PREDICTIONS OF SHORT DNA DUPLEX THERMODYNAMICS

AND EVALUATION OF NEXT NEAREST NEIGHBOR INTERACTIONS

BY

RICHARD OWCZARZY
B.S., Charles University in Prague, 1993

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Graduate College of the
University of Illinois at Chicago, 1999

Chicago, Illinois

This thesis is dedicated to my parents and Jana Širůčková, without whom it would never have been accomplished.

ACKNOWLEDGMENTS

I would like to thank my advisor, Prof. Albert S. Benight, for his support and guidance. He was patiently teaching me the craft and art of scientific research in the past five years. I will never forget the enjoyable experience of conducting an independent biophysical research in his laboratory. I would also like to thank the other members of thesis committee–Prof. Wonhwa Cho, Prof. Paul R. Young, Prof. Robert F. Goldstein and Prof. Timothy A. Keiderling for their support and advice. My thanks belong also to the members of our lab, Peter M. Vallone, Frank J. Gallo, Peter V. Riccelli for teaching me the laboratory techniques used in nucleic acid research. Their discussions and comments of experimental results were also very appreciated and admired.

For encouragement in final stages of my Ph.D. study, I specially thank my fiancée, Jana Širůčková, and my friends, Jan Kubelka and Ahu Fatma Akin. Despite of long distance between us, my parents have always supported me in whatever and wherever I decided to do and go, therefore, I feel very obligated to them.

RO

TABLE OF CONTENTS

CHAPTER PAGE

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	Adenine
ATP	Adenosine 5'-Triphosphate
C	Cytosine
CD	Circular Dichroism
ddH$_2$O	Double Distilled Water
DNA	Deoxyribonucleic Acid
DSC	Differential Scanning Calorimeter
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
G	Guanine
MWCO	Molecular Weight Cut-off
NEB	New England Biolabs
n-n	Nearest-neighbor
n-n-n	Next-nearest-neighbor
PAGE	Polyacrylamide Gel Electrophoresis
PEG	Polyethylenglycol
SVD	Singular Value Decomposition
T	Thymine
TBE	Tris, Boric acid, EDTA buffer
TLC	Thin Layer Chromatography
Tris	2-Amino-2-(hydroxymethyl)-1,3-propanediol

LIST OF ABBREVIATIONS (continued)

UV	Ultraviolet

SUMMARY

Prediction of DNA duplex stability and thermodynamics is invaluable for many molecular biology applications involving sequence dependent hybridization reactions. Sequence dependent stability of duplex DNA plays a major role in fundamental processes of the living cell, such as replication, transcription, and recombination.

The analytical methods for calculation of DNA duplex thermodynamics from sequence and DNA concentration are based on the nearest-neighbor (n-n) model. There are 10 n-n base pair doublets in DNA duplexes. However, when the ends are taken into account, 12 n-n parameters are required to describe the thermodynamics of melting of short duplex DNAs in terms of the n-n model.

At least 11 sets of n-n sequence dependent thermodynamic parameters for DNA have been published. These sets were compared in their ability to predict free energies and melting temperatures of linear DNA duplexes 10 base pair long. Some of the sequences showing the biggest discrepancies in predictions from different sets of n-n parameters were investigated experimentally. The n-n set published from our lab, Doktycz et al. (1992), and two recently reported sets of n-n thermodynamic parameters (Allawi and SantaLucia, 1997; Sugimoto et al., 1996) provided more accurate predictions of experimental melting temperatures of short duplex DNA oligomers, than the older, more commonly used set of n-n parameters (Breslauer et al., 1986). However, none of the n-n sets was able to predict melting temperatures of several selected 10 base pair sequences that displayed the most discordance between predictions of the various n-n sets. These and other observations suggested that the general characterization of DNA thermodynamic stability in terms of a n-n model may be inadequate, and significant sequence dependent interactions in DNA duplexes may extend over distances beyond nearest-neighbors. To investigate such sequence dependent interactions and their dependence on length and salt, we performed UV-melting studies on 39 DNA dumbbells. We employed DNA dumbbells because they offer many advantages for studies of sequence dependent effects. Since DNA dumbbells consist of a duplex stem effectively crosslinked on both ends by single strand loops, the melting process of the dumbbell stem is unimolecular and concentration independent. The end-loops stabilize a dumbbell stem compared to the same sequence without loops. For these reasons the dumbbell system provides a more realistic mimic of short sequences in a long DNA environment, without the concentration dependence and other anomalies associated with melting of short duplex DNAs. Dumbbells used in the melting studies had constant end sequences and variable central sequences of four to eight base pair long. From melting data of 39 dumbbells in buffers containing total sodium ion concentrations of 25, 55, 85, 115 mM, nearest-neighbor and next-nearest-neighbor (triplet) interactions in duplex DNA were evaluated. Rigorous statistical analysis revealed that melting data of dumbbells in 85 and 115 mM Na$^+$ can be adequately described by the n-n model. However, melting data of dumbbells in 25 and 55 mM Na$^+$ cannot be adequately fitted in terms of the nearest-neighbor model within the errors of the measurements. If next-nearest-neighbor (triplet) interactions are considered, a reasonable fit of the melting data is obtained even at these low Na$^+$ concentrations. This indicates that next-nearest-neighbor (n-n-n) interactions in duplex DNAs are significant in solutions below 55 mM Na$^+$. To test the evaluated nearest-neighbor and next-nearest-neighbor parameters, two additional DNA dumbbell molecules were prepared and melted in 25 mM Na$^+$. When melting

SUMMARY (continued)

temperatures predicted with the n-n and n-n-n parameters were compared to experimental melting temperatures, more accurate predictions of melting temperatures of the dumbbells were obtained when the next-nearest-neighbor parameters were employed.

A correction for the nucleation enthalpy that is required to apply the evaluated n-n parameters to predict thermodynamics of short linear DNA duplex oligomers, was evaluated in 115 mM Na^+. The correction for the nucleation enthalpy depends on percentage of G•C base pairs and length of the linear DNA duplex.

1. GENERAL INTRODUCTION

1.1 Background of Nucleic Acid Thermodynamics

DNA was first described in 1869 by F. Mischer. However, its significance to cell function and the proof that it is responsible for inherited traits did come almost a century later. The experiments that led to the assignment of genes to DNA used bacteria and their viruses (Avery et al., 1944).

In 1953, J.D. Watson and F.H.C. Crick offered a model for the physical structure of DNA (Watson and Crick, 1953). They proposed that two antiparallel strands are coiled about one another to form a double-stranded helix. In this model the bases of one strand are hydrogen-bonded to those of the other strand to form purine-pyrimidine base pairs. These are either adenine and thymine (A•T) or guanine and cytosine (G•C). When double-stranded DNA or native DNA is heated, the bonding forces between and within the strands are disrupted, and the two strands can separate. The chemical process, DNA melting or denaturation involves breaking of three hydrogen bonds in every guanine-cytosine base pair and breaking of two hydrogen bonds in adenine-thymine base pair. The important question is, what is the thermodynamics of breaking and forming DNA, RNA, or hybrid double helices for any sequence in any solvent? This question has been approached in studies of nucleic acid thermodynamics but is far from answered.

