

1-20-2009

# Studies of DNA Hybridization Reactions and Applications in Genetic Assays

Fidelis Manyanga  
*Portland State University*

Follow this and additional works at: [https://pdxscholar.library.pdx.edu/open\\_access\\_etds](https://pdxscholar.library.pdx.edu/open_access_etds)



Part of the [Environmental Sciences Commons](#)

Let us know how access to this document benefits you.

---

## Recommended Citation

Manyanga, Fidelis, "Studies of DNA Hybridization Reactions and Applications in Genetic Assays" (2009). *Dissertations and Theses*. Paper 6154.

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

STUDIES OF DNA HYBRIDIZATION REACTIONS AND APPLICATIONS IN  
GENETIC ASSAYS

by

FIDELIS MANYANGA

A dissertation submitted in partial fulfillment of the  
requirements for the degree of


DOCTOR OF PHILOSOPHY  
in  
ENVIRONMENTAL SCIENCES AND RESOURCES: CHEMISTRY

Portland State University  
2009

DISSERTATION APPROVAL


The abstract and dissertation of Fidelis Manyanga for the Doctor of Philosophy in Environmental Sciences and Resources: Chemistry were presented January 20, 2009, and accepted by the dissertation committee and the doctoral program.


COMMITTEE APPROVALS:

  
Albert S. Benight, Chair ✓


  
David H. Peyton

  
Shankar B. Ranavavare

  
Andres H. La Rosa

  
Michael Bartlett  
Representative of the Office of Graduate Studies

DOCTORAL PROGRAM APPROVAL:

  
M. A. K. Khalil, Director  
Environmental Sciences and Resources  
Ph.D. Program

## ABSTRACT

An abstract of the dissertation of Fidelis Manyanga for the Doctor of Philosophy in Environmental Sciences and Resources: Chemistry presented January 20, 2009.

Title: Studies of DNA Hybridization Reactions and Applications in Genetic Assays

The intent of this study was to investigate two fundamental aspects of short DNA duplex stability and how that stability differs for duplex molecules consisting of either all perfect match Watson/Crick base pairs or a mixture of perfect match Watson/Crick base pairs and mismatch base pairs. Theoretical and experimental investigations of the origins of the nucleation term in the free energy of DNA duplex formation were revisited. Thermodynamic parameters ( $\Delta G$ ,  $\Delta H$ ,  $\Delta S$  and  $T_m$ ) of short DNA/DNA duplexes ranging in length from 6 to 35 base pairs were systematically evaluated by Differential Scanning Calorimetry (DSC) as a function of sodium ion concentration. Extrapolation of the  $\Delta G$  versus  $N$  plot to zero base pairs gave an estimate of the reference state free energy of a 'hypothetical duplex' having no hydrogen bonds, but still occupying precisely the same molar volume as the fully base paired duplex. This analysis provided a fresh evaluation of the free energy of duplex nucleation and new insights into the process of strand annealing.

The second part involved quantitative evaluation of the influence of tandem mismatch base pairs on short DNA duplex stability as a function of  $\text{Na}^+$ . For this

investigation, DSC melting data were collected for 25 short duplex DNA molecules having increasing numbers of mismatches in two different topologies, i.e. on the 'end' or 'interspersed' in the duplex among Watson/Crick base pairs. Results revealed there is a definite influence of the position of mismatches on short duplex DNA stability. This analysis provides evaluation of the perturbation factors of the free energy associated with multiple mismatches in different topologies. Results achieved in the two phases of the work were combined to provide a new approach of calculating the thermodynamic stability of short DNA duplexes containing multiple mismatches. Tests of the new method to predict the free energies of independently characterized duplexes containing mismatch base pairs demonstrate that the model can be used to predict DNA stability.

## ACKNOWLEDGEMENTS

I would like to thank my dissertation advisor, Professor Albert S. Benight for his support, guidance, enthusiasm and financial support throughout this research project. The experience of conducting an independent biophysical research project in his laboratory was a remarkable experience. I also wish to give my sincere thanks and gratitude to the other members of my dissertation committee: Professors, David H. Peyton, Shankar B. Ranavavare, Michael Bartlett and Andres H. La Rosa for their help, guidance and encouragement. I would also like to acknowledge the assistance, and interesting conversations from the following past and present members of the Professor Benight's research group: Dr. Todd M. Horne, Dr. Greg P. Brewood, Dr. Daniel J. Fish, Dr. Jim Goodarzi, Rebekah Dickman, Saba Alemayehu, Jevin Cutler, Clark Benight and Ivona Ristovska. Most importantly, my deepest gratitude and appreciation goes to my family, whose unconditional love, encouragement and support is given without asking. It is difficult to overstate my gratitude to my parents, Francis and Benonia Manyanga, my wife Bridgeter and daughter, Kimberley, you were my source of inspiration and this is dedicated to you. Special mention goes to my brother Francis and his family, for providing me with educational support throughout my academic career. I will always feel obligated to you. Lastly, i would like to thank all my friends, who stood by my side, asking over and over again "When will you get it done?"

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	(i)
LIST OF TABLES.....	(vii)
LIST OF FIGURES.....	(ix)
LIST OF ABBREVIATIONS AND SYMBOLS.....	(xii)
PREFACE.....	(xv)
CHAPTER 1: INTRODUCTION .....	1
1.0 Background and Motivation.....	1
1.1 The nucleation Free Energy.....	6
1.2 Tandem mismatch base pairs.....	10
CHAPTER 2: MATERIALS, EXPERIMENTAL AND THEORETICAL	
METHODS.....	14
2.0 Reagents, Chemicals and Supplies.....	14
2.1 Preparation of Buffer Solutions .....	14
2.2 Characterization of DNA Samples.....	16
2.2.1 Sample Receipt and Specifications.....	16
2.2.2 General handling of single stranded DNA oligonucleotides .....	17
2.2.3 UV/Vis Absorption Spectroscopy.....	19

2.2.4	Spectral Analysis of DNA on UV/VIS Spectrophotometer.....	20
2.2.5	Preparation of duplex DNA molecules.....	21
2.3	Polyacrylamide Gel Electrophoresis, PAGE.....	22
2.3.1	Principle of PAGE.....	22
2.3.2	Reagents and Materials.....	23
2.3.3	Preparation of PAGE.....	24
2.3.4	Sample Preparation.....	24
2.3.5	Running the gel.....	25
2.3.6	Gel Staining Procedure.....	27
2.3.7	Destaining and Visualization.....	27
2.4	DNA concentration, Desalting and Recovery.....	28
2.5	Differential Scanning Calorimeter, DSC.....	31
2.5.1	Principle.....	31
2.5.2	DSC Melting Experiments.....	35
2.5.3	Sample Preparation.....	36
2.5.4	Loading the buffer and running a baseline.....	36
2.5.5	Sample loading and experimental parameters.....	37
2.5.6	Data Analysis.....	38
2.5.7	Analysis of DSC melting data: Results Acquisition.....	42
2.5.8	Sample Storage after melting experiments.....	42
2.6	THEORETICAL METHODS.....	44



2.6.1	Standard State Equilibrium Constant for Short Duplex Hybridization: Schurr's Formalism .....	44
2.6.2	Release of sodium from duplex upon melting .....	53
CHAPTER 3: ORIGINS OF THE 'NUCLEATION' FREE ENERGY TERM IN PREDICTIONS OF THE THERMODYNAMICS OF SHORT DUPLEX DNA HYBRIDIZATION.....		
		57
3.0	INTRODUCTION .....	57
3.1	EXPERIMENTAL AND THEORETICAL METHODS.....	59
3.1.1	DNA Molecules .....	59
3.1.2	Preparation of Buffer Solutions .....	60
3.1.3	Preparation of DNA molecules.....	61
3.1.4	Analytical Polyacrylamide Gel Electrophoresis.....	62
3.1.5	DSC Melting Experiments, Data Analysis .....	62
3.2	RESULTS.....	63
3.2.1	Analysis of DSC melting curves.....	63
3.2.2	DSC Melting Data Error Propagation.....	66
3.2.3	Free-Energy versus Duplex Length, N .....	71
3.2.4	Validity of the n-n model.....	75
3.2.5	Analysis of melting temperature, $T_m$ .....	78
3.2.6	Dependence of the melting transitions on $[Na^+]$ and evaluation of the counterion release upon DNA melting, $\Delta n$ .....	82
3.3	DISCUSSION.....	87

3.3.1	Double Helix Formation .....	88
CHAPTER 4: MELTING STUDIES OF TANDEM MISMATCH BASE PAIRS		
	IN SHORT DUPLEX DNA: INFLUENCE OF MISMATCH SIZE, LOCATION, AND SODIUM ION CONCENTRATION.....	92
4.0	INTRODUCTION .....	92
4.1	MATERIALS AND METHODS.....	94
4.1.1	Rationale and Design of DNA Molecules .....	94
4.1.2	DNA Molecules .....	99
4.1.3	DNA Recovery .....	99
4.1.4	Analytical Gel Electrophoresis .....	100
4.2	THEORETICAL METHODS.....	100
4.2.1	Single base pair mismatches .....	100
4.2.2	Tandem Mismatches .....	102
4.3	RESULTS AND DISCUSSION.....	104
4.3.1	Analysis of Melting Curves .....	104
4.3.2	Evaluation of Thermodynamic Parameters.....	105
4.3.3	Entropy-Enthalpy Compensation.....	110
4.3.4	Is the n-n model valid for interpreting short DNA Duplexes containing tandem mismatches?.....	113
4.3.5	[Na <sup>+</sup> ] dependence of short DNA containing tandem mismatches .....	116
4.3.6	Analysis of Melting temperature .....	117

4.3.7	How does the stability of short duplex DNA containing contiguous tandem mismatches vary with $[Na^+]$ ?	123
4.3.8	Free Energy versus Duplex length ,N for duplexes with mismatch base pairs	127
4.3.9	The relationship between perfect match duplexes of different sizes and duplexes containing increasing amounts of tandem mismatches	132
4.3.10	Applications to probe sequence design	141
CHAPTER 5: SUMMARY		149
6.0	REFERENCE LIST	155
7.0	APPENDIX: Plots of Measured $\Delta H^{cal}$ and $\Delta S^{cal}$ versus N for DNA molecules containing tandem mismatches in four buffered $[Na^+]$	165

## LIST OF TABLES

Table 2.1: Standard buffer conductivities values for sodium ions.....	16
Table 2.2(a): Detailed list of Materials, Reagents and recipes used for buffer preparations.....	26
Table 2.2(b): Typical recipe used for preparation of 12 % PAGE resolving gel and staining procedure .....	28
Table 2.4: Specifications for DSC melting experiments. ....	41
Table 3.1: Perfectly matched DNA sequences ranging in length from 6 - 35 base pairs.....	61
Table 3.2: Experimentally measured DSC thermodynamic data for DNA duplexes with 6 – 35 base pairs in all $[\text{Na}^+]$ .....	65
Table 3.3: Mean thermodynamic transition parameters per base pair for duplexes in all $[\text{Na}^+]$ concentrations. ....	71
Table 3.4: Summary of the linear fit parameters for the plots of $\Delta G_{25^\circ\text{C}}^{\text{cal}}$ (kcal/mol) versus N as a function of $[\text{Na}^+]$ . ....	74
Table 3.5: Release of $[\text{Na}^+]$ upon DNA melting, $\Delta n$ and differential cation release per phosphate group, $\Delta\psi$ for 19 DNA molecules. ....	83
Table 3.6: Evaluated values for the helix initiation free energy DNA and RNA duplexes, obtained from different laboratories.. ....	91
Table 4.1(a): The 20 base pair DNA molecules comprising Sets A .....	96

Table 4.1(b): The 20 base pair DNA molecules comprising Sets B.....	97
Table 4.1(c): Control Duplexes.....	98
Table 4.2(a): Measured Thermodynamic Parameters for Set A and Set B duplexes in buffered 85 – 1000 mM [Na <sup>+</sup> ]. .....	107
Table 4.2(b): Measured thermodynamic data for control molecules used in this study.....	109
Table 4.3(a). Data tables for calculated values of $\Delta n$ and $\Delta\psi$ for Set A duplexes in 85 to 1000 mM [Na <sup>+</sup> ]. .....	119
Table 4.3(b). Data tables for calculated values of $\Delta n$ and $\Delta\psi$ for Set B duplexes in 85 to 1000 mM [Na <sup>+</sup> ]. .....	120
Table 4.4: Parameters obtained for linear plots of Figure 4.9. Slopes, intercepts and linear regression coefficients are listed.....	131
Table 4.5: Summary of the linear fit parameters for the plots of $\partial\Delta G$ kcal/mol versus $n$ (mismatched base pairs) as a function of [Na <sup>+</sup> ]. .....	139
Table 4.6: DNA Duplexes analyzed from Reference (15) .....	147
Table 4.7: Comparison of experimentally Measured, MFOLD predicted and modified $\Delta G_{25^{\circ}C}^{cal}$ kcal/mol for duplexes presented in Table 4.6.....	148

## LIST OF FIGURES

Figure 1.1: Schematic illustration of the nucleation and propagation steps in DNA double helix formation.....	8
Figure 2.2: Schematic Diagram of the Centrifugal Filter Devices used in this study.....	30
Figure 2.3: Schematic diagram showing essential features of a DSC..	33
Figure 2.4: Representative Calorimetric melting curves.....	43
Figure 3.2: Representative DSC melting curves plotted on the same scale.....	64
Figure 3.3: Correlation between $\Delta H^{cal}$ (kcal/mol) and $T\Delta S^{cal}$ (kcal/mol) ( $T = 298.15$ K) for DNA duplexes with 6 to 35 base pairs in four $\text{Na}^+$ environments. ....	69
Figure 3.4: Plots of Measured $\Delta G_{25^\circ C}^{cal}$ (kcal/mol) versus N for DNA molecules ranging in size from 6 to 36 base pairs in four buffered $[\text{Na}^+]$ environments..	73
Figure 3.5: Plot of free-energies at the intercepts obtained from the unconstrained fit .....	75
Figure 3.6: Comparative plots of the experimental enthalpy and entropy difference from MFOLD™.....	76
Figure 3.7: Comparative plots of calorimetric $\Delta\Delta G^{cal}$ ( $\Delta G_{25^\circ C}^{cal} - \Delta G^{MFOLD}$ ) (kcal/mol) versus N in all $[\text{Na}^+]$ environments.....	78
Figure 3.8: Plot of $T_m$ versus N for the 19 DNA duplexes of this study. ....	79

Figure 3.9: Plot of the direct comparison between predicted and measured melting temperature i.e. $\Delta T_m = T_m^{\text{DSC}} - T_m^{\text{MFOLD}}$ .....	81
Figure 3.10: Histogram for $\Delta n$ versus N DNA duplexes ranging in length from 6-35 base pairs.....	84
Figure 3.11: Plot of the average values of $\Delta n$ versus N DNA duplexes ranging in length from 6 - 35 base pairs in all four $[\text{Na}^+]$ environments. ....	85
Figure 3.12: Histogram for $\Delta\psi$ versus N for the DNA duplexes ranging in length from 6-35 base pairs.....	86
Figure 3.13: Plot of the average values of $\Delta\psi$ versus N DNA duplexes ranging in length from 6 - 35 base pairs in all four $[\text{Na}^+]$ environments. ....	86
Figure 4.1: Representative DSC Melting Curves evaluated in 600 mM $[\text{Na}^+]$ .. ....	106
Figure 4.2(a): Plot of $\Delta H^{\text{cal}}$ (kcal/mol) versus $T\Delta S^{\text{cal}}$ (kcal/mol) at 298.15K in all $[\text{Na}^+]$ environments for Set A duplexes.....	111
Figure 4.2(b): Plot of $\Delta H^{\text{cal}}$ (kcal/mol) versus $T\Delta S^{\text{cal}}$ (kcal/mol) at 298.15K in all $[\text{Na}^+]$ environments for Set B duplexes. ....	112
Figure 4.3: Plot of $\Delta\Delta G_{25}^{\circ\text{C}}$ (kcal/mol) versus n for Sets A and B duplexes.....	114
Figure 4.4: Plot of $\Delta\Delta T^{\circ\text{C}}$ ( $T^{\circ\text{C}}(\text{measured}) - T^{\circ\text{C}}(\text{predicted})$ ) versus n for duplexes of Sets A and B duplexes.....	115
Figure 4.5: Histograms of empirically evaluated values for $\Delta n$ versus sample ID for both Set A and Set B duplexes.....	121
Figure 4.6: Histograms of empirically evaluated values for $\Delta\psi$ versus number of mismatches for Set A and Set B duplexes. ....	122

Figure 4.7: Plot of average values of $\Delta n$ versus N for perfect match (PM), Set A and Set B duplexes.....	125
Figure 4.8: Plot of average values of $\Delta\psi$ versus N for perfect match (PM), Set A and Set B duplexes.....	126
Figure 4.9: Plot of N versus $\Delta G_{25^\circ C}^{cal}$ (kcal/mol) for perfect match (PM), Set A and Set B DNA duplexes.....	130
Figure 4.10: Plot of N versus $\Delta G_{25^\circ C}^{cal}$ (kcal/mol) for perfect match (PM), Set A and Set B DNA duplexes for $0 < N < 20$ base pairs in all $[Na^+]$ environments labeled (a)–(d) .....	136
Figure 4.11: Plot of n versus $\partial\Delta G$ (kcal/mol) for Set A and Set B DNA duplexes versus number of mismatches n ( $0 < n < 10$ ) in all $[Na^+]$ environments.....	138
Figure 4.12: Histogram of N versus $\Delta G^0$ (kcal/mol) for Perfect match (PM), Set A and Set B DNA duplexes respectively.....	140



## LIST OF ABBREVIATIONS AND SYMBOLS

A, T,G, C	Adenine, Guanine, Thymine and Cytosine
$\Delta C_p^{\text{ex}}$	Change in excess heat capacity
DNA	Deoxyribonucleic Acid
DSC	Differential Scanning Calorimetry
EDTA	Ethylendiaminetetraacetic Acid
$\Delta G_{25^\circ\text{C}}^{\text{cal}}$	Calorimetric Free Energy at 25°C
$\Delta\Delta G$	Difference in Free Energy
$\Delta G_{\text{nuc}}$	Nucleation or initiation free energy
$\Delta G_{25^\circ\text{C}}^{\text{cal}} (N = 0)$	Free-energy of the 'hypothetical duplex' having zero base pairs
$\Delta G^{\text{cal}} / \text{bp}$	Free energy per base pair
$\Delta H^{\text{cal}}$	Calorimetric Enthalpy
$\Delta H^{\text{cal}} / \text{bp}$	Standard Calorimetric Enthalpy per base pair
$\Delta H_{\text{ion}}$	Enthalpy of ionization
$K^\circ$	Equilibrium Dissociation Constant
$K'$	Equilibrium Dissociation Constant for the nucleated complex
MWCO	Molecular Weight Cut Off

$Mwt$	Molecular Weight
$[Na^+]$	Total sodium ion concentration
OD	Optical Density at 260 nanometers
PAGE	Polyacrylamide gel electrophoresis
$\Delta S^{cal}$	Calorimetric Entropy
$\Delta S^{cal} / bp$	Standard Calorimetric Entropy per base pair
$T_m$	Melting Temperature
$\Delta T_m$	Change in Transition Melting Temperature
TEMED	(N,N,N',N',) Tetramethylethylenediamine
UV-VIS	Ultraviolet Visible Spectrophotometry
5'X-Y-3'	Orientation of the top strand NN doublet
% G-C	Molar percentage ratio of guanine and cytosine bases
N	Number of perfectly matched Watson Crick bases
$N$	Total number of particles under consideration
n	Number of mismatched bases
IDT	Integrated DNA Technologies
APS	Ammonium Persulfate
Tris	2-Amino-2-2 (hydroxymethyl)-1,3-propanediol
TBE	Tris, Boric Acid, EDTA buffer
$s^N$	Statistical Weight
$N_A$	Avogadro's number
$\kappa$	Conductivity

$S1$	DNA Single Strand
$S2$	Complementary DNA Single Strand
$D$	DNA duplex Strand ( $S1 + S2$ )
FW	Formula Weight
v/v	Volume to volume
w/v	Weight to volume
mS/cm	MilliSiemens per centimeter
$\Delta n$	Number of sodium ions released/absorbed
$\Delta \psi$	Number of sodium ions released/absorbed per phosphate
~	Approximately
$\equiv$	Equivalent to
$\lambda$	Wavelength
$\mu\text{L}$	Microlitres
kcal	Kilocalorie
e.u.	Entropy unit (cal/deg-mol)
R	Universal gas constant (1.987 cal/deg-mol)

## PREFACE

Work summarized in this dissertation is presented in five chapters. In Chapter 1, an overview of the overall goals and historical perspective leading to this project are presented. Chapter 2 contains a detailed description of the materials, experimental and theoretical methods used in this study. Simplified yet detailed expressions of the modified Schurr statistical mechanical formalism are presented. Experimental results are presented separately in Chapters 3 and 4 as follows. (i) In Chapter 3, the origins of the “nucleation” free energy in the hybridization thermodynamics of short duplex DNA is revisited and explored. A new ‘reference state’ is proposed for short duplex DNA that differs fundamentally from the current interpretations that involve the formation of at least one base pair. Conclusions are made based on thermodynamic studies of the melting transitions for 19 perfectly matched short DNA duplexes. (ii) In Chapter 4, results of melting studies for 25 short duplex DNA duplexes containing tandem mismatch base pairs are presented. Finally a reconciliation of the two projects is provided in Chapter 5, followed by a general summary consisting potential applications, major conclusions and future directions.

## CHAPTER 1

### INTRODUCTION

#### 1.0 Background and Motivation

Applications of nucleic acid-based diagnostics have evolved significantly over the past few decades. This evolution has been fueled by advances in biophysics, providing a better understanding of the thermodynamic behaviors as a function of sequence derived from well conceived studies focused on specific nucleic acid duplex oligomers. In particular, DNA oligonucleotides have been widely used in many applications, ranging from diagnostics to therapeutics (1-3,7-11,15,32). It can be convincingly argued that the cornerstone of modern molecular diagnostic approaches to gene analysis and genotyping is the sequence-dependent hybridization or annealing of two DNA or RNA single strands to form a duplex. Emerging biotechnological applications of DNA and RNA hybridization based technologies (e.g. DNA microarrays, polymerase chain reaction (PCR), RNA interference and molecular beacon assays) have heightened the need to finely tune, optimize and standardize the underlying chemical reactions. Design and construction of novel DNA oligomers with tunable structural and hybridization properties have recently been topics of intense research (2,5,10,15,32).

The concept of nucleic acid hybridization in terms of the annealing of two strands seems to be a straightforward concept whereby a specific DNA/RNA sequence will bind to its complementary sequence by virtue of Watson/Crick (W/C) base pairing (1-2). However, in a multiplex environment, where many different strands and their complements are present, the situation is much more complicated (2,15,34). Recently, there have been several studies that take a complex systems approach to DNA hybridization (1-2,15,34,46,72,92). Multiplex assays (1-2,5,15,34,46) requiring the simultaneous and high fidelity binding of many surface-linked probe molecules to a large number of potential target molecules, are emerging as attractive high throughput tools for genetic analysis. In a multiplex situation, the probability of mismatch duplex formation can be quite large (1-2,100), depending on the actual concentrations and types of DNA strands present, and their sequences (2,46). A number of fundamental features of the multiplex chemical reactions and specific sequences of molecules and their interactions involved are not well understood. Unwanted cross-hybridization between strands that contain mixtures of mismatch and perfect match base pairs represents a major nemesis in hybridization technology (2,32,100). Such cross-hybridization can lead to increased noise, lower sensitivity and even false positives in DNA assays (2,15,79,98). Ultimately, performance and outcome of a multiplex hybridization depends on reliable predictions of the melting stability of short DNA duplexes. The focus of this thesis work is on explicit characterization of fundamental

components of the helix-coil transition in short duplex DNAs, with the primary aim of improving and optimizing nucleic acid-based molecular diagnostic assays.

The phenomenon of the helix-coil or melting transition in DNA has been the subject of extensive theoretical and experimental investigations for many years (6,11-14,18-21,26). The most successful approaches to modeling the helix-coil transition in DNA have foundations in the statistical mechanical thermodynamic formalism of the Ising model (6,26,33). To date, the most widely applied and reliable thermodynamic calculations underlying the prediction of short DNA stability are based on the nearest neighbor (n-n) model (3-14,19,26,97,105). The n-n model pioneered by Zimm and Bragg (14) and further developed by Crothers *et al.* (12) Tinoco *et al.* (12-13), Wartell *et al.* (26) and Flory *et al.* (13,19) has been extensively employed to describe the helix-coil transition in DNA. If accurately parameterized, the n-n model can provide quite accurate predictions of the sequence dependent thermodynamic stability of short duplex DNAs (8-11,16-17). It is based on the assumption that the stability of a given base pair depends on the identity and orientation of n-n base pairs on either side. According to the n-n model, the free-energy change of helix melting can be partitioned into three terms. These are: (i) the free energy change for helix initiation or nucleation; (ii) the free energy change for helix propagation, which is the sum of the appropriate n-n parameters for the particular n-n doublets present in the duplex; and (iii) the free energy change associated with the reduced entropy when the duplex is composed of two self-complementary single strands. The stability of a DNA double helix is

estimated as the weighted sum of the stabilities of the combination of the appropriate 10 unique W/C n-n base pair doublets that its sequence contains (i.e. AA/TT, AT/TA, TA/AT, CA/GT, GT/CA, CT/GA, GA/CT, CG/GC, GC/CG and GG/CC). Here, the slash, /, separates strands in anti-parallel orientation, meaning that AA/TT represents the sequence 5'-AA-3' paired with complementary sequence 3'-TT-5'. Thus the total standard free energy of duplex melting is given by the following equation (8-9,11),

$$\Delta G_{duplex}^o = \Delta G_{initiation}^o + \sum_i n_i \Delta G_{i,i+1}^o + \Delta G_{symmetry}^o \quad (1.1)$$

where  $\Delta G_i^o$  is the free energy change for the 10 possible W/C base pairs,  $n_i$  is the number of occurrences of each n-n,  $i$ , and  $\Delta G_{initiation}^o$  is termed the initiation or nucleation free-energy for the duplex formation, which is considered unfavorable.  $\Delta G_{initiation}^o$  has been assigned to account for the relative difficulty of forming the first W/C base pair and “nucleating” the duplex versus formation subsequently of the remaining base pairs. It includes all other sequence independent effects, including those associated with counterions (8). The term  $\Delta G_{symmetry}^o$  is a symmetry correction free energy that accounts for the degeneracy in self-complementary versus non-self complementary sequences. In order to account for stability differences between DNA duplexes with terminal A-T versus G-C base pairs, two initiation parameters have been introduced. In addition, melting



temperature,  $T_m$  and the relative thermodynamic stability of the DNA duplex depend on the salt concentration, pH of the solvent and strand concentrations.

The thermodynamics of nucleic acids, based on n-n model have been studied by many research laboratories (2-17,26,31,60,63-64,66,74-75,77,97). However, even after over 40 years of work, there are yet concerns with this model and some problems still remain. Many fundamental questions remain unanswered when considering the n-n model to predict DNA duplex stability, because the chemistry is still not well characterized. In particular, even though the common theme is the n-n model, different laboratories use different oligonucleotide design strategies (6,10,77), different polymers (6-8,10-11), different techniques for data analysis and different interpretations of the salt dependence to short DNA hybridization (8,10,56,77). Further, there still remains some controversy about the 'length dependency' on the salt concentration (10) and to the best of our knowledge, no length dependence parameter currently exists for the n-n parameters. For these reasons, differences in thermodynamic trends for DNA oligonucleotides have been observed. There is need to refine and improve this model of DNA melting. In fact, the current state of n-n predictive models for DNA/RNA hybridization is such that researchers can obtain different thermodynamic results, depending on the routine applied (31,63-64). For nucleic acid hybridization reactions to be quantitative and reproducible, the major drawbacks of the n-n predictive algorithms need to be addressed. In order to predict the outcome of multiplex hybridization reactions accurately, there is need for n-n to be

continuously augmented and parameterized by experimental models. This thesis work is aimed at improving our quantitative understanding and optimization of the DNA multiplex assay design process. In particular, the concept of the initiation free energy or nucleated complex, denoted  $\Delta G_{nuc}$  for short DNA hybridization is revisited. Results suggest a new perspective for characterizing in more detail thermodynamic descriptions of the nucleated complex during the DNA melting process. Both theoretical and empirical treatments that describe the nucleation parameter during the helix-coil transition are developed. To do that, two specific aspects of the n-n parameters are revisited and examined in detail as summarized below.

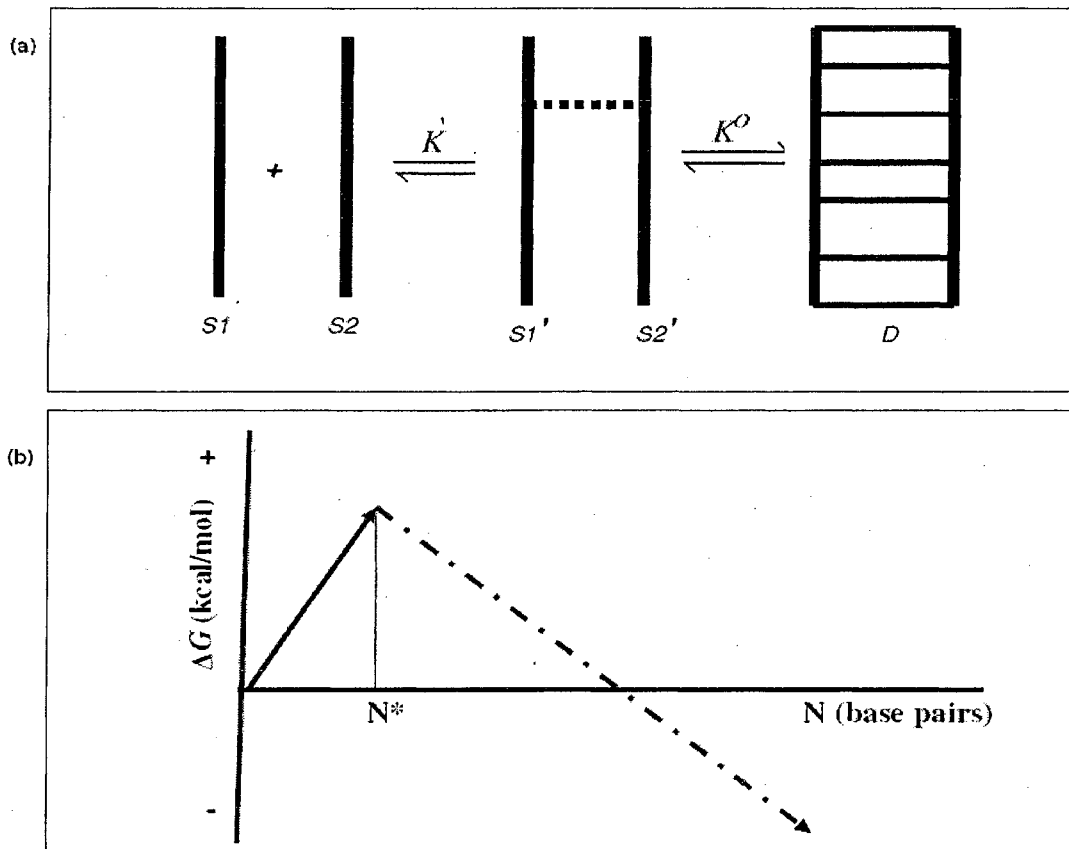
## **1.1 The Nucleation Free Energy**

The primary intent of this project is to improve the concept of duplex nucleation in short DNA hybridization. This is approached through melting studies of short duplex DNAs as a function of length and sodium ion concentration,  $\text{Na}^+$ . The concept of a nucleated or initiation complex in the process of duplex formation is not new (10,18-22,43,47,66-67,91). As early as 1963, the duplex nucleation in the formation of short DNA duplexes was introduced by Applequist and Damle (18). Their theoretical analysis suggested that DNA double helix formation occurs in two steps; namely the helix initiation followed by a growth or propagation. It is now well established that DNA double helix formation comprises of the following steps (10,12,20,47); (i) bringing the single strands into close proximity (largely

determined by the strand concentration, salt concentration and solvent viscosity); (ii) formation of the 'nucleated' complex, which begins with helix initiation followed by critical nucleus formation; and lastly (iii) quick 'zippering' of the helix also known as the propagation step (18-22). In 1971 Craig *et al.* (21) showed that only after formation of the third base pair does the nucleus become stable enough to allow addition of additional base pairs with a significantly higher probability of association than dissociation. In addition, quantitative studies of the kinetics of DNA duplex formation performed by Wetmur and Davidson (20) and Craig *et al.* (21), concluded that the first step is the unfavorable formation of a nucleation complex through bimolecular reaction of two individual single strands, followed by the rapid propagation step. In their treatment of helix-coil equilibrium, both strands were assumed to be the same and the equilibrium constant for duplex formation was given by  $K_{eq} = \beta s^N$ , where  $\beta$  is the nucleation parameter and  $s^N$  is the statistical weight of the fully intact base paired duplex containing N base pairs. The quantities are related to the standard state free-energy, i.e.  $\Delta G^\circ = -k_B T \ln s^N$ , where  $s = \exp(-\Delta G^\circ / bp / k_B T)$  and  $\Delta G^\circ / bp$  is the free-energy per base pair, which in their case was assumed to be the same for every base pair in the duplex.

The nucleation parameter is an essential component of models of the helix-coil transition (4-5,8-11,22,43,47). A simplified schematic diagram illustrating the presumed mechanism for the association of two single strands,  $S1$  and  $S2$ , to form a duplex,  $D$ , is shown in Figure 1.1(a). A bimolecular equilibrium annealing reaction between two single strands, designated  $S1$  and  $S2$  results in the formation of a

nucleated complex ( $S1' - S2'$ ) followed by the duplex complex formation,  $D$ . The quantity  $K'$  represents the equilibrium dissociation constant for the nucleated complex and  $K^o$  standard equilibrium constant for the overall reaction. As shown in Figure 1.1(b), nucleation free energy accounts for the energetically unfavorable interactions between  $S1$  and  $S2$  that must be overcome before duplex formation occurs.



**Figure 1.1:** Schematic illustration of the nucleation and propagation steps in DNA double helix formation. (a) A bimolecular equilibrium annealing reaction between two single strands, designated  $S1$  and  $S2$  results in the formation of a nucleated complex ( $S1'-S2'$ ) followed by the duplex formation,  $D$ . (b) Nucleation steps start with a slow and energetically unfavorable helix initiation, followed by the critical nucleus formation. A threshold of base pairs ( $N^*$ ) must be nucleated in order for helix 'zippering' to occur in a spontaneous propagation step.

Slightly different n-n values for nucleation parameters have been reported and employed by several authors (8-9,22,62,73). The studies by Sugimoto *et al.* (62) and Breslauer *et al.* (73) assigned helix free energy values,  $\Delta G_{nuc}^o_{37^\circ C}$  of 2.79 and 6.23 kcal/mol, respectively to DNA duplexes containing A-T and G-C base pairs, respectively. In more recent studies, Sugimoto *et al.* (22) determined the free energy of the initiation factor from analysis of melting curves of 50 DNA/DNA duplexes and found the  $\Delta G_{nuc}^o_{37^\circ C}$  to be 3.40 kcal/mol. This was essentially the same value obtained for RNA/RNA duplexes reported by Freier *et al.* (62) and only slightly different from the value of 3.10 kcal/mol obtained for the RNA/DNA duplexes (74). A number of probe design routines based on the n-n model are available (31,63,64,73). There have been reports of clear discrepancies between some predicted n-n calculations and experimentally measured thermodynamic parameters for some short DNA containing both perfect match and mismatch base pairs (2-3,10,23). The current interpretation of precisely what comprises  $\Delta G_{nuc}$  is not clear. Currently, values of the nucleation free energies reported in the literature all include formation of at least one intact base pair (A-T or G-C), and presumably comprise different enthalpic and entropic contributions, depending on the identity of the terminal base pair.

In this study, thermodynamic parameters: enthalpy,  $\Delta H^{cal}$ , entropy,  $\Delta S^{cal}$ , and free energy at 25°C,  $\Delta G_{25^\circ C}^{cal}$ , as well as melting temperature,  $T_m$  of the

melting transitions for 19 short DNA/DNA duplexes ranging in length from 6 to 35 base pairs were systematically evaluated by DSC melting experiments carried out at four salt concentrations from 85 to 1000 mM total sodium ion,  $[\text{Na}^+]$ . In essence, this analysis suggests an initiation complex, or “reference” state, comprises two strands existing in the same volume as in the duplex but without any of the favorable interactions that stabilize the duplex, i.e. hydrogen bonding and stacking. The standard states and statistical thermodynamic formalism proposed by Schurr (40) and Gilson *et al.* (42) for the association of two strands to form a duplex complex accounting for displaced solvent, was utilized to relate  $\text{Na}^+$  dependent free energy values to configuration integrals for the complex. The analysis presented here provides a new vantage point to view what has historically been referred to as the helix initiation or nucleation parameter and provides an alternate interpretation and mechanism for the nucleation complex in duplex formation.

## **1.2 Tandem mismatch base pairs**

In another related study, thermodynamic effects of well defined tandem mismatches (two or more contiguous non-W/C base pairs) as a function of relative location on DNA duplex stability were investigated. As mentioned earlier, published thermodynamic parameters for perfect match base pairs, based on the n-n model are available (8-11,31,60,63,66,70-77). Significant milestones have been achieved in evaluating thermodynamic parameters for single base pair mismatches (3-4,7,10,23-28), although the underlying structural chemistry is still not well understood. Different approaches are used to treat tandem mismatches comprised

of multiple or contiguous mismatch base pairs. For example in the current approach, calculations of the thermodynamic stabilities of some short DNA oligomers containing mismatches ignore the potential sequence-dependent thermodynamic contributions of tandem mismatches. The standard method assumes that tandem mismatches make no favorable contribution to thermodynamic stability of the duplex (10). This assumption means that the n-n stacking interactions are considered to be fully disrupted by tandem mismatches, and if bounded by W/C base pairs the mismatched region is considered to be a disordered loop. In thermodynamic calculations, effects of these loops on duplex stability are accounted for solely by a loop energy term,  $\Delta G^{\circ}_{loop}(n)$ , which depends on the number, n, of contiguous mismatched bases comprising the loop (herein abbreviated  $\Delta G^{\circ}_{loop}(n)$ ). Further, since mismatches on the ends are not assigned such loop energies, the n-n model calculations may differ for two duplexes, both containing same tandem mismatches except some mismatch base pairs are on the 'ends' versus in the 'interspersed' in the duplex.

The quantitative range of stabilities of duplexes containing tandem mismatches of different sizes in different orientations have not been investigated. In its current rendition, the apparent deficiency of the predictive power of the n-n model underscores the need to obtain a more quantitative understanding of the thermodynamic contributions of tandem mismatches to duplex stability. In this thesis, the experimental repertoire of the sequence-dependent thermodynamic melting parameters for short DNA containing tandem mismatches was examined

via DSC. Results of the studies show that stability of short DNA duplexes containing mismatches depends strongly on whether tandem mismatches are contiguous on the 'end' or 'interspersed' throughout the duplex.

In large part, mostly by convention, dependence of short nucleic acid hybridization has been traditionally studied in 1.0 M sodium chloride/phosphate buffer (3-11,70-77). In the final component, the stabilities of the DNAs were also examined as a function of buffered sodium ions,  $\text{Na}^+$ .

In this thesis, two fundamental questions addressed for a well defined set of well resolved DNA molecules were as follows: (i) What is the value of the nucleation or initiation free energy of short duplex DNA as a function of sodium ion,  $\text{Na}^+$ ? (ii) How does the presence and distribution of mismatches in a short duplex affect duplex stability?

The primary experimental technique employed throughout this study is DSC. Generally, thermodynamic parameters for DNA hybridization are evaluated using either of two standard techniques: (i) Ultraviolet-visible (UV/VIS) spectrophotometry (optical absorbance measurements at 260 nm measured as a function of temperature); and (ii) DSC which provides a direct measurement of the excess heat capacity,  $\Delta C_p^{\text{ex}}$ , versus temperature (111). Thermodynamic parameters evaluated by DSC offer a number of advantages over those obtained via optical absorbance measurements. Analyses of,  $\Delta C_p^{\text{ex}}$ , versus temperature curves using DSC provide a direct evaluation of the melting thermodynamics. Evaluation of thermodynamic parameters from UV-absorbance melting curves inherently



assumes a two-state transition. For this reason, conventional UV/VIS spectrophotometry probably lacks sufficient sensitivity to detect subtle influences of the melting buffer (94,97,103-104). In contrast to UV spectrophotometry, DSC provides a direct measurement of the melting thermodynamics without invoking a model of the transition, (i.e. model-independent parameter evaluation).

## CHAPTER 2

### MATERIALS, EXPERIMENTAL AND THEORETICAL METHODS

#### 2.0 Reagents, Chemicals and Supplies

All reagents used to prepare buffers were purchased from Sigma Chemical Company (Sigma-Aldrich, St. Louis, MO) and were all molecular biology grade. Glassware, pipettes and other chemicals were obtained from Fisher Scientific Company (Fisher, Pittsburg, PA), Nalge Nunc International (Nalgene, Rochester, NY) and VWR Scientific Corporation (VWR, West Chester, PA). A Nanopure 11 ultra-pure water purification system (Barnstead Corporation, Boston, MA), having a 0.220  $\mu\text{m}$  filter that produced water with a resistivity of 18.0 Mega ohm centimeter ( $\text{M}\Omega\text{-cm}$ ) was used in all experiments. All solvents, pipettes, microcentrifuge tubes and glassware were autoclaved for twenty minutes at 121  $^{\circ}\text{C}$  and 1.41 kg/cm pressure. All chemicals and reagents were used as purchased without further purification, unless otherwise specified.

#### 2.1 Preparation of Buffer Solutions

The buffer used in all melting experiments was sodium phosphate,  $\text{Na}_3\text{PO}_4$  combined with varying amounts of sodium chloride,  $\text{NaCl}$ . This buffer was chosen largely because during the course of a melting experiment, changes in  $\text{pK}_a$  with

temperature are known to be small (30). In addition, the enthalpy of ionization,  $\Delta H_{ion}$  and heat capacity changes for buffer ionization are small, and the pH is relatively constant as temperature is varied (12,30,44). Buffers used in all melting studies contained 10 mM  $\text{Na}_2\text{HPO}_4$ , 0.1 mM  $\text{Na}_2\text{EDTA}$  and varying amounts of NaCl, pH adjusted to 7.40 - 7.50 with 1.0 M NaOH, that resulted in final total sodium ion concentration,  $[\text{Na}^+]$  ranging from 85 to 1000 mM. Disodium ethylenediaminetetraacetic acid ( $\text{Na}_2\text{EDTA}$ ), was added to chelate trace metals and multivalent cations that could bind to the DNA and possibly cause influence the melting process. The  $[\text{Na}^+]$  of melting buffers were independently verified by taking electrical conductivity ( $\kappa$ ) measurements, made using an Orion 4 Star pH/Conductivity Meter (Thermo Electron Corporation, Beverly, MA) and standard values are as summarized in Table 2.1. Generally, conductivities change rapidly with temperature and for this reason, the conductivity data was fitted to a standard calibration curve (data not shown). Measured conductivities were in good agreement with published conductivities for  $\text{Na}_2\text{HPO}_4$  (30). Good reproducibility was obtained and melting buffers with conductivity differences greater than 6 % were discarded.

**Table 2.1:** Standard buffer conductivities values for sodium ions.

<b>[Na<sup>+</sup>] mM</b>	<b>Conductivity <math>\kappa</math> [mS/cm]</b>
85	10.6 ± 0.92
300	33.1 ± 1.3
600	61.0 ± 2.7
1000	88.3 ± 4.1

pH = 7.40 - 7.50 at 25 °C, Orion Conductivity standards used were 12.9 milliSiemens per centimetre, mS/cm and 1413 microSiemens per centimeter,  $\mu$ S/cm.

## **2.2 Characterization of DNA Samples**

### **2.2.1 Sample Receipt and Specifications**

All synthetically prepared and purified single stranded DNA oligonucleotides used in various aspects of the melting studies were purchased from Integrated DNA Technologies (IDT, Coralville, IA). After their proprietary extraction and precipitation protocol (31), purity levels were reported by the supplier to be greater than 90 %. When received from the supplier, lyophilized single stranded DNA oligonucleotides were in 2.0 mL microcentrifuge tubes and appeared as a translucent film, white powder or glassy pellet. As noted by the supplier, small variations in physical appearances of the oligonucleotides did not affect quality of the DNA. Each synthetic single strand DNA oligonucleotide was accompanied by a specification sheet listing the sequence, desalting and purification procedure, molecular weight, Guanine to Cytosine (G-C) ratio and predicted  $T_m$ . In addition, secondary structure predictions for the DNA molecules done using the OligoAnalyser 3.0™ (IDT, Coralville, IA) server were also provided. Oligonucleotides are susceptible to chemical and enzymatic degradation by nucleases, introduced during handling. As a precaution, gloves were worn at all times and nuclease free conditions were maintained when handling DNA oligonucleotides. Upon receipt, oligonucleotides were stored at -20 °C.

