

Thermodynamic parameters for DNA sequences with dangling ends

Salvatore Bommarito, Nicolas Peyret and John SantaLucia Jr*

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

Received January 7, 2000; Revised and Accepted March 7, 2000

ABSTRACT

The thermodynamic contributions to duplex formation of all 32 possible single-nucleotide dangling ends on a Watson–Crick pair are reported. In most instances, dangling ends are stabilizing with free energy contributions ranging from +0.48 (G_A^T) to –0.96 kcal/mol (A_G^C). In comparison, Watson–Crick nearest-neighbor increments range from –0.58 (TA/AT) to –2.24 (GC/CG) kcal/mol. Hence, in some cases, a dangling end contributes as much to duplex stability as a Watson–Crick A–T base pair. The implications of these results for DNA probe design are discussed. Analysis of the sequence dependence of dangling-end stabilities show that the nature of the closing base pair largely determines the stabilization. For a given closing base pair, however, adenine dangling ends are always more or equally as stable as the other dangling nucleotides. Moreover, 5' dangling ends are more or equally as stabilizing as their 3' counterparts. Comparison of DNA with RNA dangling-end motifs shows that DNA motifs with 5' dangling ends contribute to stability equally or more than their RNA counterparts. Conversely, RNA 3' dangling ends contribute to stability equally or more than their DNA counterparts. This data set has been incorporated into a DNA secondary structure prediction algorithm (DNA MFOLD) (<http://mfold2.wustl.edu/~mfold/dna/form1.cgi>) as well as a DNA hybridization prediction algorithm (HYTHER™) (<http://jsl1.chem.wayne.edu/Hyther/hythermenu.html>).

INTRODUCTION

DNA nearest-neighbor parameters have already been determined for Watson–Crick base pairs (1,2) and all single internal mismatches (1,3–6). However, there is a significant lack of data in the literature for the stability of other DNA motifs, including dangling ends, terminal mismatches and various bulge and loop structures. Knowledge of the thermodynamic contribution of these motifs is important to accurately predict nucleic-acid hybridization and secondary structure (7). In this work, we report the thermodynamic contribution to duplex formation of the 32 DNA dangling ends with a closing Watson–Crick base pair. The available literature data indicate that dangling ends can significantly stabilize duplexes (8–11).

Therefore, knowledge of the thermodynamic contributions of all dangling-end contexts will be useful to improve primer design and optimize the design of immobilized probes in DNA microarrays (12) and other diagnostics such as molecular beacons (13). Moreover, this data set will improve the accuracy of predictions of hybridization and of DNA secondary structure. The empirical data reported here also presents an opportunity to evaluate various theoretical models for nucleic-acid stacking (14–16).

MATERIALS AND METHODS

DNA synthesis and purification

Oligonucleotides were synthesized on solid supports using standard phosphoramidite chemistry (17). Oligonucleotides were detached from the solid support and the base blocking groups were removed by treatment with concentrated ammonia at 50°C overnight. Each oligonucleotide sample was dried and dissolved in 250 μ l of water. The solution was then purified by TLC on a Si500F plate (Baker) using an elution mixture of *n*-propanol/ammonia/distilled deionized water in volumetric proportions 55:35:10, respectively (18). Bands were visualized under a UV lamp, the least mobile one was scraped off and oligonucleotides were extracted with distilled deionized water. The oligonucleotides were further purified and desalted with Sep-pak C-18 cartridges (Waters).

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 thermo-electric controller as described previously (19). The buffer used for the melting curves was 1.0 M NaCl, 20 mM sodium cacodylate and 0.5 mM Na₂EDTA, pH 7.0. Each duplex was melted at 8–10 different concentrations covering an 80- to 100-fold concentration 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 calculated with the nearest-neighbor model (20).

Calculation of thermodynamic parameters

Thermodynamic parameters were determined from melting curves using the program MELTWIN v.3.0 (21). Thermodynamic parameters were calculated by two methods: (i) enthalpy and entropy changes from fits of individual melting curves at different concentrations were averaged (22); (ii) plots of