Disruption and formation of base pairs is essential for biological functions of nucleic acids in processes such as transcription and duplication of genes. Thus, the study of the thermodynamics of DNA duplex interactions in reactions analogous to these biochemical processes is fundamental to understanding basic biological processes of the cell.

Pioneer melting experiments (Marmur and Doty, 1962; Thomas, 1954) showed that DNA duplex melting and formation can be easily monitored through changes in the ultraviolet spectrum. Absorbance in the ultraviolet region increases as the double helix "melts" to two single strands. Reversibly, upon DNA duplex formation, the absorbance decreases. The hypochromicity effect (the reduction of absorbance of chromophores arranged in an ordered array) ranges in magnitude from 20-40% depending on duplex length. The structural changes in the strands that result from formation of Watson-Crick base pairs and associated interactions between neighboring bases (stacking interactions) cause the hypochromism (Sauer, 1995). When duplex DNA unwinds (denatures), it is assumed that the changes in the extinction coefficient are proportional to the extent of denaturation. In other words, the change in absorbance at around 268 nm is directly proportional to the fraction of broken base pairs, θ_B. At this wavelength, the change in extinction coefficients upon melting for G•C and A•T base pairs is approximately the same (Blake and Hydorn, 1985; Wartell and Benight, 1985). Therefore, when θ_B increases from 0.0 (intact duplex) to 1.0 (completely melted, separated single-strands), it is presumed that the total absorbance increases proportionally. The optical UV-melting experiment is usually performed in the following way. Absorbance at 268 nm and temperature is monitored while the solution is slowly heated at a steady rate (6-60 °C / hour) from room temperature to 100 °C and cooled back to room temperature. The graph of absorbance vs temperature, the melting curve, is analyzed to obtain melting temperature (T_m), standard Gibbs free energy ($\Delta G°$), standard enthalpy ($\Delta H°$) and standard entropy ($\Delta S°$), of the helix-coil transition. The melting temperature is often defined as the temperature when half of molecules are "melted", that is the temperature at the midpoint of the double helix to single strands transition. In this thesis, we used UV-melting experiments to study melting transitions of DNA dumbbells and short duplex DNAs.

Over 40 years ago, it was found that DNA stability depends on the chain length and the percentage of guanine-cytosine base pairs. The higher the G•C content, the greater the stability of DNA. The simplest interpretation is that when a guanine-cytosine base pair is formed, three hydrogen bonds are created in contrast to a adenine-thymine base pair that contains only two hydrogen bonds. Hence, higher G•C content increases the melting temperature and stability of duplex DNA.

High resolution melting experiments showed that thermodynamic stability depends not only on G•C content but also on the sequence order or context (Wada et al., 1980). Two sequences with the same content of G•C base pairs can have different stabilities. The effect is less dramatic in comparison with two sequences having different G•C content, but is significant. To consider these effects, the nearest-neighbor (n-n) model has been employed (Wartell and Benight, 1985; Delcourt and Blake, 1991). The n-n model of DNA predicts thermodynamics from n-n parameters for each doublet and assumes that there are no significant interactions beyond n-n doublets. In DNA duplex, there are 10 n-n doublets (5'-AG-3'/5'-CT-3', AA/TT, AT/AT, etc.). Consequently, the thermodynamics of double helix formation from single strands for any sequence can be calculated from 10 nearest-neighbor thermodynamic parameters. For linear duplex oligomers, additional parameters for the initiation of duplex formation are necessary. Initiation parameters were at first introduced to account for the thermodynamic difference between formation of the first base pair between separate single-strands and formation of all other subsequent base pairs. They have been introduced in various formats (Breslauer et al., 1986; Sugimoto et al., 1996). Later it was recognized (Gray, 1997a) that initiation parameters also include sequence dependent interactions between the end base pairs and the surrounding solvent, i.e. end interactions. Analysis of the different treatments of the end interactions is presented in Chapter 4.

The ten nearest-neighbor parameters represent the thermodynamic values for formation of 10 possible n-n doublets in DNA duplexes. To evaluate the n-n parameters a complete database of molecules must be constructed so that all possible combinations of nearest neighbors are included in their sequences. Then the melting properties of these molecules must be measured. The thermodynamic parameters can then be obtained by a least-squares fit of the set of measured values. Unfortunately, two constraints limit the number of independent values that can be deduced from any set of molecules (Goldstein and Benight, 1992; Gray and Tinoco, 1970; Gray, 1997a; 1997b). Consequently, evaluated parameters for the 10 nearest-neighbors are not unique and cannot be directly compared between various laboratories and research groups. However, unique sets of linear combinations of nearest-neighbor parameters can be constructed that are useful for comparison of different published sets. This mathematical fact that parameters for 10 n-n doublets cannot be determined uniquely is a source of great confusion. More detailed explanations are provided in Chapter 4, where published modifications of the nearest-neighbor model are described and comparison of n-n parameters from different laboratories is made.

The influence of metal ions on nucleic acid melting is dramatic and was carefully investigated earlier in DNA polymers (Dove and Davidson, 1962; Gruenwedel and Chi-Hsia, 1969). They found that the melting temperature of DNA increases linearly with logarithm of alkali metal cation concentration (Owen et al., 1969; Frank-Kamenetskii, 1971) up to 1.0 M. A nucleic acid duplex is a highly negatively charged polyelectrolyte due to ionized phosphate groups located in the backbone. Long-range electrostatic interactions between nucleic acids and counter-ions and co-ions in solution significantly effects the thermodynamics of nucleic acids. The local cation concentration at the DNA surface is predicted to be very high, in the molar range and the concentration of anions near the surface is essentially zero. Strong repulsive electrostatic interactions exist between the negatively charged phosphate groups. Higher monovalent ion concentrations screen this interaction more effectively which results in higher stability of the DNA. These effects on stability may be explained by differences in ion-ion interactions in the native, intact duplex compared to denatured single strands. Denaturation of nucleic acids reduces their structural charge density, which reduces local ion concentration gradients. Consequently, denaturation is accompanied by release of counterions. As salt concentration of solution is reduced, contributions from counterion release to the observed free energy difference between denatured and native states become more important. This effect is entropic in origin and can be described as a dilution of counterions (Record et al., 1981).

At monovalent salt concentrations above ~ 1.5 M, the stability and denaturation temperature of nucleic acids level off or decrease, depending on the anion (Schildkraut and Lifson, 1965; Gruenwedel et al., 1971). This effect is generally attributed to denaturation by anions which, at very high concentration, act as hydrophobic bond-breaking agents. The effects of polyvalent ions is more complex, Mg^{2+} ions increase T_m but Cu^{2+} ions have the opposite effect and destabilize DNA duplexes (Eichhorn and Shin, 1968). Perhaps, Cu^{2+} exhibits a different mode of binding than Mg^{2+}, and might prefer binding to the electron-donor groups on the bases instead of binding to phosphate moieties.