### **2.2.2 General handling of single stranded DNA oligonucleotides**

Prior to opening the 2.0 mL microcentrifuge tubes, single strand synthetic DNA oligonucleotides were equilibrated to room temperature for at least one hour. To ensure that oligonucleotides were fully collected at the bottom,

microcentrifuge tubes were centrifuged for ten minutes at 3000 revolutions per minute (rpm). Each single strand DNA sample was suspended in 2.0 mL buffer containing 85 mM  $\text{Na}_2\text{HPO}_4$  (pH 7.40 - 7.50). This buffer was made from mixing 75 mM  $\text{NaCl}$ , 10.0 mM  $\text{Na}_2\text{HPO}_4$  and 0.10 mM  $\text{Na}_2\text{EDTA}$  to make for a total of 85 mM  $[\text{Na}^+]$  in the stock solution. Each dissolved DNA stock solution was mixed by vortexing for ten minutes, and incubating at 25 °C for at least five hours. The single stranded DNA samples were then briefly vacuum centrifuged on a SpeedVac SVC100 concentrator (Savant Instruments, Holbrook, NY) for ten seconds to remove air bubbles. Samples were then allowed to dissolve completely and equilibrated overnight at -5 °C.

It was necessary to spectrophotometrically determine the exact concentration of the DNAs used in our preparations. Accurate determination of DNA concentration was an essential step required for melting curve analysis and downstream applications. DNA concentrations were determined by measuring the UV absorption at 260 nm, denoted  $A_{260}$ . Sample size, purity and integrity was determined using vertical polyacrylamide gel electrophoresis (PAGE) at four different stages of the study. These steps were: (i) upon initial dissolution of single strands in required buffer; (ii) after preparation of duplex DNA molecules; (iii) upon completion of the DSC melting experiments; and (iv) immediately after buffer exchange. Detailed descriptions of spectral analysis, DNA quantification, and electrophoresis are described in the next section.

### 2.2.3 UV/Vis Absorption Spectroscopy

UV/Vis Spectroscopy is a common technique used to identify and quantify nucleic acids by characteristic absorption spectra (6,12,33). The aromatic ring of purine and pyrimidine moieties absorbs UV light and nitrogenous bases in nucleotides have an  $A_{260}$  (12-13). The absorption spectra can be used to obtain quantitative information about the amount of nucleic acid in a sample (12-13,33). The  $A_{260}$  is routinely used to determine the concentration of nucleic acids present in a solution. The Optical Density, OD, is the  $A_{260}$ , measured in a 1.0 cm path length cell. It is related to the extinction coefficient and sample concentration through the Beer-Lambert law,

$$A_{260} = \text{Log} \left( \frac{I_o}{I} \right) = \epsilon L [C] \quad (2.1)$$

where  $A_{260}$  is the optical density measured at 260 nm,  $\epsilon$  is the molar extinction coefficient of the DNA sample ( $\text{L}/(\text{mol}\cdot\text{cm})$ ).  $I_o$  and  $I$  are the intensities of incident and transmitted light, respectively. The letter  $L$  represents the path length of the light through the sample (cm), and  $[C]$  is the molar concentration of the DNA species (mol/L). The molar extinction coefficient is a physical constant that is unique for each sequence and describes the  $A_{260}$  of 1.0 mol/L.

#### 2.2.4 Spectral Analysis of DNA on UV/VIS Spectrophotometer

Concentrations of diluted single strand solutions were determined from the  $A_{260}$  and the value of molar extinction coefficient provided by the supplier. Spectra were recorded on a Hewlett-Packard 8452A Diode Array Spectrophotometer (Hewlett-Packard Corporation, Palo Alto, CA) equipped with thermostated cell holders. Quartz cuvettes with path lengths of 1.0 cm were used. Before use in spectral analysis, cuvettes were thoroughly rinsed at least five times with bench top ethanol followed by ten or more rinses with nanopure water. In addition, UV lamps were allowed to warm up for at least one hour before taking measurements. Afterwards, a single reference wavelength was measured using 1.2 mL of desired buffer solution alone before DNA sample spectra were measured. This reference buffer measurement was also used to determine the effective dynamic range of the instrument, which is a function of the lamp intensity and the absorbance of the reference sample. The compatible software used was UV-Visible Chemstation (Agilent Technologies, Santa Clara, CA) and it automatically subtracted the blank from the subsequent sample spectra. This spectrum was used for background correction.

In order to determine concentrations of single stranded DNA oligonucleotides, stock solutions were routinely diluted approximately 100 fold. About 20  $\mu\text{L}$  of stock was mixed with 1980  $\mu\text{L}$  of 85 mM  $[\text{Na}^+]$  melting buffer, and the absorbance spectrum from 220 to 340 nm were recorded. Single strand DNA stock solutions were diluted in aliquots that gave  $A_{260}$  between 0.20 - 0.80 OD



units. Between sample runs, cuvettes were thoroughly rinsed twice with bench top ethanol followed by ten times with nanopure water. Aqua regia (50 % nitric acid and 50 % hydrochloric acid in a 1:3 ratio) was used to thoroughly clean cuvettes after each buffer exchange and at least once a week.

### 2.2.5 Preparation of duplex DNA molecules

After obtaining acceptable characteristic DNA spectra, individual single strands were mixed in a 1:1 molar stoichiometric ratio to form desired DNA duplexes. To form the duplex DNA exactly 180 nanoMoles, nM, of each single strand DNA were mixed in a 2.0 mL microcentrifuge tube. To determine the concentrations and volumes of single strand DNAs required in an exact 1:1 stoichiometric ratio, the following procedure was used: (i) Molar concentration of the stock single strand DNA solutions was determined from the  $A_{260}$  reading according to the following procedure:

$$[ssDNA]_{nM} = \left[ \frac{A_{260} [ssDNA] * Dilution Factor}{Extinction Coefficient * 10^9} \right] \quad (2.2)$$

where  $[ssDNA]_{nM}$  is the amount of single strand DNA stock solution containing 180 nM and  $10^9$  is a conversion factor to nM. (ii) Required volumes of each single strand sufficient to give exactly 180 nM of DNA were mixed together in exact 1:1

stoichiometric ratio to make the duplex DNA. (iii) Exact duplex DNA concentration was calculated from the two single strands according to the formula;

$$\left[ \text{duplex DNA} \right]_{\text{mg/mL}} = \left[ \frac{180 * 10^{-9} * \sum M_{wt} (S1 + S2) * 1000}{V_T} \right] \quad (2.3)$$

where  $\sum M_{wt} (S1 + S2)$  is the summation of the molecular weight of single strands,  $S1$  and  $S2$  respectively, in grams.  $V_T$  is the total volume in mL of the duplex DNA containing 180 nM of material.

Duplex solutions were thoroughly mixed by vortexing and let stand at 25 °C for approximately five hours to equilibrate. Afterwards, exact concentrations for the duplexes formed were evaluated by UV spectroscopy. Samples were then run on PAGE to check purity, size and integrity, and stored at -20 °C until use.

## **2.3 Polyacrylamide Gel Electrophoresis, PAGE**

### **2.3.1 Principle of PAGE**

Briefly, PAGE is a high resolution technique used to identify, separate and purify low molecular weight nucleic acids and proteins that differ in size, charge and conformation (35,37-38). When negatively charged DNA molecules are placed in an electric field, they migrate toward the positive pole. Owing to their uniform shape and

charge to mass ratio, DNA molecules are easily separated on the basis of size by gel electrophoresis. Their rate of migration is inversely related to the log of the molecular weight (35-38). Polyacrylamide is a cross-linked polymer of mono-acrylamide and N, N'-methylene bis-acrylamide. The length of the polymer matrix is dictated by the concentration of acrylamide used and N,N'-methylene-bis-acrylamide which is typically between 3 and 20 %. In the presence of free radicals, provided by ammonium persulfate, APS, and catalyzed by tetramethylethylenediamine, TEMED, a reaction occurs such that the acrylamide monomers polymerize to form chains and the bis-acrylamide molecules provide cross-links between the chains. This forms a regular matrix with holes that serve as pores in the polyacrylamide gel (35,38). The size of the holes or pores is determined by two parameters; (i) the amount of acrylamide used and (ii) the degree of cross linkage. The protocol described below has been optimized for the analysis of DNA oligonucleotides six to fifty nucleotides in length, which spans size range for oligonucleotides typically used in this project. DNA sample size, purity and integrity were characterized by PAGE before and after melting analysis. The reagents and gel preparations were as described by Maniatis (35), modified to optimize separation and resolution required for this project.

### **2.3.2 Reagents and Materials**

Materials, reagents, buffers and amounts used in the preparation of PAGE gel are summarized in Table 2.2(a) and (b). Electrophoresis was performed in a single and compact Hoefer MiniVE™ (Hoefer, Inc, Holliston, MA) vertical electrophoresis system.

### **2.3.3 Preparation of PAGE**

Gel concentrations used were between 12 – 16 %. The recipe for a typical 12 % gel is given in Table 2.2(b). Typically, 25.0 mL of working polyacrylamide gel contained 10.7 mL of 28 % Acrylamide:Bis-Acrylamide (37.5:1), 2.5 mL of 10X tris-borate-EDTA (TBE) buffer (pH 8.1- 8.3), 2.5 mL of glycerol and 1.6 mL of freshly prepared 3 % (w/v) APS (see Table 2.2(b)). About 20.0  $\mu$ L of freshly made TEMED in catalytic amounts was added and mixed with other reagents just prior to pouring the gel. For optimum results, the gel solution was degassed for at least ten minutes before adding TEMED. Care was taken not to introduce air bubbles while pouring the gel. After being poured, the gel solution was allowed to polymerize for at least one hour.

### **2.3.4 Sample Preparation**

For gel analysis, DNA concentrations ranging between 2.5  $\mu$ g - 5.0  $\mu$ g were mixed with 1.5  $\mu$ L of 10X loading buffer, 1.0  $\mu$ L of BlueJuice™ (bromophenol blue marker dye) indicator solution and nanopure water. Each well was thoroughly rinsed with the running buffer. Using a pipette with a long and thin tip, about 10.0 – 15.0  $\mu$ L of sample containing between 2.5  $\mu$ g - 5.0  $\mu$ g of DNA was loaded at the bottom of each well submerged in the buffer. About 1.0 – 2.0  $\mu$ L (0.1 - 0.2  $\mu$ g) of ready to use GeneRuler™ (Fermentas, Glen Burnie, MD) DNA ladder premixed

with 6X Orange Loading Dye solution was also loaded in one of the wells in order to quantify and track the DNA fragments.

### **2.3.5 Running the gel**

Electrophoresis was performed using a Hoeffer MiniVE, vertical electrophoresis mini gel system. Constant voltage in the range of 100 - 150 V was supplied from a POWER-PAC300™ (BIO-RAD Company, Hercules, CA) power supply. Gels were subjected to electrophoresis until the Blue Juice™ migrated about three quarters down the length of the gel.

**Table 2.2(a):** Detailed list of materials, reagents and recipes used for PAGE preparations.

Number	Reagent	Amount/ Notes
1	<b>10XTBE Buffer</b>	
	Trizma® base (FW = 121.14)	60.5 g
	Boric acid (FW = 61.83)	30.9 g
	Na <sub>2</sub> EDTA (FW =372.24)	3.70 g
	Nanopure water	1.0 L
	pH of buffer	8.10 – 8.40
2	<b>3 % Ammonium Persulfate</b>	
	Ammonium Persulfate	0.30 g
	Nanopure water	10.0 mL
3	<b>28 % Acrylamide/Bis-Acrylamide</b>	
	Acrylamide	270.0 g
	Bis-Acrylamide	10.0 g
	Nanopure water	1.0 L
4	<b>Blue Juice (Bromophenol Blue)</b>	
	30 % Ficoll®	30.0 g
	0.25 % Bromphenol Blue+ 0.25 % Xylene Cyanol	0.3 g
	10X TBE	100.0 mL
	0.2 M EDTA (pH = 8.00)	20.0 mL
5	<b>6X Loading Dye™ Solution Composition</b> 0.03 % Bromophenol Blue, 0.03 % Xylene Cyanol.	10.0 mM Tris-HCl (pH = 7.60), 60 % Glycerol and 60.0 mM EDTA
6	<b>Storage and Loading Buffer</b> 0.005% Bromophenol Blue, 0.005 % Xylene Cyanol, 10 % Glycerol	10.0 mM Tris-HCl (pH = 7.60).

### **2.3.6 Gel Staining Procedure**

Following electrophoresis, bands were visualized by staining with 3, 3'-diethyl-9-methyl-4, 5, 4', 5'-dibenzothiacarbocyanine dye solution, Stains-All (37,38) (Sigma-Aldrich). Stains-All differentiates nucleic acids and proteins as follows, DNA (blue,  $\lambda = 675$  nm, RNA (bluish purple) and proteins (red,  $\lambda = 510$  nm). Stains-All™ stock solution of 0.1 % (w/v) was used to prepare the working solution in formamide and buffer, given in Table 2.2(b). Typically, a 100 mL volume comprised 0.005 % (w/v) Stains-All™, 10 % (v/v) formamide, 25 % (v/v) isopropanol, 15.0 mL trizma HCl (pH 8.80) and 65 % (v/v) nanopure water. Gels in a Pyrex dish were soaked in a staining solution and covered with aluminium foil. Then, gels were left in the dark for eight to twelve hours on an Orbital Shaker (Mandel Technology Group, Ontario, Canada).

### **2.3.7 Destaining and Visualization**

Bands on gels were detected by gently rinsing and soaking completely in a bath of nanopure water for approximately sixty minutes. Gels were exposed to light at room temperature until bands were observable. An alternative to destaining by water was accomplished by immersing the gel in 40 % (v/v) methanol and 60 % (v/v) glacial acetic acid. Gels were then immediately scanned on an HP Scanjet 4800 series Photo Scanner (Hewlett-Packard Company, Houston, TX) with resolution of 3600 dpi. Scanned images of gels were stored for further reference and subsequent analysis and reference.

**Table 2.2(b):** Typical recipe used for preparation of 12 % PAGE resolving gel and staining.

Reagents in Buffers for 12% PAGE	Amount in 25.0 mL	Recipe for Staining	Amount in 100.0 mL
28 % Acrylamide/Bis-Acrylamide (28.0 g Acrylamide/1.0 g Bis-Acrylamide)	10.7 mL	Stains-All	0.005 % (w/v)
10X TBE Buffer-Stock (pH = 8.30)	2.5 mL	Formamide	10 % (v/v)
10 % (v/v) Glycerol (FW = 92.10)	2.5 mL	Isopropanol	25 % (v/v)
3 % (w/v) Ammonium Persulfate (FW = 228.20)	1.6 mL	Trizma-HCl	15.0 mM (pH = 8.80)
0.0325 % TEMED (FW=116.21)	30.0 $\mu$ L	nanopure water	65 % (v/v)

Stock stain: 0.1 % (w/v) Stains All™ in Formamide, where w/v denotes weight per volume, v/v denotes volume per volume.

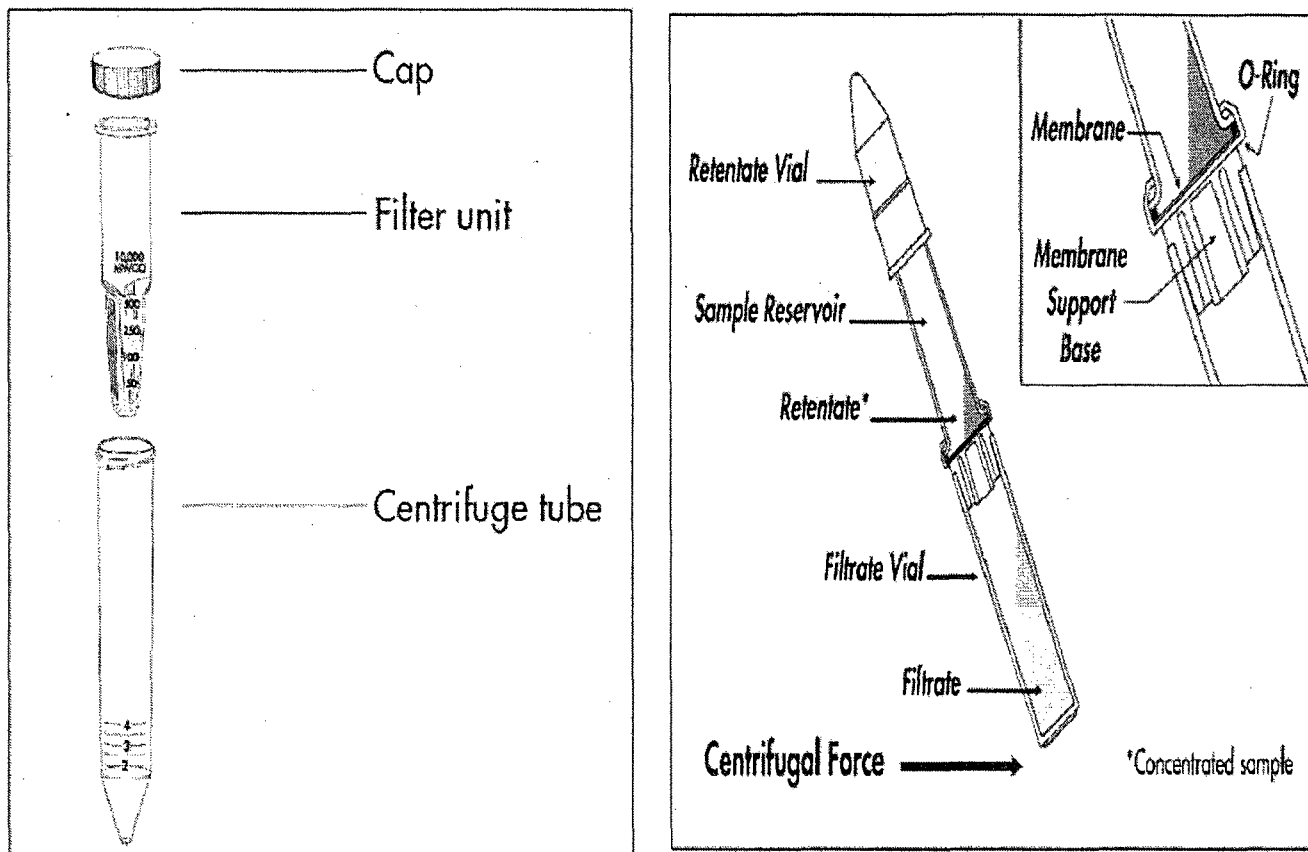
## 2.4 DNA concentration, Desalting and Recovery

Generally, desalting of DNA samples was accomplished by first concentrating DNA samples by Eppendorf 5804R Table Top Centrifuge (Eppendorf, Westbury, NY), then reconstituting the retentate to the original sample volume with the appropriate buffer. DNA samples were transferred between buffers by concentrating and desalting on Centricon YM-3® Centrifugal cellulose filters



(106) or the Amicon<sup>®</sup> Ultra-4 Centrifugal Filter Devices (Millipore, Bedford, MA) with a molecular weight cut off of 3000 Daltons. A centrifugal force drove solvents and low molecular weight solutes through the membrane into the filtrate vial.

Typically, the DNA sample to be concentrated was put into a sample reservoir and diluted with the appropriate melting buffer to 2.0 mL. After centrifugation at 4000 rpm for ninety nine minutes, volume of the sample solution was reduced to about 30 - 60  $\mu$ L. DNA duplexes were retained above the membrane inside the sample reservoir. The sample was washed and centrifuged at least two times to remove any extraneous salt and impurities. To collect the concentrated DNA solution, the retentate vial was placed over the sample reservoir, inverted and centrifuged at 500 g for five minutes. After desalting, the DNA was dried down in a concentrator under vacuum. Concentrated samples were brought up to an appropriate volume of 2.0 mL of the required melting buffer in microcentrifuge tubes.



*Note: Diagrams were adapted from Millipore® Centrifugal Filter Devices User Guide*

**Figure 2.2:** Schematic diagram of the centrifugal filter devices (106) used in this study (a) Left - Amicon® Ultra-4 Centrifuge Filter Unit: The concentrate is collected from the filter unit sample reservoir using a pipette. (b) Right- Centrifuge Filter Assembly: A sample is added to the sample reservoir. After spinning, the centrifugal filter assembly is removed from the centrifuge and then filtrate vial is separated from the membrane support base. The retentate vial is placed over sample reservoir as shown and the unit is inverted to recover the retentate.

An alternative sample recovery procedure was done using Amicon Ultra 4 model, made by the same manufacturer. Up to 4.0 mL of sample was added to the sample reservoir and spinning was done at maximum 4000 rpm for approximately ninety minutes. To recover the concentrated solute, a pipette was inserted into the bottom of the filter unit and the sample was withdrawn using a side-to-side sweeping motion to ensure total recovery. To minimize losses, concentrated samples were immediately removed directly after the centrifuge stopped spinning. Absorbance spectra of the duplex DNA samples were recorded and used in data analysis for that particular salt environment.

## **2.5 Differential Scanning Calorimeter, DSC**

The primary biophysical technique employed in this project was DSC. Some of the basic principle underlying this method and instrumentation are briefly described in this section. General aspects of the DSC will be described according to the following categories: (i) Principle; (ii) Instrumentation; (iii) Sample preparation; (iv) Sample loading and controls; (v) Experimental parameters; and (vi) Data Analysis.

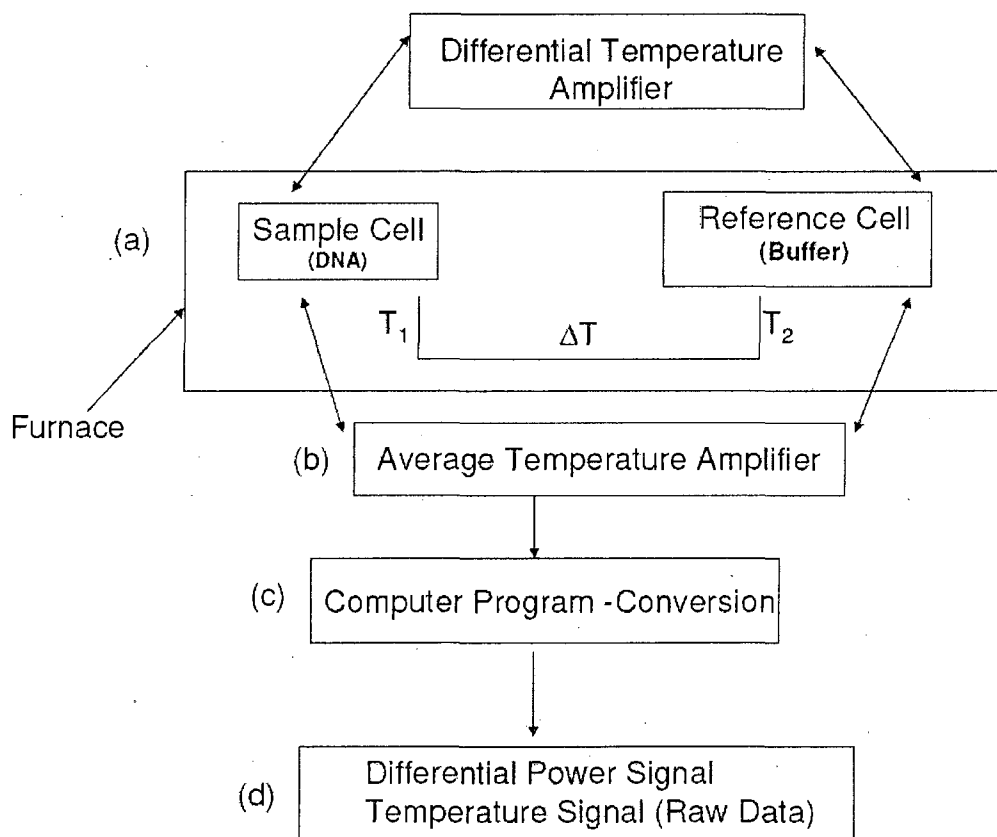
### **2.5.1 Principle**

DSC directly measures the heat energy absorbed or released by the biological sample as it is heated or cooled at a defined, constant rate. Sample temperature and

heat flow are compared with a reference sample. In the adiabatic heat flux method used here, power is determined by measuring the temperature difference across a thermal resistance. For a macromolecular sample heated at a constant rate, the heat capacity of the sample is compared to that of an exactly equal volume of buffer. DSC provides qualitative and quantitative information about physical and chemical changes associated with nucleic acid conformational transitions. The heat capacity is defined as the energy required to raise the temperature of a sample by one degree. As the double stranded DNA macromolecule denatures, there is a sharp increase in heat capacity when the sample begins to denature, because the duplex unwinding process is endothermic process. Observation of an exotherm upon cooling and a reproducible endotherm on re-scanning is evidence of reversibility of the transition. DNA melting is a highly cooperative process and DSC melting curves are represented by the height and breadth of the transition (60,80). Melting curves are generally sharper for perfect matches and tend to broaden as the number of mismatches increases.

The Nano-DSC instrument used in this study is equipped with a pair of gold or tantalum cells with semiconductor sensors. Instrument control is through an external computer that also provides analysis of data. Cells are designed to minimize temperature gradients in the heating and cooling scans and to permit easy washing and reloading of the sample without the introduction of air bubbles, which are known to interfere with the melting process. A schematic diagram of a DSC is shown in Figure 2.3. Briefly, the operating procedure is as follows: (a) The thermal jacket is heated and cooled at a constant rate. (b) Using a known reference voltage, the

temperature of the jacket is controlled by the computer. (c) The temperature of the control circuit then compares the reference voltage to that of the platinum thermometer inside the calorimeter. (d) After comparison, the computer controls the temperature of the jacket by adjusting the power applied to the heating and cooling elements through digital to analog converters to yield the raw melting curves. To convert heat flow into excess heat capacity, the following steps are followed.



**Figure 2.3:** Schematic diagram showing essential features of a DSC. Instrument design, external computer control and raw thermograms for data analysis are depicted in the flow diagram above. (a) A micro furnace is heated at a constant rate, heat is transferred to the sample and reference cells through a thermoelectric disk. (b) Average temperature amplifier increases the signal (c) A computer monitors temperature and regulates heat flow (d) Raw data is collected as Power, in microwatt ( $\mu\text{W}$ ) versus Temperature.

Heat is supplied at the same rate to the two matched cells. The sample cell will generally absorb more heat than the reference cell, causing a slight difference in the temperature,  $\Delta T$  i.e.  $(T_1 - T_2)$  between the two cells. A feedback loop monitoring temperature supplies a small amount of heat,  $q$ , to the solution cell, so as to equalize the temperatures. To obtain excess heat capacity, the raw voltages are converted to power,  $\frac{dq_p}{dt}$ , where  $q_p$  is the heat absorbed at constant pressure.

$$\left( \frac{dq_p}{dt} \right) = \text{heat flow} \quad (2.4)$$

$$\left( \frac{dT}{dt} \right) = \text{heating rate} \quad (2.5)$$

Heat capacity,  $\Delta C_p$ , refers to the ability of material to store heat. The power data are then converted to molar heat capacity,  $\Delta C_p$ , via

$$\left( \frac{dq_p}{dt} \right) / \left( \frac{dT}{dt} \right) = \left( \frac{dq_p}{dT} \right) = C_p \quad (2.6)$$

Finally,  $\Delta C_p^{\text{ex}}$ , is determined as follows,

$$\Delta C_p^{\text{ex}} = \left( \frac{dq_p}{dT} \right) \left( \frac{1}{\sigma M} \right) \quad (2.7)$$

where  $\sigma$  = scan rate,  $\frac{dT}{dt}$  and  $M$  = number of moles of the sample.

### 2.5.2 DSC Melting Experiments

Measurements of the excess heat capacity,  $\Delta C_p^{\text{ex}}$ , versus temperature for DNA solutions were made using either one of four ultra-sensitive, Nano II or CSC Model 6100 Differential Scanning Calorimeters (Calorimetry Sciences Corporation, Provo, UT). These DSC instruments consist of a matched pair of sample and reference cells with an effective cell volume of approximately 0.5 mL. Cells are heated by electric heaters in thermal contact with them, and power to the heaters is adjusted by means of a control circuit, which maintains them at equal temperatures. Measurements were performed under a constant excess pressure of 3.0 atmospheres (atm), which prevents formation of gas bubbles and boiling of aqueous solutions above 100 °C. As reported by the supplier, the noise level of the heating and cooling power difference signal was relatively low ( $\pm 0.015 \mu\text{W}$  standard deviation), with a response half time of five seconds. In addition, power regulation and noise control was done using a Series 800A Commercial Power Purification System (Controlled Power Company, Troy, MI). The general specifications for the DSCs are summarized in Table 2.3.

### **2.5.3 Sample Preparation**

For each DSC experiment, DNA concentrations were determined from absorbance measurements before and after melting experiments as described in Section 2.1.2. Absorbance spectra were measured independently at least twice and DNA concentrations were determined from the average of at least two measurements. Sample requirements were at least 20 OD/ml in a total volume of 2.0 mL buffer. To prepare samples for DSC measurements, sample and buffer solutions were degassed by bubbling with a fine stream of dry helium gas (Polar Cryogenics, Portland, OR) for at least twenty minutes just before use to remove residual air bubbles. Air bubbles can interfere with the melting process and lead to excessive noise, thermal fluctuations, poor baselines and lack of scan reproducibility.

### **2.5.4 Loading the buffer and running a baseline**

Prior to melting the sample the buffer base line was first established by loading both sample and reference cells with the appropriate buffer. In addition to baseline measurements the buffer was also used for rinsing the cells. Additional care was taken in loading the DSC cells with buffer solution. It was absolutely critical that both cells were filled with precisely the same volume of degassed buffer or sample. Further, samples and buffers were loaded without introducing air bubbles. Ten buffer versus buffer scans were made over temperature range 0 - 120 °C. Equilibration of the sample and reference cells was required after sample loading. The cylindrical cells were pressurized at 3.0 atm over the entire temperature range



for all experiments. Data collection began after fluctuations in power leveled off and remained steady for at least 5 minutes.

### **2.5.5 Sample loading and experimental parameters**

After each DSC experiment the sample and reference cells were exhaustively cleaned before loading the new buffer and DNA samples. Generally, each cell was flushed with 20.0 mL of ethanol followed by 1.0 L of nanopure water. Any excess solution remaining in the reservoir was removed by gently wiping with Cotton Swabs (Eastwest Company, Deerfield, IL) and Kimwipes (Kimberly Clark, Waukegan, IL). Depending on the DSC model, filling of the cells was done using a 1.0 mL Hamilton Syringe with a 20 gauge needle or using a micropipette, fitted with a short length of silicone rubber tubing. After loading the sample and before pressurizing the DSC cells, equilibration of the sample and reference cells was required. Equilibration was deemed complete when the microwatt,  $\mu\text{W}$ , reading settled to a constant value.

A total of five heating and five cooling scans were collected for each DNA duplex molecule. The scan rate was 2.0  $^{\circ}\text{C}/\text{min}$ , for both heating and cooling. At least two complete independent sets of DSC melting curves were collected for each DNA sample in each  $[\text{Na}^+]$  environment on different DSC instruments.

### 2.5.6 Data Analysis

Analysis of DSC data was performed using the CpCalc 2.1 (Applied Thermodynamics, Middlesex, NY) software package supplied by the manufacturer for use with the DSC instrument. The buffer versus buffer baseline was used to correct the resultant sample versus buffer scans. Resulting baseline-corrected,  $\Delta C_p^{\text{ex}}$ , versus temperature curves were normalized for total DNA strand concentration, molecular mass, and cell volume. Values for DNA concentration (mg/mL), sample cell volume (0.3269 mL), and partial specific volume of the macromolecule ( $0.55 \text{ cm}^3/\text{g}$ ) were input, and the software automatically subtracted the buffer versus buffer baseline from the DNA versus buffer curves. For these baseline-corrected curves, a progressive polynomial baseline was fit to connect linear regions in the lowest and highest temperature portions of the curve. The DSC melting curve was then constructed by subtracting the fitted base line. Integrated area under the baseline-corrected curve provided a measurement of the calorimetric transition enthalpy,  $\Delta H^{\text{cal}}$ , given by

$$\Delta H^{\text{cal}} = \int \Delta C_p^{\text{ex}} dT \quad (2.8)$$

and the corresponding DSC entropy,  $\Delta S^{\text{cal}}$ , was determined by dividing the excess heat capacity by the temperature and integrating over temperature, thus

$$\Delta S^{cal} = \int \left( \frac{\Delta C_p^{ex}}{T} \right) dT \quad (2.9)$$

For all DSC melting experiments, values of the excess heat capacity,  $\Delta C_p^{ex}$ , versus temperature were measured from 0 - 100 °C or 0 - 120 °C depending on the nature of the sample. For these analyses, it was assumed that overall the difference in excess heat capacity from the beginning to the end of the melting transition is small, (i.e.  $\Delta C_p^{ex} = 0$ ). The calorimetric free energy,  $\Delta G_{25^\circ C}^{cal}$ , was determined at  $T = 298.15$  K by

$$\Delta G_{25^\circ C}^{cal} = \Delta H^{cal} - (T)\Delta S^{cal} \quad (2.10)$$

It is well known that the melting temperature,  $T_m$  of a DNA oligonucleotide duplex refers to the temperature at which the oligonucleotide is 50 % annealed to its exact complement. The  $T_m$  for this particular study was determined as the temperature at the peak height maximum on the baseline corrected,  $\Delta C_p^{ex}$ , versus temperature curve. Finally,  $T_m$  is related to enthalpy and entropy through;

$$T_m = \left( \frac{\Delta H^{cal}}{\Delta S^{cal}} \right) \quad (2.11)$$

For duplexes comprised of two individual strands, the  $T_m$  is not necessarily dependent on strand concentration.

**Table 2.4:** Specifications for DSC melting experiments.

<b>Parameter</b>	<b>Specifications/Unit(s)</b>
Temperature Range	0 -120 °C
Temperature Scan Rate (Heating and Cooling)	2.00 °C/minute
Cell Pressure	3.00 atm
Sample volume loaded in DSC	5.50 mL
Cell Volume for MHC	0.327 mL
Partial Specific Volume (PSV)	0.550 cm <sup>3</sup> /g
DNA concentration units for MHC	mg/ml
Molecular Weight	KiloDaltons (kDa)

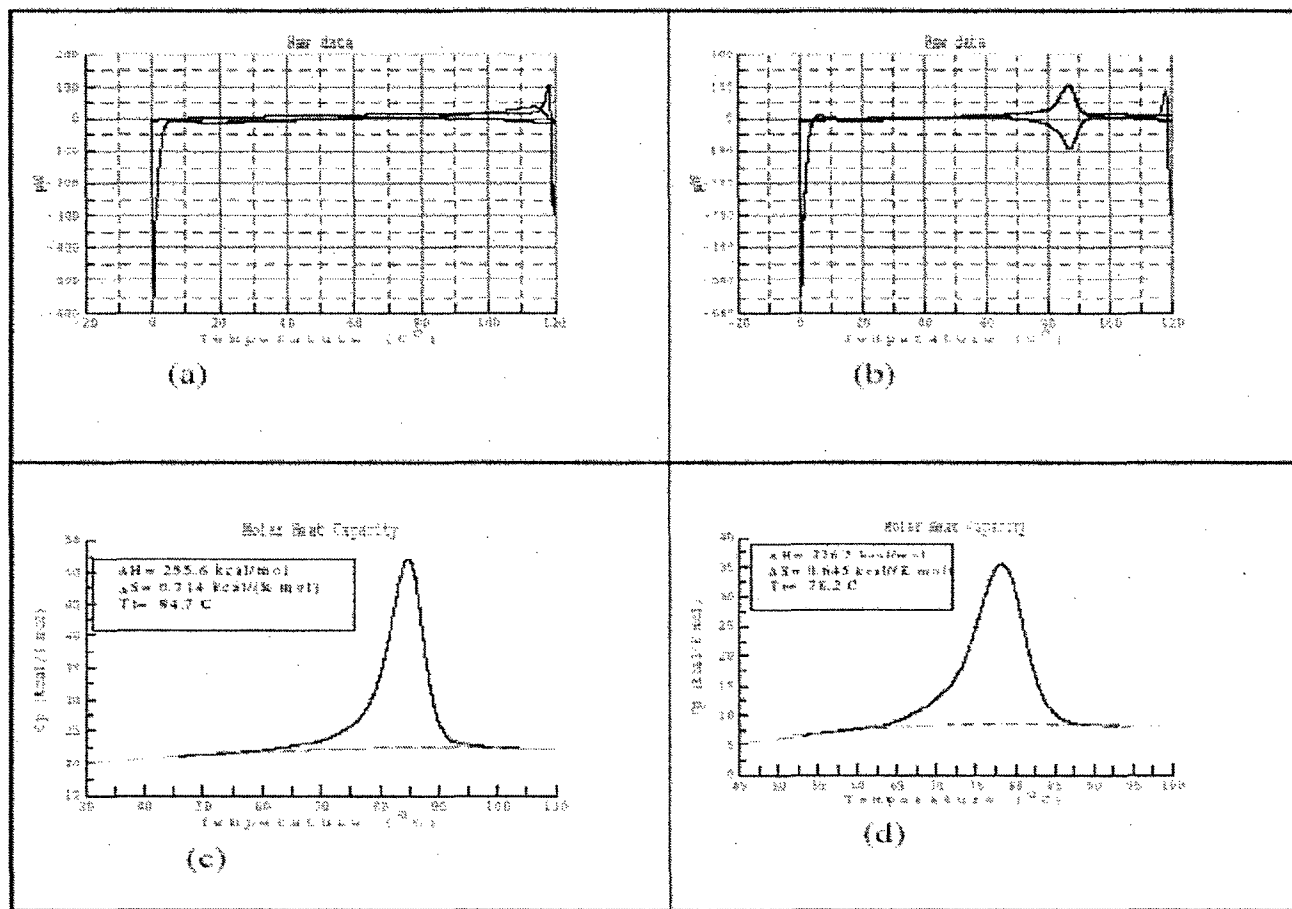
### **2.5.7 Analysis of DSC Melting Data: Results Acquisition**

Examples of typical normalized DSC plots for DNA molecules in 100 mM [Na<sup>+</sup>] salt are shown in Figure 2.4. The raw buffer versus buffer run is shown in Figure 2.4(a). The raw DNA versus buffer run is shown in Figure 2.4(b). As illustrated, at least five DNA versus buffer scans were collected for every sample. Subtraction of the buffer versus buffer scan yields the melting curve for the heating transition, shown in Figure 2.4(c) and the cooling transition shown in Figure 2.4(d). As can be observed, the integrated area under the curve yields  $\Delta H^{cal}$ .

### **2.5.8 Sample Storage after Melting experiments**

Following complete melting curve analysis, the integrity of the DNA samples were affirmed by PAGE. Samples were then centrifuged at 5000 g for ninety nine minutes to desalt and concentrate the DNA to a volume of about 50  $\mu$ L. Afterwards, this volume was vacuum dried and DNA was stored as a pellet at -20 °C.

Power ( $\mu\text{W}$ )



Temperature ( $^\circ\text{C}$ )

Figure 2.4: Representative Calorimetric melting curves (a) Raw buffer versus buffer scan (b) Raw buffer versus sample scan. (c) Representative DSC melting curve for heating transition. (d) Normalized melting curve for the cooling transition.

## 2.6 THEORETICAL METHODS

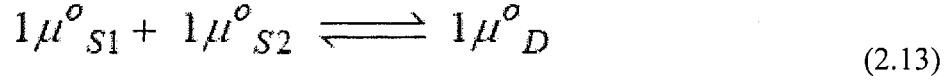
### 2.6.1 Standard State Equilibrium Constant for Short Duplex Hybridization: Schurr's Formalism

Consider a dilute population of DNA single strands comprised of  $N$  base pairs, in solution, at constant temperature and pressure, in a very large volume  $V$ . The simple two-state strand annealing reaction can be represented by the following equilibrium,



where  $S1$  and  $S2$ , are complementary single strands,  $D$  represents the intact duplex, (each species in their standard state), and  $K^o$  is the standard state equilibrium constant for forming the duplex. In this approach the DNA molecules and associated solvent counterions are jointly considered as the solute, while the surrounding water molecules and unassociated ions are collectively treated as the solvent. All solute species are assumed to be sufficiently dilute such that Henry's Law is applicable (40,42). For two single strands forming a duplex as written in Equation (2.12), the equilibrium reaction can be represented in terms of the corresponding chemical potentials as,





where  $1\mu^{\circ}_D, 1\mu^{\circ}_{S1}$  and  $1\mu^{\circ}_{S2}$  are the ‘hypothetical’ standard state chemical potentials for one duplex  $D$  formed from one single strand  $S1$  and one single strand  $S2$ , respectively. The standard state free energy,  $\Delta G^{\circ}$ , is given in terms of the differences of the standard state chemical potentials of the duplex configuration and the single strands from which it formed (40,42). That is,

$$\Delta G^{\circ} = 1\mu^{\circ}_D - 1\mu^{\circ}_{S1} - 1\mu^{\circ}_{S2} \quad (2.14)$$

Because the solutes are assumed to obey Henry’s law, the standard state equilibrium constant for the reaction in Equation (2.12) is also given by,

$$K^{\circ} = \frac{X_D}{X_{S1}X_{S2}} = e^{-(\mu^{\circ}_D - \mu^{\circ}_{S1} - \mu^{\circ}_{S2})/k_B T} \quad (2.15)$$

where  $X_D / X_{S1}X_{S2}$  is the mole fraction equilibrium constant and  $e$  denotes the base of natural logarithms. The quantity in parenthesis in the exponential function is just the standard state free energy change for the reaction in Equation (2.15). The

quantity  $k_B$  is the Boltzmann's constant and,  $T$ , is the absolute temperature. Thus, Equation (2.14) can also be written as

$$K^o = e^{-\Delta G^o/k_B T} \quad (2.16)$$

To obtain an expression for the mechanical standard state equilibrium constant,  $K^o$ , the classical statistical mechanical approach (39-42) is employed. Following that approach, the chemical potential of a molecule of type  $j$ , in its standard state of 1.0 molecule/cm<sup>3</sup> in a solvent,  $s$ , is given by,

$$\mu_j^o = -k_B T \ln \left[ \frac{Q(1_j, N_s, V, T)}{Q(N_s, V, T)} \right] \quad (2.17)$$

where  $Q(1_j, N_s, V, T)$  is the canonical partition function for one  $j$  molecule plus  $N_s$  solvent molecules in a volume,  $V = 1.0 \text{ cm}^3$ , at temperature  $T$ , in Kelvin. The term  $Q(N_s, V, T)$  is the corresponding partition function for the  $N_s$  solvent molecules alone.

Within this model, consider the arbitrary situation in which the subunits of the solvent molecules do not interact with the solute molecules. Consider also that the non-covalent association of  $S1$  and  $S2$ , is stabilized by relatively weak forces

that do not alter the interaction potentials within single strands upon association to form duplex  $D$  (40,41). Lastly, suppose the fixed subunits of  $S1$  and  $S2$  are chosen to be those most proximal and strongly interacting and dominating in the complex. Now, incorporating these assumptions, applying the general definition in Equation (2.17) for  $j = D, S1$  and  $S2$  to the specific reaction given in Equation (2.12) it can be shown that Equation (2.15) can be expressed as,

$$K^o = \frac{X_D}{X_{S1}X_{S2}} = \frac{Q(1_D, N_s, V, T)Q(N_s, V, T)N_s}{Q(1_{S1}, N_s, V, T)Q(1_{S2}, N_s, V, T)} \quad (2.18)$$

where  $Q(1_D, N_s, V, T)$  is the canonical partition function of the duplex state. Quantities  $Q(1_{S1}, N_s, V, T)$  and  $Q(1_{S2}, N_s, V, T)$  are the canonical partition functions for the single strands  $S1$  and  $S2$ , respectively. The effect of  $N_s$  is to cancel the configurational degeneracy of the center of the solute molecule, which could occupy the same space as the center of any given solvent molecule. Thus, the equilibrium constant only depends upon the contributions from solutes in single lattice sites, including solvent-solute interactions. The canonical partition function can be expressed as (40-42),

$$Q(1_j, N_s, V, T) = \frac{1}{N_j!} \left( \frac{2\pi m_j k_B T}{h^2} \right)^{\frac{3N}{2}} Z(1_j, N_s, V, T) \quad (2.19)$$

where,  $1/N_j!$  represents the number of molecules of type  $j$ , for every kind of molecule or ion in the system for this classical partition function representation.

