mismatches in the CTA context were studied by NMR by two groups and were observed in a wobble structure involving two imino to carbonyl hydrogen bonds (*63, 66*). Imino proton NMR of T·T mismatches in the CTG context recorded by Arnold et al. (*64*) also indicated wobble base pairing. The NMR solution structure of C·C mismatch in the CCA context has led to a structural model involving only one hydrogen bond in the mismatch (*67*). In another structure, the C·C mismatch formed no hydrogen bonds but instead the cytosine bases protruded into the grooves of the helix to form the "e motif" (*68*).

We can rationalize the relative stability of these mismatches in terms of structure by considering stacking interactions and hydrogen bonding. G·G mismatches, which are the most stable, have a high stacking potential due to their purine rings and are stabilized by two hydrogen bonds in most contexts (Figure 2a) (*57–59, 62*). A·A mismatches have a high stacking potential but generally form only one hydrogen bond (Figure 2c) (*63–65*). T·T mismatches have a lower stacking potential (pyrimidine ring) but are stabilized by two hydrogen bonds (Figure 2b) (*63, 64, 66*) in most contexts. This rationalizes the lower stability of A·A and T·T compared to G·G. C·C mismatches have a low stacking propensity and are stabilized by only one hydrogen bond (Figure 2d) (*67*). Consequently, C·C mismatches are the least stable of the four like with like base mismatches.

The trend in stability for the Watson–Crick pairs on the 5′ side of the mismatch is G·C ≥ C·G ≥ A·T ≥ T·A. This result is consistent with the fact that G·C pairs are more stable than A·T pairs due to the formation of an extra hydrogen bond. The higher stability of G·C vs C·G is consistent with observations for other mismatches from our laboratory. Due to the symmetry of like with like mismatches, the trend on the 3′ side of the mismatch is C·G ≥ G·C ≥ T·A ≥ A·T.

Comparison with the Stabilities of Other Internal Mismatches. Figure 3 shows the detailed stability trend in all 10 possible trimer contexts with all 16 possible X·Y pairs. The overall trend in pairing stability is dependent on context, but to a first approximation, the trend averaged over all contexts is G·C > A·T > G·G > G·T ≈ G·A > A·C+ > T·T ≈ A·A ≈ C·C+ > T·C ≥ A·C ≥ C·C. This trend qualitatively agrees with the trend observed by Ke and Wartell (*69*) in temperature gradient gel electrophoresis studies of mismatches in DNA polymers. G·G is stabilizing in 7 out of 10 trimer contexts, with $\Delta G°_{37}$ values ranging from −2.22 (GGC/CGG) to 0.88 kcal/mol (TGA/AGT). The GGC/CGG trimer is the most stable of all the possible internal mismatch trimers. G·G is more stable than all other mismatches except in four contexts where G·T mismatches are more stable (ATG/TGA, CTA/GGT, CGA/GTT and CGT/GTA). A·A is destabilizing in all contexts with $\Delta G°_{37}$ values ranging from 0.34 (GAC/CAG) to 1.38 kcal/mol (TAA/AAT). For all contexts A·A is less stable than G·T, G·G, and G·A, more stable than C·C and C·A, and more or equally stable than C·T. The relative stability of A·A and T·T is context dependent, and no clear trend is observed. T·T is destabilizing in all contexts but one (CTG/GTC) with $\Delta G°_{37}$ values ranging from −0.24 (CTG/GTC) to 1.38 kcal/mol (ATA/TTT). For all contexts, T·T is more stable than C·C, C·A, and C·T except in ACT/TTA and ACA/TTT. T·T is usually less stable than G·G, G·T, and G·A except in CXA/GYT and CXG/GYC, where G·A is less stable. C·C is destabilizing in all contexts with $\Delta G°_{37}$ values ranging from 1.40 (CCG/GCC) to 2.66 kcal/mol (ACT/TCA). This latter value is the most destabilizing value obtained for any mismatch in any context. The relative stability of C·C, and A·C, and C·T is context dependent and does not follow a clear trend.

