Volume 13 Number 13 1985 Nucleic Acids Research

Base-base mismatches. Thermodynamics of double helix formation for $dCA_3XA_3G + dCT_3YT_3G$ (X, Y = A,C,G,T)

Fareed Aboul-ela, David Koh and Ignacio Tinoco, Jr.
Department of Chemistry and Laboratory of Chemical Biodynamics, University of California, Berkeley, CA 94720, and

Francis H.Martin
Amgen, Inc., 1900 Oak Terrace Lane, Thousand Oaks, CA 91320, USA

Received 18 March 1985; Accepted 4 June 1985

ABSTRACT

Thermodynamic parameters for double strand formation have been measured for the sixteen double helices of the sequence $dCA_3XA_3G \cdot dCT_3YT_3G$, with each of the bases A, C, G and T at the positions labelled X and Y. The results are analyzed in terms of nearest-neighbors and are compared with thermodynamic parameters for RNA secondary structure. At room temperature the sequence $\frac{-A-A-}{-T-T-}$ is more stable than $\frac{-A-A-}{-U-U-}$ and is similar in stability to $\frac{-A-C-}{-T-G-}$, $\frac{-C-A-}{-G-T-}$, $\frac{-A-G-}{-T-C-}$, $\frac{-G-A-}{-C-T-}$; $\frac{-A-T-}{-T-A-}$ and $\frac{-T-A-}{-A-T-}$ are least stable. At higher temperatures the sequences containing a G•C base pair become more stable than those containing only A•T. All molecules containing mismatches are destabilized with respect to those with only Watson-Crick pairing, but there is a wide range of destabilization. At room temperature the most stable mismatches are those containing guanine (G•T, G•G, G•A); the least stable contain cytosine (C•A, C•C). At higher temperatures pyrimidine-pyrimidine mismatches become the least stable.

INTRODUCTION

The thermodynamic stability of mismatched bases affects the probability of incorporating the wrong base during replication, and of repairing the mistake during proofreading by the polymerase (1). Thus, thermodynamics is important in the study of mechanisms of mutations. Moreover, thermodynamic studies can help in understanding the sequence dependence and the polymorphism in secondary structure of DNA, which has been shown so clearly by Dickerson (2). Distinct secondary structures may be involved in the recognition of base sequences by proteins which bind to DNA.

For some years systematic studies of double strand formation have been done using oligonucleotides of specific base sequences, with the goal of being able to predict thermodynamic parameters for DNA and for RNA secondary structure (3-8). The usual method has been to measure melting temperatures (T_m) at several concentrations by monitoring absorbance as a

function of temperature, and to obtain thermodynamic values from a van't Hoff analysis. The oligonucleotide studies have led to values for free energies, enthalpies, and entropies for double strand formation in RNA and DNA based on nearest neighbor interactions. The effects of extra bases (9-12) and of mismatched bases (12-21) have also been studied in oligonucleotides and polynucleotides. However, no systematic thermodynamic studies have been reported for oligonucleotides which contain mismatches.

Here we present measurements of thermodynamic parameters for the 16 double helices of $dCA_3XA_3G \cdot dCT_3YT_3G$, with each of the bases A, C, G, T at the positions labeled X and Y. We have thus measured the effect of specific base substitutions, including all possible base-base mismatches, in an otherwise unaltered sequence. The helix with a potential bulged nucleotide ($dCA_6G \cdot dCT_3CT_3G$) has also been studied.

EXPERIMENTAL PROCEDURES

Melting curves were obtained by a method similar to that described earlier (6). The buffer in all cases contained 1 M NaCl, 10 mM phosphate, and 0.1 mM EDTA in H_2O at pH 7.

Thermodynamic values were obtained using the van't Hoff method (3). Absorbance (A) *vs.* temperature (T) curves for several concentrations of one duplex are shown in Figure 1. For a two state model, the equilibrium constant can be written as

$$K = \frac{2f}{(1 - f)^2 C_T} \qquad (1)$$

Here f is the fraction of strands in the double-stranded state and C_T is the total concentration of all single strands. At any temperature f can be obtained from

$$1 - f = \frac{A(T) - A_d(T)}{A_s(T) - A_d(T)}$$

where $A_s(T)$ and $A_d(T)$ are the absorbances of the (single strand) upper baseline and (double strand) lower baseline at temperature T. The method for obtaining baselines is explained below. K can also be written

$$K = \exp\left(\frac{-\Delta G^\circ}{RT}\right) = \exp\left(\frac{-\Delta H^\circ}{RT} + \frac{\Delta S^\circ}{R}\right) \qquad (2)$$