The quantity  $\left(\frac{2\pi m_j k_B T}{h^2}\right)^{\frac{3N}{2}}$  is the translational partition function term, and

$Z(1_j, N_s, V, T)$  is the classical configurational integral for a  $j^{\text{th}}$  molecule, defined as (40)

$$Z(1_j, N_s, V, T) = \int \dots \int dr_1 \dots dr_N e^{-\beta U(r_1, \dots, r_N)} \equiv Z(N) \quad (2.20)$$

where  $U$  is the potential energy,  $\beta$  is the inverse thermal energy,  $\beta(T) = 1/k_B T$ . Now, incorporating Equations (2.19) and (2.20), the mole fraction equilibrium constant in Equation (2.18) is expressed as

$$K^o = \frac{\left(\frac{1}{N_s}\right)^2 \left[ \prod_{j=\alpha_0}^{\delta_0} \left(\frac{2\pi m_j k_B T}{h^2}\right)^{\frac{3n^0 j}{2}} \right] \left[ \prod_{k=\alpha_0}^{\delta_0} \left(\frac{2\pi m_k k_B T}{h^2}\right)^{\frac{3nk}{2}} \right]^{2N_s} Z(1_{D_0}, N_s) Z(N_s) N_s}{\left(\frac{1}{N_s}\right)^2 \left[ \prod_{j=\alpha_1}^{\delta_1} \left(\frac{2\pi m_j k_B T}{h^2}\right)^{\frac{3n^1 j}{2}} \right] \left[ \prod_{j=\alpha_2}^{\delta_2} \left(\frac{2\pi m_j k_B T}{h^2}\right)^{\frac{3n^2 j}{2}} \right] \left[ \prod_{k=\alpha_0}^{\delta_0} \left(\frac{2\pi m_k k_B T}{h^2}\right)^{\frac{3nk}{2}} \right]^{2N_s} Z(1_{S1}, N_s) Z(1_{S2}, N_s)} \quad (2.21)$$

For this classical system, every molecule is assumed to consist of completely rigid subunits. Following this assumption, the kinetic energy contribution to the total partition function for this system can be expressed in terms of masses and moments

of inertia of their rigid subunits, and is always the same, independent of the potential energy function (40-42). Consequently, the unperturbed solvent contributions cancel and the configurational degeneracy of the centers of different species is removed in Equation (2.21). Moreover, according to the law of conservation of matter, a chemical reaction must conserve atoms for each kind  $j$  of atom, thus

$$n_j^D = n_j^{S1} + n_j^{S2} \quad (2.22)$$

Following these assumptions, the term  $\frac{1}{N_s}$  and the kinetic energy terms for the solvent cancel exactly. As a result, the kinetic energy factors of  $S1$  and  $S2$  exactly cancel those in  $D$ . Equation (2.21) is expressed as

$$K^o = \frac{X_D}{X_{S1}X_{S2}} = \frac{Z(1_D, N_s)Z(N_s)N_s}{Z(1_{S1}, N_s)Z(1_{S2}, N_s)} \quad (2.23)$$

Thus, from classical statistical mechanics, the equilibrium constant,  $K^o$ , is independent of all masses and depends only on the configurational integrals (40). Standard states for bimolecular association reactions have been discussed by Gilson *et al.* (42) but here, the classical statistical mechanics approach of Schurr (40-41), is

followed. Following that development, the standard state equilibrium constant,  $K^o$  can be represented by,

$$K^o = \left( \frac{Z_{config}^{rel}(T)}{\bar{V}_s 8\pi^2} \right) e^{(-\Delta G_m^o/k_B T)} \quad (2.24)$$

where  $\bar{V}_s \equiv V/N_s$ , i.e. the volume divided by the number of solvent molecules,  $N_s$ , and  $Z_{config}^{rel}(T)$  is a six dimensional integral, 6-D, with having units of spatial and Euler volume weighted by  $\exp(-U(R,\Omega)/k_B T)$ , where  $U(R,\Omega)$  is the potential energy which is function with  $R$  and  $\Omega$  corresponding to the relative translational and Euler coordinates, respectively. For this integral, it is necessary to specify an arbitrarily defined domain,  $D_m$  of the 6-D (Euler + translational) coordinate space occupied by the pair of complexed single strands. Specification of this domain is necessary because the duplex complex is defined for predominately repulsive forces, which is the case for the  $N = 0$  standard state of the duplex, and it is necessary to restrict the integration to this domain. All real potentials, whether attractive or repulsive, do not rise indefinitely with increasing separation. Thus, all complexes require some definition of their domain in the relative coordinate space.

In the duplex complex,  $D$ , the 1-subunit of  $S1$  is in the vicinity of the 1-subunit of  $S2$ , while all other subunits of  $S1$  and  $S2$  in the complex,  $D$ , and all subunits of the  $N_s$  solvent molecules adopt all configurations with weights given

by  $\exp[-U(R, \Omega)/k_B T]$ . Separating the potential energy into translational and Euler terms, (i.e.  $U(R, \Omega) = U_{S1:S2}^R(R) U_{S1:S2}^\Omega(\Omega)$ ), two integrals comprising  $Z_{config}^{rel}(T)$  emerge. The first integral, weighted by  $\exp[-U_{S1:S2}^R/k_B T]$  for all other subunits, is the volume available to the 1-subunit of  $S1$  in the vicinity of the 1-subunit of  $S2$  in the complex,  $D$ . The first term in the denominator,  $\bar{V}_s$ , is just the standard state molecular volume available to the solute for the mole fraction 1.0. The second integral, weighted by  $\left(\exp[-U_{S1:S2}^\Omega/k_B T]\right)$  for all other subunits, is the Euler volume available to the 1-subunit of  $S1$  in the vicinity of the 1-subunit of  $S2$  in the complex,  $D$ . The second term in the denominator,  $8\pi^2$ , is the Euler volume available to the solute in its standard state.

The standard state equilibrium constant,  $K^o$ , reflects the relative probability of finding strands  $S1$  and  $S2$  in their individual standard states compared to finding them in the duplex in its standard state. It is determined entirely by weighting the integrals over relative translational and Euler coordinates. In essence, the first term on the right side of Equation (2.24) contains the relative translational and Euler angular orientations of strands ( $S1$  and  $S2$ ) and solvent displaced from both single strands during formation of the duplex complex. The exponential argument of the second term in Equation (2.24), (i.e.  $-\Delta G_{int}^o/k_B T$ ) corresponds to the favorable attractive interactions between strands  $S1$  and  $S2$  that form the duplex. Such interactions are thought to include hydrogen bonding, stacking and solvent

interactions that may also depend on  $[\text{Na}^+]$ . To make the connection with terminology in the literature we define  $\Delta G_{int}^o \equiv \Delta G_{duplex}^o$ , so that Equation (2.24) becomes,

$$K^o = \left( \frac{Z_{config}^{rel}(T)}{\overline{V}_s 8\pi^2} \right) e^{(-\Delta G_{duplex}^o / k_B T)} \quad (2.25)$$

The parenthetical expressions on the right sides of Equations (2.24) and (2.25) suggest the existence of a “fictitious” reference state,  $K^o(N=0)$ , where the complementary single strands ( $S1$  and  $S2$ ) reside in precisely the same spatial configuration and Euler volume as in the duplex complex,  $D$ , only without the accompanying attractive interactions (i.e. hydrogen bonding, stacking) so that  $\Delta G_{duplex}^o = 0$ . More generally, Equation (2.25) can be written as,

$$K^o = e^{(-\Delta G_{reference}^o / k_B T)} e^{(-\Delta G_{duplex}^o / k_B T)} \quad (2.26)$$

with

$$\Delta G_{reference}^o = -k_B T \ln \left( \frac{Z_{config}^{rel}(T)}{\overline{V}_s 8\pi^2} \right) \quad (2.27)$$



Thus, the standard state free energy of the reaction,  $\Delta G_{reaction}^0$  is given by,

$$\Delta G_{reaction}^0 = \Delta G_{duplex}^0 + \Delta G_{reference}^0 \quad (2.28)$$

Historically, the free-energy of duplex formation has been partitioned into two contrasting terms (3-4,7,10,16,20) viz.

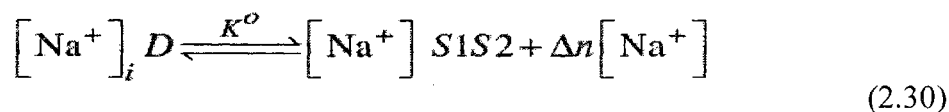
$$\Delta G_{reaction}^0 = \Delta G_{duplex}^0 + \Delta G_{initiation}^0 \quad (2.29)$$

Comparison of Equations (2.28) and (2.29) reveals that  $\Delta G_{initiation}^0 \equiv \Delta G_{reference}^0$  but (as discussed before), our interpretation of  $\Delta G_{reference}^0$  and the historical interpretation of  $\Delta G_{initiation}^0$  are fundamentally different. Experimental estimates on the values of  $\Delta G_{reference}^0$  as a function of  $[Na^+]$  were made in the experimental studies described in Chapters 3 and 4.

### 2.6.2 Release of sodium from duplex upon melting

Pioneering experiments on the effects of monovalent cations on DNA were carried out by Marmur and Doty (51) and Schildkraut *et al.* (56), between 1959 and 1962. The effects of  $[Na^+]$  and monovalent cations on the stability of DNA have been reported by several authors (8-11;29,44,56-62). DNA polyions are characterized by an extremely high linear charge density due to the negative charge

of phosphate groups and allows for the mobile counterions to surround the double helix. In the particular case of the reaction represented in Equation (2.1), the equilibrium reaction between single strands  $S1$  and  $S2$  and the duplex,  $D$ , is characterized by release and uptake of  $\text{Na}^+$ , i.e.



where  $\Delta n = i' - i$  is the stoichiometric counterion release upon melting the DNA. Generally, when DNA duplex strands denatures, there is a net release of  $\text{Na}^+$  due to reduction in the charge density of the single strands and the rearranged solvent compared to the double strand (15,54,61). Generally, DNA in an intact, double helical form has a higher charge density as compared to the single-stranded, random coil form. Because of this differential charge density, more  $\text{Na}^+$  binds more strongly to the duplex form. It is well established that DNA helix formation is accompanied by an increase in counterion association with a favorable entropic contribution (53-61). An increase in bulk salt concentration stabilizes the duplex state with higher charge density compared to the coiled state, resulting in an increase in  $T_m$  of the melting transition at higher salt concentration. During the melting process, the differential number of bound counterions for duplex and single stranded states, of DNA remains constant over a certain salt concentration range.

In this thesis, the relative contribution of the sodium ions,  $\text{Na}^+$ , to the overall DNA hybridization energetics is explored. The dependence of short DNA melting was evaluated as a function of duplex length and  $\text{Na}^+$  through the free energy and melting temperature. Generally, if  $\text{Na}^+$  is changed at constant water activity, the equilibrium constant for DNA melting depends on  $\text{Na}^+$  of the buffer. From the plot of  $T_m^{-1}$  ( $\text{K}^{-1}$ ) versus  $\ln[\text{Na}^+]$ , the release of ions that occurs upon melting,  $\Delta n$  can be evaluated according to the equation (11,57,60-61)

$$\alpha\Delta n = \left( \frac{-\Delta H^\circ(d(T_m^{-1}))}{R(d\ln[\text{Na}^+])} \right) \quad (2.31)$$

where  $\Delta H^\circ$  is the transition enthalpy of melting per nucleotide,  $R$  is the gas constant, expressed in  $[\text{cal}/(\text{mol K})]$ , and  $\alpha = 0.90$ , is a correlation factor that accounts for the changes in the mean ionic activity coefficient of  $\text{Na}^+$  in the range examined. The plot of  $T_m^{-1}$  ( $\text{K}^{-1}$ ) versus  $\ln[\text{Na}^+]$  yields a slope given by

$d(T_m^{-1})/d\ln[\text{Na}^+]$  and is used to estimate the amount of  $\text{Na}^+$  released upon

melting of the duplex sequence,  $\Delta n$ . Alternatively, the value  $dT_m/d\log[\text{Na}^+]$ , is a

characteristic property of a particular a polynucleotide system. A more comparable

quantity is the differential counterion release per phosphate group, denoted  $\Delta\psi$  is given by the following expression (57)

$$\frac{dT_m}{d\log[\text{Na}^+]} = 0.9 \left( \frac{2.303RT_m^2}{\Delta H^\circ} \right) (\Delta)\psi \quad (2.32)$$

For the two-state melting transition of DNA oligonucleotide duplexes, this cooperative unit is equal to the entire duplex. The DNA phosphate charge neutralized by the monovalent cations can be evaluated by using the ratio

$\left( \frac{2.303RT_m^2}{\Delta H^\circ} \right)$ , and has been found experimentally to be constant (11,57).

Thus for a DNA molecule with N base pairs, the following expression holds

$$\Delta\psi = \frac{\Delta n}{N} \quad (2.33)$$

From Equation (2.33), values of  $\Delta\psi$  and  $\Delta n$  will be calculated from the empirically determined  $T_m$  values in different salt environments.

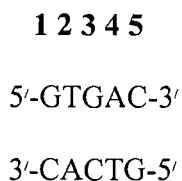
## CHAPTER 3

### ORIGINS OF THE 'NUCLEATION' FREE ENERGY TERM IN PREDICTIONS OF THE THERMODYNAMICS OF SHORT DUPLEX DNA HYBRIDIZATION

#### 3.0 INTRODUCTION

The association and melting of short oligoduplex DNAs and DNA fragments has been the subject of considerable investigation over the past 50 years (6,8-18,44). Most recently DNA probe hybridization and wide use and application of the polymerase chain reaction have sparked great interest in DNA probe design and thermodynamics of short duplex DNA stability. The equilibrium formation of short DNA oligoduplexes is of central importance in many molecular diagnostics and biotechnology applications. Sequence dependent predictions of short duplex DNA stability commonly employ the n-n model parameters. For short DNA sequences containing perfectly matched or single base pair mismatches, the n-n model provides reasonably accurate predictions of their sequence dependent thermodynamic stabilities (6-7,15-16,20-22). The n-n model is based on the assumption that thermodynamic stability of a given base pair depends on both the

identity and orientation of the n-n base pairs. As explained earlier, the n-n model is based on the assumption that thermodynamic stability of a given base pair depends on both the identity and orientation of the n-n base pairs. To demonstrate the use of n-n model in the calculation of duplex stability, consider the following duplex DNA sequence, comprised of five base pairs:



This DNA duplex sequence contains four n-n doublets (numbered 1-5 in bold), with corresponding free-energies;  $\Delta G_{1,2}$ ,  $\Delta G_{2,3}$ ,  $\Delta G_{3,4}$ ,  $\Delta G_{4,5}$ , where  $\Delta G_{1,2} = \Delta G^\circ(\text{GT/CA})$ ;  $\Delta G_{2,3} = \Delta G^\circ(\text{TG/AC})$ ;  $\Delta G_{3,4} = \Delta G^\circ(\text{GA/CT})$ , and  $\Delta G_{4,5} = \Delta G^\circ(\text{AC/TG})$  are the n-n doublet parameters. Thus, the calculated n-n free-energy of this sequence segment is given by

$$\begin{aligned}
 \Delta G^\circ_{duplex} = & \Delta G^\circ(\text{GT/CA}) + \Delta G^\circ(\text{TG/AC}) + \Delta G^\circ(\text{GA/CT}) + \\
 & \Delta G^\circ(\text{AC/TG}) + \Delta G^\circ_{initiation}
 \end{aligned} \tag{3.1}$$

From Equation (2.29) note that the total free energy change for helix melting is partitioned into the free energy for helix initiation, the free energy change for base pairing in helix propagation. The n-n doublet parameters are available and were

recently tabulated (31,89). The purpose of this study is to provide a more quantitative, experimentally determined basis for interpreting the nucleated or initiation complex in DNA duplex formation. In this work, calorimetric thermodynamic parameters, ( $\Delta H^{cal}$ ,  $\Delta S^{cal}$  and  $\Delta G_{25^{\circ}C}^{cal}$ ) as well as  $T_m$  were measured for 19, well defined DNA duplexes ranging in length from 6 – 35 base pairs. Changes in counter ion release, ( i.e.  $\text{Na}^+$  ), upon melting are evaluated as a function of duplex length.

### 3.1 EXPERIMENTAL AND THEORETICAL METHODS

A general description of the Experimental and Theoretical Methods has already been presented in Chapter 2. However, only methods specific to this project will be presented in this chapter.

#### 3.1.1 DNA Molecules

Sequences of the 19 DNA duplexes that were studied are listed in Table 3.1. Length of DNA duplexes was varied incrementally from 6 to 35 base pairs and sequences were judiciously designed to have representative combinations of the 10 possible n-n DNA base pair stacks. Sequences were also designed with variable n-n doublet base pairs composition to maintain the % GC in the range 36 - 55 %. The ten unique n-n doublets appear in the DNA duplexes shown in Table 3.1 with the following frequencies; TA/AT [21], GC/CG [17], AA/TT [30], CT/GA [28],

CG/GC [19], TG/AC [43], AT/TA [28], AC/TG [46], GG/CC [30], GA/CT [47]. Predicted thermodynamic stabilities for perfect match duplexes, potential hairpin and dimer complexes were obtained using publicly available methods, most notably HYTHER (63), OligoAnalyser (31) and MFOLD (64). DNA oligonucleotides were purchased from Integrated DNA Technologies. As received from the supplier, lyophilized single stranded DNA appeared as a translucent film or white powder. Upon receipt, DNA oligonucleotides were stored at -20 °C.

### **3.1.2 Preparation of Buffer Solutions**

Buffers were prepared as described in Materials and Methods section.



**Table 3.1:** Perfectly matched DNA sequences ranging in length from 6 - 35 base pairs. N corresponds to the number of W/C base pairs in the DNA sequence and % GC corresponds to the percentage of guanine and cytosine present in the duplex.

N	Sample ID	DNA Sequence	%GC
6	DNA6PM	5'-GGATGT-3' 3'-CCTACA-5'	50.0
8	DNA8PM	5'-GGATGTTA-3' 3'-CCTACAAT-5'	46.0
9	DNA9PM	5'-GGATGTTAG-3' 3'-CCTACAATC-5'	48.0
10	DNA10PM	5'-GGATGTTAGC-3' 3'-CCTACAATCG-5'	50.0
11	DNA11PM	5'-GGATGTTAGCG-3' 3'-CCTACAATCGC-5'	54.5
12	DNA12PM	5'-GGATGTTAGCGA-3' 3'-CCTACAATCGCT-5'	50.0
13	DNA13PM	5'-GGATGTTAGCGAC-3' 3'-CCTACAATCGCTG-5'	53.8
14	DNA14PM	5'-GGATGTTAGCGACA-3' 3'-CCTACAATCGCTGT-5'	50.0
15	DNA15PM	5'-GGATGTTAGCGACAA-3' 3'-CCTACAATCGCTGTT-5'	46.7
16	DNA16PM	5'-GGATGTTAGCGACAAG-3' 3'-CCTACAATCGCTGTTCC-5'	50.0
17	DNA17PM	5'-GGATGTTAGCGACAAGG-3' 3'-CCTACAATCGCTGTTCC-5'	52.9
18	DNA18PM	5'-GGATGTTAGCGACAAGGT-3' 3'-CCTACAATCGCTGTTCCA-5'	50.0
19	DNA19PM	5'-GGATGTTAGCGACAAGGTC-3' 3'-CCTACAATCGCTGTTCCAG-5'	52.6
20	DNA20PM	5'-GGATGTTAGCGACAAGGTCA-3' 3'-CCTACAATCGCTGTTCCAGT-5'	50.0
23	DNA23PM	5'-GGATGTTAGCGACAAGGTCATTG-3' 3'-CCTACAATCGCTGTTCCAGTAAC-5'	47.8
25	DNA25PM	5'-GGATGTTAGCGACAAGGTCATGTCG-3' 3'-CCTACAATCGCTGTTCCAGTACAGC-5'	52.0
27	DNA27PM	5'-GGATGTTAGCGACAAGGTCATGTCGAT-3' 3'-CCTACAATCGCTGTTCCAGTACAGCTA-5'	48.1
30	DNA30PM	5'-GGATGTTAGCGACAAGGTCATGTCGATACC-3' 3'-CCTACAATCGCTGTTCCAGTACAGCTATGG-5'	46.7
35	DNA35PM	5'-GGATGTTAGCGACAAGGTCATGTCGATACCTATGC-3' 3'-CCTACAATCGCTGTTCCAGTACAGCTATGGATACG-5'	48.6

### 3.1.3 Preparation of DNA molecules

Specifically for this project, DNA duplex concentration  $[DNA_{duplex}]$

was calculated from the two single strands using the following equation;

$$[DNA_{duplex}] = \left[ \frac{180 * \sum Mwt (S1 + S2)}{V_T * 10^6} \right] \quad (3.2)$$

where  $[DNA_{duplex}]$  is given in mg/mL,  $\sum Mwt (S1 + S2)$  is the sum of the molecular weights of single strands  $S1$  and  $S2$  in grams per mole, and  $V_T$  is the total volume in mL of the duplex DNA solution containing 180 nM of material. Duplex concentrations ranged from 0.46 to 1.54 mg/mL.

### 3.1.4 Analytical Polyacrylamide Gel Electrophoresis

Details of PAGE have been described in Chapter 2.

### 3.1.5 DSC Melting Experiments, Data Analysis

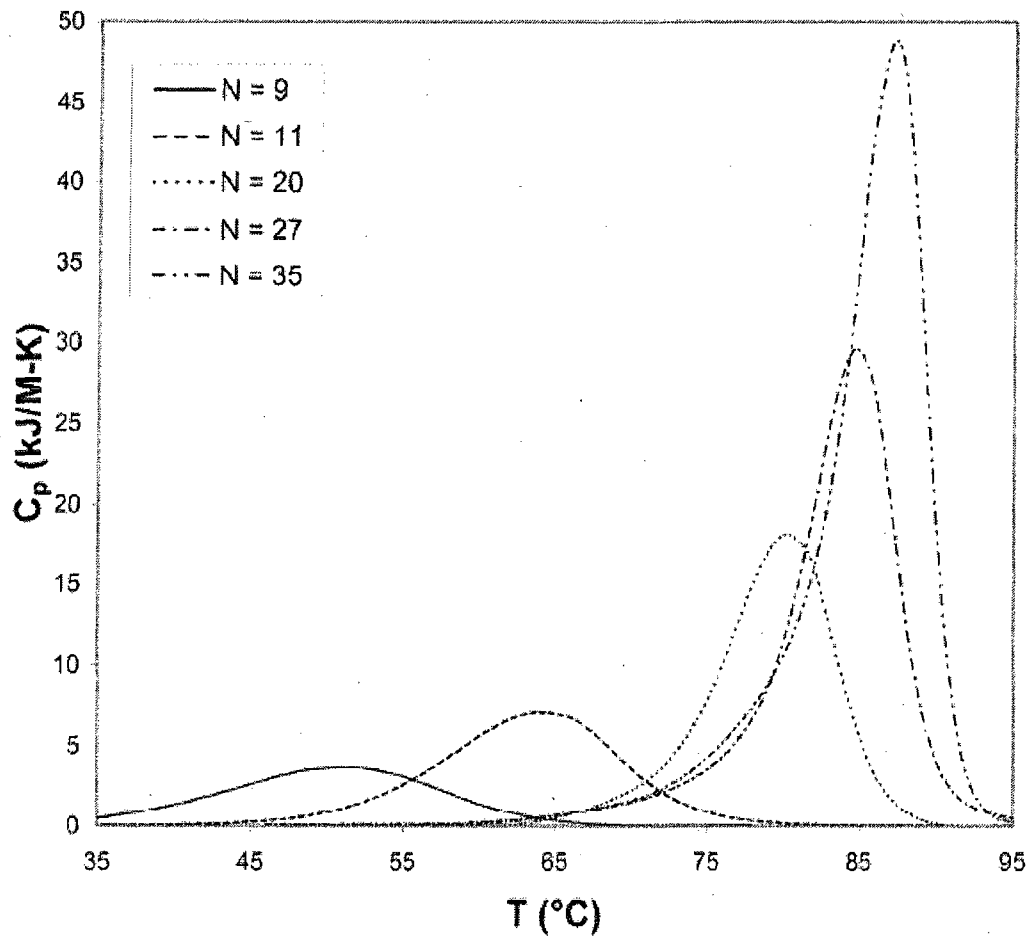
Four Ultra-sensitive, CSC Model 6100, Nano II-Differential Scanning Calorimeters were employed to measure excess heat capacity,  $\Delta C_p^{ex}$ , versus temperature curves for DNA solutions. Details of data analysis were as described in Chapter 2. Desalting of DNA samples was done using strictly Amicon<sup>®</sup> Ultra-4 Centrifugal Filter Devices.

## 3.2 RESULTS

### 3.2.1 Analysis of DSC melting curves

For all DSC melting curve experiments, values of the excess heat capacity,  $\Delta C_p^{\text{ex}}$ , versus temperature thermograms were measured from 0 - 120 °C. Over 740 DSC melting curves were measured. In the analysis of DSC melting data, it was assumed the overall change in excess heat capacity from the beginning to the end of the melting transition is nil, (i.e.  $\Delta C_p^{\text{ex}} = 0$ ). In all cases, multiplex scans were recorded and the first heating transition was discarded largely because of higher noise due to sample degassing.

That the melting and cooling transition overlapped indicated the measured transitions occur under equilibrium conditions. Examples of the normalized,  $\Delta C_p^{\text{ex}}$ , versus temperature plots for DNA molecules in all  $\text{Na}^+$  concentrations are shown in Figure 3.1. The integrated area under the curve yields  $\Delta H^{\text{cal}}$ . The DSC melting curves are displayed in two ways. The normalized differential melting curves for the 9, 11, 20, 27 and 35 base pairs DNA molecules in 1.0 M  $[\text{Na}^+]$ . At lower N, melting curves are broader and less symmetric with lower peak height (i.e. as illustrated by the 9 and 11 base pairs DNA molecules). At higher N, the DSC melting curves get sharper and more symmetric. DSC melting curves for some of the DNAs were also plotted as a function  $[\text{Na}^+]$  (data not shown). Melting transitions were very sensitive to  $\text{Na}^+$  concentrations. As expected, thermodynamic stabilities of the DNA duplexes increased with length and increasing  $[\text{Na}^+]$ .



**Figure 3.2:** Representative DSC melting curves plotted together on the same scale. Normalized DSC plots as a function of  $N$  for 9, 11, 20, 27 and 35 base pairs DNAs respectively in 1.0 M  $[\text{Na}^+]$ .

**Table 3.2:** Experimentally measured DSC thermodynamic data for DNA duplexes with 6 – 35 base pairs in all [Na<sup>+</sup>].

DSC - 1 M [Na <sup>+</sup> ]										Mfold			
ID	N	$\Delta H$ kcal/mol	$\sigma$	$\Delta S$ e.u.	$\sigma$	$\Delta G_{25}$ kcal/mol	$\sigma$	$T_m$ °C	$\sigma$	$\Delta H$ kcal/mol	$\Delta S$ e.u.	$\Delta G_{25}$ kcal/mol	$T_m$ °C
DNA6PM	6	-72.1	5.6	-219.0	18.1	-6.8	0.4	25.4	0.7	-42.4	-121.2	-6.3	15.7
DNA8PM	8	-92.4	3.8	-280.8	10.1	-8.7	0.8	43.7	1.03	-53.4	-149.3	-8.9	32.1
DNA9PM	9	-94.1	5.2	-282.6	14.9	-9.8	0.8	51.6	0.6	-63.0	-177.9	-9.9	36.3
DNA10PM	10	-117.0	9.2	-352.5	27.4	-11.9	1.1	56.7	0.5	-72.8	-202.3	-12.5	46.2
DNA11PM	11	-121.3	6.7	-360.0	25.7	-13.9	1.0	62.2	0.4	-83.4	-229.5	-15.0	53.7
DNA12PM	12	-122.2	9.2	-360.2	28.9	-14.8	0.7	66.3	0.2	-89.5	-244.3	-16.7	58.4
DNA13PM	13	-130.4	7.8	-376.3	15.8	-18.2	3.1	69.6	0.3	-100.0	-274.2	-18.2	60.4
DNA14PM	14	-139.0	3.2	-403.0	9.7	-18.8	0.4	71.1	0.4	-110.8	-304.4	-20.0	62.6
DNA15PM	15	-145.3	1.9	-422.7	5.8	-19.3	0.2	73.8	0.1	-114.2	-312.2	-21.1	64.9
DNA16PM	16	-157.9	5.5	-456.0	16.9	-21.9	0.5	74.3	0.2	-124.2	-340.2	-22.8	66.4
DNA17PM	17	-158.3	4.3	-446.2	12.2	-25.3	0.7	75.3	0.2	-132.2	-360.1	-24.9	69.6
DNA18PM	18	-163.2	6.6	-457.7	19.0	-26.8	1.0	77.5	0.4	-142.9	-390.3	-26.5	70.4
DNA19PM	19	-165.3	12.1	-460.3	32.9	-28.0	2.4	78.6	0.2	-148.8	-404.7	-28.1	72.6
DNA20PM	20	-182.4	6.5	-513.3	18.4	-29.3	1.0	80.2	0.1	-159.6	-435.0	-29.9	73.4
DNA23PM	23	-215.6	6.4	-609.1	16.3	-34.0	1.5	80.9	0.1	-180.9	-492.8	-34.0	75.8
DNA25PM	25	-247.3	6.4	-703.8	16.3	-37.5	1.5	83.7	0.4	-200.2	-542.4	-38.5	79.3
DNA27PM	27	-256.5	4.7	-716.1	13.3	-43.0	0.8	85.5	0.2	-213.4	-578.1	-41.0	80.3
DNA30PM	30	-275.8	9.3	-770.4	26.8	-46.1	1.4	85.9	0.5	-239.2	-648.7	-45.8	81.6
DNA35PM	35	-290.0	12.9	-807.0	34.9	-49.4	2.5	87.0	0.3	-279.7	-758.6	-53.5	83.5

DSC - 600 mM [Na <sup>+</sup> ]										Mfold			
ID	N	$\Delta H$ kcal/mol	$\sigma$	$\Delta S$ e.u.	$\sigma$	$\Delta G_{25}$ kcal/mol	$\sigma$	$T_m$ °C	$\sigma$	$\Delta H$ kcal/mol	$\Delta S$ e.u.	$\Delta G_{25}$ kcal/mol	$T_m$ °C
DNA6PM	6	-63.7	3.8	-198.2	11.0	-4.6	1.0	25.1	0.7	-42.4	-122.1	-6.0	13.8
DNA8PM	8	-73.4	2.9	-224.9	8.5	-6.3	0.7	42.5	0.25	-53.4	-150.6	-8.5	29.8
DNA9PM	9	-82.9	0.6	-251.0	1.8	-8.0	0.1	47.6	0.2	-63.0	-179.4	-9.5	34.0
DNA10PM	10	-106.7	3.7	-324.0	12.7	-10.1	0.9	54.3	0.3	-72.8	-204.0	-12.0	43.9
DNA11PM	11	-103.6	3.1	-307.3	9.6	-12.0	0.8	64.2	0.1	-83.4	-231.4	-14.4	51.3
DNA12PM	12	-111.0	10.1	-328.9	29.9	-13.0	2.5	64.8	0.2	-89.5	-246.4	-16.0	55.9
DNA13PM	13	-120.6	1.9	-344.5	4.8	-17.9	0.6	67.5	0.4	-100.0	-276.4	-17.6	57.9
DNA14PM	14	-131.4	1.4	-378.8	4.1	-18.5	0.4	73.6	0.1	-110.8	-306.8	-19.3	60.1
DNA15PM	15	-134.0	4.3	-387.0	12.4	-18.6	1.1	73.2	0.0	-114.2	-314.9	-20.3	62.2
DNA16PM	16	-144.7	4.5	-418.5	13.1	-19.9	1.1	72.5	0.3	-124.2	-343.0	-21.9	63.8
DNA17PM	17	-144.8	7.2	-413.7	20.7	-21.4	1.9	76.8	0.1	-132.2	-363.1	-24.0	67.0
DNA18PM	18	-163.4	4.6	-466.7	13.3	-24.2	1.2	77.1	0.3	-142.9	-393.5	-25.6	67.8
DNA19PM	19	-164.3	9.5	-466.6	27.2	-25.2	2.4	79.3	0.3	-148.8	-408.1	-27.1	69.9
DNA20PM	20	-178.1	4.4	-506.1	12.6	-27.1	1.1	78.3	0.5	-159.6	-438.5	-28.9	70.7
DNA23PM	23	-199.3	8.2	-560.8	23.3	-32.1	2.2	82.4	0.1	-180.9	-497.0	-32.7	73.0
DNA25PM	25	-226.2	4.6	-637.0	12.4	-36.3	1.3	82.0	0.1	-200.2	-547.0	-37.1	76.5
DNA27PM	27	-250.9	6.7	-708.8	17.9	-39.5	1.9	84.0	0.2	-213.4	-583.0	-39.6	77.5
DNA30PM	30	-265.9	7.5	-743.4	28.0	-44.2	2.0	84.5	0.5	-239.2	-654.2	-44.2	78.7
DNA35PM	35	-279.0	8.0	-772.3	22.3	-48.7	2.1	86.0	0.1	-279.7	-765.0	-51.6	80.6

DSC - 300 mM [Na <sup>+</sup> ]										Mfold			
ID	N	$\Delta H$ kcal/mol	$\sigma$	$\Delta S$ e.u.	$\sigma$	$\Delta G_{25}$ kcal/mol	$\sigma$	$T_m$ °C	$\sigma$	$\Delta H$ kcal/mol	$\Delta S$ e.u.	$\Delta G_{25}$ kcal/mol	$T_m$ °C
DNA6PM	6	-55.2	1.4	-180.8	3.4	-1.3	0.4	24.7	1.2	-42.4	-123.4	-5.6	11.4
DNA8PM	8	-61.0	2.7	-194.8	8.3	-3.0	0.3	38.7	1.25	-53.4	-152.4	-8.0	26.8
DNA9PM	9	-80.0	5.8	-250.1	18.5	-5.4	0.3	44.7	0.1	-63.0	-181.5	-8.9	31.0
DNA10PM	10	-99.8	9.2	-301.5	27.7	-9.9	1.0	55.4	0.2	-72.8	-206.3	-11.3	40.7
DNA11PM	11	-101.6	8.9	-307.2	26.5	-10.0	1.1	61.7	0.2	-83.4	-233.9	-13.7	48.2
DNA12PM	12	-104.4	2.8	-311.3	8.3	-11.6	0.3	62.5	0.3	-89.5	-249.2	-15.2	52.5
DNA13PM	13	-110.1	3.9	-324.5	9.2	-13.4	1.1	66.0	0.2	-100.0	-279.5	-16.7	54.6
DNA14PM	14	-125.2	6.7	-364.5	18.1	-16.5	1.3	68.4	0.2	-110.8	-310.2	-18.3	56.8
DNA15PM	15	-129.9	8.3	-378.0	24.0	-17.2	1.1	70.7	0.2	-114.2	-318.4	-19.3	58.8
DNA16PM	16	-137.6	9.4	-403.2	27.3	-17.3	1.3	69.7	0.0	-124.2	-346.8	-20.8	60.3
DNA17PM	17	-151.7	4.6	-436.8	13.1	-21.5	0.7	70.1	0.1	-132.2	-367.1	-22.7	63.4
DNA18PM	18	-158.2	8.4	-453.8	24.3	-22.9	1.2	74.6	0.0	-142.9	-397.9	-24.3	64.3
DNA19PM	19	-164.3	8.5	-471.3	24.3	-23.8	1.2	75.4	0.0	-148.8	-412.7	-25.8	66.3
DNA20PM	20	-176.6	5.4	-504.3	15.4	-26.2	0.8	77.0	0.1	-159.6	-443.4	-27.4	67.2
DNA23PM	23	-197.9	1.8	-557.8	5.1	-31.6	0.3	78.5	0.3	-180.9	-502.6	-31.1	69.3
DNA25PM	25	-222.5	9.2	-629.8	25.9	-34.7	1.5	79.9	0.1	-200.2	-553.1	-35.3	72.8
DNA27PM	27	-251.4	4.2	-714.0	12.0	-38.5	0.7	79.8	0.5	-213.4	-589.6	-37.6	73.7
DNA30PM	30	-265.9	6.1	-746.0	15.6	-43.5	1.4	81.2	0.3	-239.2	-661.6	-42.0	74.9
DNA35PM	35	-274.5	9.7	-771.3	35.0	-44.5	0.7	84.0	0.4	-279.7	-773.6	-49.0	76.8

DSC - 85 mM [Na <sup>+</sup> ]										Mfold			
ID	N	$\Delta H$ kcal/mol	$\sigma$	$\Delta S$ e.u.	$\sigma$	$\Delta G_{25}$ kcal/mol	$\sigma$	$T_m$ °C	$\sigma$	$\Delta H$ kcal/mol	$\Delta S$ e.u.	$\Delta G_{25}$ kcal/mol	$T_m$ °C
DNA6PM	6	-40.8	1.1	-134.3	4.2	-0.8	0.2	22.2	1.3	-42.4	-125.7	-4.9	7.0
DNA8PM	8	-63.2	3.9	-208.8	8.1	-0.9	1.5	30.8	0.85	-53.4	-155.7	-7.0	21.4
DNA9PM	9	-79.8	5.0	-251.0	15.4	-4.9	0.5	40.8	0.5	-63.0	-185.2	-7.8	25.7
DNA10PM	10	-79.2	2.6	-240.8	5.3	-7.4	1.1	47.9	0.5	-72.8	-210.5	-10.0	35.2
DNA11PM	11	-102.3	0.3	-311.0	4.3	-9.6	1.0	55.3	0.1	-83.4	-238.6	-12.3	42.5
DNA12PM	12	-88.8	2.8	-267.3	5.8	-9.1	1.1	57.3	0.1	-89.5	-254.3	-13.7	46.6
DNA13PM	13	-91.7	1.9	-274.5	4.9	-9.9	0.5	61.3	0.2	-100.0	-285.1	-15.0	48.7
DNA14PM	14	-110.0	1.5	-322.8	4.2	-13.8	0.3	63.0	0.1	-110.8	-316.2	-16.5	51.0
DNA15PM	15	-115.2	4.7	-337.3	12.5	-14.6	1.0	63.7	0.2	-114.2	-324.9	-17.3	52.6
DNA16PM	16	-118.3	2.9	-348.6	14.2	-14.4	1.3	66.2	1.1	-124.2	-353.8	-18.7	54.2
DNA17PM	17	-132.4	4.8	-393.1	13.8	-15.2	0.8	67.3	0.1	-132.2	-374.5	-20.5	57.2
DNA18PM	18	-139.0	2.6	-406.4	11.9	-17.9	1.0	67.5	0.3	-142.9	-405.7	-21.9	58.1
DNA19PM	19	-123.9	0.6	-344.5	0.6	-21.2	0.4	65.8	0.0	-148.8	-421.0	-23.3	60.0
DNA20PM	20	-143.7	3.9	-402.8	11.0	-23.6	0.7	73.5	0.3	-159.6	-452.2	-24.8	60.9
DNA23PM	23	-159.6	2.3	-451.3	2.4	-25.0	1.6	73.9	0.0	-180.9	-512.8	-28.0	62.9
DNA25PM	25	-221.0	4.6	-639.0	10.7	-30.5	1.5	74.2	0.2	-200.2	-564.2	-32.0	66.3
DNA27PM	27	-249.3	7.0	-717.0	25.5	-35.5	0.7	74.9	0.3	-213.4	-601.7	-34.0	67.0
DNA30PM	30	-255.8	2.4	-735.0	11.9	-36.7	1.2	74.9	0.2	-239.2	-675.0	-38.0	68.3
DNA35PM	35	-271.7	9.9	-771.0	31.2	-41.8	0.8	76.1	0.3	-279.7	-789.4	-44.3	70.0

### 3.2.2 DSC Melting Data Error Propagation

Experimentally determined thermodynamic parameters, for the 19 short (6 to 35 base pairs) DNA duplexes, evaluated at four different [Na<sup>+</sup>], namely 85, 300, 600 and 1000 mM are summarized in Table 3.2. Reported  $\Delta H^{cal}$ ,  $\Delta S^{cal}$  and  $T_m$

values are averages of at least eight forward and reverse scans of the excess heat capacity versus temperature. Experimental errors on  $\Delta H^{cal}$ ,  $\Delta S^{cal}$  and  $T_m$  reported in Table 3.2 are the standard deviations from at least two independent melting experiments. Errors from  $\Delta H^{cal}$  and  $\Delta S^{cal}$  were propagated to  $\Delta G_{25^\circ C}^{cal}$  and determined using established methods (65). That is,

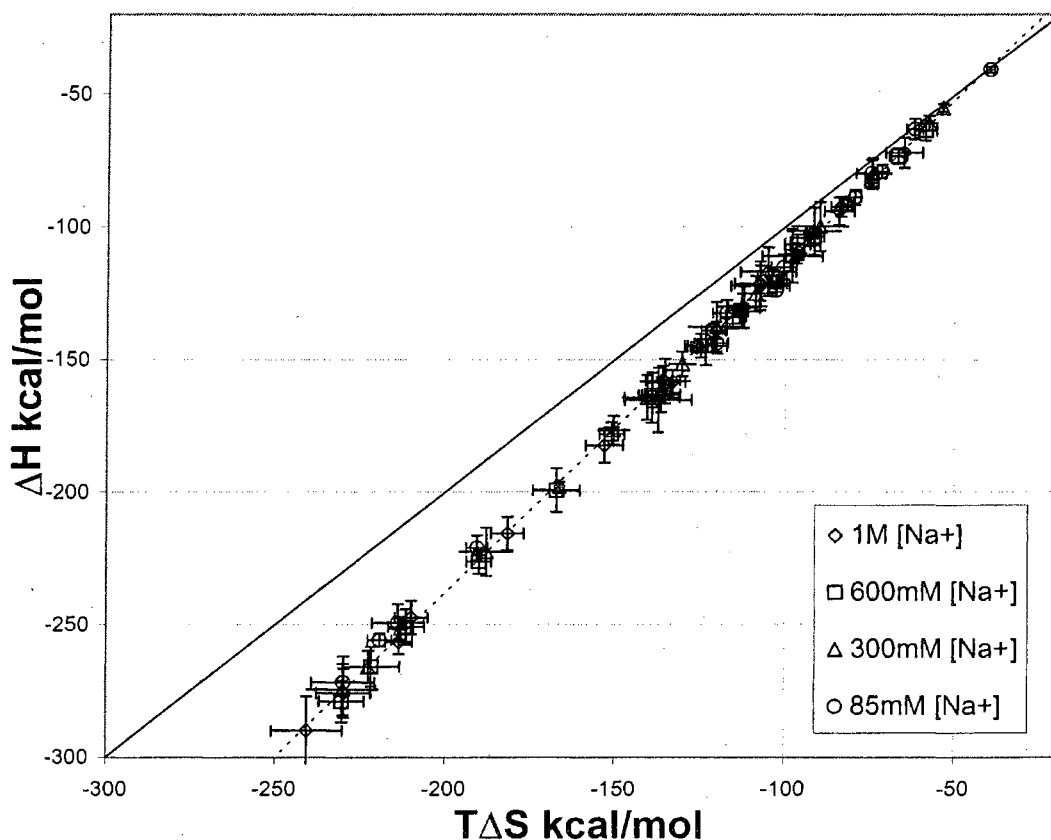
$$\left( \sigma_{\Delta G_{25^\circ C}^{cal}} \right)^2 = \left( \sigma_{\Delta H^{cal}} \right)^2 + \left( \sigma_{\Delta S^{cal}} \right)^2 - 2T(R_{\Delta H^{cal} \Delta S^{cal}}) \sigma_{\Delta H^{cal}} \sigma_{\Delta S^{cal}} \quad (3.3)$$

where the  $\sigma$ 's are the standard deviations in the measurements of at least eight melting curve experiments,  $\sigma_{\Delta H^{cal}}$  and  $\sigma_{\Delta S^{cal}}$  are the standard deviations in the DSC calorimetric enthalpy and entropy changes, respectively. The term  $R_{\Delta H^{cal} \Delta S^{cal}}$  is the correlation coefficient for  $\Delta H^{cal}$  versus  $T\Delta S^{cal}$ .