pH Dependence of the Mismatch Stabilities. For each X·X mismatch two duplexes were also melted at pH 4.9 to evaluate the influence of the pH on mismatch stability (see Supporting Information). Duplexes with single A·A and T·T
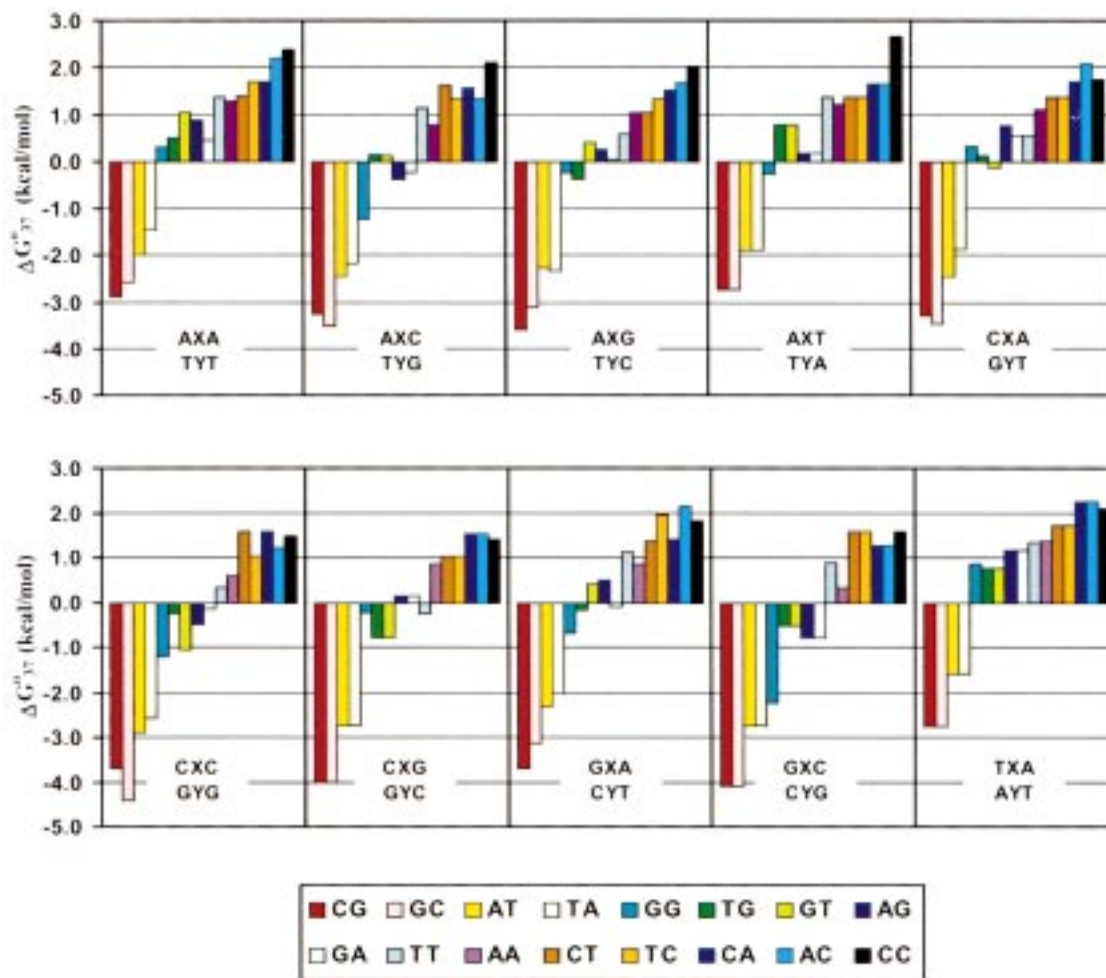
FIGURE 3: Relative stabilities of the 10 possible trimer contexts containing the 16 possible central mismatches or Watson−Crick base pairs. Data are deduced from this work (Table 2) and refs *31−35*. See Table 2 and references above for error bars. Note that only 136 out of 160 trimers shown are unique. For example, GGC/CAG is equal to GAC/CGG.

mismatches were on average 0.7 kcal/mol less stable at pH 4.9 than at pH 7. This is a much larger destabilization than we have observed for Watson−Crick G·T and G·A mismatches. The duplex GGCTGAGATCGC·GCGTTCTCTGCC contains two T·T mismatches, and a 2.20 kcal/mol total destabilization is observed at pH 4.9 compared with pH 7.0 (see Supporting Information). This suggests that destabilization of duplexes with A·A or T·T may be due to specific protonation of mismatch heteroatoms. Further investigation is required to reach a conclusive explanation. The stability of duplexes with G·G mismatches was not significantly affected by decreasing the pH. Duplexes with C·C mismatches were on average 0.7 kcal/mol more stable at pH 4.9 than at pH 7. This observation is probably related to the cytosine protonation upon lowering the pH (Figure 3d,e) as described previously (*38, 66, 67, 70*).