Besides UV-melting experiments additional techniques are sometimes employed to study DNA denaturation. The DNA denaturation process involves absorption of heat, which is necessary for breaking of hydrogen-bonding and stacking interactions, and therefore can be measured directly in a calorimeter (Ladbury and Chowdry, 1996). Namely, differential scanning calorimetry (DSC) is employed very often (Sturtevant, 1987). This technique has the advantage of providing direct, model-independent measurements of thermodynamic functions (Marky and Breslauer, 1987). Differential scanning calorimetry is not contingent on assumptions about the nature of melting transition. Therefore, calorimetric measurements can be used to test the accuracy of the two-state assumption of DNA melting. Primary data is collected as the change of excess heat capacity (ΔC_p) vs temperature (T). From these curves, the enthalpy of the melting transition is found from the integrated area under the curve. Similarly, the transition entropy is equal to the area under the plot of $\Delta C_p/T$ vs T. Other methods, including UV-melting experiments, are indirect and require some assumptions about the number of different states or intermediates during the DNA melting process in order to deduce the enthalpy, entropy, free energy of the transition. The disadvantage of the calorimetric technique is that it usually requires substantially larger amounts of DNA samples than optical spectroscopic measurements. However, the latest advances in microcalorimeters enable calorimetric measurements on smaller amounts of samples. From as little as ~250 μg of DNA reliable results of DSC melting experiments can be obtained (Vallone, Peter, unpublished communication). For an excellent review of DSC calibration procedures, see Gmelin and Sarge (1995).

1.2 Why Is It Important to Know the Thermodynamics of DNA Duplexes?

The thermal denaturation of DNA is a topic of fundamental interest and practical importance. The polymerase chain reaction (PCR), cloning and hybridization experiments are commonly used techniques in biochemistry and molecular biology. Successful design of these methods depends on knowledge of DNA duplex stability. Direct experimental determinations of melting temperature and evaluation of the thermodynamics of duplex formation are time-consuming and costly tasks. Duplexes must be synthesized and T_m must be determined. Measurements of melting temperature are low-throughput and last several hours. If the objective is to find an optimal probe with a specific melting temperature, only a single probe from many examined ones will be used in subsequent molecular biology experiments. Accurate predictions of DNA stability from sequence would provide a convenient, simple way in which to predict the outcome of hybridization reactions and thereby improve design of these experiments. Although modern models for predictions of DNA stability have been available since the 1980's , more inaccurate, old methods are still used among molecular biology specialists to calculate T_m (Wetmur, 1991). Some old methods for prediction of melting temperature of duplex DNA are based solely on G•C content and ignore deviations due to nearest-neighbor interactions. Because such predictions of T_m are not accurate enough, PCR reactions are to date for the most part optimized by trial-and-error.

In hybridization experiments, it is essential to know the minimal length required for an oligomer probe to form a stable duplex with a target under a given set of conditions. Current models of DNA denaturation provide accurate predictions in buffers containing 85 mM to 1 M total Na^+. In this work, we evaluated thermodynamic parameters in lower sodium ion concentrations, 25 and 55 mM. Nearest-neighbor thermodynamic parameters in solvents of polyvalent ions are of significant biological interest (Mg^{2+}, Ca^{2+}), but yet to be determined. Ideally, we would like to conduct hybridization reaction under stringent conditions, that is, below T_m of matched duplex and above T_m of single base-pair mismatched duplexes. Under these conditions, only perfect matches of target-probe can form duplexes and are detected. Therefore, under such conditions, false signals arising from stable mismatched duplexes are suppressed and sensitivity of DNA detection is increased. Attempts to predict stringency of hybridization conditions from DNA thermodynamics have been marginally successful (Ivanov and AbouHaidar, 1995; Kunitsyn et al., 1996). But the thermodynamic parameters of all possible single base pair mismatches are yet to be measured and evaluated.

Marketable medical diagnostic tests using DNA probes have been developed recently to detect bacterial and viral pathogens. These tests have distinct advantages over conventional antibody assays. They can detect very low amounts of pathogens immediately after infection. More conventional methods using viral or bacterial antibodies are often only detectable after an immune response to the presence of the organism. This can take much longer time. For example, the immune response to HIV-1 does not present for up to 6 months after initial infection. Early detection by other means (DNA probes) could enhance the safety of the general blood supply. Attempts have been made to test for several pathogens simultaneously using DNA probes. Detection and quantification of both Chlamydia trachomatis and Neisseria gonorrhoeae in one sample by nucleic acid based diagnostics was reported (Nelson, 1996). Reliable knowledge of nucleic acid thermodynamics will speed up the development of these and other medical tests.

Light-directed DNA synthesis on microchips using photolithographic masks has been achieved (Pirrung et al., 1998; McGall et al., 1996; Pease et al., 1994). This sensational breakthrough brings amazing new features and possibilities that seem to be limited only by our imagination. Genome-wide RNA expression monitoring (Lockhart et al., 1996), polymorphism determination (Wang et al., 1998), pathogen detection, fast sequencing by hybridization (Drmanac et al., 1998) are just a few examples. Hundreds of thousands of different sequences can be hybridized and tested on a single biochip. For practical reasons, it is important to hybridize sequences with very different G•C content (therefore very different stability) under the same conditions in a single experiment. The selected conditions (salt, buffer, temperature) must allow the hybridization of perfect matches and inhibit the hybridization of stable mismatches so that the correct signal is detected (Lipshutz et al., 1995). It is important to know how one base mismatch will destabilize hybridization. The detailed knowledge of duplex thermodynamics will help us in the selection of experimental conditions for various DNA microchips.

It was recently suggested that the inherent stability of a DNA duplex may influence protein binding. Binding constants and rates of endonuclease cleavage by BamHI and Alu I to various DNA duplexes were measured (Benight et al., 1995; Riccelli et al., 1998). The consensus binding site of the restriction endonucleases was kept constant and the flanking sequences were modified to adjust the total stability of DNA duplexes. The results showed a negative correlation between the free energy of DNA duplex denaturation and binding free energy of the enzyme binding reaction. If such an effect is proved to be omnipresent, protein binding to DNA sites could be predicted from DNA duplex stability alone.

1.3 Brief Description of Thesis

The focus of this thesis is aimed at prediction of DNA thermodynamics for short DNA duplexes and dumbbells. Chapter 2 provides a thorough description of all experimental techniques that were used in experiments. Chapter 3 describes DNA thermodynamics and melting theory. The two-state (all-or-none) model is presented. Calculation procedures for DNA dumbbells and linear duplex DNAs in terms of this model are explained.

Thermodynamic predictions of DNA duplex stability based on various published nearest-neighbor sets are systematically compared in Chapter 4. Differences and similarities for various sets of thermodynamic parameters are revealed in predictions of all possible duplexes that are 10 nucleotides long. A few of the biggest predicted discrepancies are tested directly in melting experiments. The accuracy of different sets of nearest-neighbor parameters is assayed by comparison of predictions of melting temperatures for 251 duplexes for which the melting temperatures have been determined experimentally.

Methods employed to evaluate the thermodynamic parameters from a measured database of DNA dumbbells are explained in Chapter 5. The n-n model is amended to include next-nearest-neighbor (n-n-n) interactions, i.e., interactions within triplets. Results of melting experiments conducted as a function of [Na^+] for 22 DNA dumbbells are reported. From this set and published melting data of 17 DNA dumbbells (Doktycz et al., 1992), nearest-neighbor parameters and new next-nearest-neighbor parameters are evaluated. Next-nearest-neighbor parameters are discovered to be necessary to describe the thermodynamics of DNA denaturation (within the errors) at low salt environment (below 55 mM Na^+). Application of n-n-n parameters improves prediction of two DNA dumbbell molecules in 25 mM Na^+.