At the melting temperature, T_m, f = 1/2 and we can combine Eqs. (1) and (2) to write

$$R\ln(C_T/4) = (\Delta H^\circ/T_m) - \Delta S^\circ$$

Thus if ΔH°, the difference in standard enthalpy and ΔS°, the difference in

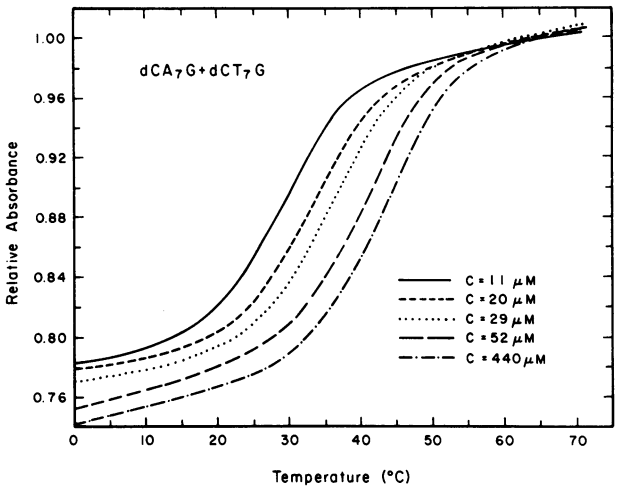


Fig. 1. Melting curves showing absorbance vs. temperature for five concentrations of $dCA_7G \cdot dCT_7G$ in 1 M NaCl, pH 7, 10 mM phosphate, 0.1 mM EDTA. The total concentrations of single strands range from 11 μM to 440 μM. All the curves are normalized to absorbance of 1 at 65°C.

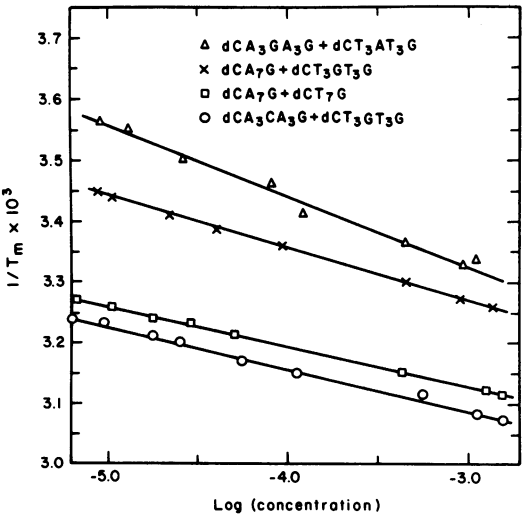


Fig. 2. van't Hoff plots of $1/T_m$ vs. log C_T where T_m is the melting temperature and C_T is the total strand concentration. The lines shown are the best least squares fit to the data.

standard entropy between the double strand and the single strand are assumed to be independent of temperature, they can be obtained from a ln C_T versus $1/T_m$ plot. Such a plot is shown in Figure 2.

The baselines were obtained from a linear least-squares fit to ten points chosen near 0°C for the lower baseline and near 65°C for the upper baseline. The same upper baseline was used for the melting curves at different concentrations for the same molecule; this minimizes the effect of choice of baseline on the thermodynamic parameters. Because hypochromicity was found to depend on concentration, the lower baseline for each experiment was chosen based on the absorbance recorded at 0°C. Data points from the melting curves at the lowest concentrations were used to obtain all upper baselines, and data from curves taken at the highest concentrations provided the slopes for lower baselines. For helices that melt at temperatures too low to provide data for lower baselines, calculations were done using an assumed flat lower baseline. For all helices, the standard free energy was calculated from the relation between ΔG° and T_m.

$$\Delta G^\circ\ (T_m) = RT_m \ln(C_T/4)$$

At 25°C ΔG° is obtained by extrapolation from the least squares fit for $\ln(C_T/4)$ *vs.* $1/T_m$ to the concentration, C_T, for which the melting temperature is 25°C.

RESULTS AND DISCUSSIONS

Measured thermodynamic parameters for double helix formation are given in Table I for all molecules of the sequence $dCA_3XA_3G \cdot dCT_3YT_3G$ with A, C, G, T substituted for X, Y. Also included is the helix with one less A•T base pair ($dCA_6G \cdot dCT_6G$) and two helices with an extra nucleotide on one strand ($dCA_3CA_3G \cdot dCT_6G$ and $dCA_6G \cdot dCT_3CT_3G$). The values of ΔH°, ΔS° and ΔG° (25°C) are referred to standard condition (1 molar concentration of each single strand reacting to form 1 molar concentration of duplex) in 1 M NaCl, pH 7, 10 mM phosphate, 0.1 mM EDTA. The melting temperatures, Tm, are given in this buffer at a total single strand concentration of 400μM. The duplexes are arranged in order of thermodynamic stability as measured by their free energy of formation from the single strands at 25°C.

