THERMODYNAMIC STUDIES OF TANDEM MISMATCHES AND OTHER STRUCTURAL ELEMENTS IN HAIRPIN AND DUPLEX NUCLEIC ACIDS

A thesis
Presented to
The Academic Faculty

by

Brooke Nicole Bourdélat-Parks

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in the
School of Biology

Georgia Institute of Technology
November 2003

THERMODYNAMIC STUDIES OF TANDEM MISMATCHES AND OTHER STRUCTURAL ELEMENTS IN HAIRPIN AND DUPLEX NUCLEIC ACIDS

Approved by:

Roger M. Wartell, Chairman

Jung H. Choi

Terry W. Snell

Nicholas V. Hud

Loren D. Williams

Date Approved: November 24, 2003

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to those people who have been a part of my life during my graduate school experience. First and foremost, thank you to Dr. Roger Wartell, for so many things. He has pushed me to explore things I never thought I was interested in, served as my voice of caution at times, patiently explained things I didn't understand, given me the space to make mistakes, and, most importantly, accepted my unfortunate lack of math skills. These are just a few of the things for which he deserves credit. Thank you also to the other members of my committee, Dr. Jung Choi, Dr. Terry Snell, and Dr. Nick Hud. A special thanks to Dr. Loren Williams with whom I had numerous scientific conversations, and who almost always provided a thought-provoking explanation. Additionally, I would like to express my gratitude to all the undergraduate students who assisted me over the years, in particular, Gemma Ellis and Karissa McClinic.

Graduate school would be an arduous task if it weren't for fellow students who can truly understand the trials and tribulations. I appreciate the friendship of those students with whom I entered Georgia Tech in 1997, particularly Cody Cain, Nikolle Reyes, and Joy Ting. My former roommate and fellow student Denise Jones led the way for me. The last few years have been made more enjoyable by my labmate Xueguang Sun, and my friends in other labs: George Lountos, Derrick Watkins, Tinoush Moulaei, Tatsuya Maehigashi, and Christine Conwell. I appreciate all the scientific and not-so-scientific conversations over the years. Finally, a very sincere thank you to Kris Woods

for always going above and beyond what is necessary, as both a fellow scientist and as my great friend.

Thank you to my parents, Richard and Renée Parks, for their love and support over the years. You have been both patient and encouraging through it all. I am so grateful to Barbara Miller Hall for her support, encouragement, excitement, and voice of reason for over 13 years. Thank you, also, to all my friends in Atlanta and around the country who have no idea what I do, but support me anyway. You have all helped to keep me sane through the years. I am so blessed to have the friends and family that I do. Without the individuality of each of the people listed here, I would not have been able to complete this work, and hope you all realize the role you play in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	adenine
C	cytosine
cDNA	complementary DNA
cm	centimeter
DEPC	diethylpyrocarbonate
dH_2O	deionized water
DNA	deoxyribonucleic acid
DSC	differential scanning calorimetry
EDTA	ethylenediamine tetraacetic acid
G	guanine
H_2O_2	hydrogen peroxide
HPLC	high performance liquid chromatography
M	molar
Mg^{2+}	magnesium
min	minute
ml	milliliter
mm	millimeter
Na^+	sodium
nm	nanometer
NMR	nuclear magnetic resonance
NN	nearest-neighbor

OD_{260}	optical density at 260 nm
PCR	polymerase chain reaction
RNA	ribonucleic acid
RNase H	ribonuclease H
T	thymine
TBE	tris-borate EDTA electrophoresis buffer
TE	tris-EDTA buffer
TGGE	temperature gradient gel electrophoresis
T_m	melting temperature
U	uracil
UV	ultraviolet
°C	degrees Celsius
ΔCp	change in heat capacity
ΔG	Gibbs free energy
ΔH	enthalpy
ΔS	entropy
ε	extinction coefficient
2' OMe	2'-O-methyl modified nucleotide

SUMMARY

An accurate understanding of the thermodynamic parameters of nucleic acids is important for the design of experiments using a number of molecular biology techniques, understanding biological processes, and predicting secondary structures. The thermodynamic properties for completely duplexed nucleic acids are well characterized. Some structural elements, such as single-nucleotide bulges and single mismatches, have been examined in some detail. In this work, some less well-studied structural elements have been examined. These include tandem mismatches in both DNA and RNA, single base dangling ends in DNA, and chimeric molecules involving 2'-O-methyl modified nucleotides.

To evaluate tandem mismatch stability in DNA, small hairpin molecules were studied by optical denaturation and differential scanning calorimetry (DSC). A series of mismatch sequences covering a range of estimated stabilities were examined by UV melting in 1.0 M Na^+, 0.1 M Na^+, and 0.1 M Na^+ with 5 mM Mg^{2+}. DSC was used to evaluate the molecules in 0.1 M Na^+. From these results, the thermodynamic contribution of the mismatch sequences could be determined. A hierarchy of stability was established and compared to previous studies. The mismatches were found to have a range of free energies of about 5 kcal / mol, based on the specific sequence. This emphasizes that it is important to account for specific sequences, not simply structural motifs, in structure prediction and experimental design.

Tandem mismatches were also examined in RNA hairpins and in RNA duplexes. For the hairpins, molecules with a similar sequence to the DNA's were designed. The

thermodynamic contribution of the mismatches was evaluated and a hierarchy of stability determined. The range of free energies at 37 °C was about 1.0 kcal / mol. The mismatches examined in duplex molecules contained a U-U mismatch adjacent to another non-canonical base pair. Each of these mismatches was studied within the context of two different sets of closing base pairs to determine the effect of context on the thermodynamic properties. One interesting discovery was that one of the single strands was able to form a hairpin with a three base pair stem, two dangling bases, and a two base hairpin loop. This is a novel hairpin motif that has not been considered in sequence-dependent predictions of RNA secondary structure.

Single-base dangling ends were evaluated in DNA hairpins. Previous work in duplexes indicated that dangling ends could stabilize a molecule. In this work, the dangling ends were found to be slightly destabilizing to the hairpins. Chimeras were also generated with a DNA segment surrounded by 2'-O-methyl modified nucleotides. Similar sequences have been used in mutation correction studies. Duplexes of the chimeras and complementary DNA strands were examined to evaluate their thermodynamic properties and compare them to the corresponding DNA / DNA duplexes. The closing base pairs surrounding the DNA segment were found to be important to the stability of the molecule. For the chimeras with two 2' OMe A-U base pairs or a 2' OMe A-U and a 2' OMe G-C adjacent to the DNA, the chimera / DNA duplex was less stable than the DNA / DNA duplex. When two 2' OMe G-C's were used as the closing base pairs, however, the chimera / DNA duplex had similar stability to the DNA / DNA duplex.

Results reported here provide valuable thermodynamic data about specific structural elements in nucleic acids. This information can increase the accuracy of a number of types of molecular biology techniques and of structure prediction databases.

CHAPTER 1

INTRODUCTION

One area that has been of interest to a number of investigators recently is the stability of DNA and RNA structural elements. In this work, a review will be made of studies done in this area to date, followed by new information providing thermodynamic parameters for some less well-studied elements.

DNA and RNA structure

Deoxyribonucleic acid, or DNA, is the main carrier of genetic information in the cell. It is typically found in the form of a double helix resembling a spiral ladder. The building blocks of DNA are nucleotides, which consist of a phosphate, a five carbon sugar called deoxyribose, and one of four nitrogenous bases, adenine (A), cytosine (C), guanine (G), and thymine (T). The nucleotides link together through phosphodiester bonding, forming a polymeric molecule with sugar and phosphate elements forming the backbone structure, or upright edges of the ladder. The bases interact with one another across the helix through hydrogen bonding to form the rungs of the ladder. This interaction is referred to as base pairing.

Ribonucleic acid, or RNA, has multiple roles in the biochemical functions of the cell. There are several classes of RNA, which each serve a different purpose. While RNA is similar in structure to DNA, there are some fundamental differences. The sugar found in RNA is ribose, rather than deoxyribose. Thymine is not used in polymeric RNA

molecules. Instead, the base uracil (U) is present in RNA. Finally, RNA is generally found in a single stranded form, although it can adopt complex secondary structures containing double stranded regions and loops of unpaired bases, allowing for greater structural variations and functionalities. Double stranded RNAs are found in some viruses and duplex formation between different RNA molecules are involved in controlling DNA replication initiation in bacterial plasmids.

Watson-Crick base pairing

The classical hydrogen-bonding pattern between bases is known as Watson-Crick base pairing, after the scientists who initially elucidated the structure of DNA. Adenine and guanine are known as purines, while cytosine, thymine, and uracil are pyrimidines. In Watson-Crick base pairs, adenine pairs with thymine in DNA and with uracil in RNA. They interact through two hydrogen bonds. Cytosine pairs with guanine in both DNA and RNA. The interaction between these bases involves three hydrogen bonds. For this reason C-G pairs are expected to be more stable than A-T pairs. Regions of high G-C content within a polymeric molecule generally have increased stability.

For DNA and RNA molecules that are completely duplexed, meaning they contain all Watson-Crick base pairs, the stability may be estimated using an empirical nearest-neighbor (NN) model [1-8]. This model assumes that the stability of a given base pair is dependent on the identity and orientation of adjacent base pairs. It is possible to determine helix stability in this manner because stacking appears to be the major energetic contributor to helix formation. Models of nearest-neighbor interactions must include thermodynamic parameters for ten unique base pair dimers and helix initiation from single strands [1]. These parameters are dependent on salt concentration. The idea of applying a nearest-neighbor model to nucleic acid helix-coil transitions and

other properties was introduced in the 1960's with work done in the laboratories of Zimm and Tinoco [2-8]. Since then, a number of theoretical and experimental papers have been published evaluating the thermodynamic parameters of DNA and RNA in terms of nearest-neighbor interactions.

For DNA, NN interactions have been studied using a variety of DNA lengths and sequences. Wartell and Benight [9], Gotoh and Tagashira [10], Vologodskii *et al.* [11], McCampbell *et al.* [12], and Delcourt and Blake [13] utilized optical melting to study the helix-coil transitions of synthetic DNA polymers and DNA restriction fragments. Results of optical melting studies on short synthetic DNA's were published by Breslauer [14], SantaLucia [15, 16], and Sugimoto *et al.* [17]. Other investigations focused on long repeating DNA polymers using both optical and calorimetric experimental methods and theoretical analysis [14, 18, 19]. Work by Doktycz *et al.* examined NN parameters in a series of DNA dumbbells [20]. Although these studies provided a great deal of information about nearest-neighbor interactions, it was difficult to compare the results because of differences in ionic strength and molecular environments. In 1998, two independent investigations by Owczarzy *et al.* [21] and SantaLucia [22] compared the results from a number of previously published papers, and attempted to come up with methods for unifying the different NN thermodynamic parameters. Owczarzy's paper determined a salt-dependent correction for scaling the melting temperature of a DNA between 115 mM Na^+ and 1.0 M Na^+, taking into account the fraction of G-C base pairs in the sequence. Similarly, SantaLucia's paper provided a unified set of nearest-neighbor parameters for the ten unique base pair stacking interactions and for the initiation factor. Equations relating ΔG and ΔS measured in 1.0 M Na^+ to other sodium ion concentrations were also determined. ΔH was assumed to be independent of salt concentration. This work also gave a detailed analysis of the methods used in different

studies and compared the different sets of NN parameters after expressing them in the same format [10, 11, 13-15, 17, 20].

Initial studies on RNA stacking parameters were carried out by Borer *et al.* in the mid-1970's [5]. This group was able to determine NN parameters for six of the ten unique base stacks and for duplex initiation with and without a G-C base pair. By the mid-1980's, Freier *et al.* were able to determine all ten parameters, as well as the initiation factor for duplexes with at least one G-C base pair [23]. By the late 1990's, RNA synthesis and availability was greatly improved. Xia *et al.* examined a large series of RNA duplexes and reevaluated the NN parameters for RNA duplexes [24].

The nearest-neighbor model, while not exact, is a reasonable method for calculating the stability of nucleic acid oligomers with all Watson-Crick base pairs. Xia *et al.* provided data in which the thermodynamic parameters for the prediction of RNA duplex formation were derived [24]. They utilized 90 two-state sequences to derive nearest neighbor parameters. Ten additional sequences were found to not fit a two-state model and were excluded from the regression. If the nearest-neighbor model were exact, duplex oligonucleotides with identical base composition, nearest neighbors, and ends, but with difference sequences, should have thermodynamic parameters within experimental error. On average, it was found that the thermodynamic parameters predicted for RNA duplexes with identical nearest neighbors and ends gave ΔG_{37} values within approximately 6% of one another [24]. Similar results have been seen for DNA duplexes [25] and for DNA/RNA hybrids [26]. This demonstrates the limits of the model. Nearest-neighbor thermodynamic parameters for base pair stacking interactions in DNA and RNA are listed in Table 1.1.

Non-Watson-Crick base pairing

Non-Watson-Crick base pairs occur when bases other than guanine and cytosine

Table 1.1: Nearest neighbor thermodynamic parameters for Watson-Crick base pair formation in 1.0 M Na^+ [22, 24].

Sequence		RNA			DNA		
RNA	DNA	ΔH (kcal / mol)	ΔS (eu)	ΔG (kcal / mol)	ΔH (kcal / mol)	ΔS (eu)	ΔG (kcal / mol)
→ GC CG ←	→ GC CG ←	-14.88	-36.9	-3.42	-9.8	-24.4	-2.24
→ GG CC ←	→ GG CC ←	-13.39	-32.7	-3.26	-8.0	-19.9	-1.84
→ CG GC ←	→ CG GC ←	-10.64	-26.7	-2.36	-10.6	-27.2	-2.17
→ GA CU ←	→ GA CT ←	-12.44	-32.5	-2.35	-8.2	-22.2	-1.30
→ GU CA ←	→ GT CA ←	-11.40	-29.5	-2.24	-8.4	-22.4	-1.44
→ CA GU ←	→ CA GT ←	-10.44	-26.9	-2.11	-8.5	-22.7	-1.45
→ CU GA ←	→ CT GA ←	-10.48	-27.1	-2.08	-7.8	-21.0	-1.28
→ UA AU ←	→ TA AT ←	-7.69	-20.5	-1.33	-7.2	-21.3	-0.58
→ AU UA ←	→ AT TA ←	-9.38	-26.7	-1.10	-7.2	-20.4	-0.88
→ AA UU ←	→ AA TT ←	-6.82	-19.0	-0.93	-7.9	-22.2	-1.00
Initiation		3.61	-1.5	4.09	0.2	-5.6	1.96
Terminal AU or AT*		3.72	10.5	0.45	2.2	6.9	0.05

* For each terminal A-U or A-T. Differential treatment of G-C and A-T (A-U) ends accounts for the dependence of the total number of Watson-Crick hydrogen bonds on base composition of a duplex.

or adenine and thymine pair with each other. When surrounded by segments of Watson-Crick duplexes, these base pairs are called mismatches or loops. A large number of internal loops of non-Watson-Crick pairs are possible ranging in size from single bases to multiple mismatches.

One type of loop is called a bulge loop. In bulges one or more unpaired nucleotides are present on one strand of a double helix [1]. In general, this has a destabilizing effect on the helix. Work published on single base bulges in RNA and DNA has shown that both the identity of the base and the surrounding base pairs affect the free energy of the bulge [27-31]. In 2002, Znosko *et al.* published a detailed model for the formation of RNA single nucleotide bulges. This work took into account a number of characteristics including whether the bulge was a purine or pyrimidine, nearest-neighbor effects, non-nearest-neighbor effects, the length of the duplex, and the location of the bulge within the helix [31].

Internal loops occur when one or more non-Watson-Crick base pairs occur surrounded by double helical stems. Thermodynamic parameters for single mismatches in DNA were determined in a series of investigations by SantaLucia's laboratory. These papers studied DNA oligomers with the single mismatches: G-A, A-C, C-T, G-T, A-A, C-C, G-G, and T-T [16, 32-35]. The authors examined these mismatches in different positions within a duplex, with one or two single mismatches present, and with different helix lengths. Benight's laboratory published work that examined the effects of sequence context on the stability of single mismatches in DNA [36].

Studies on RNA single mismatches have been carried out by Kierzek *et al.*, in which single mismatches were examined by optical melting and NMR [37], Meroueh and Chow, who used optical melting and circular dichroism to study mismatches in hairpins [38], and Zhu and Wartell, where all single mismatches were studied in several NN contexts of long RNA's using temperature gradient gel electrophoresis [39].

Thermodynamic parameters for single mismatch sequences may be found in these papers, although not all possible sequences and contexts are represented. In general, single mismatches are destabilizing. The thermodynamic parameters for a mismatch depend on both the sequence and the closing base pairs.

A few studies have been made on larger internal loops. Allawi and SantaLucia's work on G-T single mismatches included results for three molecules with tandem G-T mismatches [16]. Ke and Wartell examined tandem mismatches using temperature gradient gel electrophoresis. This method allowed them to estimate free energies of the molecules and establish a hierarchy of stability for the tandem mismatches examined [40]. Zuker's *mfold* web server [41], which predicts RNA and DNA structure based on energy minimization, utilizes free energy parameters for tandem mismatches in 1.0 M Na^+ from SantaLucia's laboratory, cited as unpublished data.

One DNA internal loop that has been studied in considerable detail contains tandem G-A mismatches. This particular sequence has been shown to have an unexpectedly high stability. Li *et al.* first discovered that the sequence d(CGAG)-d(CGAG) in a DNA duplex oligomer was more stable than expected [42]. Systematic changes were made in the sequences of the closing base pairs and thermal denaturation was used to determine thermodynamic parameters. Ebel *et al.* studied similar molecules with tandem G-A mismatches in different contexts [43]. Li and Agrawal followed this with a paper that examined tandem G-A mismatches in very short (6-mer) DNA duplexes. All possible closing base pairs were examined. Both UV melting and NMR were used to study these molecules [44].

More work has been done on internal loops in RNA. Zhu estimated free energy values for a large set of tandem mismatches in RNA using temperature gradient gel electophoresis [45]. SantaLucia *et al.* investigated a series of 12 duplexes containing tandem mismatches. These molecules were examined by optical melting, circular

dichroism, and NMR [46]. Walter *et al.* used UV melting and NMR to examine the effect of the closing base pairs on tandem G-A mismatches in RNA [47]. While this type of tandem mismatch was seen to provide unusual stability to DNA molecules, it has less of an effect in RNA. As with DNA, the closing base pairs adjacent to the G-A tandem mismatches affects the stability of the molecule. Wu *et al.* published data on the thermodynamics of a set of 18 tandem mismatches [48]. They examined the effects of closing base pairs on stability using identical mismatches. Other investigators have studied tandem G-U mismatches and internal loops containing a G-G mismatch next to another non-canonical base pair [49, 50]. Asymmetric internal loops, where the number of bases on each side of the loop differs, have also been examined [51, 52].

Applications of nucleic acid thermodynamics

An accurate thermodynamic description of DNA and RNA structural elements is needed for a number of applications. *In vivo*, mismatches may occur naturally from errors during replication [53] or genetic recombination [54], or as the result of mutagenic chemicals [54, 55] or ionizing radiation [56]. Understanding the thermodynamics of these mismatches may help in elucidating information about mismatch repair and polymerase fidelity [35]. Additionally, this information on stability is useful for a number of molecular biology techniques. Polymerase chain reaction (PCR) [35, 57], Southern blotting [58], and DNA sequencing [59] depend on the hybridization of a short DNA primer to a template sequence. Antisense gene targeting involves hybridizing short DNA's to an RNA target [60]. Gene chip technologies involve a reaction between cDNA's and synthetic DNA targets [61]. For each technique it is necessary to optimize sequence, temperature, and ionic strength conditions in order to avoid incorrect amplification or detection of the wrong target. Determining the conditions needed to obtain desired DNA hybridizations while avoiding undesirable complexes requires

accurate information on the thermodynamic parameters of specific sequences and the structural elements they form.

Effects of salt on nucleic acid helix-coil transitions

The canonical ionic condition for thermal denaturation studies is 1.0 M Na^+. This condition was selected to enable studies on short duplex oligomers (<20 base pairs) without electrostatic end effects dominating the transition. It has been shown that there is a relationship between the solvent's salt concentration and the melting temperature of a DNA or RNA duplex. As ionic strength increases, melting temperature increases. Several empirical equations have been determined that relate DNA length, % GC content, and monovalent ion concentration to a DNA's melting temperature. One representative equation for DNA developed by Bolton and McCarthy [62] and later modified by Baldino *et al.* [63] is:

$$T_m = 81.5°C + 16.6°C \times \log_{10}[Na^+] + 0.41°C \times (\%GC) - 675/N, \qquad (1.1)$$

where $[Na^+]$ is the sodium concentration in molarity, (%GC) is the G-C content of the molecule, and N is the number of bases. Equation 1.1 is an approximation and does not take sequence effects into account.

Thermodynamic properties as well as T_m values are affected by sodium concentration. The effect of ion concentration on ΔG, ΔH, and ΔS values in DNA has been considered in reference [22]. As salt concentration increases, the free energy of a duplex DNA decreases due to the added stability from shielding repulsive electrostatic interactions and entropy effects. It was assumed that this change is not sequence dependent, but is length dependent. The change in free energy of a DNA oligomer duplex as it relates to salt concentration is expressed by the equation [22]:

$$\Delta G°_{37}(\text{oligomer}, [Na^+]) = \Delta G°_{37}(\text{unified oligomer, 1M NaCl}) - 0.114 \times N \times \ln[Na^+], \quad (1.2)$$

where $[Na^+]$ is the salt concentration in molarity and N is the number of phosphates in the duplex divided by 2. For example, for an eight base duplex without terminal phosphates, N is equal to 7.

Entropy correction is presented in a similar manner. As salt concentration increases, the net entropy change decreases. Again, the assumption is made that the change in entropy is sequence independent, but is length dependent. An equation expressing the entropy change in a DNA as it relates to salt concentration is:

$$\Delta S°(\text{oligomer}, [Na^+]) = \Delta S°(\text{unified oligomer, 1M NaCl}) + 0.368 \times N \times \ln[Na^+], \quad (1.3)$$

where $[Na^+]$ is the salt concentration in molarity and N is the number of phosphates in the duplex divided by 2. In general, enthalpy (ΔH) is assumed to be salt independent, although this may not be the case [64].

Strand association

Duplexed DNA's and RNA's have been commonly used to evaluate the sequence-dependent thermodynamic parameters of the NN model. One feature of duplex formation that is not well understood is the association of the two strands. This involves cooperative intermolecular association of two single strands to form the double helix through base pairing, interstrand stacking, and changes in exposed polar and nonpolar surfaces. Although thermodynamic values have been assigned to the helix

initiation event for DNA's with either terminal G-C or A-U base pairs by a variety of groups [14, 15, 17, 22], these values are somewhat inconsistent with one another, suggesting a complex process.

Holbrook *et al.* have studied this process in more detail using a single 14-mer DNA duplex [65]. This group recognized that the association event may be described as a process of strand docking if the single strands form ordered, stacked helices prior to the strand interactions. Alternately, the two processes of the single strands becoming ordered then associating to form the helix may be coupled. The group decomposed these two steps and calculated the overall enthalpy of forming the duplex in terms of docking, forming ordered single strands from unstructured strands, and intramolecular unfolding of hairpin structures in single strands. This analysis points out that the initial association of single strands can depend on secondary structures. Assuming a single association parameter may not always be appropriate for the formation of duplexes from different complementary pairs of single strands. The group found that the enthalpy of duplex formation changed slightly over a temperature range. More work is necessary to determine if these results are typical for all duplexes.

The nearest neighbor model assumes the initial association event is entropic in nature. If this is assumed, i.e. the enthalpy change is very small relative to the entropy change of initial strand association, then the equilibrium constant for the duplex formation is temperature independent. A temperature independent constant is used for duplex oligomers. Three values are used, one for duplexes with two G-C terminal pairs, one for duplexes with two A-T terminal pairs, and one for duplexes with one G-C and one A-T terminal pair. This formulation does not take into account potential sequence-dependent structures of single strands, which are likely to be related to temperature.

Nucleic acid structural motifs: The hairpin loop

Hairpin DNA molecules were employed for some experiments in this work to avoid the uncertainty in the appropriate value of the association factor and its influence on evaluated stacking interactions. Hairpin DNA's occur when a single-stranded nucleic acid folds back on itself to create a Watson-Crick base paired stem with a region of unpaired bases at one end. More thermodynamic data is known about hairpin loops than other types of loops. The stability of a loop is determined by loop size, sequence, and closing base pair [1].

Two investigations in the 1980's examined DNA hairpin loops with potentially zero to seven thymidines using NMR, temperature-jump, and UV melting experiments [66, 67]. Thermodynamic parameters were determined and it was found that the stability of hairpin DNA molecules reached a maximum with four to five nucleotides in the loop. The group also found that hairpin loops could be formed with a lower number of thymidines in the loop, although these hairpins were often in equilibrium with the corresponding duplex molecule containing an internal loop.

A number of studies have been carried out on loops with four nucleotides, called tetraloops, which are the most commonly occurring hairpin loops in ribosomal RNA's. A variety of sequences and closing base pairs for the hairpin loops were characterized [68-74]. Some tetraloop sequences have been shown to be exceptionally stable due to non-Watson-Crick interactions across the loop. For example, UNCG loops, where N represents any nucleotide, have been shown to be highly stable in RNA. The U and G create a wobble base pair with interactions across the loop. The UUCG loop has a melting temperature approximately 20 °C higher than a hairpin with normal thermodynamic stability [72]. A similar interaction occurs for GNRA loops in RNA, where N is any nucleotide, and R is a purine [75]. A GAAA hairpin, for example, has a melting temperature about 14 °C higher than a hairpin with normal stability. In DNA, a TTCG

was not shown to be highly stable. GNRA, GATA, and CTTG loops were found to be among the most stable tetraloops in DNA [75]. GAAA loops were tested with all possible combinations of two base pair stems. These hairpins had T_m's 30 °C to 40 °C higher than loops with identical stems and hairpin loops consisting of TTTT, GCCC, or GTTT [76].

While tetraloops are the best characterized hairpin loops, and it has long been known that larger hairpins can form, recent papers have shown smaller loops can form stable stem-loop structures in both RNA and DNA. A GNA trinucleotide loop can form a stable stem-loop structure, even with a stem containing only two Watson-Crick base pairs [77]. In this work, the authors examined the 64 possible sequences of d(GCNNNGC) where NNN represents a triloop. They found that the four fragments with a GNA loop formed highly stable minihairpins. These four hairpins had faster gel mobility than a single-stranded fragment of the same length in a 7 M urea gel and also demonstrated resistance to a single-strand exonuclease. Evidence has also been presented indicating that two base DNA loops are formed *in vivo* [78]. These structural motifs were characterized using NMR [79]. An evaluation of crystal structures of ribosomal RNA's has revealed three base hairpin loops with only a single base pair stem [80]. TGGE has also been used to study thermodynamically stable and unstable RNA triloops [81]. In these electrophoresis experiments, those triloops considered to be stable had melting temperatures approximately 5 °C higher than the loops of normal thermodynamic stability. The loops considered to be unstable had T_m's approximately 1.5 °C lower than loops of normal stability. A paper by Jucker and Pardi has shown an RNA tetraloop with the sequence CUUG has Watson-Crick interactions between the C and G, suggesting this tetraloop is actually present as a biloop [82].

Nucleic acid structural motifs: Dangling ends

In addition to loops, another form of structural element is a dangling end. This occurs when a duplexed sequence has one extra base on the 5' or 3' end. Knowledge of this type of motif is important to accurately predict the stability of DNA or RNA hybridization and secondary structure. Limited knowledge is available about the effects of dangling ends, although dangling ends appear to stabilize DNA duplexes [83-86].

As microarray technology becomes more widely used, one area of particular interest will be dangling end hairpins. One application of this technique involves the use of oligonucleotides bound to a solid support to allow for effective hybridization reactions. In order to attach oligonucleotides to a solid support, a number of factors should be considered. The first of these involves the effective and practical means to carry out the attachment. The surface should not interfere with hybridization. The length of a spacer used to attach the probes to the surface also has a significant effect on hybridization efficiencies.

Dangling end hairpins offer the possibility of being used as a target. The hairpins offer an alternative method for attaching the DNA to the surface via the dangling ends. Also, it has been shown that single strands on the end of duplexes are more stabilizing than a simple single strand in hybridization reactions. Finally, it has been demonstrated that hairpins hybridize to linear probes more efficiently than a reaction between two single strands [87]. Due to these characteristics, it is of interest to examine the effects of both 5' and 3' dangling end hairpins.

Nucleic acid modifications: 2'-O-methyl nucleotides

2'-O-methyl RNA is a nucleic acid analogue that occurs naturally in ribosomal RNA's and other small functional RNA's [88, 89]. RNA's of this nature have generated interest in recent years, however, because 2'-O-methyl RNA's have been shown to

hybridize to RNA targets more quickly and with higher melting temperatures [90]. Additionally, 2'-O-methyl RNA is resistant to ribonuclease cleavage. One investigation demonstrated the ability to use *E. coli* ribonuclease H (RNase H) to cleave RNA in a site-specific manner by generating a chimeric 2'-O-methyl RNA and DNA probe [91]. This molecule consisted of nine total bases where three to five bases in the interior of the duplex were deoxynucleotides and the surrounding bases were 2'-O-methyl bases. When this molecule was hybridized to an RNA target, the RNA was cleaved at a specific site.

Chimeric molecules may also be useful in mutation correction. Liu *et al.* have shown that a double-stranded, RNA/DNA hybrid molecule folded into a hairpin may direct a cell's DNA repair machinery to repair mutations [92]. In these chimeras, the DNA section was found to direct the repair while the RNA regions provided complex stability. The authors also employed molecules with 2'-O-methyl modified nucleotides in place of the unmodified RNA regions for some studies. These chimeras were utilized in mammalian and plant cell-free extracts. Because extracts such as these contain nuclease activity, simple RNA/DNA chimeras were susceptible to degradation prior to mutation repair. By using modified 2'-O-methyl nucleotides, the degradation could be avoided, however when large numbers of 2'-O-methyl residues surrounded the DNA region, the chimera was inactive in gene repair. As long as the number of 2'-O-methyl residues was kept small, the chimeras were able to maintain some gene repair activity while also being resistant to nuclease degradation. The thermodynamic properties of the chimeras may play a role in the effectiveness of hybridization to DNA, and thus in mutation correction, making it important to determine these values.

Objectives

In order to accurately predict a nucleic acid's secondary structure from its sequence, to appropriately design molecular biology hybridization experiments, and to understand a number of biological processes, it is necessary to have an accurate understanding of the thermodynamic characteristics of nucleic acid structures. The goals of this thesis are to determine the stability of particular DNA and RNA structural elements. First, DNA tandem mismatches were examined in the context of hairpin molecules. This allowed comparison to other published results using several techniques, and new information will be presented about sequences not previously studied. RNA tandem mismatches were examined in two types of molecules. Hairpins containing tandem mismatches with the same sequences as the DNA mismatches were examined. Additionally, internal loops containing a U-U mismatch adjacent to another non-canonical pair were studied in duplexed RNA molecules. Thermodynamic parameters were determined for each of these mismatches within the context of two different sets of closing base pairs. Finally, hairpins with both 3' and 5' single base dangling ends were examined, as were chimeric 2'-O-methyl RNA's. UV-monitored denaturation and differential scanning calorimetry were used to monitor the helix-coil transitions and several types of data analysis were carried out, in order to obtain the most meaningful data possible. The results presented here add to the current knowledge of thermodynamics in nucleic acid structural elements.

CHAPTER 2

MATERIALS AND METHODS

Nucleic acid preparation

DNA molecules were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA) at an appropriate scale. RNA molecules were synthesized by Dharmacon, Inc. (Lafayette, CO). Molecules examined only by UV denaturation were generally used unpurified. DNA's and RNA's studied using differential scanning calorimetry were purified by high performance liquid chromatography (HPLC).

High Performance Liquid Chromatography

Purification was done on an Agilent 1100 series machine with a manual injector, degasser, quaternary pump, and variable wavelength detector. The column used was a Dionex DNAPac PA-100 anion exchange column. All buffers were filtered before initial use with a 0.2 μm bottle-top filter. The system was set up such that buffer A was HPLC-grade water, buffer B was 0.25 M Tris-HCl, pH 8.0, and buffer C was 0.375 M sodium perchlorate. Buffer D was a storage buffer consisting of 80% water and 20% methanol. Prior to beginning the purification, each buffer was purged by pumping at 5 ml / min with the pump valve open to allow for the removal of any dissolved gases in the buffers. Water (100% buffer A) was then run through the column at a rate of 1.5 ml / min with the pump valve closed for 5 minutes to ensure the removal of storage buffer. Finally, the

conditions were set to 88% buffer A, 10% buffer B, and 2% buffer C for 10 minutes at a rate of 1.5 ml / min to prepare the column for injection.

DNA was prepared for injection by resuspending the lyophilized synthesis product in dH_2O. Because injections were 100 µl each, a multiple of this value was chosen for the volume of dH_2O added. In general, the final DNA concentration was 150 to 250 OD_{260} / ml, or 15 to 25 OD_{260} of DNA per 100 µl injection.

RNA was deprotected according to the manufacturer's protocol prior to purification. In short, each tube was centrifuged briefly, and 400 µl of the provided 2'-deprotection buffer (100 mM acetic acid, adjusted to pH 3.8 with TEMED) added. Samples were pipetted up and down to completely dissolve the pellet and then vortexed for ten seconds, followed by centrifugation for ten seconds. Tubes were then placed in the Speedvac with the heat off and spun until dry. Following this centrifugation, the dry product was resuspended with DEPC-treated water and the tubes combined to give a final concentration of 150 to 250 OD_{260}/ml. Again, the volume of water in which the sample was resuspended was chosen to be a multiple of 100 µl.

Injection was performed once the column had been prepared with the 88:10:2 mixture of buffers A, B, and C. The pump was set with the conditions found in Table 2.1. The cycle time was set to 36 minutes to allow for column regeneration and reequilibration in preparation for additional injections [93]. The detector was set at 260 nm and peaks were collected manually. Once all injections were completed, the column was washed with 100% buffer A for 5 minutes to remove salt and then stored by running 100% buffer D through at 1.5 ml / min for 10 minutes.

Desalting

Once purification was complete, desalting was necessary. Due to the sensitivity of the differential scanning calorimeter to both salt and contaminants, two forms of

Table 2.1. HPLC program for separation of nucleic acids with 8 to 30 bases.

Time (min)	Flow rate (mL/min)	Buffer A (%)	Buffer B (%)	Buffer C (%)
0.0	1.5	88	10	2
20.0	1.5	52	10	38
21.5	1.5	0	0	100
25.5	1.5	0	0	100
25.6	1.5	88	10	2

desalting were necessary. The first desalting method used a reverse phase desalting cartridge (Sep-Pak C-18, Catalog #020515, Waters, Inc.) It was necessary to equilibrate the cartridge prior to use. The plunger was removed from a 10 ml syringe and the SepPak cartridge attached. 10 ml of acetonitrile was added to the syringe and the plunger replaced. The acetonitrile was passed through the cartridge at a rate of 2 to 3 drops per second. The cartridge was then removed, followed by the syringe plunger, and the cartridge returned to the syringe between each flush with solution. 10 ml of a 50% acetonitrile: 50% 0.3M sodium acetate, pH 5.2 solution was added to the syringe, the plunger replaced, and the liquid passed through at the same rate as before. Next, 10 ml of 0.3 M sodium acetate was used to equilibrate the cartridge.

The volume of nucleic acid solution collected from the HPLC was estimated and 0.1 volumes of 3M sodium acetate, pH 5.2 added to the tube. This was vortexed to mix well, then added to the syringe used for desalting. This solution was passed through at a rate of 1 to 2 drops per second to allow time for the nucleic acid to bind to the C18 packing, and recollected. To ensure that as little nucleic acid as possible would be lost, the solution was again added to the syringe and run over the SepPak cartridge. The cartridge was then washed with 5 ml of 50 mM triethylammonium bicarbonate (TEAB) at a rate of 2 to 3 drops per second. Finally the DNA or RNA was eluted with 5ml of a solution of 35% methanol, 35% acetonitrile, and 30% 50 mM TEAB. This solution was collected into four approximately equal fractions and then spun to dryness in the SpeedVac with the heat off.

Although SepPak cartridges remove the majority of the salt present, it was necessary to perform an additional desalting step on the nucleic acid molecules used for differential scanning calorimetry. A column with a 1.7 cm diameter was packed with Sephadex G-10 to a height of 11 cm. The four tubes containing the nucleic acid were resuspended in a total volume of 300 µl of dH_2O and carefully added to the Sephadex

column. Because Sephadex is a size exclusion method of separation, an additional 12 milliliters of dH_2O was used to elute the DNA or RNA from the column. One-milliliter fractions were collected and UV readings taken at 260 nm to determine those fractions which contained nucleic acid. These fractions were then spun to dryness in a SpeedVac and resuspended in 1X tris ethylene diaminetetraacetic acid (TE) and combined. In the case of RNA samples, the TE was made using DEPC-treated dH_2O.

Concentration determination

Concentration was determined by one of two methods. The first of these utilized the OD_{260} taken at 95 °C and nearest neighbor extinction coefficients. Absorbance and extinction coefficient are linked to concentration by the following equation:

$$C = A/\varepsilon \tag{2.1}$$

where C is concentration in molarity, A is the absorbance at 260 nm, and ε is the extinction coefficient. The extinction coefficient of a nucleic acid oligomer may be determined by the following equation:

$$\varepsilon(N_1N_2N_3N_4\ldots N_i) = 2\,[\varepsilon\,(N_1N_2) + \varepsilon\,(N_2N_3) + \varepsilon\,(N_3N_4) + \ldots\,\varepsilon\,(N_{i-1}N_i)] - [\varepsilon\,(N_2) + \varepsilon\,(N_3) + \varepsilon\,(N_4) + \ldots + \varepsilon\,(N_{i-1})] \tag{2.2}$$

Values for the extinction coefficients of the dinucleoside phosphates and nucleotides may be found in table 2.2. This method is a somewhat less accurate method for concentration determination and was therefore used for molecules that were examined by UV denaturation only.

Table 2.2. Extinction coefficients per nucleotide (M^{-1} cm^{-1} X 10^{-3}) at 260 nm. 25°C, 0.1M ionic strength, pH 7.

	RNA	DNA		RNA	DNA
Ap	15.34	15.34	CpG	9.39	9.39
Cp	7.60	7.60	CpU (CpT)	8.37	7.66
Gp	12.16	12.16	GpA	12.92	12.92
Up (Tp)	10.21	8.70	GpC	9.19	9.19
ApA	13.65	13.65	GpG	11.43	11.43
ApC	10.67	10.67	GpU (GpT)	10.96	10.22
ApG	12.79	12.79	UpA (TpA)	12.52	11.78
ApU (ApT)	12.14	11.42	UpC (TpC)	8.90	8.15
CpA	10.67	10.67	UpG (TpG)	10.40	9.70
CpC	7.52	7.52	UpU (TpT)	10.11	8.61

The extinction coefficients are estimated to be accurate to ±0.10 for the monomers and ±4% for the dimers.

The second method for determining nucleic acid concentration utilized the enzyme *Croatalus adamanetus* venom phosphodiesterase I (Worthington Biochemical Corporation or USB, Incorporated). Because accurate concentration of the nucleic acid is necessary for evaluating enthalpy change by differential scanning calorimetry, this method of concentration determination was used for experiments in the DSC. A stock phosphodiesterase solution was prepared by diluting the enzyme to a 1 mg / ml concentration (approximately 31 units / ml) in a 1X buffer consisting of 110mM NaCl, 100mM Tris-HCl, pH 8.9, and 15mM $MgCl_2$. DNA's or RNA's were diluted to approximately 0.3 OD_{260} / ml utilizing a 2X stock of the above buffer and deionized water to achieve the final 1X buffer concentration. Phosphodiesterase was added to a final concentration of approximately 0.5 units / ml and allowed to incubate at room temperature until an absorbance plateau was achieved. Incubation was generally extended to four hours. Following incubation, an OD_{260} was determined using appropriate reference solutions. The concentration was found using the absorbance value of the digested DNA or RNA and an extinction coefficient determined by the sum of single nucleotide extinction coefficients.

$$\varepsilon\,(N_1N_2N_3\ldots N_i) = \varepsilon\,(N_1) + \varepsilon\,(N_2) + \varepsilon\,(N_3) + \ldots + \varepsilon\,(N_i) \qquad (2.3)$$

Polyacrylamide gel electrophoresis

Non-denaturing polyacrylamide gel electrophoresis was carried out on a vertical gel apparatus. A gel with the appropriate polyacrylamide (37.5:1 acrylamide: bisacrylamide ratio) concentration, as described in later chapters, was prepared using 0.5X TBE, pH 8.3 (0.045 M Tris, 0.045 M boric acid, 1 mM EDTA) as the buffer. The gel dimensions were 16.5 cm X 20 cm. Samples were loaded with a 6X dye containing 15% ficoll, 0.025% bromophenol blue, and 0.025% xylene cyanol. Gels were run at 10 volts /

cm until the bromophenol blue was within 3 cm of the end of the gel. The gel was then stained using a solution of 2 μg / ml ethidium bromide for ten minutes. It was briefly rinsed in deionized water and visualized using an Alpha Innotech AlphaImager 2200.

Denaturing gels were prepared in a similar manner. In addition to the polyacrylamide and 0.5 X TBE, the gel contained 7 M urea as a denaturant. Nucleic acid samples were heated to 90 °C for one minute and then plunged into an ice bath prior to loading. The gel was run in a similar manner to the non-denaturing gels. Because single-stranded nucleic acids are more difficult to visualize when staining with ethidium bromide, gels were stained for 30 minutes, then rinsed briefly prior to visualization.

UV denaturation

UV denaturation was employed to monitor the melting transition of the nucleic acids. These experiments were carried out on a Cary 1E spectrophotometer (Varian, Inc.) with the Windows software package and a Peltier heating system. The absorbance of samples was measured at 268 nm as a function of temperature in a masked quartz cuvette with a 1 cm pathlength. In some instances, absorbance was also monitored at 280 nm. Temperature was measured with a platinum resistance probe inserted into a solvent cell adjacent to the sample. Unless otherwise noted, samples were diluted to approximately 0.45 OD_{260} / ml using dH_2O and a concentrated stock of the buffer solution. DNA samples were covered with mineral oil to reduce evaporation at high temperatures. This was not done for RNA samples because in general they were not heated to temperatures as high as the DNA samples and there was no reliable way to ensure the oil to be ribonuclease-free. Details on buffer concentration are given in later chapters for specific experiments, but in short, buffers used were 1.0 M Na^+, 0.1 M Na^+, or 0.1 M Na^+ and 5 mM Mg^{2+}. For experiments that required a higher DNA concentration, a 2 mm pathlength cuvette was employed.

Absorbance readings were taken every 0.1 °C over the temperature range. Heating rates were 0.5 °C / minute or 1.0 °C / minute, described below for each set of experiments. Specific temperature ranges and are also given in later descriptions of experiments. For some molecules, the melting transition occurred at low temperatures, making it necessary to begin the experiment at temperatures lower than 15 °C. At these low temperatures, high humidity in the laboratory caused condensation to form on the cuvettes making accurate readings difficult. For the experiments where it was necessary to begin at a temperature lower than 15 °C, the spectrophotometer and all accessories were placed in a 4 °C cold room with a constant humidity of 65%.