As shown in Figure 3.3, the plot of  $\Delta H^{cal}$  versus  $T\Delta S^{cal}$  for all  $[\text{Na}^+]$  was fitted by a linear regression function, with mean correlation coefficient,  $R^2 = 0.997$ . Several groups have reported either a similar linear or rectangular hyperbolic relationship between enthalpy and entropy (67-70,83,104). Favorable (i.e. negative  $\Delta G_{25^\circ C}^{cal}$ ) values demonstrate the enthalpy and entropy terms were almost equal in magnitude for all DNA duplexes studied, although more favorable enthalpic factors were able to offset unfavorable dissociation entropies. Overall, data reveal that compensating increases in enthalpic stabilization were larger than increases in

entropic destabilization, resulting in favorable  $\Delta G_{25^\circ C}^{cal}$  values. Similar analysis was performed for  $\Delta H^{cal}$ ,  $\Delta S^{cal}$  and  $\Delta G_{25^\circ C}^{cal}$  versus  $T_m$  (results not shown). Results also support the assumption that for these molecules,  $\Delta C_p^{ex} = 0$ . However results in the lowest  $[Na^+]$  show a slightly stronger dependence of  $\Delta H^{cal}$ ,  $\Delta S^{cal}$  and  $\Delta G_{25^\circ C}^{cal}$  on melting temperature. Inspection of thermodynamic data presented in Tables 3.2 reveals an increase in  $\Delta H^{cal}$  and  $\Delta S^{cal}$  with increasing  $[Na^+]$ .





**Figure 3.3:** Correlation between  $\Delta H^{cal}$  and  $T\Delta S^{cal}$  ( $T = 298.15$  K) for DNA duplexes with 6 to 35 base pairs in four  $\text{Na}^+$  environments. Error bars denote experimental reproducibility between 8 measurements.

As shown in Figure 3.3, the plot of  $\Delta H^{cal}$  versus  $T\Delta S^{cal}$  for all  $[\text{Na}^+]$  was fit to a straight line, with mean correlation coefficient,  $R^2 = 0.997$ . Evaluated negative and favorable  $\Delta G_{25^\circ\text{C}}^{cal}$  values demonstrated that the enthalpy and entropy terms were almost equal in magnitude for all DNA duplexes studied, although more favorable enthalpic factors were able to offset unfavorable dissociation enthalpies. Overall, data reveal that compensating increases in enthalpic stabilization were larger than entropic destabilization, resulting in favorable  $\Delta G_{25^\circ\text{C}}^{cal}$  values. Several

different groups have reported either a rectangular hyperbola or straight relationship between enthalpy and entropy plots (59,67-70).

Inspection of thermodynamic data presented in Table 3.2 reveals that DNA in high  $[\text{Na}^+]$  exhibited the highest magnitude of enthalpic stability,  $\Delta H^{cal}$  and entropic destabilization,  $\Delta S^{cal}$ . Clearly, the experimentally derived  $\Delta G_{25^\circ\text{C}}^{cal}$  values vary significantly from one salt concentration to another.

Thermodynamic parameters per base pair were deduced for DNA duplexes in all salt environments. Table 3.3 summarizes the mean thermodynamic parameters per base pair (bp). Values of thermodynamic parameters per base pair determined in each  $[\text{Na}^+]$  from the data in Tables 3.1 and 3.2 are summarized in Table 3.3. These values are in reasonable agreement with those reported by Fritz (43). For the formation of (dG-dC) oligoduplexes, in 1.0 M NaCl values of  $\Delta G^\circ/bp$ ,  $\Delta H^\circ/bp$  and  $\Delta S^\circ/bp$  were found to be -2.2 kcal/mol,  $-11.1 \pm 0.5$  kcal/mol and  $-27.9 \pm 1.0$  e.u. at 25 °C, respectively. The free energy for the formation of the first base pair (duplex nucleation) was reported to be  $2.22 \pm 0.1$  kcal/mol (43).

**Table 3.3:** Mean thermodynamic transition parameters per base pair for duplexes in all  $[\text{Na}^+]$  concentrations.

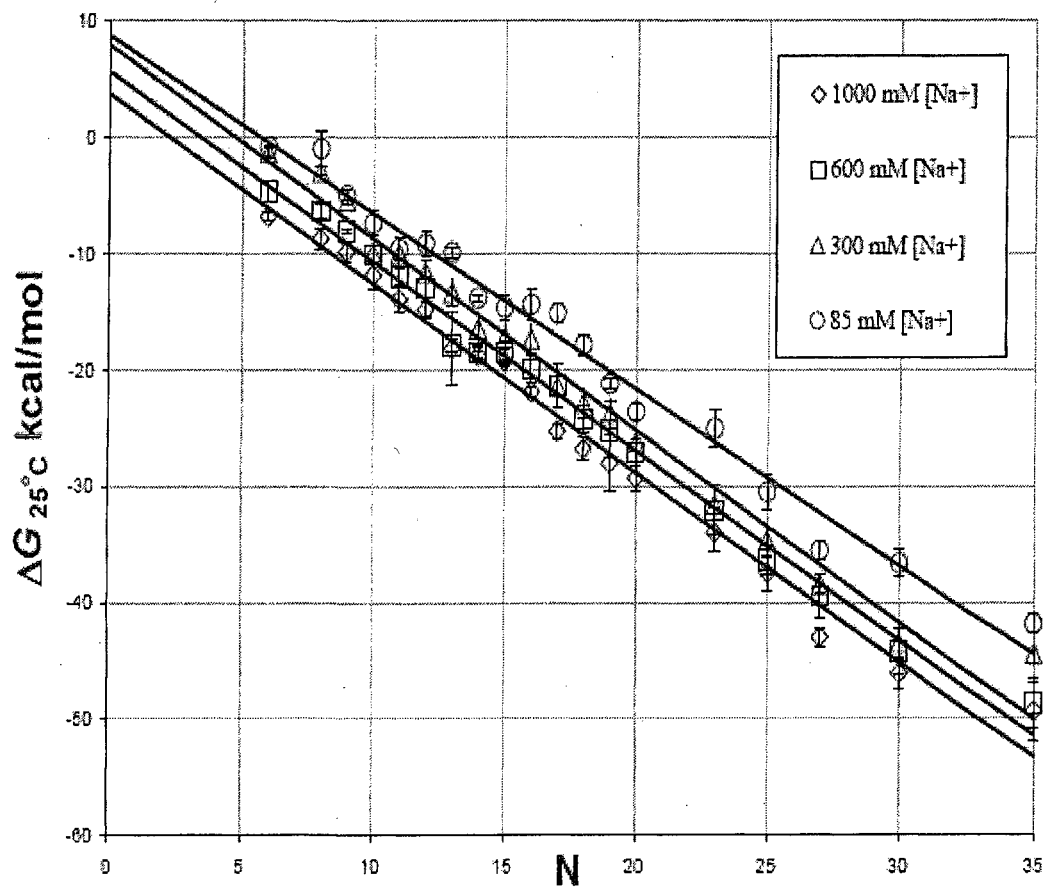
$[\text{Na}^+]$ mM	$\Delta H$ (kcal/mol) per base pair	$\Delta S$ (e.u.) per base pair	$\Delta G_{25}$ (kcal/mol) per base pair
1000	$-9.98 \pm 1.02$	$-28.90 \pm 3.84$	$-1.37 \pm 0.17$
600	$-9.16 \pm 0.62$	$-26.60 \pm 2.63$	$-1.22 \pm 0.22$
300	$-8.75 \pm 0.54$	$-25.76 \pm 2.06$	$-1.07 \pm 0.34$
85	$-7.83 \pm 0.80$	$-23.50 \pm 2.72$	$-0.90 \pm 0.34$

### 3.2.3 Free-Energy versus Duplex Length, N

Values of  $\Delta G_{25^\circ\text{C}}^{\text{cal}}$  are plotted versus duplex length, N, in Figure 3.4. Linear fits of the plots of  $\Delta G_{25^\circ\text{C}}^{\text{cal}}$  versus N, extrapolated to  $N = 0$ , yielded values for intercepts, henceforth referred to as  $\Delta G_{25^\circ\text{C}}^{\text{cal}}(N=0)$ , slopes and correlation coefficients. These parameters are summarized in Table 3.4. Values of the

extrapolated intercepts are interpreted as the free energy for the ‘hypothetical duplex’ having zero base pairs, but occupying the same volume with the same amount of displaced solvent as a duplex with base pairs intact. Values for  $\Delta G_{25^{\circ}\text{C}}^{\text{cal}}(\text{N} = 0)$  were found to be 3.69, 5.59, 7.86 and 8.68 kcal/mol in 1000, 600, 300 and 85 mM  $[\text{Na}^+]$ , respectively. The mean slope describing all data points was  $1.61 \pm 0.06$ . The average correlation coefficient for all linear fits was  $0.986 \pm 0.01$ .

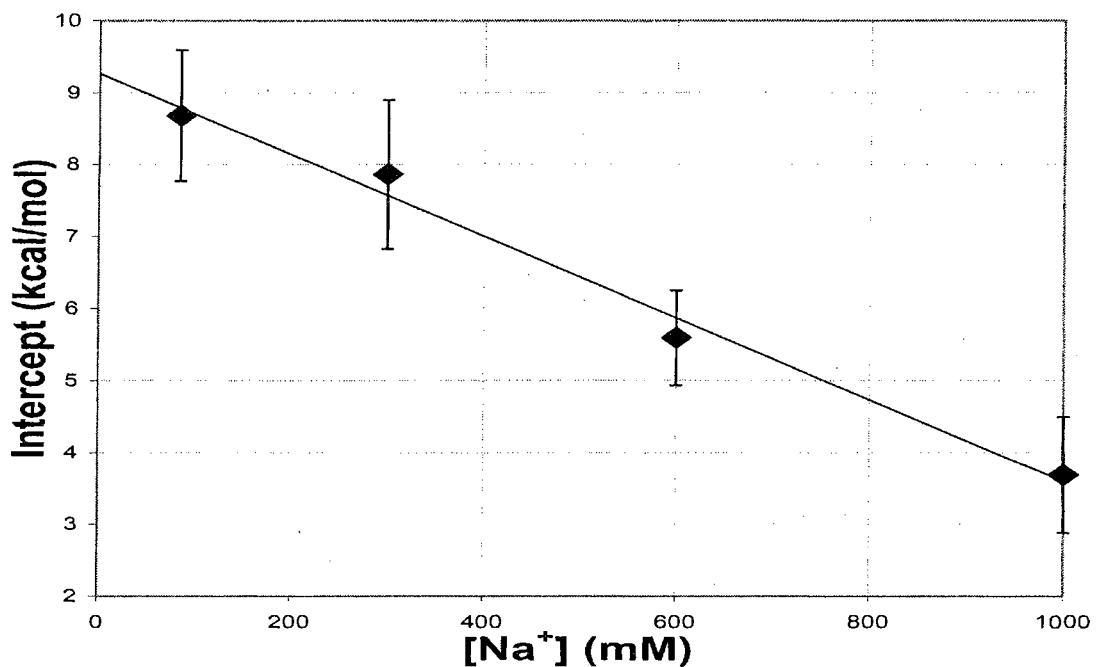
Linear extrapolations of  $\Delta G_{25^{\circ}\text{C}}^{\text{cal}}$  versus N plots reveal several interesting results. First, the values of  $\Delta G_{25^{\circ}\text{C}}^{\text{cal}}(\text{N} = 0)$  are always positive. Second, slopes of the plots are very similar in all  $[\text{Na}^+]$ ; in fact, the higher the  $[\text{Na}^+]$ , the lower (less positive) the value of  $\Delta G_{25^{\circ}\text{C}}^{\text{cal}}(\text{N} = 0)$ . Values of  $\Delta G_{25^{\circ}\text{C}}^{\text{cal}}(\text{N} = 0)$  are plotted versus  $[\text{Na}^+]$  in Figure 3.4 and reveal a linear relationship between the variables.



**Figure 3.4:** Plots of Measured  $\Delta G_{25^\circ C}^{cal}$  versus N for DNA molecules ranging in size from 6 to 36 base pairs in four buffered  $[Na^+]$  environments. Error bars are standard deviations,  $\sigma$ , calculated using Equation (3.1) and indicate experimental reproducibility.

**Table 3.4:** Summary of the linear fit parameters for the plots of  $\Delta G_{25^{\circ}C}^{cal}$  versus N as a function of  $[Na^+]$ . Slopes, intercepts and regression correlation coefficients are given for the linear equations of the individual fits.

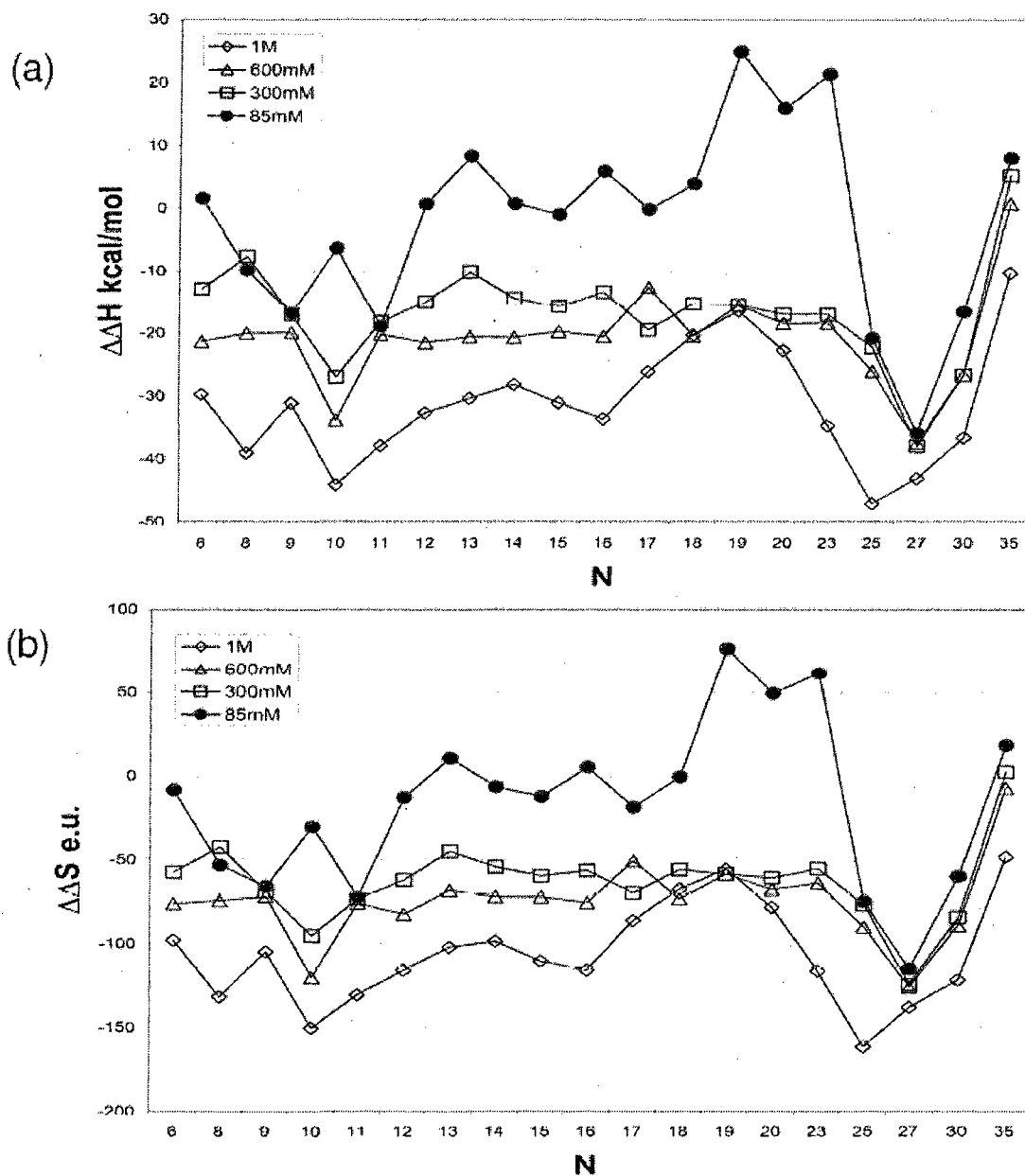
$[Na^+]$ mM	Slope	Intercept	$R^2$
1000	-1.63	3.69	0.988
600	-1.63	5.59	0.992
300	-1.65	7.86	0.98
85	-1.52	8.68	0.982
Average in all $[Na^+]$	-1.61	6.46	0.983



**Figure 3.5:** Plot of free-energies at the intercepts obtained from the unconstrained fit and summarized in Table 4. The linear fit gives Intercept =  $-5.70 \times 10^{-3}[\text{Na}^+] + 9.2677$ ,  $R^2 = 0.988$

### 3.2.4 Validity of the n-n model

A direct comparison of the thermodynamic transition parameters measured by DSC versus values calculated using the MFOLD n-n routine was made. Parameters predicted and measured were compared for all molecules in the four  $[\text{Na}^+]$  environments. Plots of  $\Delta\Delta X^{cal}$  versus N were constructed where  $X = H, S$  and  $G$ , and  $\Delta\Delta X^{cal} = (\Delta X^{cal} - \Delta X^{MFOLD})$ . These plots are shown in Figures 3.6 (a) and (b).



**Figure 3.6:** Comparative plots of the experimental enthalpy and entropy difference from MFOLD™. (a) Plot of  $\Delta\Delta H^{cal}$  ( $\Delta H^{cal} - \Delta H^{MFOLD}$ ) versus N. (b) Plot of  $\Delta\Delta S^{cal}$  ( $\Delta S^{cal} - \Delta S^{MFOLD}$ ) versus N.

Generally, the experimental values,  $\Delta H^{cal}$  and  $\Delta S^{cal}$ , determined in this study were systematically higher than predicted by MFOLD™ values. Even though predicted



and experimental values of the enthalpies and entropies were not in agreement, much better agreement was obtained in 85 mM  $[\text{Na}^+]$ . Plots of  $\Delta\Delta G^{cal}$  versus N were also prepared as shown in Figure 3.7. Generally, values of  $\Delta G^{\circ}$  predicted by MFOLD vary significantly over a larger range than experimentally observed. As seen in Figure 3.7, differences in  $\Delta\Delta G^{cal}$  values fall within a range of  $\pm 3.0$  kcal/mol for 600 and 1000 mM  $[\text{Na}^+]$ . In contrast, poor agreements between predicted and measured values were obtained in 85 and 300 mM  $[\text{Na}^+]$ . The greater differences between  $\Delta\Delta G^{cal}$  and predictions for DNA molecules in 85 mM  $\text{Na}^+$  underscore the need for improve parameters in low salt ( $< 100$  Mm).

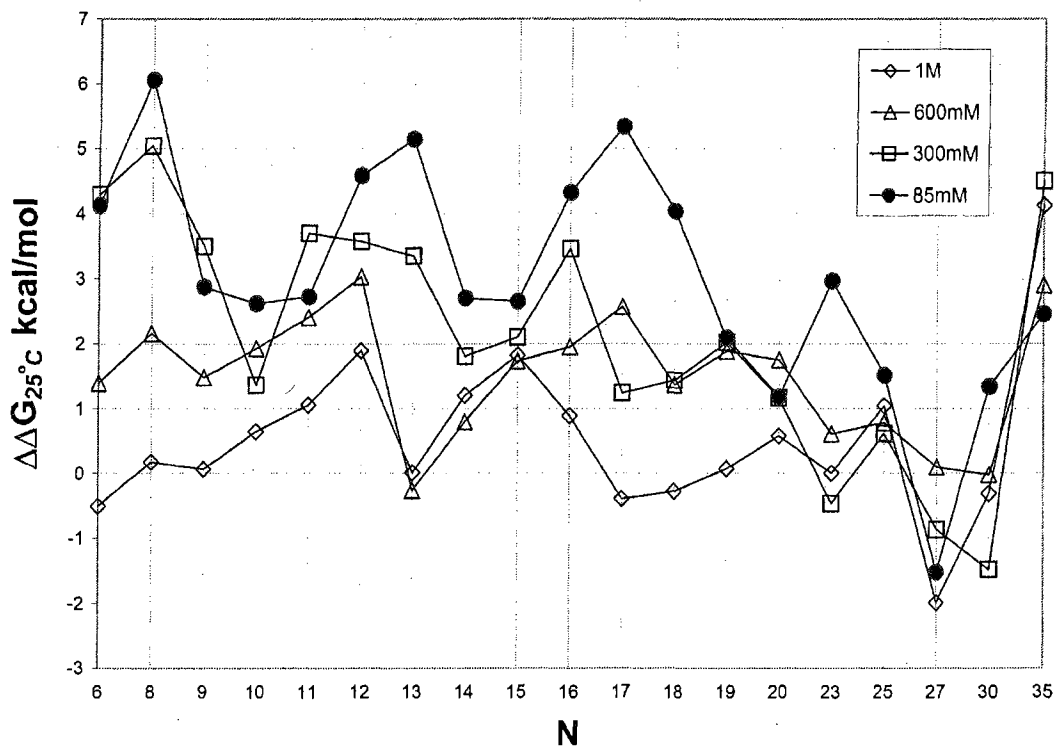
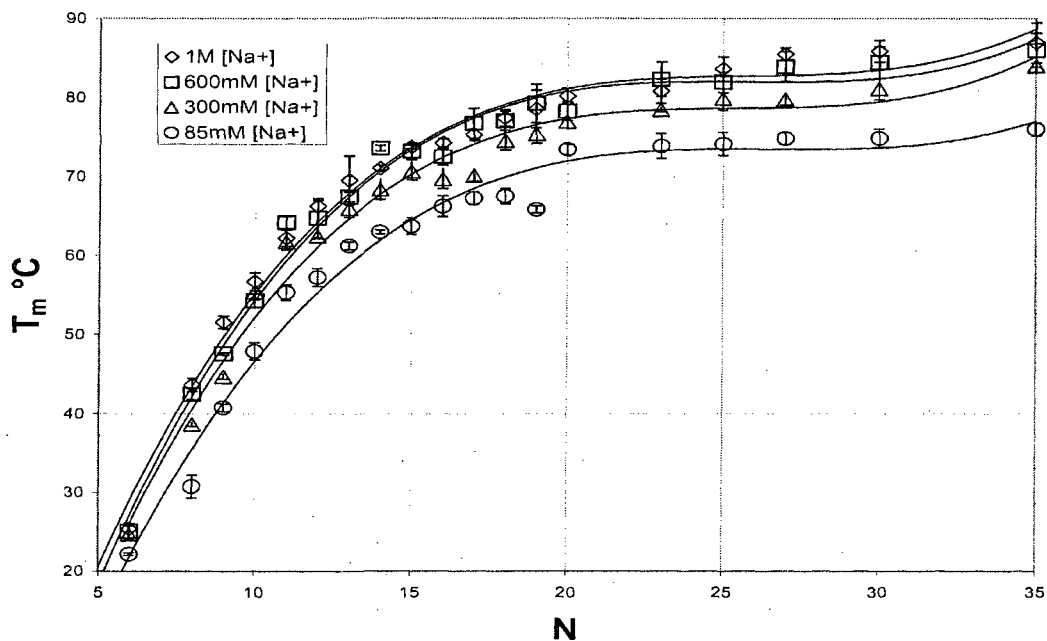


Figure 3.7: Comparative plots of calorimetric  $\Delta\Delta G^{cal}$  ( $\Delta G_{25^{\circ}C}^{cal} - \Delta G^{MFOLD}$ ) versus N in all  $[Na^+]$  environments.

### 3.2.5 Analysis of melting temperature ( $T_m$ )

Behaviors of DNA melting temperature are an much studied property and several empirical correlations have been suggested by different authors (6,11,56-60,66,70). Measured  $T_m$  values for each DNA duplex, obtained from DSC and predicted using MFOLD™. Recall that the  $T_m$  of these transitions were determined from the temperature at peak height maximum of the normalized,  $\Delta C_p^{ex}$ , versus temperature profiles. These  $T_m$  values were measured for DNA samples at strand concentrations ranging from 0.424 - 1.56 mg/mL. Results indicate that  $T_m$  was

essentially over a less than four fold range in strand concentrations. The plot of  $T_m$  versus  $N$  is shown in Figure 3.8 which clearly shows that the DNAs are increasingly destabilized with decreases in  $[\text{Na}^+]$  and  $T_m$ .



**Figure 3.8:** Plot of  $T_m$  versus  $N$  for the 19 DNA duplexes of this study. A quadratic polynomial was fitted to each set of data points.

Several general observations can be made. (i) Values for  $T_m$  increased with increasing DNA length,  $N$ , from 6 to 35 base pairs in regardless of the  $[\text{Na}^+]$  concentration;  $T_m$  seem to level off at higher  $N$  ( $> 20$  base pairs). Recent results obtained by Chen *et al.* (71) revealed that at low  $[\text{Na}^+]$  and  $N$ , binding thermodynamic parameters for short DNAs were very weak and phosphate-phosphate repulsive forces dominate the electrostatic interactions. In addition, this repulsive, destabilizing effect was stronger for double stranded than for single

stranded DNAs. In our case, the duplexes that were studied were more enthalpically stable in 1000 mM  $[\text{Na}^+]$  than at lower salt concentration, but any further increase in ionic strength concentration could have resulted in of helix destabilization. In addition, the electronic component of the free energy was weakly dependent on both the electrostatics and N. (ii) There was destabilization of the DNA duplexes in low salt, as indicated by their significantly lower  $T_m$  values in 85 mM  $[\text{Na}^+]$ . This destabilization of transition parameters in low salt ( $< 100$  mM  $[\text{Na}^+]$ ) is consistent with findings of Chen *et al.* (71,75) who attributed it to a large decrease in entropy, weak charge neutralization and fewer bound ions in lower ionic strength environments.

The plot in Figure 3.9 shows results of a direct comparison between the DSC measured  $T_m$  and values predicted using MFOLD™. Generally, predictions and measured  $T_m$  values differ significantly. However, as N increases, better agreement was observed between measured and predicted values. These findings indicate some shortcomings of the current predictive algorithms, where most experimental measurements for helix-coil equilibrium are evaluated in 1000 mM  $[\text{Na}^+]$ . Further, most of the experiments reported in the n-n model have been done using the UV visible spectrophotometer which assumes a two state model (8-11,22-24,31,48,56,60).

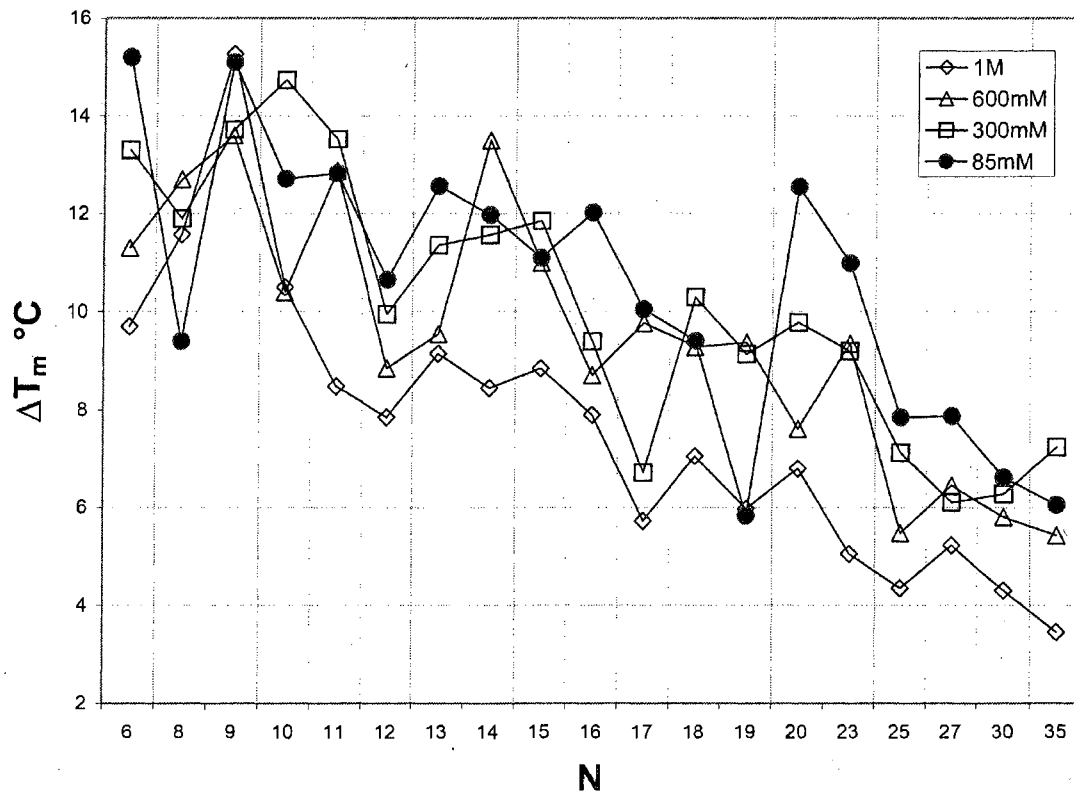


Figure 3.9: Plot of the direct comparison between predicted and measured melting temperature i.e.

$$\Delta T_m = T_m^{\text{DSC}} - T_m^{\text{MFOLD}}$$

### 3.2.6 Dependence of the melting transitions on $[\text{Na}^+]$ and evaluation of the counterion release upon DNA melting, $\Delta n$

Analysis of melting transition parameters evaluated as a function of  $[\text{Na}^+]$  provided a means of quantitatively estimating the counterion release/association upon melting for the short DNAs as a function of length. According to the Equation (2.31), the plot of  $T_m^{-1}$  ( $K^{-1}$ ) versus  $\ln[\text{Na}^+]$  for each of the 19 duplex DNAs provided an estimate of the release of  $\text{Na}^+$  ions upon melting,  $\Delta n$ , as a function of duplex length. Average linear fits of  $T_m$  versus  $\ln[\text{Na}^+]$  and  $T_m^{-1}$  versus  $\ln[\text{Na}^+]$  generated from the melting transitions obtained for each DNA molecule were made (data not shown). Generally slopes (given by given by  $\left[ \frac{dT_m^{-1} (K^{-1})}{d\ln[\text{Na}^+]} \right]$ ) gradually increased with increasing length up to about 12 base pairs and then increase less with lengths greater than 12 base pairs. The observed linear dependence of  $T_m$  versus  $\ln[\text{Na}^+]$  and  $T_m^{-1} (K^{-1})$  versus  $\ln[\text{Na}^+]$  is best explained in Equation (2.31).

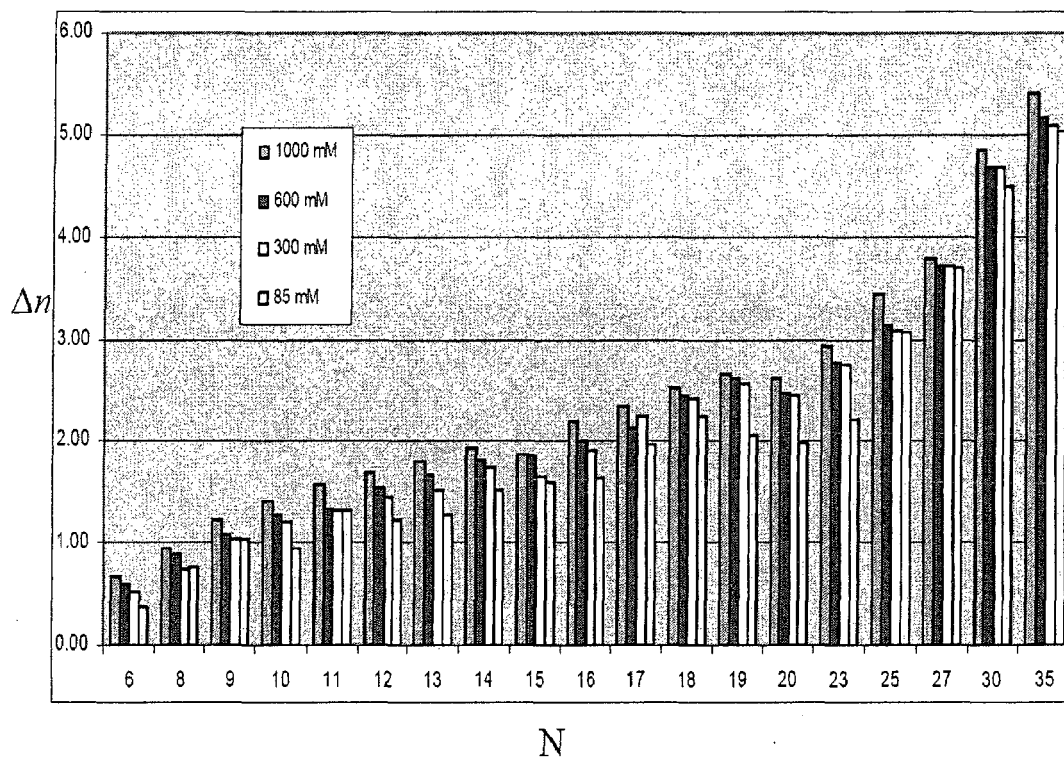
Similarly, estimates of the fractional number of counterion released per phosphate group,  $\Delta \psi$ , were also determined and summarized in Table 3.4 and Figure 3.11. The differences in  $\Delta n$ , evaluated for these DNAs reveal differential  $[\text{Na}^+]$  release as a function of length. Generally,  $\Delta n$  increases with (i) increasing length of oligomers and (ii) increasing  $\text{Na}^+$  from 6 - 35 base pairs. In some cases,

observed differences can be attributed to differences in % G-C which influences the both enthalpy and  $T_m$  from which  $\Delta n$  is evaluated. Greatest differences are observed at  $N > 20$  base pairs. Clear differences in  $\Delta n$ , evaluated for these sequences reveal differential  $[\text{Na}^+]$  release for these DNA molecules with increasing  $N$ .

**Table 3.5:** Release of  $[\text{Na}^+]$  upon DNA melting,  $\Delta n$  and differential cation release per phosphate group,  $\Delta\psi$  for 19 DNA molecules.

N	$\Delta n$	$\Delta n$	$\Delta n$	$\Delta n$	$\Delta\psi$	$\Delta\psi$	$\Delta\psi$	$\Delta\psi$
	1000 mM $[\text{Na}^+]$	600 mM $[\text{Na}^+]$	300 mM $[\text{Na}^+]$	85 mM $[\text{Na}^+]$	1000 mM $[\text{Na}^+]$	600 mM $[\text{Na}^+]$	300 mM $[\text{Na}^+]$	85 mM $[\text{Na}^+]$
6	0.67	0.59	0.51	0.38	0.11	0.10	0.09	0.06
8	0.94	0.88	0.73	0.76	0.12	0.11	0.09	0.10
9	1.22	1.07	1.04	1.03	0.14	0.12	0.12	0.11
10	1.41	1.28	1.20	0.95	0.14	0.13	0.12	0.10
11	1.57	1.34	1.32	1.33	0.14	0.12	0.12	0.12
12	1.70	1.54	1.45	1.23	0.14	0.13	0.12	0.10
13	1.81	1.67	1.53	1.27	0.14	0.13	0.12	0.10
14	1.93	1.83	1.74	1.53	0.14	0.13	0.12	0.11
15	1.87	1.86	1.66	1.60	0.12	0.12	0.11	0.11
16	2.19	2.01	1.91	1.64	0.14	0.13	0.12	0.10
17	2.35	2.14	2.25	1.96	0.14	0.13	0.13	0.12
18	2.52	2.45	2.41	2.25	0.14	0.14	0.13	0.13
19	2.65	2.61	2.56	2.07	0.14	0.14	0.13	0.11
20	2.62	2.47	2.45	2.00	0.13	0.12	0.12	0.10
23	2.93	2.77	2.75	2.22	0.13	0.12	0.12	0.10
25	3.44	3.14	3.09	3.07	0.14	0.13	0.12	0.12
27	3.79	3.72	3.72	3.69	0.14	0.14	0.14	0.14
30	4.85	4.68	4.68	4.50	0.16	0.16	0.16	0.15
35	5.40	5.17	5.08	5.03	0.15	0.15	0.15	0.14

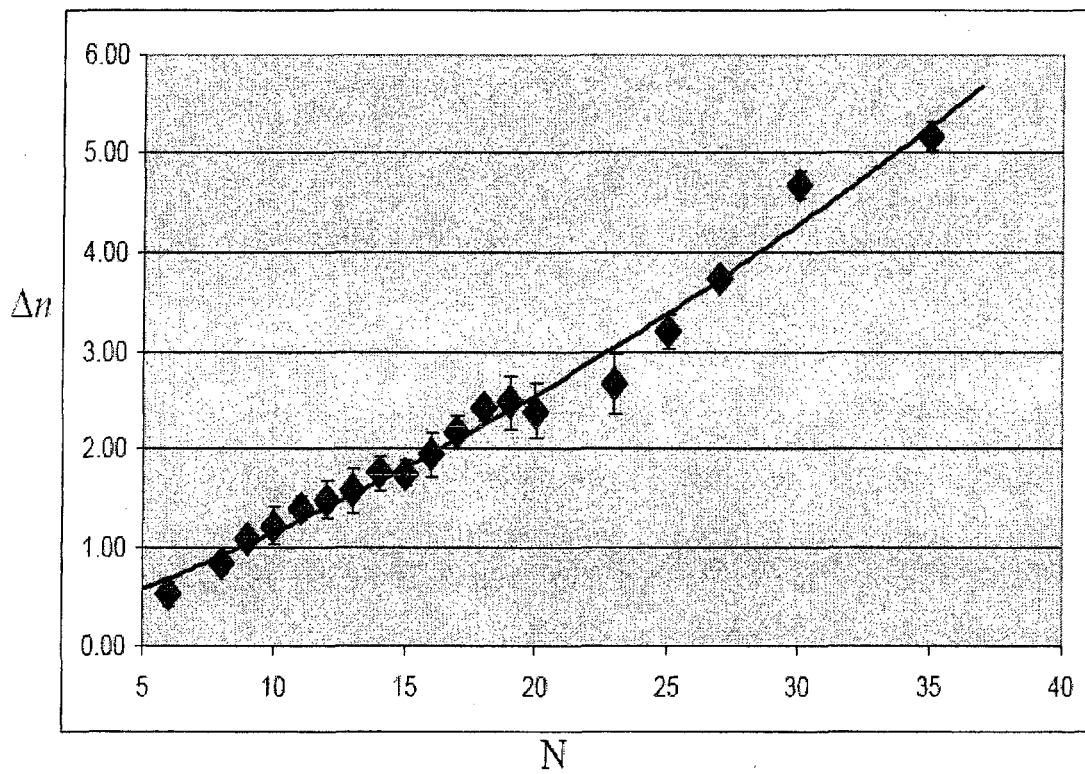
NB:  $\Delta n$  refers to the Release of  $[\text{Na}^+]$  upon DNA melting.  $\Delta\psi$  refers to the fractional number of counterions released per phosphate group.



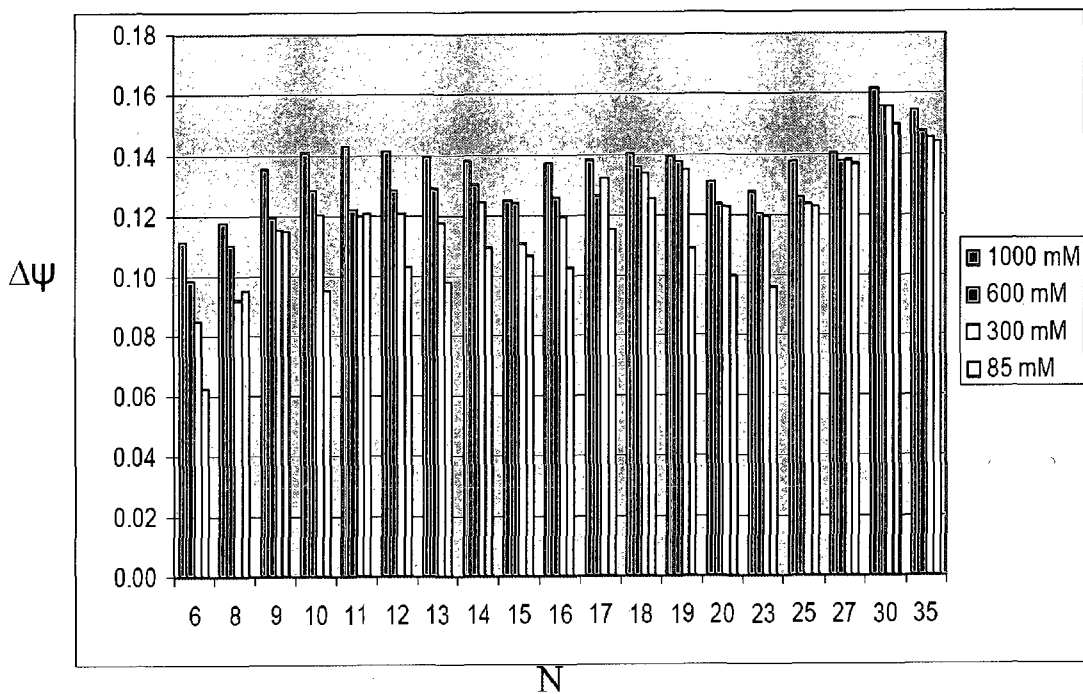
**Figure 3.10:** Histogram for  $\Delta n$  versus  $N$  DNA duplexes ranging in length from 6-35 base pairs.

However, values for  $\Delta\psi$  are essentially constant (Figure 3.12 and 3.13) indicating that except for perhaps the shortest DNAs, they behave in essentially the same manner as regards to their  $\text{Na}^+$  binding properties. These results validate our assumption regarding evaluations of the “nucleation volume” by extrapolation to zero length.

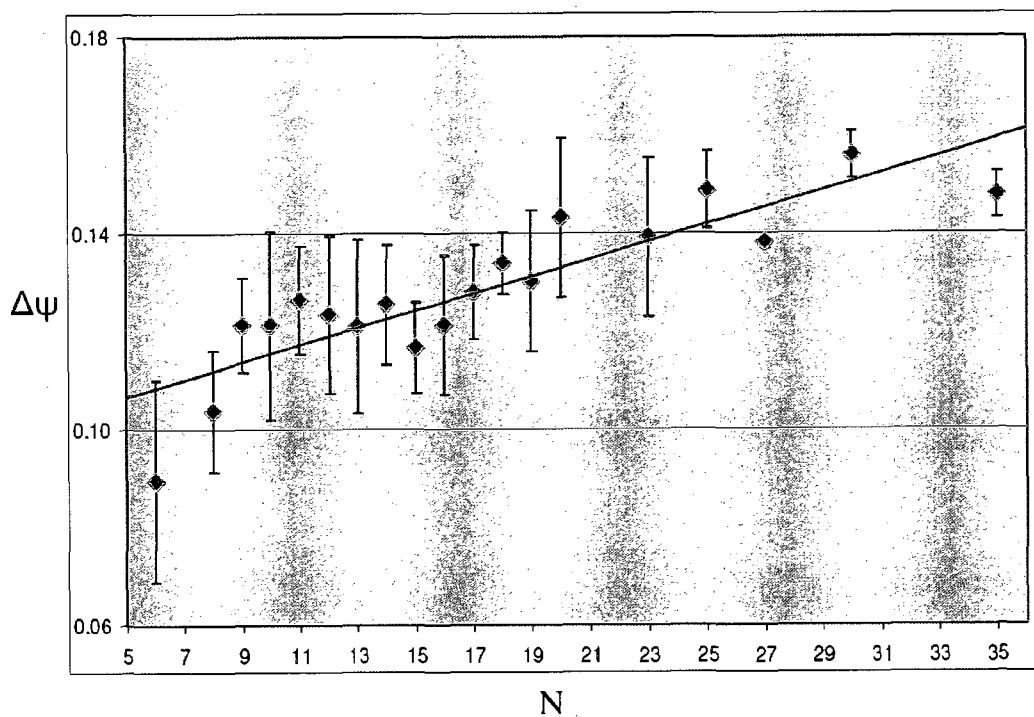




**Figure 3.11:** Plot of the average values of  $\Delta n$  versus  $N$  DNA duplexes ranging in length from 6-35 base pairs in all four  $[Na^+]$  environments.



**Figure 3.12:** Histogram for  $\Delta\psi$  versus  $N$  for the DNA duplexes ranging in length from 6-35 base pairs.



**Figure 3.13:** Plot of the average values of  $\Delta\psi$  versus  $N$  DNA duplexes ranging in length from 6-35 base pairs in all four  $[Na^+]$  environments.

### 3.3 DISCUSSION

The experimental design was aimed at evaluations of the standard state free energies for the 'hypothetical duplex' having zero base pairs, (i.e.  $\Delta G_{25^\circ C}^{cal}(N=0)$ ), in different  $[Na^+]$  environments. The nucleation free energy was presented without the partitioning of enthalpic and entropic components. As discussed by Amzel *et al.* (39), the nucleation process must surely have an unfavorable entropic cost due to the loss of translational and rotational freedom of the reacting single strand. The actual enthalpic and entropic components of the nucleation free energy is beyond the scope of this study.

