# Thermodynamics of internal C·T mismatches in DNA

## Hatim T. Allawi and John SantaLucia, Jr*

Department of Chemistry, Wayne State University, Detroit, MI 48202, USA

## ABSTRACT

Thermodynamics of 23 oligonucleotides with internal single C·T mismatches were obtained by measuring UV absorbance as a function of temperature. Results from these 23 duplexes were combined with three measurements from the literature to derive nearest-neighbor thermodynamic parameters for seven linearly independent trimer sequences with internal C·T mismatches. The data show that the nearest-neighbor model is adequate for predicting thermodynamics of oligonucleotides with internal C·T with average deviations for $\Delta G°_{37}$, $\Delta H°$, $\Delta S°$ and $T_m$ of 6.4%, 9.9%, 10.6%, and 1.9°C respectively. C·T mismatches destabilize the duplex in all sequence contexts. The thermodynamic contribution of C·T mismatches to duplex stability varies weakly depending on the orientation of the mismatch and its context and ranges from +1.02 kcal/mol for GCG/CTC and CCG/GTC to +1.95 kcal/mol for TCC/ATG.

## INTRODUCTION

DNA mismatches occur as a result of errors during replication (1), due to heteroduplex formation during homologous recombination (2) and mutagenic chemicals and ionizing radiation or spontaneous deamination (3). Mismatches also occur in the secondary structures of single-stranded DNA viruses (4–6). In addition to stable canonical Watson–Crick base pairs (G·C and A·T) there are eight possible mispairs of varying stability and structure, namely A·A, A·C, C·C, C·T, G·G, G·A, G·T and T·T. In order to understand the origins of various mismatch occurrences and to help in our interpretation of mismatch recognition and repair mechanisms, thermodynamics and structures of these mismatches need to be determined.

Several molecular biological techniques require accurate predictions of matched versus mismatched hybridization thermodynamics, such as PCR (7), sequencing by hybridization (8), gene diagnostics (9) and antisense oligonucleotide probes (9–11). In addition, recent developments of oligonucleotide chip arrays as means for biochemical assays and DNA sequencing requires accurate knowledge of hybridization thermodynamics and population ratios at matched and mismatched target sites (8,12,13).

We and others showed that a nearest-neighbor model is sufficient to accurately predict the stability and thermodynamics of DNAs with Watson–Crick pairs (14–20). Thereafter, we derived nearest-neighbor thermodynamic parameters for internal G·T and G·A mismatches and showed that, when combined with the thermodynamics of Watson–Crick pairs, accurate predictions of thermodynamics of duplexes with G·T and G·A mismatches can be determined with average standard deviations for $\Delta G°_{37}$, $\Delta H°$, $\Delta S°$ and $T_m$ of 5%, 8%, 8%, and 1.5°C respectively (17,21). To add to our mismatch parameter database and to test whether the nearest-neighbor model is applicable to unstable mismatches, such as C·T mismatches (22–24), we obtained thermodynamic measurements of 28 DNA duplexes containing internal C·T mismatches and combined them with three literature values (24,25) to derive nearest-neighbor parameters for internal C·T mismatches in DNA. The availability of internal C·T mismatch nearest-neighbor parameters along with Watson–Crick nearest-neighbors allows reliable prediction of duplex stability from sequence.

## MATERIALS AND METHODS

### Absorbance versus temperature melting curves

Oligonucleotides were synthesized on solid supports using standard phosphoramidite techniques (26) and deblocked and purified as described previously (17). Absorbance versus temperature profiles were determined using an AVIV 14DS UV-vis spectrophotometer with a heating rate of 0.8°C/min as described previously (16). Oligonucleotides were dissolved in 1.0 M NaCl, 20 mM sodium cacodylate and 0.5 mM Na₂EDTA, adjusted to pH 7.0 or 5.0 with 1.0 M HCl. Prior to the beginning of each melt, the samples were annealed and degassed by raising the temperature to 85°C for 5 min and then slowly cooling the samples to –1.5°C. While at high temperature, oligonucleotide absorbances at 260 nm were recorded and used to calculate single-strand total concentrations ($C_T$) using extinction coefficients calculated for dinucleoside monophosphates and nucleosides (27). Absorbance melting curves for each duplex were measured at 260 and 280 nm from 0 to 85 or 90°C at 8–10 different concentrations.

### Data analysis

Thermodynamic parameters for duplex formation were obtained from UV melting curves using the program MELTWIN v2.1 (28) assuming a two-state transition (i.e. duplex and random coil) by two methods: (i) averages of $\Delta H°$ and $\Delta S°$ from fits of 8–10 melting curves at different concentrations as described (29); (ii) plots of reciprocal melting temperature ($T_m^{-1}$) versus $\ln C_T$ according to the equation (30)

*To whom correspondence should be addressed. Tel: +1 313 577 0101; Fax: +1 313 577 8822; Email: jsl@chem.wayne.edu

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

For self-complementary sequences, $N = 1$ and for non-self-complementary sequences, $N = 4$. For the two-state model to apply, agreement of the parameters obtained using the two different methods is a necessary, but not sufficient, condition (17,31).

### Sequences design and rationale

Sequences were designed to have a melting temperature between 30 and 55°C and to minimize the potential of forming alternative competing secondary structures (i.e. hairpins or 'slipped' duplexes), which maximizes the likelihood of observing two-state transitions. Throughout this paper nearest-neighbors are represented in an antiparallel fashion with a slash separating the two stands and an underline indicating the position of C·T mismatches. For example, the sequence AC/TT means 5′-AC-3′ paired with 3′-TT-5′. In this study, the eight different C·T mismatch-containing dimers are evenly represented and occur with the following frequencies: AC/TT = 6, AT/TC = 9, CC/GT = 9, CT/GC = 8, GC/CT = 10, GT/CC = 9, TC/AT = 9, TT/AC = 8. In addition, all 16 possible Watson–Crick surrounding contexts are represented at least once in the data set.

### Determination of C·T mismatch contribution to duplex stability

Van't Hoff analysis of melting curves provides total $\Delta G°_{37}$, $\Delta H°$ and $\Delta S°$ for the duplex to random coil transition. Applying the nearest-neighbor model to each duplex allows determination of the internal C·T mismatch contribution to duplex stability. For example, the internal C·T mismatch contribution to the duplex GGACCGACG·CGTCTGTCC [which has a $\Delta G°_{37}$(expt) of –6.58 kcal/mol] is the sum of initiation and nearest-neighbor propagation terms

$$
\begin{array}{l}
5′\text{-GGAC}\underline{C}\text{GACG-}3′ \\
3′\text{-CCTG}\underline{T}\text{CTGC-}5′
\end{array} =
$$
$$
\text{Init.} + \tfrac{GG}{CC} + \tfrac{GA}{CT} + \tfrac{AC}{TG} + \tfrac{CC}{G\underline{T}} + \tfrac{CT}{G\underline{C}} + \tfrac{GA}{CT} + \tfrac{AC}{TG} + \tfrac{CG}{GC} \qquad \textbf{2}
$$