Salt Dependence Corrections. Gaffney and Jones (*71*) reported thermodynamic data for the sequences GGTTXTTGG· CCAAYAACC, with X and Y = A, C, G, and T, melted in 0.11 M Na⁺ buffer. The data reported by Gaffney and Jones are in excellent agreement with our predicted values (Table 3). The predicted thermodynamics in 1 M NaCl were calculated using the mismatch and Watson−Crick nearest-neighbor parameters derived in this work and our previous work (*31, 33−35*). $\Delta H°$ is assumed to be independent of sodium concentration. $\Delta G°_{37}$ and $\Delta S°$ were corrected for salt

dependence according to the empirical equations (*32*):

$$\Delta G°_{37}{}^{[Na]} = \Delta G°_{37}{}^{1\,M\,Na} - 0.114N \ln [Na^+] \qquad (6)$$

$$\Delta S°^{[Na]} = \Delta S°^{1\,M\,Na} + 0.368N \ln [Na^+] \qquad (7)$$

where $\Delta G°_{37}{}^{1\,M\,Na}$ is the prediction for 1 M sodium concentration and $\Delta G°_{37}{}^{[Na]}$ is the prediction for a given sodium concentration. N is the total number of phosphates in the duplex divided by 2 ($N = 8$ for all duplexes in Table 3). On average, for the 12 mismatch-containing sequences studied by Gaffney and Jones, $\Delta G°_{37}$ and T_M are predicted within 0.32 kcal/mol and 2.7 °C, respectively. For comparison, for the four duplexes that contain only Watson−Crick pairs, $\Delta G°_{37}$ and T_M are predicted within 0.42 kcal/mol and 1.7 °C, respectively. The predictions for the two sequences with A·C mismatches (Table 3) are not as good as the predictions for other mismatches. This is consistent with our previous observations for A·C mismatches (*35*). Since the Gaffney and Jones mismatch data (*71*) were used neither to derive the mismatch nearest-neighbor parameters nor to derive the salt dependence equations, they provide independent verification of the mismatch parameters and the salt dependence equations. The fact that the salt dependence of eqs 6 and 7 is applicable to both Watson−Crick pairs and mismatches is fully consistent with counterion condensation theory (*72,*

Table 3: Experimental and Predicted Thermodynamic Parameters of Duplex Formation of Oligonucleotides in 0.11 M Na+ Solution

sequences[a]	$\Delta H°$ (kcal/mol)		$\Delta S°$ (eu)		$\Delta G°_{37}$ (kcal/mol)		$T_M{}^d$ (°C)	
	experiment[b]	prediction[c]	experiment[b]	prediction[c]	experiment[b]	prediction[c]	experiment[b]	prediction[c]
GGTTATTGG CCAATAACC	−66.0	−62.9	−192.2	−183.3	−6.39	−6.06	40.4	38.9
GGTTCTTGG CCAAGAACC	−67.0	−64.5	−191.8	−184.8	−7.50	−7.18	45.7	44.4
GTTGTTGG CCAACAACC	−71.0	−65.4	−202.9	−186.7	−8.06	−7.49	47.8	45.8
GGTTTTTGG CCAAAAACC	−68.0	−64.3	−196.5	−186.0	−7.04	−6.60	43.4	41.5
GGTTATTGG CCAAAAACC	−53.0	−42.7	−161.7	−126.9	−2.86	−3.30	21.4	20.6
GGTTATTGG CCAAGAACC	−53.0	−48.6	−156.0	−143.3	−4.63	−4.16	31.0	27.6
GGTTGTTGG CCAAAAACC	−61.0	−46.1	−186.8	−136.7	−3.06	−3.72	24.2	24.4
GGTTATTGG CCAACAACC	−51.0	−39.8	−153.3	−119.0	−3.46	−2.91	24.1	16.8
GGTTCTTGG CCAAAAACC	−56.0	−38.6	−173.7	−116.8	−2.12	−2.39	18.5	12.7
GGTTCTTGG CCAACAACC	−57.0	−42.4	−178.1	−129.5	−1.76	−2.22	17.1	13.6
GGTTCTTGG CCAATAACC	−56.0	−48.5	−171.1	−147.0	−2.95	−2.90	22.6	20.2
GGTTTTTGG CCAACAACC	−61.0	−46.8	−187.5	−140.6	−2.85	−3.21	23.3	21.5
GGTTGTTGG CCAAGAACC	−54.0	−53.1	−160.0	−157.5	−4.38	−4.29	29.7	29.1
GGTTGTTGG CCAATAACC	−65.0	−51.1	−194.9	−151.5	−4.56	−4.10	31.8	27.7
GGTTTTTGG CCAAGAACC	−54.0	−48.8	−160.7	−143.8	−4.17	−4.19	28.6	27.8
GGTTTTTGG CCAATAACC	−58.0	−51.4	−176.1	−155.4	−3.39	−3.23	25.2	22.9