Values of  $\Delta G_{25^\circ C}^{cal}(N=0)$  were obtained by extrapolation of the plots in Figure 3.4. These values, 3.69, 5.59, 7.86 and 8.68 kcal/mol are plotted versus  $[Na^+]$  in Figure 3.5. From Equation (2,24), if  $\Delta G_{25^\circ C}^{cal}(N=0)$  then  $K^o = \left( \frac{Z_{config}^{rel}(T)}{V_s 8\pi^2} \right)$  depends only on the configurational integrals for the single strands and the duplex in the same Euler volume. As stated earlier,  $K^o$  is the relative probability of two single strands in their standard states compared to the duplex in its standard state. The fact that  $\Delta G_{25^\circ C}^{cal}(N=0)$  increases with decreasing  $[Na^+]$  indicates that the value of the configuration integral  $\left( Z_{config}^{rel}(T) \right)$  is greater in higher  $[Na^+]$ . An electrostatic argument provides an obvious explanation. Recall from the discussion following Equation (2.27),

$$Z_{config}^{rel}(T) = \int_{Dm} \exp(-U_{S1.S2}^R/k_B T) dR \cdot \int_{Dm} \exp(-U_{S1.S2}^\Omega/k_B T) d\Omega \quad (3.4)$$

where  $Dm$  is the domain of the 6-D relative (Euler plus translational) coordinate space, within which the pair of neighboring strands is considered to be complexed. This ‘arbitrary’ domain is introduced because the complex is defined for predominantly repulsive forces (as is the case for  $N=0$  standard state of the duplex). The integration is restricted over the relative coordinates, defined that domain only, still with the same weighing that was described before.

Naturally, at lower  $[Na^+]$  greater electrostatic repulsion between phosphate groups would be expected, which in turn likely affects values of the potential energy arguments of both integrals. Apparently this relatively greater phosphate repulsion leads to larger values of  $U_{S1.S2}^R$  and  $U_{S1.S2}^\Omega$ , resulting in smaller values of  $Z_{config}^{rel}(T)$ . Overall, the consequence is greater values of  $\Delta G_{25^\circ C}^{cal}(N=0)$  with decreasing  $[Na^+]$ .

### 3.3.1 Double Helix Formation

The concept of duplex nucleation or initiation was introduced over 45 years ago (18,20-21). In 1963 Applequist and Damle reported results from theoretical statistical thermodynamic analysis of melting curves of oligo-dA DNA strands at low pH (18). From their analysis they proposed that short DNA double helix

formation occurs in two steps, (i.e. unfavorable) nucleation followed by (favorable) helix zippering. In their treatment both strands were assumed to be the same and the equilibrium constant for duplex formation was given by the product of two terms, ( i.e.  $K_{eq} = \beta s^N$  ), where  $\beta$  is the nucleation parameter or association factor, with units of volume per mole ( $M^{-1}$ ) and  $s^N$  is the statistical weight of the fully intact duplex, which is the product of the N base pair stability constants,  $s$ , one for each base pair. Since then values of the initiation parameter,  $\beta$ , and base pair stability constant have been shown to be critical for obtaining reasonable agreement with measurements or prediction of experimental results (18,43). It has been universally assumed the part of the equilibrium constant denoted by  $\beta$  corresponds to the more difficult formation of the first base pair from single strands compared to zippering formation of the remaining base pairs. In this manner, the  $\beta$  parameter corresponds to the free-energy of duplex nucleation,  $\Delta G = -k_B T \ln \beta$ .

From analysis of their melting curves of duplexes ranging from 8 to 11 base pairs, depending on whether staggered or non-staggered zipper models were employed, Applequist and Damle reported mean values of  $\beta$  ranging from  $2.20 \times 10^{-3}$  to  $7.50 \times 10^{-3}$ , corresponding to  $\Delta G_{nuc}$  values of 3.70 and 2.90 kcal/mol, respectively. Over a decade later, Fritz (43) reported formation of the first base pair in short (dG-dC) duplex DNA oligomers involved a nucleation parameter,  $\beta = 1.60 \times 10^{-4} M^{-1}$ , corresponding to  $\Delta G_{nuc} = 5.20$  kcal/mol at 25°C. The free-energy of base pairing is defined by the base pair stability constants,  $s = \exp(-\Delta G_{bp} / k_B T)$ ,

where,  $\Delta G_{bp}$ , is the free-energy per base pair, which in these cases was assumed to be the same for every base pair in the duplex. The total free energy of duplex formation is given by  $\Delta G^o = -k_B T \ln s^N$ . The nearest-neighbor parameters and recipes for calculating thermodynamic stability of short duplex DNA/DNAs, DNA/RNAs and RNA/RNAs from their sequences have been reported (4,8-11,22,43,56,64,66,70-71,73-76). For these studies,  $\Delta G_{nuc}$  was varied as a fitting parameter in evaluation of the nearest-neighbor parameters, and even assigned a sequence dependence (4,8-9,73-74,77). Values of  $\Delta G_{37^oC}^{nuc}$  ranging from 1.80 to 6.00 kcal/mol have been reported (4,8,11,22,43,66,73-74).

**Table 3.6:** Evaluated values for the helix initiation free energy DNA and RNA duplexes, obtained from different laboratories. The total salt environment was 1.0 M NaCl.

Research Group	DNA $\Delta G_{nuc}$ [A-T] kcal/mol	DNA $\Delta G_{nuc}^{37^\circ C}$ [G-C] kcal/mol	$\Delta G_{nuc}$
Breslauer <i>et al.</i> (1986)	6.23 (at 25°C)	5.21 (at 25°C)	-
Sugimoto <i>et al.</i> (1996)	2.79 (at 37°C)	1.83 (at 37°C)	-
SantaLucia <i>et al.</i> (1997)	1.03 (at 37°C)	0.98 (at 37°C)	-
Sugimoto <i>et al.</i> (1995)	3.39 (at 37°C)	2.7 - 2.90 (at 37°C)	-
Fritz ( 1973)	<b>RNA</b> 3.40 (at 37°C)	5.2 (at 25°C)	-
Applequest and Damle (1963)	3.70 (at 25°C)	2.39 (at 25°C)	-
Freier <i>et al.</i> (1986)	<b>RNA</b> 3.10 (at 37°C)	-	-
Manyanga <i>et al.</i> (2009)	-	-	6.46 ± 2.3

## CHAPTER 4

### MELTING STUDIES OF TANDEM MISMATCH BASE PAIRS IN SHORT DUPLEX DNA: INFLUENCE OF MISMATCH SIZE, LOCATION, AND SODIUM ION CONCENTRATION

#### 4.0 INTRODUCTION

Characterization of mismatches via their thermodynamic parameters of melting provides a foundation for analysis of effects of these structures in a variety of biological and biotechnological applications. An examination of the current literature reveals a renewed interest in the structural characterization of mismatches with focus on biological applications, especially regulation of gene expression (1-3,15,47). Published structural and experimental data suggests that in a manner similar to DNA binding ligands, tandem mismatches can cause significant structural perturbations in DNA helical structure (5,16,81). These perturbed structures have been shown to be minor variants of B-form or A-form DNA and may play a role in the regulation of gene expression (16,46,48,78).

In addition, mismatches are fundamental to any assay design process, especially multiplex hybridization and model simulations of the inherent reactions. Specifically, perturbations of the duplex DNA thermodynamic stability associated



with mismatches have been shown to be important in probe sequence design processes (50). The n-n theory has been expanded to include single base pair mismatches with some degree of accuracy (3-11). However, calculations of the thermodynamic stabilities of short DNA oligomers containing mismatches ignore the potential sequence-dependent contributions of two or more contiguous mismatches. Neither has the range of stabilities of tandem mismatches in different sequence contexts established. Thermodynamic studies of systematically designed mismatches provide deeper insight into various forces that govern relative affinity of DNA strands in hybridization reactions.

Previously, results obtained in the Benight laboratory have demonstrated that depending on the sequence, stabilities of some tandem mismatches can be from 30 to 70 % as stable as a W/C base pair doublet (1-2,15,89). In addition, explicit for thermodynamic stability parameters for a few of the tandem mismatch sequences have been shown to affect the accuracy of model simulations of multiplex hybridization reactions (2). To further characterize tandem mismatch states, calorimetrically derived thermodynamic transition parameters of 25 designed DNA molecules were evaluated by DSC in different buffered salt environments ranging from 85 to 1000 mM  $[\text{Na}^+]$ . The molecules were designed to systematically address the following questions; (i) How does the presence and distribution of tandem mismatches in a short duplex DNAs affect thermodynamic stability?; (ii) Is the nearest-neighbor model accurate in predicting measured thermodynamic stabilities of short duplexes containing contiguous mismatch base pairs?; (iii) How does the

stability of short duplex DNA containing contiguous tandem mismatches vary with  $[\text{Na}^+]$ ?

## **4.1 MATERIALS AND METHODS**

Details of the sample preparation, instrumentation, procedures for data acquisition, analysis routines for obtaining DSC melting curves for DNA have been described in Chapter 2. Only specific methods employed for this present study are presented here.

### **4.1.1 Rationale and Design of DNA Molecules**

Sequences of duplex DNA molecules that were analyzed by DSC melting are listed in Table 4.1(a) and (b). Two uniquely distinguishable sets of DNA molecules consisting of tandem mismatches in various configurations were designed and prepared for DSC melting analysis. As depicted in Tables 4.1(a) and (b), each set contained increasing numbers of tandem mismatch base pairs in the different sequence contexts. Set A: includes molecules, 20 base pairs in length with mismatches incrementally introduced contiguously from one 'end' of the duplex. Set B: includes molecules, 20 base pairs in length with mismatches incrementally introduced at different positions in the middle of the duplex and 'interspersed' amongst the remaining W/C base pairs. In table 4.1(a) and (b) and throughout this study, all mismatched bases in Sets A and B duplexes are indicated by bold lower

case letters that are underlined. Sequences of control molecules are shown in Table 4.1(c).

Depending on the degree of self-complementarity in their sequences, single strand DNA molecules have the potential to form cruciform, triplex, tetraplex and hairpin complexes (5,9-14,36,73). For this reason, single strands used to prepare the duplexes comprising Sets A and B were designed to discourage undesired secondary structure formation in them. During the design process of the sequences, the possibility of hairpin formation in the resident single strands was assessed by estimates of thermodynamic parameters for possible intramolecular structures that might form in their sequences. For these estimates the available parameters were employed (31,63-64). All the designed strands consisted of base composition varying from 40 and 60 % G-C. Any single strand suspected of forming a hairpin with predicted  $T_m$  above 50 °C or melting free energy lower than -5.0 kcal/mol was excluded and re-designed.

**Table 4.1(a):** The 20 base pair DNA Molecules comprising Sets A duplexes.

Sample ID	n	N	Set A DNA Sequences-20 base pairs
DNA-pm	0	20	5' -GTGACAGCACAATGGTGACG-3' 3' -CACTGTCGTGTTACCACTGC-5'
DNA1mm	1	19	5' - <u>CT</u> GACAGCACAATGGTGACG-3' 3' - <u>a</u> ACTGTCGTGTTACCACTGC-5'
DNA2mm	2	18	5' - <u>ct</u> GACAGCACAATGGTGACG-3' 3' - <u>ag</u> CTGTCGTGTTACCACTGC-5'
DNA3mm	3	17	5' - <u>cta</u> ACAGCACAATGGTGACG-3' 3' - <u>aga</u> TGTCGTGTTACCACTGC-5'
DNA4mm	4	16	5' - <u>ctag</u> CAGCACAATGGTGACG-3' 3' - <u>agat</u> GTCGTGTTACCACTGC-5'
DNA5mm	5	15	5' - <u>ctagc</u> AGCACAATGGTGACG-3' 3' - <u>agatc</u> TCGTGTTACCACTGC-5'
DNA6mm	6	14	5' - <u>ctagct</u> GCACAATGGTGACG-3' 3' - <u>agatccc</u> GTGTTACCACTGC-5'
DNA7mm	7	13	5' - <u>ctagctt</u> CACAATGGTGACG-3' 3' - <u>agatcct</u> GTGTTACCACTGC-5'
DNA8mm	8	12	5' - <u>ctagctta</u> ACAATGGTGACG-3' 3' - <u>agatcctc</u> TGTTACCACTGC-5'
DNA9mm	9	11	5' - <u>ctagcttag</u> CAATGGTGACG-3' 3' - <u>agatcctct</u> GTTACCACTGC-5'
DNA10mm	10	10	5' - <u>ctagcttaga</u> AATGGTGACG-3' 3' - <u>agatcctcta</u> TACCACTGC-5'

**Table 4.1(b):** The 20 base pair DNA Molecules comprising Sets B.

Sample ID	u	N	Set B DNA Sequences–20 base pairs
DNA-pm	0	20	5' -GTGACAGCACAATGGTGACG-3' 3' -CACTGTCGTGTTACCACTGC-5'
DNA1mm	1	19	5' -GT <u>c</u> ACAGCACAATGGTGACG-3' 3' -CA <u>a</u> TGTCGTGTTACCACTGC-5'
DNA2mm	2	18	5' -GT <u>ct</u> CAGCACAATGGTGACG-3' 3' -CA <u>ag</u> GTCGTGTTACCACTGC-5'
DNA3mm	3	17	5' -GT <u>cta</u> AGCACAATGGTGACG-3' 3' -CA <u>aga</u> TCTGTGTTACCACTGC-5'
DNA4mm	4	16	5' -GT <u>ctag</u> GCACAATGGTGACG-3' 3' -CA <u>agat</u> CGTGTGTTACCACTGC-5'
DNA5mm	5	15	5' -GT <u>ctag</u> GCACAATGGTGACG-3' 3' -CA <u>agat</u> CGT <u>ct</u> TACCACTGC-5'
DNA6mm	6	14	5' -GT <u>ctag</u> GC <u>act</u> ATGGTGACG-3' 3' -CA <u>agat</u> CGT <u>cct</u> TACCACTGC-5'
DNA7mm	7	13	5' -GT <u>ctag</u> GC <u>actt</u> TGGTGACG-3' 3' -CA <u>agat</u> CGT <u>cct</u> ACCACTGC-5'
DNA8mm	8	12	5' -GT <u>ctag</u> GC <u>actta</u> GGTGACG-3' 3' -CA <u>agat</u> CGT <u>cctc</u> CCACTGC-5'
DNA9mm	9	11	5' -GT <u>ctag</u> GC <u>actta</u> GGT <u>g</u> ACG-3' 3' -CA <u>agat</u> CGT <u>cctc</u> CCATGC-5'
DNA10mm	10	10	5' -GT <u>ctag</u> GC <u>actta</u> GGT <u>ga</u> CG-3' 3' -CA <u>agat</u> CGT <u>cctc</u> CCAT <u>a</u> GC-5'

DNA-pm refers to the perfect match duplex sequence. n refers to the number of mismatch base pairs in the DNA. Duplex sequences for Set A and Set B (all listed 5'-3' for the top strand and 3'-5' for the bottom strand) are shown in column, 4 and 5, respectively. All mismatches are shown as bold, lower case and underlined. Each single strand DNA was hybridized to complementary single strand to form the duplex.

**Table 4.1(c):** Control Duplexes.

Sample ID	n	N	Control DNA Sequences
DNA-pmC	0	20	5' -CAGTGAGACAGCAATGGTCG-3' 3' -GTCACTCTGTCGTTACCAGC-5'
DNA7mmC	7	13	5' - <u>ctagctg</u> CACAATGGTGACG-3' 3' - <u>agatcca</u> GTGTTACCACTGC-5'
DNA11C	8	12	5' -GT <u>ct</u> CA <u>ag</u> AC <u>ct</u> TG <u>ta</u> GACG-3' 3' -CA <u>ag</u> GT <u>at</u> TG <u>cc</u> ACT <u>ct</u> CTGC-5'
DNA25C	2	14	5' -CTC <u>t</u> CATATG <u>cg</u> GAG-3' 3' -GAG <u>cg</u> TATACT <u>ct</u> CTC-5'

DNA-pmC refers to the 20 base pair DNA duplex of reference (57). DNA11C has four loops separated by W/C base pairs. DNA25C has two single base pair mismatches separated by six W/C base pairs and was characterized elsewhere (23).

Molecules used as control are shown in Table 4.1(c). DNA-pmC is a perfectly matched 20 base pair DNA duplex that has been characterized by Owczarzy *et al.* (57). DNA25C contains 14 base pairs and has two single base pair mismatches that has been characterized by Allawi *et al.* (23). DNA7mmC is nearly identical to DNA7mm in Set A (Table 4.1(a)), except that the t/t mismatch has been replaced by a g/a mismatch. The molecule DNA11C has four interspersed tandem mismatches, corresponding to eight tandem mismatches.

#### **4.1.2 DNA Molecules**

Synthetically prepared and purified DNA single strands used in melting studies were obtained from Integrated DNA Technologies. Purity levels reported by the supplier were greater than 90 %. As received, the dry powdered single strand synthetic DNA oligonucleotides were centrifuged for ten minutes at 3000 rpm. Single strands were then dissolved in 1.0 – 2.0 mL of 85 mM  $[\text{Na}^+]$  buffer (pH 7.40 - 7.50). Dissolved DNA stock solutions were thoroughly mixed by vortexing and incubated at 25 °C for at least five hours to equilibrate. Briefly, equimolar amounts of complementary single strand oligonucleotides were combined to form individual duplexes with  $A_{260}$  varying from 0.20 - 0.80 OD/mL.

#### **4.1.3 DNA Recovery**

DNA samples this particular study were transferred between different  $[\text{Na}^+]$  solvents by concentrating and desalting using strictly the Centricon YM-3 Centrifugal cellulose filters with a molecular weight cut off of 3000 Daltons. In these filters, a centrifugal force drives solvents (salts) but not solute (DNA) through the membrane into a filtrate vial. Concentrated samples were brought up in a volume of the desired melting buffer such that the  $A_{260}$  was within the 0.20 - 0.80 OD/mL range. Absorbance spectra of the duplex DNA samples were recorded and used in data analysis. This process was repeated to prepare samples for each  $[\text{Na}^+]$  buffer solvent.

#### 4.1.4 Analytical Gel Electrophoresis

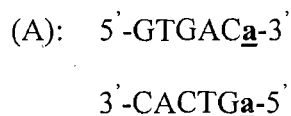
Procedures for PAGE were essentially those described in Chapter 2. In addition, purity of single strand DNAs was also characterized by denaturing PAGE (8 %) analysis. Typically, denaturing gels contained 7.0 M urea (98 %), 20  $\mu$ L TEMED, 10 % TBE, 10 % APS, 10 % acrylamide/bis acrylamide (38 %/2 %) and nanopure water.

## 4.2 THEORETICAL METHODS

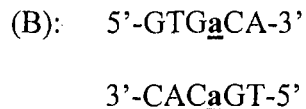
A detailed account of theoretical methods employed was presented in Chapter 2. Here, only methods specific to this project are presented.

### 4.2.1 Single base pair mismatches

Within the nearest-neighbor model, thermodynamic parameters for single base pair mismatches have been evaluated and tabulated in the literature (3-11,28,45,48,73). To see how the n-n model parameters can be utilized to predict the thermodynamic contributions of single base pair mismatches to duplex stability, consider the following sequences, labeled (A) and (B), each containing a single base pair mismatch







For sequence (A), containing a single base pair mismatch on the 'end' of the duplex, the free-energy is given by,

$$\Delta G^{\circ}_{duplex} = [\Delta G^{\circ}(GT/CA) + \Delta G^{\circ}(TG/AC) + \Delta G^{\circ}(GA/CT) + \Delta G^{\circ}(AC/TG) + \Delta G^{\circ}(Ca/Ga)] + \Delta G^{\circ}_{initiation} \quad (4.1)$$

where the first four terms in brackets on the right hand side of Equation (4.1) correspond to the resident n-n doublets comprised of W/C base pairs. The last term in brackets,  $\Delta G^{\circ}(Ca/Ga)$ , is the free energy parameter for an a/a single base pair mismatch bounded on the 5' side by an intact C-G base pair. For sequence (B) with an internal a/a single base mismatch pair bounded by a G-C and C-G base pair, the free-energy is given by,

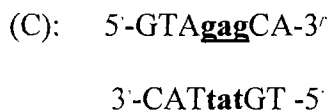
$$\Delta G^{\circ}_{duplex} = [\Delta G^{\circ}(GT/CA) + \Delta G^{\circ}(TG/AC) + \Delta G^{\circ}(Ga/Ca) + \Delta G^{\circ}(Ca/Ga) + \Delta G^{\circ}(CA/GT)] + \Delta G^{\circ}_{initiation} \quad (4.2)$$

where  $\Delta G^{\circ}(Ga/Ca)$  is the free-energy parameter for an a/a single base pair mismatch bounded by an intact G-C base pair on the 5' side, and  $\Delta G^{\circ}(Ca/Ga)$  is the free-energy parameter for an a/a single base pair mismatch bounded on the 3' side by an intact C-G base pair. Due to symmetry and the anti-parallel nature of

duplex DNA structure, note that  $\Delta G^\circ(\underline{\text{G}}\underline{\text{a}}/\underline{\text{C}}\underline{\text{a}}) = \Delta G^\circ(\underline{\text{a}}\underline{\text{C}}/\underline{\text{a}}\underline{\text{G}})$ . As described before, thermodynamic stability parameters for most of the n-n sequence dependent interactions for single base pair mismatches (i.e. flanked by intact W/C base pairs) have been evaluated.

#### 4.2.2 Tandem Mismatches

Although the n-n thermodynamic parameters for W/C perfect match base pair doublets, doublets containing single base pair mismatches and doublets containing single base pair dangling ends are available in the published literature (3-5,7-13,17,22-25,27-28,48,60-63) sequence-dependent stability parameters for the stability of tandem mismatches have not been determined (2,10,15). Even so, several approximate strategies applying the n-n model have been developed to approximate thermodynamic stability of tandem mismatches in duplexes. As explained earlier, the most commonly employed approximate method assumes that tandem mismatches make no favorable contribution to duplex stability (10,15). For example, consider the following duplex DNA sequences, denoted (C) and (D), both containing tandem mismatches;



(D): 5'-GTGAgag-3'

3'-CACTtat-5'

Duplex sequence denoted (C) contains three tandem mismatch base pairs 'interspersed' in the middle of the duplex, while sequence (D) is comprised of tandem mismatches on the 'end' of the duplex. In the current n-n model method, the calculated n-n free energy of sequence (C) is given by,

$$\begin{aligned} \Delta G^{\circ}_{duplex} = & \Delta G^{\circ}(GT/CA) + \Delta G^{\circ}(TA/AT) + \Delta G^{\circ}(A\mathbf{g}/T\mathbf{t}) + \\ & \Delta G^{\circ}_{loop}(n = 3) + \Delta G^{\circ}(\mathbf{g}C/\mathbf{t}G) + \Delta G^{\circ}(CA/GT) + \Delta G^{\circ}_{initiation} \end{aligned} \quad (4.3)$$

Notice that the tandem mismatch loop (**gag/tat**) is assigned a loop parameter,  $\Delta G^{\circ}_{loop}$  ( $n = 3$ ), where  $n$  denotes the number of mismatches comprising the loop. In comparison the n-n calculated free energy of segment (D) is given by,

$$\begin{aligned} \Delta G^{\circ}_{duplex} = & \Delta G^{\circ}(GT/CA) + \Delta G^{\circ}(TG/AC) + \Delta G^{\circ}(GA/CT) + \\ & \Delta G^{\circ}(A\mathbf{g}/T\mathbf{t}) + \Delta G^{\circ}_{initiation} \end{aligned} \quad (4.4)$$

where  $\Delta G^{\circ}(A\mathbf{g}/T\mathbf{t})$  is the free energy parameter for a single base pair **g/t** mismatch bounded on the 5' side by an intact A-T base pair. In the current n-n method, tandem mismatches on the ultimate 'end' of the DNA duplex are not assigned a particular free-energy parameter, and their potential effects are essentially ignored.

Herein may be a deficiency of the n-n model in accurately predicting thermodynamic stability of short duplex DNAs containing tandem mismatches. In contrast, for tandem mismatches in the middle of the DNA duplex, some loop energy terms have been published (10) and these were employed in our predictive calculations.

## 4.3 RESULTS AND DISCUSSION

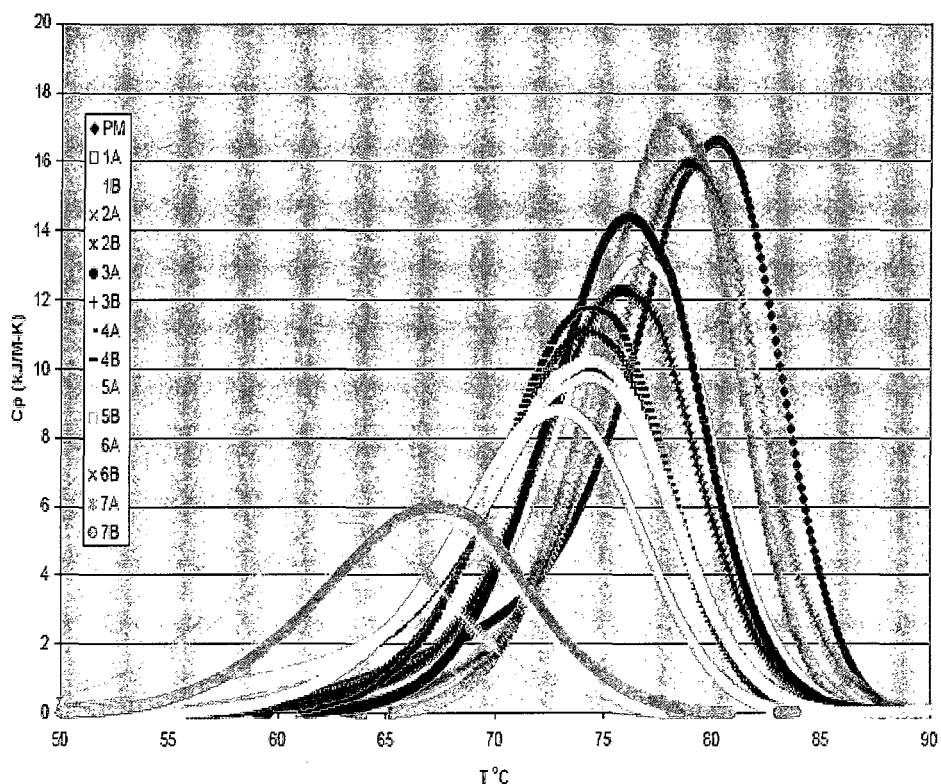
### 4.3.1 Analysis of Melting Curves

Experimental DSC melting curves of the excess heat capacity,  $\Delta C_p^{\text{ex}}$ , versus temperature (DSC melts) were measured over the range from 0 to 120 °C. Typical baseline-corrected DSC curves obtained for the DNAs are shown in Figure 4.1. Recall that in Set A, mismatches occur on the 'end' while Set B has mismatches 'internal' to the DNA duplex. Following are some general observations; DSC melting curves are relatively sharper for the perfectly matched duplexes than for duplexes containing mismatched bases. Data shown in Figure 4.2 reveal that increasing the number of mismatches causes  $T_m$  to shift progressively to lower temperatures. Essentially, as the number of mismatches increases, melting profiles become broader. For some sequences (especially those containing tandem mismatches (i.e. at least a c/t or t/c)), heating and cooling transitions were not exactly equivalent, suggesting some hysteresis in the experiments. Subtle

differences in the shapes of the DSC melting curves were observed in Sets A and B duplexes. Generally, more DSC melting curves could be reliably acquired for duplexes in Set A than in Set B molecules. This is indicative of mismatch base pairs better tolerated when they are towards the 'ends' of the duplex. For Set A duplexes, DSC melting curves were monophasic in all cases and DSC melting curves for duplexes containing up to ten mismatches were collected and analyzed. For Set B, DSC melting curves were collected for duplexes containing up to six mismatches. This is indicative of higher stability associated with contiguous mismatches introduced from the 'end' of the DNA duplex. In contrast, for Set B duplexes, there seems to be a threshold number of mismatches above which thermodynamic transition parameters suddenly plummet. This was evidenced by DSC curves becoming increasingly broad and evaluated thermodynamic parameters were generally smaller than for the Set A duplexes. In addition, shapes of progressive baselines differed for the duplexes in Sets A and B. Baselines depended on the linearity of the low and high temperature regions of the acquired curves. For Sets B duplexes, pre and post-melting transition baselines were difficult to fit after addition of some types of mismatches which were apparently highly destabilizing.

#### **4.3.2 Evaluation of Thermodynamic Parameters**

Thermodynamic parameters of thermally induced melting transitions evaluated by DSC measurements carried out in four  $[\text{Na}^+]$  environments are listed in Tables 4.2(a)



**Figure 4.1:** Representative DSC Melting Curves evaluated in 600 mM  $[\text{Na}^+]$ . Normalized DSC melting curve profiles shown correspond to fifteen samples (illustrated in different colors). As shown on the legend the numbers and letters correspond to the mismatches and systems studied, respectively. Note the three highest peaks correspond to Set A duplexes (i.e. DNA-PM, DNA1mmA, DNA2mmA). Broad curves had peaks shifted to lower melting temperatures, and these correspond to DNAmm6B, DNAmm7B that contain destabilizing  $\underline{c/c}$  and  $\underline{c/t}$  mismatches.

and (b). These results show that the range of evaluated thermodynamic parameters were higher in for Set A than Set B duplexes. It is also clear from Table 4.2(a) that the extent of helix stabilization associated with tandem mismatches seems to be greater for Set A molecules in all  $[\text{Na}^+]$  environments. This is consistent with previous results obtained by Allawi *et al.* (4,24) of DNA duplexes containing only **ga** and **ac** mismatches. In these studies, ‘terminal’ mismatches were generally more stable than ‘internal’ mismatches.

Further, duplexes were stabilized in higher  $[Na^+]$  and mismatches were destabilizing in all sequence contexts. This observation is consistent with previous studies that have shown that increased concentrations of  $Na^+$  ions leads to the stabilization of a duplex structure (10,12-13,44,56).

**Table 4.2(a):** Measured Thermodynamic Parameters for Set A and Set B duplexes in buffered 85 – 1000 mM  $[Na^+]$ .

1000 mM $[Na^+]$																	
Sample ID	N	$-\Delta H$ (kcal/mol)				$-\Delta S$ (e.u)				$-\Delta G_{25}$ (kcal/mol)				$T_m$ (°C)			
		Set A	$\sigma$	Set B	$\sigma$	Set A	$\sigma$	Set B	$\sigma$	Set A	$\sigma$	Set B	$\sigma$	Set A	$\sigma$	Set B	$\sigma$
DNA-pm	20	183.0	5.8	183.0	5.8	515.8	16.2	515.8	16.2	29.2	2.7	29.2	2.7	81.6	0.5	81.6	0.5
DNA1mm	19	163.7	4.5	161.4	6.3	457.5	10.6	457.1	16.3	27.3	0.7	25.1	3.1	80.8	0.3	78.4	0.3
DNA2mm	18	159.8	0.7	152.2	5.7	451.5	2.1	439.5	22.3	25.2	1.2	21.2	2.0	78.4	0.2	80.0	0.5
DNA3mm	17	133.9	2.3	129.4	6.3	380.6	6.5	371.1	3.4	20.4	3.5	18.8	1.7	78.4	0.1	77.9	0.2
DNA4mm	16	134.4	4.5	116.1	21.1	384.0	3.5	334.1	12.4	19.9	1.3	16.5	2.0	75.6	0.1	76.8	0.3
DNA5mm	15	131.2	16.9	70.8	12.9	376.3	0.4	202.9	8.9	19.1	2.4	10.3	4.2	75.6	0.0	75.3	0.1
DNA6mm	14	95.7	9.9	65.3	11.6	268.3	13.1	190.5	2.5	15.7	2.2	8.5	4.1	72.7	0.1	73.8	0.2
DNA7mm	13	88.5	4.4	55.9	4.2	251.8	11.3	164.1	12.2	13.6	1.4	7.0	1.9	68.5	0.1	66.9	0.7
DNA8mm	12	82.6	4.6	52.9	0.6	242.5	16.3	159.3	2.5	10.3	2.4	5.4	1.5	65.2	0.1	64.7	0.1
DNA9mm	11	78.7	8.2	41.8	1.7	239.7	0.4	124.8	5.2	7.2	3.2	4.8	1.8	64.4	0.1	65.2	0.2
DNA10mm	10	68.7	17.8	39.8	8.0	210.4	15.3	123.6	8.2	6.0	1.2	2.9	1.2	64.5	0.1	58.8	0.6

600 mM $[Na^+]$																	
Sample ID	N	$-\Delta H$ (kcal/mol)				$-\Delta S$ (e.u)				$-\Delta G_{25}$ (kcal/mol)				$T_m$ (°C)			
		Set A	$\sigma$	Set B	$\sigma$	Set A	$\sigma$	Set B	$\sigma$	Set A	$\sigma$	Set B	$\sigma$	Set A	$\sigma$	Set B	$\sigma$
DNA-pm	20	166.0	10.2	169.0	10.2	475.6	6.2	475.6	6.2	26.2	2.8	26.2	2.8	80.0	0.2	80.0	0.2
DNA1mm	19	150.4	4.4	136.7	11.0	430.0	6.9	390.4	31.3	22.2	2.0	20.3	1.9	78.4	0.2	77.0	0.4
DNA2mm	18	148.0	3.5	127.9	12.9	421.3	9.9	366.3	36.8	22.4	1.5	18.7	2.9	79.0	0.1	78.2	0.4
DNA3mm	17	131.4	4.5	116.2	12.7	376.6	12.7	335.5	35.4	19.1	1.2	16.2	1.2	76.2	0.2	74.5	0.3
DNA4mm	16	114.7	1.2	113.5	2.2	331.0	9.0	331.9	6.2	16.0	0.7	14.5	2.6	74.3	0.1	74.6	0.3
DNA5mm	15	110.1	0.6	68.5	8.8	321.8	9.0	188.4	24.2	14.1	2.3	9.4	2.1	75.6	0.1	73.3	1.0
DNA6mm	14	96.5	2.1	64.3	4.4	278.6	5.5	187.0	13.0	13.4	2.1	8.5	4.9	73.0	0.1	67.4	0.1
DNA7mm	13	93.0	14.0	61.1	0.9	269.5	7.0	185.8	3.0	12.6	3.3	6.7	0.7	67.2	0.1	66.2	0.1
DNA8mm	12	79.9	3.7	51.8	2.3	247.5	7.1	156.7	3.8	8.1	1.5	5.1	1.2	63.1	0.1	62.2	0.1
DNA9mm	11	68.4	1.9	49.2	0.6	197.3	5.5	149.7	2.5	7.6	1.0	3.6	1.6	64.5	0.2	58.2	0.4
DNA10mm	10	55.7	4.3	44.6	10.0	171.2	6.2	142.9	21.3	4.6	1.2	2.0	0.6	61.8	0.2	57.9	0.0

300 mM [Na<sup>+</sup>]

Sample ID	N	-ΔH(kcal/mol)				-ΔS(e.u)				-ΔG <sub>25</sub> (kcal/mol)				T <sub>m</sub> (°C)			
		Set A	σ	Set B	σ	Set A	σ	Set B	σ	Set A	σ	Set B	σ	Set A	σ	Set B	σ
DNA-pm	20	155.6	15.6	155.6	15.6	430.3	22.9	430.3	22.9	24.6	2.3	24.6	2.3	77.3	0.2	77.3	0.2
DNA1mm	19	150.4	2.0	146.6	21.7	430.3	6.8	429.6	22.1	22.1	1.7	18.5	2.9	75.8	0.2	74.7	0.2
DNA2mm	18	144.9	2.6	145.9	13.9	418.0	7.3	429.1	21.5	20.2	3.7	18.0	1.4	77.0	0.1	73.2	0.3
DNA3mm	17	124.3	3.6	116.0	16.4	359.3	10.6	345.5	5.0	17.2	1.6	13.0	2.0	72.8	0.0	70.7	0.0
DNA4mm	16	115.0	2.3	88.9	8.4	331.8	6.6	261.0	3.9	16.0	0.5	11.1	1.0	73.2	1.4	67.1	0.0
DNA5mm	15	115.4	4.5	90.8	14.4	338.0	31.4	269.8	1.4	14.6	1.5	10.1	2.6	70.8	0.0	67.5	0.1
DNA6mm	14	99.6	1.4	85.3	6.1	292.0	4.2	254.7	25.5	12.5	1.6	9.4	2.0	68.4	0.0	67.4	0.0
DNA7mm	13	87.9	1.0	57.7	6.9	261.0	2.9	169.0	19.6	10.1	1.2	7.3	1.4	68.1	0.1	67.6	0.9
DNA8mm	12	62.2	0.7	49.4	12.3	195.0	2.2	151.4	17.4	7.1	1.4	4.3	0.9	63.7	0.1	60.4	0.2
DNA9mm	11	52.5	3.1	38.2	7.9	158.0	9.9	122.3	23.5	5.4	2.2	1.7	1.3	58.9	0.0	58.8	0.1
DNA10mm	10	41.8	0.3	34.7	6.8	124.6	0.7	117.2	9.0	4.7	1.6	-0.2	0.2			57.3	0.3

85 mM [Na<sup>+</sup>]

Sample ID	N	-ΔH(kcal/mol)				-ΔS(e.u)				-ΔG <sub>25</sub> (kcal/mol)				T <sub>m</sub> (°C)			
		Set A	σ	Set B	σ	Set A	σ	Set B	σ	Set A	σ	Set B	σ	Set A	σ	Set B	σ
DNA-pm	20	146.6	5.3	146.6	5.3	423.6	14.9	423.6	14.9	20.3	1.8	20.3	1.8	72.5	0.2	72.5	0.0
DNA1mm	19	141.7	5.1	142.4	0.2	410.8	11.7	411.8	0.8	19.2	1.7	19.7	2.4	70.1	0.1	72.5	0.3
DNA2mm	18	126.2	2.8	125.5	2.8	360.5	15.0	368.3	8.2	16.7	0.3	15.7	0.8	70.2	1.3	67.6	0.1
DNA3mm	17	108.1	5.8	90.9	7.0	318.5	15.9	270.3	20.6	13.1	3.4	10.3	1.0	66.9	0.0	63.3	0.2
DNA4mm	16	96.7	9.1	86.8	10.6	282.9	15.7	252.8	30.6	12.3	1.3	11.4	0.6	67.7	0.2	60.5	0.1
DNA5mm	15	78.3	31.2	50.3	2.0	233.0	5.7	144.1	4.4	8.8	1.2	7.4	2.0	67.3	0.1	61.8	0.0
DNA6mm	14	67.7	1.2	44.4	2.9	203.0	3.5	134.0	9.9	7.2	1.2	4.4	1.9	60.7	0.1	61.0	0.3
DNA7mm	13	68.6	0.3	49.1	2.5	210.5	0.7	152.0	14.8	6.8	0.8	3.8	0.5	59.2	0.2	59.4	0.4
DNA8mm	12	52.3	1.3	50.1	14.3	162.3	15.8	159.0	10.5	3.9	0.1	3.0	1.3	58.8	0.2	58.3	0.0



**Table 4.2(b):** Measured thermodynamic data for control molecules used in this study.

1000 mM [Na <sup>+</sup> ]								
Sample ID	$-\Delta H(\text{kcal/mol})$	$\sigma$	$-\Delta S(\text{e.u.})$	$\sigma$	$-\Delta G_{25}^{\circ}(\text{kcal/mol})$	$\sigma$	$T_m(^{\circ}\text{C})$	$\sigma$
DNA-pmC	175.3	0.2	494.3	0.6	28.0	0.4	81.3	0.3
DNA7mmC	106.3	0.7	302.3	5.4	16.2	2.3	28.0	0.3
DNA11C	41.8	3.2	129.5	8.3	3.2	0.9	16.2	0.2
DNA25C	76.3	2.2	242.1	11.9	4.1	0.1	3.2	0.3
600 mM [Na <sup>+</sup> ]								
Sample ID	$-\Delta H(\text{kcal/mol})$	$\sigma$	$-\Delta S(\text{e.u.})$	$\sigma$	$-\Delta G_{25}^{\circ}(\text{kcal/mol})$	$\sigma$	$T_m(^{\circ}\text{C})$	$\sigma$
DNA-pmC	172.6	8.9	487.7	14.3	27.2	2.2	81.3	0.3
DNA7mmC	101.7	5.1	294.5	15.3	14.0	0.7	27.2	0.2
DNA11C	34.8	2.2	111.8	5.8	1.5	0.1	14.0	0.8
DNA25C	63.0	4.3	196.2	6.4	4.5	0.6	1.5	0.1
300 mM [Na <sup>+</sup> ]								
Sample ID	$-\Delta H(\text{kcal/mol})$	$\sigma$	$-\Delta S(\text{e.u.})$	$\sigma$	$-\Delta G_{25}^{\circ}(\text{kcal/mol})$	$\sigma$	$T_m(^{\circ}\text{C})$	$\sigma$
DNA-pmC	162.4	3.3	457.5	21.6	25.9	0.3	81.2	0.1
DNA7mmC	90.7	11.7	262.0	8.9	12.6	1.4	25.9	0.1
DNA11C	45.6	4.6	146.4	11.8	1.9	0.4	12.6	0.3
DNA25C	52.4	2.7	165.2	16.2	3.1	0.8	1.9	0.2
85 mM [Na <sup>+</sup> ]								
Sample ID	$-\Delta H(\text{kcal/mol})$	$\sigma$	$-\Delta S(\text{e.u.})$	$\sigma$	$-\Delta G_{25}^{\circ}(\text{kcal/mol})$	$\sigma$	$T_m(^{\circ}\text{C})$	$\sigma$
DNA-pmC	148.4	7.3	420.0	10.3	23.2	1.2	80.4	0.2
DNA7mmC	100.3	11.2	289.0	7.8	14.1	3.1	23.2	0.1
DNA11C	25.3	6.3	79.0	9.4	1.7	0.4	14.1	0.1
DNA25C	41.9	4.9	133.0	7.3	2.2	0.1	1.7	0.5

Thermodynamic transition parameters derived from DSC measurements of “control” DNA molecules are shown in Table 4.2(c). Results for thermodynamic parameters for the 14 base pair sequence, DNA25C containing two c/t mismatches were also evaluated in this study. Measured values for free energy transitions differed from those evaluated by SantaLucia *et al.* (23) by about 11.2 %. Such differences are expected since these values were mostly evaluated using spectrophotometry, versus DSC utilized in these studies. As expected, DNA7mmC in which the more stable g/a mismatch replaces the t/t mismatch in DNA7Amm, was found to be more stable and had higher values in magnitude for ( $\Delta H^{cal}$ ,  $\Delta S^{cal}$

and  $\Delta G_{25^{\circ}\text{C}}^{\text{cal}}$ ) as well as  $T_m$ . Results for these control molecules are in general agreement with several investigations in the literature in which the stability of duplexes containing various **g** and or **a** mismatches. Ebel *et al.* (7) and Li *et al.* (88) found that DNA duplexes consisting of at least **ga/ag** mismatch to be more stable than some perfect matched AA/TT and AT/TA W/C base pair doublets. As expected, the perfectly matched DNA-pmC was the most enthalpically stable species (most exothermic), with a compensating increase in entropic stabilization. Results for these control molecules provide an independent basis for comparison with our findings.

### 4.3.3 Entropy-Enthalpy Compensation

Enthalpy-entropy compensation is a common phenomenon in molecular recognition (67-70,84,104), and is also present in the melting of short duplex DNA molecules containing tandem mismatches. As can be seen in Figures 4.2(a) and (b), plots of  $\Delta H^{\text{cal}}$  versus  $T\Delta S^{\text{cal}}$  exhibit a linear relationship ( $R^2 > 0.946$ ). Thus, entropy changes (i.e. favoring disordered single stranded states) reflect greater increases of rotational and translational degrees of freedom upon duplex melting and are directly related to enthalpically destabilizing forces of DNA duplex structure (59,67,104) associated with tandem mismatches.