Note that for self-complementary sequences a symmetry penalty is also included in the calculation (32). The C·T mismatch contribution, $\Delta G°_{37}$(mismatch), is calculated by rearranging equation **2** to give

$$
\tfrac{CC}{G\underline{T}} + \tfrac{CT}{G\underline{C}} =
\begin{array}{l}
5′\text{-GGAC}\underline{C}\text{GACG-}3′ \\
3′\text{-CCTG}\underline{T}\text{CTGC-}5′
\end{array} -
$$
$$
\text{Init.} - \tfrac{GG}{CC} - \tfrac{GA}{CT} - \tfrac{AC}{TG} - \tfrac{GA}{CT} - \tfrac{AC}{TG} - \tfrac{CG}{GC} \qquad \textbf{3}
$$

Substitution of nearest-neighbor parameters for Watson–Crick and initiation terms, which have been previously determined (17), into equation **3** gives

$$
\tfrac{CC}{G\underline{T}} + \tfrac{CT}{G\underline{C}} =
$$
$$
-6.58 - (1.96) - (-1.84) - (-1.30) - (-1.44) - (-1.30) - (-1.44) - (-2.17)
$$
$$
\Delta G°_{37}(\text{mismatch}) = \Delta G°_{37}(\tfrac{C\underline{C}}{G\underline{T}} + \tfrac{C\underline{T}}{G\underline{C}}) = +0.95 \text{ kcal/mol} \qquad \textbf{4}
$$

Therefore, the C·T mismatch in the context CCG/GTC destabilizes the free energy of duplex formation by 0.95 kcal/mol. Similar

calculations are also performed for $\Delta H°$ and $\Delta S°$ to determine $\Delta H°$(mismatch) and $\Delta S°$(mismatch).

### Determination of thermodynamics of linearly independent sequences with C·T mismatches

The contribution of a mismatch to DNA duplex stability depends on the location of the mismatch, its nearest neighbors and its orientation (17). A mismatch located in the middle of the duplex is less stable than a mismatch located at the termini (17). For C·T mismatches, imposing a restriction on the location of the mismatch, such as forming all internal C·T mismatches, reduces the number of independent parameters that can be derived from the data set from eight to seven (17,33). Instead of the usual dimer sequences format for nearest-neighbor parameters, internal mismatches should be represented in terms of trimer sequences with the mismatch being in the middle position. There are 16 unique possibilities for such trimer duplexes with internal C·T mismatches and, according to the nearest-neighbor model, seven of them are linearly independent. Note that the trimer sequences reported here do not account for next-nearest-neighbor interactions. To derive the seven linearly independent trimer sequences one simply adds an arbitrary base pair to the end of all of the eight dimer nearest-neighbors (in this study we arbitrarily chose to add a C·G pair to the 3′-end of all of the dimer sequences). Upon adding the third base pair, two of the trimer sequences will be the same (GCC/CTG and GTC/CCG), therefore reducing the number of unique independent trimer sequences to seven. Duplexes with internal C·T mismatches are expressed as a linear combination of Watson–Crick dimer nearest-neighbors and mismatch trimers. Thus, equation **2** can be written as

$$
\begin{array}{l}
5′\text{-GGAC}\underline{C}\text{GACG-}3′ \\
3′\text{-CCTG}\underline{T}\text{CTGC-}5′
\end{array} =
$$
$$
\text{Init.} + \tfrac{GG}{CC} + \tfrac{GA}{CT} + \tfrac{AC}{TG} + \tfrac{CCG}{G\underline{T}C} + \tfrac{GA}{CT} + \tfrac{AC}{TG} + \tfrac{CG}{GC} \qquad \textbf{5}
$$

where the trimer sequence CCG/GTC is accounted for using

$$
\tfrac{CCG}{G\underline{T}C} = \tfrac{CCC}{G\underline{T}G} + \tfrac{CTC}{G\underline{C}G} - \tfrac{GCC}{C\underline{T}G} \qquad \textbf{6}
$$

### Linear regression analysis of C·T mismatch nearest-neighbors

Thermodynamic parameters derived from averages of the fits and $T_m^{-1}$ versus $\ln C_T$ are equally reliable (16,17,20), thus their averages were used to determine C·T mismatch contributions to the total $\Delta G°_{37}$, $\Delta H°$ and $\Delta S°$ of all 26 duplexes (see equation **4** above). The solution to these 26 equations for seven unknowns was determined by carrying out multiple linear regression by singular value decomposition (SVD) analysis (34) using the program MATHEMATICA v2.1 (Wolfram research) as described (16,17). The data in the SVD analysis were weighted by their errors (see below). A similar SVD calculation was performed to determine $\Delta H°$ values for the seven trimers. The solutions obtained were used to calculate $\Delta S°$ using the equation

$$\Delta S° = (\Delta H° - \Delta G°_{37})/310.15 \qquad \textbf{7}$$

To verify our results, we performed SVD analysis for $\Delta S°$ and obtained nearest-neighbor parameters that are in agreement with those obtained using equation **7**.

### Error analysis of the data

To determine the error associated with C·T mismatch contributions to each $\Delta G°_{37}$, $\Delta H°$ and $\Delta S°$ measurement, we used standard error propagation methods (35). The uncertainty in measured $\Delta G°_{37}$, $\Delta H°$ and $\Delta S°$ parameters were assumed to be 4%, 8% and 8% respectively. The measurement errors, in combination with reported errors for Watson–Crick nearest-neighbors (17), were propagated to obtain an error estimate for C·T mismatch trimer contributions using the equation

$$[\sigma_{\Delta G°_{37}(mismatch)}]^2 = [\sigma_{\Delta G°_{37}(measured)}]^2 + \sum_{NN} (\sigma_{\Delta G°_{37}})^2 \qquad \mathbf{8}$$

where $\sigma_{\Delta G°_{37}(mismatch)}$ is the propagated error associated with $\Delta G°_{37}(mismatch)$, $\sigma_{\Delta G°_{37}(measured)}$ is the uncertainty in the measured free energy for the duplex (4%) and $\sum_{NN}(\sigma_{\Delta G°_{37}})^2$ is the squared sum of errors for the Watson–Crick nearest-neighbors represented in the duplex (17). The error in the initiation parameter is negligible due to large correlation terms (17). Similar calculations were carried out to obtain $\sigma_{\Delta H°(mismatch)}$ and $\sigma_{\Delta S°(mismatch)}$. The SVD analysis propagates the mismatch errors to the determined nearest-neighbor parameters in the variance–covariance matrix in a rigorous fashion (34).

### Resampling analysis of the data

To independently evaluate the error in the obtained C·T mismatch nearest-neighbors and to point out sequences that are either outliers in the fit or that have a substantial effect on the solution obtained by SVD analysis, we performed a resampling analysis of the data. The solution obtained by performing SVD analysis on all 26 sequences is over-determined (i.e. 26 equations with seven unknowns). This resampling analysis has the advantage that it can determine the uncertainties of C·T mismatch nearest-neighbors separate of any previous assumption made about the errors in the measurements (17,36). The resampling analysis was performed for $\Delta G°_{37}$, $\Delta H°$ and $\Delta S°$. We performed 30 resampling trials in which eight randomly selected sequences were removed. For each resampling trial, the number of non-zero singular values was confirmed to be seven. For each nearest-neighbor, the 30 resampling trials were averaged and standard deviations determined. The averaged nearest-neighbors from resampling trials were within round-off error of the values obtained for an SVD analysis with all 26 sequences. The standard deviations from resampling agree with the errors propagated in SVD.