[a] Listed by mismatch type. Top strands are conventionally given in the 5' to 3' orientation. Underlined residues are mismatched. [b] Experimental $\Delta G°_{37}$ data are calculated from $\Delta G°_{25}$ and $\Delta H°$ reported in ref 71 assuming $\Delta C_p° = 0$. [c] Predicted values are corrected for salt concentration using eqs 6 and 7 (see text). [d] Calculated for 4×10^{-4} M oligomer concentration.

73), since the axial charge density for sequences with and without single mismatches is very similar.

Comparison with RNA Studies. To our knowledge, no systematic study of RNA internal single mismatch stabilities in solution has been carried out. However, some mismatches in selected contexts have been studied. Alkema et al. (74), studying mismatches in the GXC/CXG context, observed that U·U is less stable than A·A and C·C mismatches. In DNA, the order of stability is GAC/CAG > GTC/CTG > GCC/CCG. Moreover, the data reported for RNA mismatches in the CXG/GXC context (70, 75) indicate that A·A is more stable than U·U. In DNA, CTG/GTC is more stable than CAG/GAC. On the basis of these limited data, it appears that the stability trends for mismatches in DNA and RNA are different.

NMR Data. More stable contexts have T·T imino proton resonances (10−12 ppm) that are sharper and downfield shifted. Presumably, the imino proton resonances of more stable mismatches are sharper because they exchange less with solvent. The chemical shifts of the imino protons, however, are difficult to interpret because both hydrogen-bonding and ring current effects are important. Interestingly, the T6 and T7 imino resonances of (GGAAGTTCC)₂ and

G4 of (GGAGGCTCC)₂ are split (Figure 1). Further, these spectra show two peaks for the G5 imino proton corresponding to the mismatch. This observation could be either the consequence of a breakage of symmetry or the result of the coexistence of two differently populated conformations in slow exchange on the NMR time scale. At different temperatures (data not shown) the areas under the G5 peaks are constant and sum to one proton, which would be unlikely if this sequence were forming an alternative base-paired structure such as a slipped duplex or hairpin. Moreover, the UV melts appear to be two state, which would also be unlikely if alternative structures were formed. Thus, it appears that the slow exchange between asymmetric G^syn· G^anti is the most likely explanation for the observation of two resonances for G5. Further NMR studies of these sequences are required to obtain a better understanding of the structures involved.

Comparison of Mismatch Stabilities with Polymerase Fidelity. The fidelity of *Escherichia coli* DNA polymerases II and III have been extensively studied (76, 77). The ratio of polymerase insertion of a nucleotide next to various base pairs resulting in like with like base mismatches is not in good correlation with the stability of the corresponding

nearest neighbors. Therefore, we conclude that thermodynamics plays a secondary role in polymerase fidelity. As suggested elsewhere (*31, 78, 79*), geometric factors are most likely to be responsible for the different ratios observed for mismatch insertion.

Comparison of Mismatch Stabilities with Mismatch Repair Efficiency. The efficiency of postreplicative repair of like with like base mismatch has been investigated in vivo for different systems (*80−83*). These different studies concluded that the efficiency of mismatch repair depends on the mismatch type and its context. The consensus trend deduced from these investigations shows that G·G and G·T mismatches are generally the most efficiently repaired mismatches whereas G·A and C·C are generally the least efficiently repaired mismatches. The repair efficiency of the four other mismatches is intermediate. Even though G·G and G·T mismatches are the most stable mismatches and C·C is the least stable, repair efficiency does not seem directly correlated to thermodynamic stability. For instance, G·A, which is a relatively stable mismatch, constitutes a poor substrate for repair. Therefore, it is likely that both thermodynamic and structural factors play an important role in mismatch repair mechanisms (*35*).