Between experiments, cuvettes were cleaned using 95% ethanol and a cotton swab to remove any particles adhering to the inside walls. They were then rinsed thoroughly with dH_2O. For experiments using RNA samples, all buffers were made with DEPC-treated water. In addition to the ethanol, cuvettes used for RNA experiments were soaked in 3% hydrogen peroxide for ten minutes and rinsed thoroughly prior to use.

Sedimentation equilibrium

Sedimentation equilibrium experiments were used to determine the molecularity of the ga/ga hairpin. This method was based on the approach described by Bothwell *et al.* [94] and later modified by Dripps and Wartell [95]. These experiments were performed with an Optima TLX preparative ultracentrifuge (Beckman Coulter, Palo Alto, CA). Known single-stranded DNA standards of 19, 26, and 33 nucleotides in length were prepared in a 200 µl volume at approximately 10 µM. The solvent was 0.08 M NaCl, 0.01 M Na_2HPO_4 with 5 mg / ml ficoll. The DNA concentration employed should assure the molecules are monomers rather than aggregates that could be present at higher concentrations. Each standard was divided into two 100 µl aliquots in 30 mm X 5

mm straight-walled tubes. One tube of each standard was set aside for 24 hours. The other tube was centrifuged for 24 hours at 18,000 rpm using the TLS-55 rotor and straight-wall tube adaptors (Beckman Coulter catalog # 358614).

Following centrifugation, the tubes were carefully removed from the centrifuge and the top 40 µl of each tube were transferred to a new tube. The top 40 µl of the tubes that had been set aside were also transferred to new tubes. Each sample was then diluted 1:2 with water. The absorbance of these samples at 260 nm was measured in a Hewlett Packard 8453 UV spectrophotometer using a 1 cm path length, 70 µl volume capacity masked quartz cuvette. The fraction of molecules (F) remaining in the top 40 µl of each centrifuged sample was compared to the absorbance reading of the non-centrifuged samples. A plot was generated of log(F) versus length in base pairs and a linear regression used to fit a straight line.

For the ga/ga hairpin, the sample was prepared at a concentration of 50 µM, similar to the concentration in initial calorimetry experiments, in 0.08 M NaCl, 0.01 M Na_2HPO_4 with 5 mg / ml ficoll. The samples were treated as described above, except that the 40 µl taken after centrifugation was diluted 1:10 due to the higher concentration of the original sample. Again the fraction remaining in the top 40 µl was calculated and the log(F) ascertained. From this value and the standard curve, an apparent molecular weight of the sample could be determined [94, 95].

Differential scanning calorimetry

Differential scanning calorimetry allows for direct measurement of thermodynamic parameters. It was found that a 100 µM concentration of hairpin or duplex DNA or RNA provided a strong, reproducible transition curve using a minimal amount of nucleic acid. An effort was made to conduct DSC experiments in 1.0 M sodium, but high melting temperatures made it difficult to define the end of the transition,

making evaluation of the enthalpy change challenging and uncertain. For that reason, DSC experiments utilized a 0.1 M sodium buffer comprised of 78 mM NaCl, 10 mM Na_2HPO_4, and 1 mM Na_2-EDTA. DNA samples were diluted approximately 1:10 using dH_2O and a 10X concentration of the sodium buffer. For RNA experiments, all buffers and dilutions were made using DEPC-treated dH_2O.

A minimum of four buffer versus buffer scans, where both the reference and sample cells contained melting buffer only, were run prior to the acquisition of sample data. These scans used the same temperature ranges and heating rates as the sample runs. Following these scans, the buffer was removed by aspiration and buffer again loaded into the reference cell, but DNA or RNA solution loaded into the sample cell. At least four sets of data were collected for each sample at a constant pressure of three atmospheres to prevent gas formation at high temperatures. The heating rate used was 1 °C / minute. Temperature ranges will be given with the descriptions of specific experiments in later chapters. Each run included a ten-minute equilibration at the initial temperature prior to the start of heating.

In initial experiments, data from the first heating was discarded due to irreproducibility. It was later found that stringent cleaning of the instrument and degassing of the samples for at least 30 minutes could help alleviate this problem. Cleaning the cells employed 600 µl of 50% formic acid, which was loaded into both the sample and reference cell. The machine was heated from 25 °C to 80 °C one time. When the instrument cells had returned to room temperature, they were rinsed with 500 ml dH_2O. For RNA experiments, 2 ml of hydrogen peroxide (H_2O_2), which inactivates ribonucleases, were loaded into the cells and the pipette tips used to load the solution left attached to the cell stems to hold the excess. The H_2O_2 was allowed to incubate for 10 minutes. Two to three times during the incubation period the pipette was reattached to the tips and the H_2O_2 flushed back and forth to remove any gas that built up in the cell.

The cell was again flushed with 500ml dH_2O. The most reproducible results were seen when the experiment was carried out in the order of cleaning, buffer versus buffer scans, sample scans, and then repeating the process starting with the cleaning procedure. Another important procedure for obtaining reproducible results was degassing the sample to help prevent air bubbles from forming at high temperatures. Although the instrument manufacturer recommends degassing samples for five to ten minutes prior to loading, it was found that 30 minutes was necessary to obtain consistent data.

Calculation of thermodynamic parameters from UV denaturation curve

Absorbance versus temperature data obtained from the UV denaturation curves was graphed using SigmaPlot software. The pre- and post-transition baselines were drawn manually and the slope of each line determined. The hyperchromicity of the transition was found by dividing the last data point by the first. Values for the hyperchromicity at 268 nm fell in the range 1.15 to 1.25. The raw data was then smoothed using a 13-point Lowess smooth using SigmaPlot. In the SigmaPlot transform for Lowess, the user must input a value for the term *f*, which represents the fraction of total points used in the regression and therefore the amount of smoothing. To determine the value of *f* for this smoothing, the number 13 is divided by the total number of raw data points.

Absorbance versus temperature data was then expressed in terms of the fraction of molecules in the single stranded coil state, θ_B, using the following equation [9]:

$$\theta_B(T) = [A(T) - A_{pre}(T)]/[A_{post}(T) - A_{pre}(T)] \qquad (2.4)$$

A(T) is the absorbance of the sample at temperature T and $A_{pre}(T)$ and $A_{post}(T)$ are the pre-transition and post-transition linear baselines of the denaturation curves. The

derivative melting curve was determined by dividing the increments $\Delta\theta_B(T)$ by the temperature increments, ΔT. An 11-point Lowess smooth was performed on the derivative curve, where *f* was determined by dividing the total number of data points into 11.

Assuming that the helix-coil transitions are two-state, a van't Hoff analysis was used to evaluate the enthalpy and entropy of the hairpin molecules at the T_m using the equations [96]:

$$\Delta H^\circ = (2+2n)\, RT_m^2\, (d\theta_B(T) / dT)_{T=Tm} \tag{2.5}$$

$$\Delta S^\circ = \Delta H^\circ / T_m \tag{2.6}$$

ΔH° and ΔS° are the enthalpy and entropy changes of the transition, n is the molecularity of the transition, R is the gas constant (1.987 cal-K^{-1}-mol^{-1}), and $(d\theta_B(T)/dT)_{T=Tm}$ is the height of the derivative melting curve at the T_m. Free energies may then be determined from the equation:

$$\Delta G^\circ = \Delta H^\circ - T\, \Delta S^\circ \tag{2.7}$$

Where ΔG° represents the Gibbs free energy and T is temperature. This equation does not take into account any changes in heat capacity because it assumes ΔH° is temperature-independent. All ΔG° values reported in this work are at 37 °C.

A second van't Hoff method was also used to determine ΔH° and ΔS° in the region of the transition. In this approach, a non-linear least squares regression was employed to evaluate ΔH° and ΔS° pairs that best fit the shape of the DNA melting curves. For the hairpin nucleic acids, assuming a two-state monomolecular transition,

the fraction of strands in the coiled state, $\theta_B(T)$, can be related to the enthalpy and entropy change through the equation [96].

$$-\Delta H^\circ + T\Delta S^\circ = RT \ln [(\theta_B(T)/(1- \theta_B(T))] \quad (2.8)$$

For each transition curve the ΔH° and ΔS° that best fit equation 2.8 were determined. The average values for each DNA were generally very similar to those evaluated from equations 2.5 and 2.6.

The change in heat capacity between duplex and single stranded states was calculated for several RNA duplex molecules. For these experiments, UV denaturation experiments were conducted at varying strand concentrations and the thermodynamic parameters determined as described above. These results were then used to plot a linear regression of the enthalpy change with respect to melting temperature. Using this plot and the following equation:

$$\Delta Cp_H = d\Delta H/dT_m \quad (2.9)$$

a value was determined for the heat capacity change in enthalpy. Similarly, the value for heat capacity change in entropy was determined from a linear regression of the entropy change with respect to the logarithmic scale of the melting temperature and the equation:

$$\Delta Cp_S = d\Delta S/d\ln T_m \quad (2.10)$$

This value gives the change in heat capacity for the entire molecule. By dividing these values by the number of base pairs, one determines the change in heat capacity per base pair for the formation of the helix [97].

Calculation of thermodynamic parameters from differential scanning calorimetry

Analysis of the differential scanning calorimetry results was carried out using CpCalc software (Calorimetry Sciences Corporation). Data files containing the experimental results and the corresponding buffer-buffer scans were inserted into the data sheet. The raw data was converted to molar heat capacity by subtracting a buffer-buffer scan from the sample-buffer scan in the software's "converter object". The partial specific volume used for the nucleic acids was 0.55 g / ml and the cell volume was 0.299 ml. A quadratic polynomial baseline was fit to the molar heat capacity plot to obtain the excess heat capacity curve. Data was then exported to SigmaPlot to generate a ΔC_p versus T plot. $\Delta H°_{cal}$ was obtained by integrating the area under the curve. T_m values were determined from the peak of the excess heat capacity curve. These values were used to obtain ΔS and ΔG values from the calorimetry data.

In some instances it was desired to obtain a value for ΔCp using the calorimetry data. To do this, a plot was made of the molar heat capacity curve. Both the pre-transition and post-transition baselines were extrapolated to the midpoint temperature of the transition and the difference at that point used as the change in heat capacity.

Calculation of tandem mismatch loop thermodynamic parameters

The thermodynamic parameters of a tandem mismatch loop were determined by calculating their contribution to the stability of the hairpin DNA molecules containing the tandem mismatch. These calculations will be illustrated in detail in chapter 3. In general, the parameters are calculated according to the following equation:

$$\Delta G^\circ_{loop} = \Delta G^\circ_{hairpin\ with\ loop} - (\Delta G^\circ_{hairpin\ without\ loop} - \Delta G^\circ_{interrupted\ base\ pair}) \quad (2.11)$$

The last term of the equation ($\Delta G^\circ_{hairpin\ without\ loop} - \Delta G^\circ_{interrupted\ base\ pair}$) subtracts the energetic contribution of the parts of the molecule not involved in the loop. The experimental free energy value of the hairpin DNA without a tandem mismatch loop was used. The free energy of stacking for the base pair that is interrupted by the mismatches is subtracted from this value. The ΔH° and ΔS° for the loop may be determined in a similar manner.

CHAPTER 3

THERMODYNAMIC PARAMETERS OF SHORT DNA HAIRPINS CONTAINING TANDEM MISMATCHES

Introduction

Accurate prediction of secondary structure in DNA is important to a number of molecular biology techniques used in research such as polymerase chain reaction, Southern blotting, and antisense drug design. In each of these cases a DNA strand hybridizes to a complementary nucleic acid target site in the presence of a large number of other sites. Some of these competing sites may have sequences very similar to the target site. In order to select conditions that maximize stringently selective hybridization between the probe DNA strand and the target site it is necessary to know the effect of single and tandem mismatches on the thermal stability of a hybrid.

Previously, Ke and Wartell used temperature gradient gel electrophoresis (TGGE) to determine the melting temperature and estimate the free energy of a series of tandem mismatches contained within duplexed DNA several hundred base pairs in length [40]. This work showed that the stability of a tandem mismatch was influenced by the sequence of the mispairs and the adjacent base pairs. The DNA transitions in the Ke study occurred in a polyacrylamide gel involving a solvent with approximately 20% formamide, 0.045 M Tris, and only 2 mM Na^+. Thus the relative free energy values evaluated for the tandem mismatches may not be appropriate or scaleable to the typical

0.1 to 1.0 M Na^+ conditions that are conventionally used. Here we have chosen some of the same tandem mismatches to study in an effort to compare TGGE, UV melting, and DSC.

In the Ke and Wartell studies, and in many other thermodynamic studies on nucleic acids, duplexed DNA's are utilized which contain mismatches or mispaired structural motifs, such as bulges and internal loops. In the instance of two single stranded molecules coming together, there is an initiation or docking factor that contributes to the total enthalpy of the molecule, but which is generally assumed to be close to zero. Holbrook *et al.* have shown that as temperature increases, there are also small increases in $\Delta H°_{dock}$ [65]. In order to avoid any discrepancies in the data due to this factor, the molecules used here were designed to be hairpins.

It has been shown previously that tandem G-A mismatches are of particular interest for a number of reasons. The G-A mismatch base pairs are among the more stable DNA mismatches [98, 99] and can lead to mutations with different repair efficiencies [100]. These mismatches are among the most studied by X-ray and NMR techniques. Li *et al.* examined duplexes containing tandem G-A mismatches in different closing base pair contexts [42]. They found that 5'-pyrimidine-GA-purine-3' sequences were considerably more stable than 5'-purine-GA-pyrimidine-3' sequences. In some cases, the difference in melting temperature between these two types of sequences was 10 °C. Ebel *et al.* found that 5'-pyrimidine-GA-purine-3' sequences are also sometimes more stable than 5'-pyrimidine-TA-purine-3' sequences [43].

Methods

The tandem mismatches examined were selected from those studied by Ke and Wartell and spanned the highest and lowest free energies evaluated. The tandem mismatches were placed in a DNA hairpin with an eight-base pair stem and a five-base

hairpin loop. Initially experiments were conducted using a four-base hairpin loop, but these molecules were found to be too stable to conveniently analyze using a 1.0 M Na^+ solvent. The sequence of the oligodeoxynucleotides used was 5'-GATCXYGCTGATCTACAGCWZGATC-3', where XY and WZ represent mismatched or complementary base pairs. In two cases one or both of the XY/WZ pairs were omitted.

The oligonucleotides were obtained commercially and purified as described in Chapter 2. All DNA oligomers produced a single band on a 16.5 cm x 20 cm 20% polyacrylamide gel. Based on ethidium bromide staining sensitivity, they were estimated to be at least 95% pure. The hairpin DNA molecules are referred to by the bases located at positions 5, 6 and 21, 22 (XY/WZ) in the above core sequence. The DNA hairpin designated --/-- (figure 3.1) has no bases in the positions represented by X, Y, W, and Z. The A/T DNA hairpin DNA contains one A-T base pair at the location of what is generally a tandem mismatch (figure 3.1). DNA molecules with all Watson-Crick paired bases in the stem are denoted by capital letters throughout the paper while mismatch bases are shown with lower case letters.

Concentrations were determined using phosphodiesterase degradation as described in Chapter 2. Concentrations were also estimated using extinction coefficients calculated for single strands based on dinucleotide frequencies [101] using absorbance values of denatured oligomers measured at 90 °C. The latter approach produced concentrations similar to those determined by enzyme degradation for some oligonucleotides (±6%) but not others. For this reason only concentration determined by phosphodiesterase degradation was utilized.

UV absorbance was employed to monitor the melting transitions of the hairpin molecules. The DNA was diluted to a concentration of approximately 0.45 OD / mL at 260 nm (approximately 1.8 μM strand) in either 1.0 M NaCl, 0.01 M MOPS, pH 7.0, or 0.08 M NaCl, 0.01 M Na_2HPO_4, pH 7.0. Prior to loading into the spectrophotometer,

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

--/-- hairpin

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

aa/gc hairpin

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

A/T hairpin

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

ca/gc hairpin

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

AA/TT hairpin

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

ga/ga hairpin

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

GA/TC hairpin

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

ta/ac hairpin

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

tc/tc hairpin

Figure 3.1. Expected structure of DNA hairpins containing tandem mismatches. Each hairpin has a five base hairpin loop. The control molecules have stems containing eight, nine, or ten Watson-Crick paired bases. The experimental molecules have stems with eight Watson-Crick paired bases interrupted by two non-canonical paired bases.

samples were denatured at 95 °C for 3 minutes, then slowly cooled to room temperature to allow annealing. Samples were then placed in a cuvette and absorbance readings were taken every 0.1 °C over an appropriate temperature range. This was determined by utilizing a wide range, such as 25 – 95 °C, for initial studies on a molecule and later narrowing the range if possible. Four or more melting transitions were obtained for each sample. Reverse transitions closely followed forward transitions.

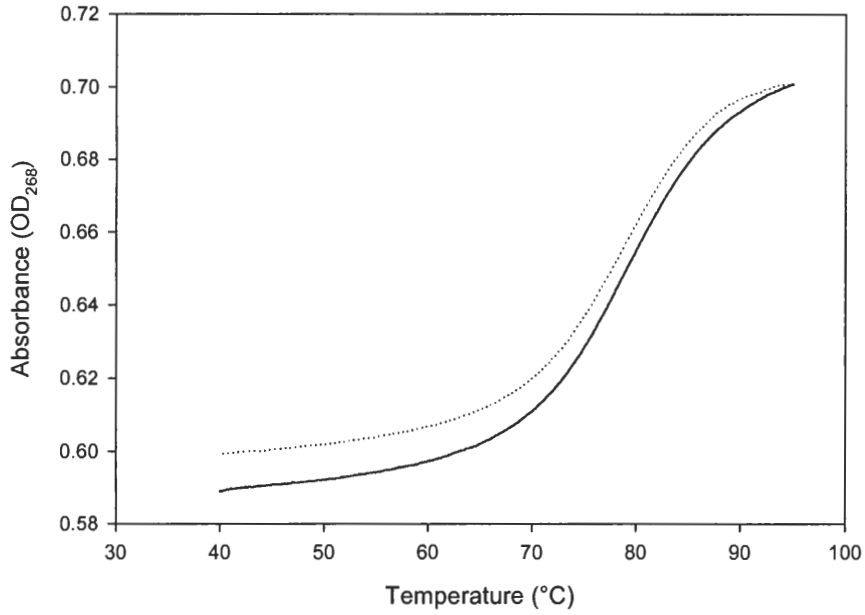
The ga/ga, AA/TT, and --/-- molecules, predicted to be among the most stable, were also prepared at concentrations of 0.8 μM and 40 μM in 1.0 M NaCl, 0.01 M MOPS and melting curves obtained. The 40 μM samples were denatured in a 2 mm path length cell. The midpoint temperature of the transitions, T_m, for these three molecules did not change significantly over the 50-fold concentration range. The ga/ga molecule was also examined by the sedimentation equilibrium method described in Chapter 2.

Results

Figure 3.1 shows the expected structure of the nine DNA's examined in this study. The sequence was designed to facilitate folding into a hairpin molecule with a five base hairpin loop and a stem of eight Watson-Crick paired bases, interrupted by two tandem mismatch pairs. Control molecules were designed to have only the eight paired bases in the stem, or one or two additional Watson-Crick pairs in place of the tandem mismatches.

Figure 3.2 shows typical UV melting curves for control and tandem mismatch DNA's in 1.0 M sodium. This buffer is the canonical salt concentration used for most thermal denaturation studies. The transitions were generally highly reversible. The solid lines represent the denaturation curves while the dotted lines represent the renaturation curves. When denaturation and renaturation curves were each normalized between 0 and 1.0, the normalized curves were essentially identical.

A.



B.

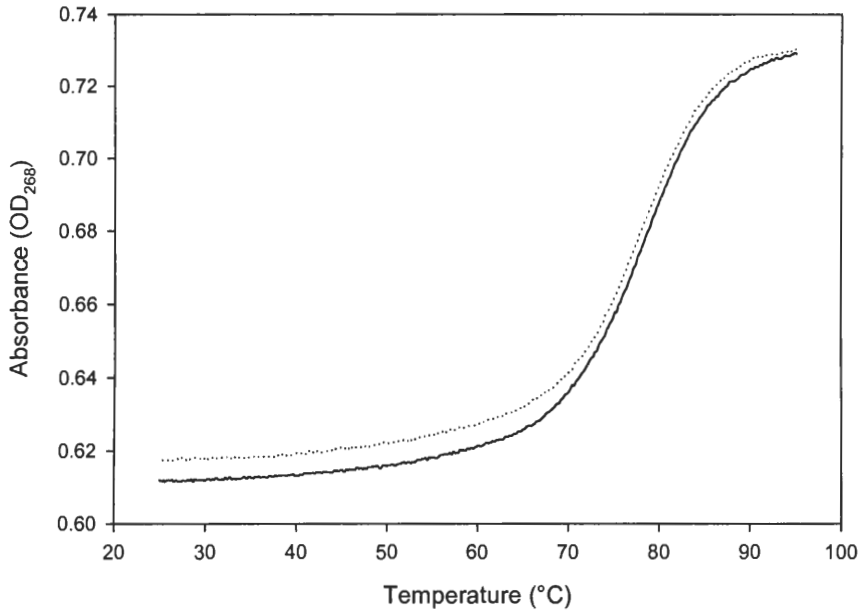
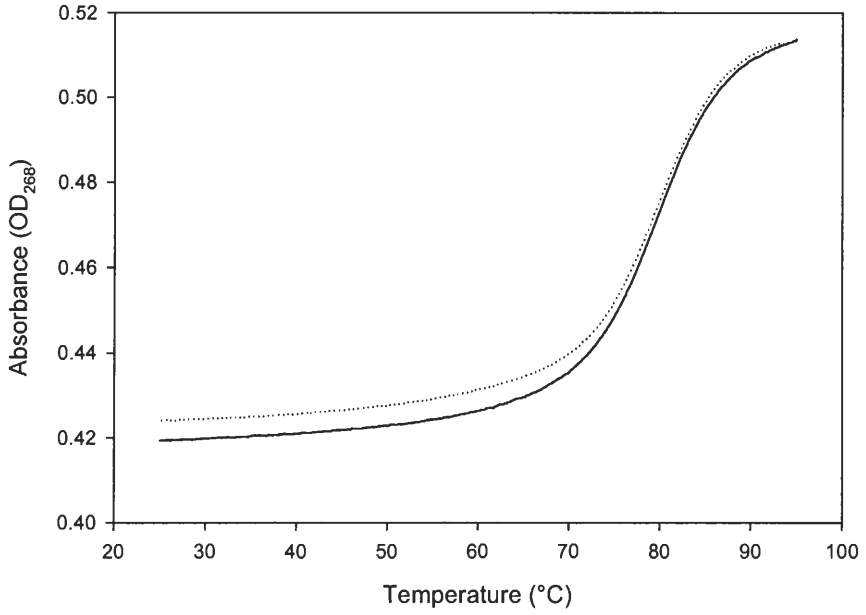


Figure 3.2. Typical UV denaturation curves for short DNA hairpins containing tandem mismatches in 1.0 M Na^+. Denaturation curves are represented by solid lines while renaturation curves are represented by dotted lines. A. --/--. B. A/T.

C.



D.

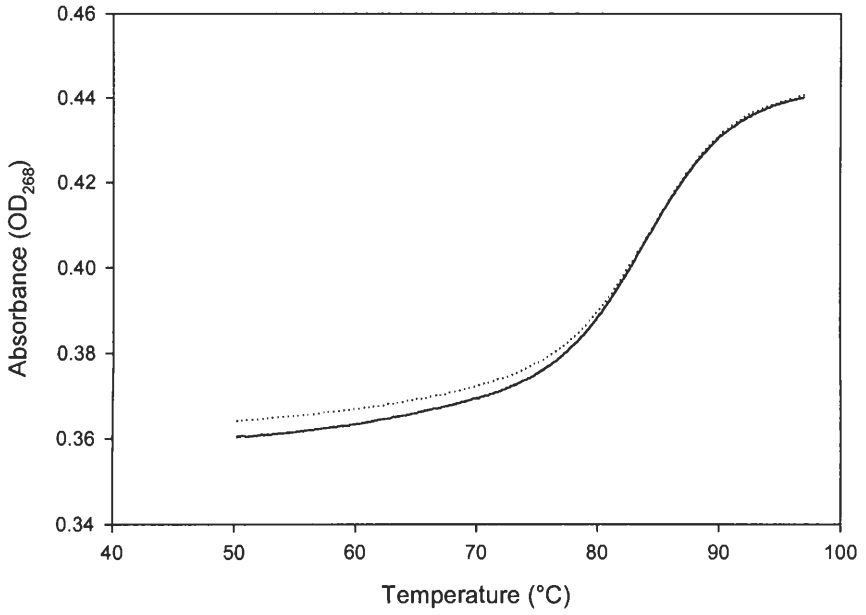
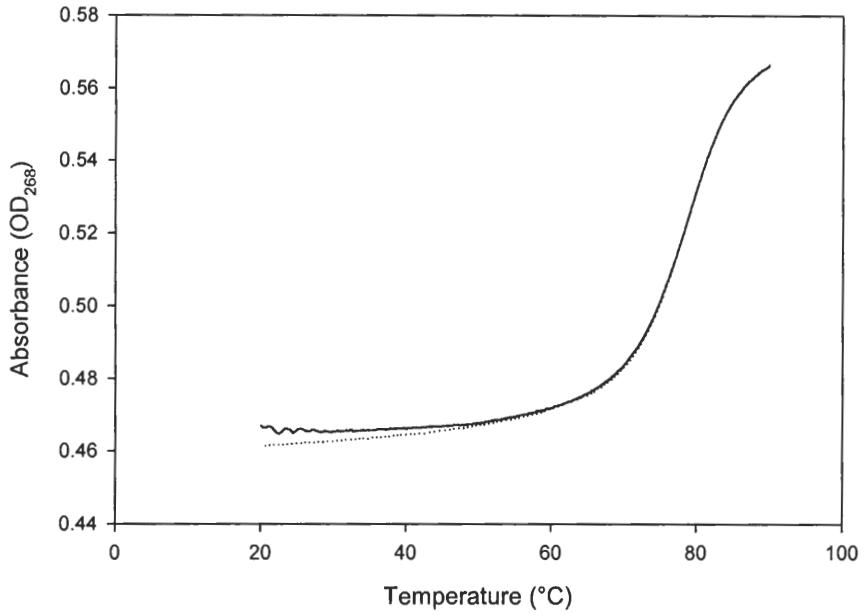


Figure 3.2 (continued). Typical UV denaturation curves for short DNA hairpins containing tandem mismatches in 1.0 M Na^+. Denaturation curves are represented by solid lines while renaturation curves are represented by dotted lines. C. AA/TT. D. GA/TC.

E.



F.

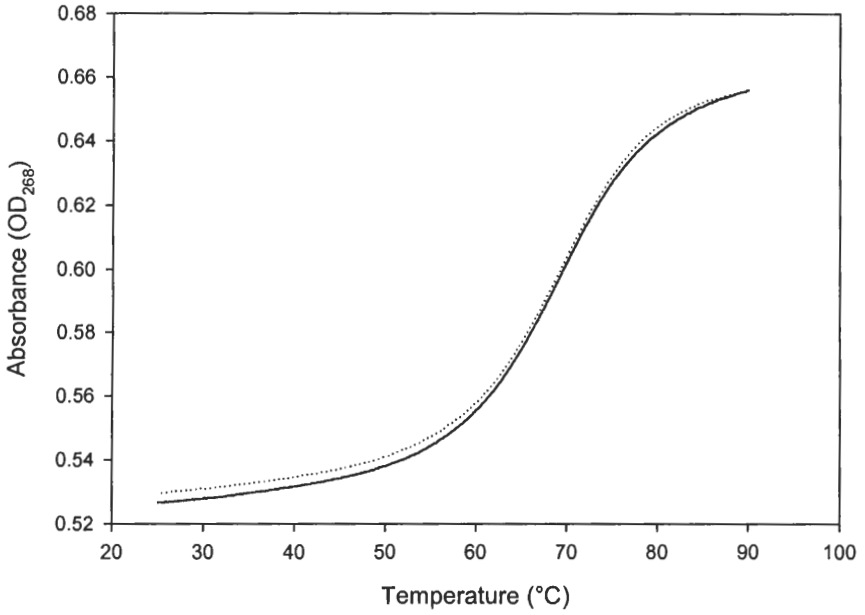
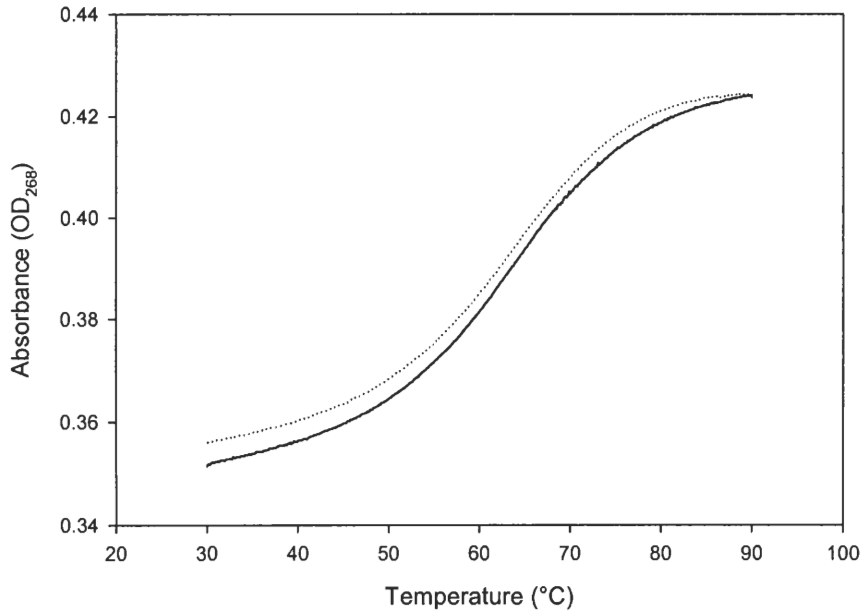


Figure 3.2 (continued). Typical UV denaturation curves for short DNA hairpins containing tandem mismatches in 1.0 M Na^+. Denaturation curves are represented by solid lines while renaturation curves are represented by dotted lines. E. ga/ga. F. aa/gc.

G.



H.

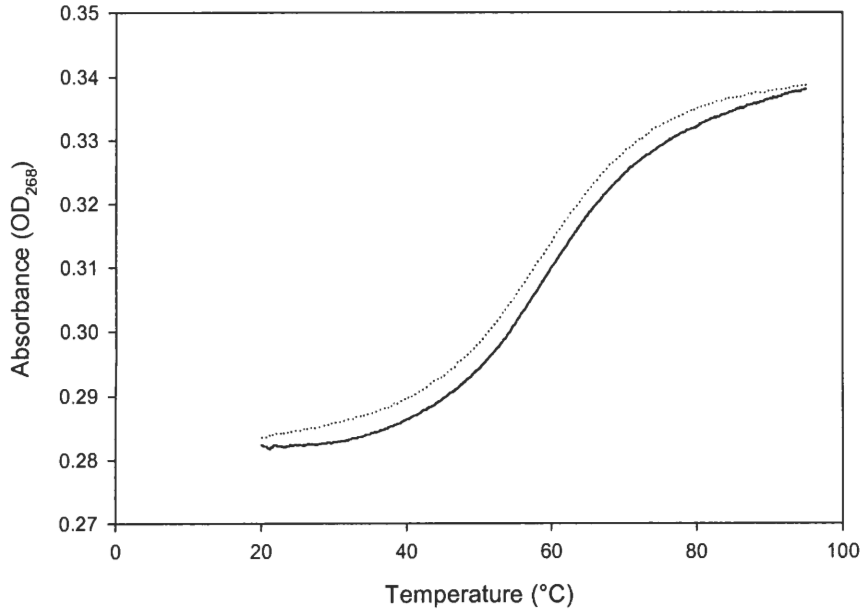


Figure 3.2 (continued). Typical UV denaturation curves for short DNA hairpins containing tandem mismatches in 1.0 M Na^+. Denaturation curves are represented by solid lines while renaturation curves are represented by dotted lines. G. ca/gc. H. ta/ac.

I.

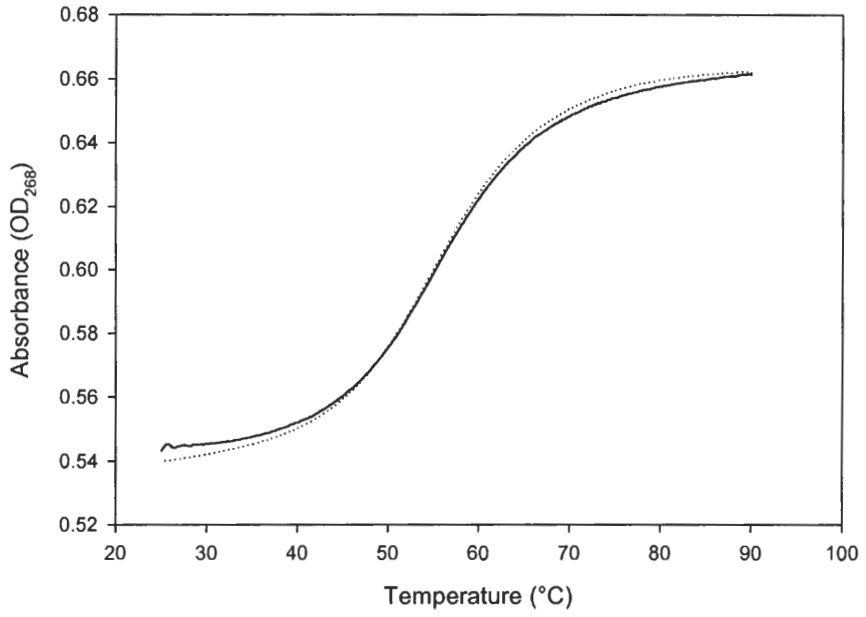


Figure 3.2 (continued). Typical UV denaturation curves for short DNA hairpins containing tandem mismatches in 1.0 M Na^+. Denaturation curves are represented by solid lines while renaturation curves are represented by dotted lines. I. tc/tc.

Table 3.1 compares the two methods of obtaining van't Hoff thermodynamic values. The free energy was calculated using the assumption that enthalpy and entropy are independent of temperature, or that the change in heat capacity between helix and coil states is zero. In general, the results obtained from the two methods were quite similar. Notably, the molecules which showed the biggest difference in enthalpy between the two methods were the GA/TC hairpin and the ga/ga hairpin. The characteristic that distinguishes these molecules from the other hairpins is a GpA step in positions 5 and 6. This may indicate a unique feature of the sequence that causes the thermodynamic values evaluated by the two methods to differ. Additionally, the AA/TT molecule shows somewhat divergent values for the two methods, albeit not as great as the differences for the GA/TC and ga/ga values. The above three DNAs are the most stable of the hairpins studied. The greater disparity between the two van't Hoff methods for these three DNA's may indicate that their transitions are not completely two-state. In general, the standard deviations for the peak height method are smaller than the deviations for the curve fit method. The average of the standard deviations for the peak height method was 1.2 kcal / mol while the average for the curve fit method was 2.7 kcal / mol. Differential scanning calorimetry studies were difficult to obtain in the 1.0 M Na^+ buffer. Since the DNA hairpins denatured at a relatively high temperature in this buffer and some transitions extended beyond 90 °C, it was difficult to define a post-transition baseline.

Although melting experiments are generally conducted in 1.0 M sodium buffer, the monovalent cation concentration in a cell is generally considered to be lower (approximately 0.15 to 0.25 M) and magnesium is generally present [102]. For this reason, the DNA hairpin transitions were also examined in 0.1 M NaCl plus 5 mM $MgCl_2$. Typical denaturation curves for the molecules in these salt conditions may be found in Appendix 1. Table 3.2 shows the thermodynamic parameters for DNA hairpin formation

Table 3.1. Van't Hoff thermodynamic parameters for DNA hairpin formation in 1.0 M Na^+. Peak height and curve fit refer to the method of van't Hoff analysis used to obtain results.

	Tm (°C)	Peak height -ΔH (kcal / mol)	Curve fit -ΔH (kcal / mol)	Peak height -ΔS (eu)	Curve fit -ΔS (eu)	Peak height $-\Delta G_{37}$ (kcal / mol)	Curve fit $-\Delta G_{37}$ (kcal / mol)
--/--	77.7 ± 0.4	52.6 ± 0.4	52.7 ± 2.6	150 ± 3	150 ± 8	6.11 ± 0.15	6.22 ± 0.44
A/T	78.1 ± 0.4	59.9 ± 1.7	58.6 ± 4.4	175 ± 14	167 ±13	7.21 ± 0.63	6.68 ± 0.51
AA/TT	79.2 ± 0.2	64.0 ± 1.1	59.7 ± 2.8	182 ± 3	170 ± 8	7.66 ± 0.12	6.95 ± 0.38
GA/TC	82.6 ± 0.2	65.0 ± 1.6	56.4 ± 2.7	183 ± 5	159 ± 7	8.33 ± 0.22	7.05 ± 0.46
ga/ga	77.7 ± 0.2	66.9 ± 1.4	58.8 ± 3.1	191 ± 4	168 ± 9	7.77 ± 0.17	6.57 ± 0.43
aa/gc	68.5 ± 0.3	47.2 ± 2.4	46.7 ± 4.1	138 ± 7	137 ± 12	4.35 ± 0.21	4.19 ± 0.44
ca/gc	62.9 ± 0.3	39.6 ± 1.4	41.4 ± 2.0	118 ± 4	123 ± 5	3.05 ± 0.13	3.32 ± 0.56
ta/ac	58.7 ± 1.1	32.6 ± 0.3	33.0 ± 1.2	99 ± 1	100 ± 4	2.14 ± 0.09	1.93 ± 0.13
tc/tc	54.9 ± 0.4	41.4 ± 0.8	41.5 ± 1.4	126 ± 3	127 ± 4	2.25 ± 0.05	2.19 ± 0.12

Table 3.2. Van't Hoff thermodynamic parameters for DNA hairpin formation in 5 mM Mg^{2+} plus 0.1 M Na^{+}. Peak height and curve fit refer to the method of van't Hoff analysis used to determine results.

	Tm (°C)	Peak height -ΔH (kcal / mol)	Curve fit -ΔH (kcal / mol)	Peak height -ΔS (eu)	Curve fit -ΔS (eu)	Peak height $-\Delta G_{37}$ (kcal / mol)	Curve fit $-\Delta G_{37}$ (kcal / mol)
--/--	72.3 ± 0.1	56.0 ± 1.3	53.9 ± 2.9	162 ± 4	157 ± 8	5.73 ± 0.13	5.40 ± 0.31
A/T	ND	ND	ND	ND	ND	ND	ND
AA/TT	71.4 ± 0.1	66.5 ± 1.7	63.4 ± 3.5	193 ± 5	184 ± 10	6.64 ± 0.17	6.27 ± 0.36
GA/TC	77.1 ± 0.2	67.8 ± 1.5	62.4 ± 1.9	194 ± 4	179 ± 5	7.76 ± 0.17	7.01 ± 0.23
ga/ga	71.5 ± 0.3	65.3 ± 2.0	62.0 ± 4.7	190 ± 6	181 ± 14	6.54 ± 0.17	6.01 ± 0.50
aa/gc	63.6 ± 0.5	53.4 ± 1.0	53.2 ± 2.0	161 ± 6	159 ± 6	4.30 ± 0.24	4.08 ± 0.18
ca/gc	57.5 ± 0.3	41.8 ± 1.4	42.8 ± 2.4	127 ± 4	130 ± 8	2.59 ± 0.09	2.58 ± 0.15
ta/ac	53.4 ± 0.8	35.6 ± 1.3	36.3 ± 1.3	109 ± 4	112 ± 4	1.78 ± 0.11	1.70 ± 0.06
tc/tc	47.8 ± 0.4	38.2 ± 1.4	38.1 ± 2.5	119 ± 5	119 ± 8	1.29 ± 0.02	1.27 ± 0.11

evaluated by the peak height and the curve fit methods. The two methods gave very similar results. The biggest differences between the two methods were again seen in the more stable molecules, although the percent difference was smaller in this buffer than in 1.0 M Na^+. For example, the difference between the enthalpy values determined by the two methods in 1.0 M Na^+ was 13% for the GA/TC molecule, while the percent difference for that same molecule in the lower salt and magnesium buffer was 8%. The standard deviations in this buffer also followed similar trends to the 1.0 M Na^+. The average standard deviation for the peak height data was 1.5 kcal / mol while the average for the curve fit data was 2.7 kcal / mol.

Comparing the average enthalpy changes for the 1.0 M Na^+ and 0.1 M Na^+ / 5 mM Mg^{2+} buffers shows a difference less than 12% for each of the DNAs studied. As with the 1.0 M Na^+ studies, the melting temperatures of these molecules were too high in the 0.1 M Na^+ / 5 mM Mg^{2+} buffer to allow for DSC experiments.

UV denaturation analysis and DSC experiments were employed to determine thermodynamic parameters in 0.1 M Na^+. The lower T_m in this buffer allowed DSC experiments to be performed. Typical UV melting curves for the hairpins in 0.1 M Na^+ may be found in Appendix 2. Table 3.3 shows the thermodynamic parameters evaluated from the UV denaturation curves by peak height and curve fit analyses, and the parameters determined by differential scanning calorimetry. The two van't Hoff methods for determining thermodynamic parameters produced results that are nearly identical. The largest difference observed, for the A/T molecule, was still within experimental error. As seen with the other two buffer conditions, the standard deviations for the curve fit method are slightly higher than the deviation for the peak height method.

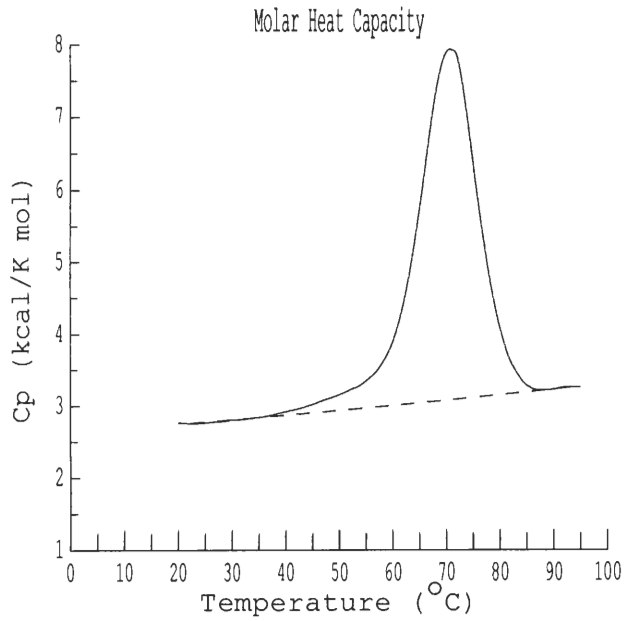
Figure 3.3 shows molar heat capacity curves from DSC experiments in 0.1 M Na^+. The area under each curve was determined after fitting a quadratic polynomial

Table 3.3. Van't Hoff thermodynamic parameters for DNA hairpin formation in 0.1M Na^+.