2. MATERIAL AND EXPERIMENTAL METHODS

2.1 Synthesis of Oligodeoxyribonucleotides

All oligodeoxyribonucleotides were synthesized on an Applied Biosystems 380B synthesizer using the standard β-cyanoethyl phosphoramidite method (Caruthers, 1985; Caruhers et al., 1992). One or two μmole synthesis were carried out. The hairpins strands synthesized for DNA dumbbell preparation required a 5' phosphate group necessary for the ligation step (described below). As part of the synthesis method, after the final deoxyribonucleotide was added to oligomer, the end was phosphorylated with 2-[2-(4,4'-dimethoxytrityloxy)ethylsulfonyl]ethyl-(2-cyanoethyl)-(N,N-diisopropyl)-phosphoramidite. Short linear DNA duplexes were formed from oligomer strands synthesized without a phosphate moiety at the 5' end, because the strands were not used for ligation.

Cleavage and deprotection from solid support was completed on the DNA synthesizer. The synthesis product was deprotected by adding three milliliters of concentrated ammonia and heating the mixture at 55 °C for 8-16 hours. The DNA sample was lyophilized on a SpeedVac SVC100 concentrator under vacuum and stored in a freezer at -20 °C.

2.2 Determination of DNA Concentration

Dried synthetic oligodeoxyribonucleotides were rehydrated in 1 ml of double distilled water (ddH$_2$O) and incubated at 36 °C for at least 15 minutes. The sample was vortexed and the amount of DNA was estimated by UV spectroscopy. Typically, 100:1 dilutions of stock DNA samples were prepared. The 10 μl of stock solution was mixed with 990 μl ddH$_2$O and the absorbance of diluted solution at 260 nm was measured on the Hewlett-Packard 8452 spectrophotometer. The molar concentration of DNA, c_{DNA}, was calculated from the Lambert-Beer Law,

$$c_{DNA} = \frac{A_{260}}{\varepsilon_{260} . l} \tag{2.1}$$

where A_{260} is the absorbance at 260 nm, l is the pathlength of the cuvette (1 cm) and ε_{260} denotes the extinction coefficient of single stranded DNA. The published nearest-neighbor values (Cantor and Warshaw, 1970; Fasman, 1975) were used to determine extinction coefficients of single strand DNAs at 25 °C, 260 nm and neutral pH. The calculation was programed in FORTRAN77 and C languages.

2.3 Analytical Gel Electrophoresis

Rehydrated oligodeoxyribonucleotides were checked for purity by analytical polyacrylamide gel electrophoresis (PAGE) under denaturing conditions. Polyacrylamide gels contained TBE buffer and 7 M urea. TBE buffer is composed of 0.050 M Tris (J. T. Baker Company), 0.050 M boric acid (Amresco Company), 1 mM Na_2EDTA (Amresco Company), pH = 8.1-8.3. Electrophoresis was performed using a Hoeffer electrophoresis stand. The total acrylamide (Eastman Kodak Company) monomer concentration of solutions varied between 15-20 percent weight depending on the sample. For shorter DNA chain lengths, higher gel concentrations were used. N,N'-methylenebisacrylamide cross-linker was added so that it was 1.57% of the monomer molar concentration. Dry crystalline acrylamide and bisacrylamide were combined, dissolved in ddH_2O and filtered to form a 30 percent weight solution. 24 g of urea (J.

T. Baker Company), 5 ml 10 x TBE and 25 ml 30% acrylamide solution were mixed. After urea dissolution, 350 μl of 10% ammonium persulfate and 20 μl of N,N,N'N'-tetramethylethylenediamine (Bio-Rad Laboratories or Sigma) was added and the solution was poured between glass plates. An appropriate comb was pressed into the solution to form wells. The solution was completely polymerized after 1 hour and the gel was positioned in the electrophoresis stand. Thickness of the gel was 1 mm. DNA was rehydrated in 8 μl of loading buffer (7 M urea and 1 x TBE) and 2 μl of indicator (0.25 percent weight of bromophenol blue and 0.25 percent weight of xylene cyanole). To remove persulfate ions, gels were pre-electrophoresed for 15-30 minutes. Approximately 2.1 μg of DNA was loaded in each well. Electrophoresis was run at 230 V, 20-30 mA, for about 60 minutes until the bromophenol blue dye moved 80% of the gel length. Intensity of the electric field was about 13 V/cm. The gel was soaked in Stains-all dye solution in the dark for 8-16 hours. Stains-all dye solution was prepared from 5 mg Stains-all (3,3'-diethyl-9-methyl-4,5,4',5'-dibenzothiacarbocyanine) dissolved in 300 ml of N,N-dimethylformamide and 300 ml of ddH_2O. After destaining in ddH_2O for approximately 15 minutes, the gel was photographed under bright visible light with a Polaroid MP4 land camera using Polaroid type-52 film. Alternatively, gels were scanned on a flatbed, black-white scanner (Visioneer) with resolution of 150-300 dpi.

2.4 Synthesis of Short DNA Duplexes

If analytical gel electrophoresis of samples showed that samples were pure and contained mostly the expected synthesis product (≥95%) with minimal contamination by shorter failure sequences, then the DNA was not subjected to further purification. To remove excess salt, the total amount of sample from a single synthesis was dialyzed against ddH_2O for 2 days. The dialysis tubing (Cellu•Sep from Membrane Filtration Products Inc.) had a molecular weight cut-off 2000 g/mol. After desalting, the DNA was dried down in a SpeedVac SVC100 concentrator under vacuum and rehydrated for an hour in 120 μl of denaturing electrophoresis buffer (0.89 M Tris, 0.89 M boric acid, 0.05 M Na_2EDTA, 7 M urea, pH = 8.3). Then 25 μl of indicator (0.25 percent weight bromophenol blue in water) was added and the sample was loaded on a preparative (6 mm thick) 20% polyacrylamide gel containing 7 M urea and 1 X TBE buffer. Prior to loading, the gel was pre-electrophoresed at 220 V for 1 hour. The electrophoresis was run at 220V, 60-80 mA for about 90 minutes until the bromophenol blue dye traveled 80% of the length of the gel. The gel was carefully removed from the glass plates and placed on a white TLC plate (Whatman) that was covered in plastic wrap. The gel was briefly shadowed with a 254 nm ultraviolet lamp to minimize the formation of UV-photoadducts (Widlak et al., 1995). Oligodeoxyribonucleotides were separated based on chain length (Maniatis et al., 1975). The major (darkest) band that had the slowest mobility was sliced out with a razor. The long narrow gel strip was placed into a dialysis bag (2000 MWCO) filled with 1 x TBE buffer and DNA was separated from the gel by electroelution. In this procedure, the dialysis bag was placed into a horizontal electrophoresis tank. Direct electric current was run perpendicularly to the gel strip. Intensity of the electric field was about 4.5 V/cm. The electroelution tank contained 1 x TBE buffer, pH = 8.3. After 90 minutes, the direction of current was briefly reversed for 1 minute to detach any DNA stuck to the side of the dialysis membrane. The bag was rinsed with 1 x TBE buffer and filled with fresh buffer. Then a second round of electroelution was done. About 15-20% more DNA was obtained during the second electroelution. To remove the gel debris, DNA was collected and filtered through a 0.2 μm nylon filter (Milipore or Gelman Sciences) attached to a 60 ml syringe. To remove the extraneous acrylamide and buffer, filtered purified samples were dialyzed exhaustively versus ddH_2O.