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

Table 1. Thermodynamic parameters for duplex formation in 1 M NaCl³

Duplexes with Dangling Ends ^a	$\Delta H^{\circ b}$ (kcal/mol)	$\Delta S^{\circ b}$ (cal/mol K)	$\Delta G_{37}^{\circ b}$ (kcal/mol)	T_M^c (°C)
AAGAGCTCT ^c TCTCGAGAA	-52.4 ± 3.7	-141.6 ± 11.3	-8.51 ± 0.26	54.6
ACGATATCG ^d GCTATAGCA	-64.4 ± 1.3	-179.3 ± 4.1	-8.80 ± 0.05	52.8
AGTAGCTAC ^d CATCGATGA	-59.0 ± 1.0	-163.9 ± 3.3	-8.16 ± 0.04	50.5
ATGAGCTCA ^d ACTCGAGTA	-56.2 ± 0.8	-152.9 ± 2.3	-8.72 ± 0.03	54.8
CAGATATCT ^c TCTATAGAC	-47.8 ± 3.3	-135.2 ± 10.8	-5.88 ± 0.18	38.4
CCGATATCG ^d GCGATAGCC	-60.7 ± 1.6	-170.4 ± 4.9	-7.92 ± 0.08	48.8
CGTAGCTAC ^d CATCGATGC	-59.7 ± 0.9	-167.7 ± 2.9	-7.70 ± 0.02	47.9
CTTAGCTAA ^d AATCGATT C	-44.0 ± 1.8	-124.1 ± 5.7	-5.53 ± 0.03	36.0
GATAGCTAT ^c TATCGATAG	-51.8 ± 3.6	-146.1 ± 11.7	-6.50 ± 0.20	41.6
GCGATATCG ^d GCTATAGCG	-62.0 ± 0.7	-173.2 ± 2.3	-8.32 ± 0.02	50.9
GGTAGCTAC ^d CATCGATGG	-59.5 ± 2.5	-165.6 ± 7.6	-8.13 ± 0.12	50.3
GTAGCTCA ^d ACTCGAGTG	-55.0 ± 2.7	-155.3 ± 8.8	-6.83 ± 0.05	43.6
GTTAGCTAA ^d AATCGATTG	-46.3 ± 0.9	-135.4 ± 3.1	-4.44 ± 0.05	28.1
TACAGCTGT ^d TGTCGACAT	-62.9 ± 1.9	-172.7 ± 5.8	-9.35 ± 0.10	56.2
TCGATATCG ^d GCTATAGCT	-59.8 ± 1.4	-166.9 ± 4.5	-8.05 ± 0.04	49.9
TGTAGCTAC ^d CATCGATGT	-61.4 ± 1.5	-171.6 ± 4.5	-8.22 ± 0.04	50.4
TTCAGCTGA ^c AGTCGACTT	-54.7 ± 3.8	-150.4 ± 12.0	-8.04 ± 0.24	51.0
AGAGCTCTA ^c ATCTCGAGA	-54.2 ± 3.8	-148.7 ± 11.9	-8.46 ± 0.25	51.7
CGATATCGA ^d AGCTATAGC	-56.2 ± 0.8	-152.9 ± 2.3	-8.72 ± 0.03	54.8
GTAGCTACA ^d ACATCGATG	-63.5 ± 0.5	-178.3 ± 1.5	-8.65 ± 0.05	49.7
TGAGCTCAA ^d AACTCGAGT	-51.4 ± 1.2	-140.0 ± 3.8	-7.97 ± 0.06	51.5
ATAGCTATC ^c CTATCGATA	-40.7 ± 2.8	-113.1 ± 9.0	-5.63 ± 0.17	36.7
CGATATCGC ^d CGCTATAGC	-52.4 ± 1.8	-145.2 ± 5.8	-7.34 ± 0.04	47.2
GTAGCTACC ^d CCATCGATG	-56.8 ± 0.8	-158.6 ± 2.4	-7.64 ± 0.02	48.1
TCAGCTGAC ^c CAGTCGACT	-44.8 ± 3.1	-121.0 ± 9.7	-7.28 ± 0.22	48.5
ATAGCTATG ^c GTATCGATA	-52.9 ± 3.7	-150.2 ± 12.0	-6.27 ± 0.19	40.6
CGATATCGG ^d GGCTATAGC	-59.7 ± 0.9	-167.5 ± 2.8	-7.77 ± 0.02	48.3
GTAGCTACG ^d GCATCGATG	-58.1 ± 0.7	-164.6 ± 2.2	-7.02 ± 0.01	44.4

reciprocal melting temperatures (T_m^{-1} versus $\ln C_T$) were fitted to the following equation (23) (all sequences in this study are self-complementary):

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

Both methods assume a two-state model and $\Delta C_p^\circ = 0$ for the transition equilibrium (22,24). The two-state approximation was assumed to be valid for sequences in which the ΔH° values

Table 1. Continued

	$\Delta H^{\circ b}$ (kcal/mol)	$\Delta S^{\circ b}$ (cal/mol K)	$\Delta G_{37}^{\circ b}$ (kcal/mol)	T_M^c (°C)
TTAGCTAAG ^d GAATCGATT	-44.1 ± 2.0	-124.4 ± 6.6	-5.51 ± 0.07	35.7
ATAGCTATT ^c TATCGATA	-43.7 ± 3.1	-122.4 ± 9.8	-5.84 ± 0.18	37.6
CGATATCGT ^d TGCTATAGC	-60.7 ± 1.4	-170.7 ± 4.3	-7.58 ± 0.04	48.2
GTAGCTACT ^d TATCGATG	-62.0 ± 1.0	-175.7 ± 2.9	-8.05 ± 0.05	46.5
TTAGCTAAT ^d TATCGATT	-43.5 ± 1.6	-123.4 ± 5.1	-5.22 ± 0.06	33.7
Core sequences				
TTAGCTAA ^d AATCGATT	-35.9 ± 1.8	-98.1 ± 6.0	-5.48 ± 0.06	35.5
ACAGCTGT ^d TGTCGACA	-49.1 ± 3.1	-132.7 ± 9.5	-7.92 ± 0.13	51.8
CGATATCG ^f ^d GCTATAGC	-51.9 ± 0.6	-145.3 ± 1.4	-6.89 ± 0.09	44.1
GTAGCTAC ^f ^d CATCGATG	-51.6 ± 0.6	-143.7 ± 1.3	-7.01 ± 0.08	45.6
TGAGCTCA ^g ^d ACTCGAGT	-50.5 ± 0.5	-137.7 ± 1.3	-7.73 ± 0.04	50.4