UV Denaturation							
	Tm (°C)	Peak height -ΔH (kcal / mol)	Curve fit -ΔH (kcal / mol)	Peak height -ΔS (eu)	Curve fit -ΔS (eu)	Peak height $-\Delta G_{37}$ (kcal / mol)	Curve fit $-\Delta G_{37}$ (kcal / mol)
--/--	71.0 ± 0.2	62.7 ± 1.7	61.7 ± 1.4	182 ± 5	180 ± 4	6.20 ± 0.18	6.01 ± 0.15
A/T	69.1 ± 0.2	63.8 ± 1.6	57.8 ± 4.7	187 ± 5	170 ± 14	5.98 ± 0.14	5.21 ± 0.48
AA/TT	68.3 ± 0.3	69.4 ± 1.8	68.3 ± 3.2	204 ± 5	200 ± 10	6.37 ± 0.17	6.19 ± 0.26
GA/TC	73.6 ± 0.3	75.6 ± 3.3	75.7 ± 4.7	218 ± 10	219 ± 14	7.98 ± 0.36	7.88 ± 0.52
ga/ga	68.9 ± 0.2	69.5 ± 0.9	66.7 ± 2.9	203 ± 3	196 ± 9	6.49 ± 0.06	6.12 ± 0.30
aa/gc	58.0 ± 0.2	56.6 ± 1.6	56.0 ± 3.2	171 ± 5	169 ± 10	3.58 ± 0.12	3.52 ± 0.17
ca/gc	50.8 ± 0.4	43.8 ± 2.0	46.6 ± 2.7	135 ± 6	144 ± 8	1.87 ± 0.14	2.08 ± 0.15
ta/ac	45.4 ± 0.4	39.1 ± 1.6	40.4 ± 2.9	120 ± 5	127 ± 9	1.03 ± 0.08	1.18 ± 0.13
tc/tc	45.4 ± 0.2	49.2 ± 0.5	48.6 ± 1.2	155 ± 2	153 ± 4	1.29 ± 0.03	1.32 ± 0.04

Differential Scanning Calorimetry				
	Tm (°C)	-ΔH (kcal / mol)	-ΔS (eu)	$-\Delta G_{37}$ (kcal / mol)
--/--	73.4 ± 0.5	55.5 ± 2.3	160 ± 7	5.88 ± 0.20
A/T	70.7 ± 0.1	67.2 ± 1.1	196 ± 3	6.53 ± 0.13
AA/TT	71.4 ± 0.1	64.3 ± 2.1	187 ± 6	6.42 ± 0.20
GA/TC	75.5 ± 0.2	75.1 ± 4.8	215 ± 14	8.35 ± 0.48
ga/ga	70.3 ± 0.2	62.2 ± 4.4	181 ± 13	6.04 ± 0.38
aa/gc	60.2 ± 0.1	54.0 ± 1.9	162 ± 6	3.76 ± 0.17
ca/gc	53.6 ± 0.1	43.3 ± 1.5	133 ± 4	2.26 ± 0.16
ta/ac	48.2 ± 0.5	36.2 ± 3.2	113 ± 10	1.22 ± 0.16
tc/tc	46.9 ± 0.4	39.6 ± 1.4	124 ± 4	1.20 ± 0.15

A.



B.

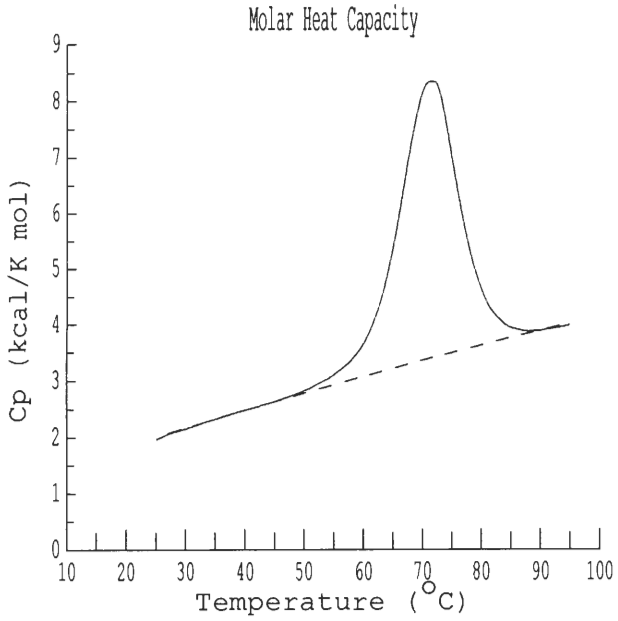
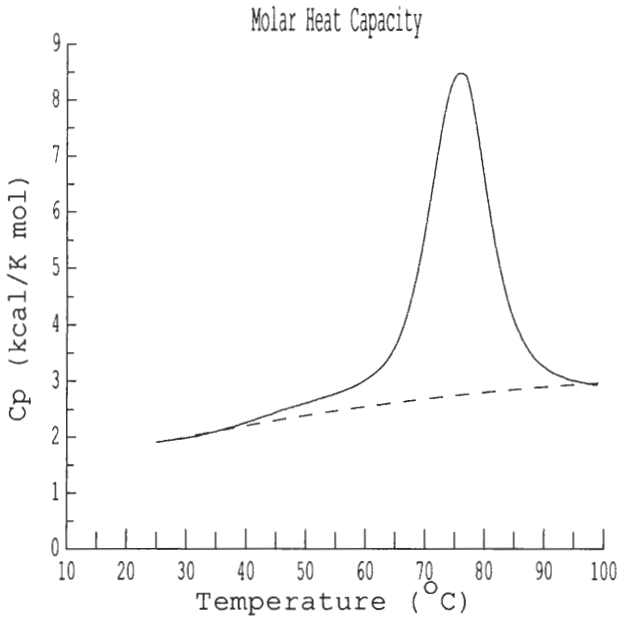


Figure 3.3. Typical molar heat capacity curves from DSC experiments on short DNA hairpins containing tandem mismatches in 0.1 M Na^+. A. A/T. B. AA/TT.

C.



D.

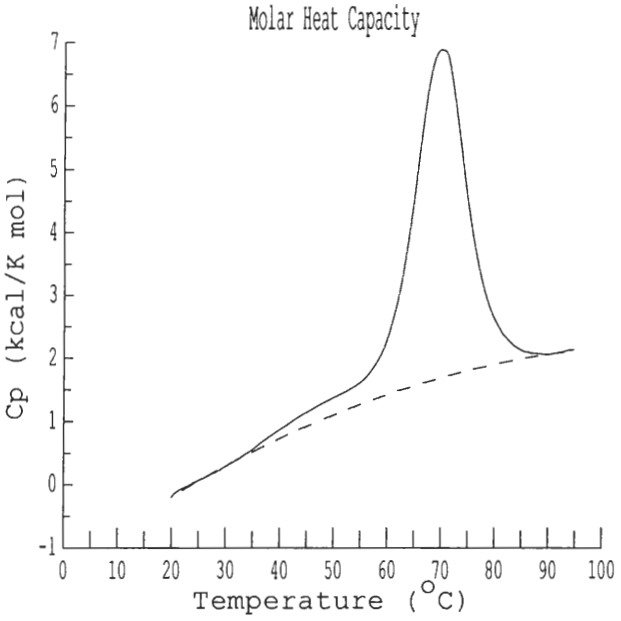
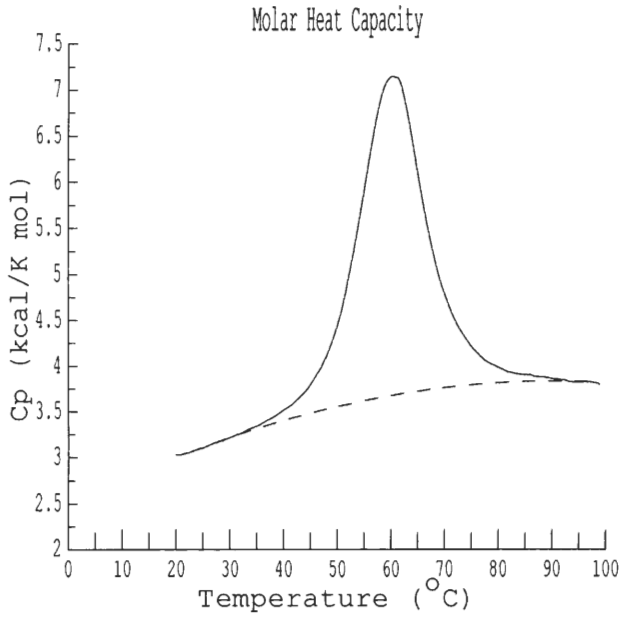


Figure 3.3 (continued). Typical molar heat capacity curves from DSC experiments on short DNA hairpins containing tandem mismatches in 0.1 M Na^+. C. GA/TC. D. ga/ga.

E.



F.

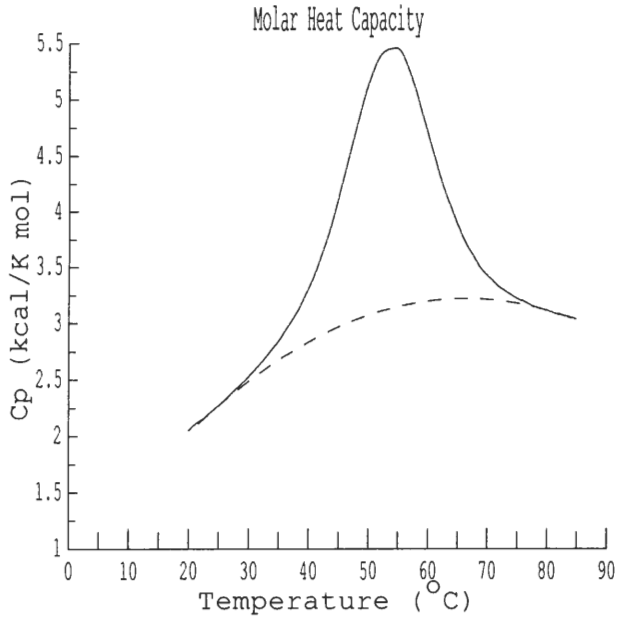
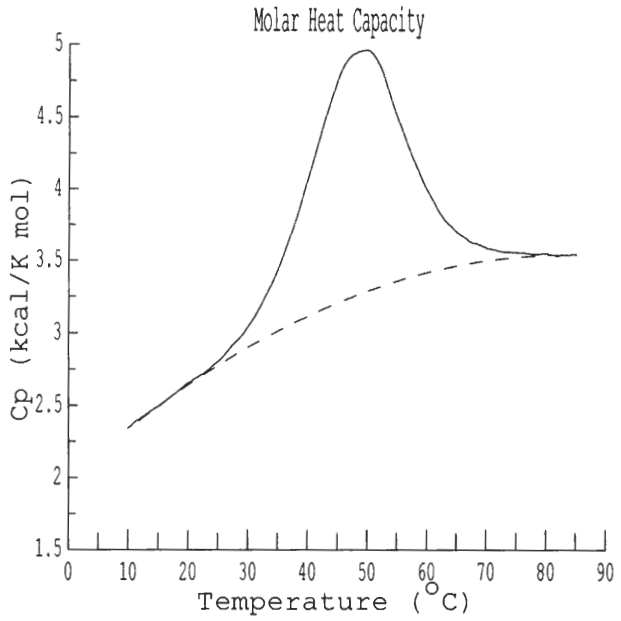


Figure 3.3 (continued). Typical molar heat capacity curves from DSC experiments on short DNA hairpins containing tandem mismatches in 0.1 M Na^+. E. aa/gc. F. ca/gc.

G.



H.

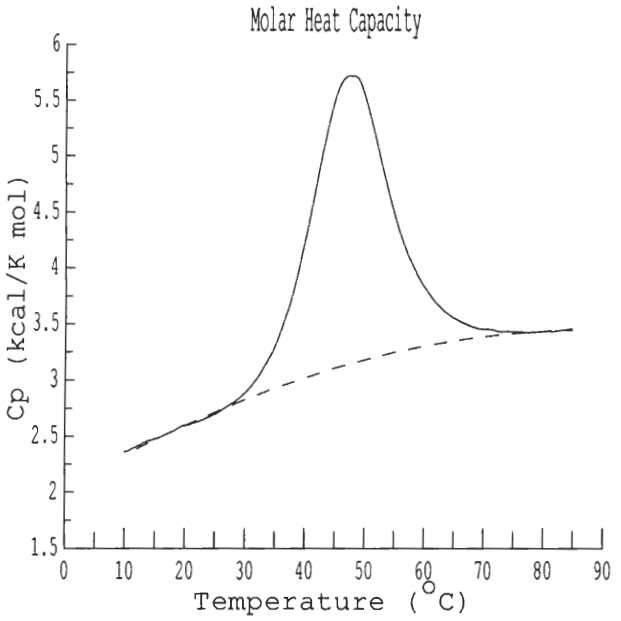


Figure 3.3 (continued). Typical molar heat capacity curves from DSC experiments on short DNA hairpins containing tandem mismatches in 0.1 M Na^+. G. ta/ac. H. tc/tc.

baseline (dashed lines). This area, along with the temperature value at the curve peak was used to obtain calorimetric thermodynamic parameters.

The melting temperatures of the samples run in the DSC were always slightly higher than the samples examined spectrophotometrically (2.1 ± 0.6°C). This is likely to be due to a difference in calibration of the temperature probes. This was verified by evaluating the T_m of a single sample preparation by DSC and UV denaturation. The T_m difference determined by DSC was 2 °C higher than the T_m determined by UV. The thermodynamic parameters were similar to those evaluated by van't Hoff analysis.

Figure 3.4 shows the van't Hoff and calorimetric enthalpies determined for the hairpin molecules. The van't Hoff values are represented by solid circles while the calorimetric enthalpies are open circles. Gray shading shows a 15% range on either side of the van't Hoff values. For each of the molecules except tc/tc, the calorimetric enthalpy falls within 15% of the van't Hoff enthalpy. For the tc/tc molecule the difference is 20%. Previous studies have considered a 15% or less difference in enthalpy values evaluated by the two different approaches utilizing UV denaturation curves as indicative of a two-state transition [52]. Since the values obtained by DSC are model-independent, the close agreement observed in figure 3.4 supports the assumed two-state nature of the transition.

Table 3.4 lists the thermodynamic parameters for the internal loops created by the tandem mismatches. These parameters are calculated according to equation 2.11. The last term of the equation ($\Delta G^\circ_{\text{hairpin without loop}} - \Delta G^\circ_{\text{interrupted base pair}}$) subtracts the energetic contribution of the parts of the molecule not involved in the loop. To determine a value for this term, the DNA hairpins with all Watson-Crick base pairs in the stem were used. For the --/-- molecule, the published free energy of the CG/CG stacking interaction was subtracted from the experimental free energy value of the hairpin [22]. This approach was utilized in a similar manner for the other duplex molecules. For the

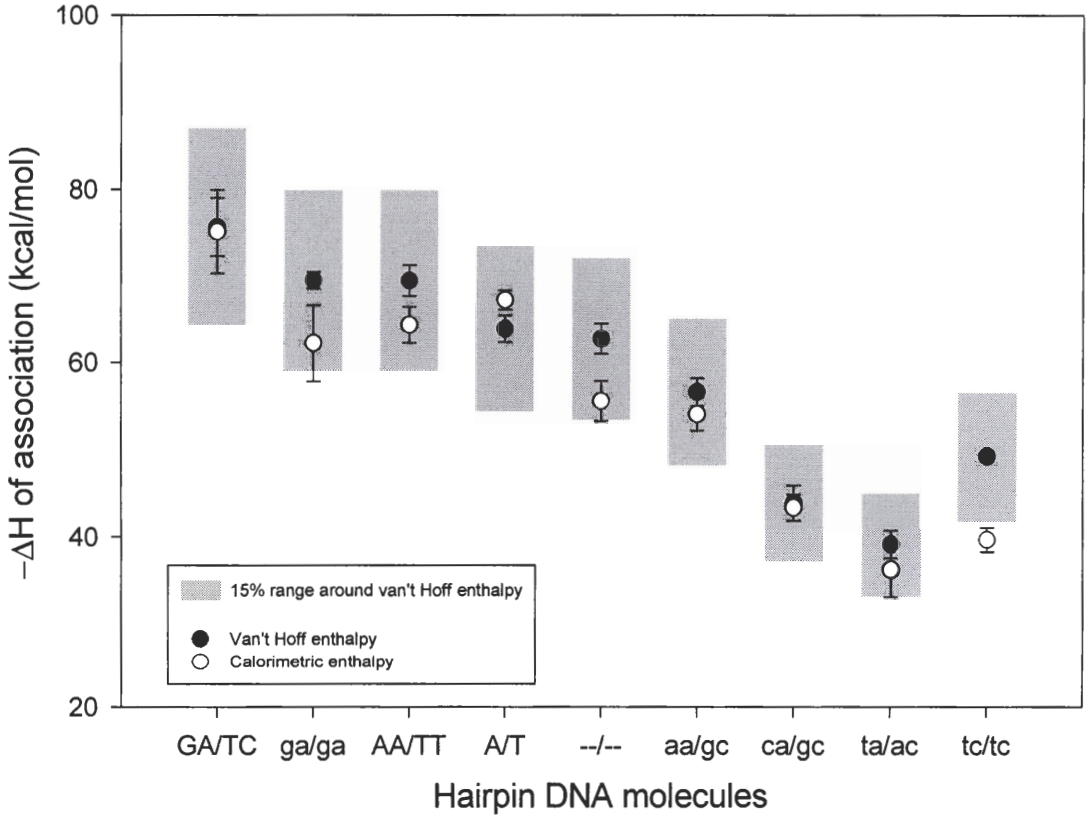


Figure 3.4. Comparison of van't Hoff enthalpy and calorimetric enthalpy. Filled circles represent the van't Hoff values as determined by the peak height method while open circles represent the calorimetric data. Error bars represent the standard deviation of enthalpies determined from a minimum of four denaturation curves. Gray boxes show a 15% range on either side of each of the van't Hoff values, not including error bars. The 15% range was determined by multiplying the peak height enthalpies found in table 3.3 (p. 47) by 0.85 and 1.15.

Table 3.4A. Thermodynamic properties of tandem mismatches in 1.0 M Na^+

Mismatch	ΔG_{37} (kcal / mol)	ΔS (cal/mol-K)	ΔH (kcal / mol)
ga/ga	-3.79	-70	-25.5
aa/gc	-0.37	-17	-5.8
ca/gc	0.93	3	-1.8
ta/ac	1.84	22	8.8
tc/tc	1.73	-5	0.0

Table 3.4B. Thermodynamic properties of tandem mismatches in 0.1 M Na^+.

Mismatch	ΔG_{37} (kcal / mol)		ΔS (cal/mol-K)		ΔH (kcal / mol)	
	UV	DSC	UV	DSC	UV	DSC
ga/ga	-2.62	-2.00	-59	-45	-21.0	-16.1
aa/gc	0.29	0.28	-27	-26	-8.1	-7.9
ca/gc	2.00	1.78	9	3	4.7	2.8
ta/ac	2.84	2.82	24	23	9.4	9.9
tc/tc	2.58	2.84	-11	12	-0.7	6.5

A/T DNA, it is necessary to subtract stacking interactions for CA/TG and AG/CT from the experimental free energy value of the DNA hairpin. In the case of the GA/TC molecule, the stacking free energies which must be subtracted are CG/CG, GA/TC, and AG/CT. Similarly, three stacking values are utilized for the AA/TT molecule. The values obtained by subtracting the stacking interactions represent the free energy contributed by the rest of the molecule. The four values, determined from the four hairpins with completely duplexed stems, were averaged and used for the last term in the equation ($\Delta G°_{\text{hairpin without loop}} - \Delta G°_{\text{interrupted base pair}}$). The enthalpy and entropy for the mismatches may be determined in a similar manner. The values determined for the last term of the equation in 1.0 M Na^+ were $\Delta G = -3.98 \pm 0.37$ kcal / mol, $\Delta S = 121 \pm 9$ eu, and $\Delta H = -41.4 \pm 3.0$ kcal / mol. In 0.1 M Na^+ the values were $\Delta G = -3.87 \pm 0.37$ kcal / mol, $\Delta S = 144 \pm 8$ eu, and $\Delta H = -48.5 \pm 2.9$ kcal / mol.

The thermodynamic values for the mismatches were also determined for UV denaturation experiments in 0.1 M sodium and for DSC experiments in the same solvent. To account for the difference in sodium concentration, the stacking parameters in 1.0 M Na^+ were adjusted using equations 1.2 and 1.3 (p.10). These equations provided corrected ΔG and ΔS values. ΔH was assumed to be independent of salt concentration. As would be expected, the tandem mismatch loops are more stable in the 1.0 M Na^+ buffer than in the 0.1 M Na^+ buffer. The values for the thermodynamic parameters were similar in the 0.1 M Na^+ when examined by UV denaturation and by DSC. Additionally, the hierarchy of stability based on free energy measurements was the same in both salt conditions.

To ensure that the transitions reflect hairpin to coil transitions and that duplexes with internal loops are not intermediates, UV melting curves were performed in 1.0 M sodium on selected DNA's at a strand concentration of 0.8 μM and 40 μM, a 50-fold difference. The AA/TT , --/--, and ga/ga hairpins were treated in this manner. For the

AA/TT hairpin, the melting temperatures were determined to be 78.5 ± 0.2 °C and 78.0 ± 0.1 °C at 0.8 and 40 μM strand concentration, respectively. For the --/-- molecule, these values were 78.7 ± 0.4 °C and 78.7 ± 0.4 °C. For the ga/ga hairpin, the melting temperatures were 69.3 ± 0.2 °C and 69.8 ± 0.3 °C at the two concentrations. Since the melting temperature of multi-stranded molecules increases with an increase in strand concentration, the lack of change in the T_m for the hairpin DNA's at different concentrations is indicative of the presence of a single-stranded entity. These results suggest that each of the molecules is in fact present as only a monomer.

Similar results may be seen by comparing the melting temperatures in 0.1 M sodium as determined by UV melting and DSC. Calorimetric experiments were carried out at 100 uM strand concentrations while UV experiments were done at approximately 1.8 uM strand concentrations. Although there is some discrepancy between the two temperatures for each molecule, the previously mentioned experiment (p. 52) indicates this is a systematic difference in temperature calibration.

Additionally, sedimentation equilibrium experiments were performed with known standards to determine the apparent molecular weight of the ga/ga molecule. In the initial analysis, a large difference was seen between the results from van't Hoff analysis of UV melting and from calorimetric results due to inaccurate evaluations of the DNA concentrations. For this reason, it was desirable to determine if the ga/ga molecule was present as a monomolecular hairpin or a duplex. From the centrifugation experiments, it was determined that the molecule is present as a twenty-five base molecule, consistent with it being a monomer rather than dimer (Figure 3.5).

In the initial studies, concentration had been determined by measuring the absorbance at a high temperature, where it should be single-stranded. This value was utilized in conjunction with the single-stranded extinction coefficient to determine concentration. This method was found to be accurate for some molecules and not

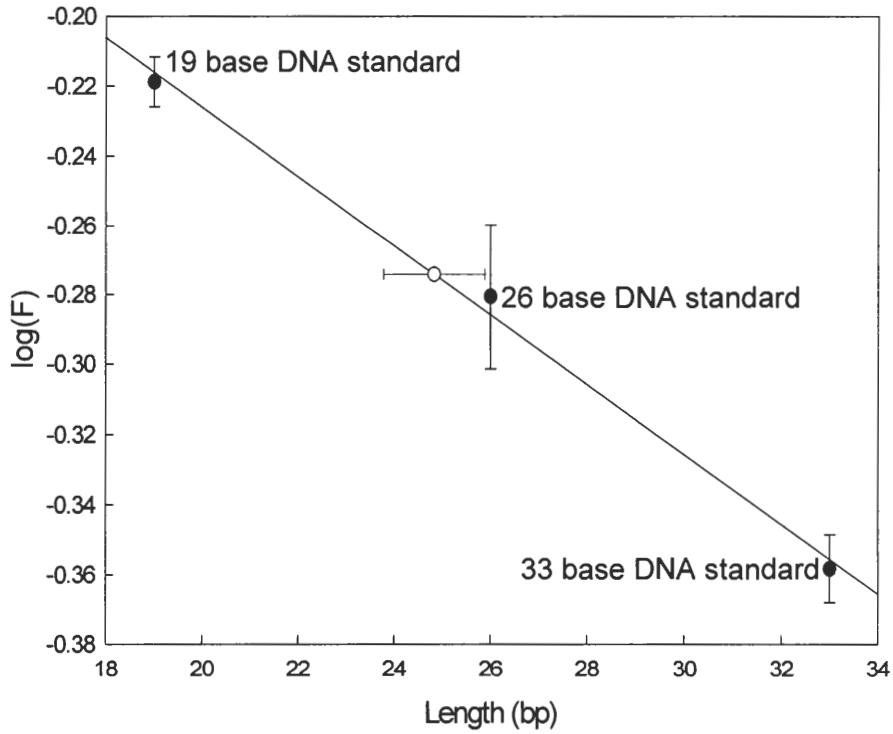


Figure 3.5. Results of sedimentation equilibrium experiments. Solid circles represent DNA standards of known length, with the number of bases indicated. The open circle represents the ga/ga molecule. From this, it was determined that the ga/ga molecule is present as a 25 base molecule, consistent with the monomer hairpin conformation.

others. DSC utilizes the strand concentration to determine molar heat capacity in a direct correlation. For example, if the concentration value that is input is smaller by 7% than the actual concentration, the resulting enthalpy values will be smaller by 7%. For this reason it is very important to use an accurate method, such as phosodiesterase degradation, to determine concentration for DSC experiments.

Discussion

The work done on short DNA hairpins containing tandem mismatches has led to several conclusions. One goal of this work was to compare results of TGGE experiments, where estimated free energies were determined, to the results of UV-monitored denaturation curves and DSC. The results of the optical melting experiments in 1.0 M and 0.1 M Na^+ found the hierarchy of stability for these mismatches, according to free energy values at 37 °C, from most stable to least stable to be ga/ga > aa/gc > ca/gc > ta/ac > tc/tc. This order is the same as the hierarchy of stability determined by TGGE by Ke and Wartell [40]. In that investigation, the estimated free energies for the tandem mismatch loops were as follows: ga/ga was 0.4 kcal / mol, aa/gc was 1.4 kcal / mol, ca/gc was 1.9 kcal / mol, ta/ac was 2.5 kcal / mol, and tc/tc was 3.0 kcal / mol. The estimated free energies reported are different from those determined here, however this is likely due to the differences in solution conditions.

Each of the molecules here were examined in three solution conditions, 1.0 M Na^+, 0.1 M Na^+, and 0.1 M Na^+ plus 5 mM Mg^{2+}. 1.0 M sodium is the traditional buffer condition for performing optical melting studies. Arguments have been presented previously justifying the application of thermodynamic results in 1.0 M Na^+ conditions to investigate DNA stability, but no experiments have been carried out to show this. The results show similar thermodynamic values for the DNA hairpins in 1.0 M Na^+ and in 0.1 M Na^+ with 5 mM Mg^{2+}. This suggests that it is appropriate to utilize thermodynamic

parameters of base pair stacking, tandem mismatches, and hairpin loops determined in 1.0 M Na^+ to analyze DNA stability under more physiological conditions.

Programs that predict secondary structure and thermodynamics for nucleic acids frequently assign a given free energy for a particular structural motif, such as a tandem mismatch, regardless of the sequence of that mismatch. In this work, mismatches were chosen based on the idea that they would cover a range of stabilities. In examining the free energy contributed by the tandem mismatch alone, it was found that there is a range of just over 5.5 kcal / mol for the five mismatch molecules in 1.0 M Na^+. Similar results were seen in the other two buffer conditions studied. By changing only the four bases of the tandem mismatch, the free energy of the molecule may change by over 5 kcal / mol. This is particularly significant given that the free energy range for the entire hairpin molecules is from approximately 2 kcal / mol to 8 kcal / mol in 1.0 M Na^+.

Little data has been published regarding thermodynamic parameters for tandem mismatches. A paper by Li and Agrawal compared a reference duplex molecule of the sequence 5'GCGAGC3' / 5'GCTCGC3' to a sample molecule of the sequence 5'GCGAGC3' / 5'GCGAGC3', which contained tandem G-A mismatches with closing base pairs like those studied here. The studies were performed in 1.0 M Na^+ [44]. By subtracting stacking energies for the interrupted base pairs, values were determined for the remainder of the molecule. From this, thermodynamic values for the loop were obtained. From the Li and Agrawal results, ΔH_{loop} = -23.6 kcal / mol, ΔS_{loop} = -63.4 cal/mol-K, and ΔG_{loop} = -3.8 kcal / mol. Their results agree very well with the 1.0 M Na^+ parameters presented here of ΔH_{loop} = -25.5 kcal / mol, ΔS_{loop} = -70 cal/mol-K, and ΔG_{loop} = -3.79 kcal / mol.

Optical studies on nucleic acids serve as an indirect means of determining thermodynamic parameters because change in absorbance is being measured, rather than heat. Van't Hoff analyses performed on UV data also rely on the assumption that a

molecule follows a two-state model. In order to obtain a direct measurement of heat, and eliminate the assumption of two-statedness, differential scanning calorimetry experiments were performed. Although experiments to determine thermodynamic parameters are traditionally carried out in 1.0 M Na^+, the high melting temperatures of the molecules in this salt made DSC studies difficult. For this reason, the molecules were examined by UV and DSC in 0.1 M Na^+. Each method gave similar thermodynamic results. In general, researchers compare curve fit and peak height methods of van't Hoff analysis and if the results fall within 15% of one another, the molecules are taken to be folding in a two-state manner. Here, the two indirect methods of van't Hoff analysis and the direct method of DSC were compared. The methods gave results within 15% of one another with the exception of the tc/tc molecule. For this hairpin the van't Hoff and DSC methods gave results within 20% of one another. It is therefore reasonable to say that the short hairpins with tandem mismatches studied here are folding in a two-state manner.

Determining change in heat capacity for hairpin DNA's

It has generally been assumed that for short nucleic acid molecules, there is little change in heat capacity when going from a duplexed state to a coil state. Recent papers have addressed this issue and shown that there is a change in heat capacity which, although small, can have a significant effect on thermodynamic parameters when extrapolated over a long temperature range [64, 65, 97]. When molecules have melting temperatures that are much higher than 37 °C, such as the hairpin molecules studied here, if there is a change in heat capacity it can have a significant effect on the resulting thermodynamic parameters. In hairpins it is not possible to use the same methods to determine change in heat capacity as can be used for duplexes. In duplexes, the higher

the strand concentration, the higher the T_m, which allows thermodynamic parameters to be evaluated at different temperatures.

One method for determining change in heat capacity in hairpins is to use the molar heat capacity curves from differential scanning calorimetry. The pre- and post-transition baselines may be extrapolated to the midpoint of the transition and any difference between the baselines is the ΔCp (figure 3.6). This method was used for each of the hairpins studied here. The results from this evaluation were not always reproducible, but in general indicated very small, if any, change in heat capacity. One possible explanation is there is no change in heat capacity for the hairpins. Previous studies on ΔCp have been carried out with duplex molecules, and it is not clear whether hairpins would have a similar change in heat capacity. Another possible explanation is the instrument used for the calorimetry studies is not sensitive enough to detect any changes in heat capacity. This method of determining ΔCp relies on accurate baselines and large signal-to-noise ratios. If the instrument is not sensitive enough to obtain reliable baselines, an accurate ΔCp may not be determined.

It is possible to calculate a ΔCp for the hairpins by utilizing results obtained from the control molecules. The thermodynamic parameters for completely Watson-Crick paired bases have been determined fairly accurately. By comparing the enthalpy values determined for the control molecules at the transition midpoints to enthalpies estimated at 37 °C, a calculation may be made for ΔCp. The *mfold* program [41], by Michael Zuker, provides secondary structure predictions and estimated thermodynamic parameters for single-stranded nucleic acids. The assumption must be made that the base pairs in the Watson-Crick paired stem provides the only enthalpic contribution, and that there is no enthalpic contribution from the hairpin loop. The calculation is performed using the following equation:

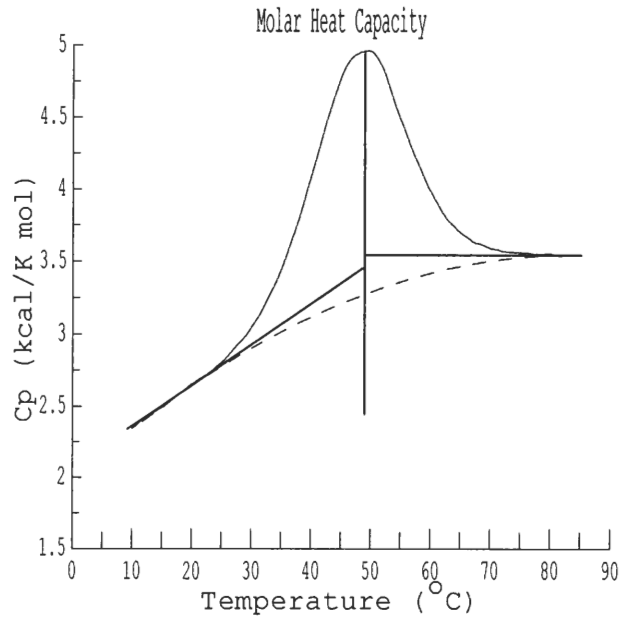


Figure 3.6. Determination of change in heat capacity from DSC data. The pre- and post-transition baselines are extrapolated to the midpoint of the transition, represented by the vertical line. The difference between the two baselines at the midpoint is the change in heat capacity.

$$\Delta Cp = - [(\Delta H_{observed} - \Delta H_{Zuker}) / (N) (T_m - 310)] \quad (3.1)$$

where ΔCp represents the change in heat capacity for the formation of the duplex, ΔH_{Zuker} is the enthalpy determined by the *mfold* program at 37 °C, $\Delta H_{observed}$ is the experimental enthalpy for a control molecule in this work, N is the number of paired bases in the hairpin stem, and T_m is the experimental melting temperature for the molecule in Kelvin. 310 K is 37 °C.

When this calculation was carried out for each of the control hairpins in 1.0 M Na^+, a similar ΔCp was determined for each. For the --/-- molecule, the ΔCp is -43 cal / mol-K per base pair. The A/T molecule gives a change in heat capacity of -33 cal / mol-K per base pair. Both the GA/TC and AA/TT molecules give a ΔCp of -38 cal / mol-K per base pair. The average change in heat capacity for the formation of the four control molecules is -38 cal / mol-K per base pair. This fits well with previously published experimental estimates of ΔCp for RNA and DNA duplexes, which range from 0 to -100 cal / mol-K per base pair [65].

Each of the tandem mismatch molecules may then be adjusted to reflect this change in heat capacity using the equation:

$$\Delta H_{37} = \Delta H_{observed} + \Delta Cp (N) (T_m - 310) \quad (3.2)$$

ΔH_{37} represents the enthalpy adjusted to 37 °C for the change in heat capacity. ΔCp is the change in heat capacity for formation of the molecule, here calculated to be -38 cal / mol-K per base pair. N is the number of base pairs in the stem, including the mismatches bases. T_m and $\Delta H_{observed}$ are the experimental values for the melting temperatures and enthalpies of the mismatch molecules.

When this is done, the changes in enthalpy are significant for the hairpin molecules. The corrected enthalpy for the ga/ga molecule is –82.4 kcal / mol, compared to the original value of –66.9 kcal / mol, a difference of 23%. For the aa/gc molecule, the corrected enthalpy is –59.2 kcal / mol, compared to –47.2 kcal / mol originally. This represents a 25% difference. The ca/gc and ta/ac molecules also showed differences of about 25% between the corrected and original values. For the ca/gc molecule, the corrected ΔH is -49.4 kcal / mol while the original was -39.6 kcal / mol. For the ta/ac molecule the corrected and original enthalpies were -40.8 kcal / mol and –32.6 kcal / mol. For the tc/tc molecule the corrected enthalpy was –47.9 kcal / mol compared to the original –41.4 kcal / mol, a difference of 16%. From these results, it is apparent that the change in heat capacity can have a significant effect on thermodynamic parameters, particularly for those molecules with T_m's much higher than 37 °C.

CHAPTER 4

STUDIES ON RNA TANDEM MISMATCHES IN SHORT HAIRPINS AND DUPLEXES

Introduction

The previous chapter discussed results on tandem mismatches in DNA hairpins. It is important to understand the thermodynamic parameters of structural elements in RNA as well. Because RNA is often present in cells as single-stranded molecules with helical regions, there is a great deal more secondary structure in RNA. RNA also forms tertiary structures with interactions between secondary elements. It is thought that a given RNA folds into secondary structure prior to tertiary structure, making an accurate understanding of thermodynamics important for predictions of both types of structures. The stabilities of different RNA structures are important for understanding their function, and for applications such as RNA interference (RNAi).

In this chapter, studies on RNA tandem mismatches will be discussed. First, a series of short RNA hairpins similar in sequence to the DNA hairpins from the previous chapter were designed. These hairpins were studied by UV denaturation to determine the free energy of tandem mismatches with various sequences.

Secondly, a series of short RNA duplexes with tandem mismatches were designed. The tandem mismatches consisted of one U-U mismatch adjacent to another non-canonical base pair. In previous studies utilizing TGGE to estimate free energies for a variety of RNA tandem mismatches [45], the results for tandem mismatches containing a U-U mismatch differed significantly from values estimated by Turner *et al.* [103]. The

latter values were based on parameters determined for a limited number of RNA tandem mismatches and assumptions that extrapolated this data to unexamined sequences. For each of the tandem mismatches chosen, two sets of closing base pairs were examined in order to evaluate the effects of context on the stability of the loops.

Methods

RNA hairpins containing tandem mismatches (figure 4.1) were studied by UV denaturation. The sequence of the hairpins is 5'-GACXYGCGAUCUACGCZWGUC-3', where X, Y, Z, and W represent mismatched bases. Samples were prepared to have a starting absorbance of 0.3 to 0.4 absorbance units at 268 nm. Experiments conducted in 1.0 M sodium buffer (0.978 M NaCl, 0.01 M Na_2HPO_4, 1 mM Na_2-EDTA) were heated over an appropriate temperature range, generally 25 °C – 95 °C or 15 °C – 90 °C. Experiments conducted in 0.1 M sodium buffer (0.078 M NaCl, 0.01 M Na_2HPO_4, 1 mM Na_2-EDTA) were carried out over a temperature range of 13 °C – 85 °C. The heating rate was 1°C per minute. The transition for all molecules was monitored at 268 nm. Several molecules were also monitored at 280 nm to ensure agreement of results at the two wavelengths.

RNA duplexes containing tandem mismatches were studied. The first set (set 1) of molecules utilized two single-stranded molecules with the sequences 5' CGAGXYAGGC 3' (top strand) and 5' GCCUWZCUCG 3' (bottom strand). This provided a G-C and an A-U closing base pair surrounding the mismatch. The second set (set 2) of molecules utilized single-stranded RNA's with the sequences 5' CCUCXYGUGA 3' (top strands) and 5' UCGCWZGAGG 3' (bottom strands). For these molecules the closing base pairs were a C-G and a G-C. Some strands were used to generate more than one duplex. Table 4.1 shows the sequences of each of the single stranded molecules used and their designated name. Each molecule is named

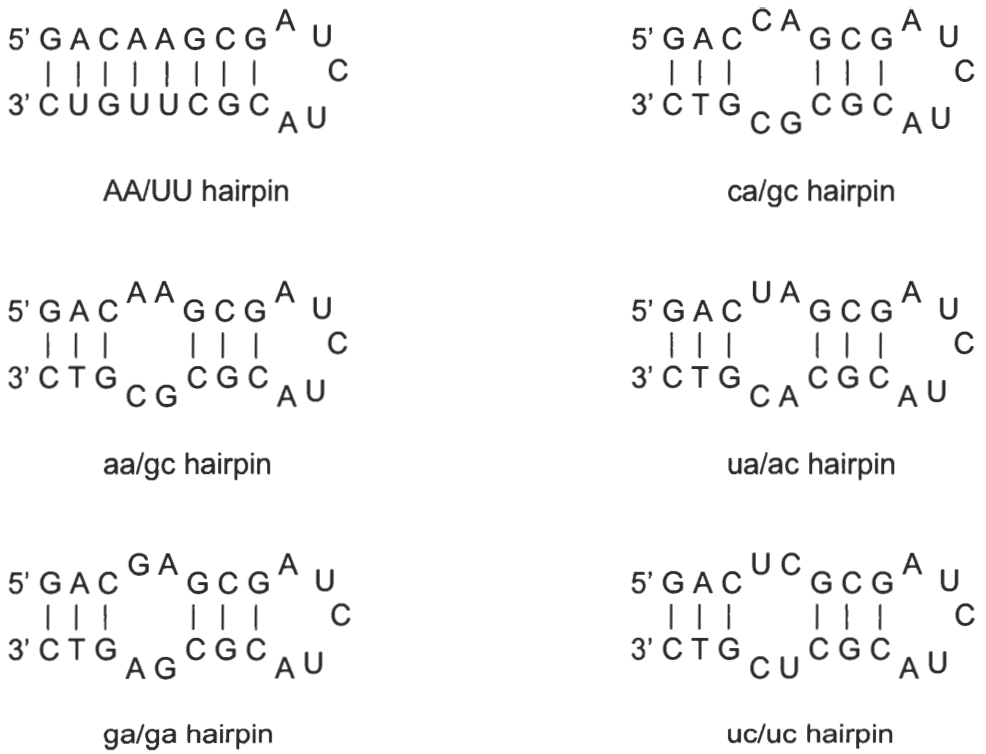


Figure 4.1. Expected structure of short RNA hairpins containing tandem mismatches. Each hairpin is characterized by a stem consisting of six Watson-Crick base pairs interrupted by two mismatch pairs and a hairpin loop of five bases. The control molecule (AA/UU) has two additional Watson-Crick base pairs in place of the tandem mismatches for a total of eight Watson-Crick base pairs in the stem.

Table 4.1: Sequences of single-stranded molecules used to form duplexes.

Name	Sequence	Name	Sequence
1-B-AU	5' GCCUAUCUCG	2-B-AU	5' UCGCAUGAGG
1-B-UA	5' GCCUUACUCG	2-B-UA	5' UCGCUAGAGG
1-B-UC	5' GCCUUCCUCG	2-B-UC	5' UCGCUCGAGG
1-T-AU	5' CGAGAUAGGC	2-T-AU	5' CCUCAUGUGA
1-T-UC	5' CGAGUCAGGC	2-T-UC	5' CCUCUCGUGA
1-T-UA	5' CGAGUAAGGC	2-T-UA	5' CCUCUAGUGA
1-T-GU	5' CGAGGUAGGC	2-T-GU	5' CCUCGUGUGA
1-T-CU	5' CGAGCUAGGC	2-T-CU	5' CCUCCUGUGA
1-T-UU	5' CGAGUUAGGC	2-T-UU	5' CCUCUUGUGA

according to the set, strand, and bases involved in the mismatches. For example, 1-B-AU is a molecule from set 1 (containing G-C and A-U closing base pairs) that served as a bottom strand, with the A and U in positions 5 and 6 involved in the mismatches. Underlined bases represent the portions of the molecule that form the mismatched bases.