Short linear DNA duplexes that were studied were only 10 nucleotides long. Their stability was relatively low and as a result the duplexes could not be cleanly separated by polyacrylamide gel electrophoresis. Thus, the individual single strands were purified by PAGE and dialyzed against 115 mM melting buffer containing 100 mM NaCl, 5 mM Na_2HPO_4, 5 mM NaH_2PO_4, 1 mM Na_2EDTA, pH adjusted to 6.80 with NaOH. The purified complementary single strands were then mixed in a 1:1 molar ratio to form DNA duplexes.

Concentrations of oligodeoxyribonucleotides were determined from their absorbance at 260 nm and molar extinction coefficients (see Chapter 2.2). Identical molar amounts of each complementary strand in 115 mM melting buffer were then mixed together. At this point the linear duplexes were ready for melting experiments. To transfer the duplexes into 1 M melting buffer (1 M NaCl, 5 mM Na_2HPO_4, 5 mM NaH_2PO_4, 1 mM Na_2EDTA, pH adjusted to 6.80 with NaOH), Centricon-3 concentrators could not be used because their membranes have a tendency to break at salt concentrations above 1 M Na^+. Plus, their molecular weight cut-off was close to the molar mass of the single strands. Therefore, DNA duplexes were dialyzed exhaustively against 1 M buffer in dialysis bags with a molecular weight cut-off of 2000 g/mol.

2.5 Construction of DNA Dumbbells

DNA dumbbells are self-complementary closed circular molecules. Procedures of DNA dumbbells preparation have been published (Wemmer and Benight, 1985). Our dumbbell construction scheme is depicted on Figure 1. DNA dumbbells have a double-hairpin structure prepared by enzymatic ligation (Fareed et al., 1971) of two dangling-ended hairpins. Two DNA hairpins with 5' sticky ends were synthesized, deprotected and desalted. The 5' phosphate group was attached during chemical synthesis. The amounts of each oligodeoxyribonucleotide were estimated from the absorbance measured at 260 nm (see Chapter 2.2). Purity of the hairpin strands was checked by analytical gel electrophoresis. Because most of the synthesis products were the proper, full-length oligodeoxyribonucleotides and shorter sequences than expected (attributed to failures on the DNA synthesizer) did not react in ligation reactions, the hairpins were not purified before ligation reaction.

Enzymatic ligation was preferred to chemical synthesis with coupling agents (Sokolova et al., 1988; Ashley and Kushlan, 1991). Although cyanogen bromide or carbodiimide derivates could be efficient, there is also the danger of unwanted byproduct formation. Some coupling agents can modify nucleotide bases or hydroxy and phosphate groups. For this reason, all dumbbells made for experiments in this thesis were ligated enzymatically using T4 DNA ligase.

Complementary, sticky ended hairpins were rehydrated in 500 µl ddH_2O and equal molar amounts were mixed together. In the case of large scale pot ligation the entire amount of DNA sample obtained from a single synthesis (50-100 A_{260} units) was used. Ligation reactions were carried out in 1-2 ml solutions containing 1-2 mM ATP (FisherBiotech), 20 mM DTT, 10 mM $MgCl_2$, 50 mM Tris-HCl (pH = 7.8), 0.4-1 mM DNA oligomers. The whole mixture was

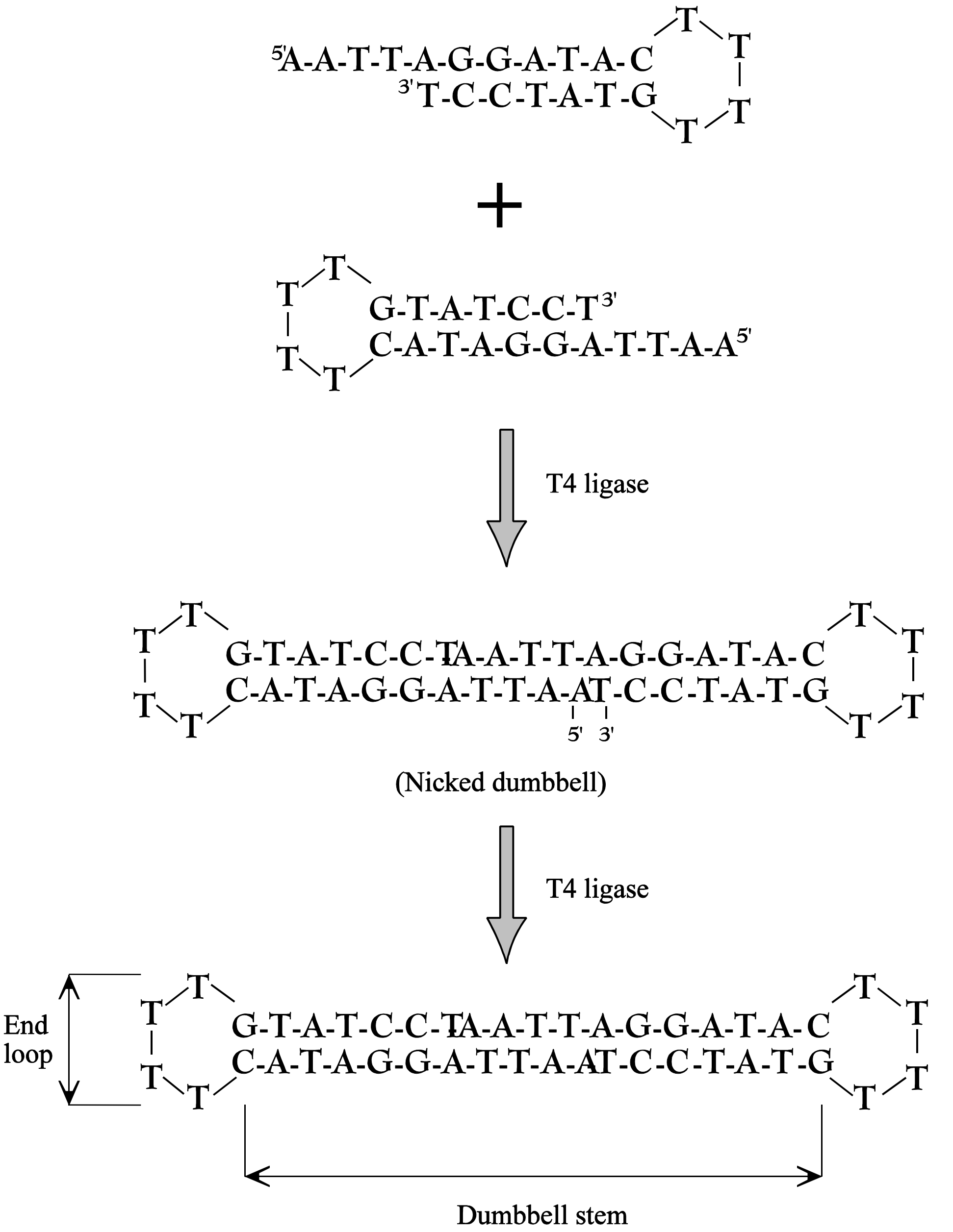


Figure 1. Synthesis scheme for DNA dumbbells. Two complementary sticky-ended hairpins are ligated at two sites to form a closed circular DNA dumbbell.