Application to Triplet Repeat Stability Prediction. DNA triplet repeat sequences have been shown to form hairpin structures (*7−12*) that may be responsible for genome expansion or deletion during replication (*13−16*). The triplet repeat hairpin structures involve consecutive symmetric single mismatches separated by two Watson−Crick base pairs. The thermodynamic stability of the following duplex is well predicted by our nearest-neighbor parameters (Table 1):

```
5'  G G A C A G C A G C A G A G G  3'
    | | | | × | | × | | × | | | |
3'  C C T G A C G A C G A C T C C  5'
```

This observation seems to rule out significant next-nearest-neighbor effects for closely spaced single A·A mismatches. Our parameters also make good $\Delta G°_{37}$ predictions for $(CAG)_{10}$ and $(CAG)_{25}$: $\Delta G°_{37 \text{ predicted}} = -2.7$ and -6.9 kcal/mol whereas the values reported by Gacy and McMurray (*14*) are -2.2 and -5.3 kcal/mol, respectively. The predictions were made using the mismatch parameters in Table 2, the Watson−Crick nearest-neighbor parameters (*31*), the salt dependence in eq 6 (assuming $N = 12$ or $N = 33$), estimated hairpin loop penalties of 2.4 or 3.2 kcal/mol (J. SantaLucia, Jr., unpublished results), and a terminal mismatch contribution of -1.9 kcal/mol (S. Varma and J. SantaLucia, Jr., unpublished results). On the other hand, the transition $\Delta H°$ values were not well predicted for $(CAG)_{10}$ and $(CAG)_{25}$. The poor $\Delta H°$ predictions suggest that the transitions are likely not two state. To further investigate the thermodynamic basis of triplet repeat diseases, we intend to test our parameters with sequences containing a larger number of repeats as well as different types of triplet repeats.

ACKNOWLEDGMENT

SUPPORTING INFORMATION AVAILABLE

One table showing thermodynamic parameters of duplex formation derived from $1/T_M$ vs ln C_T plots and from averaging the fits of individual melting curves of oligonucleotides with A·A, C·C, G·G, and T·T mismatches in 1 M NaCl at pH 7 and 4.9, one figure showing typical $1/T_M$ vs ln C_T plots, and seven figures showing 1D NOE difference spectra. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Goodman, M. F., Creighton, S., Bloom, L. B., and Petruska, J. (1993) *Crit. Rev. Biochem. Mol. Biol. 28*, 83−126.
2. Bhattacharyya, A. (1989) *J. Mol. Biol. 209*, 583−597.
3. Leonard, G. A., Booth, E. D., and Brown, T. (1990) *Nucleic Acids Res. 18*, 5617−5623.
4. Plum, G. E., Grollman, A. P., Johnson, F., and Breslauer, K. J. (1995) *Biochemistry 34*, 16148−16160.
5. Brown, T. (1995) *Aldrichimica Acta 28*, 15−20.
6. Lindahl, T. (1993) *Nature 362*, 709−715.
7. Smith, G. K., Jie, J., Fox, G. E., and Gao, X. (1995) *Nucleic Acids Res. 23*, 4303−4311.
8. Hopkin, K. (1995) *Journal of NIH Research 7*, 45−48.
9. Mariappan, S. V. S., Garcia, A. E., and Gupta, G. (1996) *Nucleic Acids Res 24*, 775−783.
10. Zheng, M., Huang, X., Smith, G. K., Yang, X., and Gao, X. (1996) *J. Mol. Biol. 264*, 323−336.
11. Chen, X., Mariappan, S. V. S., Catasti, P., Ratliff, R., Moyzis, R. K., Laayoon, A., Smith, S. S., Bradbury, E. M., and Gupta, G. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 5199−5203.
12. Mitas, M., Yu, A., Dill, J., Kamp, T. J., Chambers, E. J., and Haworth, I. S. (1995) *Nucleic Acids Res. 23*, 1050−1059.
13. Gacy, A. M., Goellner, G., Juranic, N., Macura, S., and McMurray C. T. (1995) *Cell 81*, 533−540.
14. Gacy, A. M., and McMurray, C. T. (1998) *Biochemistry 37*, 9420−9434.
15. Bowater, R. P., Jaworski, A., Larson, J. E., Parniewski, P., and Wells, L. D. (1997) *Nucleic Acids Res. 25*, 2861−2868.
16. Mitas, M. (1997) *Nucleic Acids Res. 25*, 2245−2253.
17. Ashley, T. C., Jr., and Warren, S. T. (1995) *Annu. Rev. Genet. 29*, 709−728.
18. Timchenko, L. T., and Caskey, C. T. (1996) *FASEB J.* 1589−1597.
19. Sutherland, G. R., and Richards, R. I. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 3636−3641.
20. 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.
21. Southern, E. M. (1975) *J. Mol. Biol. 98*, 503−517.
22. Kanazawa, H., Noumi, T., and Futai, M. (1986) *Methods Enzymol. 126*, 595−603.
23. Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., and Sekiya, T. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2766−2770.
24. Nielsen, D. A., Novoradovsky, A., and Goldman, D. (1995) *Nucleic Acids Res. 23*, 2287−2291.
25. 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.
26. Mirzabekov, A. D. (1994) *Trends Biotechnol. 12*, 27−32.
27. Freier, S. M. (1993) in *Antisense Research and Applications* (Crooke, S. T., and Lebleu, B., Eds.) pp 67−82, CRC Press, Boca Raton, FL.
28. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol. 154*, 367−382.
29. Wallace, R. W. (1997) *Mol. Med. Today 3*, 384−389.
30. SantaLucia, J., Jr., Allawi, H. T., and Seneviratne, P. A. (1996) *Biochemistry 35*, 3555−3562.
31. Allawi, H. T., and SantaLucia, J., Jr. (1997) *Biochemistry 36*, 10581−10594.
32. SantaLucia, J., Jr. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 1460−1465.