^aThe top strand of each system is represented in the 5' to 3' orientation. Nucleotides involved in dangling ends are represented in bold.

^b ΔH° , ΔS° and ΔG_{37}° are the error-weighted averages of the of the $1/T_M$ versus $\ln C_T$ plot and average curve fit methods in Table S1.

^c T_M calculated using 10^{-4} M total strand concentration.

^dErrors reflect the precision of the data (see text).

^eErrors reflect the accuracy of the data (see text).

^fJ.SantaLucia, Jr *et al.* (19).

^gS.Varma and J.SantaLucia, Jr, unpublished results.

derived from the two methods agreed within 15% (25). Since the two methods are equally reliable, the error-weighted average of the data obtained by the two methods was calculated as follows and reported in Table 1 (26):

$$\Delta G_{37}^\circ = \{[\Delta G_{37}^\circ(i)/\sigma(i)^2] + [\Delta G_{37}^\circ(ii)/\sigma(ii)^2]\} / \{[1/\sigma(i)^2] + [1/\sigma(ii)^2]\} \quad 2$$

where $\Delta G_{37}^\circ(i)$ and $\sigma(i)$ represent the free energy and the corresponding precision obtained by method (i) and $\Delta G_{37}^\circ(ii)$ and $\sigma(ii)$ represent the free energy and the corresponding precision obtained by method (ii). Errors in the weight-averaged free energies were calculated as follows (26):

$$\sigma_{\text{weighted average}} = \{1/[1/\sigma(i)^2] + [1/\sigma(ii)^2]\}^{1/2} \quad 3$$

where $\sigma(i)$ and $\sigma(ii)$ represent the precision of the data obtained by (i) average of enthalpy and entropy changes from fits of individual melting curves at different concentrations (22) and (ii) plots of reciprocal melting temperatures (T_m^{-1} versus $\ln C_T$).

Design of sequences

Oligonucleotides were designed to have a melting temperature between 30 and 65°C and to minimize formation of undesired hairpin or slipped-duplex conformations that might result in a non-two-state transition (1). Sequences were also designed to be self-complementary so that the error in the dangling-end

Table 2. Thermodynamic parameters for 3' and 5' dangling ends

Dangling Ends ^a	$\Delta H^{\circ b}$ (kcal / mol)	$\Delta S^{\circ b}$ (cal / mol K)	$\Delta G_{37}^{\circ b}$ (kcal / mol)
AA	0.2 ± 2.6	2.3 ± 8.1	-0.51 ± 0.17
TT			
AC	-6.3 ± 0.7	-17.1 ± 2.2	-0.96 ± 0.05
CG			
AG	-3.7 ± 0.6	-10.0 ± 1.8	-0.58 ± 0.05
CC			
AT	-2.9 ± 0.5	-7.6 ± 1.3	-0.50 ± 0.02
AA			
CA	0.6 ± 2.4	3.3 ± 7.8	-0.42 ± 0.12
TT			
CC	-4.4 ± 0.9	-12.6 ± 2.5	-0.52 ± 0.06
GG			
CG	-4.0 ± 0.5	-11.9 ± 1.6	-0.34 ± 0.04
CC			
CT	-4.1 ± 1.3	-13.0 ± 4.1	-0.02 ± 0.03
AA			
GA	-1.1 ± 2.5	-1.6 ± 8.2	-0.62 ± 0.13
TT			
GC	-5.1 ± 0.5	-14.0 ± 1.4	-0.72 ± 0.05
GG			
GG	-3.9 ± 1.3	-10.9 ± 3.9	-0.56 ± 0.07
CC			
GT	-4.2 ± 0.8	-15.0 ± 2.7	0.48 ± 0.02
AA			
TA	-6.9 ± 1.8	-20.0 ± 5.6	-0.71 ± 0.08
TT			
TC	-4.0 ± 0.8	-10.9 ± 2.4	-0.58 ± 0.05
GG			
TG	-4.9 ± 0.8	-13.8 ± 2.4	-0.61 ± 0.05
CC			
TT	-0.2 ± 2.7	-0.5 ± 8.5	-0.10 ± 0.17
AA			
AA	-0.7 ± 2.6	-0.8 ± 8.3	-0.48 ± 0.17
AT			
C	-2.1 ± 0.5	-3.9 ± 1.4	-0.92 ± 0.05
AG			
G	-5.9 ± 0.4	-16.5 ± 1.0	-0.82 ± 0.05
AC			
T	-0.5 ± 0.7	-1.1 ± 2.0	-0.12 ± 0.04
AA			
A	4.4 ± 2.2	14.9 ± 7.3	-0.19 ± 0.12
CT			
C	-0.2 ± 1.0	-0.1 ± 3.0	-0.23 ± 0.05
CG			
G	-2.6 ± 0.5	-7.4 ± 1.4	-0.31 ± 0.04
CC			
T	4.7 ± 2.5	14.2 ± 7.7	0.28 ± 0.16
CA			
A	-1.6 ± 2.5	-3.6 ± 8.3	-0.50 ± 0.12
GT			
C	-3.9 ± 0.5	-11.2 ± 1.6	-0.44 ± 0.05
GG			
G	-3.2 ± 0.5	-10.4 ± 1.3	-0.01 ± 0.04
GC			