To form duplexes, the single strands were mixed in equimolar concentrations to give a concentration of 5 μM duplex. The helix-coil transition of each duplex molecule was examined in 1.0 M sodium buffer. The temperature range for the molecules with a G-C and A-U as the closing base pairs around the mismatches was 15 °C – 75 °C or 15 °C – 85 °C in the case of the fully duplexes molecule. The temperature range for the duplexes with a C-G and G-C as the closing base pairs was generally 5 °C – 65 °C. The completely duplexed RNA's with these closing base pairs was heated over a range of 15 °C – 85 °C. The heating rate was 1 °C per minute and the wavelength at which the transitions were monitored was 268 nm.

Two RNA duplexes were also utilized to estimate the change in heat capacity between the duplex and coil states. Duplex concentrations ranging from 1 μM to 75 μM were employed. The thermodynamic parameters for each sample were determined assuming a two-state transition using the peak height methods (chapter 3). Plots were then made of T_m (°C) vs. $-\Delta H°$ (kcal / mol-duplex) and of $\ln(T_m)$ vs. $-\Delta S°$ (cal / K-mol-duplex). ΔCp determined by the enthalpy change was derived from a linear regression of the enthalpy change with respect to the melting temperature ($\Delta Cp_H = d\Delta H/dT_m$). ΔCp as it relates to entropy change was derived from a linear regression of the entropy change with respect to the logarithmic scale of the melting temperature ($\Delta Cp_S = d\Delta S/d\ln T_m$) [97].

Thermodynamic evaluation of duplexes with two transitions

Three duplexes exhibited two transitions. Based on experimental results of the individual strands, the first step of these transitions is interpreted to represent a duplex to hairpin and single strand transition (equation 4.1). The second step is believed to correspond to the hairpin to single strand transition. The experimental transitions were analyzed by comparing them with a theoretical curve representing the coupled equilibria. It was assumed that upon duplex denaturation, 100% of the strand that could form a hairpin did so. The overall transition is represented by:

$$S_1 \cdot S_2 \underset{K_2}{\rightleftarrows} H_1 + S_2 \underset{K_1}{\rightleftarrows} S_1 + S_2 \qquad (4.1)$$

where $S_1 \cdot S_2$ is the duplex formed by strand 1 and strand 2, H_1 is the hairpin formed by strand 1, S_2 is the single-stranded form of strand 2, and S_1 represents the single-stranded form of strand 1. K_1 and K_2 are the temperature-dependent equilibrium constants for the formation of the hairpin and the duplex (the reverse reactions of the two steps of equation 4.1). The fraction of molecules that are duplex θ_2 is given by:

$$\theta_2 = \frac{S_1 \cdot S_2}{S_{T2}} = \frac{K_2 K_1 (1 - \theta_2) S_{T2}}{K_2 K_1 (1 - \theta_2) S_{T2} + (1 + K_1)} \qquad (4.2)$$

where $S_1 \cdot S_2$ is the concentration of the duplex strand and S_{T2} represents the total concentration of strand 2. The fraction of molecules that are hairpins θ_1 is given by:

$$\theta_1 = \frac{H_1}{S_{T1}} = \frac{K_1 (1 - \theta_2)}{1 + K_1} \qquad (4.3)$$

where H_1 is the concentration of the hairpin and S_{T1} is the concentration of strand 1 available to be a hairpin. Solving these coupled equations requires initial estimates of K_1 and K_2, and thus estimates of the ΔH and T_m for each transition. For the hairpin transition, the ΔH_1 and T_{m1} were determined through UV-monitored denaturation studies of the hairpin alone. An initial estimate of T_{m2} for the duplex to hairpin H_1 and single strand S_2 transition was determined from the experimental melting curve. An initial ΔH_2 value was based on the enthalpy change for the duplex to strand transition of similar molecules. These estimates were iteratively refined by comparing the normalized melting curve $A_N(T)$ to theoretically calculated curves using equations 4.2, 4.3, and the following equation:

$$A_N(T) = \frac{A(T) - A_{dup}(T)}{A_{ss}(T) - A_{dup}(T)} = (1 - \theta_2) - \beta\,\theta_1 \qquad (4.4)$$

where $A_N(T)$ is the normalized absorbance of the sample at temperature T. A(T) is the total absorbance at temperature T. $A_{dup}(T)$ and $A_{ss}(T)$ represent the pre- and post-transition linear baselines of the denaturation curves. β is the ratio of the absorbance change of the hairpin to the absorbance change overall (see figure 4.6, p.96). Values input for ΔH_1, ΔH_2, T_{m1}, and T_{m2} were varied to give the best fit to the experimental curves.

Results

Figure 4.1 (p. 67) shows the predicted structure of each of the RNA hairpins studied. These molecules are characterized by a stem of six Watson-Crick base pairs interrupted by two mismatch pairs and a hairpin loop consisting of 5 bases. In the

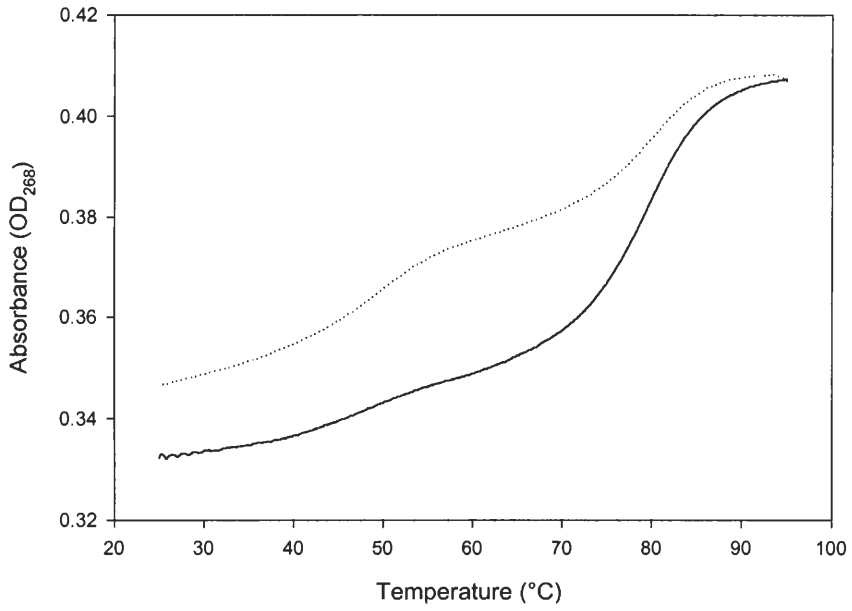
control molecule, the stem has two additional Watson-Crick base pairs in place of the tandem mismatches for a total of eight paired bases. Each molecule is named according to the mismatches in the 5' to 3' direction following the example XY/WZ. In this work, mismatch molecules will be referred to by lower case letters while the control molecule will be referred to by capital letters.

Figure 4.2 provides typical melting curve profiles for each of the hairpins studied in 1.0 M Na^+ buffer with solid lines representing the denaturation curves and dotted lines showing the renaturation curves. In general, the denaturation and renaturation were similar. Although some differences are seen between the starting and ending points for several of the curves, this is probably due to slight evaporation. Often mineral oil is used to cover samples to prevent evaporation however, because mineral oil is not guaranteed to be ribonuclease-free, none was used for RNA samples.

The denaturation and renaturation curves of the AA/UU RNA show two transitions in 1.0 M Na^+, unlike the homologous DNA sequence (Chapter 3). This behavior is consistent with the presence of duplex molecules with 16 base pairs and an internal loop of five mispairs at the lower temperatures. Based on previous studies of nucleic acid oligomers with partial self-complementary sequences, the first transition is likely due to the duplex to hairpin conversion and the second transition corresponds to unwinding RNA hairpins to single strands [104]. This result illustrates that the relative contribution of base pair stacking and internal and hairpin loops to the conformational free energy of hairpin versus duplex forms differs for homologous RNA and DNA sequences. Thermodynamic parameters were evaluated from the second transition assuming it reflects the hairpin to single strand transformation.

Except for the RNA with the tandem G-A mismatches, the RNA melting curves in 1.0 M Na^+ show monophasic transitions. Although not as apparent as the AA/UU RNA

A.



B.

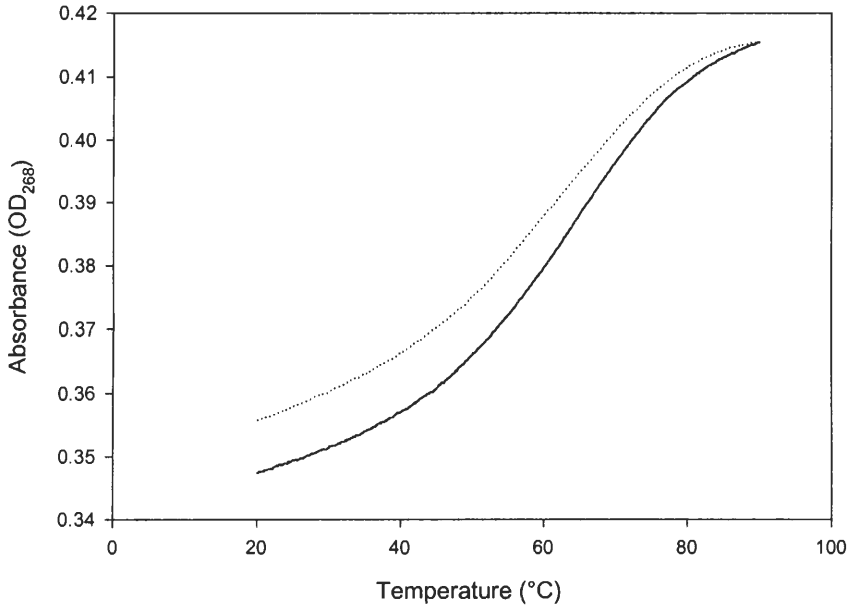
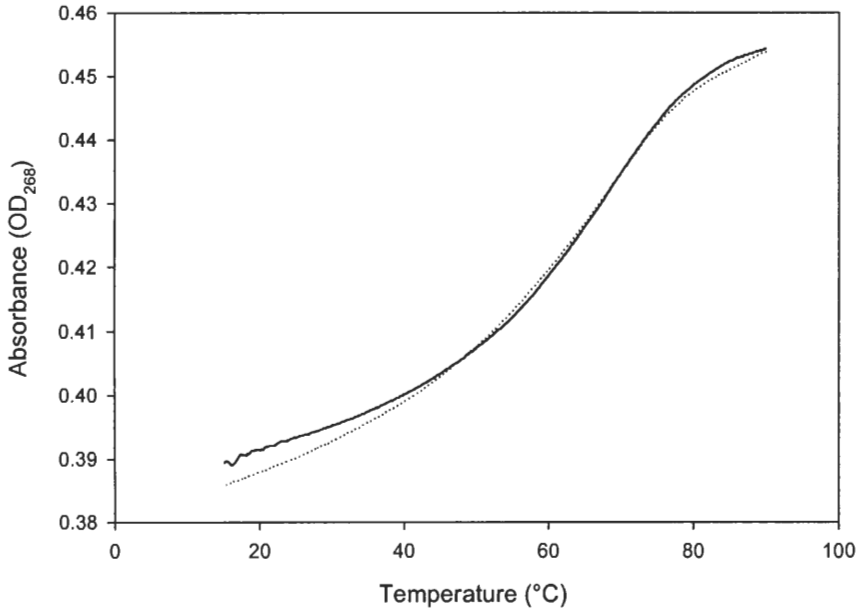


Figure 4.2. Typical UV denaturation curves of RNA hairpins containing tandem mismatches in 1.0 M Na^+. Denaturation curves are shown as solid lines and renaturation curves are represented as dotted lines. A. AA/UU molecule. B. aa/gc molecule.

C.



D.

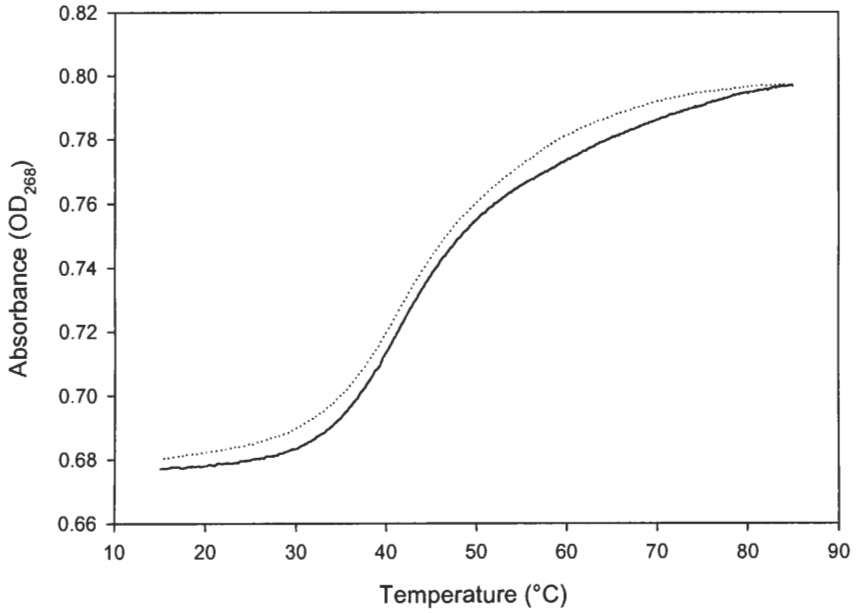
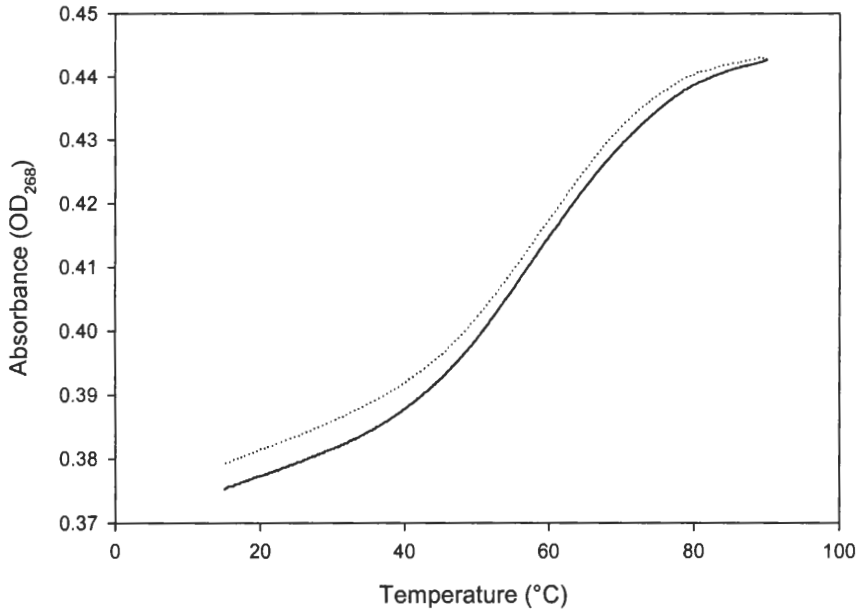


Figure 4.2 (continued). Typical UV denaturation curves of RNA hairpins containing tandem mismatches in 1.0 M Na^+. Denaturation curves are shown as solid lines and renaturation curves are represented as dotted lines. C. ca/gc molecule. D. ga/ga molecule.

E.



F.

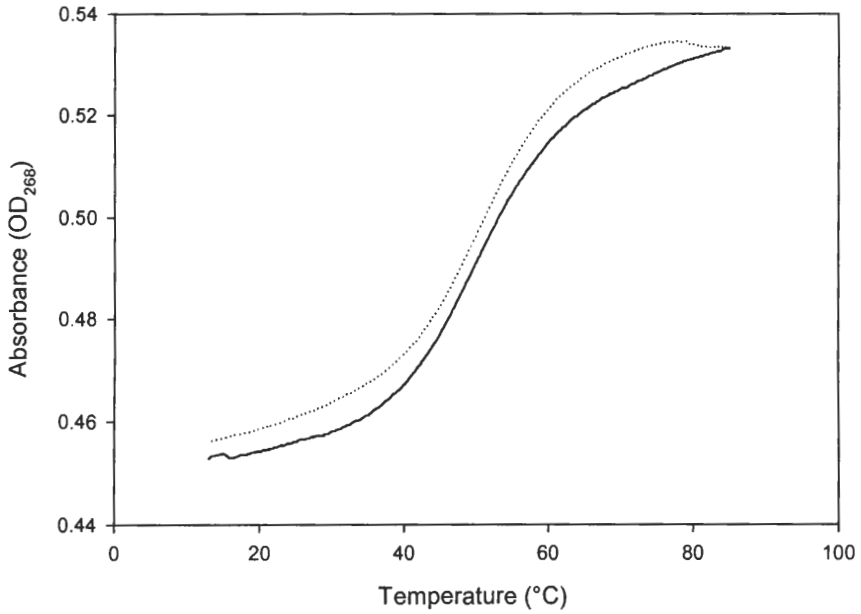


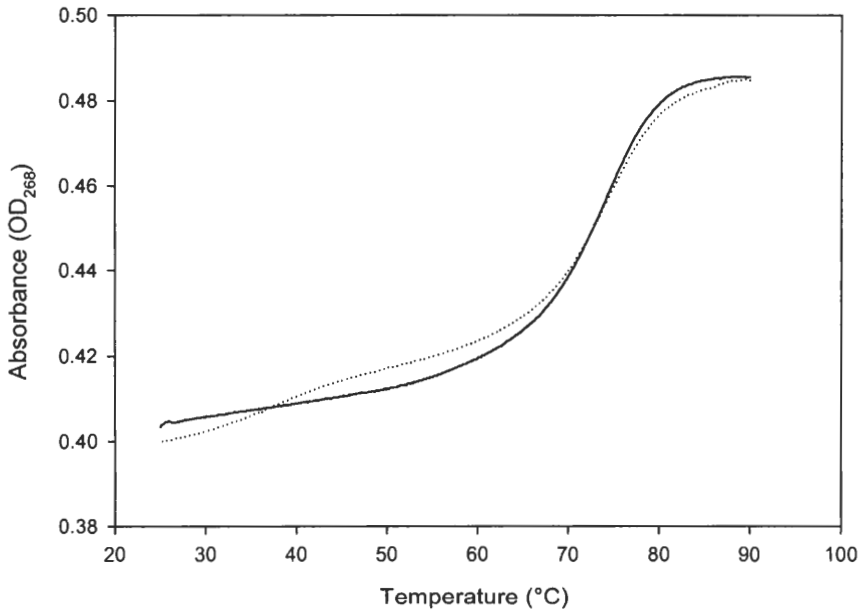
Figure 4.2 (continued). Typical UV denaturation curves of RNA hairpins containing tandem mismatches in 1.0 M Na^+. Denaturation curves are shown as solid lines and renaturation curves are represented as dotted lines. E. ua/ac molecule. F. uc/uc molecule.

transition, the denaturation curve of the ga/ga RNA (figure 4.2D) reproducibly showed a break in the transition slope around 50 °C. This may indicate an overlap of duplex to hairpin and hairpin to single strand transitions. Previous studies on tandem G-A mismatches in RNA indicate it is a relatively stable tandem mismatch [48]. The two sets of tandem mismatches that would occur for a duplex may be sufficiently stable to enable this state.

Due to the uncertainty regarding the nature of the ga/ga RNA transition, evaluation of thermodynamic parameters was not carried out for this RNA in 1.0 M Na^+. The normalized denaturation curves of the other RNA's were analyzed assuming they correspond to the hairpin to single strand transition. Denaturation curves of the ca/gc RNA hairpin obtained at both 0.8 uM and 40 uM concentrations gave similar T_m values, indicating this assumption is valid. At the lower strand concentration, the melting temperature was found to be 55.9 °C. At the higher strand concentration the T_m was 55.4 °C. If significant amounts of duplex were present, the T_m would be expected to shift by several degrees.

Figure 4.3 shows typical denaturation and renaturation curves for each of the RNA molecules in 0.1 M Na^+. Again there are small differences in the starting and ending absorbances due to evaporation of the solvent, however the denaturation and renaturation curves of most molecules give similar transition curves after normalizing between 0 and 1. In some cases there was a shift in T_m by approximately 1.5 °C, however the overall shape of the transition remained the same. This may indicate the presence of small amounts of duplex, however, one of the molecules that showed a shift was the ca/gc molecule. This was described above as having a constant melting temperature over a 50-fold concentration difference, indicating that if duplex is present, it is in very small amounts.

A.



B.

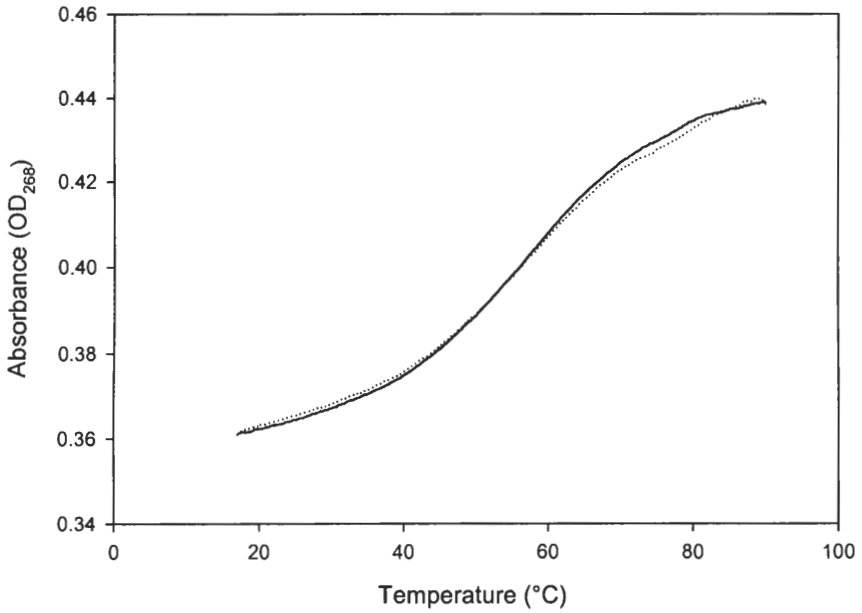
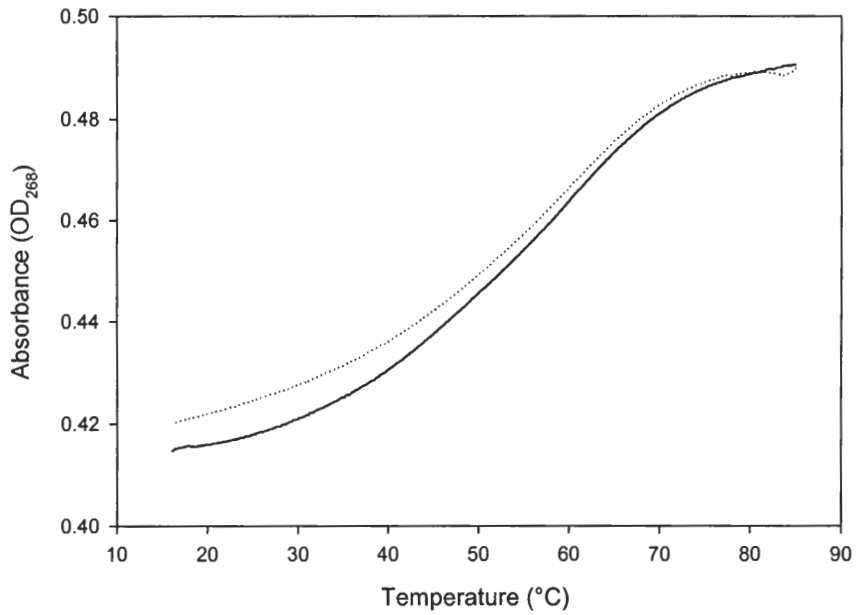


Figure 4.3. Typical UV denaturation curves of RNA hairpins containing tandem mismatches in 0.1 M Na^+. Denaturation curves are shown as solid lines and renaturation curves are represented as dotted lines. A. AA/UU molecule. B. aa/gc molecule.

C.



D.

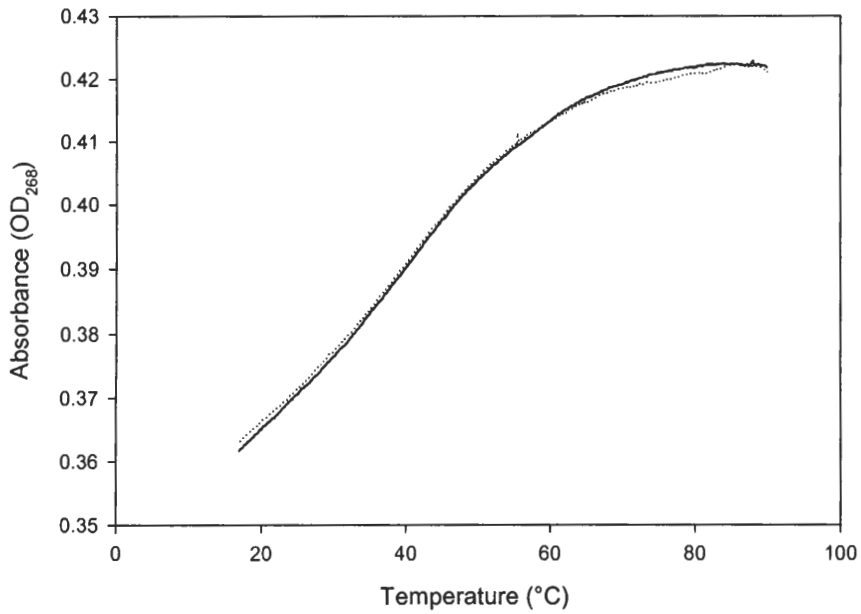
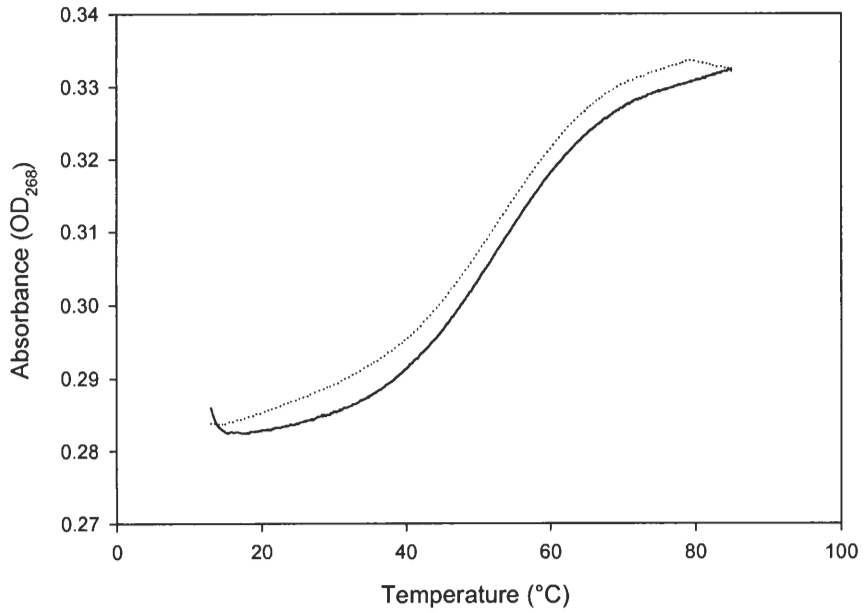


Figure 4.3 (continued). Typical UV denaturation curves of RNA hairpins containing tandem mismatches in 0.1 M Na^+. Denaturation curves are shown as solid lines and renaturation curves are represented as dotted lines. C. ca/gc molecule. D. ga/ga molecule.

E.



F.

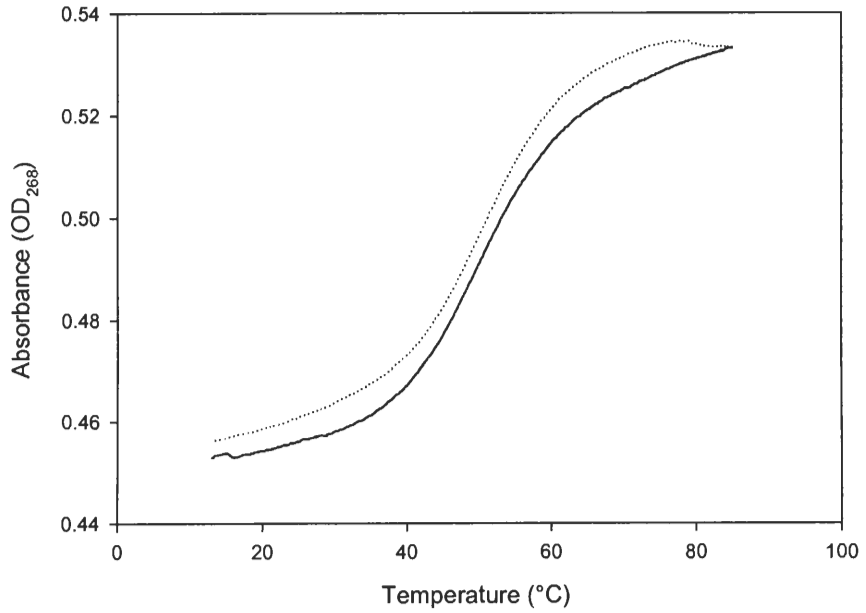


Figure 4.3 (continued). Typical UV denaturation curves of RNA hairpins containing tandem mismatches in 0.1 M Na^+. Denaturation curves are shown as solid lines and renaturation curves are represented as dotted lines. E. ua/ac molecule. F. uc/uc molecule.

Two transitions are apparent in the renaturation curve of the AA/UU molecule but not in the denaturation curve in 0.1 M Na^+. Lower salt concentrations are known to favor the hairpin state [104], and the hairpin to duplex step in the renaturation curve is indeed relatively small when compared to the 1.0 M Na^+ transition curves. In 0.1 M Na^+, two distinct folded conformations are apparent only when the AA/UU molecules are cooled slowly. For this reason, thermodynamic analysis was carried out on the initial denaturation curves of AA/UU RNA molecules. It was assumed that the absorbance change from 40 °C to 90 °C represented the hairpin to single strand transition.

The transition of the ga/ga molecule in 0.1 M Na^+ appears to begin at a temperature below the starting temperature shown in figure 4.3D. Due to difficulties in obtaining absorbance data below 15 °C, the conventional analysis was not carried out. Another approach may be employed for transitions where it is difficult to define the lower baseline. A derivative curve was obtained on the smoothed absorbance data, which was not normalized from 0 – 1. A plot was made for (dA) / (d 1/T) versus temperature. The following equation was then used to obtain an enthalpy value [96]:

$$\Delta H_{VH} = B' / [(1 / T_{max}) - (1 / T_2)] \qquad (4.5)$$

B' is a constant with the value –3.5 for a monomolecular transition. T_{max} is the temperature at the maximum of the derivative curve. T_2 is the upper temperature at which the change in the observable Y-axis values is one-half of $[(dA) / (d\ 1/T)]_{max}$.

Table 4.2A and 4.2B provides the thermodynamic parameters for the formation of RNA hairpins containing tandem mismatches in both 1.0 M and 0.1M Na^+. As would be expected, the AA/UU molecule with all Watson-Crick base pairs in the stem has the highest melting temperature and is the most stable. The RNA molecules containing tandem mismatches had a range of melting temperatures of 55.2 °C to 64.1 °C in 1.0 M

Table 4.2A: Melting temperatures and thermodynamic parameters for formation of RNA hairpins containing tandem mismatches in 1.0 M Na^+.

Molecule	T_m (°C)	ΔH (kcal / mol)	ΔS (eu)	ΔG_{37} (kcal / mol)
AA/UU	79.5 ± 0.5	-74.6 ± 3.4	-211 ± 10	-9.00 ± 0.43
aa/gc	63.6 ± 0.7	-31.7 ± 2.0	-94 ± 6	-2.50 ± 0.14
ca/gc	64.1 ± 0.7	-30.3 ± 2.7	-90 ± 8	-2.44 ± 0.27
ga/ga	ND	ND	ND	ND
ua/ac	58.0 ± 0.4	-29.5 ± 1.3	-89 ± 4	-1.87 ± 0.09
uc/uc	55.2 ± 0.4	-33.7 ± 2.9	-103 ± 9	-1.86 ± 0.13

Table 4.2B: Melting temperatures and thermodynamic parameters for formation of RNA hairpins containing tandem mismatches in 0.1 M Na^+.

Molecule	T_m (°C)	ΔH (kcal / mol)	ΔS (eu)	ΔG_{37} (kcal / mol)
AA/UU	72.8 ± 0.1	-69.8 ± 2.6	-202 ± 7	-7.22 ± 0.27
aa/gc	56.3 ± 0.4	-35.2 ± 2.2	-107 ± 7	-2.06 ± 0.14
ca/gc	56.5 ± 0.8	-27.8 ± 1.1	-84 ± 3	-1.64 ± 0.11
ga/ga	40.6 ± 0.8	-17.8 ± 1.4	-57 ± 5	-0.20 ± 0.04
ua/ac	52.5 ± 0.4	-33.0 ± 1.2	-101 ± 4	-1.57 ± 0.03
uc/uc	50.1 ± 0.2	-40.6 ± 1.6	-126 ± 5	-1.64 ± 0.04

Na^+. In the lower salt concentration the range of melting temperatures was 50.1 °C to 56.5 °C.

Table 4.3 shows the thermodynamic parameters of loop formation for the RNA tandem mismatches in both salt conditions. In 1.0 M Na^+, the enthalpy is very similar for all molecules with a range of only 4.2 kcal / mol. The range of free energies for the different sequences was 0.64 kcal / mol in the high salt. The order of stability based on free energy for the tandem mismatches in 1.0 M Na^+ was as follows: CaaG/CgcG > CcaG/CgcG > CuaG/CacG > CucG/CucG. The ca/gc and aa/gc molecules showed similar free energy values, as did the ua/ac and uc/uc molecules.

For the lower salt buffer, the range of the enthalpies was much larger, ranging from 3.1 kcal / mol for the uc/uc molecule to 15.9 kcal / mol for the ca/gc molecule. This large range was compensated by the entropy, however, because the range of free energies in the lower salt was 0.49 kcal / mol, which is similar to the range in the 1.0 M Na^+ buffer. In the 0.1 M Na^+, the aa/gc molecule was the most stable based on free energy, while the three remaining molecules had nearly identical stabilities.

RNA duplex molecules with tandem mismatches

The expected structures of the 18 RNA duplexes containing a U-U mismatch adjacent to another non-canonical base pair are shown in figure 4.4. Two sets of molecules were synthesized with different closing base pairs in order to examine their effect on the stability of tandem mismatches. The first column shows the duplexes with closing base pairs of G-C and A-U. These molecules contain eight Watson-Crick base pairs interrupted by the mismatches. The second column shows the structure of the duplexes with a C-G and G-C as the closing base pairs. These molecules contain the tandem mismatches surrounded by a stem comprised of seven Watson-Crick base pairs and one G-U base pair. Because both closing base pairs of these molecules were

Table 4.3: Thermodynamic parameters for formation of tandem mismatches in the context 5' CXYG 3' / 5' CWZG 3'.

Molecule	1.0 M Na^+			0.1 M Na^+		
	ΔH (kcal/mol)	ΔS (eu)	ΔG_{37} (kcal/mol)	ΔH (kcal/mol)	ΔS (eu)	ΔG_{37} (kcal/mol)
aa/gc	15.2	45	1.38	8.4	27	1.84
ga/ga	ND	ND	ND	ND	ND	ND
ca/gc	16.6	49	1.44	15.9	50	2.26
ua/ac	17.4	50	2.01	10.7	33	2.33
uc/uc	13.2	36	2.02	3.1	8	2.26

Set 1

```
5' C G A G A U A G G C 3'
   | | | | | | | | | |
3' G C U C U A U C C G 5'
```

GAUA / UAUC

```
5' C G A G U U A G G C 3'
   | | | |     | | | |
3' G C U C C U U C C G 5'
```

GuuA / UucC

```
5' C G A G G U A G G C 3'
   | | | |     | | | |
3' G C U C A U U C C G 5'
```

GguA / UuaC

```
5' C G A G U C A G G C 3'
   | | | |     | | | |
3' G C U C U A U C C G 5'
```

GucA / UauC

```
5' C G A G A U A G G C 3'
   | | | |     | | | |
3' G C U C A U U C C G 5'
```

GauA / UuaC

Set 2

```
5' C C U C A U G U G A 3'
   | | | | | | | | | |
3' G G A G U A C G C U 5'
```

CAUG / CAUG

```
5' C C U C U U G U G A 3'
   | | | |     | | | |
3' G G A G C U C G C U 5'
```

CuuG / CucG

```
5' C C U C G U G U G A 3'
   | | | |     | | | |
3' G G A G A U C G C U 5'
```

CguG / CuaG

```
5' C C U C U C G U G A 3'
   | | | |     | | | |
3' G G A G U A C G C U 5'
```

CucG / CauG

```
5' C C U C A U G U G A 3'
   | | | |     | | | |
3' G G A G A U C G C U 5'
```

CauG / CuaG

Figure 4.4. Expected structure of RNA duplexes containing tandem mismatches. Each molecule is characterized by 8 Watson-Crick base pairs interrupted by two mismatched bases. Each tandem mismatch was examined in the context of two sets of closing base pairs, represented by the two columns and designated set 1 and set 2.

Set 1

```
5' C G A G   C U   A G G C 3'
   | | | |       | | | |
3' G C U C   A U   U C C G 5'
```

GcuA / UuaC

```
5' C G A G   A U   A G G C 3'
   | | | |       | | | |
3' G C U C   C U   U C C G 5'
```

GauA / UucC

```
5' C G A G   C U   A G G C 3'
   | | | |       | | | |
3' G C U C   C U   U C C G 5'
```

GcuA / UucC

```
5' C G A G   U A   A G G C 3'
   | | | |       | | | |
3' G C U C   U A   U C C G 5'
```

GuaA / UauC

Set 2

```
5' C C U C   C U   G U G A 3'
   | | | |       | | | |
3' G G A G   A U   C G C U 5'
```

CcuG / CuaG

```
5' C C U C   A U   G U G A 3'
   | | | |       | | | |
3' G G A G   C U   C G C U 5'
```

CauG / CucG

```
5' C C U C   C U   G U G A 3'
   | | | |       | | | |
3' G G A G   C U   C G C U 5'
```

CcuG / CucG

```
5' C C U C   U A   G U G A 3'
   | | | |       | | | |
3' G G A G   U A   C G C U 5'
```

CuaG / CauG

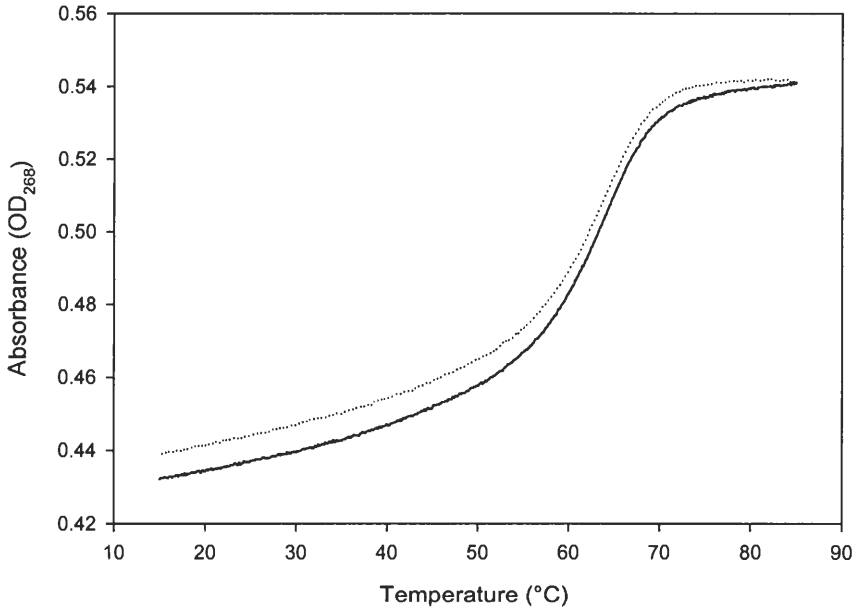
Figure 4.4. (continued) Predicted structure of RNA duplexes containing tandem mismatches. Each molecule is characterized by 8 Watson-Crick base pairs interrupted by two mismatched bases. Each tandem mismatch was examined in the context of two sets of closing base pairs, represented by the two columns and designated set 1 and set 2.

comprised of G-C, a G-U base pair was employed to prevent the duplex from having too high a melting temperature.

Figure 4.5 shows typical melting curves for each of the 18 RNA duplexes examined in 1.0 M Na^+. In general, the renaturation curves closely follow the denaturation curves. The melting temperatures for the molecules with the C-G and G-C closing base pairs was lower than the melting temperatures for the other set of molecules due to the G-U wobble base in the stem. Because the transition did occur at a lower temperature, the starting point for these curves is lower than for the other set of molecules.

The last three denaturation curves (figures 4.5P, Q, and R), for the CauG/CucG, CcuG/CucG, and CuuG/CucG molecules, are of considerable interest because they show two melting transitions. The distinction between these two-step transitions and the other RNA transitions is more apparent in figure 4.6, which compares normalized melting curves. The midpoints of the second transition occur at 47.2 °C, 46.3 °C, and 45.7 °C respectively. These three duplexes were the only ones that utilized the single-stranded molecule 2-B-UC, which has the sequence 5' UCGCUCGAGG 3'. This sequence can potentially fold into a hairpin structure shown in figure 4.7. This hairpin structure has a stem comprised of the U in position 1 paired with the A in position 8, the C in position 2 paired with the G in position 7, and the G in position 3 paired with the C in position 6. The hairpin loop has the two base sequence C – U from positions 4 and 5. Two G's in positions 9 and 10 would form a dangling end. Alternately, the sequence could form a duplex molecule with six Watson-Crick base pairs, two mismatches, and two dangling G's on each end. The higher T_m of the second transition suggests a hairpin molecule rather than a duplex, however it would need to be confirmed should further studies be carried out on this molecule.

A.



B.