33. Allawi, H. T., and SantaLucia, J., Jr. (1998) *Biochemistry 37*, 2170−2179.

34. Allawi, H. T., and SantaLucia, J., Jr. (1998) *Nucleic Acids Res. 26*, 2694−2701.

35. Allawi, H. T., and SantaLucia, J., Jr. (1998) *Biochemistry 37*, 9435−9444.

36. Aboul-ela, F., Koh, D., Tinoco, I., Jr., and Martin, F. H. (1985) *Nucleic Acids Res. 13*, 4811−4824.

37. Arghavani, M. B., SantaLucia, J., Jr., and Romano, L. J. (1998) *Biochemistry 37*, 8575−8583.

38. Brown, T., Leonard, G. A., and Booth, E. D. (1990) *J. Mol. Biol. 212*, 437−440.

39. Chou, S.-H., Flynn, P., and Reid, B. (1989) *Biochemistry 28*, 2422−2435.

40. Richards, E. G. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) p 597, CRC Press, Cleveland, OH.

41. McDowell, J. A., and Turner, D. H. (1996) *Biochemistry 35*, 14077−14089.

42. Petersheim, M., and Turner, D. H. (1983) *Biochemistry 22*, 256−263.

43. Borer, P. N., Dengler, B., Tinoco, I., Jr., and Uhlenbeck, O. C. (1974) *J. Mol. Biol. 86*, 843−853.

44. Freier, S. M., Sugimoto, N., Sinclair, A., Alkema, D., Neilson, T., Kierzek, R., Caruthers, M. H., and Turner, D. H. (1986) *Biochemistry 25*, 3214−3219.

45. Marky, L. A., and Breslauer, K. J. (1987) *Biopolymers 26*, 1601−1620.

46. 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.

47. Gray, D. M., and Tinoco, I., Jr. (1970) *Biopolymers 9*, 223−244.

48. Tinoco, I., Borer, P. N., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., and Gralla, J. (1973) *Nat. New Biol. 246*, 40−41.

49. Uhlenbeck, O. C., Borer, P. N., Dengler, B., and Tinoco, I., Jr. (1973) *J. Mol. Biol. 73*, 483−496.

50. Allawi, H. T., and SantaLucia, J., Jr. (1998) *Biochemistry 37*, 2170−2179.

51. Wolfram, S. (1992) *MATHEMATICA version 2.1*, Wolfram Research, Inc.

52. Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1989) *Numerical Recipes*, Cambridge University Press, New York.

