# Sequence Context and Thermodynamic Stability of a Single Base Pair Mismatch in Short Deoxyoligonucleotide Duplexes

Timothy S. Hall, Petr Pancoska, Peter V. Riccelli, Kathleen Mandell, and Albert S. Benight*

*Department of Chemistry, University of Illinois at Chicago 845 W. Taylor Street, Room 4500, Chicago, Illinois 60617*

*Received June 7, 2001*

**Introduction.** Perturbations of duplex DNA thermodynamic stability associated with single base pair mismatches are of primary importance in probe sequence design and of concern when analyzing results of diagnostic nucleic acid hybridization reactions. Single base pair mismatches in duplex DNA have also been associated with replication errors and mutagenesis in vivo.[1−3] Thus, for both biological and chemical reasons the ability to calculate effects of mismatches on duplex DNA hybridization is important. Perhaps most notable are applications that require estimates of single base pair mismatch stability.[4−6] For example, a significant challenge in designing and analyzing highly parallel, multiplex hybridization reactions is minimization of undesired cross annealing between different sequences. If duplexes with mismatches can form, from a design standpoint it is beneficial to know what their stabilities will be relative to the perfect matched duplexes. For most purposes sequence design is based on the use of reported sequence dependent stability parameters for duplex DNA.[7,8] Perhaps most notable among these are the parameters based on the nearest-neighbor model, evaluated by SantaLucia and co-workers[4−6] from melting studies of short linear duplex DNA oligomers. Their nearest-neighbor stacking and single base pair mismatch parameters are used in the HyTher program[9] to calculate the thermodynamic stability of short duplex DNAs. Within the nearest-neighbor model sequence specific parameters are available for calculating stabilities of both perfectly matched duplexes and duplexes having single base mismatches.

It is generally accepted that predictions from HyTher for standard duplexes, and those having single base pair mismatches, provide fairly accurate results. However, as has been pointed out,[8] using the SantaLucia parameters, or other nearest-neighbor stability parameter sets derived from melting analysis of other types of samples such as DNA dumbbells, DNA restriction fragments, or homogeneous sequence polymers, many cases can be readily found where accurate $t_m$ values cannot be reliably predicted.[8] This despite empirical corrections that have been suggested.[7] Although the actual origins of this shortcoming are not known, a strong possibility is the potential influence of sequence context (beyond nearest-neighbors) on duplex oligomer stability. Such sequence context effects have not been considered in stability predictions of short duplex DNAs.

Relationships between thermodynamic stability of flanking sequence and site specific binding by restriction enzymes have been explored and found in a number of model deoxyoligonucleotide duplex systems.[10−12] Here we report clear experimental evidence for the influence of flanking sequence context, beyond nearest-neighbors, on the thermodynamic stability of a single base pair mismatch in a short (31 base pair) linear deoxyoligonucleotide duplex.

**PM duplex**

5′-taa aag ata cca tca atg agg aag ctg cag a-3′
3′-att ttc tat ggt agt tac tcc ttc gac gtc t-5′

**AAG/TCC (Left Side)**

5′-taa aag ata cca tca atg agg aag ctg cag a-3′
3′-att tcc tat ggt agt tac tcc ttc gac gtc t-5′

**AAG/ TCC (Right Side)**

5′-taa aag ata cca tca atg agg aag ctg cag a-3′
3′-att ttc tat ggt agt tac tcc tcc gac gtc t-5′

**Figure 1.** Sequences of the 31 base pair duplex DNAs that were studied. Boxes denote positions of single base pair mismatches. Duplex samples were prepared as described in refs 11 and 12.