The four Watson-Crick paired duplexes are, as expected, considerably more stable than those containing a non-Watson-Crick base opposition. We note that the duplex containing 7 A•T pairs with the A's on the same strand

Table I. Van't Hoff Thermodynamic Values for Double Helix Formation of $dCA_3XA_3G + dCT_3YT_3G$ in 1 M NaCl, pH 7.

X•Y	ΔG°(kcal mol^{-1})[a] 25°C	50°C	ΔH° (kcal mol^{-1})[b]	ΔS° (cal $deg^{-1} mol^{-1}$)[c]	T_m(°C) C_T=400μM
C•G	-10.1	-5.5	-64.5	-183	48°
A•T	- 9.6	-4.7	-68.0	-196	45°
G•C	- 9.5	-5.0	-62.8	-179	50°
T•A	- 8.5	-4.3	-58.6	-168	41°
-•-[e]	- 7.8	-3.3	-59	-172	37°
T•G	-6.5	-2.4	-55.6	-165	31°
G•G	-6.3	-2.4	-53.5	-158	30°
G•A	-6.2	-2.3	-52.6	-156	30°
G•T	-5.8	-2.4	-46.7	-137	27°
A•G	-5.3	-2.4	-39.9	-116	24°
C•T	-5.3	-1.3	-53.2	-161	24°
T•C	-5.0	-1.2	-50.0	-151	22°
A•A	-5.0	-2.3	-36.9	-107	21°
T•T	(-5.0)[h]	-0.8	(-54.6)	(-167)	(22°)
C•-[f]	-4.9	-0.9	-53.0	-161	22°
C•A	(-4.6)	-1.6	(-40.3)	(-120)	(19°)
C•C	(-4.5)	-0.2	(-55.3)	(-171)	(20°)
A•C	(-4.4)	-1.8	(-35.8)	(-106)	(19°)
-•C[g]	(-4.2)	-0.8	(-45.0)	(-135)	(16°)

[a]Estimated precision in ΔG° is ±0.1 kcal mol^{-1}

[b]Estimated precision in ΔH° is ±3 kcal mol^{-1}

[c]Estimated precision in ΔS° is ±9 cal deg^{-1} mol^{-1}

[d]Estimated precision in T_m is ±1°

[e]$dCA_6G•dCT_6G$. Data from reference 9.

[f]$dCA_3CA_3G•dCT_6G$. Data from reference 9.

[g]$dCA_6G•dCT_3CT_3G$.

[h]Data in parentheses are significantly less accurate. An estimated flat lower base line was used to obtain the (1-f) vs. T curve.

is of comparable stability to the duplexes which replace an A•T pair with a G•C pair. The other A•T duplex is less stable by +1.0 to +1.6 kcal mol^{-1} for ΔG° (25°C); this corresponds to an order of magnitude decrease in the equilibrium constant for double strand formation. The order of duplex

Table II. Nearest-neighbor Contributions to Double Strand Formation in 1M NaCl, pH 7.[a]

Nearest Neighbor	ΔG°,25°C (kcal mol^{-1})	ΔH° (kcal mol^{-1})	ΔS° (cal $deg^{-1}mol^{-1}$)
$\begin{matrix}-A-A-\\ -T-T-\end{matrix}$	-1.5±0.2	-10.2	-29
$\frac{1}{2}\left(\begin{matrix}-A-T-\\ -T-A-\end{matrix} + \begin{matrix}-T-A-\\ -A-T-\end{matrix}\right)$	-1.0±0.2	-5.6	-15
$\frac{1}{2}\left(\begin{matrix}-A-G-\\ -T-C-\end{matrix} + \begin{matrix}-G-A-\\ -C-T-\end{matrix}\right)$	-1.5±0.2	-7.6	-21
$\frac{1}{2}\left(\begin{matrix}-A-C-\\ -T-G-\end{matrix} + \begin{matrix}-C-A-\\ -G-T-\end{matrix}\right)$	-1.8±0.2	-8.5	-22

[a]The values given are for the reaction $\begin{matrix}-W_1 & W_2-\\ -C_1 & C_2\end{matrix} \rightleftharpoons \begin{matrix}W_1- & W_2-\\ -C_1- & C_2-\end{matrix}$

stability is $dCA_3CA_3G \cdot dCT_3GT_3G$ > $dCA_7G \cdot dCT_7G \cong dCA_3GA_3G \cdot dCA_3CA_3G$ > $dCA_3TA_3G \cdot dCT_3AT_3G$.