53. Doktycz, M. J., Goldstein, R. F., Paner, T. M., Gallo, F. J., and Benight, A. S. (1992) *Biopolymers 32*, 849−864.

54. Bevington, P. R. (1969) *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York.

55. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR 2*, 661−665.

56. Lippens, G., Dhalluin, C., and Wieruszeski, J.-M. (1995) *J. Biomol. NMR 5*, 327−331.

57. Cognet, J. A. H., Gabarro-Arpa, J., LeBret, M., van der Marel, G. A., van Boom, J. H., and Fazaferley, G. V. (1991) *Nucleic Acids Res. 19*, 6771−6779.

58. Catasti, P., Gupta, G., Garcia, A. E., Ratliff, R., Hong, L., Yau, P., Moyzis, R. K., and Bradbury, E. M. (1994) *Biochemistry 33*, 3819−3830.

59. Lane, A., and Peck, B. (1995) *Eur. J. Biochem. 230*, 1073−1087.

60. Borden, K. L. B., Jenkins, T. C., Skelly, J. V., Brown T., and Lane, A. N. (1992) *Biochemistry 31*, 5411−5422.

61. Faibis, V., Cognet, J. A. H., Sowers, L. C., and Fazakerley, G. V. (1996) *Biochemistry 35*, 14452−14464.

62. Skelly, J. V., Edwards, K. J., Jenkins, T. C., and Neidle, S. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 804−808.

63. Gervais, V., Cognet, J. A. H., Le Bret, M., Sowers, L. C., and Fazakerley, G. V. (1995) *Eur. J. Biochem. 228*, 279−290.

64. Arnold, F. H., Wolk, S., Cruz, P., and Tinoco, I., Jr. (1987) *Biochemistry 26*, 4068−4075.

65. Maskos, K., Gunn, B. M., LeBlanc, D. A., and Morden, K. M. (1993) *Biochemistry 32*, 3583−3595.

66. Kouchakdjian, M., Li, B. F. L., Swann, P. F., and Patel, D. J. (1988) *J. Mol. Biol. 202*, 139−155.

67. Boulard, Y., Cognet, J. A. H., and Fazakerley, G. V. (1997) *J. Mol. Biol. 268*, 331−347.

68. Sarma, M. H., Gupta, G., and Sarma, R. H. (1986) *FEBS Lett. 205*, 223−229.

69. Ke, S.-H., and Wartell, R. M. (1993) *Nucleic Acids Res. 21*, 5137−5143.

70. SantaLucia, J., Jr., Kierzek, R., and Turner, D. H. (1991) *Biochemistry 30*, 8242−8251.

71. Gaffney, B. L., and Jones, R. A. (1989) *Biochemistry 28*, 5881−5889.

72. Manning, G. (1978) *Q. Rev. Biophys. 11*, 179−246.

73. Record, M. T., Jr., and Lohman, T. M. (1978) *Biopolymers 17*, 159−166.

74. Alkema, D., Hader, P. A., Bell, R. A., and Neilson, T. (1982) *Biochemistry 21*, 2109−2117.

75. Peritz, A. E., Kierzek, R., Sugimoto, N., and Turner, D. H. (1991) *Biochemistry 30*, 6428−6436.

76. Cay, H., McEntee, K., Kunkel, T. A., and Goodman, M. F. (1995) *J. Biol. Chem. 270*, 15327−15335.

77. Bloom, L. B., Chen, X., Fygenson, D. K., Turner, J., O'Donnel, M., and Goodman, M. F. (1997) *J. Biol. Chem. 272*, 27919−27930.

78. Goodman, M. F. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 10493−10495.

79. Allawi, H. T., and SantaLucia, J., Jr. (1998) *Nucleic Acids Res. 26*, 4925−4934.

80. Kramer, B., Kramer, W., and Fritz, H.-J. (1984) *Cell 38*, 879−887.

81. Dohet, C., Wagner, R., and Radman, M. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 503−505.

82. Su, S.-S., Lahue, R. S., Au, K. G., and Modrich, P. (1988) *J. Biol. Chem. 263*, 6829−6835.

83. Gasc, A.-M., Sicard, A.-M., and Claverys, J.-P. (1989) *Genetics 120*, 29−36.