Table 2. Continued

Dangling Ends ^a	$\Delta H^{\circ b}$ (kcal / mol)	$\Delta S^{\circ b}$ (cal / mol K)	$\Delta G_{37}^{\circ b}$ (kcal / mol)
T	-4.1 ± 1.4	-13.1 ± 4.4	-0.01 ± 0.05
GA			
A	2.9 ± 2.3	10.4 ± 7.5	-0.29 ± 0.12
TT			
C	-4.4 ± 0.8	-13.1 ± 2.3	-0.35 ± 0.05
TG			
G	-5.2 ± 0.6	-15.0 ± 1.6	-0.52 ± 0.05
TC			
T	-3.8 ± 1.2	-12.6 ± 3.9	0.13 ± 0.04
TA			

^aThese parameters and their corresponding errors are calculated from Table 1 using equations 3 and 5.

^bThe top strand of each duplex is represented in the 5' to 3' orientation.

^cParameters for this dangling end are the error-weighted average of the parameters obtained for the two sequences studied [(GTGAGCTCA)₂ and (GTTAGCTAA)₂].

core sequence thermodynamics from the sequence with dangling ends. This is equivalent to assuming that dangling-end stability can be approximated by a nearest-neighbor model (25). For instance the free energy for the dangling end ${}^{\text{AG}}_{\text{C}}$ in the sequence (AGTAGCTAC)₂ is calculated as follows:

$$\Delta G_{37}^{\circ}({}^{\text{AG}}_{\text{C}}) = \frac{1}{2} \{ \Delta G_{37}^{\circ}[(\text{AGTAGCTAC})_2] - \Delta G_{37}^{\circ}[(\text{GTAGCTAC})_2] \} \quad 4$$

The factor of 1/2 in equation 4 is introduced because two dangling ends are present in each duplex (27). When thermodynamic data were not available for core sequences the nearest-neighbor model was used to predict the stability of the core sequence (1,2). The error in the dangling-end parameters (Table 2) was calculated by standard error propagation methods (26). For instance, the error for the dangling end ${}^{\text{AA}}_{\text{T}}$ free energy is calculated using the errors reported in Table 1 as follows:

$$\sigma[\Delta G_{37}^{\circ}({}^{\text{AA}}_{\text{T}})] = \frac{1}{2} \{ \sigma^2[\Delta G_{37}^{\circ}((\text{AAGAGCTCT})_2)] + \sigma^2[\Delta G_{37}^{\circ}((\text{AGAGCTCT})_2)] \}^{1/2} \quad 5$$

where $\sigma[\Delta G_{37}^{\circ}((\text{AAGAGCTCT})_2)]$ and $\sigma[\Delta G_{37}^{\circ}((\text{AGAGCTCT})_2)]$ are the errors reported in Table 1. Analogous equations apply for the error propagation for ΔH° and ΔS° of dangling-end formation. It is important to point out that the propagated errors obtained by these methods (equation 5) reflect the accuracy of the dangling-end parameters obtained for a given sequence context. Imposing the nearest-neighbor model itself, however, introduces additional uncertainty when the parameters are used to predict dangling-end stability in different sequence contexts (i.e. non-nearest-neighbor effects). Further, additional thermodynamic contributions from additional dangling bases can sometimes be significant (8,9) but are neglected with the model presented here.

Estimated errors

All of the sequences used in this study are 8 or 9 nt long and were specifically designed to avoid competing structures. As a result, the experimental precision of the data in Table 1 are

contribution would be decreased by a factor of two compared to a duplex with only one dangling end (27).