heated to 95 °C for 3 minutes and then slowly cooled to room temperature for 20 minutes. Although the heating-cooling procedure enhanced proper formation and annealing of hairpins, it had only minor influence on the overall yield of doubly ligated dumbbells. After 15 minutes incubation on ice, 20-30 x10^3 NEB units of T4 DNA ligase (New England Biolabs) were added. Ligation reaction was carried out at 16 °C for 15-24 hours. The reaction mixture was then incubated at 90 °C for 10 minutes to denature T4 ligase. We refined the ligation procedure by systematically changing the concentrations of reactants. Relatively small volumes (about 1 ml) of the ligation mixture were necessary to achieve the high concentrations of DNA and ligase essential for obtaining reasonably high yields (70-80%) of fully ligated product (Sugino et al., 1977). If the concentration of ATP was less then 1 mM, the yield was also low (30-40%), probably due to the lack of necessary cofactor. To achieve maximum ligase efficiency, optimal concentrations of Mg^{2+} were around 10 mM. Concentrations of Tris and dithiothreitol exhibited only minor effects on the yields. It has been reported that at concentrations of 70 mM NaCl and higher T4 ligase is inhibited. Concentrations of 200 mM NaCl or 1 mM spermine were reported (Raae et al., 1975) to cause almost 100% inhibition of ligase activity. Dithiothreitol is supposed to stabilize ligase but it did not seem to be necessary in our experiments. It was demonstrated (Zimmerman and Pheiffer, 1983; Pheiffer and Zimmerman, 1983) that blunt-end ligation yield is improved by addition of polyethylenglycol (PEG) of molecular weight 8000. However, we found in the case of ligation of sticky-ended hairpins that polyethylenglycol of 5-15 percent weigh did not increase yields of doubly ligated dumbbells significantly, and thus was not used in ordinary ligation reactions. Rates of ligation increases with temperature until inactivation when denaturing temperature of the enzyme is reached. The inactivation temperature of T4 DNA ligase, assayed by closing DNA nicks, was reported to be about 40 °C (Pohl et al., 1982). We studied the temperature dependence of dumbbell ligation. The sticky-ended hairpin (having sequence 5'-TCGAGGGATACTTTTGTATCCC-3' present at 15 μM) was ligated at 4, 16 and 36 °C in 1-2 mM ATP, 20 mM DTT, 10 mM $MgCl_2$, 50 mM Tris-HCl, pH = 7.8. The amount of doubly ligated dumbbell was estimated from PAGE gels. The highest yield of dumbbell (70%) was achieved at 16 °C, with about 10% lower yield at 4 °C and significantly lower yield (30%) at 36 °C. Unlike Pohl's study which assayed T4 ligase activity in a 15 minute experiment, our ligation reaction of dumbbells lasted more than 15 hours. If the enzyme denaturing process is slow at 36 °C, it may not be detected in a 15 minute assay, but it could be significant in overnight ligations. This may be the reason why we determined the optimal temperature of dumbbell ligation (16 °C) to be lower than the optimal temperature (38 °C) reported in Pohl's study.

DNA ligase isolated from Escherichia coli (Panasenko et al., 1977; 1978) could have also been used to ligate hairpins into DNA dumbbells with similar efficiency. However, because this DNA ligase was more expensive than T4 ligase, it was not used.

After ligation, dumbbell samples were desalted using Centricon-3 concentrators (Amicon), or by dialysis against ddH_2O. Dialysis tubing with molecular weight cut-off 3500 g/mol was employed. After ligation reactions, samples contained a mixture of reactants (hairpins), nicked dumbbell (once ligated product) and doubly ligated DNA dumbbell. These components were separated by polyacrylamide gel electrophoresis. DNA was dried down on a SpeedVac SVC100 concentrators and purified on 15% denaturing (7 M urea) polyacrylamide gels in 1 x TBE buffer (1 mM Na_2EDTA, 50 mM Tris-boric acid buffer, 7 M urea, pH=8.1).

Because doubly ligated dumbbell and once ligated nicked dumbbell had the same length and they tend to have very similar gel mobilities, it was necessary to optimize the PAGE purification procedure. Interactions of DNA with the gel matrix are complex and not fully understood (Neiderweis et al., 1994). Various electrophoresis conditions have been experimentally tested (Landick et al., 1984). Initially, we employed a discontinuous PAGE with a 3% stacking gel (containing 0.125 M Tris-HCl, 7 M urea, pH = 8.9) and 15% separation gel (containing 0.375 M Tris-HCl, 7 M urea) run at 270 V for 90 minutes (Doktycz, 1993, Chrambach and Jovin, 1983; Jovin, 1973; Maurer, 1971; Zsolnai et al., 1993) with electrode buffers containing 0.025 M Tris and 0.192 M Glycine, pH = 8.3. Separation of doubly ligated dumbbells and singly ligated nicked dumbbells could not be achieved under this conditions.

Alternatively, we attempted to optimize PAGE with a gel containing 7 M urea and TBE buffer. Addition of 20% glycerol to the loading buffer provided poorer product separation in spite of decreasing diffusion and “uplifting” during loading of the sample in gel wells.

We found gel temperatures to be the most critical variable for clean separation of the dumbbell, nicked dumbbell and unreacted hairpin species. The temperature of a 15% PAGE with a gel containing 7 M urea and TBE buffer must be carefully adjusted according to the G•C base pair content of the particular dumbbell sequence. Gel mobilities vary significantly with temperature as shown on Figure 2. Molecules having a fraction of G•C base pairs, *f*(G•C) = 0.33 were purified at 48 °C. Molecules with *f*(G•C) ranging from 0.4-0.5 were purified at 51 °C, and molecules having *f*(G•C) > 0.51 were run at 53 °C. Under these moderate denaturing conditions, the nicked dumbbells melt to a single strand while the doubly ligated DNA dumbbells remain intact. Therefore, under these conditions, the migration of the compact doubly ligated DNA