### ¹H-NMR spectroscopy

Oligomers were dissolved in 90% $H_2O$ and 10% $D_2O$ with 1 M NaCl, 10 mM disodium phosphate and 0.1 mM $Na_2EDTA$ at pH 7.0 or 5.0. 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 (37,38). Spectra were recorded with the carrier placed at the solvent frequency and with high power and low power pulse widths of 10.0 and 1800 μs, a sweep width of 12 kHz and a gradient field strength of 10.0 G/cm and duration of 1 ms. 512–1024 transients were collected for each spectrum. Data were multiplied by a 4.0 Hz line

broadening exponential function and Fourier transformed with a Silicon Graphics Indigo²Extreme computer with Varian VNMR software. No baseline correction or solvent subtraction was applied. 3-Trimethylsilyl propionic-2,2,3,3-d⁴ acid (TSP) was used as the internal standard for chemical shift reference. One-dimensional NOE difference spectra were acquired as described above, but with selective decoupling of individual resonances during the 1 s recycle delay. Each resonance was decoupled with a power sufficient to saturate <80% of the signal intensity, so that spillover artifacts would be minimized. The spectra were acquired in an interleaved fashion in blocks of 16 scans to minimize subtraction errors due to long term instrument drift. 3200–6400 scans were collected for each FID.

## RESULTS

### Thermodynamics of DNA duplexes with C·T mismatches

Plots of $T_m^{-1}$ versus $\ln C_T$ for all the duplexes in this study were linear (correlation coefficient >0.99; not shown). Thermodynamic parameters for helix to coil transitions for 28 sequences using averages of the fits of melting curves and $T_m^{-1}$ versus $\ln C_T$ plots are listed in Table 1. A widely used method for determining applicability of the two-state model to melting curves is comparison of the $\Delta H°$ values obtained from the averages of the fits and the $T_m^{-1}$ versus $\ln C_T$ plots. If the $\Delta H°$ parameters from both methods agree within 15%, the duplex to random coil transition is assumed to be two-state (16,17,20). However, melts that exhibit agreement of $\Delta H°$ values of 15% do not necessarily rule out non-two-state behavior (17,31). Twenty three of the sequences in Table 1 have a $\Delta H°$ agreement from the two methods of ≤15% and showed monophasic transitions, indicating bimolecular two-state behavior. Five duplexes in Table 1 melt with non-two-state transitions. The non-two-state behavior of these duplexes is manifested in the >15% disagreement in $\Delta H°$ values derived by the two methods. These non-two-state sequences may have the ability to form alternative conformations, such as hairpins or slipped duplexes, during the duplex to random coil transition. For duplexes with two-state transitions, the thermodynamics obtained from the fits and the $T_m^{-1}$ versus $\ln C_T$ plots are equally reliable (16,17,20,21) and thus their averages are the experimental values listed in Table 2.

### Nearest-neighbor thermodynamics of unique trimer sequences with internal C·T mismatches

Table 3 lists thermodynamic parameters obtained using SVD analysis for all 16 unique trimer sequences with internal C·T mismatches. According to the nearest-neighbor model, seven of these trimer sequences are linearly independent and can be used in linear combination to obtain parameters for the other nine trimer sequences. The errors listed in Table 3 are the standard deviations from resampling analysis of the data (see Materials and Methods). These errors are the same as the errors obtained by propagating the experimental and Watson–Crick nearest-neighbor errors in the SVD analysis. The parameters listed in Table 3, along with Watson–Crick nearest-neighbor and initiation parameters (17), predict the thermodynamics of all 26 duplexes with two-state thermodynamics (Table 2) with average deviations for $\Delta G°_{37}$, $\Delta H°$, $\Delta S°$ and $T_m$ of 0.45 kcal/mol, 5.9 kcal/mol, 18.0 e.u., and 1.9°C respectively.