Determination of dangling-end contributions to duplex stability

When thermodynamic data were available for core sequences (19; S.Varma and J.SantaLucia,Jr, unpublished results) dangling-end contributions were obtained by subtracting the

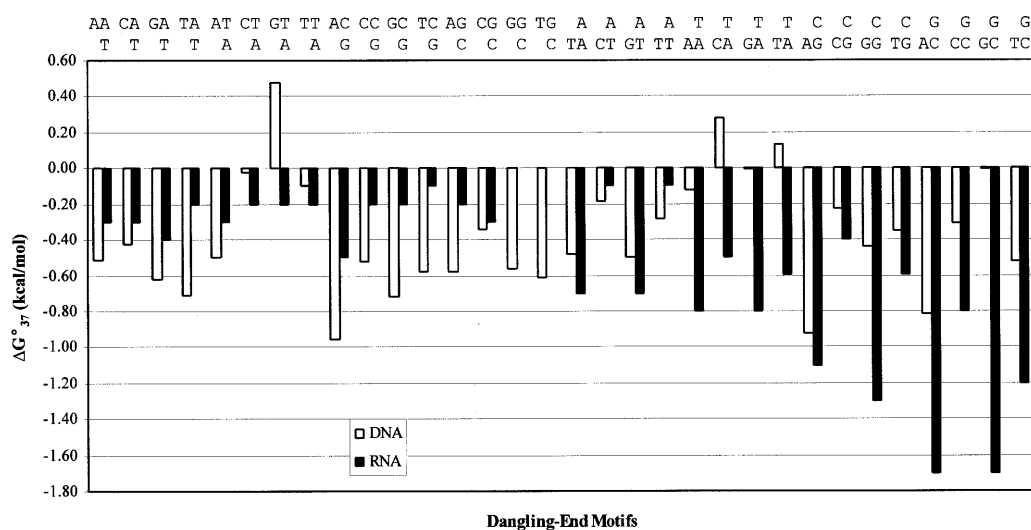


Figure 1. Comparison of DNA and RNA dangling-end free-energy contributions. Motifs are classified by closing base pair and 5' versus 3' position of the dangling end. Free energy contributions of RNA dangling-end motifs are from Freier *et al.* (25). For the corresponding RNA sequences substitute U for T. Note that the free energy contributions of the RNA sequences ${}^{\text{G}}_{\text{C}}$ and ${}^{\text{U}}_{\text{C}}$ are both 0.0 kcal/mol.

generally within 1–2% for ΔG°_{37} , 5% for ΔH° and 5–6% for ΔS° . The accuracy of the measurements is probably slightly worse than this because of systematic errors such as temperature calibration (estimated to be correct within 0.3°C) and application of the two-state approximation and the assumption that $\Delta C_p^{\circ} = 0$. Since the dangling-end and core sequences were both measured on the same instrument, however, systematic errors are likely to cancel so that errors in the calculated dangling-end contribution are largely reflected in the precision not accuracy of the measurements. Hence, when the core sequence corresponding to a given duplex with dangling ends was measured, errors reported in Table 1 for the duplex with dangling ends correspond to the precision of the data. When the core sequence thermodynamics were predicted using the nearest-neighbor model (2), errors reported in Table 1 for the corresponding sequence with dangling ends reflect the accuracy of the measurement and are conservatively estimated to be within 3, 7 and 8% for ΔG°_{37} , ΔH° and ΔS° , respectively.

RESULTS AND DISCUSSION

Thermodynamic data

Plots of T_m^{-1} versus $\ln C_T$ were linear with correlation coefficients ≥ 0.98 over the entire 80- to 100-fold range in concentration. Table S1 lists the thermodynamic parameters derived from fits of melting curves and from T_m^{-1} versus $\ln C_T$ plots (see Supplementary Material). The agreement in ΔH° values of the two methods is within 15% for all the sequences. Therefore the two-state approximation appears to be valid for these sequences (25). Since the data from both methods are equally reliable, the error-weighted average parameters (Table 1) were used in equation 4 to calculate the 32 dangling-end parameters shown in Table 2 (see Materials and Methods).

Application of dangling-end parameters

In this work, we assume that the nearest-neighbor model is an accurate approximation for sequences with terminal dangling ends. For instance, the thermodynamics for the sequence $(\text{AAGAGCTCT})_2$ is decomposed as follows:

$$\begin{aligned} \Delta G^{\circ}_{37}[(\text{AAGAGCTCT})_2] &= 2\Delta G^{\circ}_{37} \text{ initiation with terminal AT} + \\ &\Delta G^{\circ}_{37} \text{ sym} + 2\Delta G^{\circ}_{37}(\text{AA}_{\text{T}}) + 4\Delta G^{\circ}_{37}(\text{AG}_{\text{T}}) + 2\Delta G^{\circ}_{37}(\text{GC}_{\text{T}}) + \\ &\Delta G^{\circ}_{37}(\text{GC}_{\text{G}}) \end{aligned} \quad 6$$