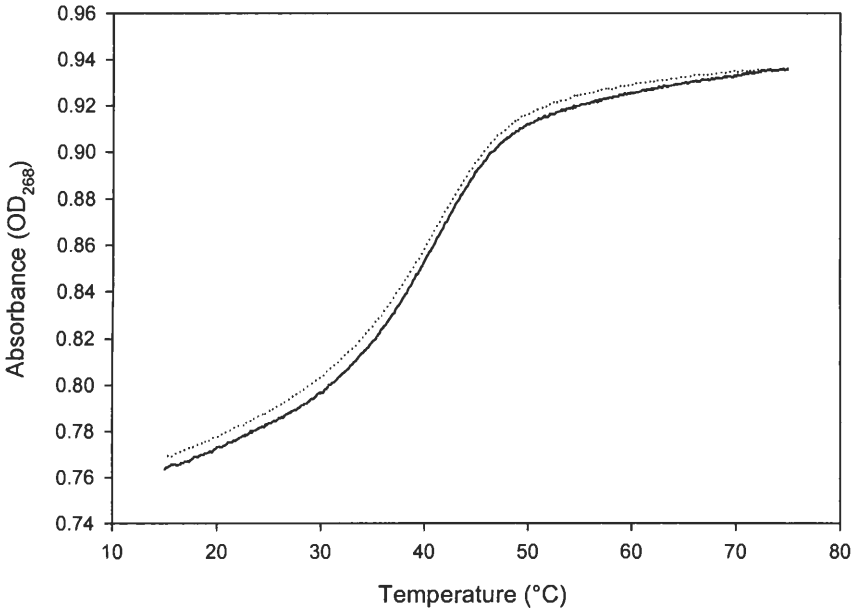
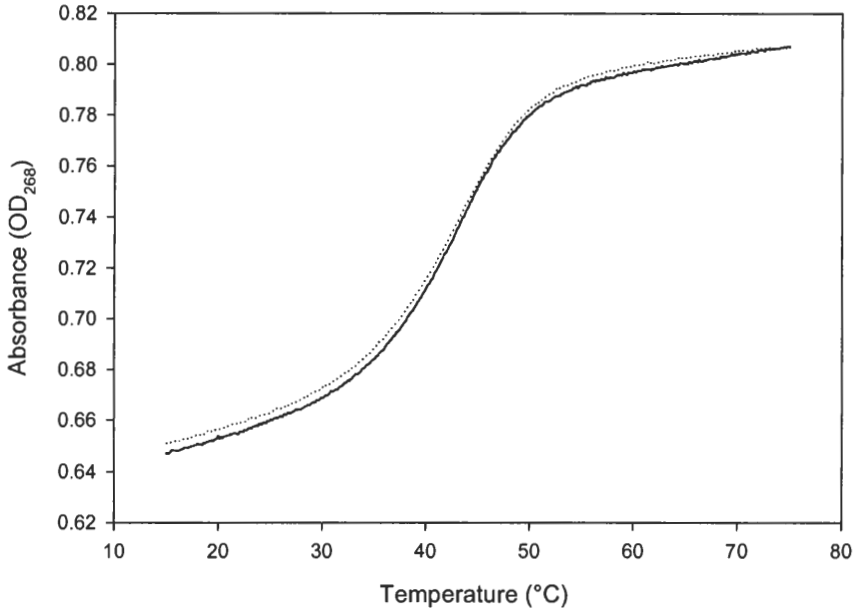


Figure 4.5. Typical UV denaturation curves for RNA duplexes containing tandem mismatches in 1.0 M Na^+. Solid lines represent denaturation curves. Dotted lines represent renaturation curves. A. GAUA/UAUC. B. GucA/UauC.

C.



D.

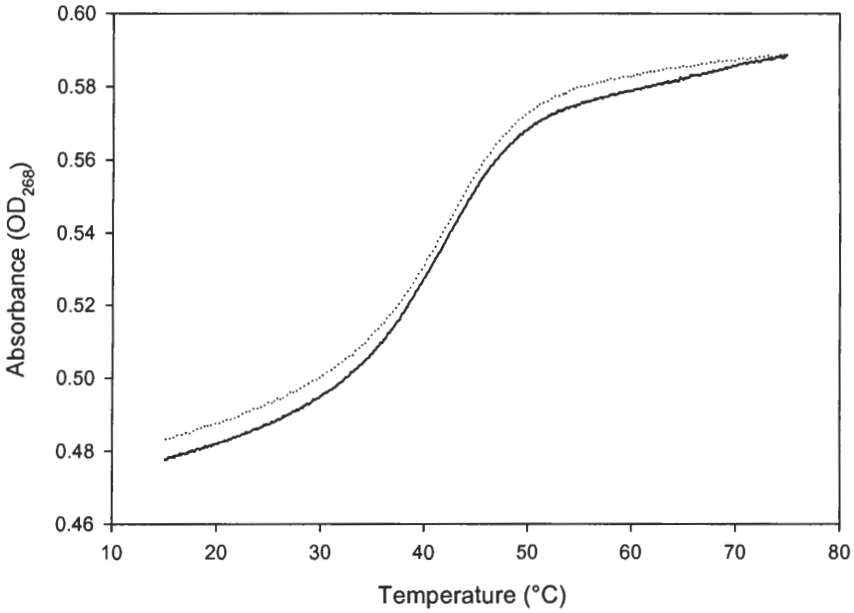
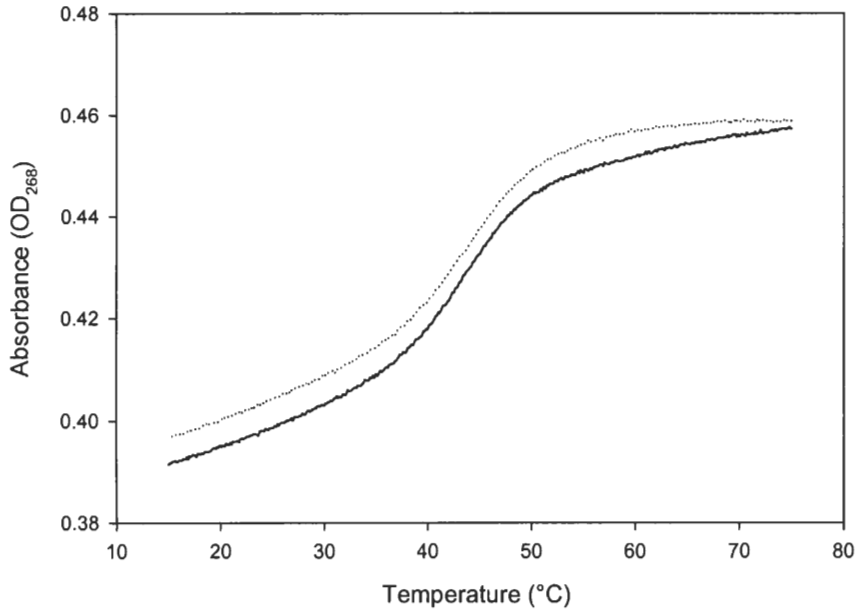


Figure 4.5 (continued). Typical UV denaturation curves for RNA duplexes containing tandem mismatches in 1.0 M Na^+. Solid lines represent denaturation curves. Dotted lines represent renaturation curves. C. GuaA/UauC. D. GauA/UuaC.

E.



F.

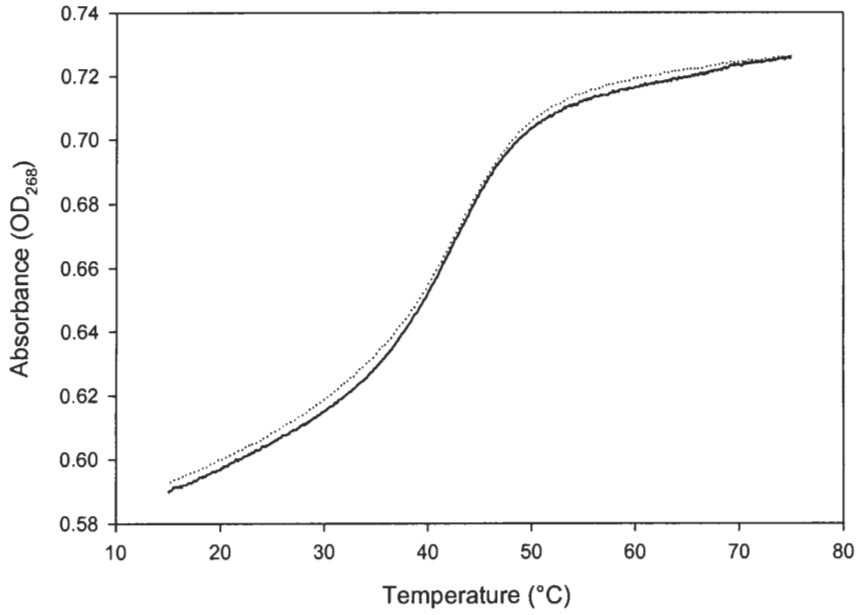
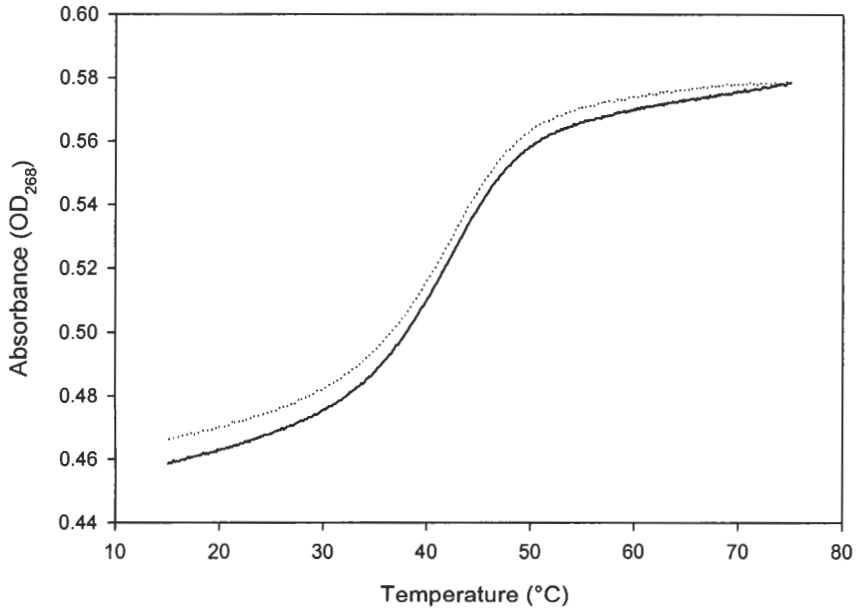


Figure 4.5 (continued). Typical UV denaturation curves for RNA duplexes containing tandem mismatches in 1.0 M Na^+. Solid lines represent denaturation curves. Dotted lines represent renaturation curves. E. GguA/UuaC. F. GcuA/UuaC.

G.



H.

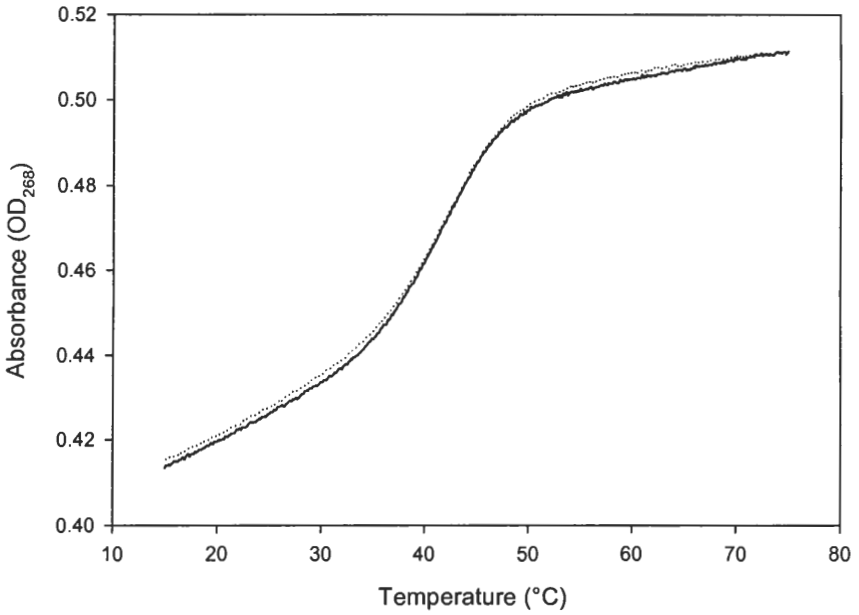
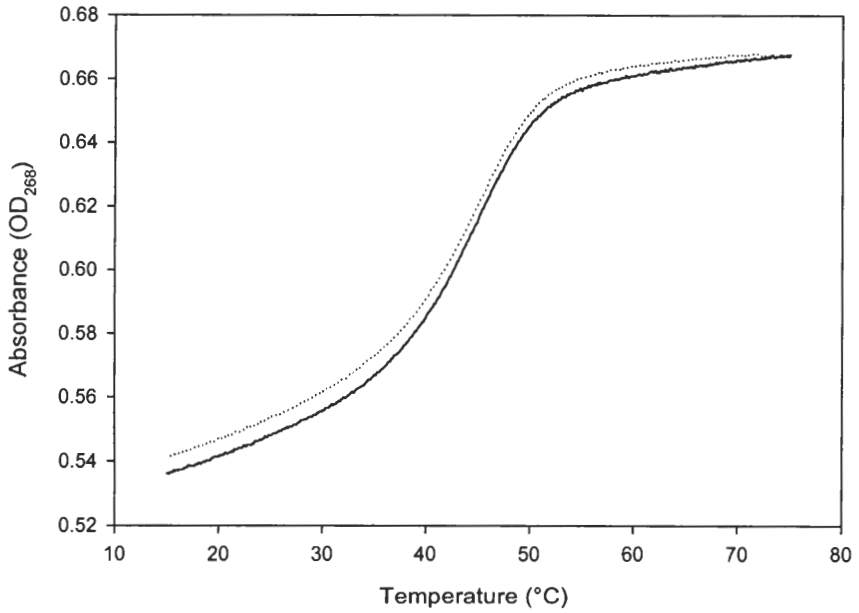


Figure 4.5 (continued). Typical UV denaturation curves for RNA duplexes containing tandem mismatches in 1.0 M Na^+. Solid lines represent denaturation curves. Dotted lines represent renaturation curves. G. GauA/UucC. H. GcuA/UucC.

I.



J.

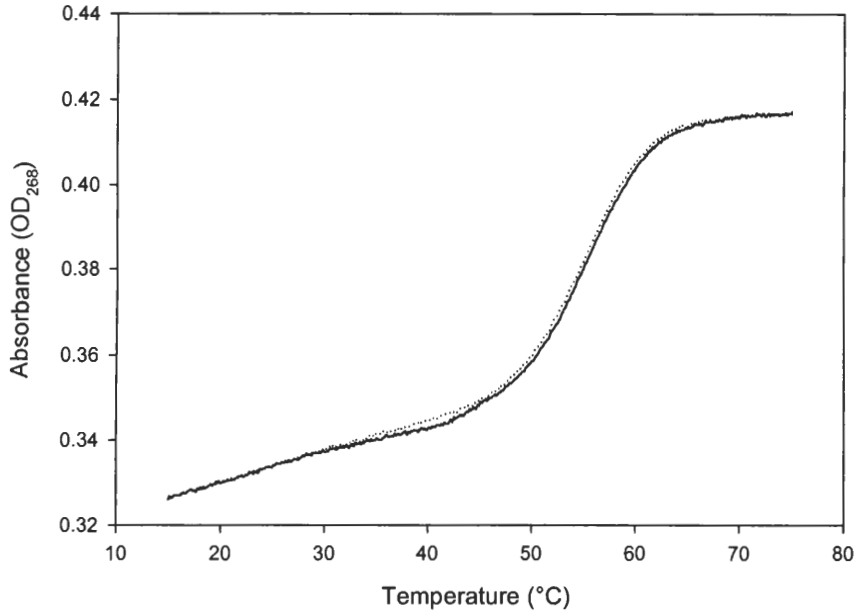
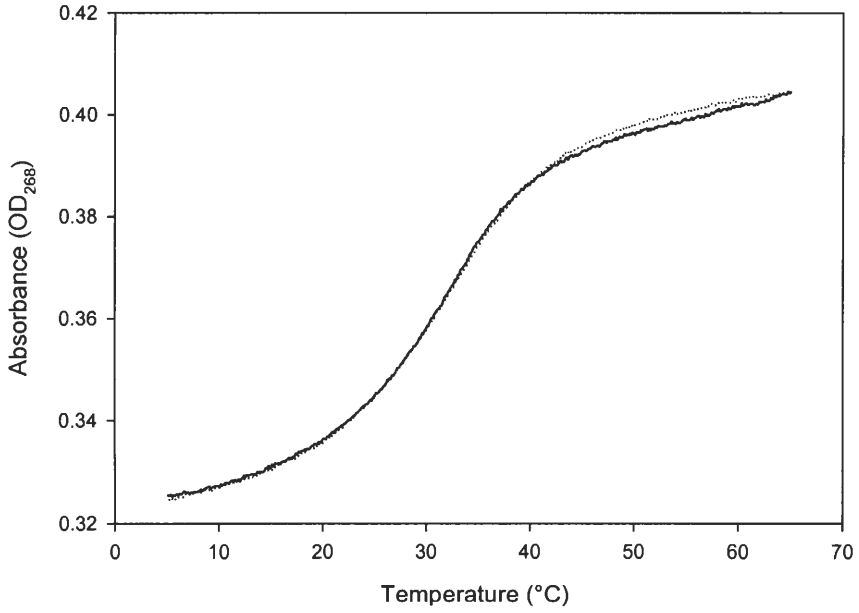


Figure 4.5 (continued). Typical UV denaturation curves for RNA duplexes containing tandem mismatches in 1.0 M Na^+. Solid lines represent denaturation curves. Dotted lines represent renaturation curves. I. GuuA/UucC. J. CAUG/CAUC.

K.



L.

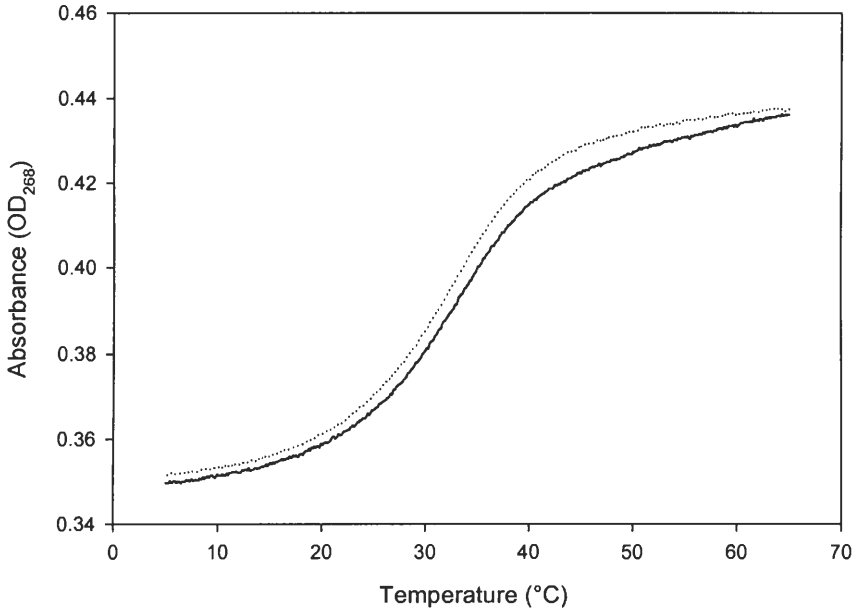
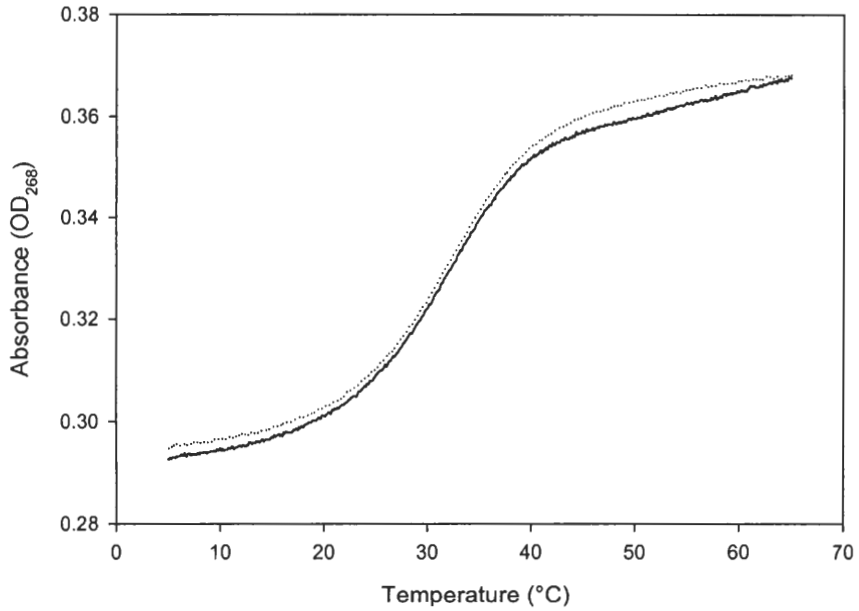


Figure 4.5 (continued). Typical UV denaturation curves for RNA duplexes containing tandem mismatches in 1.0 M Na^+. Solid lines represent denaturation curves. Dotted lines represent renaturation curves. K. CucG/CauG. L. CuaG/CauG.

M.



N.

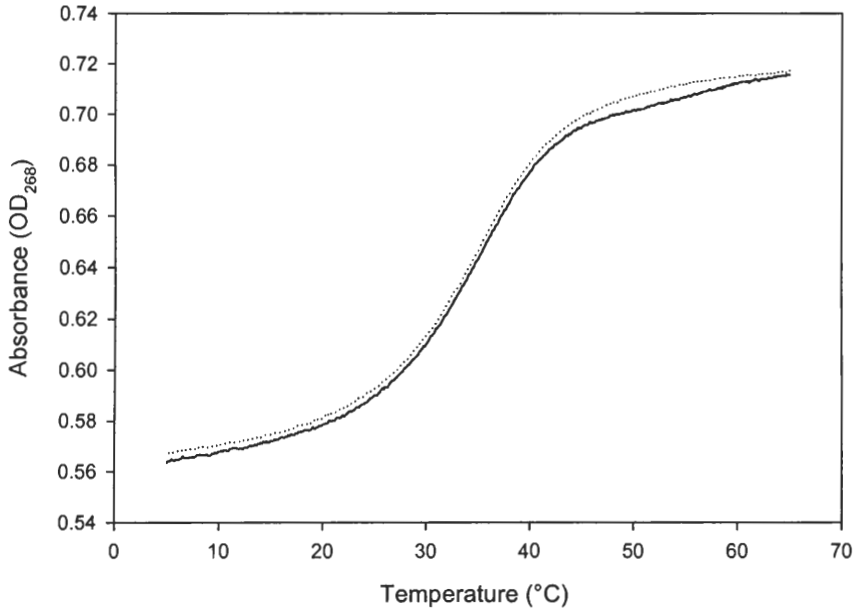
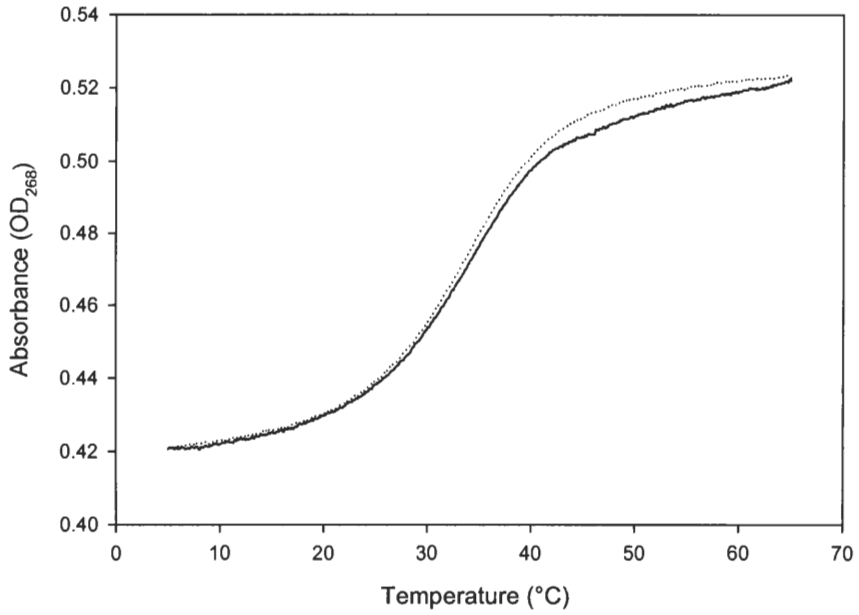


Figure 4.5 (continued). Typical UV denaturation curves for RNA duplexes containing tandem mismatches in 1.0 M Na^+. Solid lines represent denaturation curves. Dotted lines represent renaturation curves. M. CauG/CuaG. N. CguG/CuaG.

O.



P.

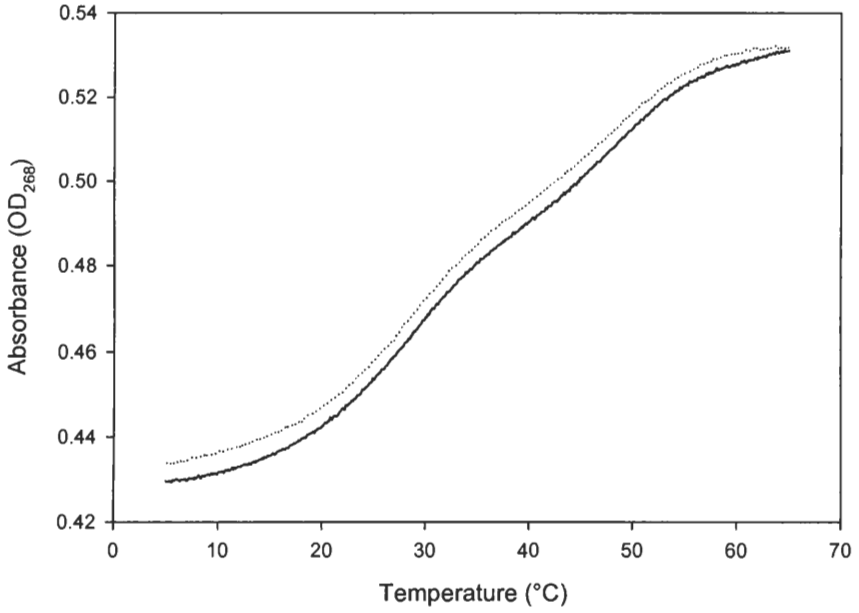
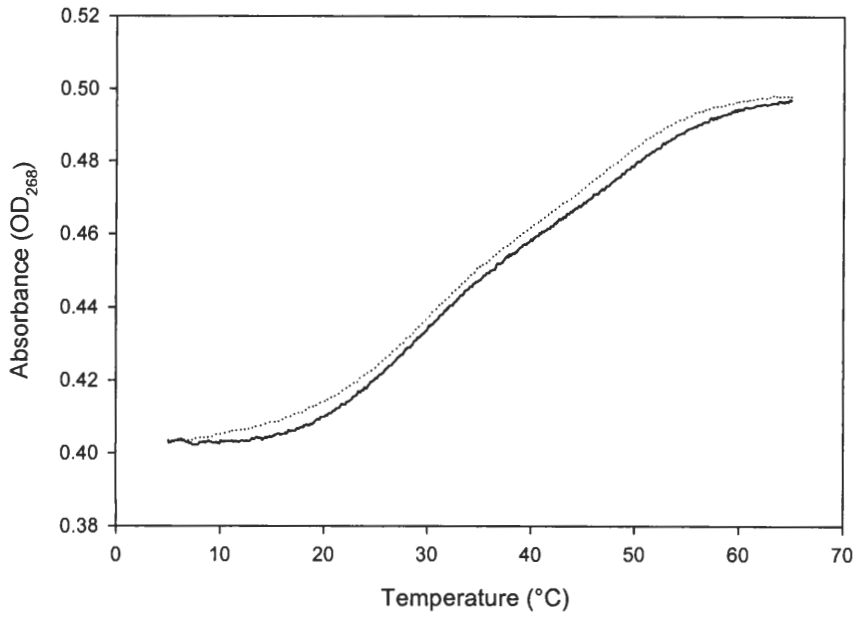


Figure 4.5 (continued). Typical UV denaturation curves for RNA duplexes containing tandem mismatches in 1.0 M Na^+. Solid lines represent denaturation curves. Dotted lines represent renaturation curves. O. CcuG/CuaG. P. CauG/CucG.

Q.



R.

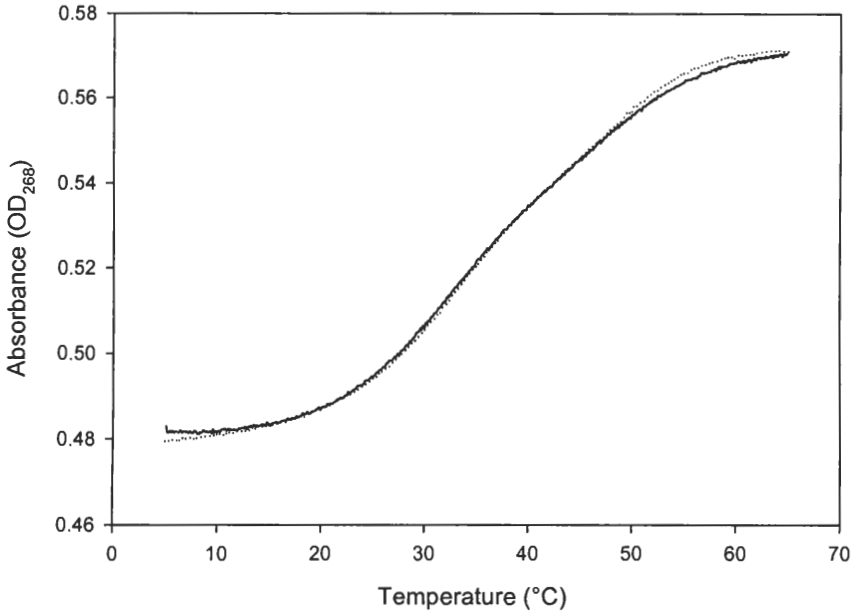


Figure 4.5 (continued). Typical UV denaturation curves for RNA duplexes containing tandem mismatches in 1.0 M Na^+. Solid lines represent denaturation curves. Dotted lines represent renaturation curves. Q. CcuG/CucG. R. CuuG/CucG.

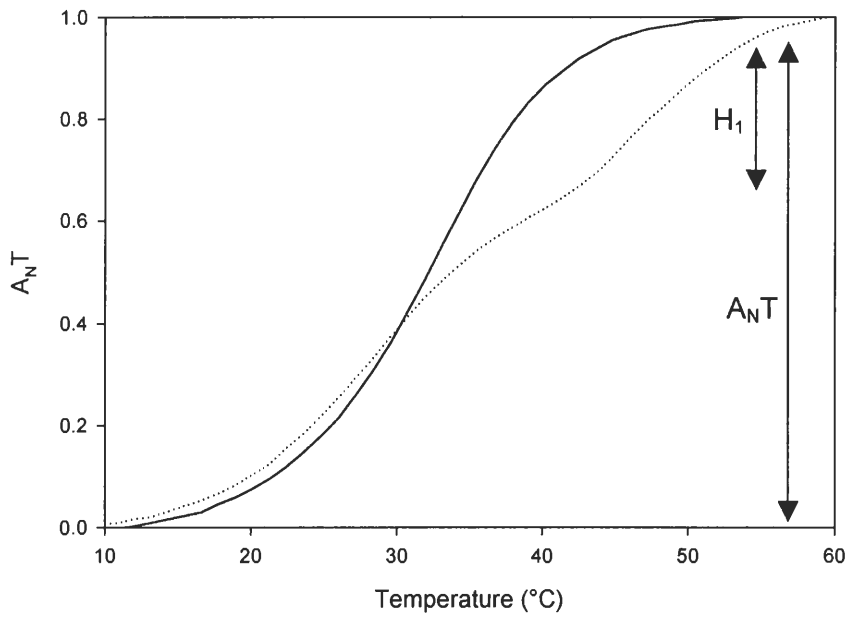


Figure 4.6. Normalized melting curves. The solid line represents the CuaG/CauG duplex, which has a single transition. The dotted line represents the CauG/CucG duplex, which shows two transitions. β (equation 4.4) is represented by the ratio of the change in absorbance for the hairpin H_1 to the total change in absorbance A_NT.

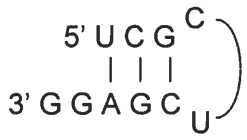


Figure 4.7. Proposed structure for the single-stranded 2-B-UC hairpin. The hairpin has three Watson-Crick base pairs in the stem, two dangling G's at the 3' end, and a two-base hairpin loop with the sequence C – U. The closing base pairs are G-C.

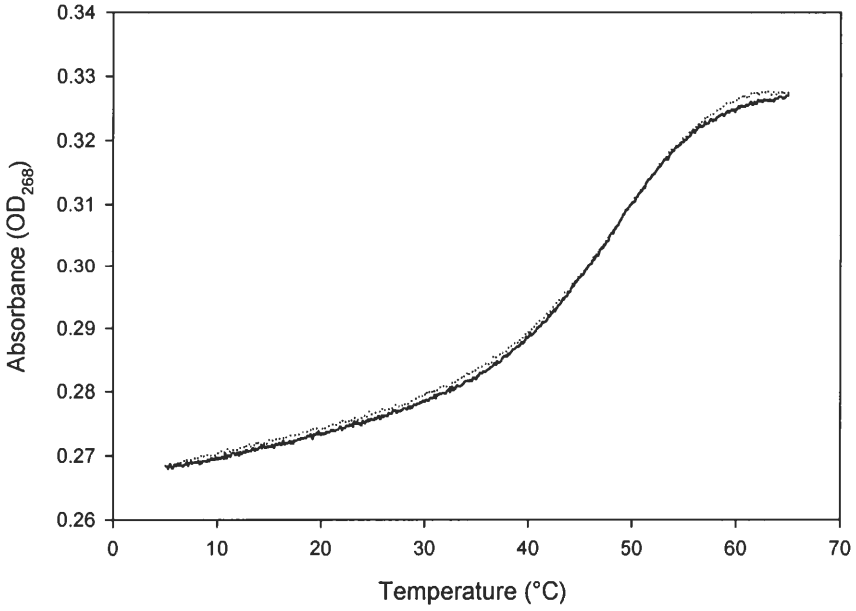
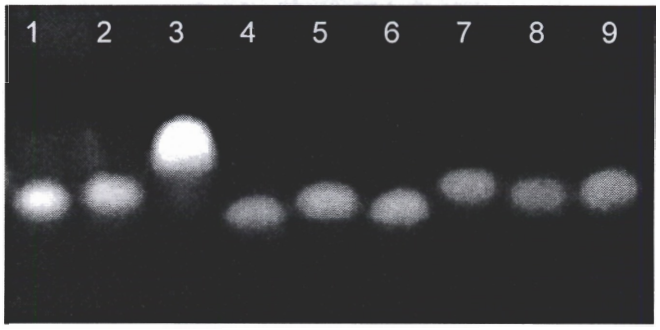


Figure 4.8. UV denaturation curve of the 2-B-UC molecule in 1.0 M Na^+. The solid line represents denaturation while the dotted line represents renaturation. The hyperchromicity of the transition is 22%, which is similar to the change in absorbance for hairpin molecules. The melting temperature for the transition is 46.6 °C, which is consistent with the second transition of the duplexes utilizing this strand.

When the single strand 2-B-UC is denatured alone in an optical melting experiment, one observes a sigmoidal transition curve (figure 4.8, p. 97). The melting temperature of the transition is 46.6 °C, nearly identical to the T_m of the second transition in the duplex melting curves. The hyperchromicity of this transition is 22%, consistent with the absorbance change noted for other hairpin nucleic acids. Figure 4.9 compares non-denaturing and denaturing polyacrylamide gels of each of the single-stranded molecules for set 2, the duplexes with C-G and G-C closing base pairs. In the non-denaturing gel, 2-B-UC shows a slower mobility than the other single-stranded molecules. This suggests a secondary structure that causes this molecule to move more slowly through the gel. This interaction was confirmed in the denaturing gel. The 2-B-UC molecule runs at the same position as the other single-stranded molecules, indicating that the native secondary structure was denatured, and not an inherent feature of this sequence. The intensity of the 2-B-UC band in the denaturing gel may also indicate the presence of a secondary structure. In ethidium bromide staining, helical molecules appear brighter than single-stranded single molecules. The 2-B-UC strand may have denatured as it traveled through the urea gel, but refolded upon staining in the ethidium bromide solution.

To determine the thermodynamic parameters for the duplexes with two transitions, the experimental curves were fit with theoretical curves derived from equations 4.2, 4.3, and 4.4 (pp. 70-1). Figure 4.10 shows the normalized melting transitions in solid lines compared to the fit regression curves in dotted lines. For the CauG/CucG molecule, the first transition, which represents the duplex to hairpin (2-B-UC) plus single strand (2-T-AU) transition, has a T_m of 27.3 °C and a ΔH of –60 kcal / mol. The second transition, which represents the hairpin to single strand transition, has a T_m of 47.2 °C and a ΔH of –57.0 kcal / mol. The melting temperature is similar to the T_m seen when the 2-B-UC strand alone is denatured. The enthalpy is significantly more

A.



B.

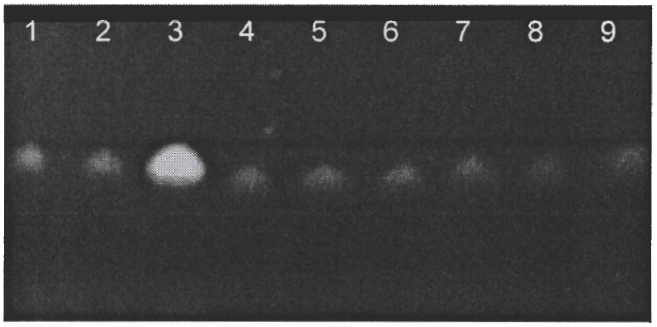
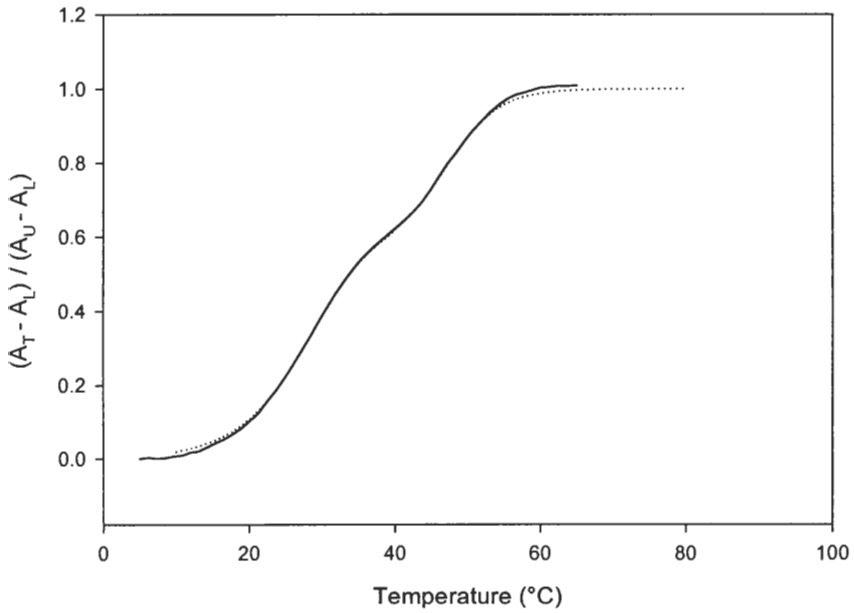


Figure 4.9. Polyacrylamide gel electrophoresis of single stranded molecules for duplexes with C-G and G-C closing base pairs. Lane 1: 2-B-AU. Lane 2: 2-B-UA. Lane 3: 2-B-UC. Lane 4: 2-T-AU. Lane 5: 2-T-UC. Lane 6: 2-T-UA. Lane 7: 2-T-GU. Lane 8: 2-T-CU. Lane 9: 2-T-UU. A. 18% non-denaturing polyacrylamide gel. 2-B-UC runs at a slower mobility than the other single strands, suggesting the presence of secondary structure. B. 18% denaturing (7M urea) polyacrylamide gel. 2-B-UC runs at the same mobility as other strands, suggesting a secondary structure has been denatured.

A.



B.

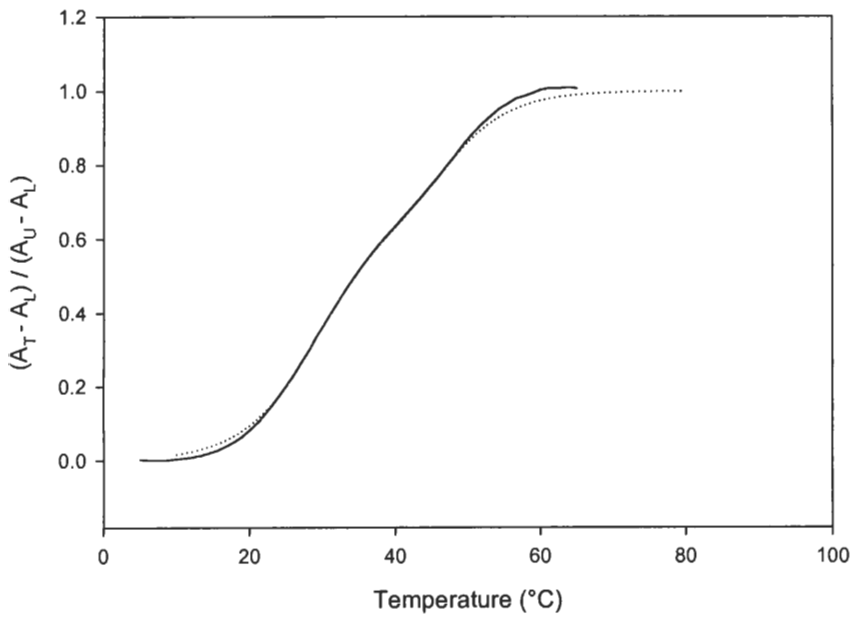


Figure 4.10. Normalized melting transitions for RNA duplexes involving the 2-B-UC strand with fit regression lines. Solid lines represent the normalized data while dotted lines represent the regression data. A. CauG/CucG. B. CcuG/CucG.

C.

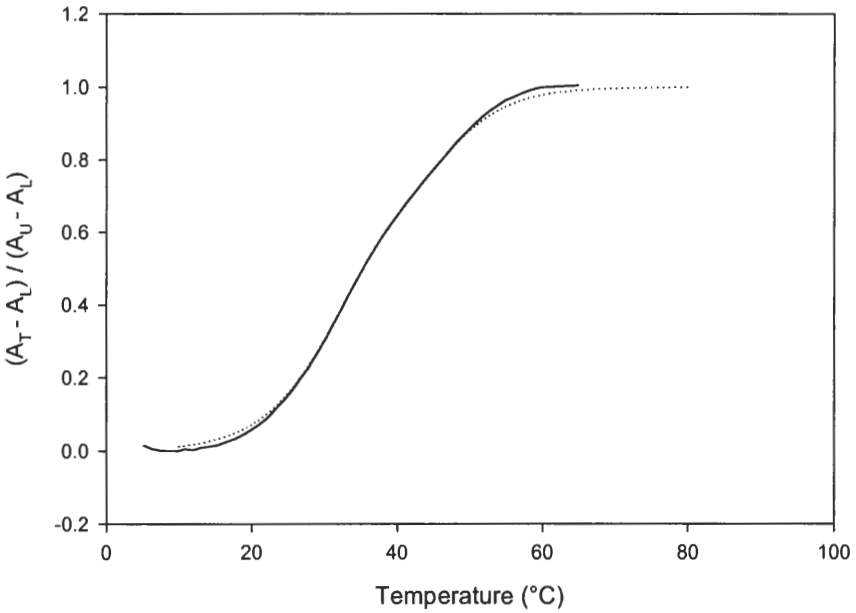


Figure 4.10 (continued). Normalized melting transitions for RNA duplexes involving the 2-B-UC strand with fit regression lines. Solid lines represent the normalized data while dotted lines represent the regression data. C. CuuG/CucG.

negative, however. This may be due to an interaction of the single strand with the hairpin strand. If this were the case, the interaction would be due to the two bases in 2-T-AU that create the tandem mismatch because other than those bases, it is identical to the other "top strand" molecules. Alternately, the increased enthalpy may be the result of 2-T-AU forming a stable secondary structure. Looking at sequence alone, it would be possible for the 2-T-AU to form a hairpin similar to the 2-B-UC molecule with two bases in the hairpin loop. This would be expected to be considerably less stable, however, because the base pairs in the stem would be an A-U, a C-G, and a U-A. Without further studies it is difficult to determine the cause of the higher enthalpy. The fit for the CcuG/CucG molecule gave a T_m of 27.8 °C and a ΔH of –60 kcal / mol for the first transition. The second transition had a T_m of 46.3 °C and a ΔH of –42.7 kcal / mol. This is very similar to the results from UV studies on the hairpin alone where the T_m was found to be 46.6 °C and the ΔH was –42.7 kcal / mol. For the CuuG/CucG molecule, the first transition had a melting temperature of 29.9 °C and an enthalpy of –60 kcal / mol. The second transition had a T_m of 45.7 °C and a ΔH of –42.7 kcal / mol. These were again very similar to the results from studies on the hairpin alone. This suggests that for these two molecules the hairpin is the sole contributor to the second transition.