Estimation of helical stability of nucleic acids has often been based on analysis of oligonucleotide duplexes in terms of nearest-neighbor contributions (6). The key assumption is that thermodynamic properties of an oligonucleotide are the sum of the properties of neighboring base pairs taken two at a time. A nearest-neighbor analysis of our oligonucleotide is

$$\begin{matrix}CA_3XA_3G\\ GT_3YT_3C\end{matrix} = \left[\begin{matrix}C-A-\\ G-T-\end{matrix} + 4\left(\begin{matrix}-A-A-\\ -T-T-\end{matrix} + \begin{matrix}-A-G\\ -T-C\end{matrix}\right] + \left(\begin{matrix}-A-X-\\ -T-Y-\end{matrix} + \begin{matrix}-X-A-\\ -Y-T-\end{matrix}\right) \qquad (3)$$

The terms in square brackets equal the nearest-neighbor contribution of $CA_5G \cdot CT_5G$. Instead of using the experimental data for this duplex, we use a best least-squares fit to the experimental data (7,9) for $CA_nG \cdot CT_nG$ (n=5,6,7). This gives a best nearest-neighbor equivalent for $CA_5G \cdot CT_5G$. By subtracting the thermodynamic parameters for this $CA_5G \cdot CT_5G$ equivalent from measured values in Table I, we obtain the nearest-neighbor thermodynamic contributions for the terms in parentheses. For example

$$\Delta G^\circ\left(\begin{matrix}-A-X-\\ -T-Y-\end{matrix} + \begin{matrix}-X-A-\\ -Y-T-\end{matrix}\right) = \Delta G^\circ(CA_3XA_3G \cdot CT_3YT_3G) - \Delta G^\circ(CA_5G \cdot CT_5G) \qquad (4)$$

By studying only these very similar duplexes we obtain data which do not require assumptions about free energies of initiation or about end effects. Of course the sequence effects are very limited.

For the non-Watson Crick base-base oppositions we can consider the

contributions in parentheses in Eq. (4) to represent an internal loop of two unpaired bases.

$$\Delta G^\circ\left(\begin{smallmatrix}-A-X-A-\\ -T-Y-T-\end{smallmatrix}\right) = \Delta G^\circ(CA_3XA_3G\cdot CT_3YT_3G) - \Delta G^\circ(CA_5G\cdot CT_5G) \quad (5)$$

The distinction is somewhat arbitrary, but in Table II we give data from Eq. (4) as the average of $\begin{smallmatrix}-A-X-\\ -T-Y-\end{smallmatrix}$ and $\begin{smallmatrix}-X-A-\\ -Y-T-\end{smallmatrix}$ where X•Y is a Watson Crick pair. In Table III are data for $\begin{smallmatrix}-A-X-A-\\ -T-Y-T-\end{smallmatrix}$ where X Y are not Watson-Crick pairs. We recognize that some of these pairs, such as G•T and G•A, are definitely hydrogen bonded.

Table II gives ΔG° (25°C), ΔH° and ΔS° for the nearest-neighbor contribution to formation of Watson-Crick base pairs in a double strand. The $\begin{smallmatrix}-A-A-\\ -T-T-\end{smallmatrix}$ contribution to ΔG° at 25°C is much greater than $\frac{1}{2}\left(\begin{smallmatrix}-A-T-\\ -T-A-\end{smallmatrix} + \begin{smallmatrix}-T-A-\\ -A-T-\end{smallmatrix}\right)$, and is more nearly equal to $\frac{1}{2}\left(\begin{smallmatrix}-A-C-\\ -T-G-\end{smallmatrix} + \begin{smallmatrix}-C-A-\\ -G-T-\end{smallmatrix}\right)$ and $\frac{1}{2}\left(\begin{smallmatrix}-A-G-\\ -T-C-\end{smallmatrix} + \begin{smallmatrix}-G-A-\\ -C-T-\end{smallmatrix}\right)$. This means that replacing an A•T base pair by a C•G base pair does not always increase the stability of the double helix in DNA. The ΔG° data in Table II are consistent with those obtained from a much more extensive set of oligonucleotides measured by Markey and Breslauer (Ref. 8 and personal communication). The ΔG° values are more directly related to the measurements ($\Delta G^\circ = -RT \ln K$) and are thus more accurate than the values of ΔH° and ΔS°, particularly for the less stable duplexes.