**Table 1.** Thermodynamics of duplex formation of oligonucleotides with internal C·T mismatches[a]

| DNA duplex | $1/T_m$ versus $\ln C_T$ parameters | | | | Curve fit parameters | | |
|---|---|---|---|---|---|---|---|
| | $-\Delta G°_{37}$ (kcal/mol) | $-\Delta H°$ (kcal/mol) | $-\Delta S°$ (eu) | $T_m$ (°C)[b] | $-\Delta G°_{37}$ (kcal/mol) | $-\Delta H°$ (kcal/mol) | $-\Delta S°$ (eu) |
| | *Molecules with two-state transitions* | | | | | | |
| CGTCCGTCC | 6.73 ± 0.32 | 56.5 ± 1.3 | 160.4 ± 3.1 | 42.9 | 6.68 ± 0.11 | 60.1 ± 1.6 | 172.3 ± 5.4 |
| CGTGCCTCC | 6.75 ± 0.15 | 57.1 ± 0.6 | 162.2 ± 1.5 | 43.0 | 6.75 ± 0.04 | 59.1 ± 1.0 | 168.6 ± 3.2 |
| | (6.28 ± 0.22) | (53.6 ± 0.9) | (152.5 ± 2.1) | (40.7) | (6.31 ± 0.03) | (53.3 ± 1.4) | (151.4 ± 4.7) |
| GGACCCTCG | 6.23 ± 0.21 | 53.2 ± 0.9 | 151.4 ± 2.1 | 40.3 | 6.21 ± 0.03 | 55.0 ± 1.1 | 157.3 ± 3.4 |
| GGACCGACG | 6.60 ± 0.51 | 54.5 ± 2.0 | 154.3 ± 4.8 | 42.4 | 6.56 ± 0.10 | 59.1 ± 1.0 | 169.4 ± 3.1 |
| GGAGCCACG | 6.59 ± 0.24 | 57.0 ± 1.0 | 162.5 ± 2.5 | 42.0 | 6.58 ± 0.04 | 59.0 ± 1.6 | 169.0 ± 5.1 |
| CACAGCAGGTC | 7.74 ± 0.22 | 65.6 ± 1.0 | 186.6 ± 2.4 | 47.0 | 7.75 ± 0.04 | 62.6 ± 3.9 | 176.8 ± 12.4 |
| CATGACGCTAC | 8.44 ± 0.86 | 73.8 ± 4.0 | 210.7 ± 10.1 | 49.1 | 8.61 ± 0.18 | 85.9 ± 4.3 | 249.3 ± 13.5 |
| | (7.93 ± 0.33) | (67.5 ± 1.5) | (192.2 ± 3.7) | (47.5) | (8.10 ± 0.23) | (80.7 ± 3.6) | (234.0 ± 10.9) |
| CATGATGCTAC | 8.00 ± 0.46 | 71.9 ± 2.1 | 206.1 ± 5.4 | 47.3 | 8.01 ± 0.07 | 73.1 ± 1.9 | 209.8 ± 6.0 |
| CATGTCCACTAC | 6.98 ± 0.10 | 66.0 ± 0.5 | 190.3 ± 1.3 | 43.3 | 6.99 ± 0.03 | 66.6 ± 2.6 | 192.2 ± 8.3 |
| CATGTTACTAC | 6.91 ± 0.15 | 65.5 ± 0.7 | 188.8 ± 1.7 | 43.0 | 6.92 ± 0.02 | 65.6 ± 1.8 | 189.1 ± 6.0 |
| GAACGCTGTCC | 8.29 ± 0.41 | 62.6 ± 1.6 | 175.1 ± 4.0 | 50.5 | 8.45 ± 0.13 | 64.8 ± 4.6 | 181.7 ± 14.4 |
| GACCTCCTGTG | 7.56 ± 0.23 | 66.5 ± 1.0 | 190.0 ± 2.6 | 46.1 | 7.55 ± 0.05 | 61.7 ± 2.5 | 174.6 ± 8.2 |
| GATCATTGTAC | 7.04 ± 0.44 | 67.2 ± 2.0 | 194.0 ± 5.0 | 43.4 | 7.01 ± 0.09 | 72.9 ± 3.2 | 212.5 ± 10.1 |
| GATGTCTGTAC | 6.61 ± 0.29 | 65.5 ± 1.4 | 189.8 ± 3.5 | 41.5 | 6.58 ± 0.04 | 69.8 ± 2.4 | 203.8 ± 7.6 |
| GCTAGCAATCC | 7.27 ± 0.17 | 67.3 ± 0.8 | 193.4 ± 2.0 | 44.8 | 7.25 ± 0.10 | 59.8 ± 2.2 | 169.6 ± 7.0 |
| CGCCAGAGCCGG | 6.73 ± 0.61 | 43.7 ± 1.9 | 119.0 ± 4.2 | 44.7 | 6.68 ± 0.13 | 49.9 ± 3.0 | 139.3 ± 9.4 |
| CGCTAGAGTCGG | 6.44 ± 0.29 | 46.7 ± 1.0 | 129.8 ± 2.3 | 42.2 | 6.43 ± 0.05 | 48.6 ± 2.0 | 136.1 ± 6.5 |
| GGCCGAGACCGC | 7.56 ± 0.59 | 65.2 ± 2.6 | 186.0 ± 6.5 | 46.2 | 7.56 ± 0.06 | 63.1 ± 3.1 | 179.2 ± 9.9 |
| GGCTGAGATCGC | 7.34 ± 0.47 | 60.9 ± 2.0 | 172.6 ± 4.8 | 45.7 | 7.34 ± 0.06 | 56.3 ± 1.4 | 157.9 ± 4.6 |
| CGACCATATGTTCG | 6.29 ± 0.32 | 53.1 ± 1.3 | 151.0 ± 3.2 | 40.6 | 6.37 ± 0.07 | 51.1 ± 5.6 | 144.1 ± 10.1 |
| CGTCTCATGATACG | 7.28 ± 0.29 | 79.1 ± 1.6 | 231.5 ± 4.4 | 43.4 | 7.23 ± 0.09 | 69.5 ± 3.3 | 200.6 ± 10.9 |
| | (6.59 ± 0.37) | (73.5 ± 2.0) | (215.7 ± 5.4) | (41.0) | (6.66 ± 0.17) | (61.2 ± 4.6) | (176.0 ± 15.3) |
| CTCCACATGTTGAG | 6.80 ± 0.34 | 72.1 ± 3.5 | 210.5 ± 9.2 | 41.9 | 6.78 ± 0.15 | 61.9 ± 4.1 | 177.8 ± 13.5 |
| | (6.39 ± 0.51) | (59.8 ± 2.3) | (172.2 ± 5.8) | (40.8) | (6.48 ± 0.11) | (54.6 ± 5.8) | (155.4 ± 18.9) |
| CTCTCATATGCGAG | 6.49 ± 0.33 | 71.1 ± 1.8 | 208.3 ± 4.7 | 40.6 | 6.52 ± 0.12 | 60.4 ± 4.1 | 173.6 ± 13.7 |
| | *Molecules with non-two-state transitions* | | | | | | |
| CGAGCGTCC | 6.20 ± 0.63 | 65.8 ± 3.1 | 192.1 ± 8.1 | 39.5 | 6.38 ± 0.17 | 51.0 ± 2.0 | 143.8 ± 6.2 |
| GAACGCAGTCC | 6.59 ± 0.96 | 26.3 ± 1.8 | 63.6 ± 2.9 | 48.2 | 6.51 ± 0.30 | 38.3 ± 7.9 | 102.6 ± 24.8 |
| GATCTTTGTAC | 7.14 ± 0.26 | 67.3 ± 1.2 | 194.1 ± 3.1 | 43.9 | 7.11 ± 0.15 | 81.0 ± 3.4 | 238.3 ± 10.6 |
| CTCTATGGTACTGC | 7.50 ± 0.59 | 88.8 ± 3.6 | 262.3 ± 9.6 | 43.5 | 7.45 ± 0.15 | 68.3 ± 1.4 | 196.3 ± 4.8 |
| GCATCTGCGGCTAG | 10.28 ± 2.10 | 46.7 ± 5.6 | 117.4 ± 11.3 | 71.0 | 9.78 ± 0.44 | 39.1 ± 5.1 | 94.4 ± 15.1 |

[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 C·T mismatch. Molecules listed as two-state had $\Delta H°$ agreement within 15% by two different methods. Molecules listed as non-two-state had $\Delta H°$ disagreement of >15%. Solutions are 1 M NaCl, 20 mM sodium cacodylate, 0.5 mM Na$_2$EDTA, pH 7.0. Values reported in parentheses were obtained under the same solution conditions as above except at pH 5.0. Errors are standard deviations from the regression analysis of the melting data. Extra significance figures are given to allow accurate calculation of $\Delta G°_{37}$ and $T_m$.
[b]Calculated for $10^{-4}$ M oligomer concentration for self-complementary sequences and $4 \times 10^{-4}$ M for non-self-complementary sequences.