Table 4.4 lists the evaluated thermodynamic parameters and melting temperatures for each of the duplexes in 1.0 M Na^+. For both sets of molecules, the completely Watson-Crick base paired duplex has a higher melting temperature and a lower free energy than the mismatched molecules. The different tandem mismatches produced similar thermodynamic effects within each set of molecules. For set 1, with G-C and A-U closing base pairs for the mismatch loop, the melting temperatures for the mismatched molecules were within 3.5 °C of each other. The free energies for the mismatched duplex molecules at 37 °C varied from –0.58 kcal / mol to –1.57 kcal / mol. The mismatched molecules of set 2 melted over a temperature range of 27.3 °C to 33.4

Table 4.4. Melting temperatures and thermodynamic values for RNA duplexes containing a U-U mismatch adjacent to another non-canonical base pair. Free energy values represent ΔG at 37 °C, assuming ΔH and ΔS are temperature independent.

Set 1				
Molecule	T_m (°C)	ΔH (kcal / mol)	ΔS (eu)	ΔG_{37} (kcal / mol)
GAUA/UAUC	61.8 ± 0.2	-90.7 ± 4.2	-271 ± 12	-6.72 ± 0.36
GucA/UauC	39.3 ± 0.4	-78.3 ± 0.9	-251 ± 3	-0.58 ± 0.1
GuaA/UauC	41.9 ± 0.1	-81.8 ± 1.5	-260 ± 5	-1.27 ± 0.04
GauA/UuaC	40.7 ± 0.1	-76.8 ± 1.0	-245 ± 3	-0.90 ± 0.02
GguA/UuaC	42.8 ± 0.1	-86.3 ± 3.2	-273 ± 10	-1.57 ± 0.05
GcuA/UuaC	40.5 ± 0.2	-79.9 ± 2.1	-255 ± 7	-0.90 ± 0.05
GauA/UucC	41.1 ± 0.1	-80.5 ± 1.8	-256 ± 6	-1.04 ± 0.03
GcuA/UucC	40.3 ± 0.2	-88.7 ± 3.0	-283 ± 10	-0.92 ± 0.07
GuuA/UucC	42.4 ± 0.3	-85.7 ± 3.4	-272 ± 10	-1.47 ± 0.04

Set 2				
Molecule	T_m (°C)	ΔH (kcal / mol)	ΔS (eu)	ΔG_{37} (kcal / mol)
CAUG/CAUG	54.5 ± 0.2	-102.3 ± 1.5	-312 ± 5	-5.46 ± 0.09
CucG/CauG	31.1 ± 0.3	-58.2 ± 3.0	-191 ± 10	1.12 ± 0.09
CuaG/CauG	32.4 ± 0.1	-61.3 ± 0.6	-201 ± 2	0.92 ± 0.02
CauG/CuaG	31.3 ± 0.3	-64.8 ± 2.4	-213 ± 8	1.21 ± 0.10
CguG/CuaG	33.4 ± 0.3	-68.5 ± 3.1	-224 ± 10	0.80 ± 0.06
CcuG/CuaG	33.0 ± 0.3	-65.1 ± 1.9	-213 ± 6	0.85 ± 0.04
CauG/CucG*	27.3	-60	-200	1.94
CcuG/CucG*	27.8	-60	-200	1.84
CuuG/CucG*	29.9	-60	-198	1.41

* Thermodynamic parameters were estimated by fitting a theoretical curve to normalized data.

°C and the free energies at 37 °C ranged from 0.80 kcal / mol to 1.94 kcal / mol. The results show that all of the tandem mismatches with one U-U pair were destabilizing. The variation of thermodynamic parameters based on the sequence of the mismatches is much smaller than in DNA. Additionally, the range of free energies is much smaller than the values estimated by extrapolation of results from other mismatches by Turner's group for set 1. The estimated range of free energies of the mismatches with a G-C and an A-U closing base pair was from 0.00 kcal / mol to 2.30 kcal / mol (http://www.bioinfo.rpi.edu/~zukerm/cgi-bin/efiles-3.0.cgi). The experimental free energies determined here were from 0.37 kcal / mol to 1.36 kcal / mol. For set 2, although the values for the free energies of the mismatches were different, the estimated range was similar. The estimated range of free energies for the mismatches with a C-G and a G-C as the closing base pairs was from 0.3 kcal / mol to 1.40 kcal / mol (http://www.bioinfo.rpi.edu/~zukerm/cgi-bin/efiles-3.0.cgi). The experimental values ranges from 0.94 kcal / mol to 2.08 kcal / mol. This set included the molecules analyzed by fitting a theoretical curve. The values for those mismatches are somewhat larger than values for the other mismatches, possibly due to this type of analysis.

Table 4.5 shows the thermodynamic parameters for tandem mismatch formation with different closing base pairs. The table is arranged such that thermodynamic parameters for a given loop with both sets of closing base pairs are on one line. Here again, the thermodynamic parameters are similar within a given set. In all cases, tandem mismatches are seen to be slightly destabilizing. It is interesting to note that in the case of set 2, with the C-G and G-C closing base pairs, both entropy and enthalpy are positive values, while in the case of set 1, with G-C and A-U closing base pairs, these values are negative. In general, it would be expected that closing base pairs with all G's and C's would be more stable than those with an A-U closing base pair. Here, the positive values may be an indication of next-nearest neighbor effects. The duplexes

Table 4.5. Thermodynamic values for tandem mismatches with different closing base pairs.

Set 1				Set 2			
Mismatch	ΔH (kcal/mol)	ΔS (eu)	ΔG_{37} (kcal/mol)	Mismatch	ΔH (kcal/mol)	ΔS (eu)	ΔG_{37} (kcal/mol)
GucA/UauC	-17.1	-60	1.36	CucG/CauG	13.8	41	1.26
GuaA/UauC	-20.6	-69	0.67	CuaG/CauG	10.7	31	1.06
GauA/UuaC	-15.6	-63	1.04	CauG/CuaG	7.2	19	1.35
GguA/UuaC	-25.1	-82	0.37	CguG/CuaG	3.5	8	0.94
GcuA/UuaC	-18.7	-64	1.04	CcuG/CuaG	6.9	19	0.99
GauA/UucC	-19.3	-65	0.90	CauG/CucG*	12	32	2.08
GcuA/UucC	-27.5	-92	1.02	CcuG/CucG*	12	32	1.98
GuuA/UucC	-24.5	-81	0.47	CuuG/CucG*	12	34	1.55

*Molecules were evaluated by fitting a theoretical melting curve to experimental data. Due to the limitations of this technique, the thermodynamic values for the mismatches evaluated in this manner may be less accurate than those values determined through the peak height method of van't Hoff analysis.

in set 2 had a G-U wobble base in the position adjacent to the closing base pair. Because this type of interaction is less stable than a Watson-Crick base pair would be, this base pair may be affecting the thermodynamic values for the tandem mismatches.

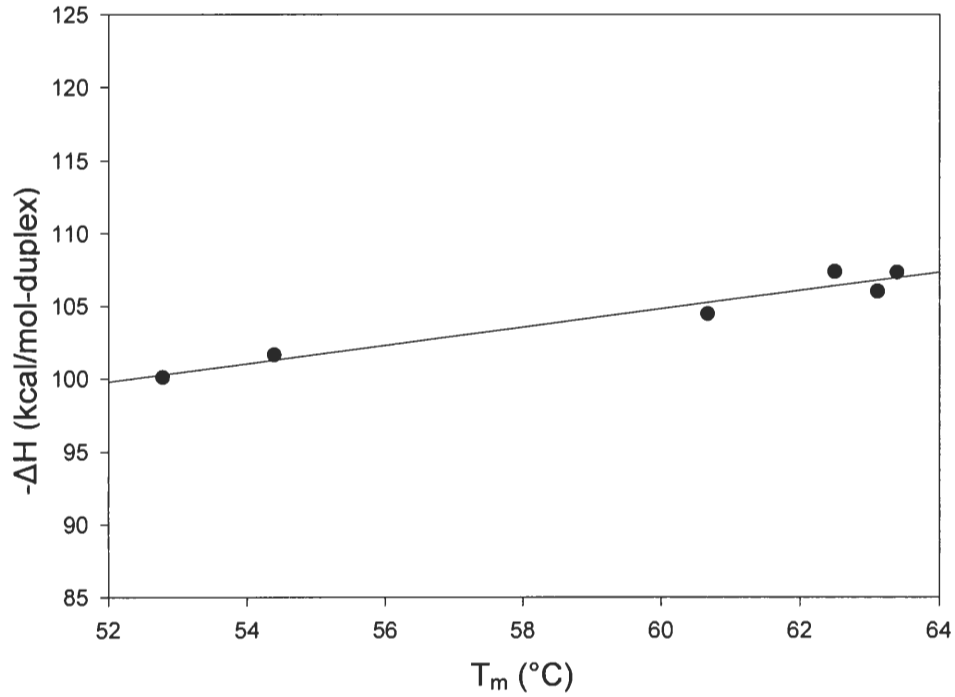
In addition to determining thermodynamic values for the duplexes, it was also desirable to estimate change in heat capacity for the formation of select molecules. Figure 4.11 shows the plots made for the formation of the completely Watson-Crick base paired molecule from set 1 (GAUA/UAUC). The slope of the regression line gives the change in heat capacity for formation of the duplex. In figure 4.11A, the ΔCp for duplex formation, as determined using the enthalpy change, was found to be -629 cal/mol-K. Because heat capacity may be affected by the length of the molecule, ΔCp is generally reported per base pair. The duplexes examined here have ten base pairs, so the ΔCp from the enthalpy is -62.9 cal/mol-K per base pair. In figure 4.11B, the change in heat capacity was determined using the entropy values for the molecule. Here, the ΔCp from the entropy change was found to be -32.9 cal/mol-K per base pair.

Figure 4.12 shows a similar analysis for the change in heat capacity of the CauG/CuaG mismatch molecule from set 2. From these graphs, ΔCp as determined by enthalpy was -39.6 cal/mol-K per base pair, while ΔCp determined by entropy was -19.7 cal/mol-K per base pair.

Discussion

RNA thermodynamic parameters are important for elucidating mechanisms of biological functions, as well as experimental design for techniques such as RNAi. In this chapter, tandem mismatches in RNA molecules have been studied. One set of experiments was carried out in short hairpin molecules, where the tandem mismatches contained a variety of sequences. Other experiments were performed on duplex RNA molecules with one U-U mismatch adjacent to another non-canonical base pair. These molecules were

A.



B.

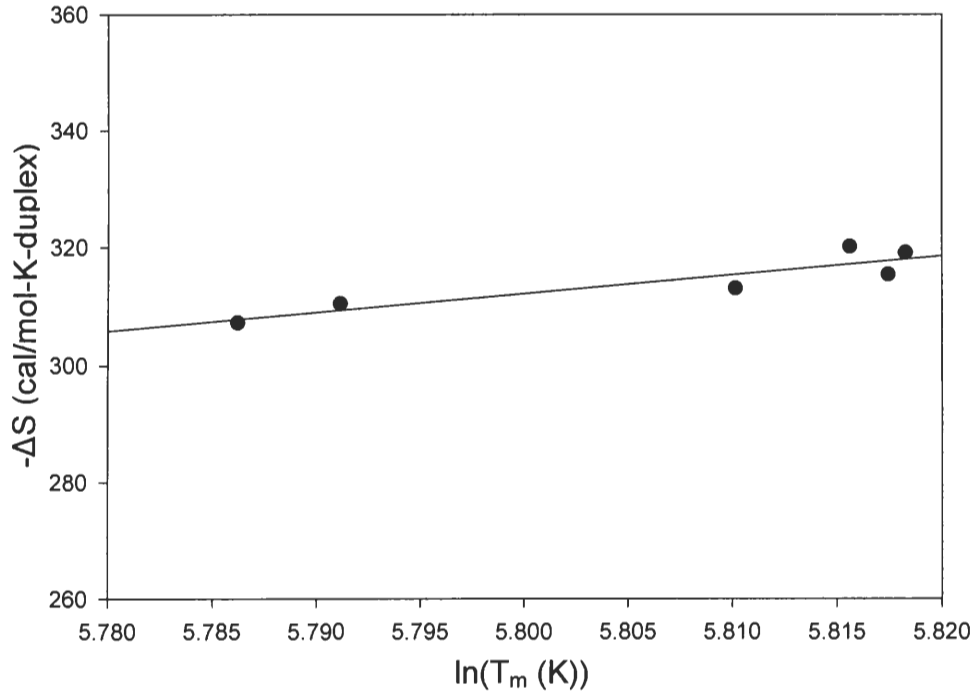
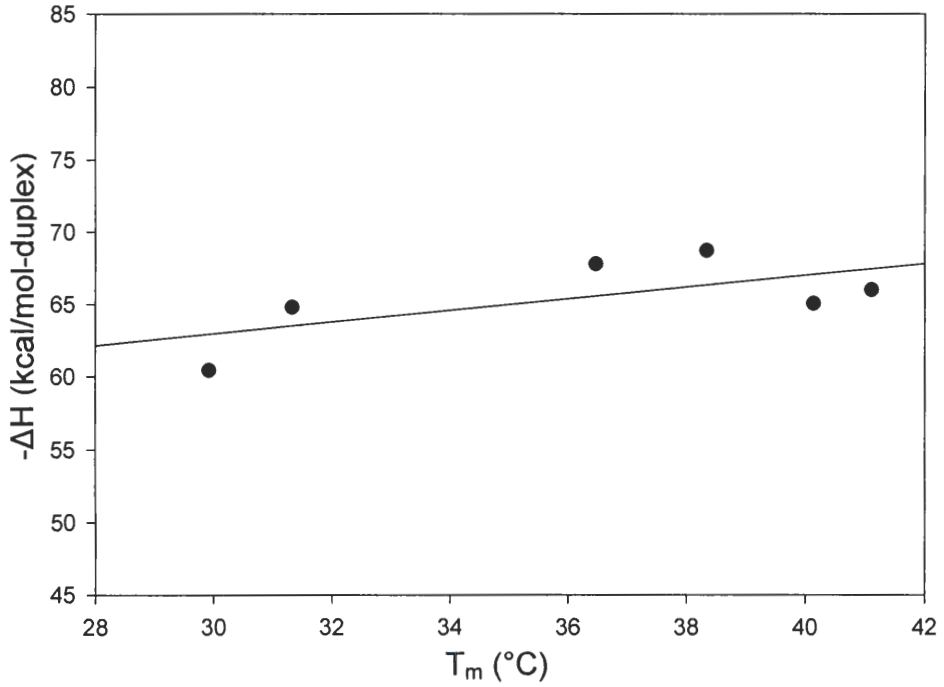


Figure 4.11. Evaluation of ΔCp for the GAUA/UAUC molecule. A. shows the linear regression of the T_m vs $-\Delta H$ plot for varying duplex concentrations in 1.0 M Na^+ as determined by the peak height evaluation of UV denaturation experiments. B. shows the linear regression of the $\ln(T_m)$ vs. $-\Delta S$ plot from the same experiments.

A.



B.

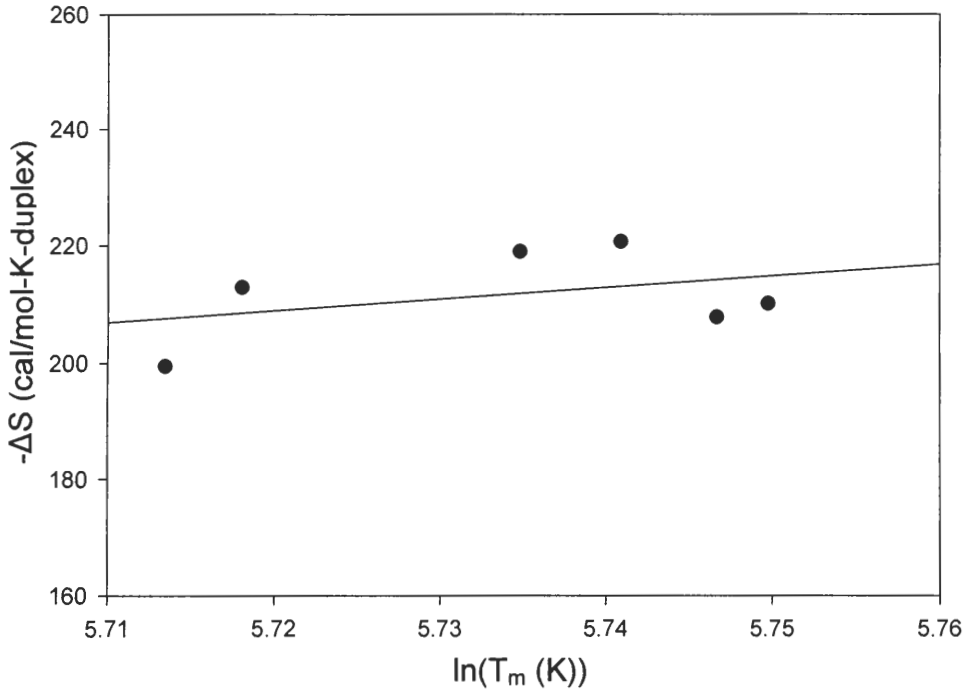


Figure 4.12. Evaluation of ΔCp for the caug/cuag molecule. A. shows the linear regression of the T_m vs –ΔH plot for varying duplex concentrations in 1.0 M Na^+ as determined by the peak height evaluation of UV denaturation experiments. B. shows the linear regression of the $\ln(T_m)$ vs. –ΔS plot from the same experiments.

examined not only for their thermodynamic parameters, but also for the effect of closing base pairs on the mismatches.

Hairpin molecules are valuable in determining thermodynamic parameters, as they avoid a duplex initiation factor. The UV denaturation curves in this study showed the hairpins to have reversible transitions, as would be expected for short hairpin molecules. The AA/UU RNA showed two transitions in the optical melting in both 1.0 M and 0.1 M Na^+ buffers. The ga/ga molecule showed two transitions in the 1.0 M Na^+ buffer. This is believed to be the result of a duplex to hairpin transition followed by a hairpin to coil transition. DNA molecules with similar sequences (chapter 3 of this work) gave only a single transition, demonstrating the differences in stability between DNA and RNA. For the AA/UU molecule, the two transitions had enough separation between them that it was still possible to extract thermodynamic data from the studies. The free energy value obtained at 37 °C for this molecule in 1.0 M Na^+ was -9.00 ± 0.43 kcal / mol. This is in reasonable agreement with the predicted estimates of free energy for this molecule, which give a value of -11.00 kcal / mol. The change in heat capacity determined for the formation of DNA hairpins (Chapter 3, p. 63) may be used to adjust the enthalpy determined for the AA/UU molecule. When this correction was made to the enthalpy and the free energy determined, the ΔG_{37} was found to be -10.6 kcal / mol. This value is very similar to the predicted estimate for this molecule, and suggests that RNA and DNA hairpins may have a similar ΔCp value. Thermodynamic predictions do not take the sequence of the hairpin loop into account, and relatively little is known about the energy of five base hairpin loops. Given that, the value obtained in these experiments appears to be accurate. The separation between transitions for the ga/ga molecule in 1.0 M Na^+ was not large enough to accurately derive thermodynamic data.

In 0.1 M Na^+, the ga/ga molecule denatured at a low T_m making it difficult to define the pre-transition baseline. This molecule was analyzed by a method that utilizes

the upper half of a melting profile to obtain thermodynamic data. Due to the reversibility of the transition, it is assumed that the curve represents only a hairpin to single-stranded transition. Prior to this work, ga/ga mismatches have not been examined in 0.1 M Na^+. Work by Walter *et al.* done in 1.0 M Na^+ attributed a free energy of –0.70 kcal / mol to the molecule [47]. This is considerably different from the free energy value obtained for the mismatches here of 3.70 kcal / mol. Utilizing equation 1.3 (p. 10) to correct the 1.0 M Na^+ data for the difference in salt concentration, a value of 0.09 kcal / mol is obtained. Although this equation is generally used for adjusting values for whole molecules, it can give a reasonable estimate for shorter nucleic acid elements as well. The difference may be due to the presence of duplex molecules that are not apparent in the melting transition.

Thermodynamic values for the formation of mismatches were determined for the remaining sequences. In all cases, the mismatches examined were found to be destabilizing. The hierarchy of stability for the mismatches examined here, from most to least stable, is aa/gc > ca/gc > ua/ac > uc/uc. Data from Turner's laboratory indicates that the range of free energies for all tandem mismatches is generally from –1.5 kcal / mol to +2.0 kcal / mol. The data seen for sequences examined here falls within that range. Turner's data also lists free energy values for specific sequences with particular closing base pairs (http://www.bioinfo.rpi.edu/~zukerm/rna/energy/). In comparing results obtained here, the data is very consistent in some cases. For the ca/gc hairpin, the experimental free energy was 1.44 kcal / mol, while Turner gives a free energy of 1.5 kcal / mol. For the ua/ac molecule, the results obtained here gave a free energy of 2.01 kcal / mol as compared to Turner's value of 1.9 kcal / mol. In the other two cases, the results differed more. For the aa/gc molecule, the experimental free energy was 1.38 kcal / mol while Turner reports a value of 0.6 kcal / mol. For the uc/uc molecule, the data here indicates a free energy of 2.02 kcal / mol compared to Turner's 1.4 kcal / mol.

This may not be surprising, though, as only Turner's uc/uc values is an experimental value [103]. The remaining Turner values are extrapolated predictions based on other results. Although the range of free energies for the different mismatch sequences is smaller for RNA's than for DNA's, it is important to obtain accurate information about the thermodynamic parameters for a variety of sequences to ensure precision in experimental design.

Duplex RNA's containing tandem mismatches

Tandem mismatches containing a U-U base pair were examined by UV melting in duplexes with two different closing base pair contexts. The curves appear to be very reversible. In examining the thermodynamic parameters for the entire duplex molecule, set 1, with a G-C and an A-U as the closing base pairs, gave slightly negative free energy values. These duplexes contained six G-C base pairs and two A-U base pairs. The duplexes in set 2, with a C-G and a G-C as the closing base pairs showed slightly positive free energy values. The stems of these duplexes contained five G-C base pairs, two A-U base pairs, and a single G-U wobble base. It is generally expected that mismatches closed by base pairs containing all G's and C's are more stable than mismatches with other closing base pairs. It is important to realize that this does not represent the thermodynamics of the entire molecule, however. In this case, although the G-C and A-U content of these duplexes is similar, the G-U wobble base caused the second set of molecules to have a lower melting temperature. When designing experiments that involve hybridization of a short molecule to a target, it is important to look at the thermodynamics of all aspects of the nucleic acid carefully.

In comparing the thermodynamic parameters of the mismatches alone, it should be noted that in the case of set 1, with G-C and A-U closing base pairs, the entropy and enthalpy for the loop is negative. In set 2, with C-G and G-C closing base pairs, these

values are positive. This may be a reflection of non-nearest-neighbor interactions. Adjacent to the G-C closing base pair, there is a G-U wobble base. The non-Watson-Crick base pair may affect the thermodynamic values for the tandem mismatches. Although this G-U is present in the control molecule, and any contribution should ideally be subtracted out, the remainder of the control molecule is completely duplexed. The wobble base may have different effects when it is present one base away from the tandem mismatches.

In some cases, there seemed to be little effect of the closing base pairs. The uc/au mismatches had a free energy value of 1.36 kcal / mol in set 1 and 1.26 kcal / mol in set 2, which is a difference of 0.1 kcal / mol. The cu/ua loop had a ΔG value of 1.04 kcal / mol in set 1 and 0.99 kcal / mol in set 2, a difference of 0.05 kcal / mol. Both of these mismatches contained a C-A pair adjacent to the U-U pair. The molecules containing an A-A pair adjacent to the U-U had similar differences as well. For the ua/au mismatch the ΔG was 0.67 kcal / mol in set 1 and 1.06 kcal / mol in set 2, a difference of 0.39 kcal / mol. Similarly, the difference for the au/ua molecules was 0.31 kcal / mol with the loop from set 1 having a value of 1.04 kcal / mol and the loop from set 2 having a free energy of 1.35 kcal / mol. These mismatches were the only ones studied where the same non-canonical base pair was found on both sides of the U-U. This may be preliminary evidence that in tandem mismatches when a U-U mismatch is present, the free energy values for the loop depend only on the sequence of the other base pair, not it's position.

A larger difference was seen between the different closing base pair sets for the gu/ua molecules. Here the loop ΔG for set 1 was 0.37 kcal / mol while the free energy for set 2 was 0.94 kcal / mol, a difference of 0.57 kcal / mol. Differences close to 1 kcal / mol were seen for the au/uc, cu/uc, and uu/uc sets of molecules. It is important to note, however, that the thermodynamic parameters for the three molecules with these

mismatches from set 2 were evaluated by fitting a theoretical curve to the experimental data due to the two transitions. This is different from the analytical methods for the same mismatches from set 1, and there may be differences in precision between the two methods. In determining the theoretical curve, if the enthalpy value were changed by –1 kcal / mol, from –60 to –61 kcal / mol, the resulting free energy value would be more in the range of both the Turner values and the other experimental free energies determined here. This change would probably not be apparent in the comparison of the theoretical curve and the experimental curve, and may represent a limitation of this technique.

The overall results suggest that in some cases closing base pairs have little effect on the stabilities of tandem mismatches containing a U-U base pair, and in other cases those mismatches with a G-C and an A-U closing base pair are slightly more stable. This may again be the result of effects of the wobble base adjacent to the G-C closing pair in set 2, as it would usually be expected that mismatches with all G's and C's making up the closing base pairs would be more stable.

In comparing these results to values published from Turner's laboratory, there are again some molecules for which the results are very similar, and others where the differences are larger. To look first at the mismatches in set 1, where there is a G-C and an A-U closing base pair, the experimental values obtained here are very similar to Turner's results for four of the molecules. For the GauA/UucC loop, the experimental results here gave a ΔG of 0.9 kcal / mol, while Turner predicts a value of 1.0 kcal / mol. For the GcuA/UucC loop the experimental and Turner's values were 1.02 and 1.0 kcal / mol. The experimental and extrapolated values for the GcuA/UuaC mismatches were 1.04 and 1.0 kcal / mol. In the case of the GucA/UauC loop, the experimental and extrapolated values were 1.36 and 1.1 kcal / mol.

The differences between the experimental and Turner's values for set 1 were larger for other mismatches. The GguA/UuaC loop had a difference between the

experimental and extrapolated values of 0.76 kcal / mol and the GuuA/UucC loop had a difference between the two of 0.47 kcal / mol. In the first case, the experimental value was 0.34 kcal / mol, which is more stable than the estimated value of 1.1 kcal / mol. In the second case, the experimental value (0.47 kcal / mol) was less stable than Turner's free energy of 0.0 kcal / mol. Differences between the experimental and predicted values of greater than 1 kcal / mol were seen in the GuaA/UauC and GauA/UuaC molecules. In both cases, Turner's value was slightly over 2 kcal / mol while the experimental value was significantly more stable with free energy values of 0.67 and 1.04 kcal / mol. The values predicted by Turner's group for the loops with G-C and A-U closing base pairs show a free energy range of 2.3 kcal / mol. Experimentally, the range was found to be smaller at just over 1 kcal / mol. Again, Turner's data was extrapolated from other experimental data, so differences may not be surprising.

For set 2, with C-G and G-C closing base pairs, there was good agreement between the published free energy value and experimental value for the CauG/CuaG molecule with a difference of only 0.05 kcal / mol. The CguG/CuaG, CuaG/CauG, and CucG/CauG molecules all showed differences between Turner's value and the experimental value of less than 0.5 kcal / mol. Turner's value was less stable than the experimental value for the first two molecules, while the experimental value was less stable than the published value for the latter of the three. The remaining four molecules showed larger differences between the experimental results and the values from Turner. In each case, Turner's value predicted the mismatches to be more stable than they were found to be. Although it is not clear which of Turner's values are theoretical and which are experimental, the free energy range for all mismatches reported by the group is from –1.50 kcal / mol to +2.00 kcal / mol. The mismatches examined here do fall within that range with the exception of the CauG/CucG mismatch where the experimental loop free energy is 2.08 kcal / mol.

For two of the duplexes, experiments were performed to determine the change in heat capacity. For each of the molecules the value obtained from the enthalpy results was higher than the value obtained from entropy results. A recent paper by Wu *et al.* indicated that when change in heat capacity is plotted in these two manners, the values should agree within 5% [97]. A paper by Rouzina and Bloomfield, however, notes that in examining the plots for a number of longer DNA's, the enthalpy value is always larger. For the examples shown, the difference ranges from the enthalpy being about 40% higher to almost 1000% higher. Citing a 1983 paper by Petersheim and Turner [105], the Rouzina paper claimed that for short DNA's of four to six bases, however, the plots gave values within 5% [64]. In the work presented here, the differences may be due to the length of the molecules. Thermodynamic values obtained for sequences of only a few base pairs are more precise than values obtained for longer molecules, such as those used here. The consequence of this would be a less accurate regression line and as a result a less accurate slope. Rouzina and Bloomfield report the change in heat capacity to be small for the formation of DNA molecules, on the order of -40 to -100 cal/mol-K per base pair. Although definitive results are not available from this work for the change in heat capacity, it is apparent that the change is small and on a similar order to the range given by Rouzina.

One interesting result obtained unexpectedly was the possibility that one of the single strands from set 2 is forming a hairpin. This was found after the three duplexes utilizing this strand showed two transitions in optical melting experiments. There are several pieces of evidence that point to this strand forming a hairpin. In a non-denaturing polyacrylamide gel, the 2-B-UC strand runs at a slower mobility than the other single stranded molecules. When the same molecules are run on a denaturing polyacrylamide gel, the 2-B-UC strand is present at a similar position to the other strands (figure 4.8). When UV denaturation is performed on the 2-B-UC strand alone, a melting

transition is seen. This curve has a similar hyperchromicity to what would be expected from a hairpin to coil transition. The melting temperature found in the single strand is at a similar position to the second transition in the duplex denaturation experiments.

If this molecule is forming a hairpin, it is predicted that the stem contains three Watson-Crick base pairs, two dangling bases on the 3' end, and a two base hairpin loop. As described in chapter 1 of this work, some highly stable four base hairpin loops have been shown to have non-Watson-Crick interactions across the loop between bases in positions 1 and 4 of the loop. Additionally, a paper by Jucker and Pardi showed that in a C-U-U-G tetraloop the C and G in positions 1 and 4 form a Watson-Crick paired base [82]. This indicates that this particular tetraloop is actually present as a biloop. The 2-B-UC molecule represents a novel form of biloop. This hairpin loop would have a G-C as the closing base pairs, rather than the C-G from Jucker and Pardi's paper. Traditionally, hairpin loops closed by a C-G are more stable than those closed by other bases, so it is interesting that the 2-B-UC molecule could be forming a hairpin with G-C closing base pairs. The bases forming the loop would be a C-U instead of the U-U from the Jucker paper. In that paper, it was seen that the first U, in position 2, was folded down into the loop and interacted with the C-G base pair, while the U in position 3 was turned out of the loop. NMR studies on the 2-B-UC molecule could show if in fact a hairpin is being formed and the orientation of the C and U in the loop.

CHAPTER 5

THE EFFECT OF DANGLING ENDS AND 2'-O-METHYL MODIFICATION ON STABILITY

Introduction

Knowledge of the thermodynamic properties of structural motifs, such as tandem mismatches, are important for accurate secondary structure prediction and the understanding of biological processes. In addition to the studies on the tandem mismatches in DNA and RNA, experiments were carried out to evaluate the influence of two other modifications on the stability of a DNA duplex.

One modification is a 3' or 5' terminal, unpaired base, also called a dangling base, added to the end of a hairpin. A study by Riccelli *et al.* examined hairpin molecules with dangling ends of 13 or 22 bases [106] while a paper by Doktycz *et al.* studied hairpins with four base dangling ends [84]. Senior *et al.* studied two duplexes with two dangling thymidines (T's) utilizing UV denaturation, circular dichroism, and nuclear magnetic resonance [83]. The authors found that dangling T's caused an increase in the melting temperature as compared to the duplex without the dangling ends. When the T's were positioned at the 5' end of the strands, the increase in melting temperature was greater than when the dangling bases were located at the 3' ends of the strands. SantaLucia's laboratory has studied the 32 possible single-nucleotide dangling ends on a Watson-Crick base pair in duplex molecules. Self-complementary

DNA strands were used, which resulted in a duplex with the same single dangling base on each end [86].

The work presented here examines single-nucleotide dangling ends on a hairpin molecule. Because the work in SantaLucia's laboratory utilized molecules with two single base dangling ends, assumptions that each end contributed equally to the thermodynamic parameters were necessary. Their results implied that in some instances a single dangling base would enhance the stability of a duplex to a greater extent than adding a Watson-Crick base pair onto the end of the duplex [83, 84, 86]. It was of interest to determine if this is correct and to quantify the effect of a single-base dangling end on the stability of a hairpin molecule. Information about the thermodynamics of dangling ends will improve the accuracy of secondary structure prediction, as well as aid in the design of immobilized probes in microarray and molecular beacon experiments.

Another type of modification examined in this chapter is the 2'-O-methyl modified sugar. Recently a number of investigators have begun using nucleic acid oligomers with this type of modified sugar. Strands with 2'-O-methyl riboses have been found to bind RNA targets faster and with a higher T_m than to corresponding DNA targets. When the target is a DNA, the oligomers with all 2'-O-methyl sugars bind more slowly and with the same melting temperature as the corresponding deoxyoligonucleotide. One application of this is mutation repair, in which a chimeric strand is directed to displace one strand of a DNA duplex and stimulate recombination repair. Little is known, however, about the relative stability of a chimeric molecule binding to a DNA strand. This information is important for accurate design of probes and targets. Studies were undertaken to examine the stability of three chimeric molecules bound to a DNA target and compare their stability to corresponding DNA/DNA duplexes.

Methods

Hairpin molecules were designed to contain a five base hairpin loop and a completely Watson-Crick base paired stem or a stem with a single, terminal unpaired base. One set of molecules had a core sequence of 5' ATAGCGACATCTAGTCGCTAT 3'. This set was designed with an A-T as the original terminal base pair. The remaining hairpins in this set added a T to the 5' end, an A to the 3' end, or both, to give a T-A as the new terminal base pair. The second set of molecules was generated in a similar manner, but with the original terminal base pair as a C-G, utilizing the sequence 5' CTAGCGACATCTAGTCGCTAG 3'. The DNA's were named according to the bases at the terminus of the hairpin, beginning at the 5' end. For example, the original molecule for the first set described above is called A/T hairpin after the terminal base pair. The molecule from the same set with a 5' thymidine is called the TA/T hairpin. Nomenclature for the remaining molecules is shown in figure 5.1.

Denaturation curves of the hairpins were monitored by UV absorbance. Samples were diluted to approximately 0.3 OD_{268}/ml (approximately 1.8 μM strand) in a buffer containing 0.078 M NaCl, 0.01 M Na_2HPO_4, and 1 mM Na_2-EDTA, pH 7.0. Each sample was heated to 95 °C for three minutes and cooled to room temperature prior to loading into the quartz cuvette. Each sample was overlaid by mineral oil to reduce evaporation. The melting transition for each molecule was monitored at 268 nm over a temperature range of 25 to 97 °C with a heating rate of 0.5 °C/min. Absorbance readings were taken every 0.1 °C. A minimum of four transitions were obtained for each molecule.

The duplexes using one chimeric strand were 14 base pairs long and are shown in figure 5.2. Each chimeric strand had two regions of five 2'-O-methyl modified nucleotides surrounding a region of four DNA bases. The complementary strand consisted of all deoxyribose bases. Chimeric molecules were synthesized by Operon

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

A/T hairpin

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

C/G hairpin

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

TA/T hairpin

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

TC/G hairpin

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

A/TA hairpin

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

C/GA hairpin

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

TA/TA hairpin

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

TC/GA hairpin

Figure 5.1. Expected structure for dangling end DNA hairpins. Each molecule has a five base hairpin loop. The stems for each molecule consist of either eight Watson-Crick base pairs, eight Watson-Crick base pairs with either one 5' or 3' dangling base, or nine Watson-Crick base pairs.

DNA duplexes	2'-O-methyl duplexes
5' GTGTTGAATTAATT 3' |||||||||||||| 3' CACAACTTAATTAA 5'	5' GUGUUGAATUAAUU 3' |||||||||||||| 3' CA CAACTTAA TTAA 5'
5040	5040 2'-O-methyl
5' CATTAAAATCCCGA 3' |||||||||||||| 3' GTAATTTTAGGGCT 5'	5' CAUUAAAATCCCGA 3' |||||||||||||| 3' GTAAT TTTAGGGCT 5'
5041	5041 2'-O-methyl
5' TGATGTGTTGAATT 3' |||||||||||||| 3' ACTACACAACTTAA 5'	5' UGAUGUGUUGAAUU 3' |||||||||||||| 3' ACTA CACA ACTTAA 5'
5020	5020 2'-O-methyl

Figure 5.2. Sequence of 2'-O-methyl chimera/DNA duplexes and corresponding DNA/DNA duplexes. Underlined sections represent the 2'-O-methyl modified nucleotides.

Technologies / Qiagen (Valencia, CA). For comparison, DNA/DNA duplexes with the same sequences were also synthesized.

Each strand was mixed in equimolar amounts for a final duplex concentration of 2 μM. To this, a concentrated buffer solution and dH_2O were added to give a final salt concentration of 0.078 M NaCl, 0.01 M Na_2HPO_4, and 1 mM Na_2-EDTA, pH 7.0. The melting transition of each duplex was monitored at 268 nm over an appropriate temperature range. This was generally 13 to 80 °C. Samples were heated at a rate of 0.5 °C/min and absorbance readings were taken every 0.1 °C. A minimum of four melting transitions were obtained for each duplex.

Results

Figure 5.1 (p. 120) shows the expected structure for the dangling end hairpin molecules. Each hairpin consisted of a five base loop and a core stem of eight Watson-Crick base pairs. Additional molecules had a 5' or 3' single, unpaired terminal base, or a ninth Watson-Crick base pair. One set of molecules contained an A-T terminal base in the core sequence while the other group of hairpins had a C-G as the terminal base pair for the core. In both cases the additional 5' base was a thymidine while the dangling 3' base was an adenine. For molecules containing nine Watson-Crick base pairs, both of the dangling end bases were added to create a T-A base pair.

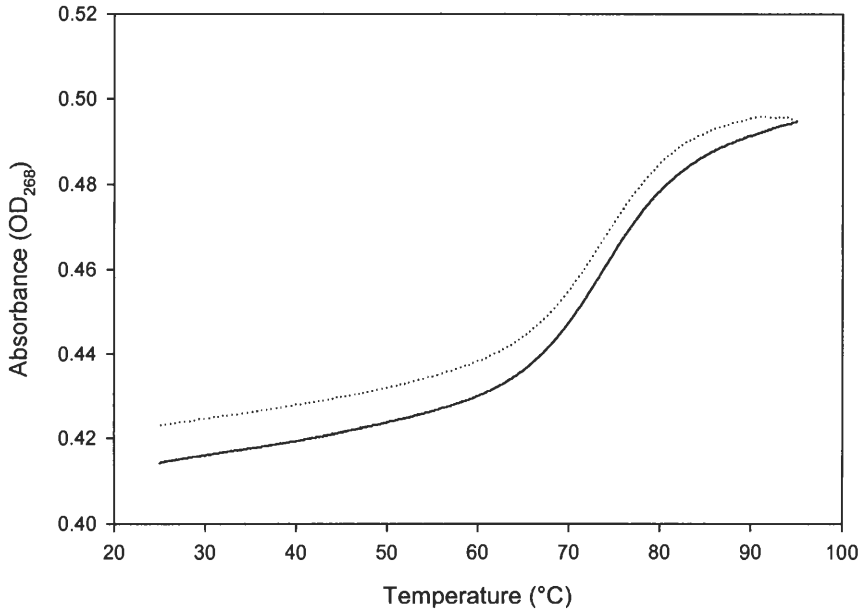
Figure 5.2 (p.121) shows the sequences of the DNA/DNA and the 2'-O-methyl chimera/DNA duplexes studied. Underlined bases represent the 2'-O-methyl modified RNA bases. The remaining bases are DNA bases. Each duplex contained 14 base pairs. For the chimeras, there were two regions of five modified bases interrupted by a four base DNA region. The molecules were selected to have closing base pairs around the DNA region of two 2' OMe U-A's, a 2' OMe A-T and a 2' OMe C-G, and two 2' OMe

G-C's. The molecules were named according to their position in a gene of interest for other experiments.

Typical melting results for each of the hairpins in the 0.1 M Na^+ buffer are shown in figure 5.3. Solid lines represent denaturation while dotted lines represent renaturation curves. In general, the curves are highly reversible. Although the samples were covered by mineral oil, condensation was observed in some cases, which led to slight increases in the absorbance readings at the beginning of the renaturation. Because the overall shape of the curves was similar for a given molecule, the differences in heating and cooling curves are believed to be due to this evaporation effect.

Table 5.1 lists the T_m values and thermodynamic parameters evaluated from the normalized denaturation curves of the dangling end hairpins. The transition enthalpies were evaluated by the peak height method, and the entropy and free energy calculated. Free energies were determined at 37 °C assuming ΔH and ΔS are temperature independent. Although heat capacity changes (Chapter 3) indicate a temperature dependent correction is needed, this study was primarily focused on the ΔG difference between pairs of hairpins. In addition, most tabulated thermodynamic parameters assume temperature-independent ΔH values. The free energies of the hairpins were in good agreement with the free energies predicted by Zuker's *mfold* program [4 1] . For the A/T hairpin, the experimental free energy value was –5.79 ± 0.31 kcal/mol. Zuker's program predicted a free energy value for this molecule in 0.1 M Na^+ of –5.7 kcal/mol. For the TA/TA hairpin, the experimental value was –6.06 ± 0.17. Zuker's program predicted two possible structures for this molecule. The first structure had Watson-Crick base pairs for all nine bases in the stem, and had a free energy estimation of –6.0 kcal/mol. The second structure had eight of the stem bases paired, with the terminal T and A unpaired. The estimated free energy for this molecule was –6.7 kcal/mol. This estimation is believed to be the result of utilizing the sum of the energetic contributions

A.