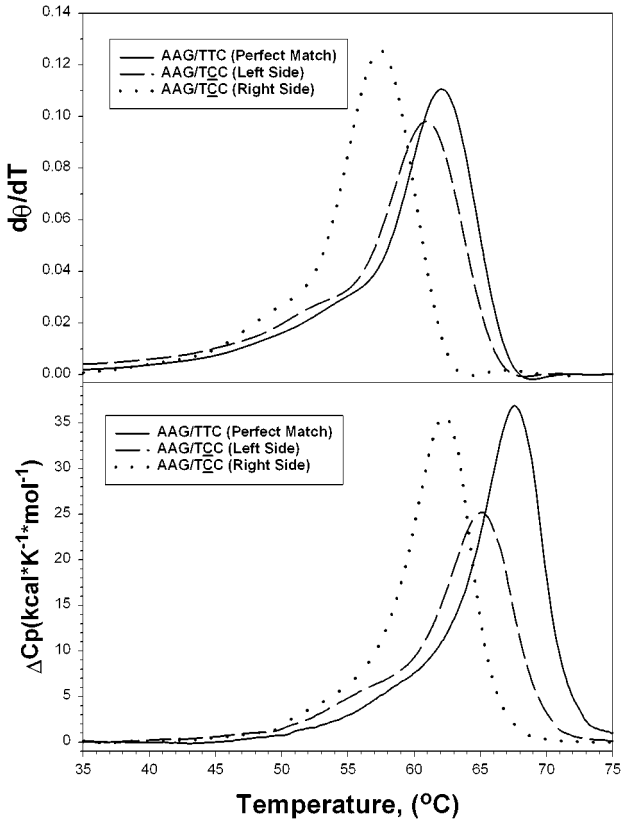


**Figure 2.** Melting Curves of the duplex DNAs. Differential optical melting curve (top) obtained by differentiation of the melting curve of $\theta$ versus temperature, where $\theta$ is the fraction of broken base pairs as a function of temperature. The DSC melting curve (bottom) is a plot of the excess heat capacity, $\Delta C_P$, versus temperature. These data were obtained and analyzed as described in refs 11 and 12.

**Results and Discussion.** Sequences of duplex DNA molecules that were studied are shown in Figure 1. Each is comprised of an annealed complementary pair of 31 base single strands. One of these is the control, a perfectly matched, PM duplex, comprised of 31 intact base pairs. Dashed boxes indicate sites where mismatches exist in the other duplexes. Each mismatch duplex contains the same single base pair mismatch (A/C), flanked by the same nearest-neighbor base pairs on the 5′ (left) side (A:T) and 3′ (right) side (G:C), (AAG/TCC), at a different position in

(1) Kunkel, T. A.; Bebenek, R. *Annu. Rev. Biochem.* **2000**, *69*, 497−529.
(2) Modrich, P.; Lahue, R. *Annu. Rev. Biochem.* **1996**, *65*, 101−133.
(3) Evans, E.; Alani, E. *Mol. Cell Biol.* **2000**, *20* (21), 7839−7844.
(4) Allawi, H. T.; SantaLucia, J., Jr. *Biochemistry* **1998**, *37*, 9435−9444.
(5) Allawi, H. T.; SantaLucia, J., Jr. *Biochemistry* **1998**, *37*, 2170−2179.
(6) Allawi, H. T.; SantaLucia, J., Jr. *Biochemistry* **1997**, *36*, 10581−10594.
(7) SantaLucia, J., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1460−1465.
(8) Owczarzy, R.; Vallone, P. M.; Gallo, F. J.; Paner, T. M.; Lane, M. J.; Benight, A. S. *Biopolymers (Nucleic Acid Sci.)* **1997**, *4*, 217−239.
(9) SantaLucia, J., Jr. Hyther web site: http://jsl7.chem.wayne.edu/home-.html.
(10) Benight, A. S.; Gallo, F. J.; Paner, T. M.; Bishop, K. D.; Faldasz, B. D.; Lane, M. J. *Adv. Biophys. Chem.* **1995**, *5*, 1−55.
(11) Riccelli, P. V.; Vallone, P. M.; Kashin, I.; Faldasz, B. D.; Lane, M. J.; Benight, A. S. *Biochemistry* **1999**, *38*, 11197−11208.
(12) Vallone, P. M.; Benight, A. S. *Biochemistry* **2000**, *39*, 7835−7846.

10.1021/ja016360j CCC: $20.00 
Published on Web 11/01/2001

**Table 1.** Summary of Thermodynamic Parameters of Three Deoxyoligonucleotide Duplexes

| duplex | $t_m$ (EXP) | $t_m$ (HyTher) | $\Delta t_m^{E}$ [a] | $\Delta t_m^{C}$ [b] | $\Delta H$ (kcal·mol$^{-1}$) | $\Delta S$ (cal·K$^{-1}$·mol$^{-1}$) | $\Delta G$ (kcal·mol$^{-1}$) | DSC $t_m$ |
|---|---|---|---|---|---|---|---|---|
| PM | 62.0 | 62.4 | 0 | 0 | 281 ± 1 | 831 ± 3 | 34 ± 1 | 67.3 |
| MM(L) | 60.9 | 58.3 | −1.1 | −4.1 | 223 ± 2 | 664 ± 4 | 25 ± 2 | 65.0 |
| MM(R) | 57.4 | 58.3 | −4.6 | −4.1 | 253 ± 2 | 758 ± 5 | 27 ± 2 | 62.0 |