Although ΔH° and ΔS° values are less reliable than ΔG° the magnitudes of the effects seen in Table I and II for Watson-Crick base pairs merit comment. The most negative enthalpy (favorable) and negative entropy (unfavorable) occur for formation of $dCA_7G\cdot dCT_7G$. The helix with the same base composition, but with the center A•T reversed ($dCA_3TA_3G\cdot dCT_3AT_3G$), has a much less favorable enthalpy and a more favorable entropy; the net effect is a less favorable free energy. The special stability of the $\begin{smallmatrix}-A-A-\\ -T-T-\end{smallmatrix}$ sequence may be linked to the unique properties of poly dA•poly dT. Experiments using the band shift method (23-25) to measure helical repeat lengths of DNA sequences in supercoiled plasmids have shown that the sequence $dA_n\cdot dT_n$ forms a helix with a pitch of 10±0.1 base pairs, compared to a pitch of 10.6±0.1 for all other sequences measured. Moreover, Wu and Crothers (26) have suggested that the sequences $dCA_5G\cdot dCT_5G$ and $dCA_6G\cdot dCT_6G$ in a restriction fragment have a non-standard secondary structure. They find that the double helix is bent at the sequence...$CA_5TGTCCA_6TAGGCA_6TGCCA_5T$... They postulate a bend of approximately 12° between the B-form and the $dA_n\cdot dT_n$ conformation. This

conformation may be related to the poly dA•poly dT fiber structure (27) which apparently has a C2' endo sugar conformation on one strand and a C3' endo sugar on the other strand. It may be that a nearest-neighbor analysis for $dA_n \cdot dT_n$ sequences in nucleic acids is not valid because the conformation changes as n increases beyond n=2. Thus the nearest-neighbor parameters obtained here may be more relevant to $dA_n \cdot dT_n$ (n>5) than to sequences with only two or three neighboring A•T pairs.

Comparison of ΔG° values for DNA with those for RNA (6,22) show significant differences. In RNA $\begin{matrix}-A-A-\\ -U-U-\end{matrix}$ is the least stable [ΔG°(25°C) = -1.2 kcal mol^{-1}], $\begin{matrix}-A-U-\\ -U-A-\end{matrix}$ and $\begin{matrix}-U-A-\\ -A-U-\end{matrix}$ are intermediate, and $\begin{matrix}-C-A-\\ -G-U-\end{matrix}$, $\begin{matrix}-A-C-\\ -U-G-\end{matrix}$, $\begin{matrix}-G-A-\\ -C-U-\end{matrix}$ and $\begin{matrix}-A-G-\\ -U-C-\end{matrix}$ are the most stable [ΔG°(25°C) > -2.5 kcal mol^{-1}].

The mismatches and wobble base pairs confer a wide range of stabilities on the duplexes studied (Table I). The two duplexes with G•T wobbles have T_m's of 31°C and 27°C (depending on sequence) at 400 μM strand concentration, whereas an A•C mismatch results in a T_m of 19°C. The most stable non Watson-Crick base oppositions in the duplexes are G•T, G•G and G•A; the least stable are A•A, A•C, C•C and C•T. The thermodynamic contributions of all these base-base oppositions can be treated in a nearest-neighbor analysis as an internal loop of two bases regardless of whether the bases are hydrogen bonded or not (see Eq. 5). It is well known in fact that G•T and G•A are hydrogen bonded (7,18,20). The results (given in Table III) indicate the wide range of stability of "an internal loop of two bases". The ΔG° values show that G•T, G•G and G•A are the most stable; this is consistent with hydrogen bonding in G•T and G•A (13,18,20) and suggest that G•G also forms hydrogen bonds. Sequence effects and stacking are very important as shown by the slight destabilization caused by $\begin{matrix}-A-G-A-\\ -T-A-T-\end{matrix}$, but the large destabilization caused by $\begin{matrix}-A-A-A-\\ -T-G-T-\end{matrix}$. The other base oppositions destabilize the helix with an unfavorable standard free energy at 25° relative to $CA_5G \cdot CT_5G$ varying from 1.0 kcal mol^{-1} to 1.9 kcal mol^{-1}. In general the most unfavorable base to have in a mismatch is C; G tends to be the least destabilizing. The order of stability is approximately G•T > G•G > G•A > C•T > A•A = T•T > A•C = C•C, but is dependent on the surrounding sequence. The order may be somewhat different still in sequences which are not $dA_n \cdot dT_n$.