B.

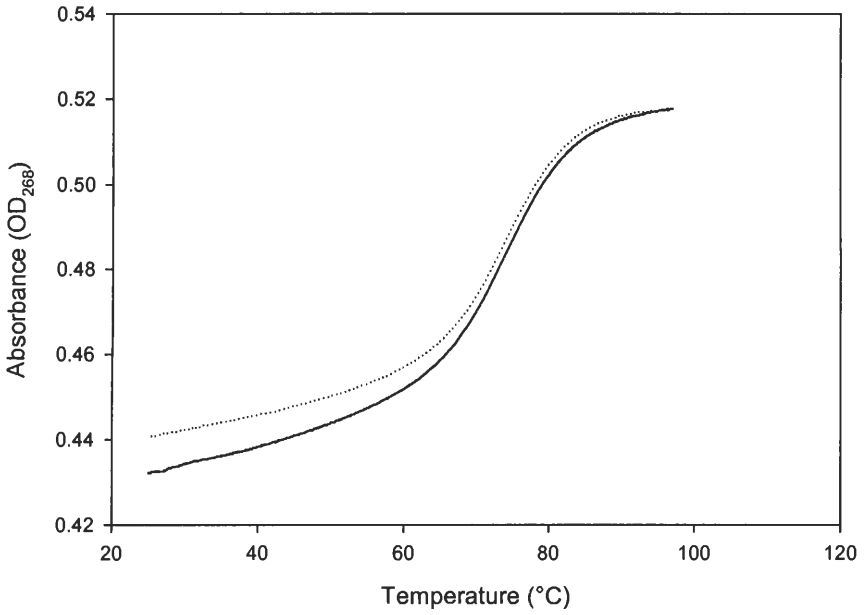
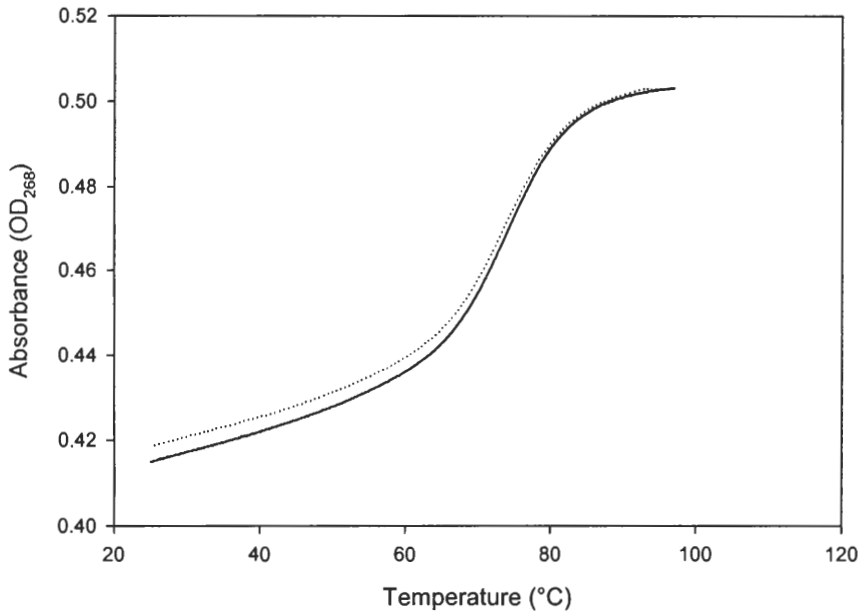


Figure 5.3. Typical UV denaturation curves of dangling end hairpins in 0.1 M Na^+. Solid lines represent denaturation curves while dotted lines represent renaturation curves. A. A/T. B. TA/T.

C.



D.

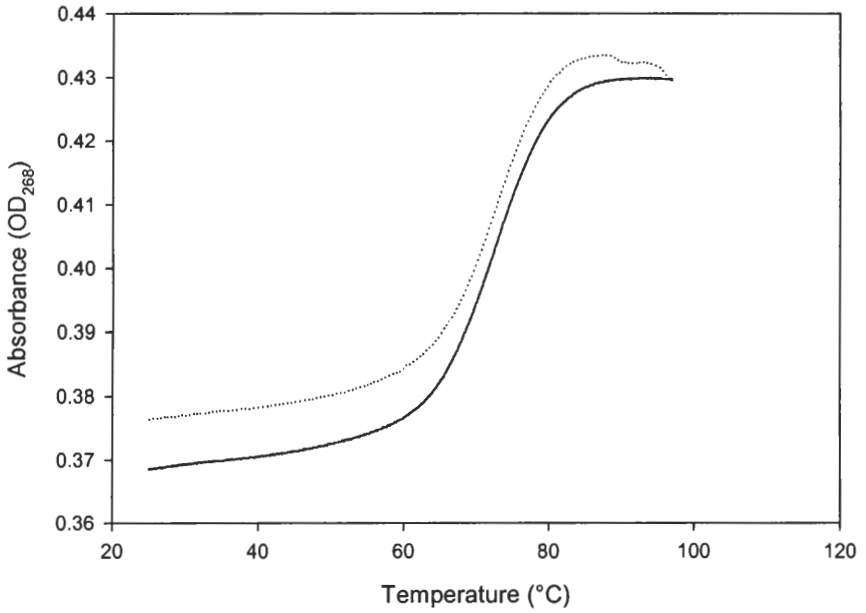
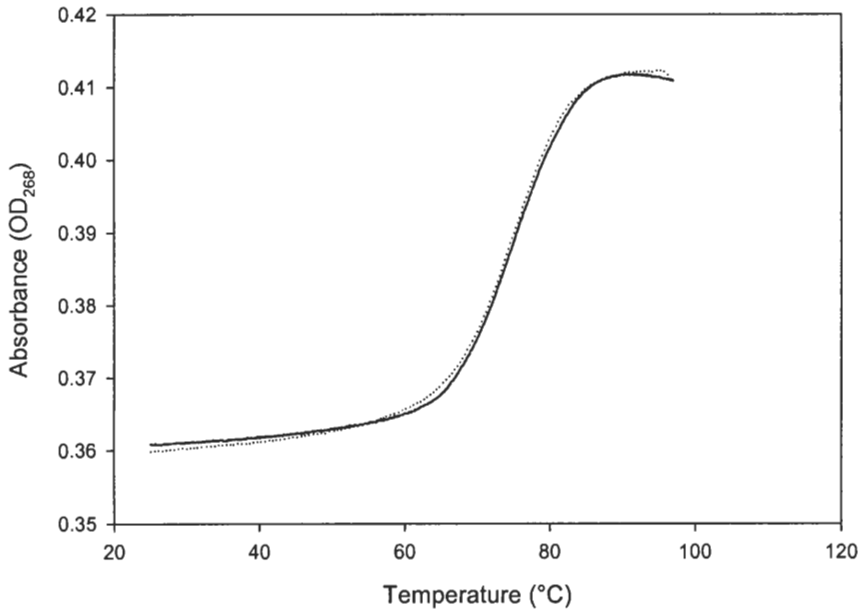


Figure 5.3 (continued). Typical UV denaturation curves of dangling end hairpins in 0.1 M Na^+. Solid lines represent denaturation curves while dotted lines represent renaturation curves. C. A/TA. D. TA/TA.

E.



F.

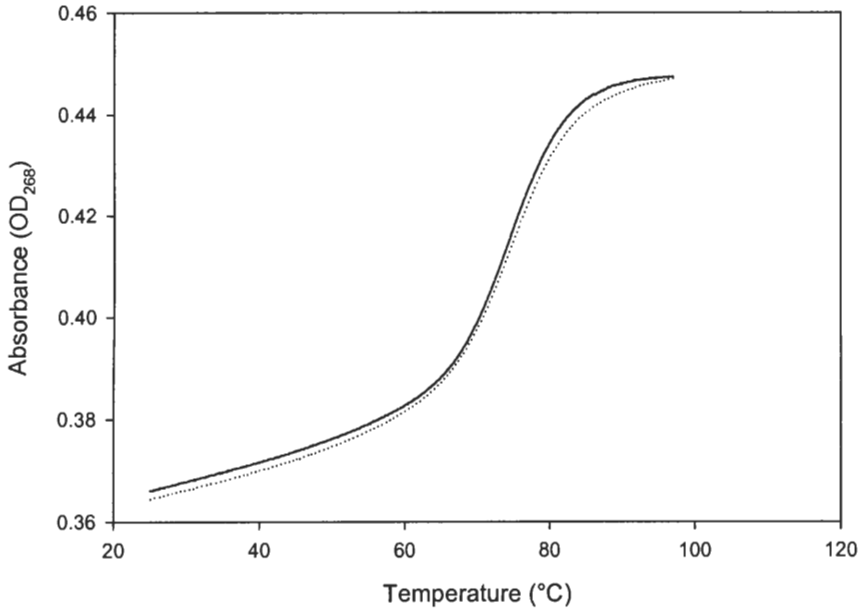
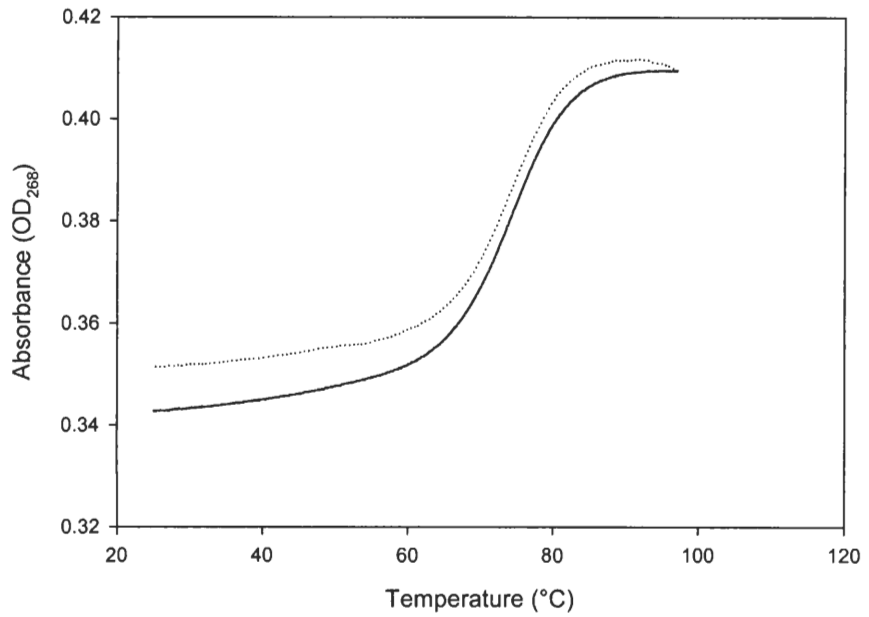


Figure 5.3 (continued). Typical UV denaturation curves of dangling end hairpins in 0.1 M Na^+. Solid lines represent denaturation curves while dotted lines represent renaturation curves. E. C/G. F. TC/G.

G.



H.

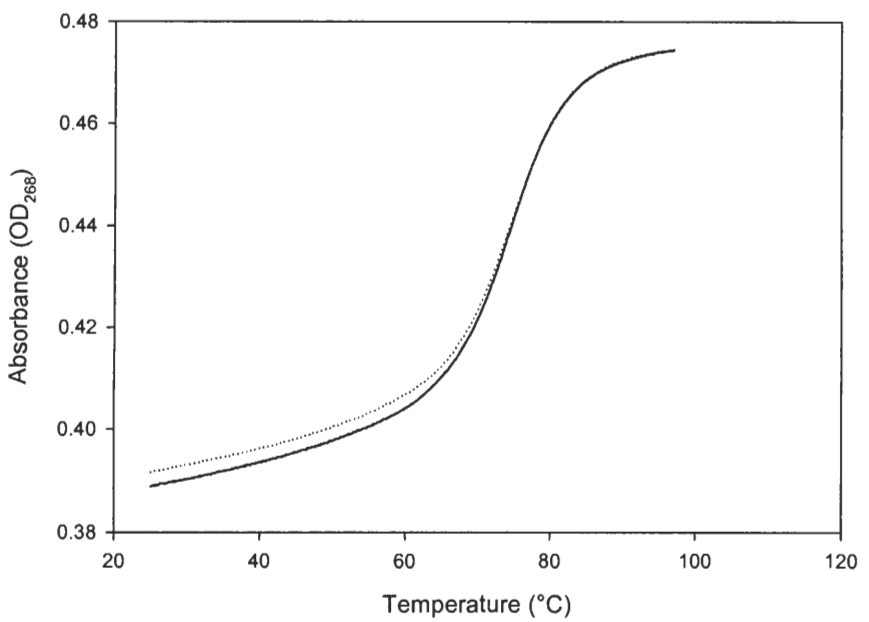


Figure 5.3 (continued). Typical UV denaturation curves of dangling end hairpins in 0.1 M Na^+. Solid lines represent denaturation curves while dotted lines represent renaturation curves. G. C/GA. H. TC/GA.

Table 5.1. Van't Hoff analysis of formation of dangling end hairpins in 0.1 M Na^+.

DNA hairpin	T_m (°C)	-ΔH (kcal/mol)	-ΔS (eu)	$-\Delta G_{37}$ (kcal/mol)
A/T	73.2 ± 0.2	55.3 ± 3.1	160 ± 9	5.79 ± 0.31
TA/T	72.5 ± 0.3	52.9 ± 1.7	153 ± 5	5.44 ± 0.19
A/TA	72.6 ± 0.4	55.2 ± 1.4	160 ± 5	5.69 ± 0.14
TA/TA	72.4 ± 0.2	59.2 ± 1.4	171 ± 4	6.06 ± 0.17
C/G	74.6 ± 0.3	59.0 ± 3.2	170 ± 9	6.38 ± 0.33
TC/G	73.6 ± 0.2	55.8 ± 2.7	161 ± 8	5.90 ± 0.29
C/GA	73.3 ± 0.2	55.3 ± 2.7	160 ± 8	5.80 ± 0.31
TC/GA	73.8 ± 0.3	58.6 ± 1.9	169 ± 6	6.22 ± 0.23

of the two corresponding dangling ends as the contribution of the paired bases. The experimental results agree well with the free energy value of the first structure containing nine Watson-Crick pairs.

The results for the C-G core sequence also agree well with free energies estimated by Zuker. For the C/G molecule the experimental and predicted results were –6.38 ± 0.33 kcal/mol and –6.2 kcal/mol. For the TC/GA hairpin, the experimental free energy was 6.22 ± 0.23 kcal/mol. Zuker's program predicted a free energy of –7.2 kcal/mol for the completely base paired stem, and a free energy of –7.4 kcal/mol for the stem with an unpaired terminus. Although the free energies do not agree as well, it is important to realize that a small difference in the entropy change for this transition would provide better agreement. The average ΔS was found to be -169 ± 6 eu. If the entropy change is taken as -166 eu (within experimental error), and the enthalpy is left as -58.6 kcal/mol, the resulting free energy would be –7.2 kcal/mol. The differences between the estimated and experimental free energies are within experimental error.

The addition of dangling bases caused a slight decrease in the melting temperature of the molecule, as does the addition of a T-A base pair. Dangling bases were found to be somewhat destabilizing in all cases. This is counterintuitive to what is expected for extending a helix. To determine the thermodynamic contribution of the dangling ends, the values for the core sequence was subtracted from the experimental values for the dangling end molecules. In the case of the TA/T end, the thermodynamic contributions were found to be ΔH = 2.4 kcal/mol, ΔS = 7 eu, and ΔG = 0.35 kcal/mol. For the A/TA end, the values were ΔH = 0.1 kcal/mol, ΔS = 0 eu, and ΔG = 0.10 kcal/mol, suggesting that a 3' adenine has little effect on the overall stablility of a hairpin with an A-T closing base pair. The addition of a ninth Watson-Crick base pair was stabilizing for the core sequence with an A-T terminus. Here the thermodynamic contributions were ΔH = -3.9 kcal/mol, ΔS = -11 eu, and ΔG = -0.27 kcal/mol. These

values do not show the additional base to be as stabilizing as nearest-neighbor stacking interactions would generally predict, however, as discussed in the previous paragraph, the results are close to the estimated values from Zuker. This may be the result of non-nearest-neighbor interactions or end effects.

For the core sequence ending in a C-G terminus, the dangling bases were destabilizing. The 5' dangling end (TC/G) was found to have the following thermodynamic contribution: ΔH = 3.2 kcal/mol, ΔS = 9 eu, and ΔG = 0.48 kcal/mol. The C/GA end contributed ΔH = 3.7 kcal/mol, ΔS = 10 eu, and ΔG = 0.58 kcal/mol. The additional base pair had little effect on the thermodynamic values with contributions of ΔH = 0.4 kcal/mol, ΔS = 1 eu, and ΔG = 0.16 kcal/mol. The differences between these values and estimated values are within experimental error.

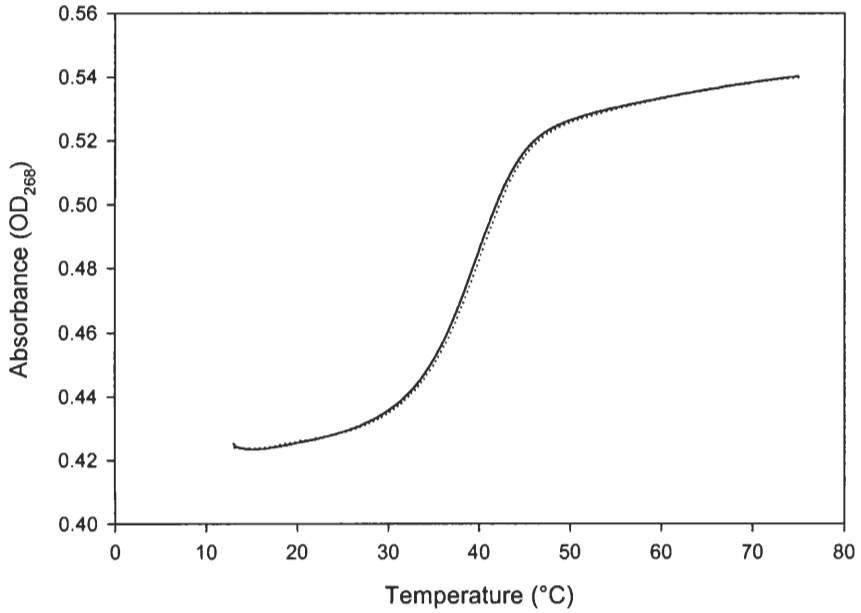
2'-O-methyl chimeras

Figure 5.4 shows typical melting curves for each of the DNA and chimeric duplexes. Solid lines represent denaturation and dotted lines represent renaturation. All molecules were highly renaturable. Little evaporation was seen because the low melting temperatures of the duplexes allowed for heating over a lower temperature range.

The thermodynamic values for each of the duplexes are given in table 5.2. The 5040 molecules contain two 2' OMe U-A closing base pairs around the DNA region. The 2'-O-methyl modification lowered the melting temperature by several degrees in comparison with the corresponding DNA duplex. Although the enthalpy and entropy for these duplexes are similar, the difference in melting temperature caused the free energy of formation of the chimera to be positive, while it is negative for the DNA/DNA duplex. The chimera is almost 2 kcal/mol less stable than the corresponding DNA duplex.

The 5041 chimera had a 2' OMe A-T and a 2' OMe C-G as the closing base pairs around the DNA region. The melting temperature of the chimeric duplex was

A.



B.

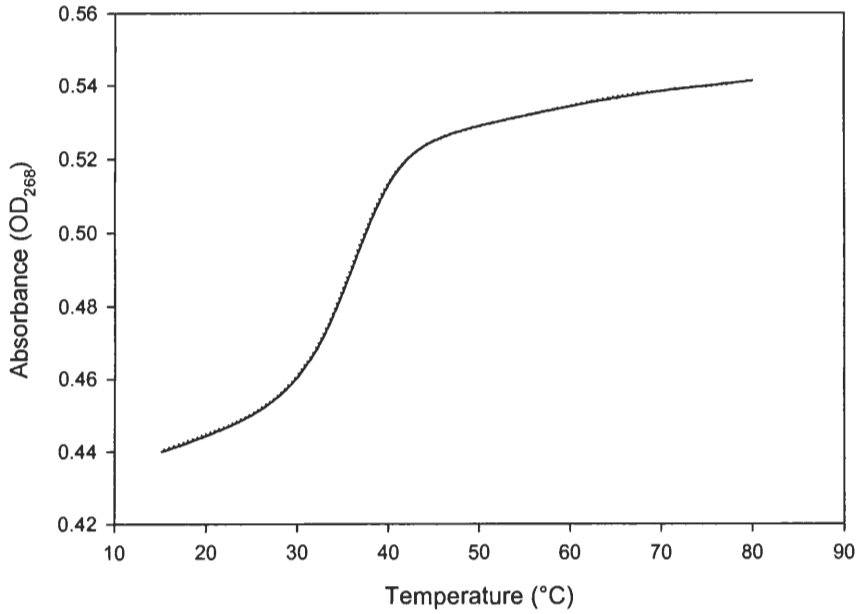
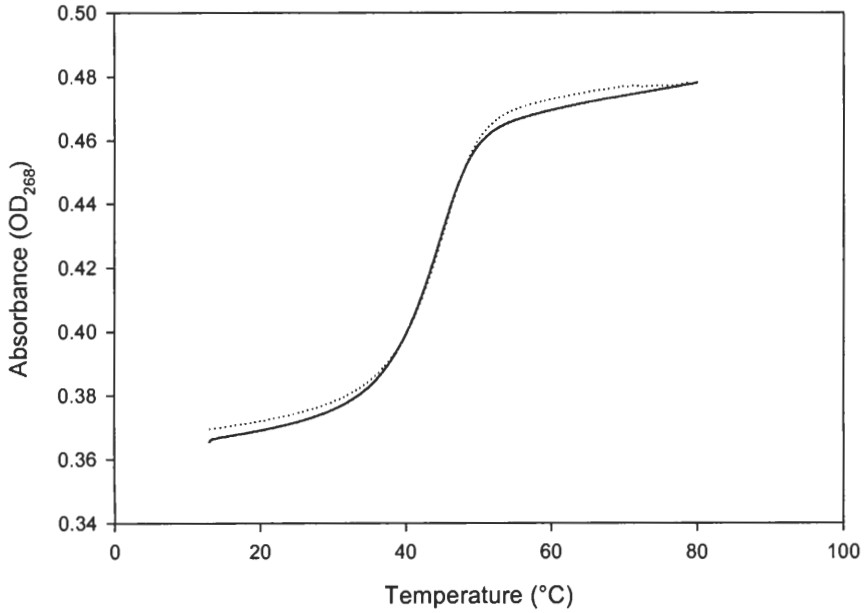


Figure 5.4. Typical UV denaturation curves for 2'-O-methyl chimera/DNA and corresponding DNA/DNA duplexes in 0.1 M Na^+. Solid lines represent denaturation while dotted lines represent renaturation. A. 5040 DNA duplex. B. 5040 2'-O-methyl duplex.

C.



D.

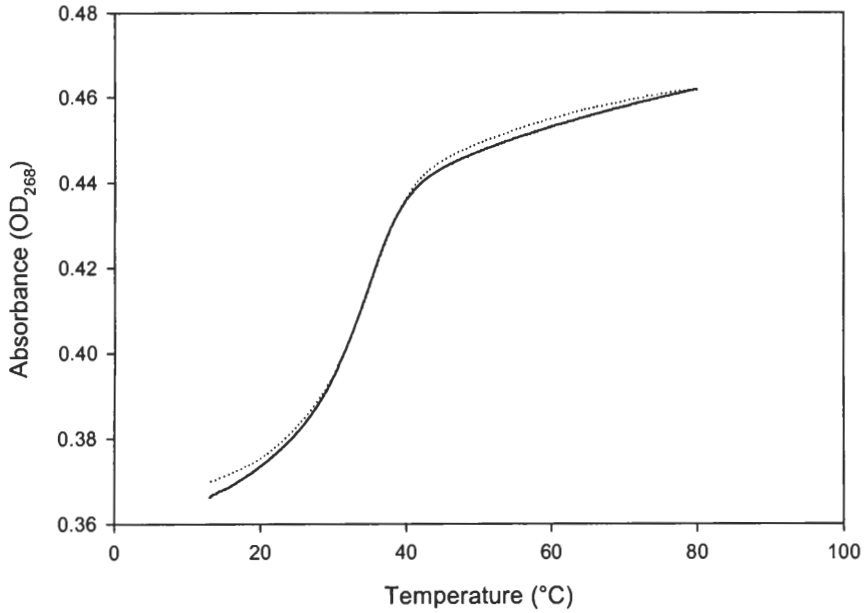
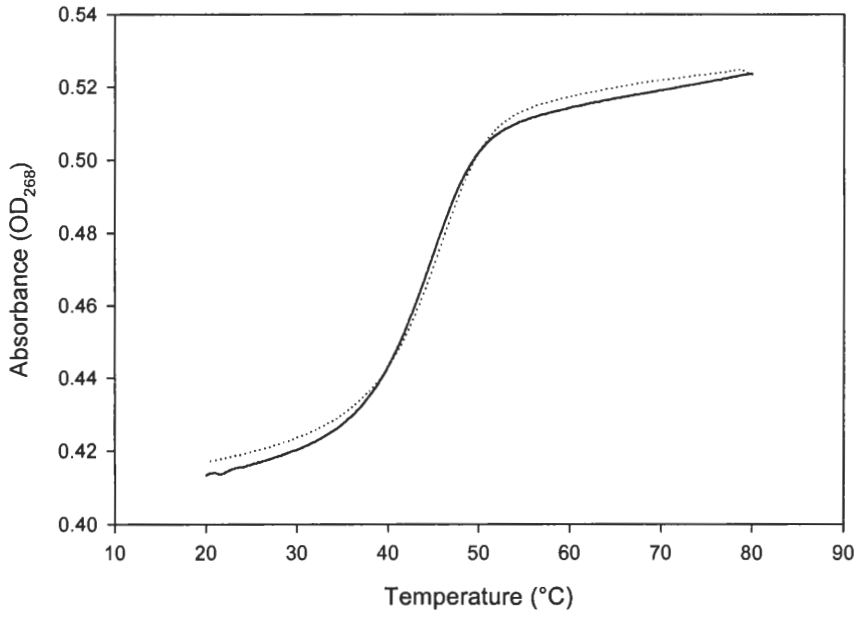


Figure 5.4 (continued). Typical UV denaturation curves for 2'-O-methyl chimera/DNA and corresponding DNA/DNA duplexes in 0.1 M Na^+. Solid lines represent denaturation while dotted lines represent renaturation. C. 5041 DNA duplex. D. 5041 2'-O-methyl duplex.

E.



F.

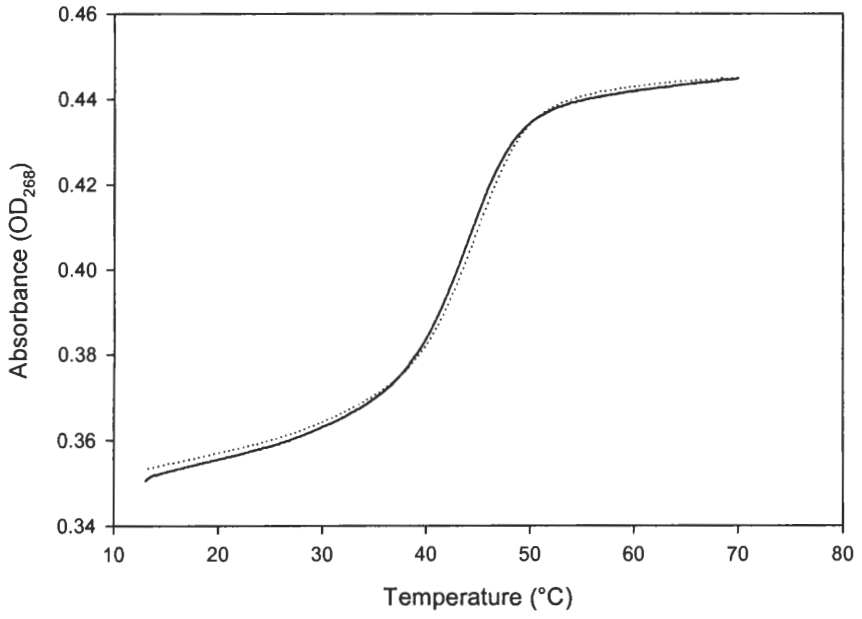


Figure 5.4 (continued). Typical UV denaturation curves for 2'-O-methyl chimera/DNA and corresponding DNA/DNA duplexes in 0.1 M Na^+. Solid lines represent denaturation while dotted lines represent renaturation. E. 5020 DNA duplex. F. 5020 2'-O-methyl duplex.

Table 5.2. Van't Hoff analysis of formation of DNA duplexes and 2'-O-methyl chimeras in 0.1M Na^+.

Duplex	T_m (°C)	ΔH (kcal/mol)	ΔS (eu)	ΔG_{37} (kcal/mol)
5040 DNA	39.3 ± 0.1	-98.7 ± 1.7	-316 ± 6	-0.73 ± 0.03
5040 2'-O-methyl	33.5 ± 0.3	-100.2 ± 3.4	-327 ± 11	1.15 ± 0.10
5041 DNA	43.5 ± 0.1	-98.2 ± 2.4	-310 ± 8	-2.03 ± 0.03
5041 2'-O-methyl	33.2 ± 0.2	-81.99 ± 2.7	-268 ± 9	1.03 ± 0.08
5020 DNA	43.8 ± 0.1	-97.9 ± 1.7	-309 ± 5	-2.10 ± 0.05
5020 2'-O-methyl	43.2 ± 0.2	-106.3 ± 4.6	-336 ± 15	-2.09 ± 0.10

approximately ten degrees lower than the T_m of the corresponding DNA. There were larger differences between the enthalpy and entropy than seen in the 5040 molecules. For the 5041 molecule, the chimera again gave a positive free energy of formation while the DNA duplex had a negative free energy of formation. The DNA duplex was 3 kcal/mol more stable than the chimeric duplex.

The 5020 chimera contained two 2' OMe C-G closing base pairs surrounding the DNA region. This type of closing base pair would be expected to be among the most stable. The difference in T_m between the chimeric and DNA duplexes was approximately 0.5 °C. Although there were slight differences in the entropies and enthalpies of the chimeric and DNA duplexes, the free energy values were nearly identical. This suggests that the thermodynamics of chimera/DNA duplexes and DNA/DNA duplexes are similar if the interface between the modified RNA regions and DNA region is stabilized using appropriate base pairs.

Discussion

The thermodynamic values of dangling ends are important for structure prediction and immobilized probe design. Previous studies have examined hairpins with long dangling ends and duplexes with single-nucleotide and longer dangling ends. This work examined 5' and 3' single, unpaired terminal bases in hairpin molecules.

Single-nucleotide dangling end thermodynamics depend on both the sequence of the closing base and the unpaired base. Bommarito *et al.* examined the 32 possible sequences in duplex molecules. The design of this experiment was such that there were two identical dangling ends, and the energetic contribution attributed equally to each [86]. By using hairpin molecules in this work, it was possible to determine the thermodynamic values of a single dangling end. The experimental free energy values for the completely Watson-Crick paired stems agree well with estimated values from

Zuker. In comparing the results for the contribution of the dangling ends to values from the Bommarito paper, however, there are significant differences. Bommarito's results imply that the sequences examined here are all slightly stabilizing to duplexes. The free energy values for the contribution of the dangling ends in the Bommarito duplexes ranged from –0.48 kcal/mol to –0.92 kcal/mol. For the hairpin molecules studied here, the dangling ends were found to be slightly destabilizing, with free energies ranging from 0.1 to 0.58 kcal/mol. These differences may be due to non-nearest-neighbor effects. The base pair next to the terminal base pair was a C-G or a G-C in the Bommarito paper, while it was a T-A in the hairpins studied here. Work by Hall *et al.* compared two 31 base pair duplexes [36]. A single mismatch was made at position 5 in one duplex and at position 23 in the other. Each of the mismatches had the same closing base pairs, however the next adjacent bases were different. The T_m's for these molecules differed by 2.5 °C. The ΔH differed by 30 kcal / mol, the ΔS by 94 eu, and the ΔG by 2 kcal / mol. This indicates that non-nearest-neighbors have influence on the thermodynamic parameters.

The *mfold* program predicts two structures for each of the hairpins with nine bases in the stem. One structure has all nine bases paired, while the other structure has eight paired bases and an unpaired terminus. The free energy results for the hairpins examined here are more similar to the estimations for the completely paired stems, suggesting that the molecules are present in that form.

2'-O-methyl modified RNA bases are of interest because they have not only been shown to bind to targets at different rates and with different melting temperatures as compared to unmodified bases, but also because chimeric molecules utilizing these bases may provide a method for mutation repair in cells. In order to properly design probes and hybridization conditions, it is necessary to understand the thermodynamics of chimeric duplexes containing the modified bases. Here, three chimeric molecules and

their corresponding DNA's were hybridized to DNA targets and their thermodynamic values compared. Each of the molecules had similar G-C content in both the DNA region and the 2'-O-methyl regions. One chimera was found to have similar thermodynamic parameters to the corresponding DNA. This chimera had two 2' OMe C-G closing base pairs around the DNA region of the strand. This may indicate that it is important to stabilize the interface between the DNA and modified RNA regions. This study suggests that less stable closing base pairs may not allow convenient conditions for hybridization reactions and should be taken into account in experimental design.

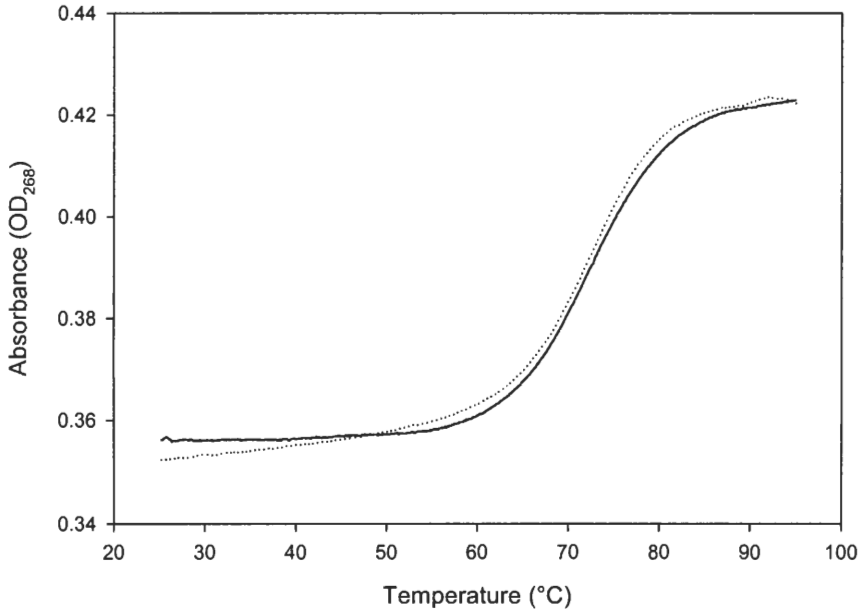
In studies on mutation correction in cells, it was found that the DNA region of a molecule directs the repair mechanism in the cell [92]. RNA bases help stabilize the repair complex, however mutation repair was less effective when a vector containing all RNA bases was utilized. These results may be due, in part, to the presence of nucleases in the cell. Both DNA and RNA strands would be susceptible to these nucleases, and could be degraded before mutation correction could occur. To avoid degradation, 2'-O-methyl nucleotides, which are resistant to nuclease cleavage, were used in chimeric molecules. The chimeras were found to be less active in mutation repair than DNA vectors or RNA/DNA chimeras, and were inactive when large numbers of 2'-O-methyl nucleotides were used. The results presented here suggest that this inactivity may be due to the thermodynamic parameters of the 2'-O-methyl chimeras. The molecules that had two 2' OMe A-U's or one 2' OMe A-U and one 2' OMe G-C as the closing base pairs surrounding the DNA region were less stable than their corresponding DNA/DNA duplexes. The mutation repair system described by Gamper *et al.* [92] relies on the ability of the chimera to compete with a cellular DNA strand to hybridize to the complementary DNA. If the thermodynamic stability of the chimera is much less than the DNA/DNA duplex, it will be unable to compete and will not be bound to the target. Mutation correction would then not occur, and may account for the lowered

levels of repair observed with the 2'-O-methyl chimeras. This emphasizes the necessity of obtaining accurate thermodynamic data about these molecules.

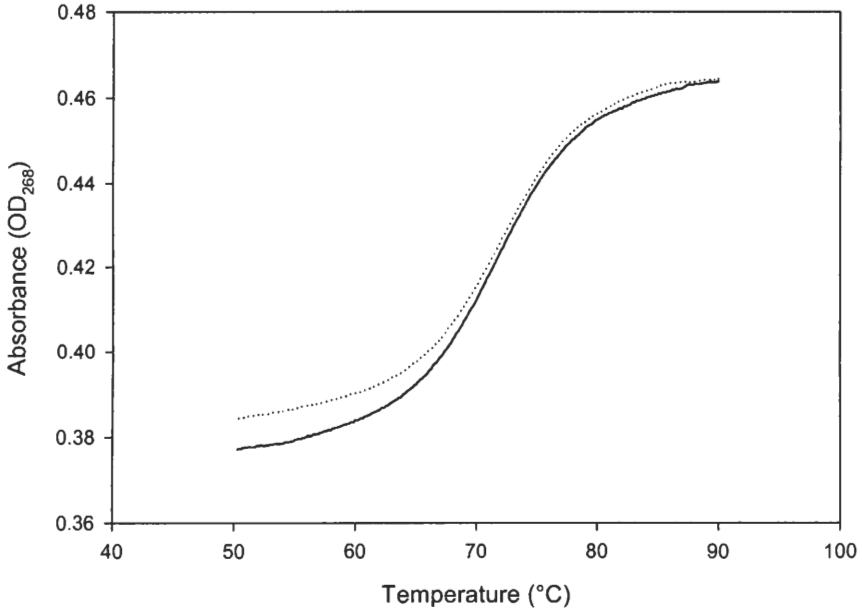
APPENDIX 1

UV DENATURATION CURVES IN 0.1 M SODIUM WITH 5 mM MAGNESIUM

A.

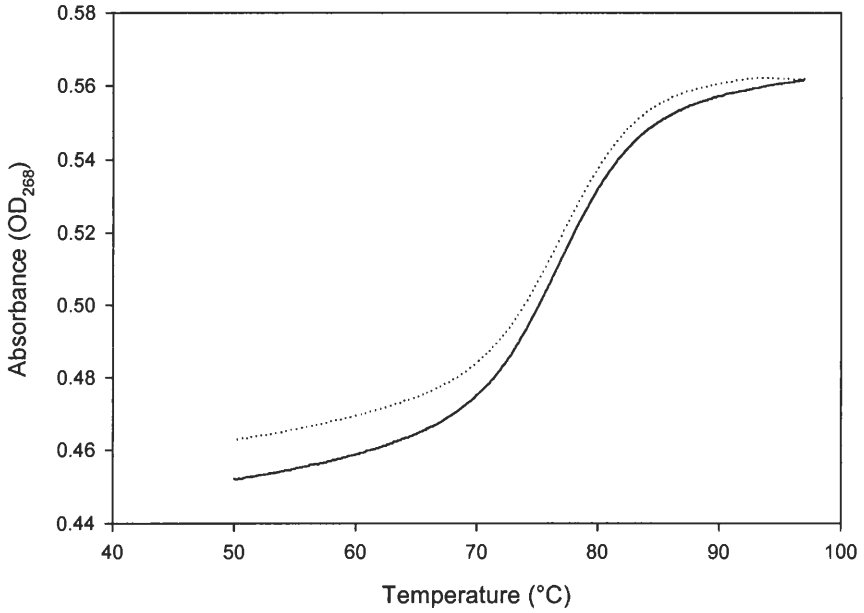


B.

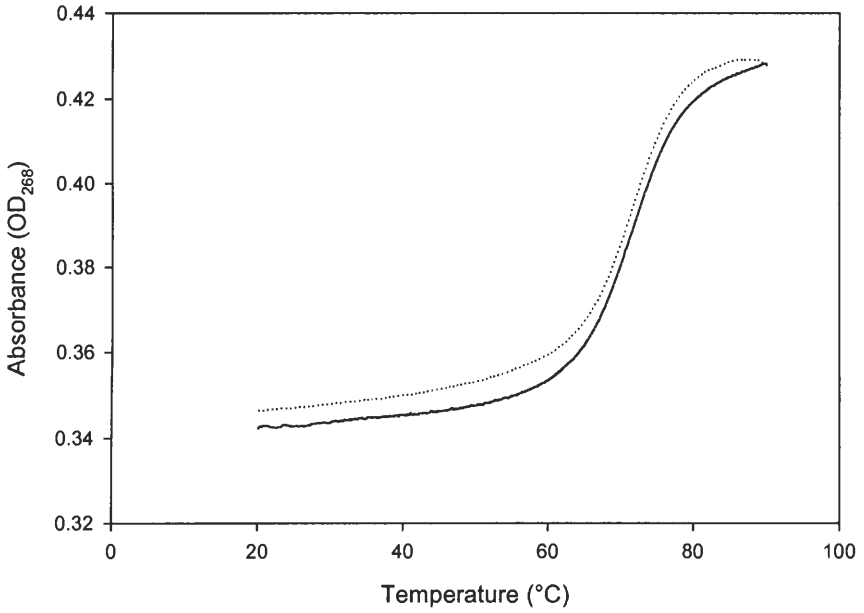


Typical UV denaturation curves of short DNA hairpins containing tandem mismatches in 5 mM Mg^{2+} plus 0.1 M Na^{+}. Solid lines represent denaturation while dotted lines represent renaturation. A. --/--. B. AA/TT.

C.

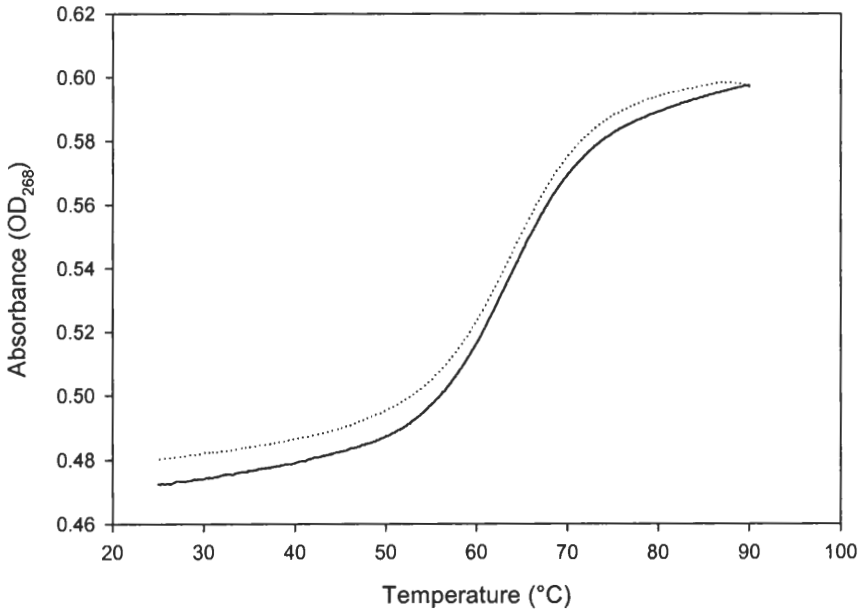


D.