**Table 2.** Experimental and predicted thermodynamics of oligonucleotides with C·T mismatches[a]

| DNA duplex | Ref.[b] | −ΔG°37 (kcal/mol)[c] | | −ΔH° (kcal/mol)[c] | | −ΔS° (e.u)[c] | | $T_m$ (°C)[d] | |
|---|---|---|---|---|---|---|---|---|---|
| | | Expt. | Predicted | Expt. | Predicted | Expt. | Predicted | Expt. | Predicted |
| *Molecules with two-state transitions* | | | | | | | | | |
| CAAACAAAG | (24) | 3.27 | 3.38 | 53.2 | 46.0 | 161.0 | 137.2 | 23.6 | 22.7 |
| CAAATAAAG | (24) | 3.17 | 3.07 | 50.0 | 47.7 | 151.0 | 143.6 | 22.2 | 21.5 |
| CGTCCGTCC | | 6.70 | 6.51 | 58.3 | 53.9 | 166.4 | 152.5 | 42.6 | 42.4 |
| CGTGCCTCC | | 6.75 | 5.92 | 58.1 | 43.8 | 165.4 | 122.1 | 42.9 | 38.8 |
| GGACCCTCG | | 6.22 | 5.77 | 54.1 | 46.6 | 154.4 | 131.5 | 40.2 | 37.9 |
| GGACCGACG | | 6.58 | 6.51 | 56.8 | 53.9 | 161.8 | 152.5 | 42.0 | 42.4 |
| GGAGCCACG | | 6.58 | 5.92 | 58.0 | 43.8 | 165.7 | 122.1 | 41.9 | 38.8 |
| CACAGCAGGTC | | 7.74 | 8.15 | 64.1 | 62.1 | 181.7 | 173.8 | 47.3 | 50.1 |
| CATGACGCTAC | | 8.52 | 7.62 | 79.9 | 66.2 | 230.0 | 188.6 | 48.5 | 46.8 |
| CATGATGCTAC | | 8.01 | 7.31 | 72.5 | 67.4 | 207.9 | 193.4 | 47.3 | 45.2 |
| CATGTCACTAC | | 6.99 | 6.28 | 66.3 | 62.0 | 191.2 | 179.5 | 43.3 | 40.3 |
| CATGTTACTAC | | 6.92 | 6.28 | 65.5 | 62.0 | 189.0 | 179.5 | 43.0 | 40.3 |
| GAACGCTGTCC | | 8.37 | 8.63 | 63.7 | 66.9 | 178.4 | 187.6 | 50.7 | 51.8 |
| GACCTCCTGTG | | 7.56 | 7.57 | 64.1 | 59.0 | 182.3 | 165.7 | 46.4 | 47.5 |
| GATCATTGTAC | | 7.03 | 6.51 | 70.1 | 64.9 | 203.2 | 187.9 | 43.1 | 41.6 |
| GATCTCTGTAC | | 6.59 | 6.01 | 67.6 | 63.7 | 196.8 | 185.7 | 41.3 | 39.1 |
| GCTAGCAATCC | | 7.26 | 7.07 | 63.5 | 60.4 | 181.5 | 171.9 | 44.9 | 44.4 |
| CGCCAGAGCCGG | | 6.71 | 7.35 | 46.8 | 54.9 | 129.1 | 153.4 | 44.0 | 46.6 |
| CGCTAGAGTCGG | | 6.44 | 7.35 | 47.7 | 55.4 | 132.9 | 155.0 | 42.0 | 46.5 |
| GGCCGAGACCGC | | 7.56 | 7.77 | 64.2 | 58.6 | 182.6 | 163.8 | 46.4 | 48.6 |
| GGCTGAGATCGC | | 7.34 | 8.04 | 58.6 | 63.4 | 165.3 | 178.3 | 46.1 | 49.3 |
| CGACCATATGTTCG | | 6.33 | 6.58 | 52.1 | 64.2 | 147.5 | 185.9 | 40.9 | 41.2 |
| CGTCTCATGATACG | | 7.25 | 7.84 | 74.3 | 78.4 | 216.0 | 227.4 | 43.7 | 45.9 |
| CTCCACATGTTGAG | | 6.78 | 6.72 | 67.0 | 72.4 | 194.2 | 211.6 | 42.2 | 41.8 |
| CTCTCATATGCGAG | | 6.51 | 6.00 | 65.7 | 68.8 | 190.9 | 202.3 | 41.0 | 38.7 |
| CAACTTGATATTAATA | (25) | 9.70 | 10.13 | 98.4 | 99.4 | 286.0 | 287.4 | 50.2 | 52.0 |
| *Molecules with non-two-state transitions* | | | | | | | | | |
| CGAGCGTCC | | 6.29 | 6.35 | 58.4 | 50.2 | 168.0 | 141.2 | 40.3 | 41.6 |
| GAACGCAGTCC | | 7.12 | 6.32 | 74.2 | 62.0 | 216.2 | 179.3 | 43.2 | 40.6 |
| GATCTTTGTAC | | 6.56 | 8.44 | 32.3 | 64.0 | 83.1 | 179.0 | 45.7 | 51.2 |
| CTCTATGGTACTGC | | 7.48 | 7.76 | 78.6 | 74.2 | 229.3 | 214.0 | 44.3 | 46.3 |
| GCATCTGCGGCTAG | | 10.03 | 9.87 | 42.9 | 75.6 | 105.9 | 211.8 | 72.1 | 55.4 |

[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 C·T mismatch. Experimental values are the averages of the $T_m^{-1}$ versus $\ln C_T$ and the curve fit parameter 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.

## Non-unique C·T mismatch nearest-neighbor thermodynamics

As stated previously, analysis of internal C·T mismatches in terms of dimer sequences results in eight nearest-neighbors that are not a unique solution. The non-uniqueness of these dimer sequences results from having all C·T mismatches located internally (17,33). Table 4 lists nearest-neighbor parameters for dimer sequences with C·T mismatches obtained by fitting the data to eight parameters. The eight dimer parameters listed in Table 4 are an alternative representation of the seven trimer parameters listed in Table 3. However, in the SVD analysis of eight dimer sequences, the number of non-zero singular values is seven, indicating that the stacking matrix is rank deficient and that the parameters are non-unique. To clarify the non-uniqueness of the parameters in Table 4 one could show that a linear combination of the parameters in Table 4 can be used to derive parameters for

the seven linearly independent trimer sequences in Table 3, but not *vice versa* unless an eighth parameter is given (SVD assumes the eighth parameter is zero) (14). Nonetheless, the parameters in Tables 3 and 4 result in the same predictions and, thus, one could use either representation of the data, keeping in mind that both apply only to internal C·T mismatches.

**Table 3.** Nearest-neighbor thermodynamic parameters for 16 trimer sequences with internal C·T mismatches in 1 M NaCl[a]

| Propagation sequence | $\Delta H°$ (kcal/mol) | $\Delta S°$ (e.u) | $\Delta G°_{37}$ (kcal/mol) |
|---|---|---|---|
| *Seven linearly independent trimers* | | | |
| ACC/TTG | 5.9 ± 1.3 | 13.8 ± 2.6 | 1.62 ± 0.10 |
| CCC/GTG | 4.4 ± 1.2 | 9.0 ± 2.5 | 1.60 ± 0.13 |
| GCA/CTT | 3.3 ± 1.1 | 6.2 ± 2.2 | 1.37 ± 0.12 |
| GCC/CTG | 7.5 ± 1.5 | 19.0 ± 3.0 | 1.60 ± 0.13 |
| GCG/CTC | 0.8 ± 1.2 | –0.7 ± 2.5 | 1.02 ± 0.11 |
| GCT/CTA | 1.1 ± 1.3 | –0.8 ± 2.0 | 1.35 ± 0.12 |
| TCC/ATG | 6.4 ± 1.3 | 14.3 ± 2.9 | 1.95 ± 0.14 |
| *The nine other trimer contexts[b]* | | | |
| ACA/TTT | 1.7 ± 1.4 | 1.0 ± 2.0 | 1.39 ± 0.11 |
| ACG/TTC | –0.8 ± 1.7 | –5.9 ± 3.0 | 1.04 ± 0.14 |
| ACT/TTA | –0.5 ± 1.3 | –6.0 ± 2.8 | 1.37 ± 0.15 |
| CCA/GTT | 0.2 ± 1.1 | –3.8 ± 2.2 | 1.37 ± 0.10 |
| CCG/GTC | –2.3 ± 1.4 | –10.7 ± 2.2 | 1.02 ± 0.13 |
| CCT/GTA | –2.0 ± 1.5 | –10.8 ± 2.3 | 1.35 ± 0.14 |
| TCA/ATT | 2.2 ± 1.3 | 1.5 ± 2.1 | 1.72 ± 0.13 |
| TCG/ATC | –0.3 ± 1.3 | –5.4 ± 3.0 | 1.37 ± 0.12 |
| TCT/ATA | 0.0 ± 1.5 | –5.5 ± 2.9 | 1.70 ± 0.15 |

[a]Errors are resampling standard deviations (see text).
[b]These other contexts can be derived from linear combinations of the seven linearly independent trimer sequences (see text). These parameters are not applicable to terminal or penultimate C•T mismatches.