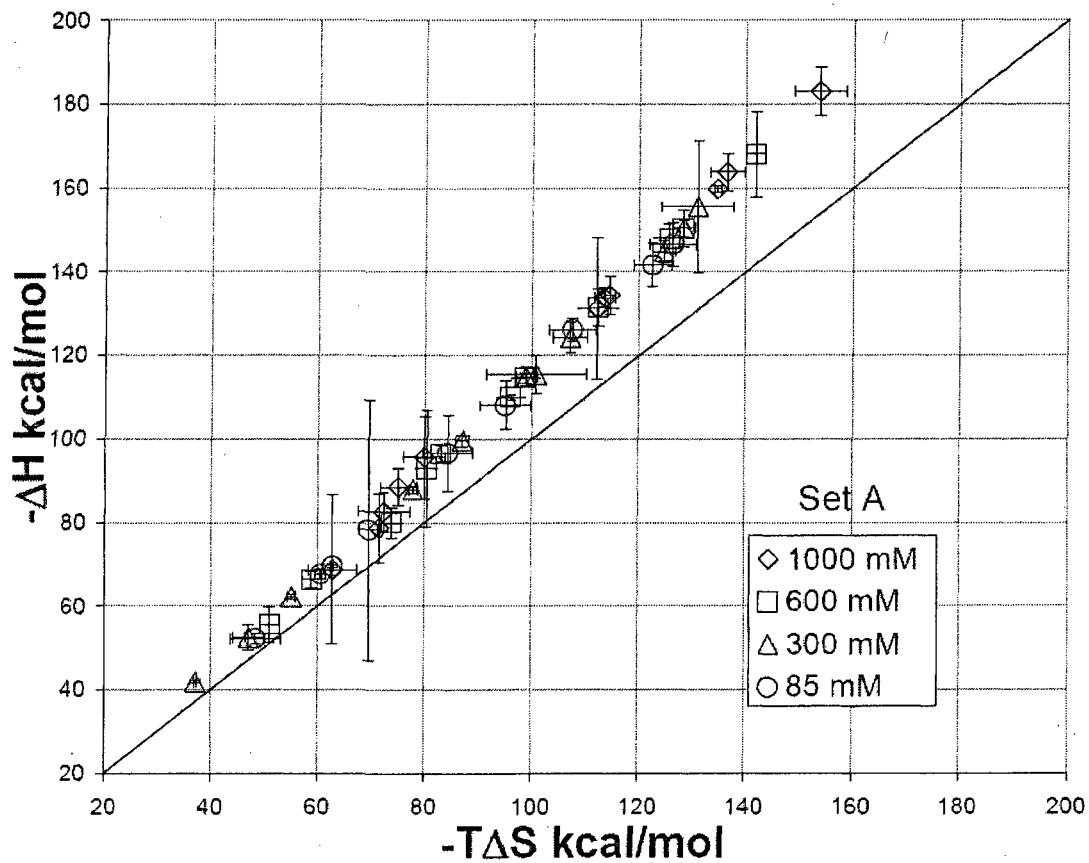


Figure 4.2(a): Plot of  $\Delta H^{cal}$  versus  $T\Delta S^{cal}$  at 298.15K in all [Na<sup>+</sup>] environments for Set A duplexes.

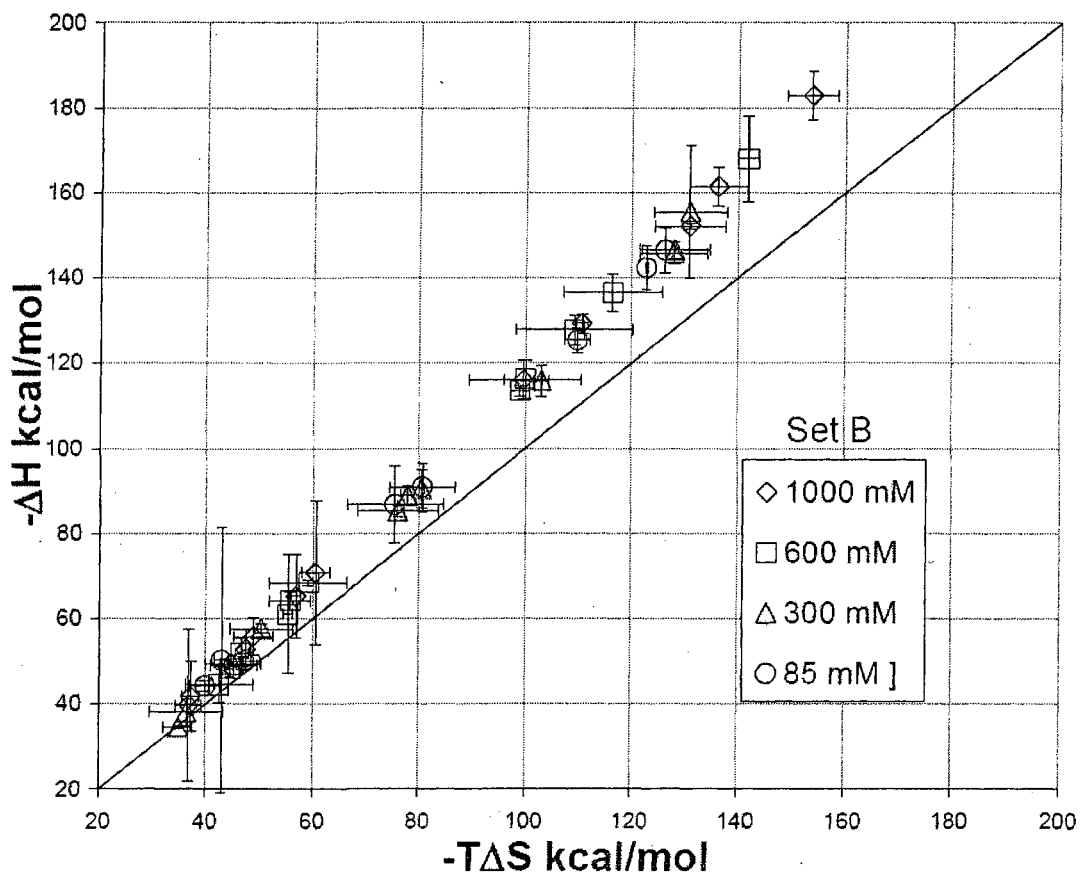
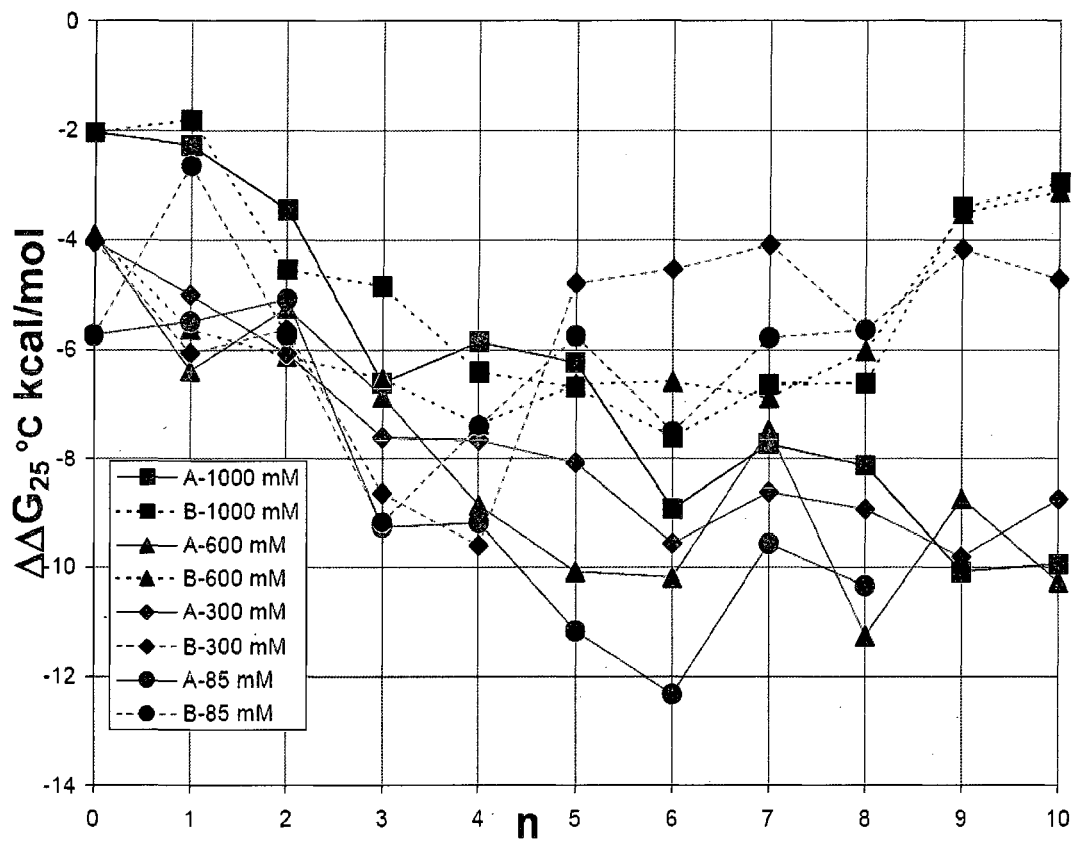


Figure 4.2(b): Plot of  $\Delta H^{cal}$  versus  $T\Delta S^{cal}$  at 298.15K in all  $[Na^+]$  environments for Set B duplexes.

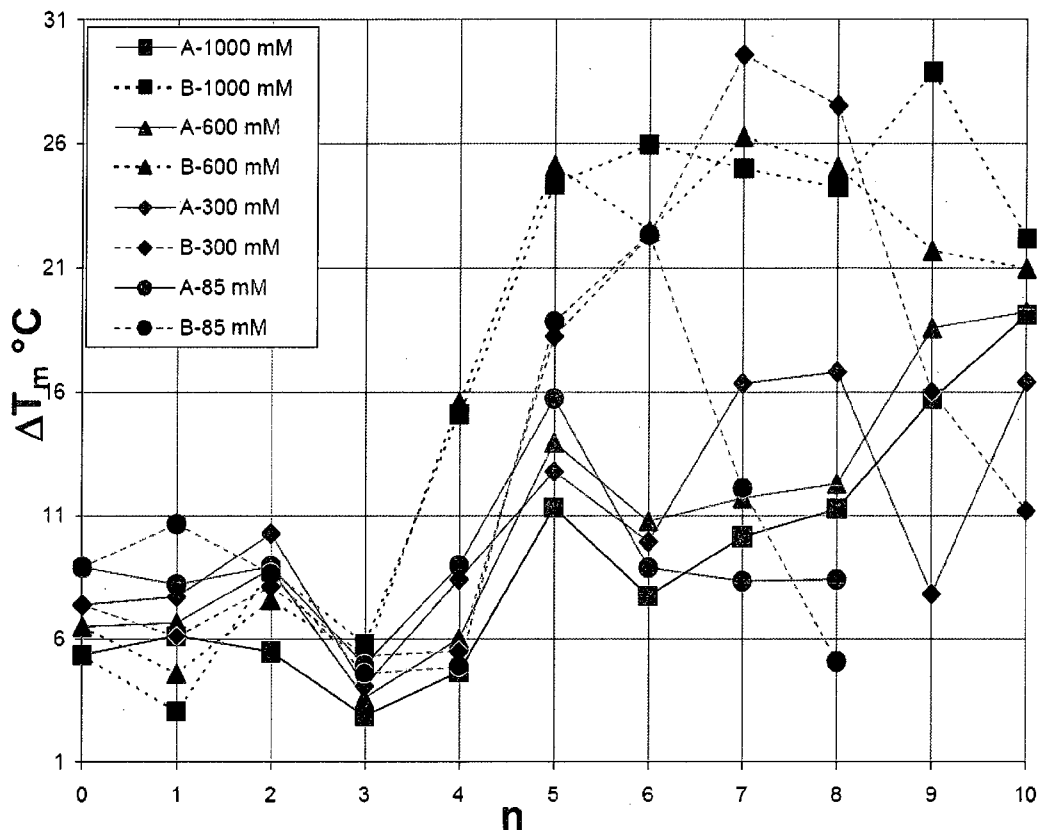
Results shown above indicate that in essence, for duplexes containing tandem mismatches, stabilizing enthalpic contributions are counterbalanced by destabilizing entropic effects. Error bars in Figures 4.2(a) and (b) denote experimental reproducibility. Figures 4.2 (a) and (b) clearly display enthalpy dominated stability of the duplexes.

#### 4.3.4 Is the n-n model valid for predicting thermodynamic stability of short DNA Duplexes containing tandem mismatches?

In order to test the validity of the n-n model to calculate the thermodynamics of short DNA molecules consisting of tandem mismatches, a direct comparison of n-n predicted and experimentally measured melting free energies was made. Differences between predicted and measured free energies is given by  $\Delta\Delta G_{25^\circ\text{C}}$  (kcal/mol) where  $\Delta\Delta G_{25^\circ\text{C}} = (\Delta G_{25^\circ\text{C}}^{\text{cal}} - \Delta G_{25^\circ\text{C}}^{\text{o(predicted)}})$  determined for Sets A and B. These values are plotted in Figure 4.3 (for all  $[\text{Na}^+]$ ). Clearly there are considerable differences between predicted and experimentally measured values revealing the poor predictive power of the n-n model for these duplexes. In particular, very significant differences are observed for molecules of Set A (having  $n > 5$ ), especially in low  $[\text{Na}^+]$  (85 and 300 mM). Apparently n-n predictions are not accurate for molecules having mismatches on the 'end'. Conversely, predictions for molecules containing 'internal' mismatches made by employing an 'internal' loop parameter are not as bad. This suggests there is a threshold number of mismatches ( $n \approx 5$ ) in the DNA duplex for which the n-n model is considerably worse. Differences in measured and predicted values of the  $T_m$  of the two sets of molecules in each salt environment are summarized in Figure 4.4. For both sets of duplexes, greater differences between n-n based predictions and experimental measurements are observed as the number of mismatches increases above five.



**Figure 4.3:** Plot of  $\Delta\Delta G_{25^\circ\text{C}}(\text{kcal/mol}) = (\Delta G_{25^\circ\text{C}}^{\text{cal}} - \Delta G_{25^\circ\text{C}}^{\text{o(predicted)}})$  versus  $n$  for Sets A and B duplexes. Predicted values were calculated using the MFOLD  $n$ - $n$  model routine (64). Set A and Set B duplexes are denoted by black and red colors, respectively.



**Figure 4.4:** Plot of  $\Delta T_m$  ( $T_m^{\text{measured}} - T_m^{\text{predicted}}$ ) versus  $n$  for duplexes of Sets A and B duplexes. Predicted values were made using the MFOLD routine. Set A and Set duplexes are denoted by black and red colors, respectively.

The experimental error in  $T_m$  determination is less than 1.0 °C. Clearly, major discrepancies between measured and predicted values of  $T_m$  are observed for all  $[\text{Na}^+]$  from 85 to 1000 mM. Moreover these differences for Set A and B duplexes increase dramatically when the number of mismatches is greater than four. These findings reveal, despite the reported  $n$ - $n$  parameters and associated empirical corrections for the dependence of  $T_m$  dependence on  $[\text{Na}^+]$  (24,44,56,60,63,108),  $\Delta G^\circ$  (22,24,48,63,73) cannot yet be accurately predicted using the  $n$ - $n$  model. Likely origins of this deficiency is the potential influence beyond  $n$ - $n$  of tandem

mismatches on the melting of short duplex DNA. Tandem mismatches are not currently considered (adequately) predictions of short DNA thermodynamic stability (64,90). Inconsistencies of the n-n model predictions of melting temperatures and free energy values have been reported by Hall *et al.* (50), Breslauer *et al.* (73) and SantaLucia *et al.* (8), suggesting the possible presence of sequence-dependent interactions beyond nearest neighbors. In addition, salt corrections reported in the literature for DNA oligomers are not applicable to ionic environments below 100 mM (11,44,56,58,71). These factors contribute to inadequacies of the n-n model.

#### **4.3.5 Dependence of short DNAs containing tandem mismatches on $[\text{Na}^+]$**

Dependence of the thermodynamic stability of short DNA duplexes containing tandem mismatches on differential  $[\text{Na}^+]$  was examined and directly compared with predictions using the n-n model. The local secondary structure is strongly influenced by the binding of solvent ( $\text{Na}^+$ ) and water. Thus, the dependence of the melting stability on  $\text{Na}^+$  can provide indications of secondary structural perturbations and fidelity.



#### 4.3.6

#### Analysis of Melting temperature

Recall from Chapter 2 that values for  $T_m$  are determined as the temperature corresponding to the maximum of the DSC melting curves of,  $\Delta C_p^{\text{ex}}$ , versus temperature. Experimentally derived  $T_m$  values for the 25 duplexes studied in this study are presented in Table 4.2. Following are noteworthy observations; (i) As expected, data shown in Figure 4.2 reveal that  $T_m$  is higher for both types of duplexes in higher salt (i.e. values for  $T_m$  in  $1000 > 600 > 300 > 85 \text{ mM } [\text{Na}^+]$ ) and decreases progressively as the number of mismatches increases, regardless of  $[\text{Na}^+]$  environment. (ii) Values for  $T_m$  are consistently higher for Set A duplexes than Set B duplexes for duplexes that have the same number of mismatches.

Next to quantify the dependence of the melting transitions on  $[\text{Na}^+]$ , plots of  $T_m^{-1} (K^{-1})$  versus  $\ln [\text{Na}^+]$  were constructed (plots not shown). As outlined in Equation (2.32), the slopes of these plots given by  $\frac{d(T_m^{-1})}{d\ln[\text{Na}^+]}$  provide estimates on the stoichiometric release of sodium ions,  $\Delta n$ , upon dissociation of duplex to single strands. From Equation (2.34) the stoichiometric release of sodium ions per phosphate upon melting,  $\Delta\psi$ , is defined. From these equations values of  $\Delta\psi$  were determined for every DNA molecule. For comparison the values of  $\Delta\psi$  and  $\Delta n$  evaluated for the Sets A and B, using the DSC measured enthalpy at each  $[\text{Na}^+]$ , are given in Table 4.3. In order to directly compare the relative behaviors of the Set A and B molecules, histograms of  $\Delta n$  and  $\Delta\psi$  were constructed. Results are shown in Figures 4.5 and 4.6. Sets A and B are uniquely

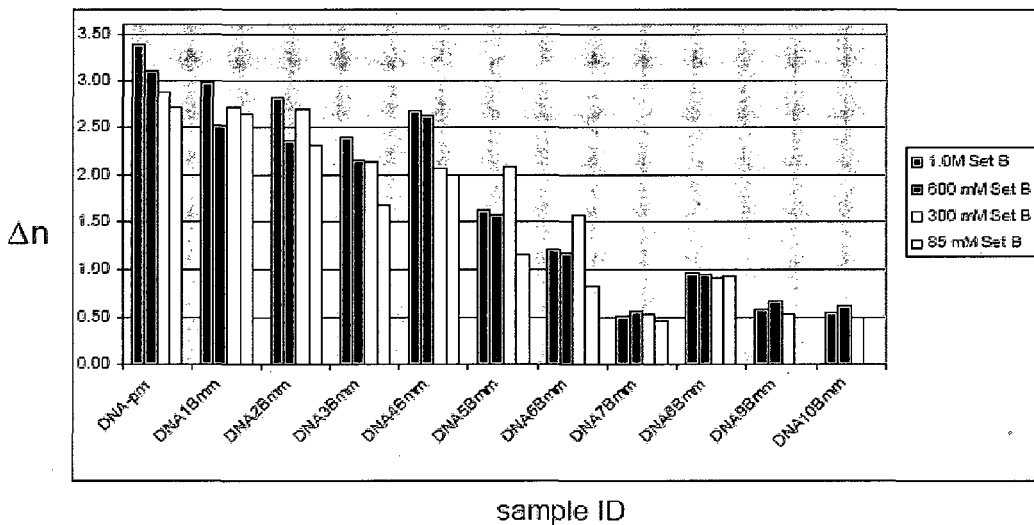
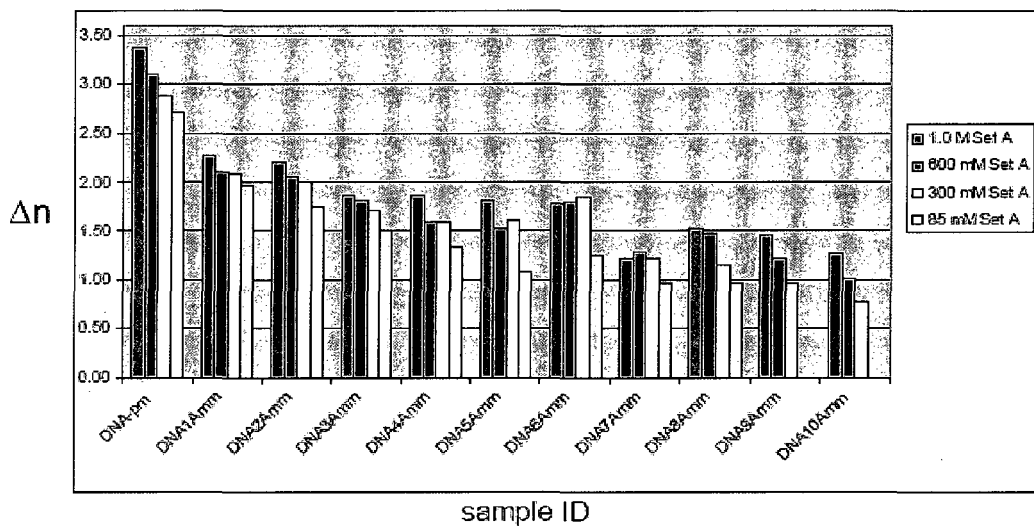
distinguished by either having tandem mismatch pairs on the end versus interspersed with W/C base pairs in the duplex. Values of  $\Delta n$  and  $\Delta\psi$  are highest for the 20 base pair perfect match duplex (DNA-pm) and decrease progressively with increased number of mismatches. Initially, values for  $\Delta n$  and  $\Delta\psi$  for Set A duplexes are systematically lower than those for Set B duplexes and progressively fall at a fairly constant rate with increasing number of tandem mismatches. As can be observed in Figures 4.5 and 4.6 values of  $\Delta n$  for Set B duplexes decrease more than the Set A with increased numbers of mismatches, to a final minimum value of 0.534 and versus 0.963, respectively.

**Table 4.3 (a).** Data tables for calculated values of  $\Delta n$  and  $\Delta \psi$  for Set A duplexes in 85 to 1000 mM  $[\text{Na}^+]$ .

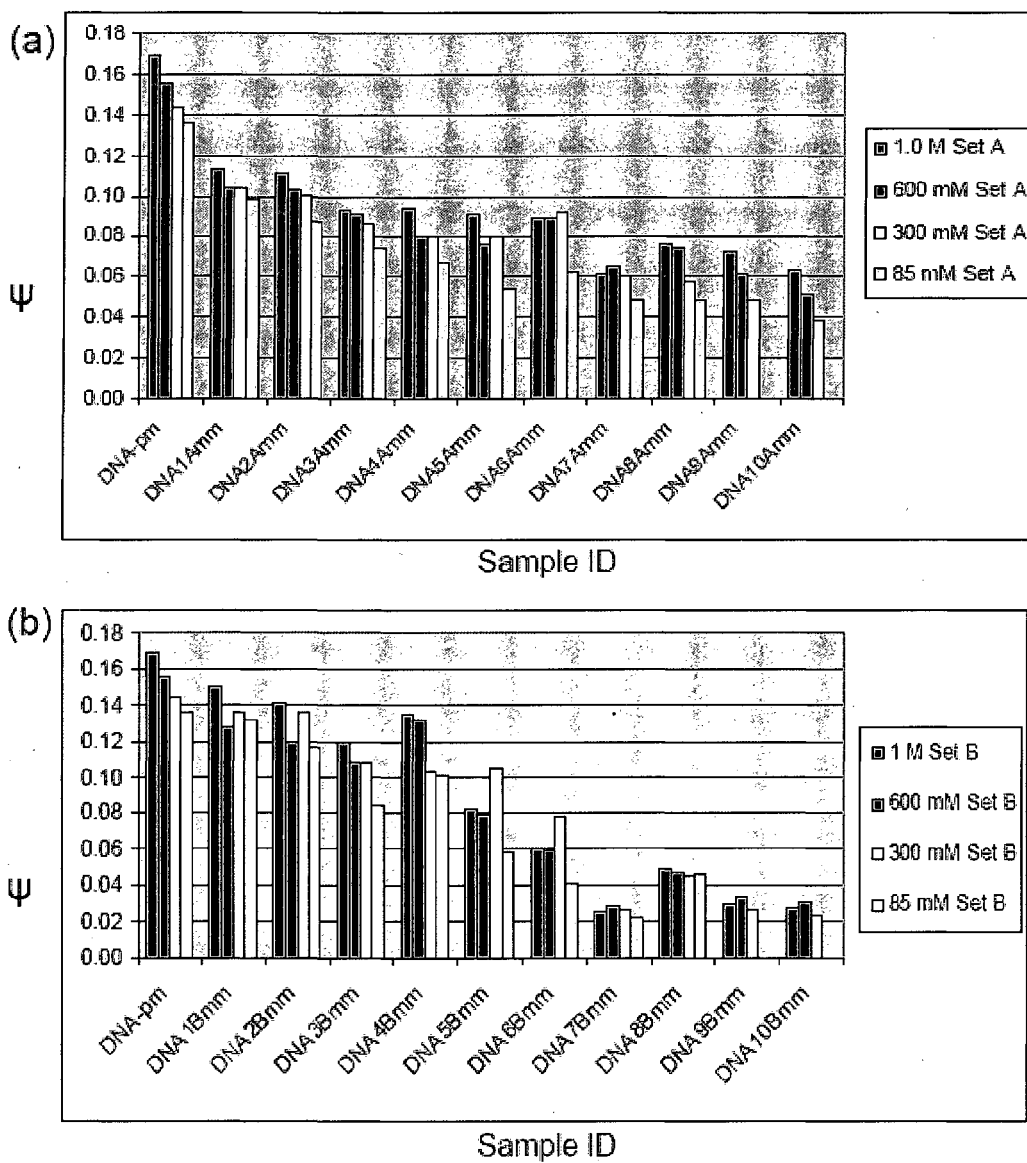
Set A Sample ID	$\Delta n$	$\Delta n$	$\Delta n$	$\Delta n$	$\Delta \psi$	$\Delta \psi$	$\Delta \psi$	$\Delta \psi$
	1000 mM [Na <sup>+</sup> ]	600 mM [Na <sup>+</sup> ]	300 mM [Na <sup>+</sup> ]	85 mM [Na <sup>+</sup> ]	1000 mM [Na <sup>+</sup> ]	600 mM [Na <sup>+</sup> ]	300 mM [Na <sup>+</sup> ]	85 mM [Na <sup>+</sup> ]
DNA-pm	3.39	3.11	2.88	2.71	0.17	0.16	0.14	0.14
DNA1mm	2.27	2.09	2.09	1.97	0.11	0.10	0.10	0.10
DNA2mm	2.22	2.06	2.01	1.75	0.11	0.10	0.10	0.09
DNA3mm	1.86	1.83	1.73	1.50	0.09	0.09	0.09	0.08
DNA4mm	1.87	1.59	1.60	1.34	0.09	0.08	0.08	0.07
DNA5mm	1.82	1.53	1.60	1.09	0.09	0.08	0.08	0.05
DNA6mm	1.77	1.79	1.84	1.25	0.09	0.09	0.09	0.06
DNA7mm	1.23	1.29	1.22	0.97	0.06	0.06	0.06	0.05
DNA8mm	1.53	1.48	1.15	0.97	0.08	0.07	0.06	0.05
DNA9mm	1.46	1.23	0.97	-----	0.07	0.06	0.05	----
DNA10mm	1.27	1.03	0.77	-----	0.06	0.05	0.04	----

**Table 4.3(b).** Data tables for calculated values of  $\Delta n$  and  $\Delta \psi$  for Set B duplexes in 85 to 1000 mM  $[\text{Na}^+]$ .

Set B Sample ID	$\Delta n$	$\Delta n$	$\Delta n$	$\Delta n$	$\Delta \psi$	$\Delta \psi$	$\Delta \psi$	$\Delta \psi$
	1000 mM [Na <sup>+</sup> ]	600 mM [Na <sup>+</sup> ]	300 mM [Na <sup>+</sup> ]	85 mM [Na <sup>+</sup> ]	1000 mM [Na <sup>+</sup> ]	600 mM [Na <sup>+</sup> ]	300 mM [Na <sup>+</sup> ]	85 mM [Na <sup>+</sup> ]
DNA- $\mu\text{m}$	3.39	3.11	2.88	2.71	0.17	0.16	0.14	0.14
DNA1mm	2.99	2.53	2.72	2.64	0.15	0.13	0.14	0.13
DNA2mm	2.82	2.37	2.70	2.32	0.14	0.12	0.14	0.12
DNA3mm	2.40	2.15	2.15	1.68	0.12	0.10	0.11	0.08
DNA4mm	2.69	2.63	2.06	2.01	0.13	0.13	0.10	0.10
DNA5mm	1.64	1.59	2.10	1.17	0.08	0.08	0.10	0.06
DNA6mm	1.21	1.19	1.58	0.82	0.06	0.06	0.08	0.04
DNA7mm	0.52	0.57	0.53	0.45	0.03	0.03	0.03	0.02
DNA8mm	0.98	0.96	0.91	0.93	0.05	0.05	0.05	0.05
DNA9mm	0.58	0.67	0.53	-----	0.03	0.03	0.03	-----
DNA10mm	0.55	0.62	0.48	-----	0.03	0.03	0.02	-----



**Figure 4.5:** Histograms of empirically evaluated values for  $\Delta n$  versus increasing number of mismatches for both Set A and Set B duplexes.



**Figure 4.6:** Histograms of empirically evaluated values for  $\Delta\psi$  versus increasing number of mismatches for Set A and Set B duplexes.

#### 4.3.7 **How does the stability of short duplex DNA containing contiguous tandem mismatches vary with $[Na^+]$ ?**

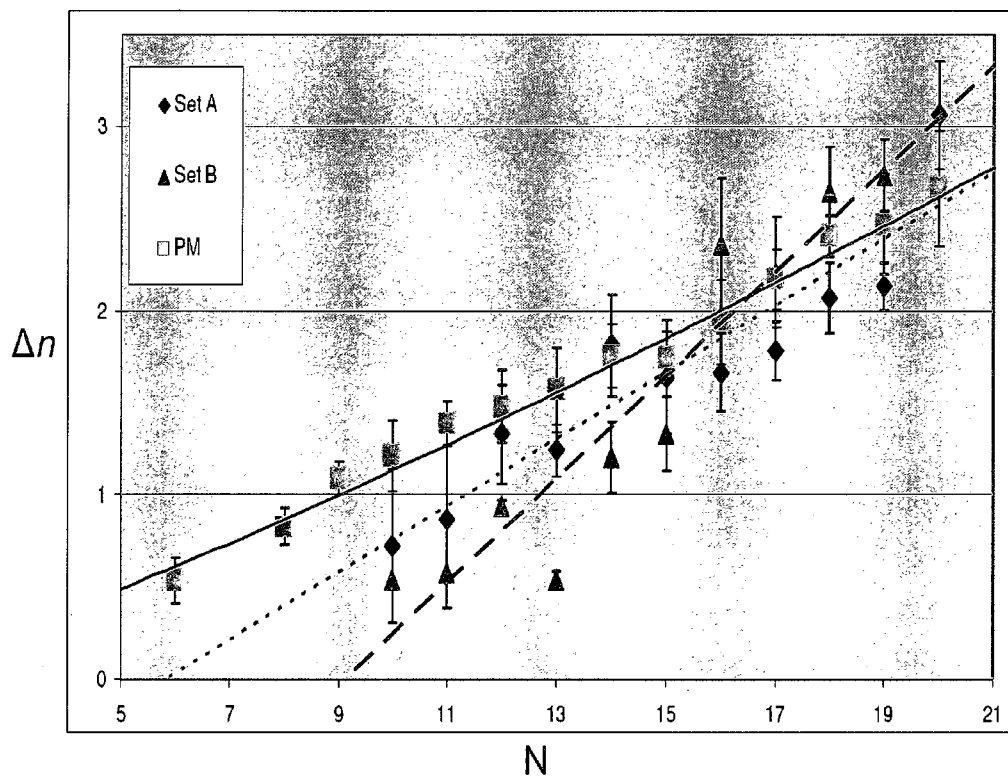
In order to directly compare the relative behaviors of the perfect match duplexes studied in Chapter 3 and with those of the Set A and B molecules, plots of the average values of  $\Delta n$  versus  $N$  and  $\Delta\psi$  versus  $N$  were calculated using the average value of the DSC measured transition enthalpy ( $\Delta H^o$  in Equations (2.31 and 2.33)) for each of the perfect match duplexes from 6 to 20 base pairs in length and the 20 base pair mer molecules comprising Sets A and B. Results are shown in Figures 4.7 and 4.8, respectively. Inspection of this data reveals the following; For perfect matches, Set A and Set B duplexes there is a linear relationship between the total counterion release,  $\Delta n$ , and the number of W/C base pairs,  $N$  (Figure 4.7). Likewise, the counterion release per phosphate,  $\Delta\psi$  versus  $N$  displays a clear linear relationship for the perfect match. However, for Set A and Set B the slopes of these plots are definitely different.

Differences are clearer in the plots of  $\Delta\psi$  versus  $N$  in Figure 4.8. For the perfect match duplexes there is a slight decrease in  $\Delta\psi$  with decreasing  $N$ , except for the smallest duplex of six bases which is slightly smaller. Provided, as suggested above that values of  $\Delta\psi$  are indicative of the fidelity of duplex secondary structure and solvent interactions, behavior of the perfect match duplexes indicates these molecules maintain essentially the same secondary structure as a function of decreasing length, as indicated by the steep slope for the perfect match plot in Figure 4.8. In contrast, the slopes of the plots of  $\Delta\psi$  versus  $N$  for Sets A and B in

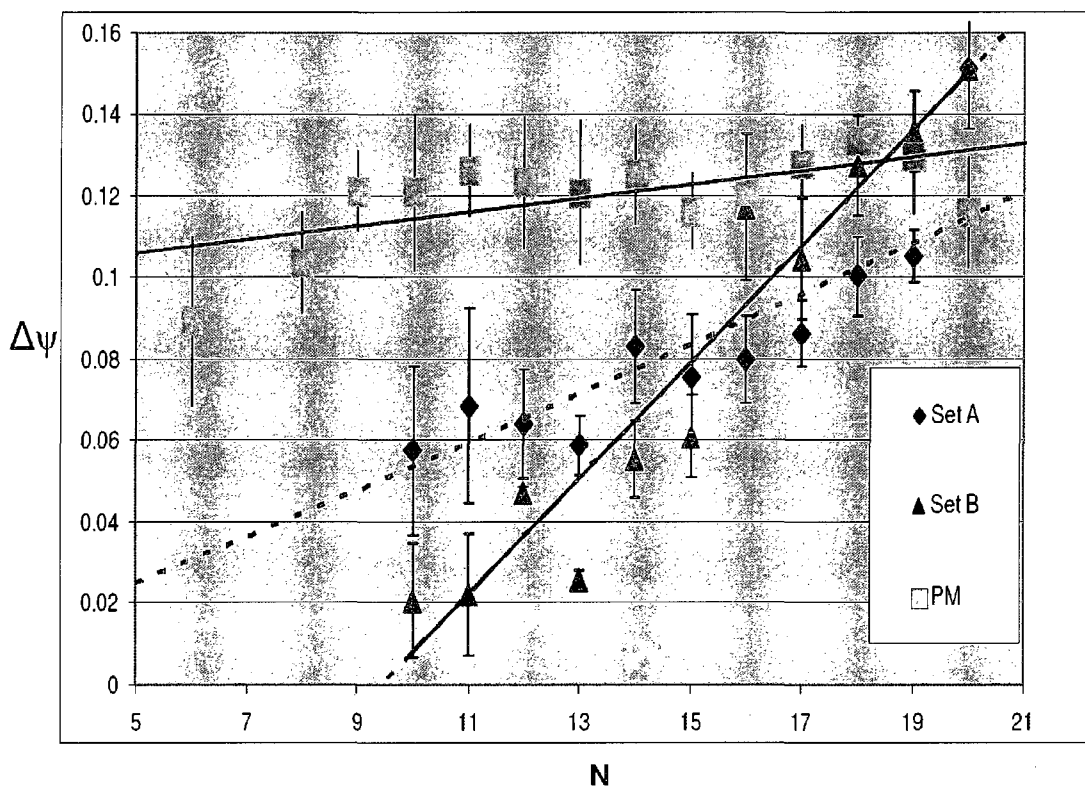
Figure 4.8 are much steeper than observed for the perfect match duplexes and different from one another, revealing the differential effects of tandem mismatches and their relative position and number on duplex stability. Interestingly, for the Set B molecules with tandem mismatches interspersed with W/C base pairs in the duplexes values of  $\Delta\psi$  are equivalent to the perfect match duplex and much greater than for the Set A duplexes until about four mismatches are introduced. Note, the slightly higher value for the perfect match 20-mer control for the Set A and B molecules compared to the 20 base pairs perfect match studied in Chapter 3, due to the difference in the transition enthalpy of the two 20 base pair DNAs. After four mismatches the plots of  $\Delta\psi$  versus N for the Set B molecules decreases with the greatest slope with decreasing N. Again, provided our assertions that values of  $\Delta\psi$  correspond to fidelity of duplex structure, the initial behavior of  $\Delta\psi$  versus N for Set B duplexes suggests that the duplex is able to better tolerate the presence of tandem mismatches 'interspersed' with W/C base pairs compared to on the 'ends' provided there are four or less mismatches. In contrast, on the plot in Figure 4.8 for the Set A molecules, there is a dramatic drop in  $\Delta\psi$  compared to the perfect match with introduction of the first tandem mismatch. This contrasting behavior for the Set A molecules suggests that the initial perturbative effect on duplex structure associated with tandem mismatches is greater when mismatches are on the 'ends'. But with increasing numbers of tandem mismatches the structure does not become as severely eroded when the tandem mismatches are located contiguously on the 'end' compared to being interspersed in the duplex. In essence this behavior



indicates that greater numbers of mismatch base pairs are better tolerated in duplex structure of the Set A compared to Set B duplexes.



**Figure 4.7:** Plot of average values of  $\Delta n$  versus  $N$  for perfect match (PM), Set A and Set B duplexes. Error bars indicate experimental reproducibility. Linear fits are shown. Note the different slopes and the dramatic initial decrease for the Set A molecule with a single mismatch on the end.



**Figure 4.8:** Plot of average values of  $\Delta\psi$  versus  $N$  for perfect match (PM), Set A and Set B duplexes. Error bars indicate experimental reproducibility. Note the different slopes and the dramatic decrease for the Set A molecule with a single mismatch on the end.

The following scenario is consistent with the observations. It is well established that the linear charge density is higher for the duplex than for the single strands. Generally, the higher charge density results in larger amounts of sodium ion binding to the duplex compared to the single strands (29,110). Thus, a net counterion release is observed upon melting. Thus, an increase in bulk  $[\text{Na}^+]$  stabilizes the (duplex) state with the higher charge density, compared to the coiled state resulting in an increased  $T_m$  with increasing with increasing  $[\text{Na}^+]$  up to about 1.3 M (29,44,49,53,56,71). Abundant support for this observation has been obtained from a number of melting studies of a variety of short DNAs of varying

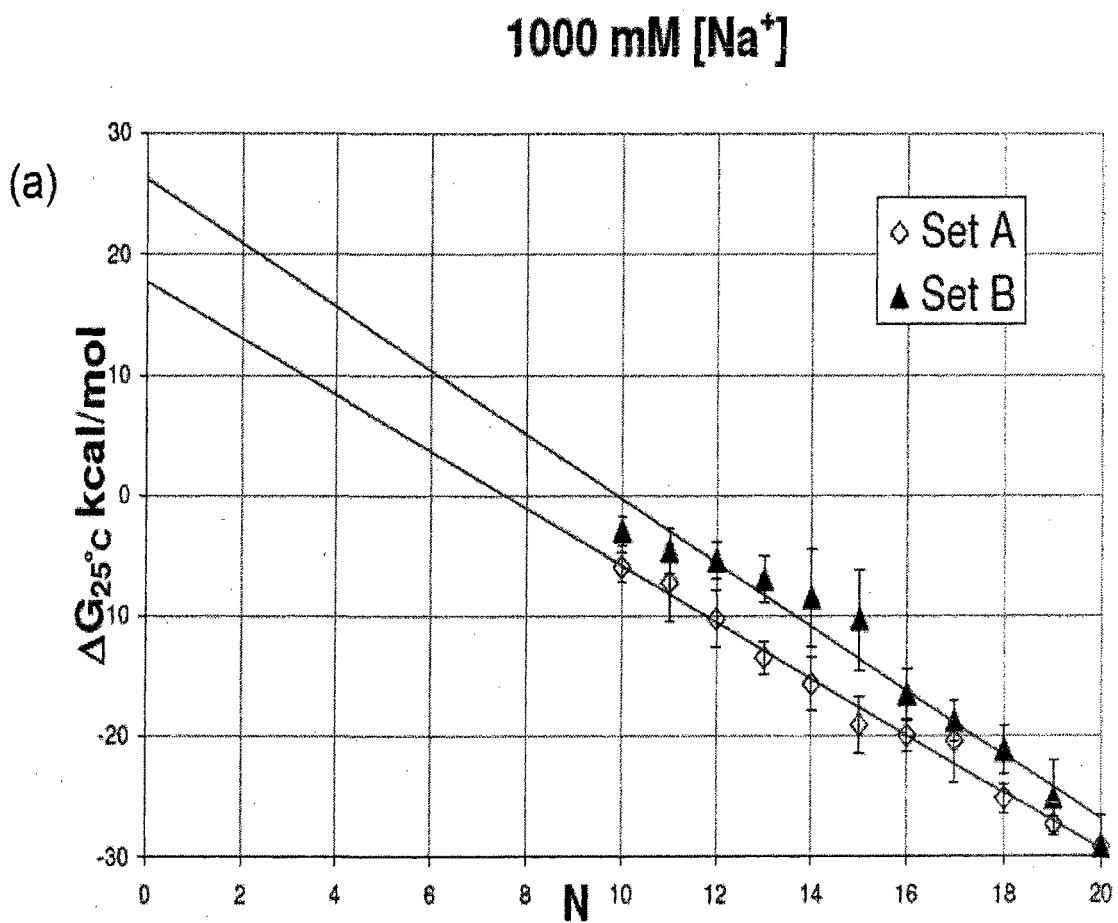
sequence composition (29,58,66,91). Therefore, it seems plausible that perturbations of duplex secondary structure that influence variations (decreases) of the linear charge density, structural integrity and order of the duplex associated with the presence of tandem matches. This would necessarily result in a reduction in the amount of sodium ion binding to the duplex form and therefore result in overall fewer counterions released upon melting as reflected in the values of  $\Delta\psi$ . If so, then the different behaviors of  $\Delta\psi$  versus  $N$  observed for Sets A and B indicate different structural behaviors associated with tandem mismatches depending on the mismatch topology, i.e. whether they exist on the ends (Set A) or interspersed in the duplex (Set B).

From the relative values of  $\Delta n$  versus  $N$  plotted in Figure 4.7 (or  $\Delta\psi$  versus  $N$  in Figure 4.8), it appears as the number of mismatches increase that the Set B duplexes tend to behave more like single strands, while the Set A duplexes apparently can tolerate more mismatches with less single strand-like behavior.