Stability trends

Dangling ends are stabilizing in 24 out of 32 cases and destabilizing in only three out of 32 cases with ΔG°_{37} values ranging from +0.48 to –0.96 kcal/mol (Fig. 1 and Table 2). Stabilization depends on four factors: the base type (A, C, G or T) and position (5' versus 3') of the dangling nucleotide and the type and orientation of the closing base pair (A–T, T–A, C–G or G–C). For a given closing base pair of given orientation, adenine dangling ends are always more or equally as stabilizing as the other dangling ends. However, for a given closing base pair, C, G and T dangling nucleotides show stabilizations that do not differ significantly from one another. Therefore, with the exception of A dangling ends, the dangling nucleotide identity plays only a minor role in the stabilization. The only exceptions to this trend are observed for ${}^{\text{G}}_{\text{GC}}$ ($\Delta G^{\circ}_{37} = -0.01$ kcal/mol), which is less stable than ${}^{\text{G}}_{\text{TC}}$ ($\Delta G^{\circ}_{37} = -0.52$ kcal/mol), and for ${}^{\text{GT}}_{\text{A}}$ ($\Delta G^{\circ}_{37} = +0.48$ kcal/mol), which is less stable than both ${}^{\text{CT}}_{\text{A}}$ ($\Delta G^{\circ}_{37} = -0.02$ kcal/mol) and ${}^{\text{TT}}_{\text{A}}$ ($\Delta G^{\circ}_{37} = -0.10$ kcal/mol). To ensure that the observation for ${}^{\text{GT}}_{\text{A}}$ is not the result of an experimental artifact, two different sequences bearing this dangling-end motif were melted. Both sequences $[(\text{GTTAGCTAA})_2$ and $(\text{GTGAGCTCA})_2]$ showed identical contributions for the dangling-end motif (Table 1). This result suggests that the unusual behavior of ${}^{\text{GT}}_{\text{A}}$ is not due to next nearest-neighbor effects or due to unusual single strand stability. At present we do not have a good explanation

for why GT_A^T is destabilizing. The 3' versus 5' position of a dangling end can result in stabilization of up to 0.98 kcal/mol (GT_A^T versus GT_T^A). Moreover, 5' dangling ends are more or equally as stable as their 3' counterparts. This observation confirms previous results obtained by Senior *et al.* (8). The only significant exceptions to this trend are observed for GT_A^T (+0.48 kcal/mol) versus GT_T^A (-0.50 kcal/mol) and AG_C^G (-0.58 kcal/mol) versus AG_G^C (-0.92 kcal/mol). Furthermore, changing the closing base pair type and orientation can lead to stabilization by up to 1.20 kcal/mol (GC_G^C versus GT_A^T).

Comparison with literature data

A few thermodynamic measurements of DNA sequences with dangling ends are reported in the literature. Guckian *et al.* (10) studied AC_G^C and TC_G^C dangling-end motifs and reported stabilizations of -1.0 ± 0.2 and -0.6 ± 0.2 kcal/mol, respectively. These measurements are in excellent agreement with our data of -0.96 ± 0.05 kcal/mol for AC_G^C and -0.58 ± 0.05 kcal/mol for TC_G^C . In another study, Doctycz *et al.* (9) studied hairpins with 5' base overhangs on their 5'-side. They observed that the stabilization depended mainly on the dangling nucleotide adjacent to the duplex, consistent with the nearest-neighbor approximation. To allow comparison with our data, the data of Doctycz *et al.* were converted from ΔG_{25}° to ΔG_{37}° . Averages of the free energy measurements obtained for all overhangs with an identical dangling nucleotide adjacent to the duplex (AG_C^G , -0.58 kcal/mol; GG_C^C , -0.53 kcal/mol; CG_C^G , -0.59 kcal/mol; TC_C^G , -0.54 kcal/mol) are within the experimental error of the results presented in this work (AG_C^G , -0.58 kcal/mol; GG_C^C , -0.56 kcal/mol; CG_C^G , -0.34 kcal/mol; TC_C^G , -0.61 kcal/mol). Since the data of Guckian *et al.* (10) and Doctycz *et al.* (9) were measured in different contexts than the sequences in this study, these data suggest that the nearest-neighbor model is a good approximation for dangling ends. Senior *et al.* (8) also reported thermodynamic measurements of duplexes with two T nucleotide overhangs. To allow comparison with our work, the data of Senior *et al.* were converted from ΔG_{25}° to ΔG_{37}° (TC_G^C , -0.93 kcal/mol; TC_C^G , -0.45 kcal/mol; TC_G^G , -1.01 kcal/mol; TC_C^G , -0.35 kcal/mol). These data of Senior *et al.* (8) are basically within the experimental error of our measurements (TC_G^C , -0.61 kcal/mol; TC_C^G , -0.52 kcal/mol; TC_G^G , -0.58 kcal/mol; TC_C^G , -0.35 kcal/mol). It is possible that the slight differences are due either to experimental error or due to the effect of tandem TT dangling ends that are not present in our sequences.