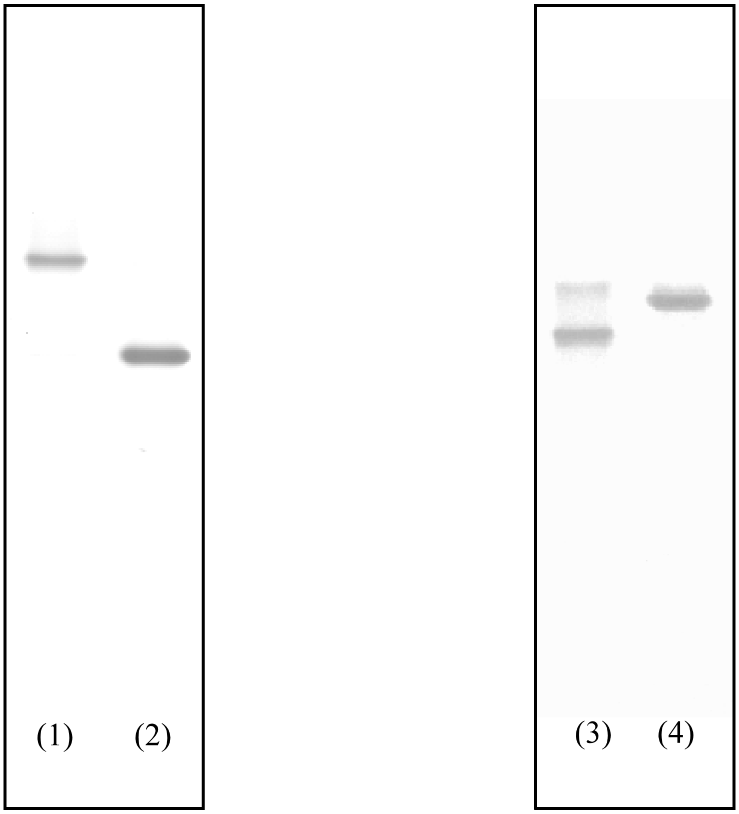


Figure 2. A dramatic effect of temperature on electrophoretic mobility of once ligated nicked dumbbell and doubly ligated dumbbell having the central sequence of 5'-AAGGTTCC-3'. Procedures for electrophoresis are described in Chapter 2.3. Both gels contain 15% acrylamide and 7 M urea, the only difference between gels is the temperature of gel during electrophoresis. DNA samples were loaded at the top and moved down during electrophoresis. The gels were scanned at 300 dpi resolution and backgrounds were subtracted. Lanes: (1) nicked dumbbell and (2) doubly ligated dumbbell at 55 °C, (3) nicked dumbbell and (4) doubly ligated dumbbell at 65 °C. As can be seen, excellent separation of nicked dumbbell and doubly ligated dumbbell species is achieved at 55 °C.

dumbbells was greater than the migration of the denatured once ligated nicked dumbbells. Thus, DNA molecules of the approximately same molar mass (doubly ligated dumbbells and once ligated nicked dumbbells) exhibited very different mobilities and could be efficiently separated by PAGE.

Other details of the purification procedure were exactly the same as for oligomer single strand purification (Chapter 2.4). After brief UV shadowing all major bands were sliced out and DNA was electroeluted from the gel in TBE buffer and filtered through a 0.45 µm nylon filter. Three different procedures were applied to remove extraneous acrylamide and TBE buffer from the eluted DNA sample, and to concentrate DNA.

Solutions of DNAs were concentrated to 1-2 ml under vacuum on a SpeedVac SVC100 concentrator. Initial sample volumes after electroelution were 25-40 ml. Concentration of the DNA sample resulted in an increased concentration of TBE buffer by 12-20 times. Sep-pak Vac C-18 Cartridges (Millipore) were initially rinsed with 10 ml of acetonitrile (Burdick and Jackson Company) and 15 ml doubly distilled water. To remove the salt and acrylamide, concentrated DNA solutions were loaded on the cartridge and 8 ml of ddH$_2$O was applied. Cartridge bound DNA was eluted with 10 ml of 40% acetonitrile solution and lyophilized on SpeedVac SVC100 concentrator. It was essential that concentrated DNA dumbbells were loaded on the cartridge in 12-20 x TBE buffer. Amounts of dumbbells bound to the Sep-pack cartridges were estimated from absorbance of the loaded solution at 260 nm and the appropriate extinction coefficients of the dumbbells. Absorbance of the loading solution was measured before loading and upon elution from the cartridge. When dumbbells were loaded in 1 x TBE buffer, they did not bind to the cartridge. However, when the loading buffer contained 0.1 M NaCl and 1 x TBE, 95% of dumbbell sample was bound. These results suggested that the binding of dumbbells to the cartridge are influenced by ionic strength of the loading buffer. Apparently, dumbbells will not bind to the Sep-pack C-18 cartridges at ionic strengths below 0.1.

Alternatively, oligodeoxyribonucleotides were concentrated in Centriprep-3 concentrators (Amicon) by centrifuging repeatedly at 3000 x g for 60 minutes. These samples were desalted in Centricon-3 concentrators (see Chapter 2.7 for details).

Some dumbbells were desalted by hydoxylapatite chromatography. A column (2.0 cm long, 1.0 cm internal diameter) was prepared from 1.75 g of hydroxylapatite (Biorad) and 0.25 g fibrous phosphocellulose (Whitman) dispersed in 100 ml ddH_2O. Phosphocellulose increased the flow of solutions through the column. After washing with 100 ml ddH_2O, electroeluted DNA samples were loaded on the column. Salts were washed out with approximately 100 ml ddH_2O and DNA was eluted with 30 ml of 0.8 M Na_3PO_4 buffer (pH = 7.0) and collected in small microcentrifuge tubes (1.5 ml) using a microfraction collector (Gilson, model 203). The absorbance of solutions in all microcentrifuge tubes was measured. Solutions showing $A_{260} > 0.1$ were placed in a dialysis bag (MWCO 3500) and dialyzed exhaustively against distilled water for 5-6 days to remove sodium phosphate. Because of high osmotic pressure, the dialyzed solution increased in volume by about 2.5 times. After dialysis, the sample was lyophilized on a SpeedVac SVC100 concentrator.

Three procedures described above yield DNA dumbbells with the identical thermodynamic properties. Purity of each dumbbell was always confirmed by denaturing analytical gel electrophoresis. Identity of all compounds was verified by Exonuclease III assay (described in the next section). Yields of purified dumbbells ranged from 20 to 60 percent.

Enzymatic ligation efficiency and losses of DNA during preparative electrophoresis seemed to be the major factors responsible for lower yields. Wide range of yields was observed but is not surprising. T4 DNA ligase is substrate specific (Harada and Orgel, 1993) and ligation is more efficient for some sequences. Because T4 DNA ligase requires a 5'-phosphate group, failure to attach it during chemical synthesis of dangling ended hairpins could also attribute to lower yields. Polynucleotide kinase can be used to attach 5'-phosphate groups to hairpins. The enzyme requires ATP and Mg^{2+} and is active in the same buffer as T4 ligase. It is possible to employ both T4 ligase and kinase at the same. To test whether DNA hairpins had 5'-phosphate groups, 150 units of T4 polynucleotide kinase (New England Biolabs) was added with 25 x 10^3 NEB units of T4 DNA ligase to ligation mixtures. Ligations were carried out at 16 °C for 15 hours as described above and monitored by PAGE. No significant increase in dumbbell yield was observed suggesting that the majority of hairpins had a 5'-phosphate moiety. Thus, the presence of polynucleotide kinase during ligation reaction was unnecessary.