**Table 4.** Nearest-neighbor thermodynamics of C·T mismatches in 1 M NaCl[a]

| Dimer sequence | $\Delta H°$ (kcal/mol) | $\Delta S°$ (e.u) | $\Delta G°_{37}$ (kcal/mol) |
|---|---|---|---|
| AC/TT | 0.7 | 0.2 | 0.64 |
| AT/TC | –1.2 | –6.2 | 0.73 |
| CC/GT | –0.8 | –4.5 | 0.62 |
| CT/GC | –1.5 | –6.1 | 0.40 |
| GC/CT | 2.3 | 5.4 | 0.62 |
| GT/CC | 5.2 | 13.5 | 0.98 |
| TC/AT | 1.2 | 0.7 | 0.97 |
| TT/AC | 1.0 | 0.7 | 0.75 |

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

## Thermodynamics of C·T mismatches at pH 5.0

To test the thermodynamic effects of protonation of a C·T mismatch (i.e. C+·T versus C·T) thermodynamic measurements were made on four duplexes with C·T mismatches at pH 5.0 and pH 7.0. The p$K_a$ of protonation for cytosine in the context of a C·T mismatch has been reported to be ~5.7 (39), thus, at pH 5.0, ~66% of C·T mismatches should be protonated. Four sequences were selected to represent different C·T mismatch nearest-neighbor contexts. On average, for the four C·T mismatch-containing sequences tested for pH effects, changing the pH from 7.0 to 5.0 decreased the stability of the duplex by 0.3 kcal/mol for $\Delta G°_{37}$ and 1.1°C for the $T_m$. The data obtained for these four sequences suggest that the thermodynamics of C·T mismatches at pH 5.0 are slightly less stable than at pH 7.0.

## NMR and pairing geometry of C·T and C+·T mismatches

C·T mismatches have been proposed to form at least four different structures depending on sequence context and solution conditions (Fig. 1; 39–42). To determine the pairing geometry for C·T mismatches in this study, one-dimensional exchangeable proton NMR spectra of five DNA duplexes with different C·T mismatch contexts were acquired at pH 7.0 and 5.0. Figures 2 and 3 show a representative imino region (9–15 ppm) of two of the duplexes studied containing C·T mismatches at pH 7.0 and 5.0. Resonances between 12–13 and 13–15 ppm are usually the imino protons of canonical Watson–Crick G·C and A·T pairs. At pH 7.0, an imino peak is observed around 11.5 ppm (Figs 2a and 3a) which broadens out at pH 5.0 (Figs 2b and 3b). Irradiation of this resonance did not show any observable NOEs (not shown), probably due to rapid chemical exchange with water. Previous structural studies on C·T and C·U mismatches in DNA and RNA showed that at neutral pH these mismatches can pair with two hydrogen bonds, one of which, due to the repulsion of the carbonyl groups of the cytosine and thymine (43), is possibly mediated via a water molecule (Fig. 1b; 39–42). Our data are most consistent with NMR observations on C·T mismatches at neutral pH and, thus, we tentatively assign the resonance at 11.5 ppm as the imino proton of thymine hydrogen bonded to N3 of cytosine via a water molecule (39,40,42). At pH 5.0, the protonation of N3 of cytosine results in a change in the pairing geometry of the C·T mismatch which broadens the imino resonance of the thymine in the C·T mismatch (11.5 ppm). This broadening of the imino resonance might be a result of chemical exchange between protonated and non-protonated C·T mispairs. Previous structural studies of C·T mismatches under acidic conditions suggest that the imino proton of thymine becomes hydrogen bonded to the carbonyl group of cytosine, possibly via a water molecule, making it exchange faster with water (Fig. 1c and d; 39,40). In contrast, C·C and A·C in RNA (44) and in DNA (H.T.Allawi and J.SantaLucia Jr, unpublished results) are often stabilized at acidic pH.

## DISCUSSION

### Applicability of the nearest-neighbor model to internal C·T mismatches

Table 2 compares experimental results of 26 duplexes with C·T mismatches with predictions made by the parameters listed in Table 3 (or Table 4) and Watson–Crick nearest-neighbor parameters (17). For single mismatches in DNA, we have previously shown that a nearest-neighbor model can accurately predict duplexes
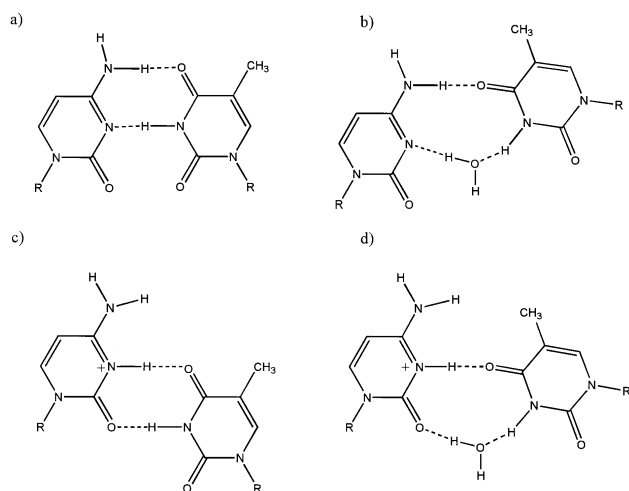
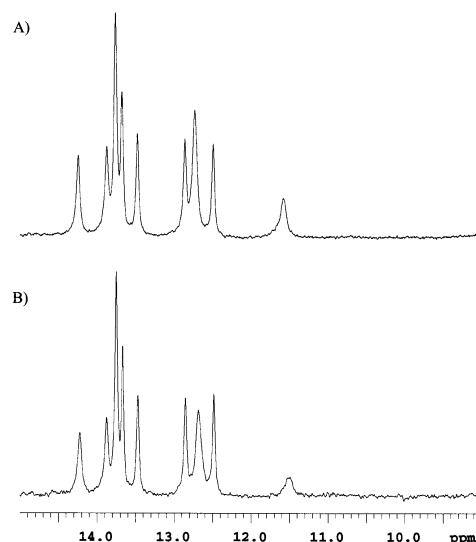**Figure 1.** Four hydrogen bonded structures of the C·T mispair at neutral pH (**a** and **b**) and at acidic pH (**c** and **d**).



**Figure 2.** 500 MHz $^1$H-NMR spectra of the exchangeable imino region (9–15 ppm) in 1 M NaCl, 10 mM disodium phosphate and 0.1 mM Na$_2$EDTA at 10°C in 90% H$_2$O/10% D$_2$O of CATGTTACTAC•GTACTCACATG at (**a**) pH 7.0 and (**b**) pH 5.0.