Comparison with terminal mismatch data

Dangling-end parameters were also compared with the 48 terminal mismatch parameters (S.Varma and J.SantaLucia,Jr, unpublished results) to test if the stability contribution of a terminal mismatch equals the sum of the contributions of the two corresponding dangling ends. For instance, the terminal mismatch motif G_C^A can be expressed as the sum of the two dangling end motifs G_C^A and G_A^C . In eight of the 48 cases, the sums of the dangling-end free energy values are >0.4 kcal/mol more stable than the corresponding terminal-mismatch free-energy values. This observation can be explained by considering the possibility for optimal stacking in the case of dangling ends, which is sterically precluded for terminal mismatches. In 29 of the 48 cases, the sums of the dangling-end free-energy contributions are within 0.4 kcal/mol of the corresponding terminal-mismatch free-energy contributions. Three explanations

for this observation are: (i) neither dangling-end motif yields any significant stabilization and hydrogen bonding in the terminal mismatch is negligible; (ii) the terminal-mismatch and dangling-end structures stack with equal efficiency in some cases; (iii) stacking of the terminal nucleotides is less efficient in the terminal-mismatch structure than in the dangling-end structures, but there is compensating favorable hydrogen bonding or electrostatic interactions between the bases of the terminal nucleotides. In 11 of the 48 cases, the sums of the dangling-end free-energy values are >0.4 kcal/mol less stable than the corresponding terminal-mismatch free-energy values. In these cases, the higher stability of the terminal-mismatch parameters may be due to interactions between the terminal bases, such as hydrogen bonding or electrostatic interactions. Further structural data on terminal mismatches are required to determine the exact nature of these interactions. Similar cases in which the sum of the dangling-end stabilities are either equal to or higher than the corresponding terminal-mismatch free energy were also observed for RNA (28). Solvation and conformational entropy may also be important determinants of dangling-end and terminal-mismatch stability (29).

Comparison with RNA

Comparison with the RNA dangling-end parameters compiled by Freier *et al.* (25,28) show that DNA motifs with 5' dangling ends are more or equally as stable as their RNA counterparts. The only exception to this trend is observed for GT_A^T . On the other hand, RNA motifs with 3' dangling ends are more or equally as stable as their DNA counterparts. A geometric explanation for this difference in stability of 3' versus 5' RNA dangling ends is that for A-form RNA structures 3' nucleotides are well stacked on top of the adjacent base pair whereas 5' dangling ends display poor stacking. For B-form DNA, however, dangling ends show efficient geometric intrastrand stacking. Thus, DNA trends (dictated by B-form geometry) are expected to differ from RNA trends (dictated by A-form geometry) (28). This expectation is verified by our DNA measurements that show 5' dangling ends are always more or equally as stable as their 3' counterparts.

Physical significance of dangling-end parameters

Each dangling-end thermodynamic parameter reflects the stacking ability of the dangling base on the duplex closing base pair. The fundamental interactions responsible for the sequence dependence of stacking are, however, not clear and probably include dipole-induced dipole (14,30) or induced dipole-induced dipole interactions (31,32) as well as solvophobic effects (30,33). Further understanding of the physical chemistry of dangling-end stabilization might be achieved by solving the structures of several duplexes with dangling ends. Theoretical approaches such as *ab initio* quantum mechanical calculations could also be used to determine the origin of the stabilization. To date, however, the various studies available have focused on base/base stacking (15,16,31,32) and none have studied stacking of an unpaired base on top of a base pair. Thus, we are currently unable to correlate the literature data (15,31,32,34) with our thermodynamic parameters. With the development of improved computational techniques, it is likely that in the future such theoretical approaches will bring insights into the structural basis of the thermodynamic data.

Implications for DNA probe design

Oligonucleotide probe hybridization to a long target DNA involves not only base pairing, but also two dangling-end contributions. This work demonstrates that the free-energy contribution of each dangling end ranges from +0.48 ($\frac{GT}{A}$) to -0.96 kcal/mol ($\frac{AC}{G}$). Thus, depending on the sequence, two dangling ends can affect probe-target hybridization by as much as +0.96 to -1.92 kcal/mol. By comparison, the stabilization conferred by addition of one Watson-Crick base pair to a probe-target duplex ranges from -0.58 (TA/AT) to -2.24 kcal/mol (GC/CG) (2). Therefore, accounting for dangling-end effects is crucial to accurately calculate probe/target binding strength and dangling-end parameters should be incorporated into current probe design programs.

CONCLUSIONS

The set of measurements reported in this paper is the first systematic analysis of dangling end contributions to DNA duplex stability. The data show that in many cases dangling ends stabilize duplexes as much as an additional A-T base pair would. These parameters will be useful to optimize probe-target-based biochemical techniques as well as DNA secondary structure prediction. The dangling end parameters reported here are now implemented in the DNA MFOLD secondary structure prediction algorithm (35; J.SantaLucia,Jr and M.Zuker, unpublished results; <http://mfold2.wustl.edu/~mfold/dna/form1.cgi>) as well as in the HYTHER™ hybridization optimization algorithm (N.Peyret and J.SantaLucia,Jr, unpublished results; <http://jsl1.chem.wayne.edu/Hyther/hythermenu.html>).

SUPPLEMENTARY MATERIAL

See Supplementary Material available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Christine Chow and May Meroueh for assistance with DNA synthesis. We thank Dr Shikha Varma for providing unpublished data on terminal mismatches and for useful discussions. This work was supported by NIH grant HG02020.