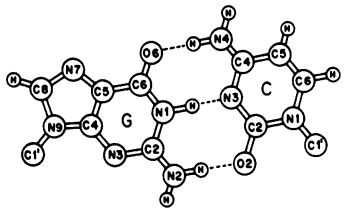
The data in Table III can be compared with the ΔG° (25°C) values used for calculating secondary structure in RNA (5,28). A G•U surrounded by A•U

Table III. Destabilization of Double Helices by Base-Base Mismatches or Wobble Base Pairs[a]

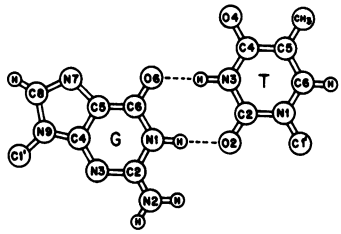
Mismatch/ Wobble	ΔG°, 25° (kcal mol^{-1})	ΔH° (kcal mol^{-1})	ΔS° (cal deg^{-1} mol^{-1})
-A-T-A- / -T-G-T-	0	-8.1	-27
-A-G-A- / -T-G-T-	0.2	-6.0	-20
-A-G-A- / -T-A-T-	0.3	-5.1	-18
-A-G-A- / -T-T-T-	0.7	0.8	1
-A-A-A- / -T-G-T-	1.2	7.6	22
-A-C-A- / -T-T-T-	1.2	-5.7	-23
-A-T-A- / -T-C-T-	1.5	-2.5	-13
-A-A-A- / -T-A-T-	1.5	10.6	31
-A-T-A- / -T-T-T-	1.5	-7.1	-29
-A-C-A- / -T-A-T-	1.9	7.2	38
-A-C-A- / -T-C-T-	2.0	-7.8	-33
-A-A-A- / -T-C-T-	2.1	11.7	32

[a]The values are obtained by subtracting nearest-neighbor contributions present in $dCA_5G \cdot dCT_5G$ from the data present in Table I. The values from a least squares fit to $dCA_5G \cdot dCT_5G$, $dCA_6G \cdot dCT_6G$, $dCA_7G \cdot dCT_7G$ are ΔG° = -6.5 kcal mol^{-1}, ΔH° = -47.5 kcal mol^{-1}, ΔS° = -138 cal deg^{-1} mol^{-1} for the nearest neighbor approximation to $dCA_5G \cdot CT_5G$.

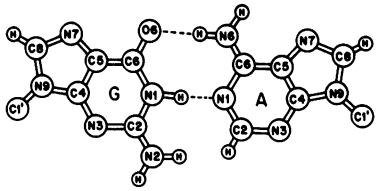
base pairs is assigned a stabilizing effect of ΔG° (25°C) = -0.6 kcal mol^{-1}. Table III gives -0.2 kcal to +0.5 kcal for the two orientations of G•T surrounded by A•T's. An internal loop of any two bases with A•U's on both sides is assigned a destabilizing effect of ΔG° (25°C) = 1.8 kcal mol^{-1} (28). This is found only for A•C and C•C surrounded by A•T's; other base oppositions are more stable. Ninio (29) has used a computer analysis of 100 transfer RNA's to estimate relative stabilities of base-base mismatches in RNA. His ranking is G•U≈G•G > C•C, U•U, C•A, A•A > G•A, U•C. Our DNA data disagree on the order of C•C, C•A vs. G•A, U•C. It is useful to note that among the most unstable duplexes of all are the ones in



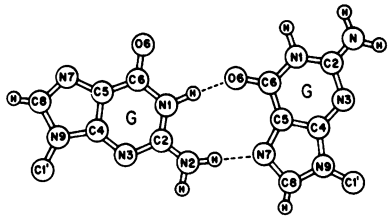
Watson Crick G (*anti*) · C (*anti*)



Wobble G (*anti*) · T (*anti*)



G (*anti*) · A (*anti*)



G (*anti*) · G (*syn*)

Fig. 3. Postulated base pairing schemes for base pairs involving guanine. The drawings are to scale; G•C is shown to indicate relative distances between C1' atoms and to give relative C1'-N bond orientations.

which an extra nucleotide (a bulge) is present on one strand. Similar results are found in RNA (5).