**Figure 3.** 500 MHz $^1$H-NMR spectra of the exchangeable imino region (9–15 ppm) in 1 M NaCl, 10 mM disodium phosphate and 0.1 mM Na$_2$EDTA at 10°C in 90% H$_2$O/10% D$_2$O of (CGTCTCATGATACG)$_2$ at (**a**) pH 7.0 and (**b**) pH 5.0.

with internal G·A and G·T with average deviations for $\Delta G^\circ_{37}$, $\Delta H^\circ$, $\Delta S^\circ$ and $T_m$ of 5.0%, 8.0%, 8.0%, and 1.5°C respectively (17,21). In this study, we find that analysis of C·T mismatch contributions to duplex stability in terms of a nearest-neighbor model results in parameters that predict the thermodynamics of sequences with two-state transitions with an average deviation for $\Delta G^\circ_{37}$, $\Delta H^\circ$, $\Delta S^\circ$ and $T_m$ of 6.4%, 9.9%, 10.6% and 1.9°C respectively. These average deviations are slightly higher than what was observed for G·A and G·T mismatches (17,21). Nonetheless, considering how unstable C·T mismatches are, one might expect that C·T mismatches are capable of disrupting double-helical DNA in a fashion that may extend to next-nearest-neighboring Watson–Crick pairs. However, results from this study suggest that if there are any next-nearest-neighbor effects for C·T mismatches they are very small and can be neglected. Hence, the nearest-neighbor parameters in Tables 3 and 4 make predictions that are adequate for most applications. An alternative way to test the applicability of the nearest-neighbor model is to synthesize oligonucleotides with different sequences but the same nearest-neighbor composition (17,45–47). In this study, three pairs of duplexes have the same nearest-neighbor composition (Tables 1 and 2). For example, the duplexes CGTGCCT-CC·GGAGTCACG and GGAGCCACG·CGTGTCTCC have different sequences but the same nearest-neighbors and their $\Delta G^\circ_{37}$, $\Delta H^\circ$, $\Delta S^\circ$ and $T_m$ agree within 0.17 kcal/mol, 0.1 kcal/mol, 0.3 e.u., and 1.0°C respectively. The average deviation from the mean between the three pairs of duplexes with the same nearest-neighbors for $\Delta G^\circ_{37}$, $\Delta H^\circ$, $\Delta S^\circ$ and $T_m$ are 0.06 kcal/mol, 0.4 kcal/mol, 1.2 e.u., and 0.3°C respectively.

## Trends in C·T mismatch thermodynamics

Trimer mismatch free energy ($\Delta G^\circ_{37}$) contributions for internal C·T mismatches vary weakly, depending on the mismatch orientation and context (Tables 3 and 4). The most stable trimer sequences (GCG/CTC and CCG/GTC) destabilize the duplex by +1.02 kcal/mol and the least stable trimer (TCC/ATG) destabilizes the duplex by +1.95 kcal/mol. This range of 0.93 kcal/mol for $\Delta G^\circ_{37}$ indicates that there is a weak stacking contribution to
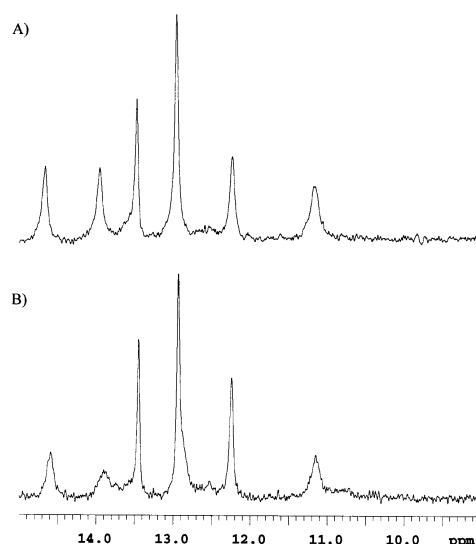
stability of a C·T mismatch. For trimer sequences with the cytosine of the C·T mismatch on the top strand, the general trend for the 5′-end closing Watson–Crick pair (with decreasing order of stability) is G·C ≈ C·G > A·T >> T·A. However, when the thymine of the C·T mismatch is on the top strand (i.e. T·C), the trend on the 5′-end becomes (with decreasing order of stability) C·G > A·T ≈ T·A > G·C. Close inspection of these trends reveals an interesting result. Generally, a G·C base pair (which has three hydrogen bonds) is expected to have a stabilizing effect on duplexes that is larger than an A·T pair (which has two hydrogen bonds). However, G·C pairs stacked on the 5′-end of a T·C mismatch destabilize the duplex by 0.98 kcal/mol and A·T pairs

stacked on the 5′-end of a T·C mismatch destabilize the duplex by 0.73 kcal/mol. Therefore, in this case, a 5′ A·T pair stabilizes T·C mismatches more than does a 5′ G·C. Thus, stacking interactions, more than hydrogen bonding, play a major role in the stability of duplexes with internal C·T mismatches. This is also evident when a G·C pair stacked on a T·C mismatch (GT̲/C̲C) is compared with a C·G (CT̲/G̲C), which are destabilizing by 0.98 and 0.40 kcal/mol respectively (Table 4).

## Comparison of thermodynamics of C·T mismatches and Watson–Crick pairs

No correlation is observed when comparing thermodynamics of trimer sequences with internal C·T mismatches with the corresponding trimer sequences with either G·C or A·T Watson–Crick base pairs (17). Free energies of Watson–Crick trimer sequences with a central A·T or G·C pair vary over a range of 2.95 kcal/mol, whereas the range of trimer sequences with internal C·T mismatches vary over 0.93 kcal/mol in $\Delta G^{\circ}_{37}$. The most stable Watson–Crick trimer sequence is GCG/CGC ($\Delta G^{\circ}_{37}$ = –4.41 kcal/mol) and the least stable is ATA/TAT ($\Delta G^{\circ}_{37}$ = –1.46 kcal/mol) (17). For internal C·T mismatches, the most stable C·T trimer sequence contexts are GC̲G/CT̲C and CC̲G/GT̲C (+1.02 kcal/mol) and is the same context as the most stable Watson–Crick sequence (GCG/CGC). However, the trimer sequence TC̲C/AT̲G, which is the least stable C·T context, is different than the least stable Watson–Crick sequence (ATA/TAT).

## Comparison of C·T, G·T and G·A mismatch thermodynamics

Comparison of internal C·T mismatches thermodynamics (Table 3) with previously published parameters for internal G·A (21) and G·T (17) mismatch thermodynamics indicates that C·T mismatches are among the most unstable mismatches in DNA consistent with previous observations (23,24). The most stable C·T trimer sequences are GC̲G/CT̲C and CC̲G/GT̲C ($\Delta G^{\circ}_{37}$ of +1.02 kcal/mol) and the most stable G·A or G·T trimer sequences are GG̲C/CA̲G and CG̲C/GT̲G ($\Delta G^{\circ}_{37}$ –0.78 and –1.05 kcal/mol respectively). Moreover, the least stable C·T trimer sequence is TC̲C/AT̲G ($\Delta G^{\circ}_{37}$ +1.95 kcal/mol) and the least stable G·A or G·T trimer sequences are TG̲A/AA̲T and AG̲A/TT̲T ($\Delta G^{\circ}_{37}$ +1.16 and +1.05 kcal/mol respectively). The average free energy contribution of all 16 unique trimer sequences with internal C·T mismatches is +1.43 kcal/mol. Average internal G·A and G·T mismatch free energy contributions for all 16 unique trimer sequences, on the other hand, are +0.17 and +0.05 kcal/mol respectively. Furthermore, stabilities of G·A and G·T mismatches are spread over a range of 1.94 and 2.10 kcal/mol respectively, while C·T mismatch stabilities are spread over a range of 0.93 kcal/mol indicating that, while contributions of internal C·T mismatch thermodynamics depend slightly on the neighboring bases, their thermodynamics are not as sensitive to the surrounding base pair context as in G·A and G·T mismatches.