[a] $\Delta t_m^{E} = t_m(\text{MM}) - t_m(\text{PM})$ for (EXP). [b] $\Delta t_m^{C} = t_m(\text{MM}) - t_m(\text{PM})$ for (HyTher).

the duplex. The sequence context beyond nearest-neighbor base pairs surrounding the A/C mismatch is different in each duplex. Optical and differential scanning calorimetry (DSC) melting curves for the three duplexes shown in Figure 1 were collected for DNA solutions in a 55 mM $Na^+$ buffer. Optical melting transitions were measured for DNA samples at an identical concentration (3.6 $\mu$M strands). DSC melting curves were collected at strand concentrations ranging from 60 to 90 $\mu$M strands. These measurements provided quantitative assessment of the influence of sequence context on perturbations of the melting transition temperatures and thermodynamic stability associated with a (AAG/TCC) mismatch.

The differential optical and DSC melting curves of the three duplexes studied, AAG/TTC (Perfect Match), AAG/TCC (Left Side), and AAG/TCC (Right Side), are shown in Figure 2. There are clear differences in the melting curves for the mismatches (dotted and broken lines), indicating the melting processes for these duplexes are different. A slight pre-transition shoulder occurs on the curves for the molecule with the mismatch on the left (dashed line), perhaps indicative of fraying of the A−T rich left end. End-fraying effects are expected to be small beyond three bases,[13] and therefore should not significantly influence the results. Interestingly, both the differential optical and DSC melting curves have nearly the same relative shapes.

Results of melting experiments are summarized in Table 1 where the transition temperatures from optical and DSC experiments and DSC thermodynamic parameters evaluated assuming $\Delta C_p = 0$ are given. Measured $t_m$(EXP) values for each duplex obtained from optical melting curves and predicted using SantaLucia's HyTher program[9] are also given in Table 1. Uncertainties in $t_m$(EXP) values are 0.5 °C, while uncertainty in the DSC $t_m$ values is slightly greater at 1.0 °C. These results show the $t_m$ values of the mismatch duplexes are less than that of the perfect match, $\Delta t_m$, but not by the same amount. For the perfect match (PM), $t_m = 62.0$ °C, while $t_m = 60.9$ °C and $\Delta t_m = 1.1$ °C for the mismatch on the left side, MM(L). For the mismatch on the right side, MM(R), $t_m = 57.4$ °C and $\Delta t_m = 4.6$ °C! The predicted $t_m$ value from HyTher is in good agreement for the perfect match duplex (62.4 versus 62.0 °C) but not in agreement for both mismatches. Not surprisingly, because identity of the mismatch and flanking nearest-neighbor base pairs are the same, HyTher (based on the nearest-neighbor model) predicts the $t_m$ of the mismatch duplexes to be identical. The thermodynamic parameters for the mismatch duplexes are also considerably different. In particular for MM(L) the transition enthalpy, $\Delta H$, and entropy, $\Delta S$, values are significantly smaller than for MM(R), yet the free energies of both duplexes, $\Delta G$ (at 25 °C), are essentially the same within experimental error. These differences between the DSC parameters for the MM(L) and MM(R) molecules clearly show the effect on context on mismatch stability.

The nearest-neighbor model has marginal validity in 55 mM $Na^+$.[14,15] Although beyond the scope of this communication, an analysis of the dependence of mismatch stability on $Na^+$ environment is required to determine relationships between [$Na^+$] and sequence dependence of mismatch thermodynamics.

Practically, our results serve to underscore the need for more effective sequence design strategies that consider effects of sequence context on DNA stability calculations. Toward this end, we have developed functional representations of DNA sequences and convenient algorithms for the quantitative treatment of sequence context and oligoduplex stability. Applications of these methods to analyze the data presented here will be the subject of a future publication.

JA016360J

(13) Amaratunga, M.; Snowden-Ifft, E.; Wemmer, D. E.; Benight, A. S. *Biopolymers* **1992**, *32*, 865−879.

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

(15) Owczarzy, R.; Vallone, P. M.; Goldstein, R. F.; Benight, A. S. *Biopolymers (Nucleic Acid Sci.)* **1999**, *52*, 29−56.