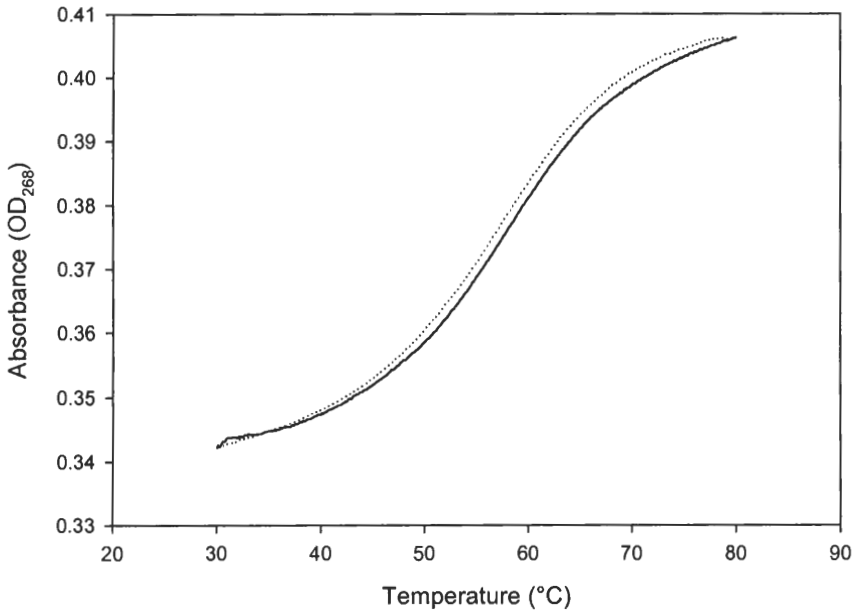


Typical UV denaturation curves of short DNA hairpins containing tandem mismatches in 5 mM Mg^{2+} plus 0.1 M Na^{+}. Solid lines represent denaturation while dotted lines represent renaturation. C. GA/TC. D. ga/ga.

E.

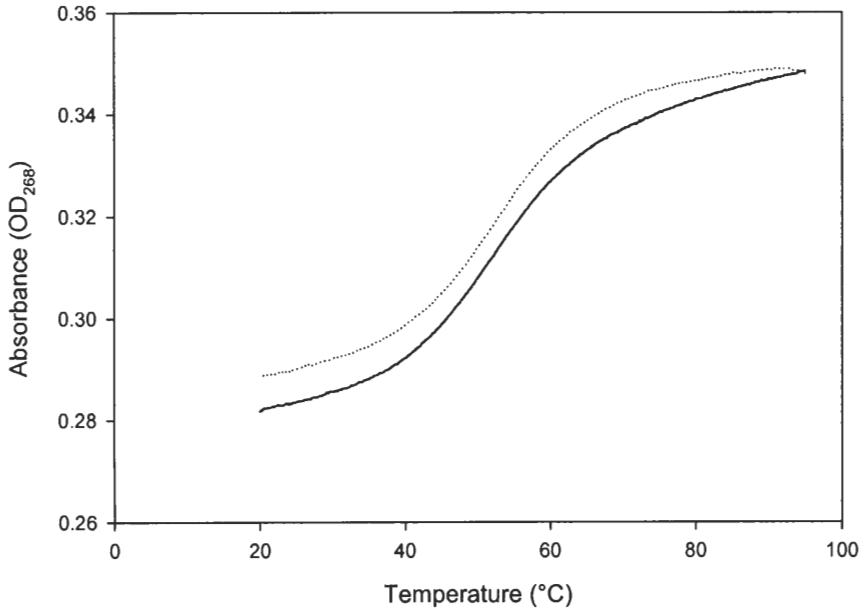


F.

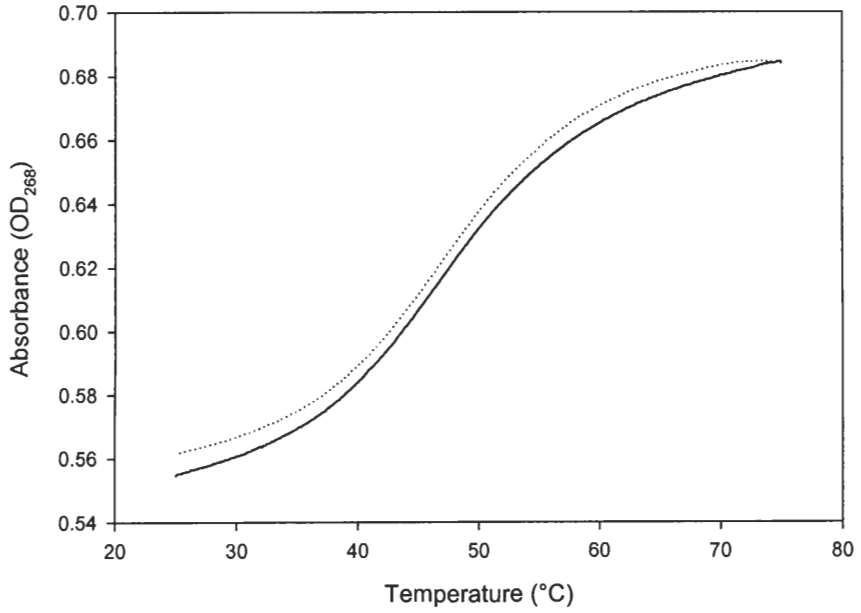


Typical UV denaturation curves of short DNA hairpins containing tandem mismatches in 5 mM Mg^{2+} plus 0.1 M Na^{+}. Solid lines represent denaturation while dotted lines represent renaturation. E. aa/gc. F. ca/gc.

G.



H.

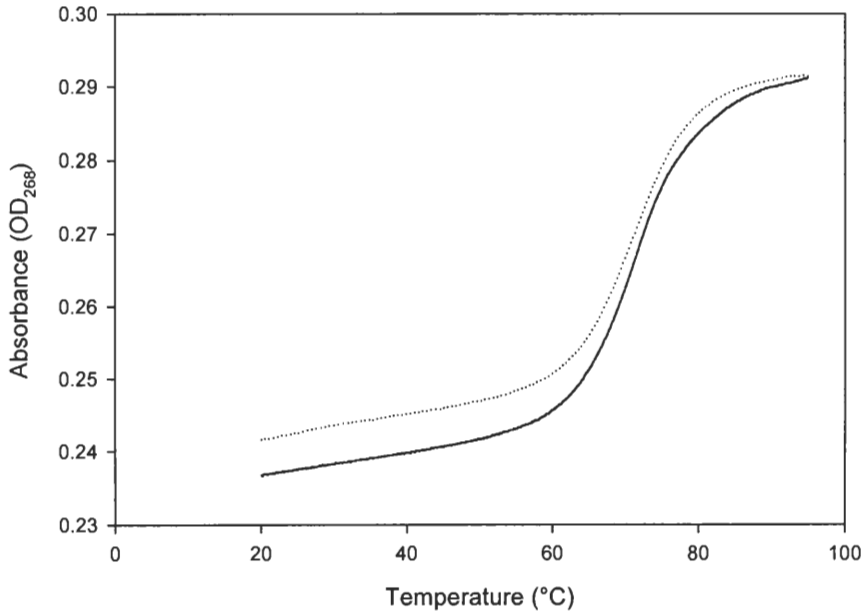


Typical UV denaturation curves of short DNA hairpins containing tandem mismatches in 5 mM Mg^{2+} plus 0.1 M Na^{+}. Solid lines represent denaturation while dotted lines represent renaturation. G. ta/ac. H. tc/tc.

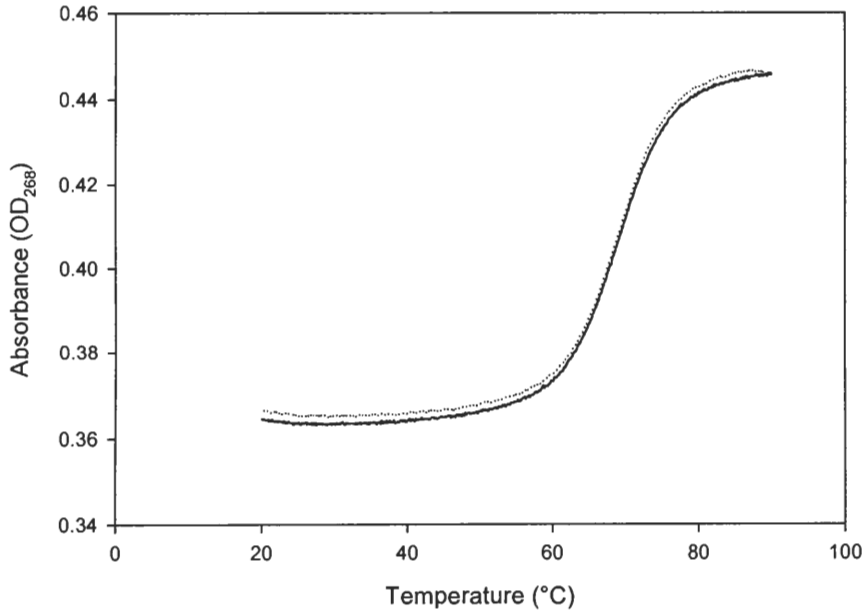
APPENDIX 2

UV DENATURATION CURVES IN 0.1 M SODIUM

A.

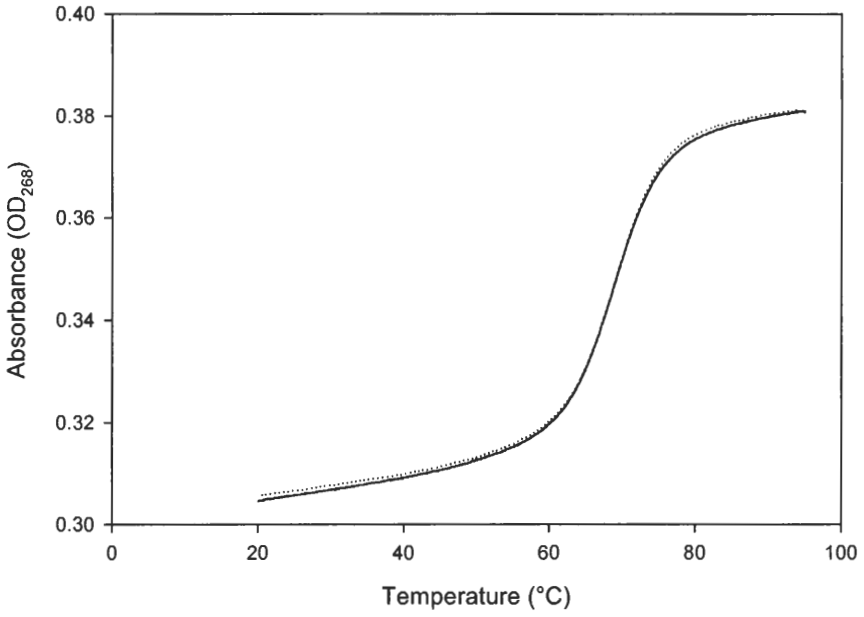


B.

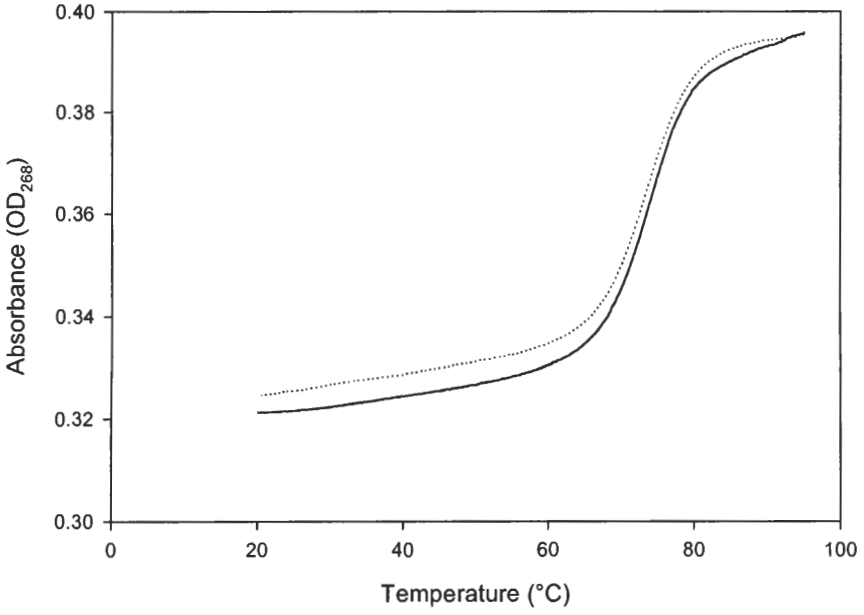


Typical UV denaturation curves of short DNA hairpins containing tandem mismatches in 0.1 M Na^+. Solid lines represent denaturation while dotted lines represent renaturation. A. --/--. B. A/T.

C.

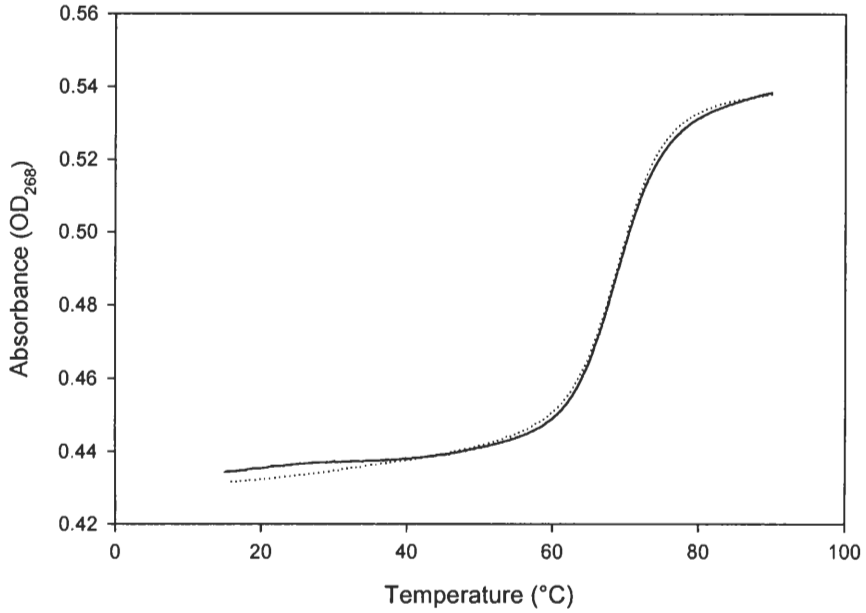


D.

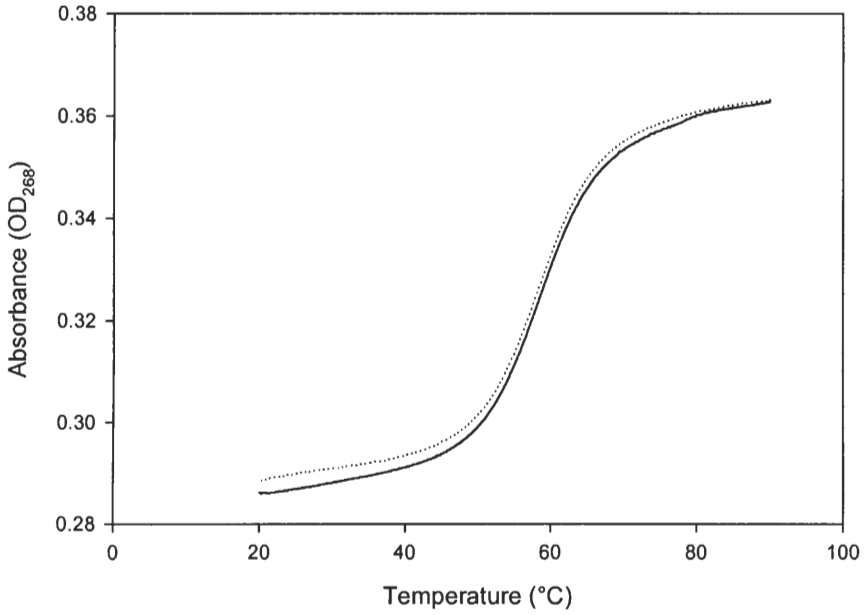


Typical UV denaturation curves of short DNA hairpins containing tandem mismatches in 0.1 M Na^+. Solid lines represent denaturation while dotted lines represent renaturation. C. AA/TT. D. GA/TC.

E.

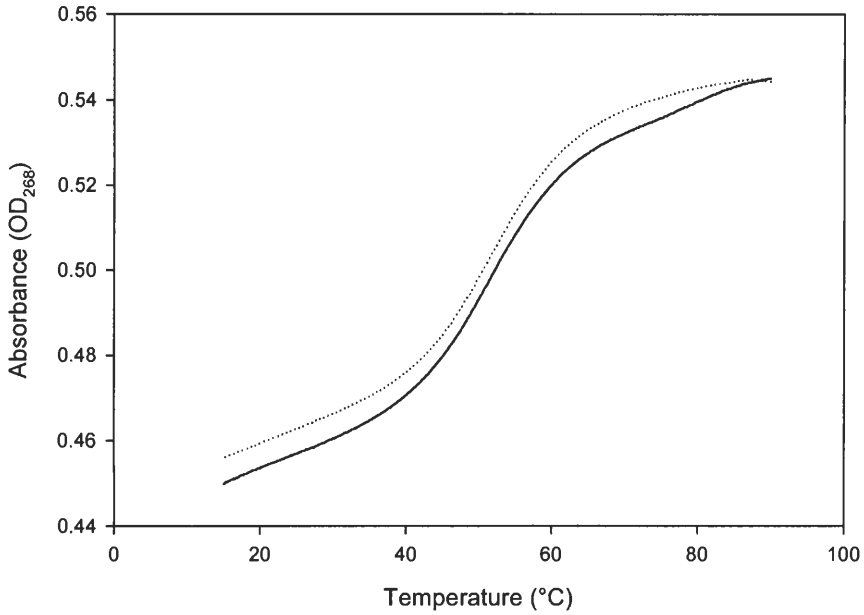


F.

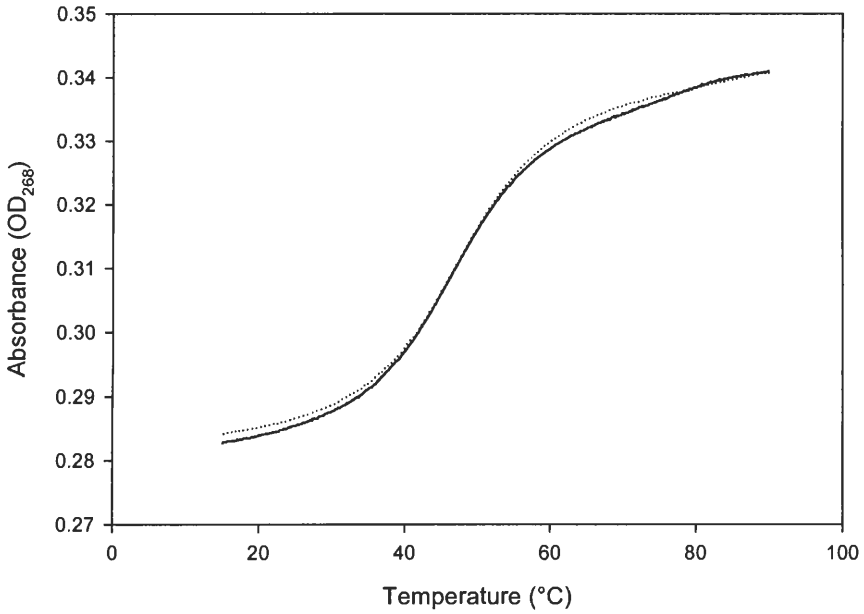


Typical UV denaturation curves of short DNA hairpins containing tandem mismatches in 0.1 M Na^+. Solid lines represent denaturation while dotted lines represent renaturation. E. ga/ga. F. aa/gc.

G.

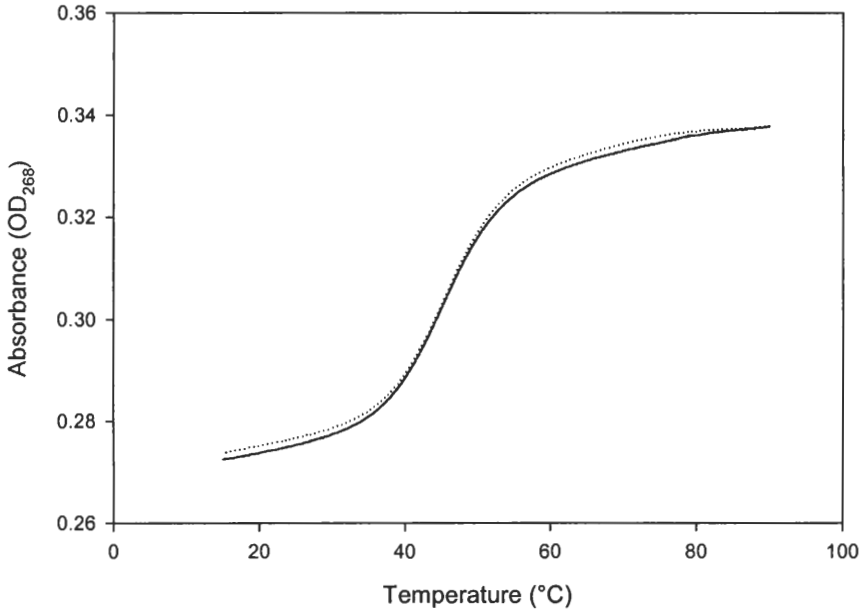


H.



Typical UV denaturation curves of short DNA hairpins containing tandem mismatches in 0.1 M Na^+. Solid lines represent denaturation while dotted lines represent renaturation. G. ca/gc. H. ta/ac.

I.



Typical UV denaturation curves of short DNA hairpins containing tandem mismatches in 0.1 M Na^+. Solid lines represent denaturation while dotted lines represent renaturation. I. tc/tc.

REFERENCES

1. Bloomfield, V.A., D.M. Crothers, and I. Tinoco, Jr., *Nucleic Acids: Structures, Properties, and Functions*. 2000, Sausalito, CA: University Science Books. 794.

2. Zimm, B.H. and J.K. Bragg, J. Chem. Phys., 1959. **31**: p. 526-35.

3. Crothers, D.M. and B.H. Zimm, J Mol Biol, 1964. **9**: p. 1-9.

4. Gray, D.M. and I. Tinoco, Jr., Biopolymers, 1970. **9**: p. 223-44.

5. Borer, P.N., et al., *Stability of ribonucleic acid double-stranded helices.* J Mol Biol, 1974. **86**(4): p. 843-53.

6. Uhlenbeck, O.C., et al., *Stability of RNA hairpin loops: A 6 -C m -U 6.* J Mol Biol, 1973. **73**(4): p. 483-96.

7. Tinoco, I., Jr., et al., *Improved estimation of secondary structure in ribonucleic acids.* Nat New Biol, 1973. **246**(150): p. 40-1.

8. DeVoe, H. and I. Tinoco, Jr., J Mol Biol, 1962. **4**: p. 500-17.

9. Wartell, R.M. and A.S. Benight, *Thermal denaturation of DNA molecules: a comparison of theory with experiment.* Physics Reports, 1985. **126**(2): p. 67-107.

10. Gotoh, O. and Y. Tagashira, Biopolymers, 1981. **20**: p. 1033-42.

11. Vologodskii, A.V., et al., *Allowance for heterogeneous stacking in the DNA helix-coil transition theory.* J Biomol Struct Dyn, 1984. **2**(1): p. 131-48.

12. McCampbell, C.R., R.M. Wartell, and R.R. Plaskon, *Inverted repeat sequences can influence the melting transitions of linear DNAs.* Biopolymers, 1989. **28**(10): p. 1745-58.

13. Delcourt, S.G. and R.D. Blake, *Stacking energies in DNA.* J Biol Chem, 1991. **266**(23): p. 15160-9.

14. Breslauer, K.J., et al., *Predicting DNA duplex stability from the base sequence.* Proc Natl Acad Sci U S A, 1986. **83**(11): p. 3746-50.

15. SantaLucia, J., Jr., H.T. Allawi, and P.A. Seneviratne, *Improved nearest-neighbor parameters for predicting DNA duplex stability.* Biochemistry, 1996. **35**(11): p. 3555-62.

16. Allawi, H.T. and J. SantaLucia, Jr., *Thermodynamics and NMR of internal G.T mismatches in DNA.* Biochemistry, 1997. **36**(34): p. 10581-94.

17. Sugimoto, N., et al., *Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes.* Nucleic Acids Res, 1996. **24**(22): p. 4501-5.

18. Aida, M., *An ab initio molecular orbital study on the sequence-dependency of DNA conformation: an evaluation of intra- and inter-strand stacking interaction energy.* J Theor Biol, 1988. **130**(3): p. 327-35.

19. Ornstein, R.L. and J.R. Fresco, *Correlation of Tm and sequence of DNA duplexes with delta H computed by an improved empirical potential method.* Biopolymers, 1983. **22**(8): p. 1979-2000.

20. Doktycz, M.J., et al., *Studies of DNA dumbbells. I. Melting curves of 17 DNA dumbbells with different duplex stem sequences linked by T4 endloops: evaluation of the nearest-neighbor stacking interactions in DNA.* Biopolymers, 1992. **32**(7): p. 849-64.

21. Owczarzy, R., et al., *Predicting sequence-dependent melting stability of short duplex DNA oligomers.* Biopolymers, 1997. **44**(3): p. 217-39.

22. SantaLucia, J., Jr., *A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics.* Proc Natl Acad Sci U S A, 1998. **95**(4): p. 1460-5.

23. Freier, S.M., et al., *Improved free-energy parameters for predictions of RNA duplex stability.* Proc Natl Acad Sci U S A, 1986. **83**(24): p. 9373-7.

24. Xia, T., et al., *Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs.* Biochemistry, 1998. **37**(42): p. 14719-35.

25. Sugimoto, N., K. Honda, and M. Sasaki, Nucleosides Nucleotides, 1994. **13**: p. 1311-7.

26. Sugimoto, N., et al., *Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes.* Biochemistry, 1995. **34**(35): p. 11211-6.

27. Longfellow, C.E., R. Kierzek, and D.H. Turner, *Thermodynamic and spectroscopic study of bulge loops in oligoribonucleotides.* Biochemistry, 1990. **29**(1): p. 278-85.

28. Groebe, D.R. and O.C. Uhlenbeck, *Thermal stability of RNA hairpins containing a four-membered loop and a bulge nucleotide.* Biochemistry, 1989. **28**(2): p. 742-7.

29. Ke, S.H. and R.M. Wartell, *Influence of neighboring base pairs on the stability of single base bulges and base pairs in a DNA fragment.* Biochemistry, 1995. **34**(14): p. 4593-600.

30. Zhu, J. and R.M. Wartell, *The effect of base sequence on the stability of RNA and DNA single base bulges.* Biochemistry, 1999. **38**(48): p. 15986-93.

31. Znosko, B.M., et al., *Thermodynamic parameters for an expanded nearest-neighbor model for the formation of RNA duplexes with single nucleotide bulges.* Biochemistry, 2002. **41**(33): p. 10406-17.

32. Allawi, H.T. and J. SantaLucia, Jr., *Nearest neighbor thermodynamic parameters for internal G.A mismatches in DNA.* Biochemistry, 1998. **37**(8): p. 2170-9.

33. Allawi, H.T. and J. SantaLucia, Jr., *Nearest-neighbor thermodynamics of internal A.C mismatches in DNA: sequence dependence and pH effects.* Biochemistry, 1998. **37**(26): p. 9435-44.

34. Allawi, H.T. and J. SantaLucia, Jr., *Thermodynamics of internal C.T mismatches in DNA.* Nucleic Acids Res, 1998. **26**(11): p. 2694-701.

35. Peyret, N., et al., *Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches.* Biochemistry, 1999. **38**(12): p. 3468-77.

36. Hall, T.S., et al., *Sequence context and thermodynamic stability of a single base pair mismatch in short deoxyoligonucleotide duplexes.* J Am Chem Soc, 2001. **123**(47): p. 11811-2.

37. Kierzek, R., M.E. Burkard, and D.H. Turner, *Thermodynamics of single mismatches in RNA duplexes.* Biochemistry, 1999. **38**(43): p. 14214-23.

38. Meroueh, M. and C.S. Chow, *Thermodynamics of RNA hairpins containing single internal mismatches.* Nucleic Acids Res, 1999. **27**(4): p. 1118-25.

39. Zhu, J. and R.M. Wartell, *The relative stabilities of base pair stacking interactions and single mismatches in long RNA measured by temperature gradient gel electrophoresis.* Biochemistry, 1997. **36**(49): p. 15326-35.

40. Ke, S.H. and R.M. Wartell, *The thermal stability of DNA fragments with tandem mismatches at a d(CXYG).d(CY'X'G) site.* Nucleic Acids Res, 1996. **24**(4): p. 707-12.

41. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Res, 2003. **31**(13): p. 3406-15.

42. Li, Y., G. Zon, and W.D. Wilson, *Thermodynamics of DNA duplexes with adjacent G.A mismatches.* Biochemistry, 1991. **30**(30): p. 7566-72.

43. Ebel, S., A.N. Lane, and T. Brown, *Very stable mismatch duplexes: structural and thermodynamic studies on tandem G.A mismatches in DNA.* Biochemistry, 1992. **31**(48): p. 12083-6.

44. Li, Y. and S. Agrawal, *Oligonucleotides containing G.A pairs: effect of flanking sequences on structure and stability.* Biochemistry, 1995. **34**(31): p. 10056-62.

45. Zhu, J., *The stabilities of RNA and DNA structural elements*, in *School of Biology*. 1998, Georgia Institute of Technology: Atlanta, GA. p. 186.

46. SantaLucia, J., Jr., R. Kierzek, and D.H. Turner, *Stabilities of consecutive A.C, C.C, G.G, U.C, and U.U mismatches in RNA internal loops: Evidence for stable hydrogen-bonded U.U and C.C.+ pairs.* Biochemistry, 1991. **30**(33): p. 8242-51.

47. Walter, A.E., M. Wu, and D.H. Turner, *The stability and structure of tandem GA mismatches in RNA depend on closing base pairs.* Biochemistry, 1994. **33**(37): p. 11349-54.

48. Wu, M., J.A. McDowell, and D.H. Turner, *A periodic table of symmetric tandem mismatches in RNA.* Biochemistry, 1995. **34**(10): p. 3204-11.

49. McDowell, J.A. and D.H. Turner, *Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: solution structure of (rGAGGUCUC)2 by two-dimensional NMR and simulated annealing.* Biochemistry, 1996. **35**(45): p. 14077-89.

50. Burkard, M.E., T. Xia, and D.H. Turner, *Thermodynamics of RNA internal loops with a guanosine-guanosine pair adjacent to another non-canonical pair.* Biochemistry, 2001. **40**(8): p. 2478-83.

51. Xia, T., J.A. McDowell, and D.H. Turner, *Thermodynamics of nonsymmetric tandem mismatches adjacent to G.C base pairs in RNA.* Biochemistry, 1997. **36**(41): p. 12486-97.

52. Schroeder, S.J. and D.H. Turner, *Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA.* Biochemistry, 2000. **39**(31): p. 9257-74.

53. Goodman, M.F., et al., *Biochemical basis of DNA replication fidelity.* Crit Rev Biochem Mol Biol, 1993. **28**(2): p. 83-126.

54. Bhattacharyya, A. and D.M. Lilley, *Single base mismatches in DNA. Long- and short-range structure probed by analysis of axis trajectory and local chemical reactivity.* J Mol Biol, 1989. **209**(4): p. 583-97.

55. Leonard, G.A., E.D. Booth, and T. Brown, *Structural and thermodynamic studies on the adenine.guanine mismatch in B-DNA.* Nucleic Acids Res, 1990. **18**(19): p. 5617-23.

56. Brown, T., Aldrichimica Acta, 1995. **28**: p. 15-20.

57. Saiki, R.K., et al., *Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase.* Science, 1988. **239**(4839): p. 487-91.

58. Southern, E.M., *Detection of specific sequences among DNA fragments separated by gel electrophoresis.* J Mol Biol, 1975. **98**(3): p. 503-17.

59. Fodor, S.P., et al., *Multiplexed biochemical assays with biological chips.* Nature, 1993. **364**(6437): p. 555-6.

60. Wickstrom, E.L., et al., *Human promyelocytic leukemia HL-60 cell proliferation and c-myc protein expression are inhibited by an antisense pentadecadeoxynucleotide targeted against c-myc mRNA.* Proc Natl Acad Sci U S A, 1988. **85**(4): p. 1028-32.

61. O'Donnell-Maloney, M.J. and D.P. Little, *Microfabrication and array technologies for DNA sequencing and diagnostics.* Genet Anal, 1996. **13**(6): p. 151-7.

62. Bolton, E.T. and C.B. Mc, *A general method for the isolation of RNA complementary to DNA.* Proc Natl Acad Sci U S A, 1962. **48**: p. 1390-7.

63. Baldino, F., Jr., M.F. Chesselet, and M.E. Lewis, *High-resolution in situ hybridization histochemistry.* Methods Enzymol, 1989. **168**: p. 761-77.

64. Rouzina, I. and V.A. Bloomfield, *Heat capacity effects on the melting of DNA. 1. General aspects.* Biophys J, 1999. **77**(6): p. 3242-51.

65. Holbrook, J.A., et al., *Enthalpy and heat capacity changes for formation of an oligomeric DNA duplex: interpretation in terms of coupled processes of formation and association of single-stranded helices.* Biochemistry, 1999. **38**(26): p. 8409-22.

66. Haasnoot, C.A., et al., *Structure, kinetics and thermodynamics of DNA hairpin fragments in solution.* J Biomol Struct Dyn, 1983. **1**(1): p. 115-29.

67. Haasnoot, C.A., et al., *On loopfolding in nucleic acid hairpin-type structures.* Biomolecular stereodynamics IV., 1986.

68. Blommers, M.J., et al., *Effects of base sequence on the loop folding in DNA hairpins.* Biochemistry, 1989. **28**(18): p. 7491-8.

69. Senior, M.M., R.A. Jones, and K.J. Breslauer, *Influence of loop residues on the relative stabilities of DNA hairpin structures.* Proc Natl Acad Sci U S A, 1988. **85**(17): p. 6242-6.

70. Paner, T.M., et al., *Analysis of melting transitions of the DNA hairpins formed from the oligomer sequences d[GGATAC(X)4GTATCC] (X = A, T, G, C).* Biopolymers, 1990. **29**(14): p. 1715-34.

71. Vallone, P.M., et al., *Melting studies of short DNA hairpins: influence of loop sequence and adjoining base pair identity on hairpin thermodynamic stability.* Biopolymers, 1999. **50**(4): p. 425-42.

72. Tuerk, C., et al., *CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes.* Proc Natl Acad Sci U S A, 1988. **85**(5): p. 1364-8.

73. Woese, C.R., S. Winker, and R.R. Gutell, *Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops".* Proc Natl Acad Sci U S A, 1990. **87**(21): p. 8467-71.

74. Proctor, D.J., et al., *Isolation and characterization of a family of stable RNA tetraloops with the motif YNMG that participate in tertiary interactions.* Biochemistry, 2002. **41**(40): p. 12062-75.

75. Antao, V.P., S.Y. Lai, and I. Tinoco, Jr., *A thermodynamic study of unusually stable RNA and DNA hairpins.* Nucleic Acids Res, 1991. **19**(21): p. 5901-5.

76. Hirao, I., et al., *Extraordinary stable structure of short single-stranded DNA fragments containing a specific base sequence: d(GCGAAAGC).* Nucleic Acids Res, 1989. **17**(6): p. 2223-31.

77. Yoshizawa, S., et al., *GNA trinucleotide loop sequences producing extraordinarily stable DNA minihairpins.* Biochemistry, 1997. **36**(16): p. 4761-7.

78. Davison, A. and D.R. Leach, *Two-base DNA hairpin-loop structures in vivo.* Nucleic Acids Res, 1994. **22**(21): p. 4361-3.

79. Sandusky, P., et al., *Occurrence, solution structure and stability of DNA hairpins stabilized by a GA/CG helix unit.* Nucleic Acids Res, 1995. **23**(22): p. 4717-25.

80. Lee, J.C., J.J. Cannone, and R.R. Gutell, *The lonepair triloop: a new motif in RNA structure.* J Mol Biol, 2003. **325**(1): p. 65-83.

81. Shu, Z. and P.C. Bevilacqua, *Isolation and characterization of thermodynamically stable and unstable RNA hairpins from a triloop combinatorial library.* Biochemistry, 1999. **38**(46): p. 15369-79.

82. Jucker, F.M. and A. Pardi, *Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif.* Biochemistry, 1995. **34**(44): p. 14416-27.

83. Senior, M., R.A. Jones, and K.J. Breslauer, *Influence of dangling thymidine residues on the stability and structure of two DNA duplexes.* Biochemistry, 1988. **27**(10): p. 3879-85.

84. Doktycz, M.J., et al., *Thermodynamic stability of the 5' dangling-ended DNA hairpins formed from sequences 5'-(XY)2GGATAC(T)4GTATCC-3', where X, Y = A, T, G, C.* Biopolymers, 1990. **30**(7-8): p. 829-45.

85. Quartin, R.S. and J.G. Wetmur, *Effect of ionic strength on the hybridization of oligodeoxynucleotides with reduced charge due to methylphosphonate linkages to unmodified oligodeoxynucleotides containing the complementary sequence.* Biochemistry, 1989. **28**(3): p. 1040-7.

86. Bommarito, S., N. Peyret, and J. SantaLucia, Jr., *Thermodynamic parameters for DNA sequences with dangling ends.* Nucleic Acids Res, 2000. **28**(9): p. 1929-34.

87. Ricelli, P.V., K.E. Mandell, and A.S. Benight, *Melting studies of dangling-ended DNA hairpins: effects of end length, loop sequence, and biotinylation of loop bases.* Unpublished.

88. Lowe, T.M. and S.R. Eddy, *A computational screen for methylation guide snoRNAs in yeast.* Science, 1999. **283**(5405): p. 1168-71.

89. Maden, B.E., *Mapping 2'-O-methyl groups in ribosomal RNA.* Methods, 2001. **25**(3): p. 374-82.

90. Majlessi, M., N.C. Nelson, and M.M. Becker, *Advantages of 2'-O-methyl oligoribonucleotide probes for detecting RNA targets.* Nucleic Acids Res, 1998. **26**(9): p. 2224-9.

91. Inoue, H., et al., *Sequence-specific cleavage of RNA using chimeric DNA splints and RNase H.* Nucleic Acids Symp Ser, 1988(19): p. 135-8.

92. Gamper, H.B., et al., *The DNA strand of chimeric RNA/DNA oligonucleotides can direct gene repair/conversion activity in mammalian and plant cell-free extracts.* Nucleic Acids Res, 2000. **28**(21): p. 4332-9.

93. Thayer, J.R., R.M. McCormick, and N. Avdalovic, *High-resolution nucleic acid separations by high-performance liquid chromatography.* Methods Enzymol, 1996. **271**: p. 147-74.

94. Bothwell, M.A., G.J. Howlett, and H.K. Schachman, *A sedimentation equilibrium method for determining molecular weights of proteins with a tabletop high speed air turbine centrifuge.* J Biol Chem, 1978. **253**(7): p. 2073-7.

95. Dripps, D. and R.M. Wartell, *DNA bending induced by the catabolite activator protein allows ring formation of a 144 bp DNA.* J Biomol Struct Dyn, 1987. **5**(1): p. 1-13.

96. Marky, L.A. and K.J. Breslauer, *Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves.* Biopolymers, 1987. **26**(9): p. 1601-20.

97. Wu, P., S. Nakano, and N. Sugimoto, *Temperature dependence of thermodynamic properties for DNA/DNA and RNA/DNA duplex formation.* Eur J Biochem, 2002. **269**(12): p. 2821-30.

98. Aboul-ela, F., et al., *Base-base mismatches. Thermodynamics of double helix formation for dCA3XA3G + dCT3YT3G (X, Y = A,C,G,T).* Nucleic Acids Res, 1985. **13**(13): p. 4811-24.

99. Dodgson, J.B. and R.D. Wells, *Synthesis and thermal melting behavior of oligomer-polymer complexes containing defined lengths of mismatched dA-dG and dG-dG nucleotides.* Biochemistry, 1977. **16**(11): p. 2367-74.

100. Fazakerley, G.V., et al., *Structures of mismatched base pairs in DNA and their recognition by the Escherichia coli mismatch repair system.* Embo J, 1986. **5**(13): p. 3697-703.

101. Gray, D.M., S.H. Hung, and K.H. Johnson, *Absorption and circular dichroism spectroscopy of nucleic acid duplexes and triplexes.* Methods Enzymol, 1995. **246**: p. 19-34.

102. Alberts, B., et al., *Molecular Biology of the Cell*. 3rd ed. 1994, New York, NY: Garland Publishing, Inc. 1294.

103. Mathews, D.H., et al., *Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure.* J Mol Biol, 1999. **288**(5): p. 911-40.

104. Marky, L.A., et al., *Salt-dependent conformational transitions in the self-complementary deoxydodecanucleotide d(CGCAATTCGCG): evidence for hairpin formation.* Biopolymers, 1983. **22**(4): p. 1247-57.

105. Petersheim, M. and D.H. Turner, *Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp.* Biochemistry, 1983. **22**(2): p. 256-63.

106. Riccelli, P.V., K.E. Mandell, and A.S. Benight, *Melting studies of dangling-ended DNA hairpins: effects of end length, loop sequence and biotinylation of loop bases.* Nucleic Acids Res, 2002. **30**(18): p. 4088-93.

VITA

Brooke N. Bourdélat-Parks was born August 24, 1973 in Greensboro, NC. She is the daughter of Richard and Renée Parks. Shortly after her birth, the family moved to Charlotte, NC, where she remained through her graduation from East Mecklenburg High School. From 1991 to 1995, she attended Oglethorpe University in Atlanta, GA and earned a B.S. in Biology. From January 1995 to May 1995, Brooke completed an internship in the Pediatric Endocrinology laboratory at Walter Reed Army Medical Center, Washington, DC. Following graduation from college, she worked as a biologist at the National Center for Human Genome Research, a part of the National Institutes of Health. From there, she went to work as an embryologist in the in vitro fertilization clinic at The George Washington University in Washington, DC. From 1997 to 2003, she attended Georgia Institute of Technology for graduate studies in Biology, under the supervision of Dr. Roger M. Wartell. Following graduation, she entered the Fellowships in Research and Science Teaching (FIRST) postdoctoral program at Emory University in the laboratory of Dr. David Weinshenker.