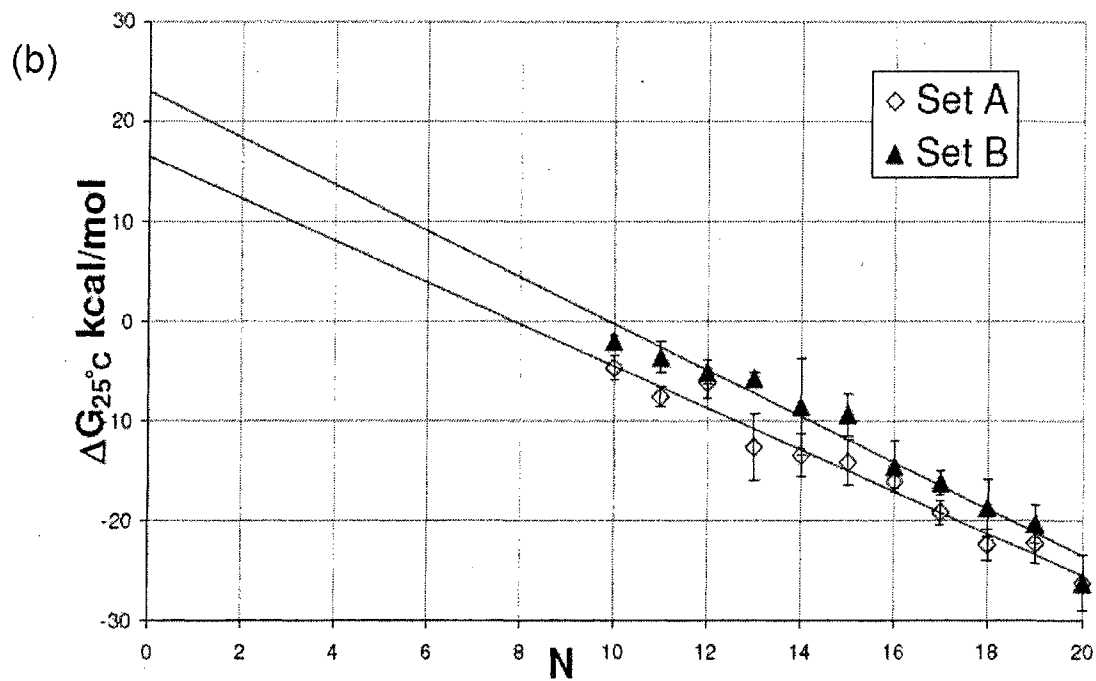
#### **4.3.8 Free-Energy versus Duplex Length, $N$ for duplexes with mismatch base pairs**

Observed destabilization of the Set B molecules as reflected in their  $T_m$  and  $\Delta G_{25^\circ C}^{cal}$  values (Table 4.2 and Figure 4.1) is presumably due to destabilizing

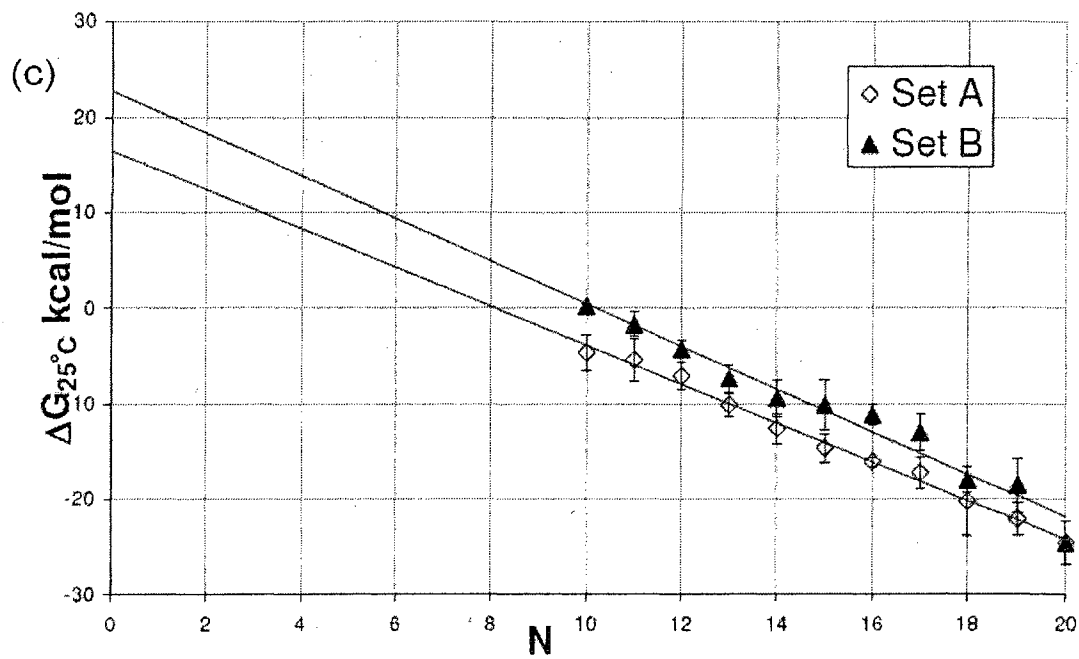
effects associated with mismatch internal loops as reported in the literature (3-4,9-10).



600 mM [Na<sup>+</sup>]



300 mM [Na<sup>+</sup>]



85 mM [Na<sup>+</sup>]

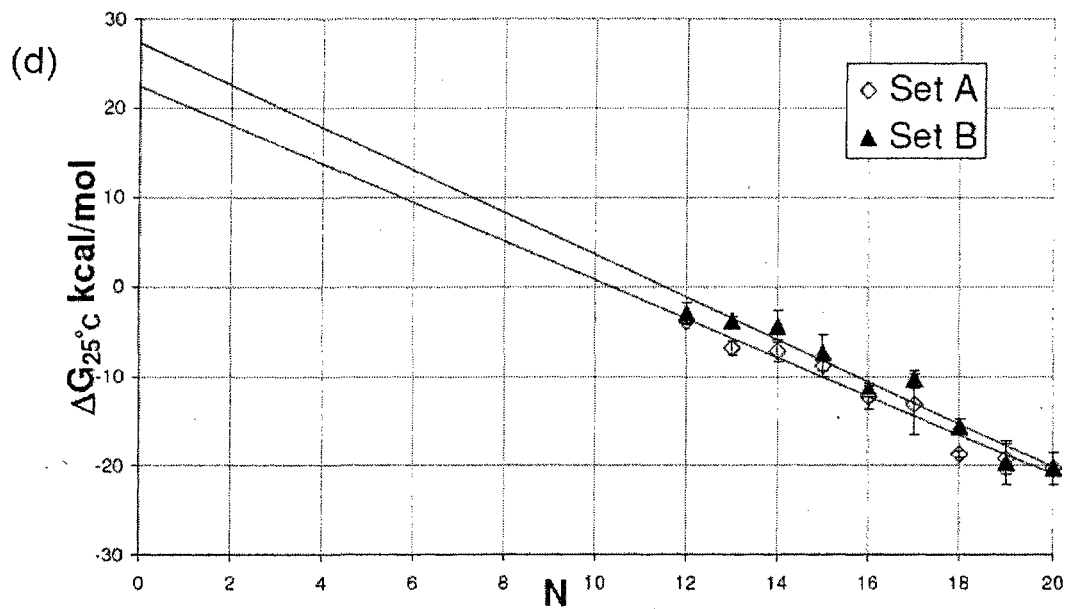


Figure 4.9: Plots of  $\Delta G_{25^{\circ}C}^{cal}$  (kcal/mol) versus N for the DNA molecules of Sets A and B duplexes in four different buffered solvents. As labeled, plots correspond to 1000, 600, 300 and mM [Na<sup>+</sup>], respectively.

**Table 4.4:** Parameters obtained for the linear fits of the plots of Figure 4.9. Slopes, intercepts and linear regression coefficients are listed.

[Na <sup>+</sup> ] mM	System	Linear equation	R <sup>2</sup>
1000	Set A	$y = -2.36x + 17.7$	0.988
	Set B	$y = -2.66x + 26.3$	0.956
600	Set A	$y = -2.10x + 16.5$	0.968
	Set B	$y = -2.23x + 23.0$	0.964
300	Set A	$y = -2.03x + 16.5$	0.992
	Set B	$y = -2.31x + 22.8$	0.966
85	Set A	$y = -2.17x + 22.4$	0.965
	Set B	$y = -2.37x + 27.3$	0.950

Parameters in Table 4.4 are reported as three significant figures. R<sup>2</sup> denotes the linear regression coefficients obtained for the linear fits of the data. Here  $y = \Delta G_{25^{\circ}C}^{cal}$  and  $x = N$  as in Figure 4.9.

Data shown in Table 4.4 reveal clear differences between the duplexes containing mismatches on the ends (Set A) and in the middle (Set B) of the duplex. In general,

values of  $\Delta G_{25^{\circ}C}^{cal}$  are systematically lower for Set A than for Set B duplexes, indicative of higher stability associated with Set A duplexes. That the extrapolated intercept values for the extrapolated for Sets A and Set B sequences differ (Figure 4.9 and Table 4.4) (i.e.  $\Delta G_{25^{\circ}C}^{cal}(N=0)_B > \Delta G_{25^{\circ}C}^{cal}(N=0)_A$ ) suggests stronger duplex destabilizing forces associated with the mismatches in the Set B molecules. Considering mismatches to be duplex perturbants, then mismatches in the middle induce stronger perturbations than when they are on the ends. Results summarized in Table 4.4 for the Set A and Set B molecules, in conjunction with those obtained for the perfect match duplexes as summarized in Figure 3.4 and Table 3.4 can be used to quantitatively determine the strength of this perturbation depending on the mismatch topology, i.e. whether the mismatches reside on the ends or in the middle of the duplex. It should be noted that his analysis is based solely on relative comparisons of experimentally evaluated thermodynamic parameters only and is not dependent or encumbered what-so-ever on the n-n model or inherent assumption thereof.

#### **4.3.9 The relationship between perfect match duplexes and duplexes containing tandem mismatches**

Comparisons of results presented in Chapter 3 for perfect match duplexes of increasing length from 6 to 35 base pairs with those obtained here for the Set A and Set B mismatch DNAs provide a rigorous means for determining the relationship



between perfect match duplexes of a certain length and mismatch duplexes that number of tandem mismatches in different topologies (i.e. contiguous on the ‘ends’, or ‘interspersed’ among W/C bas pairs in the duplex). This analysis does not consider the added influence of the specific sequences of the tandem mismatches, and their associated effects. Rather, the focus is on the tandem mismatch duplex topology. The analysis begins by considering for each sodium ion environment there are three linear expressions that describe the behavior of the experimental free-energy,  $\Delta G_{25^{\circ}C}^{cal}$  versus N. Essentially, in each  $[Na^+]$  equations for the fitted lines in Figure 4.9 as summarized in Table 4.4 are given by,

$$\Delta G_{Set A} = m_{Set A} N + \Delta G_{25^{\circ}C}^{cal} (N = 0)_{Set A} \quad (4.5)$$

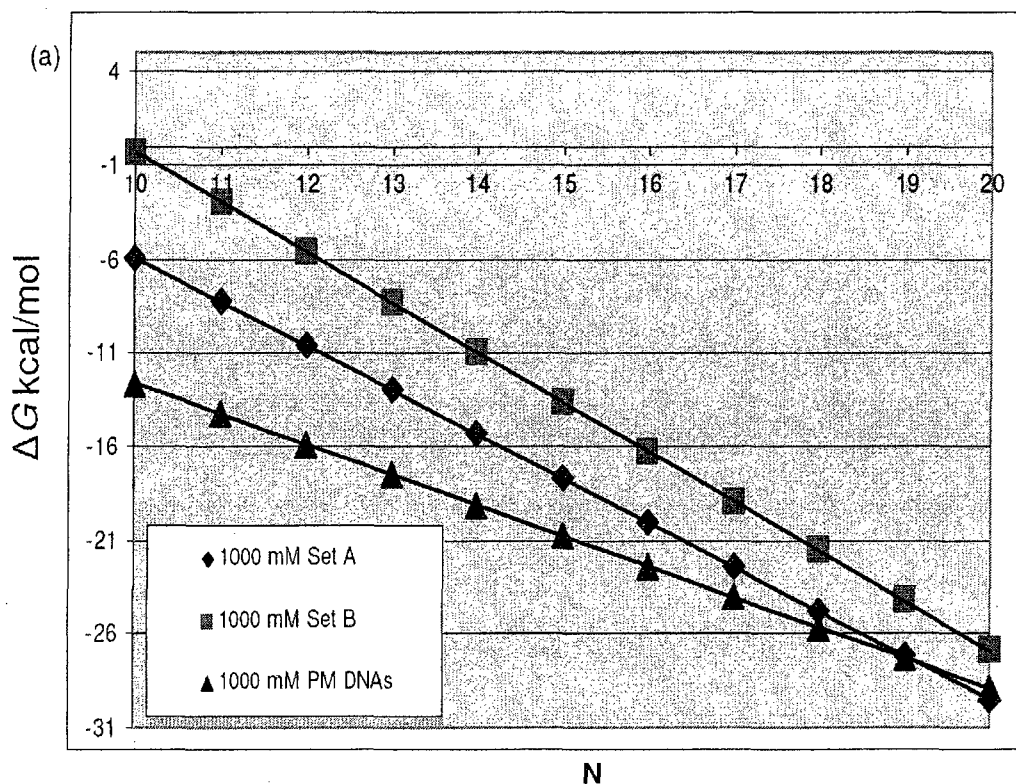
$$\Delta G_{Set B} = m_{Set B} N + \Delta G_{25^{\circ}C}^{cal} (N = 0)_{Set B} \quad (4.6)$$

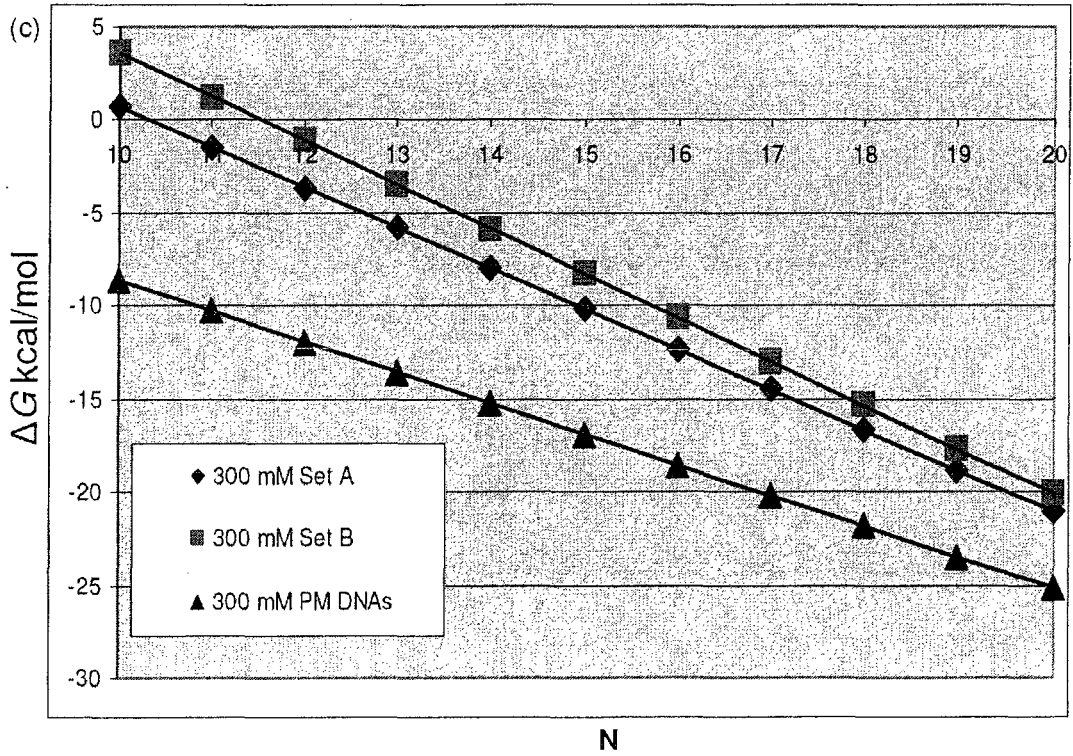
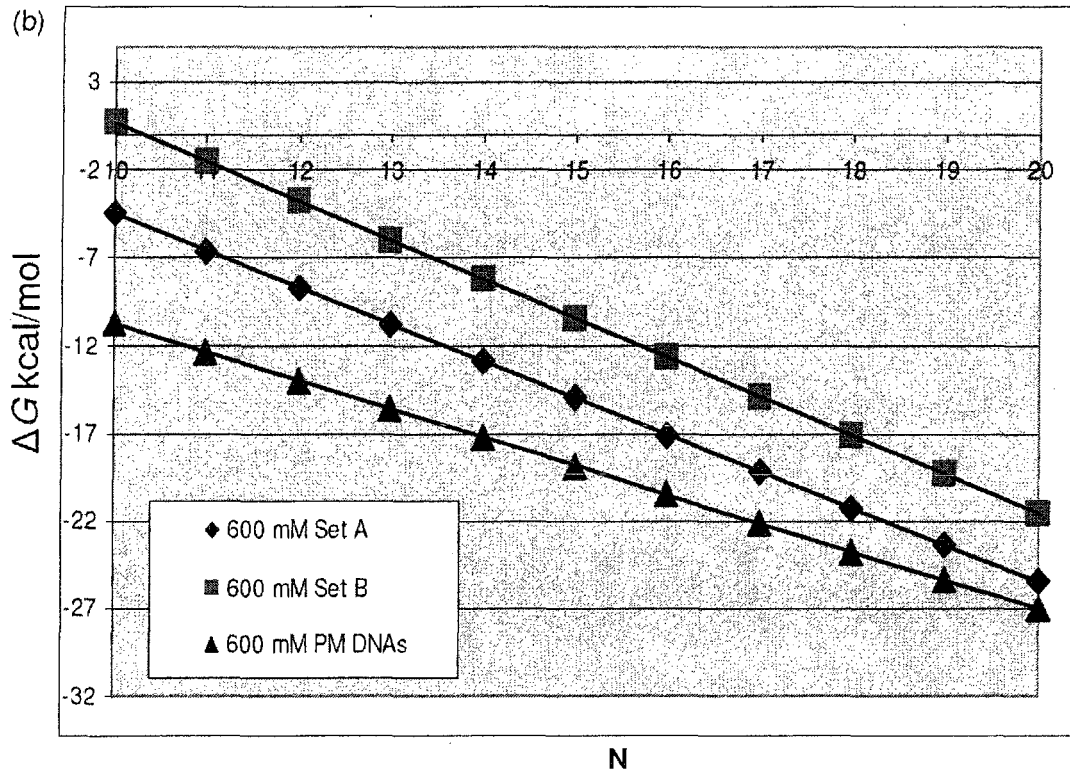
$$\Delta G_{PM DNA} = m_{PM DNA} N + \Delta G^o \quad (4.7)$$

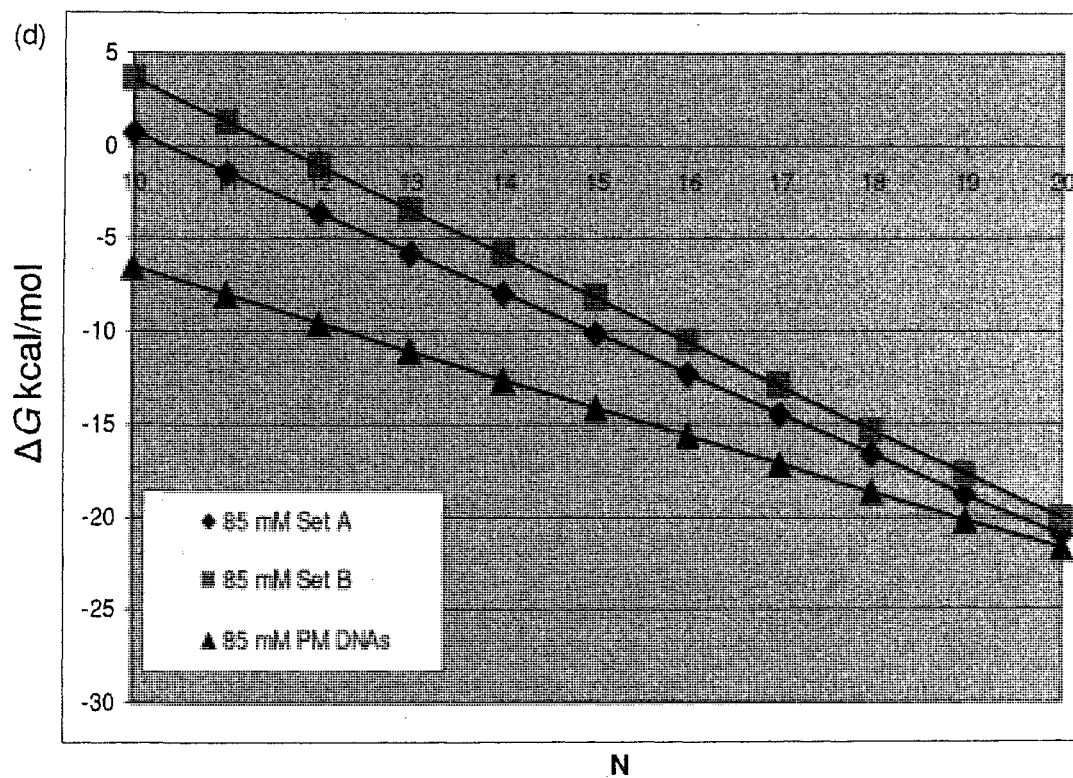
Plots of  $\Delta G$  (kcal/mol) =  $\Delta G_{Set A}$ ,  $\Delta G_{Set B}$  and  $\Delta G_{PM DNA}$  versus N ( $0 < N < 20$  base pairs) are shown in Figure 4.10 (a) - (d) for each  $[Na^+]$  environments. From these plots, the differences between  $\Delta G_{Set A}$  (or  $\Delta G_{Set B}$ ) and  $\Delta G_{PM DNA}$  represents the free energy penalty associated with the number of tandem mismatches at that

N. Now, in Sets A and B there are at each decreasing value of N, and increase in the number of mismatches, n.

The difference at each N value represents the destabilizing “repulsion” associated with the number of mismatch at that particular N value associated with the particular mismatch topology. That is,







**Figure 4.10:** Plot of  $N$  versus  $\Delta G$  (kcal/mol) for Perfect match, Set A and Set B DNA duplexes. Results are shown for  $0 < N < 20$  base pairs in all  $[\text{Na}^+]$  environments labeled (a) - (d).

That is,

$$\Delta G_{\text{PM}}(N) - \Delta G_{\text{Set A}}(N) = \partial \Delta G(N)_A \quad (4.8)$$

where  $\partial \Delta G(N)_A$  is the free energy penalty for mismatches in the Set A topology

when there are  $N$  W/C base pairs. Likewise, for the Set B molecules,

$$\Delta G_{\text{PM}}(N) - \Delta G_{\text{Set B}}(N) = \partial \Delta G(N)_B \quad (4.9)$$

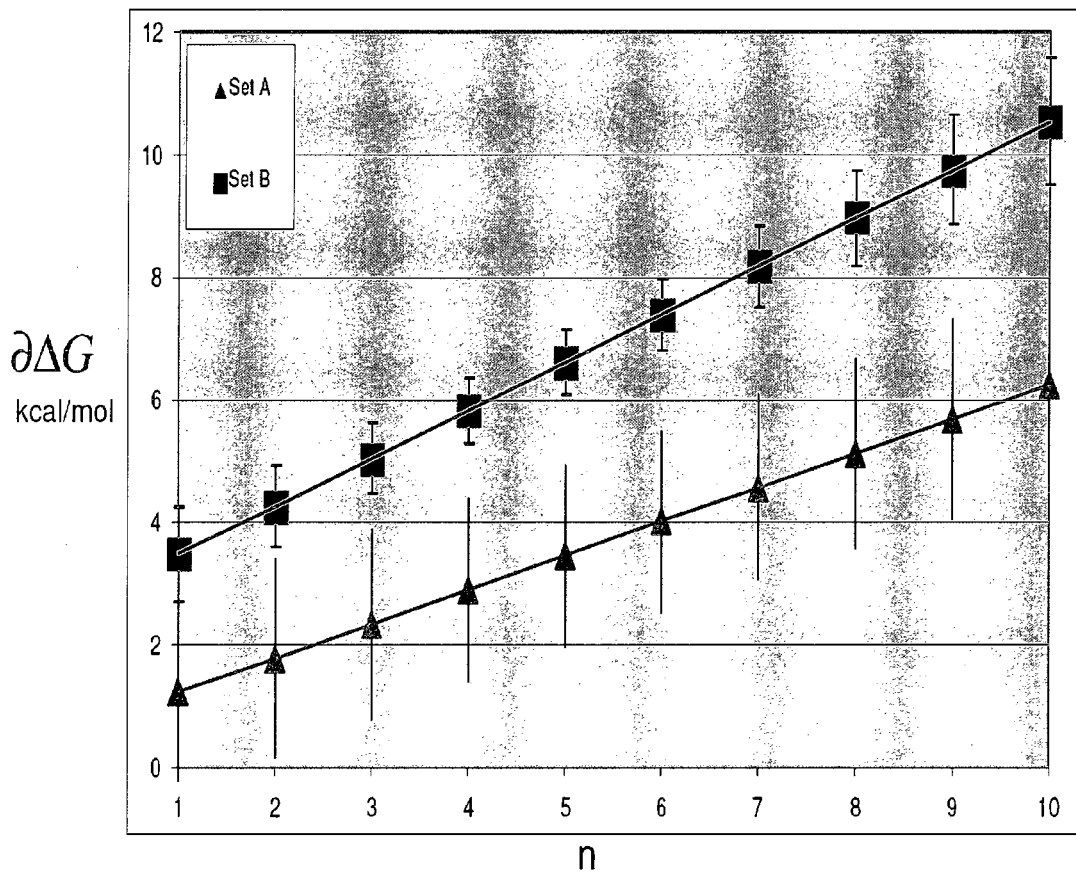
where  $\partial\Delta G(N)_B$  is the free energy penalty or ‘repulsion’ for mismatches in the Set B topology when there are N W/C base pairs. Terms of their linear equation parameters are summarized in Table 4.5.

$$\partial\Delta G_{\text{Set A}}(N) = N(m_{\text{PM}} - m_{\text{Set A}}) + \Delta G_{25^\circ\text{C}}^0 - \Delta G_{25^\circ\text{C}}^{\text{cal}}(N=0)_{\text{Set A}} \quad (4.10)$$

and,

$$\partial\Delta G_{\text{Set B}}(N) = N(m_{\text{PM}} - m_{\text{Set B}}) + \Delta G_{25^\circ\text{C}}^0 - \Delta G_{25^\circ\text{C}}^{\text{cal}}(N=0)_{\text{Set B}} \quad (4.11)$$

For the 20-mer molecules of Sets A and B, at each N there is a corresponding number of mismatches,  $n = 20 - N$ . Thus, the values for  $\partial\Delta G_{\text{Set A}}(N)$  and  $\partial\Delta G_{\text{Set B}}(N)$  in Equations (4.10) and (4.11) can be redefined in terms of the number of mismatches,  $n$  (i.e.  $\partial\Delta G_{\text{Set A}}(n)$  and  $\partial\Delta G_{\text{Set B}}(n)$ ) which are plotted versus  $n$  (in each  $\text{Na}^+$  environment) and mean values are summarized in Figure 4.11 and Table 4.5.



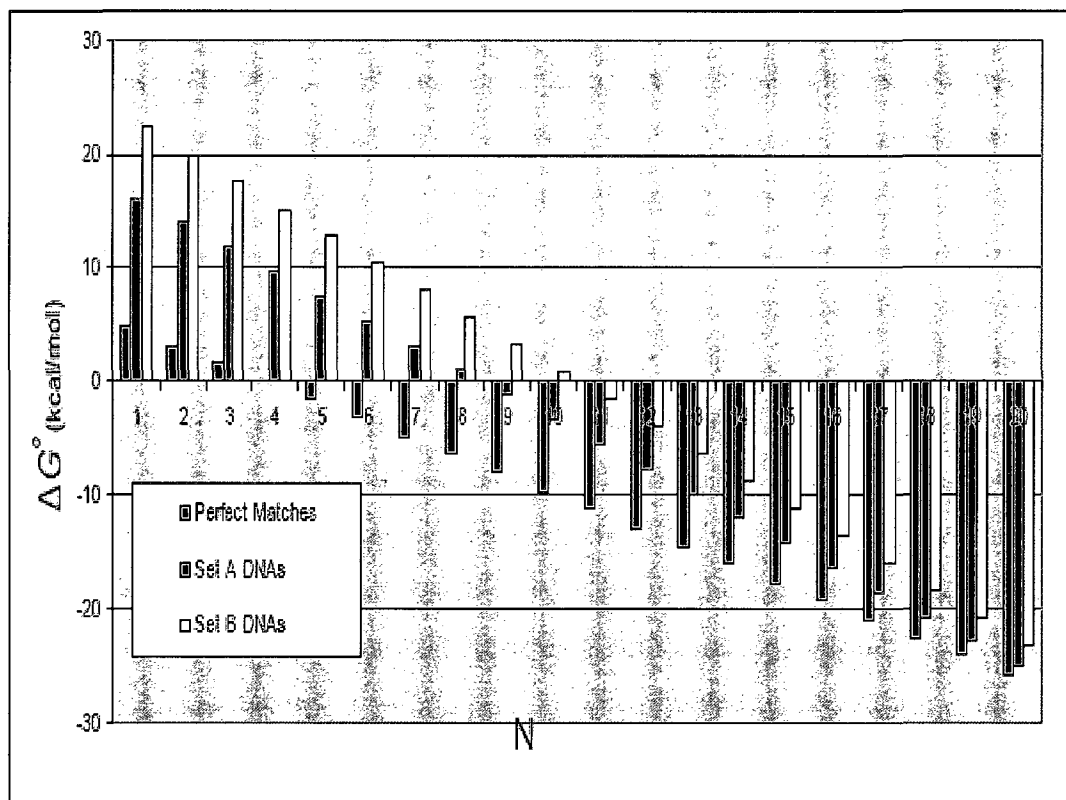
**Figure 4.11:** Plot of average  $\partial\Delta G$  (kcal/mol) for Set A and Set B DNA duplexes versus number of mismatches  $n$  ( $0 < n < 10$ ) in all  $[\text{Na}^+]$  environment (a)-(d).

The plots in Figure 4.11 and the corresponding linear equations in Table 4.5 provide a means for assigning the amount of destabilizing free-energy that should be assigned for a duplex DNA containing W/C base pairs and  $n$  mismatches in the Set A (on the end) or Set B (interspersed) topology. These values can then be used to calculate the stability of short duplex DNAs as will demonstrated below.

**Table 4.5:** Summary of the linear fit parameters for the plots of the mismatch perturbation  $\partial\Delta G$  (kcal/mol) versus  $n$  (base mismatches) at each  $[\text{Na}^+]$ . Slopes and intercepts provided correspond to the linear equations of the individual fits. Regression coefficients were unity for all fitted equations.

$[\text{Na}^+]$ mM	System	Linear equation
1000	Set A	$y = 0.73x + 0.59$
	Set B	$y = 1.03x + 2.01$
600	Set A	$y = 0.47x + 1.51$
	Set B	$y = 0.69x + 3.56$
300	Set A	$y = 0.38x + 1.04$
	Set B	$y = 0.66x + 1.74$
85	Set A	$y = 0.65x + 0.72$
	Set B	$y = 0.73x + 1.88$
Mean equation in all $[\text{Na}^+]$	Set A	$y = 0.56x + 0.67$
	Set B	$y = 0.79x + 2.69$

A comparison of the evaluated free-energies,  $\Delta G_{\text{Set A}}$ ,  $\Delta G_{\text{Set B}}$  and  $\Delta G_{\text{PM DNA}}$  is shown in histogram form in Figure 4.12. This comparison sheds some light on the interrelated helix nucleation and propagation and the relationship between the number of perfect match W/C base pairs that must be formed in a nucleation complex before duplex formation is a favorable process,  $\Delta G_{\text{PM DNA}}$ ,  $\Delta G_{\text{Set A}}$  or  $\Delta G_{\text{Set B}} < 0$ . From the histograms in Figure 4.12 the threshold free-energy for each type of duplex can be clearly seen.



**Figure 4.12:** Histogram of  $\Delta G_{\text{PM DNA}}$ ,  $\Delta G_{\text{Set A}}$  and  $\Delta G_{\text{Set B}}$  versus  $N$  versus for Perfect match, Set A and Set B DNA duplexes, respectively.



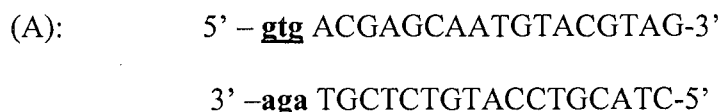
The free-energy threshold for W/C base pairs (where the values of  $\Delta G_{\text{PM DNA}}$ ,  $\Delta G_{\text{Set A}}$  and  $\Delta G_{\text{Set B}}$  versus N cross zero) that must be nucleated before duplex formation is a favorable and spontaneous process differs for the three types of duplexes. Apparently, for the perfect match duplexes, three base pairs are required for nucleation. For the Set A and Set B molecules, eight and 10 W/C base pairs, respectively, are required for duplex nucleation. Results for perfect match duplex are entirely consistent with those reported by Craig *et al.* (21). From kinetic studies of the association of short duplex DNA molecules, they found out that nucleation steps start with a slow and energetically unfavorable helix initiation, followed by the critical nucleus formation. In addition, they found a minimum threshold of three base pairs must be nucleated in order for helix “zippering” or elongation occurs in a spontaneous propagation step.

#### **4.3.10 Applications to probe sequence design**

Results of these studies have direct utility in applications of high-throughput assessments of DNA probe-target sequences that require specific DNA/DNA complexes. In the process of designing probes for use in DNA diagnostic assays based on multiplex hybridization it is essential to know the thermodynamic stabilities of the designed probe-target complexes that will form. Perhaps more importantly it is also critical to know the thermodynamic stabilities of all the potential mismatch complexes that can form from the constituents single strands comprising the designed perfect match complexes. From the experimental

results that have been presented and analyzed for the perfect match and mismatch duplexes with different mismatch topologies, a novel method emerges for quantitative prediction of the contribution of multiple mismatches on the melting thermodynamic of short duplex DNAs. For lack of a better term this new method is referred to as the 'Modified' method. This method does not rely on the n-n model. First, it is demonstrated how the method is applied to explicitly determine the free-energy of a mismatch duplex of complex topology. Then the method is used to predict measured values reported in the literature for a variety of well defined short mismatch duplexes. As will be seen, comparisons between predictions and actual experimental values are quite encouraging.

To demonstrate the 'Modified' method using the relationship derived earlier in this chapter, consider the following mismatch duplexes and calculate their melting free-energies. First, consider the duplex sequence given below labeled A with three mismatches on the end.



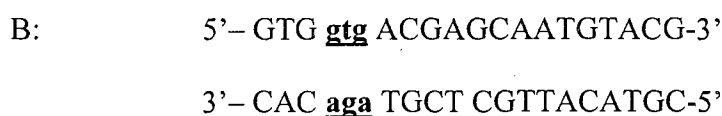
As seen earlier in this Chapter, this a mismatch duplex like those of Set A. Here the number of W/C base pairs is  $N = 17$  and the number of Set A type mismatches

$n_{\text{Set A}} = 3$ . The total free energy of duplex A (in each  $\text{Na}^+$  environment) is given by,

$$\Delta G_{\text{total}}(\text{Na}^+) = N\Delta G_{bp}(\text{Na}^+) + \partial\Delta G_{\text{Set A}}^{\text{repulsion}}(\text{Na}^+, n_{\text{Set A}} = 3) + \Delta G_{nuc}^0(\text{Na}^+) \quad (4.12)$$

where  $N\Delta G_{bp}(\text{Na}^+)$  is the free-energy of the duplex portion comprised only of W/C base pairs, which can be predicted from the explicit sequence using the n-n model parameters as described in Chapter 2. The repulsion term due to mismatches,  $\partial\Delta G_{\text{Set A}}^{\text{repulsion}}(\text{Na}^+, n_{\text{Set A}} = 3)$ , is a function of  $\text{Na}^+$ . These values were determined as a function of n for Set A (and Set B) mismatch topologies and evaluated from the plots in Figure 4.11 and depends on the number of mismatches in the Set A-type topology. Finally,  $\Delta G_{nuc}^0$  is the duplex nucleation free-energy determined from the melting studies in Chapter 2, which was also found to be a function of  $\text{Na}^+$  (see Figure 3.4 and Table 3.4)

Next, consider mismatch duplexes like those of Set B with mismatch base pairs in the duplex flanked on either side by W/C base pairs. For example, the following sequence labeled B,

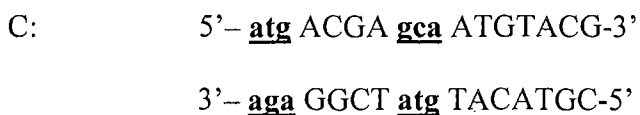


where again the number of W/C base pairs is  $N = 17$  and the number of Set B type mismatches is  $n_{\text{Set B}} = 3$ . As in Equation (4.12) the total free energy of duplex B (in each  $\text{Na}^+$  environment) is given by,

$$\Delta G_{\text{total}}(\text{Na}^+) = N\Delta G_{bp}(\text{Na}^+) + \partial\Delta G_{\text{Set B}}^{\text{repulsion}}(\text{Na}^+, n_{\text{Set B}} = 3) + \Delta G_{\text{mic}}^o(\text{Na}^+) \quad (4.12)$$

where again,  $N\Delta G_{bp}(\text{Na}^+)$  is the free-energy of the duplex portion comprised only of W/C base pairs, in the examples given below assumed to be a contiguous duplex stretch for purposes of using the n-n model predicted stability. The repulsion term associated with the Set B mismatches,  $\partial\Delta G_{\text{Set B}}^{\text{repulsion}}(\text{Na}^+, n_{\text{Set B}} = 3)$ , can be obtained from the plots in Figure 4.11.

Finally, consider a mismatch duplex containing mismatches on the end (Set A-type) and in the interior (Set B-type), for example sequence labeled C below.



Now for this particular topology,  $N = 11$ ,  $n_{\text{Set A}} = 3$ ,  $n_{\text{Set B}} = 3$ , and the total free-energy is given by,

$$\Delta G_{\text{total}}(\text{Na}^+) = N\Delta G_{\text{bp}}(\text{Na}^+) + \partial\Delta G_{\text{Set A}}^{\text{repulsion}}(\text{Na}^+, n_{\text{Set A}}=3) + \partial\Delta G_{\text{Set B}}^{\text{repulsion}}(\text{Na}^+, n_{\text{Set B}}=3) + \Delta G_{\text{ruc}}^0(\text{Na}^+) \quad (4.14)$$

The expressions in Equations (4.11) - (4.13) were used to predict the free-energies of melting for a few mismatch duplexes whose melting free-energies were determined by DSC melting curve analysis and published study (15). The explicit duplex sequences that were considered are shown in Table 4.6, along with their experimentally measured values. As seen there these molecules are examples of Set A and Set B type duplexes as well as mixture of both types. Reported free-energy values (15) were measured in 85 mM Na<sup>+</sup> so calculations were performed with parameters evaluated in this salt environment. The values summarized in Table 4.7 are, the experimental values determined in 85 mM Na<sup>+</sup>,  $\Delta G_{25^\circ\text{C}}^{\text{cal}}$ , values calculated from the n-n model via the MFOLD program and the values calculated using the expressions in Equations (4.11) - (4.14) above and appropriate values obtained from Figure 4.11,  $\Delta G^{(\text{Modified})}$ . In this treatment, sequences of the W/C duplex base pairs are not considered, and there is minimal regard to the influence of specific sequences of the mismatch pairs and their relative influence on duplex stability. The focus is more on the general influence of mismatch topology. Owing to the recently discovered strong sequence dependent stability of some tandem mismatches (89) neglecting their explicit influence could influence the accuracy of the predictions.

For the perfect match duplexes, denoted 4.1, 5.1, 6.1; 7.1, 8.1, 9.1, and 10.1 (to be consistent with their designation in the publication (15), differences between

the measured and predicted free-energies was about 0.45 kcal/mol or an average deviation of 1.8 %. Examination of the comparisons in Table 4.7 reveals the following; For the type A and B mismatches containing a tandem mismatch in the duplex, denoted 4.2, 5.2, 6.2, 7.2, 8.2, 9.2, 10.2, the values predicted by the n-n model using MFOLD are in slightly better agreement (23.2 % average percent deviation) with measured values than the  $\Delta G^{(\text{Modified})}$  predicted values (16.6 % average deviation). That the results are comparable indicates both methods provide less than perfect, but comparable, predictions of experiment results. Finally, results for the mismatch molecules containing multiple mismatches, denoted 4.3, 5.3, 6.3, 7.3, 8.3, 9.3 and 10.3, reveal predicted values of  $\Delta G^{(\text{Modified})}$  and  $\Delta G^{\text{MFOLD}}$  are also in comparable agreement with experiment. Thus for mismatch duplexes containing multiplex tandem mismatches the method developed here provides promising alternative to the n-n model with comparable predictive power at this stage of development.

Table 4.6: DNA Duplexes analyzed from Reference (15).

Sample ID	DNA Sequence	N	n
6.1	5' -TTATGAAGCAACGAAATTAATGAGAA 3'-AATACTTCGTTGCTTTAATTACTCTT	26	0
7.1	5'-AAGAAAGATTAGGACATGAGATTATG 3'-TTCTTTCTAATCCTGTACTCTAATAC	26	0
8.1	5'-TTAGTTAGATACGGAACTGTTAGTTA 3'-AATCAATCTATGCCTTTGACAATCAAT	27	0
9.1	5'-TAGTGTAGTAACGGGAAATCTAAAGTGT 3'-ATCACATCATTGCCCTTTAGATTTCACA	28	0
10.1	5'-TTATGAAATTATGAAATAGTGTAGAT 3'-AATACTTTAATACTTTATCACATCTA	26	0
6.2	5'-TTATGAAGCAACggaAATTAATGAGAA 3'-AATACTTCGTTgagTTAATTACTCTT	24	2
7.2	5'-AAGAAAGATTAGgacATGAGATTATG 3'-TTCTTTCTAATCaggTACTCTAATAC	24	2
8.2	5'-TTAGTTAGATACggaAACTGTTAGTTA 3'-AATCAATCTATGaggTTGACAATCAAT	24	3
9.2	5'-TAGTGTAGTAACgggaAATCTAAAGTGT 3'-ATCACATCATTgaggTTAGATTTCACA	24	4
10.2	5'-TTATGAAATTATgaaAATAGTGTAGAT 3'-AATACTTTAATAaggTTATCACATCTA	24	2
6.3	5'-TTATGAAGCAACGAAATTAATGagaa 3'-AATACTTCGTTGCTTTAATTACTagg	23	3
7.3	5'-AagaaagattAGgacATgagATTATG 3'-TTagTTagAATCgagTAagCTAATAC	18	6
8.3	5'-TTagTTagATACggaAACTGTTagTTA 3'-AagaaATagATgagTTGACAagaaAT	18	9
9.3	5'-TAGTGTAGTAACGGGAAATCTaaagTGT 3'-ATCACATCATTGCCCTTTAGAgagaACA	24	4
10.3	5'-TTATGAAATTATGAAagTgT agat 3'-AATACTTTAATACTT agAaAgagA	23	3

**Table 4.7:** Comparison of experimentally Measured, MFOLD predicted and Modified  $\Delta G_{25^{\circ}C}^{cal}$  kcal/mol for duplexes presented in Table 4.6.

Sample ID	Measured $\Delta G_{25^{\circ}C}^{cal}$ kcal/mol	MFOLD $\Delta G_{25^{\circ}C}^{cal}$ kcal/mol	Modified $\Delta G_{25^{\circ}C}^{cal}$ kcal/mol	Measured-MFOLD $\Delta G_{25^{\circ}C}^{cal}$ kcal/mol	Measured-Modified $\Delta G_{25^{\circ}C}^{cal}$ kcal/mol
6.2	-17.5	-20.9	-24.5	3.4	7.0
7.2	-20.4	-20.5	-24.5	0.1	4.1
8.2	-20.1	-18.7	-23.7	-1.4	3.6
9.2	-20.2	-19.5	-23.0	-0.7	2.8
10.2	-14.5	-17.8	-21.4	3.3	6.9
6.3	-18.7	-24.4	-22.2	5.7	3.5
7.3	-8.6	-6.2	-11.0	-2.4	2.4
8.3	-3.2	-7.0	-10.2	3.8	7.1
9.3	-18.7	-24.3	-23.0	5.6	4.3
10.3	-5.0	-14.6	-16.2	9.6	11.2



## CHAPTER 5

### SUMMARY

In these studies, thermodynamic melting transition parameters of well defined DNA molecules containing perfect match base pairs and tandem mismatch pairs were evaluated directly from DSC measurements. There were three goals (two fundamental and one applied) of this dissertation research. As summarized below each of these goals has been achieved.

**1 Evaluation of the influence of length on short DNA duplex stability as an function of  $[Na^+]$  and definition of the ‘reference’ state and associated the free energy of duplex nucleation.**

Theoretical and experimental investigations of the origins of the nucleation term in the free-energy of duplex formation have been revisited in DSC melting studies of 19 short DNA duplexes ranging in length from 6 to 35 base pairs measured at four different buffered  $[Na^+]$  environments. For this study as presented in Chapter 3, over 650 DSC melting curves were collected and analyzed. From results of these studies of well defined short duplex DNAs, the nucleation or initiation free energy term required for predictions of thermodynamic stability of short linear DNA duplex oligomers was evaluated in four different  $Na^+$ . Following Schurr’s theoretical approach the ‘hypothetical’ nucleation state is comprised of two single strands residing in the same volume as in the duplex but without the

favorable attractive interactions that stabilize the duplex, e.g. hydrogen bonding and stacking. Values of the nucleation free energy,  $\Delta G_{25^{\circ}\text{C}}^{\text{cal}}(N=0)$  at each  $[\text{Na}^+]$  were determined by extrapolating plots of the free-energy values measured by DSC versus the number of base pairs,  $N$ , back to zero base pairs. Relative differences between two single strands in their standard states and the duplex (in its standard state) and solvent displaced during the annealing process was taken into account. This ‘fictitious’ reference state differs fundamentally from current interpretations of the nucleation complex that requires formation of at least one base pair. A novel view of the nucleation step for DNA duplex formation was presented based on information in Figure 3.4. This analysis provided a new view of what has historically been referred to as helix initiation or nucleation parameter, which is salt dependent and suggests an alternative interpretation and mechanism for the nucleation complex in duplex formation. The significance of the formalism developed in this work is that it provides quantitatively accurate estimates in addition to a new way of thinking about the nucleation process in short duplex formation. These results will not only help to improve accuracy of predictions of the thermodynamic stability for short duplex DNAs, they also serve as the basis for comparisons with short duplex DNAs containing a mixture of W/C base pairs and mismatch pairs. As summarized next, determination of the thermodynamic stability of DNA duplexes with systematically increased numbers of mismatch pairs, as a function of  $\text{Na}^+$ , was the second goal of this dissertation research.