## 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 Brown,T. (1995) *Aldrichimica Acta.*, **28**, 15–20.
4 Ikoku,A.S. and Hearst,J.E. (1981) *J. Mol. Biol.*, **151**, 245–259.
5 Morozov,S.Y., Chernov,B.K., Merits,A. and Blinov,V.M. (1994) *J. Biomol. Struct. Dyn.*, **11**, 837–847.
6 Ozawa,K., Kurtzman,G. and Young,N. (1986) *Science*, **233**, 883–886.
7 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.
8 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.
9 Zon,G. (1989) In Cohen,J.S. (ed.), *Oligonucleotides.* CRC Press, Boca Raton, FL, pp. 233–249.
10 Freier,S.M. (1993) In Crooke,S.T. and Lebleu,B. (eds), *Antisense Research and Applications.* CRC Press, Boca Raton, FL, pp. 67–82.
11 Symons,R.H. (1989) In Symons,R.H. (ed.), *Nucleic Acids Probes.* CRC Press, Boca Raton, FL.
12 Chee,M., Yang,R., Hubbell,E., Berno,A., Huang,X.C., Stern,D., Winkler,J., Lockhart,D.J., Morris,M.S. and Fodor,S.P. (1996) *Science*, **274**, 610–614.
13 Pease,A.C., Solas,D., Sullivan,E.J., Cronin,M.T., Holmes,C.P. and Fodor,S.P.A. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 5022–5026.
14 SantaLucia,J.,Jr (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 1460–1465.
15 Vologodskii,A.V., Amirikyan,B.R., Lyubchenko,Y.L. and Frank-Kamenetskii,M.D. (1984) *J. Biomol. Struct. Dyn.*, **2**, 131–148.
16 SantaLucia,J.,Jr, Allawi,H.T. and Seneviratne,P.A. (1996) *Biochemistry*, **35**, 3555–3562.
17 Allawi,H.T. and SantaLucia,J.,Jr (1997) *Biochemistry*, **36**, 10581–10594.
18 Gotoh,O. and Tagashira,Y. (1981) *Biopolymers*, **20**, 1033–1042.
19 Doktycz,M.J., Goldstein,R.F., Paner,T.M., Gallo,F.J. and Benight,A.S. (1992) *Biopolymers*, **32**, 849–864.
20 Sugimoto,N., Nakano,S., Yoneyama,M. and Honda,K. (1996) *Nucleic Acids Res.*, **24**, 4501–4505.
21 Allawi,H.T. and SantaLucia,J.,Jr (1998) *Biochemistry*, **37**, 2170–2179.
22 Ke,S.H. and Wartell,R.M. (1996) *Nucleic Acids Res.*, **24**, 707–712.
23 Gaffney,B.L. and Jones,R.A. (1989) *Biochemistry*, **28**, 5881–5889.
24 Aboul-ela,F., Koh,D., Tinoco,I.,Jr and Martin,F.H. (1985) *Nucleic Acids Res.*, **13**, 4811–4824.
25 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.
26 Brown,T. and Brown,D.J.S. (1991) In Eckstein,F. (ed.), *Oligonucleotides and Analogues.* IRL Press, New York, NY, pp. 1–24.
27 Richards,E.G. (1975) In Fasman,G.D. (ed.), *Handbook of Biochemistry and Molecular Biology: Nucleic Acids,* 3rd Edn. CRC Press, Cleveland, OH, Vol. I.
28 McDowell,J.A. and Turner,D.H. (1996) *Biochemistry*, **35**, 14077–14089.
29 Petersheim,M. and Turner,D.H. (1983) *Biochemistry*, **22**, 256–263.
30 Borer,P.N., Dengler,B., Tinoco,I.,Jr and Uhlenbeck,O.C. (1974) *J. Mol. Biol.*, **86**, 843–853.
31 Marky,L.A. and Breslauer,K.J. (1987) *Biopolymers*, **26**, 1601–1620.
32 Cantor,C.R. and Schimmel,P.R. (1980) *Biophysical Chemistry Part III: The Behaviour of Biological Macromolecules.* W.H.Freeman, San Francisco, CA, pp. 1183–1264.
33 Gray,D.M. (1997) *Biopolymers*, **42**, 783–793.
34 Press,W.H., Flannery,B.P., Teukolsky,S.A. and Vetterling,W.T. (1989) *Numerical Recipes.* Cambridge University Press, New York, NY.
35 Bevington,P.R. (1969) *Data Reduction and Error Analysis for the Physical Sciences.* McGraw-Hill, New York, NY, pp. 164–203.
36 Efron,B. and Tibshirani,R. (1993) *An Introduction to the Bootstrap.* Chapman & Hall, London, UK.
37 Piotto,M., Saudek,V. and Sklenar,V. (1992) *J. Biomol. NMR*, **2**, 661–665.
38 Lippens,G., Dhalluin,C. and Wieruszeski,J.-M. (1995) *J. Biomol. NMR*, **5**, 327–331.
39 Boulard,Y., Cognet,J.A.H. and Fazakerley,G.V. (1997) *J. Mol. Biol.*, **268**, 331–347.
40 Bhaumik,S.R., Chary,K.V.R., Govil,G., Liu,K. and Miles,H.T. (1997) *Biopolymers*, **41**, 773–784.
41 Holbrook,S.R., Cheong,C., Tinoco,I. and Kim,S.-H. (1991) *Nature*, **353**, 579–581.
42 Patel,D.J., Kozlowski,S.A., Ikuta,S. and Itakura,K. (1984) *Fedn Proc. Fedn Am. Soc. Exp. Biol.*, **43**, 2663–2670.
43 Sponer,J., Leszczynski,J. and Hobza,P. (1996) *J. Phys. Chem.*, **100**, 1965–1974.
44 SantaLucia,J.,Jr, Kierzek,R. and Turner,D.H. (1991) *Biochemistry*, **30**, 8242–8251.
45 Sugimoto,N., Nakano,S., Katoh,M., Matsumura,A., Nakamuta,H., Ohmichi,T., Yonegama,M. and Sasaki,M. (1995) *Biochemistry*, **34**, 11211–11216.
46 Sugimoto,N., Honda,K. and Sasaki,M. (1994) *Nucleosides Nucleotides*, **13**, 1311–1317.
47 Kierzek,R., Caruthers,M.H., Longfellow,C.E., Swinton,D., Turner,D.H. and Freier,S.M. (1986) *Biochemistry*, **25**, 7840–7846.