Possible hydrogen bonding schemes for all base-base oppositions are shown in Figures 3-5. Watson-Crick base pairing is included to allow a

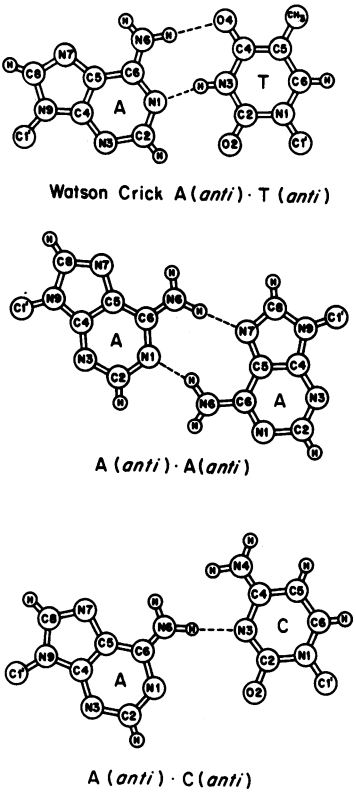
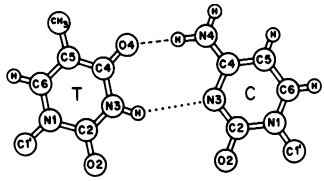
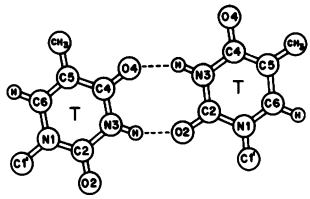


Fig. 4. Postulated base pairing schemes for base pairs involving adenine.

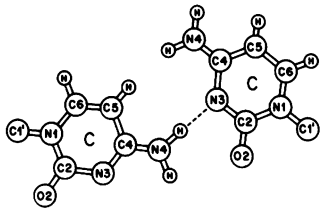
qualitative comparison of the sizes (C1'-C1' separation) and orientations of the different pairs. The structures given for G•C, G•T and G•A (17,18,20) are well established in DNA. The G (anti)•G (syn) base pair (Fig. 3) has not been seen to our knowledge, but the hydrogen bonding shown is postulated in a four-stranded poly G structure (30). Another possibility is two hydrogen bonds of the type N1-H...O6. In Fig. 4 the A•C structure has been postulated by Patel et al. (21) and by Kollman (31). The A (anti)• A (anti) structure is speculation; another possibility is A (anti)• A (syn) with two N6-H...N1 hydrogen bonds. Figure 5 contains the least stable base-base oppositions. Sugar-phosphate constraints may prevent any of these hydrogen bonds from forming, however Kollman has postulated the T•C structure shown (31). Direct proof for any of these base pairs should come from NMR studies of the exchangeable protons to show



T(*anti*) · C(*anti*)



T(*anti*) · T(*anti*)



C(*anti*) · C(*anti*)

Fig. 5. Possible pyrimidine-pyrimidine base pairs. There is no experimental evidence for these pairs. The geometries are very different from Watson-Crick base pairs.

hydrogen bonding, and NOE measurements to establish the syn or anti conformation.

The most plausible hydrogen bonding base pairs usually can be drawn for the base-base oppositions which are thermodynamically the most stable (ΔG° most negative). The data also show that the neighboring sequence is an important factor in stability. It is well known that base stacking is necessary for duplex stability. However, hydrogen bonding may be necessary to allow the bases close enough to each other inside the helix where they can stack.

Stabilities of base mismatches are important to understanding mutation frequencies and to design of oligonucleotide probes for locating genes. When only the amino acid sequence of the gene is known, the data in Tables I and III may be useful in deciding what strand of DNA to search

for, and what sequence to use. Of course, oligonucleotides with C•G base pairs on either side of the mismatch are needed to provide a more complete data set.

It should also be realized that the free energies of different base pairs vary differently with temperature, so their relative stabilities at a typical hybridization temperature may be quite different from those at 25°C. The standard free energy at any temperature can be calculated from that at 25°C using the standard entropy.

$$\Delta G°(T) = \Delta G°(25°C) - (T-25°C) \cdot \Delta S°$$

In Table I we have used this equation to provide values of $\Delta G°(50°C)$; this temperature is nearer the hybridization temperatures usually used. We note that for the $\Delta G°$ values at the higher temperatures the sequence dependence of the various base oppositions nearly vanish. The two G•C duplexes are more stable than the A•T duplexes. Pyrimidine-pyrimidine oppositions such as T•C become by far the most destabilizing. However, one should keep in mind that extrapolation to higher temperatures are least accurate for the least stable duplexes.

As our understanding of the conformations of the mismatches and of their effect on the thermodynamics improves, it may not be necessary to measure all possible sequences to predict the results. Evidence to test the hydrogen bonding shown in Figs. 3-6 will be most helpful.

ACKNOWLEDGEMENTS

We wish to thank Mr. Miguel Castro for assistance in oligonucleotide preparation. This research was supported in part by National Institute of Health grant GM10840, and by the U.S. Department of Energy, Office of Energy Research under contract 82ER60090.