**2 Determination of the influence of tandem mismatches on short duplex DNA stability as a function of  $[Na^+]$  and how this depends on the relative placement of the mismatches in the duplex.**

The second major goal of this work was to quantitatively evaluate the thermodynamic melting transition parameters for short duplex DNAs having increasing numbers of mismatches in two different topologies; contiguous on the 'end' of the duplex adjoined by W/C base pairs or; 'interspersed' with W/C base pairs in the duplex. Results are described in Chapter 4. DSC melting curves were collected for 20 short duplex DNAs comprising two sets of 20 base pairs duplexes each containing from one to 10 mismatch pairs in combination with W/C base pairs. From collection of over 850 DSC melting curves of these molecules the influence of mismatch number and topology on short duplex thermodynamic stability was quantitatively determined as a function of  $Na^+$ . Results demonstrated how the presence and distribution of tandem mismatches in short duplexes can differentially affect duplex stabilities. In effect, mismatch pairs on the 'end' of the duplex do not destabilize the duplex as much as mismatches interspersed within the duplex. Dependence of the results on  $Na^+$  indicates differences in structural perturbations induced by increasing numbers of mismatches, which is strongly dependent on the mismatch topology. Comparison of experimental results with predictions using the n-n model nearly universally revealed serious deficiencies of the n-n to accurately predict the thermodynamic stability short duplex DNAs containing multiple mismatches. This must certainly be due to improper

consideration of tandem mismatches in the n-n model and/or inappropriate application of the n-n model for duplexes containing multiple mismatches because in such cases, interactions longer than nearest-neighbors might be present. In the n-n model tandem mismatches are accounted for by a single loop entropy term  $\Delta G_{(loop)}(n)$  that depends only the number of contiguous mismatches that comprising the loop, not their relative position or topology or sequence. These results underscore the urgent need for alternatives to the n-n model in order to accurately account for and evaluate the contributions of tandem mismatches to the thermodynamic stability of short duplex DNAs. The results of Chapter 3 for short perfect match duplex DNAs and those presented in Chapter 4 for the mismatch duplexes can be combined and collectively analyzed to provide quantitative evaluations of the relative destabilizing contributions of tandem mismatch depending on their relative number and placement in the duplex. This analysis is presented at the end of Chapter 4 and provides an alternative to the n-n model for calculating the contributions of multiple mismatches to the thermodynamic stability of short duplex DNAs, as summarized next.

### **3 Application of the results to probe sequence design for multiplex hybridization reactions.**

The final goal of the dissertation work was to utilize the product of the combined work (Chapter 3 and Chapter 4) performed in achieving the first two fundamental goals in the third applied goal to utilize the result to improve the

quantitative accuracy of probe design for multiplex hybridization reactions. In multiplex hybridization reactions containing many probes that must hybridize with high fidelity and simultaneously to many independent target strands it is critical to quantitatively understand the thermodynamic stability of duplexes containing multiple mismatches. If stable enough such mismatch duplexes, or so-called cross-hybrids, can contribute significantly to the assay hybridization signal resulting in increased noise; or worse, detection of false positives or false negatives and flawed diagnoses.

At the end of Chapter 4, the evaluated “corrections” for mismatch repulsion were used to predict results of independent experiments of mismatch duplexes reported in the literature (15) with reasonable good success in most cases. However, some predictions were better than others and suggested that the influence of sequence dependence of the mismatches might be at issue. It has recently been shown in DSC studies of short duplexes containing isolated single tandem mismatches with different well defined sequences that actual identity of tandem mismatches greatly influences their effect on duplex thermodynamic stability. The actual range of the effect depends not only on the sequence identity of the mismatch but also the W/C base pairs flanking the mismatch (89). Sequences of our molecules were not systematically varied to investigate explicit effects of sequence. But this should certainly be a topic for future study. Overall, these results provide a new basis for an improved treatment and evaluation of the

thermodynamic contributions of multiple mismatches to short duplex DNA stability.

1. Fish, D. J., Horne, M. T., Searles, R. P., Brewood, G. P., and Benight, A. S. (2007) Multiplex SNP discrimination. *Biophys. J.* 92, L89-L91.
2. Horne, M. T., Fish D. J., and Benight, A. S. (2006) Statistical thermodynamics and kinetics of DNA multiplex hybridization reactions. *Biophys. J.* 91, 4133-4153.
3. Allawi, H. T., and SantaLucia, J., Jr. (1998) Nearest neighbor thermodynamic parameters for internal G·A mismatches in DNA. *Biochemistry* 37, 2170–2179.
4. Allawi, H. T., and SantaLucia, J., Jr. (1997) Thermodynamics and NMR of internal G·T mismatches in DNA. *Biochemistry* 36, 10581–10594.
5. Benight, A. S., Pancoska, P., Owczarzy, R., Vallone, P. M., Nesetril, J., and Riccelli, P. V. (2001) Calculating sequence-dependent melting stability of duplex DNA oligomers and multiplex sequence analysis by graphs. *Methods Enzymol.* 340, 165-192.
6. Wartell, R. M., and Benight, A. S. (1985) Thermal denaturation of DNA molecules: a comparison of theory with experiment. *Physics Reports* 126, 67-107.
7. Li, Y., Zon, G., and Wilson, W. D. (1991) Thermodynamics of DNA duplexes with adjacent G·A mismatches. *Biochemistry* 30, 7566-7572.
8. SantaLucia, J., Jr. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U.S.A.* 95, 1460–1465.
9. SantaLucia, J., Jr., Allawi, H. T., and Seneviratne, P. A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* 35, 3555–3562.
10. SantaLucia, J., Jr., and Hicks, D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* 33, 415–440.
11. Owczarzy, R., Vallone, P. M., Gallo, F. J., Paner, T. M., Lane, M. J., and Benight, A. S. (1997) Predicting sequence-dependent melting stability of short duplex DNA oligomers. *Biopolymers* 44, 217–239.

12. Bloomfield, V. A., Crothers, D. M., and Tinoco, I., Jr. (1974) *Physical Chemistry of Nucleic Acids*, Harper and Row, New York, U.S.A. pp 294-301.
13. DeVoe, H., and Tinoco, I., Jr. (1962) The stability of helical polynucleotides: base contributions. *J. Mol. Biol.* 4, 500-517.
14. Zimm, B. H. (1960). Theory of melting of the helical form in double chains of the DNA type. *J. Chem. Phys.* 33, 1349-1356.
15. Fish, D. J., Horne, T. M., Brewood, G. P., Goodarzi, J. P., Alemayehu, S., Bhandiwad, A., Searles, R. P., and Benight, A. S. (2007) DNA multiplex hybridization on microarrays and thermodynamic stability in solution: a direct comparison. *Nucleic Acids Res.* 35, 7197-7208.
16. Ricelli, P. V., Vallone, P. M., Kashin, I., Faldasz, B. D., Lane, M. J., and Benight, A. S. (1999) Thermodynamic, spectroscopic, and equilibrium binding studies of DNA sequence context effects in six 22-base pair deoxyoligonucleotides. *Biochemistry* 38, 11197-11208.
17. Bommarito, S., Peyert, N., and SantaLucia, J. Jr. (2000) Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res.* 28, 1929-1934.
18. Applequist, J. (1963) Theory of the concentration and chain length on helix-coil equilibrium in two stranded nucleic acids. *J. Chem. Phys.* 39, 2719-2721.
19. Flory, P. J., and Miller, G. W. (1966) A general treatment of helix-coil equilibria in macromolecular systems. *J. Mol. Biol.* 15, 284-297.
20. Wetmur, J. G., and Davidson, N. (1968) Kinetics of renaturation of DNA. *J. Mol. Biol.* 31, 349-370.
21. Craig, M. E., Crothers, D. M., and Doty, P. (1971) Relaxation kinetics of dimer formation by self complementary oligonucleotides. *J. Mol. Biol.* 1462, 383-401.
22. Sugimoto, N., Nakano, S., Yoneyama, M., and Honda, K. (1996) Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic Acid Res.* 4, 4501-4505.
23. Allawi, H. T., and SantaLucia, J., Jr. (1998) Thermodynamics of internal C.T mismatches in DNA. *Nucleic Acids Res.* 26, 2694-2701.



24. Allawi, H. T., and SantaLucia, J., Jr. (1998) Nearest-neighbor thermodynamics of internal A:C mismatches in DNA: sequence dependence and pH effects. *Biochemistry* 37, 9435-9444.
25. Glickman, B. W., Horsfall, M. J., Gordon, A. J., and Burns, P. A. (1987) Nearest neighbor affects G:C to A:T transitions induced by alkylating agents. *Environ. Health Perspect.* 76, 29-32.
26. Wartell, R. W, and Montroll, E. W. (1972) Equilibrium denaturation of natural and of periodic synthetic DNA molecules. *Adv. Chem. Phys.* 22, 129-203.
27. Tikhomirova, A., Beletskaya, I. V., and Chalikian, T. V. (2006) Stability of DNA duplexes containing GG, CC, AA, and TT mismatches. *Biochemistry* 45,10563-10571.
28. Walter, A. E., Wu, M., and Turner, D. H. (1994) The stability and structure of tandem GA mismatches in RNA depend on closing base pairs. *Biochemistry* 33, 11349-11354.
29. Anderson, C. F., and Record, M. T., Jr. (1995) Salt-nucleic acid interactions. *Annu. Rev. Phys. Chem.* 46, 657-700.
30. Good, N. E., Winget, G. D., Winter, W., Connolly, T. N., Izawa, S., and Singh, R. M. M. (1966) Hydrogen ion buffers for biological research. *Biochemistry* 5, 467-477.
31. <http://www.idtdna.com/SCITOOLS/scitools.aspx>
32. Weckx, S., Carlon, E., Vuyst, L. D., and Hummelen, P. V. (2007) Thermodynamic Behavior of Short Oligonucleotides in Microarray Hybridizations Can Be Described Using Gibbs Free Energy in a Nearest-Neighbor. *J. Phys. Chem. B.* 111, 13583-13590.
33. Poland, D. C., and Scheraga, H. A. (1970) *Theory of the Helix-Coil Transition*, Academic Press, New York, NY.
34. Henegariu, O., Heerema, N. A., Dlouhy, S. R., Vance, G. H., Vogt, P. H. (1997) Multiplex PCR: critical parameters and step-by-step protocol. *BioTechniques* 23, 504-511.
35. Maniatis, T. F., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 3, 150-201.

36. Battersby, T. R., Albalos, M., and Friesenhahn, M. J. (2007) An unusual mode of DNA duplex association: Watson-Crick interaction of all-purine deoxyribonucleic acids. *Chem. Biol.* 14, 525-531.
37. Goldberg, H. A., and Warner, K. J. (1997) The staining of acidic proteins on polyacrylamide gels: enhanced sensitivity and stability of "Stains-all" staining in combination with silver nitrate. *Anal. Biochem.* 251, 227-233.
38. Andrews, A. T. (1988) *Electrophoresis: Theory, Techniques, and Biochemical Applications*, Oxford University Press, NY., U.S.A., 2, 38-42.
39. Amzel, L. M. (1997). Loss of translational entropy in binding, folding, and catalysis. *Proteins* 28, 144-149.
40. Schurr, J. M. Chapter 9: Polyanion models of nucleic acid-metal ion interactions. In *Nucleic Acid-Metal ion interactions*, Hud, V. N. *Royal Society of Chemistry*, Cambridge. (2008) ISBN: 978-0-85404-1995-4. 307-349,
41. Schurr, J. M., and Fujimoto, B. S. (2002) Extensions of counterion condensation theory. I. Alternative geometries and finite salt concentration. *Biophys. Chem.* 101-102, 425-445.
42. Gilson, M. K., Given, J. A., Bush, B., L and McCammon., J. A. (1997). The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys. J.* 72, 1047-1069.
43. Fritz., M. P. (1974) Thermodynamics of the Helix-Coil Transition of (dG-dC) Oligomers. *Eur. J. Biochem.* 42, 495-504.
44. Owczarzy, R., Moreira, B. G., You Y., Behlke, M. A., and Walder, J. A. (2008) Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochemistry* 47, 5336-5353.
45. Peyret, N., Seneviratne, P. A., Allawi, H. T. and SantaLucia, J., Jr. (1999) Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. *Biochemistry* 38, 3468-3477.
46. Riccelli, P. V., Hall, T. S., and Pancoska, P. (2003) DNA sequence, context and multiplex hybridization reactions: Melting studies of heteromorphic duplex DNA complexes. *J. Am. Chem. Soc.* 125, 141-150.
47. Benight, A. S., Gallo, F. J., Paner, T. M., Bishop, K. D., Faldasz, B. D., and Lane, M. J. (1995) Sequence context and DNA reactivity: application to sequence-specific cleavage of DNA. *Adv. Biophys. Chem.* 5, 1-55.

48. Allawi, H. T. and SantaLucia, J., Jr. (1998) NMR solution structure of a DNA dodecamer containing single G.T mismatches. *Nucleic Acids Res.* 26, 4925-4934.
49. Korolev, N., Lyubartsev, A. P., and Nordenskiöld, L. (1998) Application of polyelectrolyte theories for analysis of DNA melting in the presence of Na<sup>+</sup> and Mg<sup>2+</sup> ions. *Biophys. J.* 75, 3041–3056.
50. Stigter, D. (1995) Evaluation of the counterion condensation theory of polyelectrolytes. *Biophys. J.* 69, 380-388.
51. Marmur, J., and Doty, P. (1962) Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *J. Mol. Biol.* 5, 109-118.
52. Manning, G. S. (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* 11, 179–246.
53. Völker, J., Klump, H. H., Manning, G. S., and Breslauer, K. J. (2001) Counterion association with native and denatured nucleic acids: an experimental approach. *J. Mol. Biol.* 310, 1011–1025.
54. Tomac, S., Sarkar, M., Ratilainen, T., Wittung, P., Nielsen, P., Borden, B., and Graslund, A. (1996) Ionic effects on the stability and conformation of peptide nucleic acid complexes. *J. Am. Chem. Soc.* 118, 5544–5552.
55. Uhlenbeck, O. C., Borer, P. N., Dengler, B., and Tinoco, I., Jr. (1973) Stability of RNA hairpin loop: A<sub>6</sub>-C<sub>m</sub>-U<sub>6</sub>, *J. Mol. Biol.* 73, 483–496.
56. Schildkraut, C. (1965) Dependence of the melting temperature of DNA on salt concentration. *Biopolymers* 3, 195-208.
57. Owczarzy, R., You, Y., Moreira, B. G., Manthey, J. A., Huang, L., Behlke, M. A. and Walder, J. A. (2004) Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures. *Biochemistry* 43, 3537–3554.
58. Khan, M. O., Mel'nikov, S. M., and Jönsson, B. (1999) Anomalous Salt Effect On DNA conformation: Experiment and Theory. *Macromolecules* 32, 8836-8840.
59. Milev, S., Bosshard, H. R. and Jelesarov, I. (2005) Enthalpic and entropic effects of salt and polyol osmolytes on site-specific protein-DNA association: the integrase Tn916-DNA complex. *Biochemistry* 44, 285-293.

60. Blake, R. D., and Delcourt, S. G. (1998) Thermal Stability of DNA. *Nucleic Acids Res.* 26, 3323-3332.
61. Mandell, K. E., Vallone, P. M., Owczarzy, R., Riccelli, P. V., and Benight, A. S. (2006) Studies of DNA dumbbells VIII. Melting analysis of DNA dumbbells with dinucleotide repeat stem sequences. *Biopolymers* 82, 199-221.
62. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. M., Neilson, T. and Turner, D. H. (1986) Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. U.S.A.* 83, 9373-9377.
63. Peyret, N., and SantaLucia, J., Jr. HYTHER, Version 1.0. Wayne State University, Detroit, MI. Available at <http://ozone3.chem.wayne.edu/Hyther/html>
64. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406-3415.
65. Bevington, P. R., and Robinson, D. K. (1992) *Data Reduction and Error Analysis for the Physical Sciences*, WCB/McGraw-Hill, Boston, MA.
66. Nakano, S., Fujimoto, M., and Hara, H. (1999) Nucleic acid duplex stability: influence of base composition on cation effects. *Nucleic Acids Res.* 27, 2957-2965.
67. Petruska, J., and Goodman, M. F. (2008) Enthalpy-Entropy Compensation in DNA Melting Thermodynamics. *J. Biol. Chem.* 270, 746-750.
68. Lafont, V., Armstrong, A. A., Ohtaka, H., Kiso, Y., Mario, A. L., and Freire, E. (2007) Compensating enthalpic and entropic changes hinder binding affinity optimization. *Chem. Biol. Drug Des.* 69, 413-422.
69. Starikov, E. B., and Norden, B. (2007) Enthalpy-entropy compensation: a phantom or something useful? *J. Phys. Chem. B*, 111, 14431-14435.
70. Wu, P., Nakano, S., and Sugimoto, N. (2002) Temperature dependence of thermodynamic properties for DNA/DNA and RNA/DNA duplex formation. *Eur. J. Biochem.* 269, 2821-2830.
71. Tan, Z. J., and Chen, S. J. (2006) Nucleic acid helix stability: effects of salt concentration, cation valence and size, and chain length. *Biophys. J.* 90, 1175-1190.

72. Watkins, N. E. Jr., and SantaLucia, J., Jr. (2005) Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes. *Nucleic Acids Res.* 33, 6258-6267.
73. Breslauer, K. J., Frank, R., Blocker, H., and Marky, L. A. (1986) Predicting DNA duplex Stability from base sequence. *Proc. Natl. Acad. Sci. U.S.A.* 83, 3746-3750.
74. Sugimoto, N., Nakano, S., Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., Yoneyama, M., and Sasaki, M. (1995) Thermodynamic Parameters To Predict Stability of RNA/DNA Hybrid Duplexes. *Biochemistry* 34, 11211-11216.
75. Tan, Z. J., and Chen, S. J. (2006) Ion-mediated nucleic acid helix-helix interactions. *Biophys. J.* 91, 518-536.
76. Antao, V., and Tinoco, I. Jr. (2008) Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucleic Acids Res.* 20, 819-824.
77. Gray, D. M. (1997) Derivation of nearest-neighbor properties from data on nucleic acid oligomers. II. Thermodynamic parameters of DNA:RNA hybrids and DNA duplexes. *Biopolymers* 42, 795-810.
78. Mendez, J., and Stillman, B. (2003) Perpetuating the double helix: molecular machines at eukaryotic DNA replication origins. *Bioessays* 25, 1158-1167
79. Koehler, R. T., and Peyret, N. (2005) Effects of DNA secondary structure on oligonucleotide probe binding efficiency. *Comput. Biol. Chem.* 29, 393-397.
80. Koehler, R. T., and Peyret, N. (2005) Thermodynamic properties of DNA sequences: characteristic values for the human genome. *Bioinformatics* 21, 3333-3339.
81. Bourdelat-Parks, B. N., and Wartell, R. M. (2004) Thermodynamic stability of DNA tandem mismatches. *Biochemistry* 43, 9918-9925.
82. Lando, D. Y., and Fridman, A. S. (2001) Role of small loops in DNA melting. *Biopolymers* 58, 374-389.
83. Leung, D. H., Bergman, R. G., and Raymond, K. N. (2008) Enthalpy-entropy compensation reveals solvent reorganization as a driving force for supramolecular encapsulation in water. *J. Am. Chem. Soc.* 130, 2798-2805.

84. Sharp, K. (2001) Entropy-enthalpy compensation: fact or artifact? *Protein Sci.* 10, 661-667.
85. Song-Hua, K., and Wartell, R. M. (1996) The thermal stability of DNA fragments with Tandem mismatches at a d(CXYG).d(CY'X'G) site. *Nucleic Acid Res.* 24, 707-712.
86. Arnold, F. H., Wolk, S., Cruz, P., and Tinoco, I., Jr. (1987) Structure, dynamics, and thermodynamics of mismatched DNA oligonucleotide duplexes d(CCCAGGG)<sub>2</sub> and d(CCCTGGG)<sub>2</sub>. *Biochemistry* 26, 4068-4075.
87. Aboul-ela, F., Koh, D., Tinoco, I. Jr., and Martin, F. H. (1985) Base-base mismatches. Thermodynamics of double helix formation for dCA<sub>3</sub>XA<sub>3</sub>G + dCT<sub>3</sub>YT<sub>3</sub>G (X, Y = A,C,G,T). *Nucleic Acids Res.* 13, 4811-4824.
88. Ebel, S., Brown, T., and Lane, A. N. (1992) Very stable mismatch duplexes: structural and thermodynamic studies on tandem G.A mismatches in DNA. *Biochemistry* 31, 12083-12086.
89. Horne, M, T. (2008) Theoretical Investigations and Microarray Applications of DNA Multiplex Hybridization Reactions, Ph.D. Thesis, Chapters 1,6 and 7, Portland State University, Portland, OR.
90. Dimitrov, R., and Zuker, M. (2004) Prediction of hybridization and melting for double-stranded nucleic acids. *Biophys. J.* 87, 215-226.
91. Peyret, N. (2000) Prediction of Nucleic Acid Hybridization: Parameters and Algorithms, Ph.D. Thesis, Section 5.4.2, pp 128, Wayne State University, Detroit, MI.
92. Rychlik, W., Spencer, W. J., and Rhoads, R. E. (1990) Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Res.* 18, 6409-6412.
93. Sugimoto, N., Nakano, M., and Nakano, S. (2000) Thermodynamics-structure relationship of single mismatches in RNA/DNA duplexes. *Biochemistry* 39, 11270-11281.
94. Owczarzy, R. (2005) Melting temperatures of nucleic acids: discrepancies in analysis. *Biophys. Chem.* 117, 207-215.
95. Wilson, R. W., Rau, D. C., and Bloomfield, V. A. (1980) Comparison of polyelectrolyte theories of the binding of cations to DNA. *Biophys. J.* 30, 317-326.

96. Manning, G. S. (1972) On the application of polyelectrolyte “limiting laws” to the helix-coil transition of DNA. II. The effect of  $Mg^{+2}$  counterions. *Biopolymers* 11, 951–955.
97. Breslauer, K. J. (1994) Extracting thermodynamic data from equilibrium melting curves for oligonucleotide order-disorder transitions. *Methods Mol. Biol.* 26, 347–372.
98. Zhang, L., Wu, C., Carta, R., and Zhao, H. (2006) Free energy of DNA duplex formation on short oligonucleotide microarrays. *Nucleic Acids Res.* 35, e18.
99. Owczarzy, R., Tataurov, A. V., Wu, Y., Manthey J. A., McQuisten, K. A., Almaguer, H. G., Pedersen, K. F., Lin, Y., Garretson, J., McEntaggart, N. O., Sailor, C. A., Dawson, R. B., and Peek, A. S. (2008) IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Res.* 36, Suppl. S, W163-W169.
100. Chavali, S., Mahajan, A., and Tabassum, R. (2005) Oligonucleotide properties determination and primer designing: a critical examination of predictions. *Bioinformatics* 21, 3918-3925.
101. Panjkovich, A., and Melo, F. (2005) Comparison of different melting temperature calculation methods for short DNA sequences. *Bioinformatics* 21, 711-722.
102. Richard, O., Isard, D., Mark, A. B., Irving, M. K., and Walder, J. A. (2003) Thermodynamic treatment of oligonucleotide duplex-simplex equilibria. *Proc. Natl. Acad. Sci. U.S.A.* 100, 14840-14845.
103. Chalikian, T. V., Volker, J., and Plum, G. E. (1999) A more unified picture for the thermodynamics of nucleic acid duplex melting: A characterization by calorimetric and volumetric techniques. *Proc. Natl. Acad. Sci. U.S.A.* 96, 7853-7858.
104. Rouzina, I., and Bloomfield, V. A. (1999) Heat capacity effects on the melting of DNA. 1. General aspects. *Biophys. J.* 77, 3242-3251.
105. Gotoh, O. (1983). Prediction of melting profiles and local helix stability for sequenced DNA. *Adv. Biophys.* 16, 1–52.
106. <http://www.millipore.com/userguides.nsf>.
107. Xia, T., SantaLucia, J. Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) Thermodynamic parameters for

an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* 37, 14719-14735.

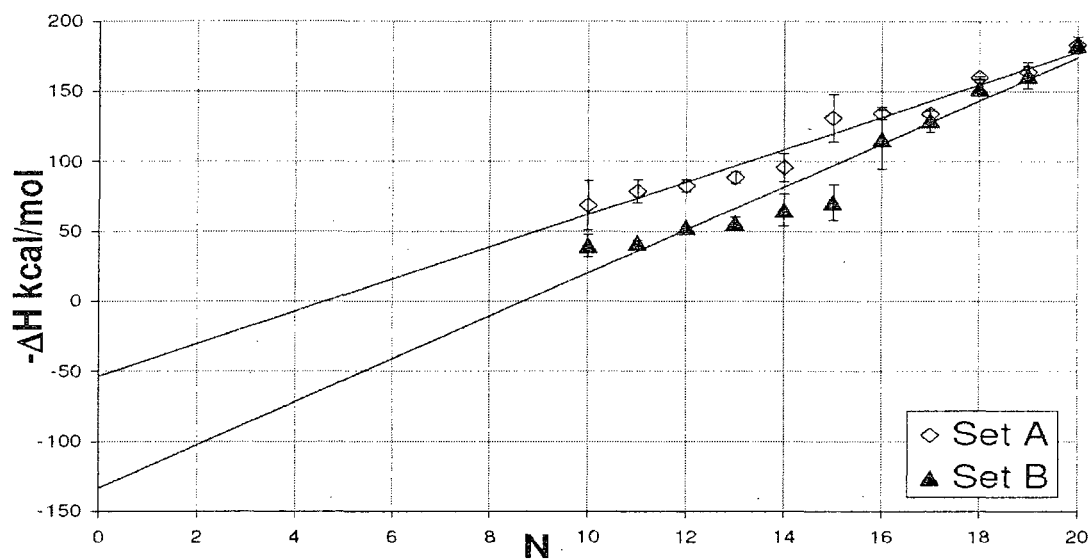
108. Kamenetskii, F., M. D. (1971) Simplification of the empirical relationship between melting temperature of DNA, its GC content and concentration of sodium ions in solution. *Biopolymers* 10, 2623-2624.
109. Goldstein, R. F., and Benight, A. S. (1992) How many numbers are required to specify sequence-dependent properties of polynucleotides? *Biopolymers* 32, 1679-1693.
110. Bond, J. P., Anderson, C. F, and Record, M. T., Jr. (1994) Conformational transitions of duplex and triplex nucleic acid helices: Thermodynamic analysis of effects of salt concentration on stability using preferential interaction coefficients, *Biophys. J.* 67, 825-836.
111. Privalov, P. L., Ptitsyn, O. B., and Birshstein, T. M. (1969) Determination of stability of the DNA double helix in an aqueous medium. *Biopolymers* 8, 559-571.
112. Gray, D. M. (1997) Derivation of nearest-neighbor properties from data on nucleic acid oligomers. *Biopolymers* 42, 783-793.



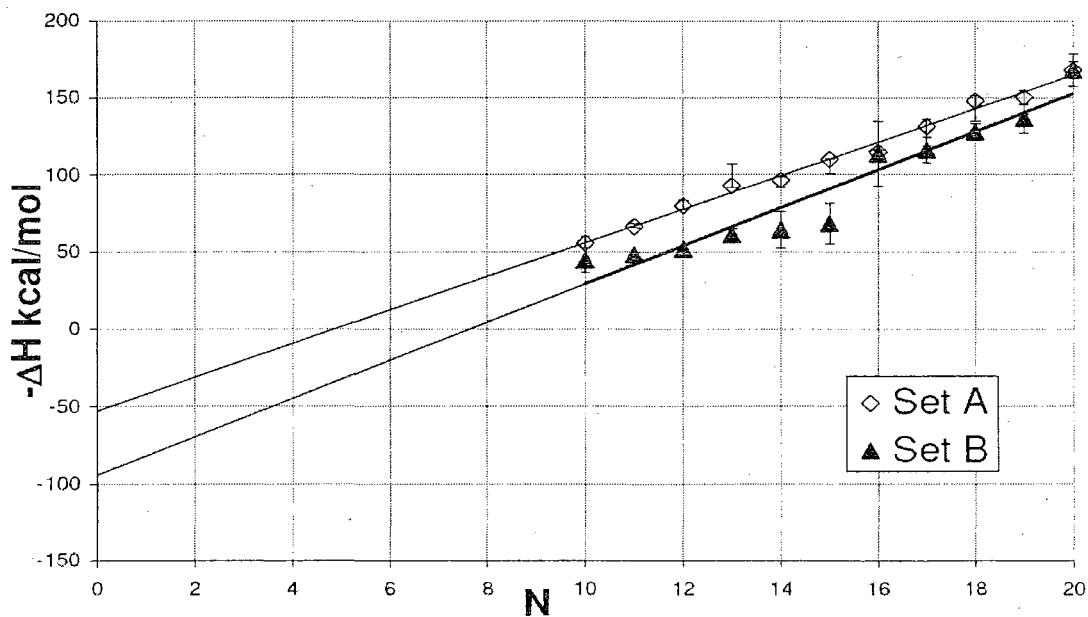
7.0

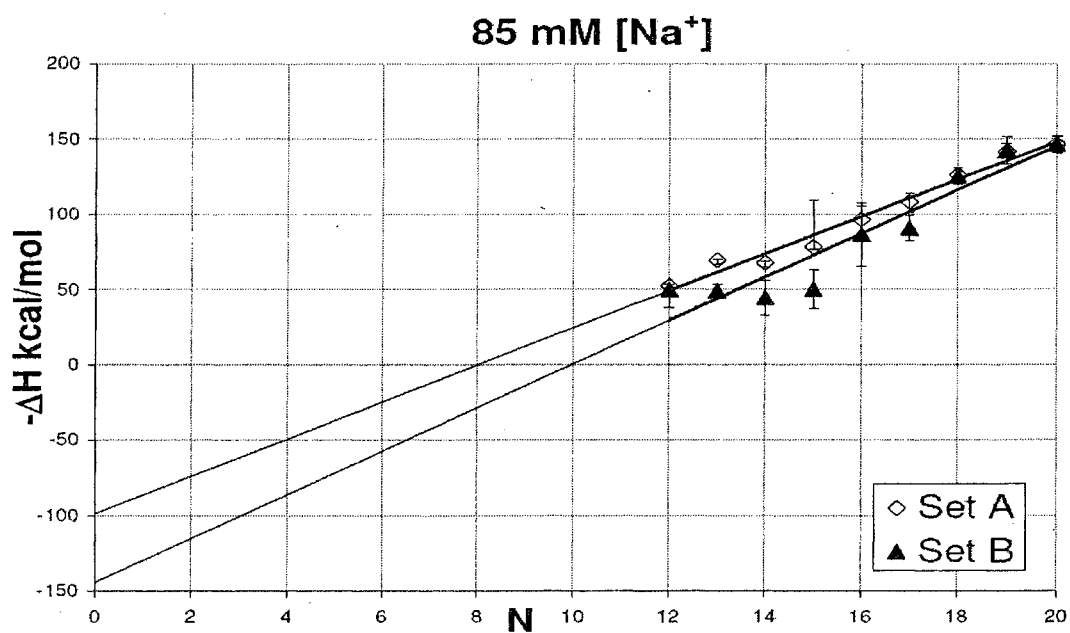
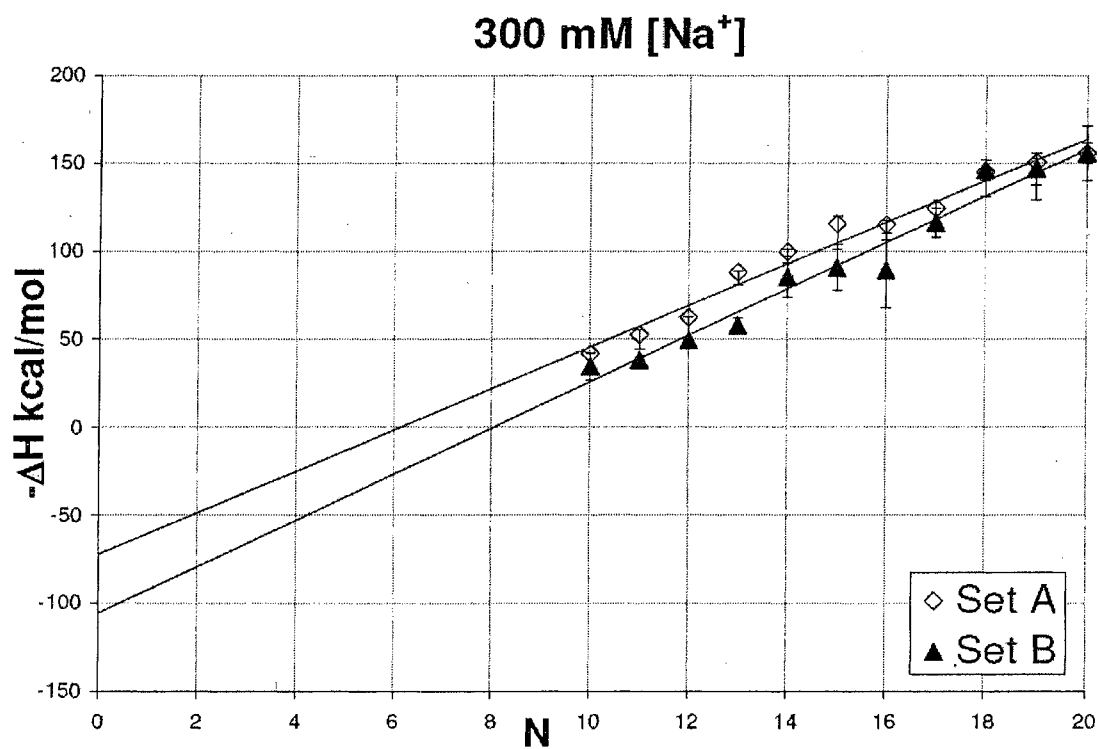
APPENDIX: Plots of Measured  $\Delta H^{cal}$  and  $\Delta S^{cal}$  versus N for DNA molecules containing tandem mismatches in four buffered  $[Na^+]$

1000 mM  $[Na^+]$



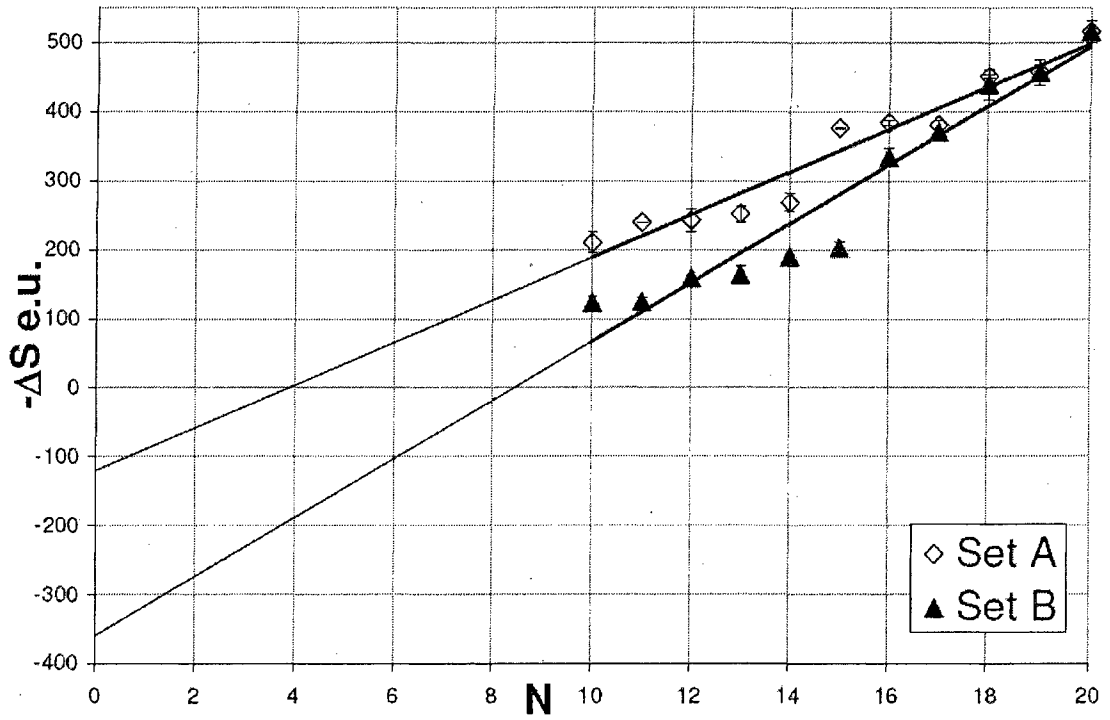
600 mM  $[Na^+]$



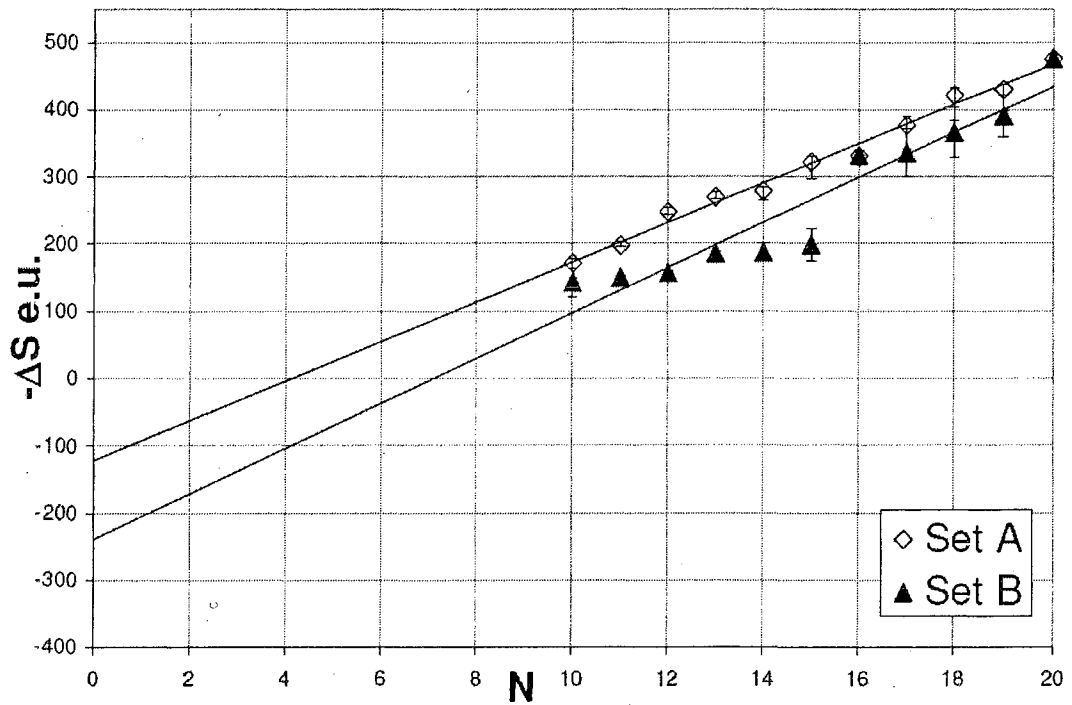


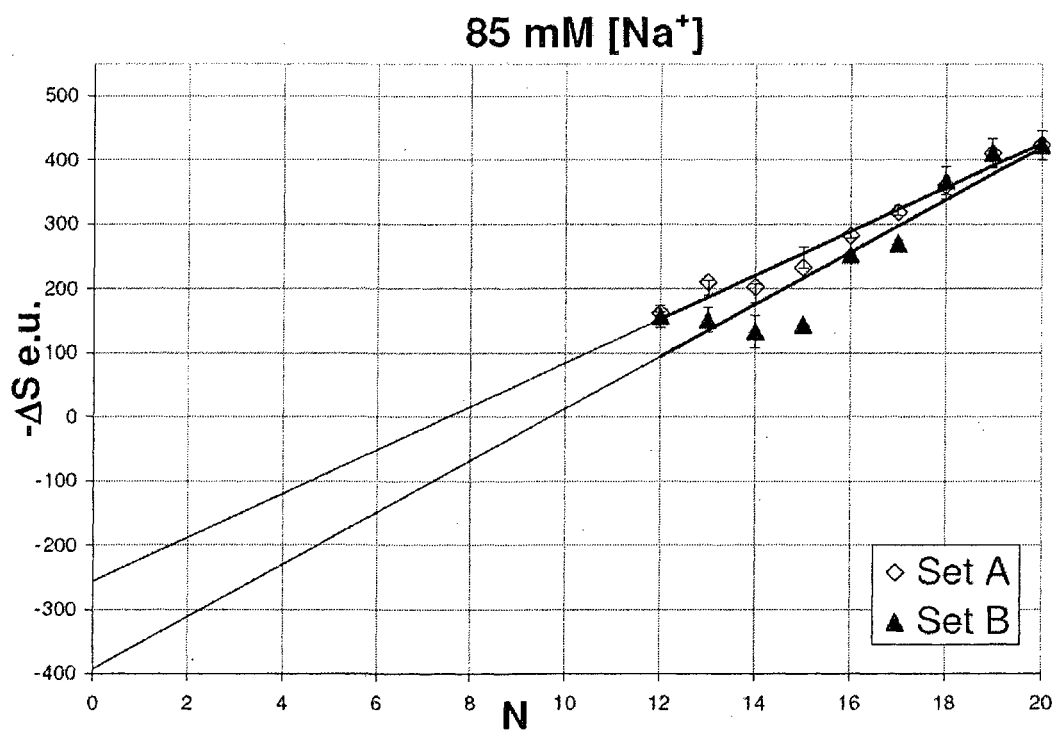
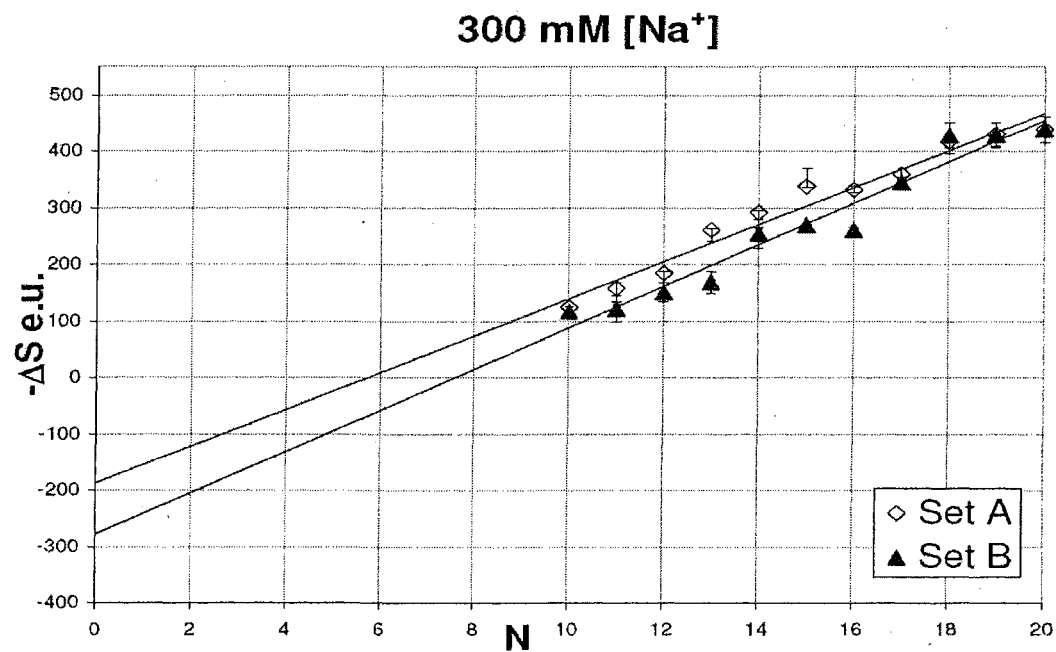
**Figure A1:** Plot of  $N$  versus mean value  $\Delta H$  (kcal/mol) for Set A and Set B DNA duplexes in all [Na<sup>+</sup>] environments. Results are shown for  $0 < N < 20$  base pairs in all [Na<sup>+</sup>] environments.

### 1000 mM [Na<sup>+</sup>]



### 600 mM [Na<sup>+</sup>]





**Figure A2:** Plot of  $N$  versus mean value  $\Delta S$  (kcal/mol) versus for Set A and Set B DNA duplexes in all [Na<sup>+</sup>] environments. Results are shown for  $0 < N < 20$  base pairs in all [Na<sup>+</sup>] environments.