REFERENCES

- Allawi,H.T. and SantaLucia,J.,Jr (1997) *Biochemistry*, **36**, 10581–10594.
- SantaLucia,J.,Jr (1998) *Proc. Natl Acad. Sci. USA*, **95**, 1460–1465.
- Allawi,H.T. and SantaLucia,J.,Jr (1998) *Biochemistry*, **37**, 2170–2179.
- Allawi,H.T. and SantaLucia,J.,Jr (1998) *Biochemistry*, **37**, 9435–9444.
- Allawi,H.T. and SantaLucia,J.,Jr (1998) *Nucleic Acids Res.*, **26**, 2694–2701.
- Peyret,N., Seneviratne,P.A., Allawi,H.T. and SantaLucia,J.,Jr (1999) *Biochemistry*, **38**, 3468–3477.
- Mathews,D.H., Sabina,J., Zuker,M. and Turner,D.H. (1999) *J. Mol. Biol.*, **288**, 911–940.
- Senior,M., Jones,R.A. and Breslauer,K.J. (1988) *Biochemistry*, **27**, 3879–3885.
- Doktycz,M.J., Paner,T.M., Amaratunga,M. and Benight,A.S. (1990) *Biopolymers*, **30**, 829–845.
- Guckian,K.M., Schweitzer,B.A., Ren,R.X.-F., Sheils,C.J., Paris,P.L., Tahmassebi,D.C. and Kool,E.T. (1996) *J. Am. Chem. Soc.*, **118**, 8182–8183.
- Quartin,R.S. and Wetmur,J.G. (1989) *Biochemistry*, **28**, 1040–1047.
- Williams,J.C., Case-Green,S.C., Mir,K.U. and Southern,E.M. (1994) *Nucleic Acids Res.*, **22**, 1365–1367.
- Tyagi,S. and Kramer,F.R. (1996) *Nature Biotechnol.*, **14**, 303–308.
- Gellman,S.H., Haque,T.S. and Newcomb,L.F. (1996) *Biophys. J.*, **71**, 3523–3526.
- Sponer,J., Gabb,H.A., Leszczynski,J. and Hobza,P. (1997) *Biophys. J.*, **73**, 76–87.
- Florian,J., Sponer,J. and Warshel,A. (1999) *J. Phys. Chem. B*, **103**, 884–892.
- Brown,T., Leonard,G.A. and Booth,E.D. (1990) *J. Mol. Biol.*, **212**, 437–440.
- Chou,S.-H., Flynn,P. and Reid,B. (1989) *Biochemistry*, **28**, 2422–2435.
- SantaLucia,J.,Jr, Allawi,H.T. and Seneviratne,P.A. (1996) *Biochemistry*, **35**, 3555–3562.
- Richards,E.G. (1975) In Fasman,G.D. (ed.), *Handbook of Biochemistry and Molecular Biology: Nucleic Acids*. 3rd Edn. CRC Press, Cleveland, OH, Vol. 1, pp. 597.
- McDowell,J.A. and Turner,D.H. (1996) *Biochemistry*, **35**, 14077–14089.
- Petersheim,M. and Turner,D.H. (1983) *Biochemistry*, **22**, 256–263.
- Borer,P.N., Dengler,B., Tinoco,I.,Jr and Uhlenbeck,O.C. (1974) *J. Mol. Biol.*, **86**, 843–853.
- Freier,S.M., Sugimoto,N., Sinclair,A., Alkema,D., Neilson,T., Kierzek,R., Caruthers,M.H. and Turner,D.H. (1986) *Biochemistry*, **25**, 3214–3219.
- Freier,S.M., Kierzek,R., Jaeger,J.A., Sugimoto,N., Caruthers,M.H., Neilson,T. and Turner,D.H. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 9373–9377.
- Bevington,P.R. (1969) *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, New York, NY, pp. 56–65 and 66–91.
- SantaLucia,J.,Jr, Kierzek,R. and Turner,D. (1991) *J. Am. Chem. Soc.*, **113**, 4313–4322.
- Freier,S.M., Alkema,D., Sinclair,A., Neilson,T. and Turner,D.H. (1985) *Biochemistry*, **24**, 4533–4539.
- Turner,D.H., Sugimoto,N., Kierzek,R. and Breiker,S.D. (1987) *J. Am. Chem. Soc.*, **109**, 3783–3785.
- Newcomb,L.F. and Gellman,S.H. (1994) *J. Am. Chem. Soc.*, **116**, 4993–4994.
- Hobza,P., Sponer,J. and Polásek,M. (1995) *J. Am. Chem. Soc.*, **117**, 792–798.
- Friedman,R.A. and Honig,B. (1995) *Biophys. J.*, **69**, 1528–1535.
- Kankia,B.I. and Marky,L.A. (1999) *J. Phys. Chem. B*, **103**, 8759–8767.
- Norberg,J. and Lennart,N. (1996) *Biopolymers*, **39**, 765–768.
- Zuker,M. (1989) *Science*, **244**, 48–52.