REFERENCES

1. Lomant, A.J. and Fresco, J.R. (1975) in Progress in Nucleic Acid Research Molecular Biology, eds., Vol 15, pp. 185-218, Academic Press, New York.
2. Dickerson, R.E. and Drew, H.R. (1981) J. Mol. Biol. 149, 761-786.
3. Martin, F.H., Uhlenbeck, O.C. and Doty, P. (1971) J. Mol. Biol. 57, 201-215.
4. Gralla, J. and Crothers, D.M. (1973). J. Mol. Biol. 78, 301-319.
5. Tinoco, I., Jr., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers and Gralla, J., Nature New Biology 246, 40-41 (1973).
6. Borer, P.N., Dengler, B., Tinoco, I., Jr. and Uhlenbeck, O.C. (1974) J. Mol. Biol. 86, 843-853.
7. Nelson, J.W., Martin, F.H. and Tinoco, I., Jr. (1981) Biopolymers 20, 2509-2531.
8. Marky, L.A. and Breslauer, K.J. (1982) Biopolymers 21, 2185-2194.

9. Morden, K.M., Chu, Y.G., Martin, F.H. and Tinoco, I., Jr. (1983) Biochemistry 21, 428-436.
10. Patel, D.J., Kozlowski, S.A., Marky, L.A., Broka, C., Rice, J.A., Itakura, k. and Breslauer, K.J. (1982) Biochemistry 21, 428-436.
11. Patel, D.J., Kozlowski, S.A., Marky, L.A., Rice, J.A., Broka, C., Itakura, K. and Breslauer, K.J., (1982) Biochemistry 21, 445-451.
12. Pardi, A., Morden, K.M., Patel, D.J., and Tinoco, I., Jr. (1982) Biochemistry 21, 6567-6574.
13. Dodgson, J.B. and Wells, R.D. (1977) Biochemistry 16, 2367-2374.
14. Early, T.A., Olmsted, J. III, Kearns, D.R. and Lezius, A.G., (1978) Nucleic Acids Res. 5, 1955-1970.
15. Cornelis, A.G., Haasnoot, J.H.J., den Hartog, J.F., deRooij, M., van Boom, J.H. and Cornelis, A. (1979) Nature 281, 235-236.
16. Wallace, R.B., Shaffer, J., Murphy, R.F., Bonner, J., Hirose, T. and Itakura, K. (1979) Nucleic Acids Res. 6, 3543-3557.
17. Patel, D.J., Kozlowski, S.A., Marky, L.A., Rice, J.A., Broka, C., Dallas, J., Itakura, K. and Breslauer, K.J. (1982) Biochemistry 21, 437-444.
18. Kan, L., Chandrasegaran, S., Pulford, S.M. and Miller, P.S. (1983) Proc. Natl. Acad. Sci. USA 80, 4263-4265.
19. Tibanyenda, S.H., DeBruin, C., Haasnoot, A.G., Van Der Marel, G.A., Van Boom, J.H. and Hilbers, C.W. (1984) Eur. J. Biochem. 139, 19-27.
20. Patel, D.J., Kozlowski, S.A., Ikuta, S., and Itakura, K. (1984) Biochemistry 23, 3207-3217.
21. Patel, D.J., Kozlowski, S.A., Ikuta, S., and Itakura, K. (1984) Biochemistry 23, 3218-3226.
22. Freier, S.M., Petersheim, M., Hickey, D.R. and Turner, D.H. (1984) J. of Biomol. Structure and Dynamics 1, 1229-1242.
23. Peck, L.J. and Wang, J.C. (1981) Nature 292, 375-377.
24. Rhodes, D. and Klug, A. (1981) Nature 292, 378-380.
25. Strauss, F., Gaillard, C. and Prunell, A. (1981) European J. Biochemistry 118, 215-222.
26. Wu, H.M. and Crothers, D.M. (1984) Nature 308,509-513.
27. Arnott, S., Chandresekaran, R., Hall, I.H. and Puigjaner, L.C. (1983) Nucleic Acids Research 11, 4141-4155.
28. Salser, W., (1977) Cold Spring Harbor Symposia, 42, 985-1002
29. Ninio, J. (1979) Biochimie 61, 1133-1150.
30. Sasisekharan, V., Zimmerman, S. and Davis, D.R. (1975) J. Mol. Biol. 92, 171-179.
31. Keepers, J.W., Schmidt, P., James, T.L. and Kollman, P.A. (1984) Biopolymers 23, 2901-2929.