2.6 The Exonuclease III Assay

This assay confirmed that dangling ended hairpins were properly ligated and the desired samples , doubly ligated closed circular molecules, DNA dumbbells, were obtained. Because Exo III requires a 3'-hydroxyl in a DNA duplex, exposure of the ligation mixture to this enzyme results in digestion of all unligated hairpins and singly ligated nicked dumbbells. The doubly ligated dumbbell does not contain any ends, therefore, it is not a substrate that can be degraded by exonuclease III (Weiss, 1976). This assay was used to verify that samples recovered from preparative gel electrophoresis were in fact pure doubly ligated dumbbells. An assay of this type was required because the relative positions of electrophoretic bands corresponding to the doubly ligated dumbbell and once ligated nicked dumbbell were sequence specific, and depending on the temperature, these molecules could have reversed gel mobilities during electrophoresis.

Assays were performed in 20 µl containing 0.6 nmol of DNA mixed with 0.6 nmol of added hairpin standard (formed from the sequence 5'-TTGAAGGATACTTTTGTATCCT-3') in 6.3 mM $MgCl_2$, 6.3 mM DTT, 63 mM Tris-HCl, pH = 7.5, 37 °C. The added hairpin served as an internal standard for enzyme activity and displayed much faster gel mobility than any dumbbell. Rapid degradation of the hairpin standard verified enzyme activity. At the outset a 4 µl aliquot (1/5 of the reaction mixture) was removed and 2.0 µl of Exo III (New England Biolabs, 100,000 U/ml) was added to the reaction mixture. Subsequently, 4.5 µl aliquots were taken at intervals of 2, 10 and 60 min, and 20 hrs. Exonuclease activity in each aliquot was arrested by addition of 5 µl of 0.1 M EDTA solution and heating at 85 °C for 10 minutes. After addition of 2 µl of 8 M urea and 2 µl of indicator (0.25% bromophenol blue and 0.25% xylene cyanol), samples were loaded on denaturing polyacrylamide gels run at 65 °C. A representative picture of one such test is shown in Figure 3. Since DNA dumbbells are circular and have no free ends, they are not attacked and digested by the enzyme. Intensity of electrophoretic bands corresponding to dumbbells did not differ before or after treatment with Exo III. This is not the case for the hairpin standard and once ligated nicked dumbbells. Both of these molecular species have 3'-hydroxyl termini in DNA duplex. Therefore, they are good substrates for the enzyme and are completely degraded within several hours of exposure, as evidenced by the diminishing intensity of their gel bands with time of exposure. Immunity from Exo III digestion confirmed the presence of pure, doubly ligated dumbbell samples.

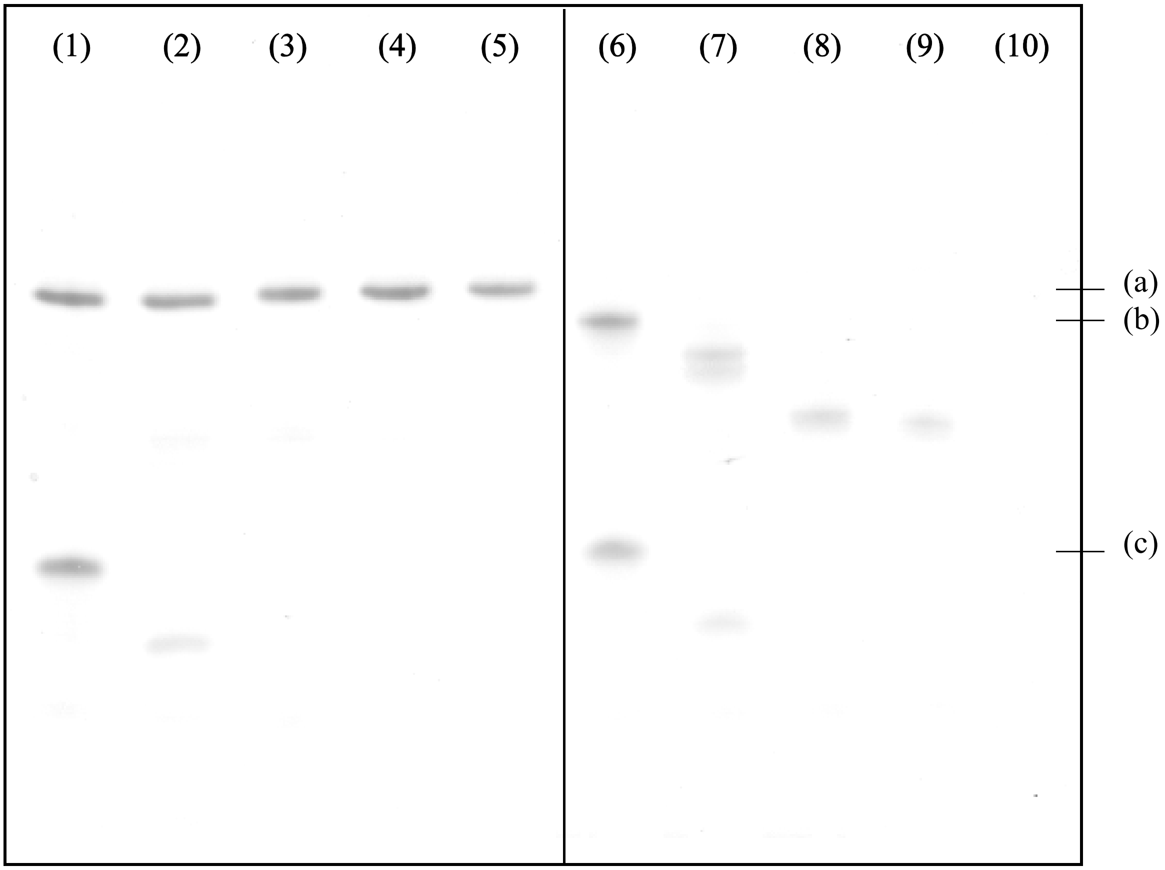


Figure 3. Results of an exonuclease III assay for doubly ligated dumbbell and once ligated nicked dumbbell with the central sequence 5'-AAGCTT-3'. The gel was scanned at 300 dpi resolution and background noise was subtracted. Lanes (1) - (5) contain mixtures of the dumbbell and hairpin standard (control). Lanes (1), (2), (3), (4), (5) correspond to aliquots taken at the intervals of 0 min, 2 min, 10 minutes, 1 hour, and 18.5 hours, respectively, after addition of exonuclease III. Lanes (6) - (10) contain mixtures of once ligated nicked dumbbell and hairpin standard taken at 0 min, 2 min, 10 minutes, 1 hour, and 18.5 hours, respectively. After addition of exonuclease III, intensity of bands corresponding to doubly ligated dumbbell (a) in lanes (1)-(5) did not differ from before exposure to Exo III. The once ligated nicked dumbbell (b) in lanes (6)-(10) and hairpin standard (c) are completely degraded within several hours of exposure as shown by their gel bands that have diminishing intensity with increased time of exposure.

2.7 Ultraviolet Melting Experiments

Melting buffer must be chosen carefully (Good et al., 1966). It should have maximum water solubility, it should be as stable as possible and its pK_a should be around 7 since this is the region where we planned to investigate DNA melting behavior. The absorbance of buffer about 260 nm should also be very low. Finally, the dependence of pK_a on temperature should be minimal to maintain a stable pH over the temperature range (15-100 °C). The acidic dissociation constant ($K_a(T)$) for ionization of buffer depends on temperature (T) as follows,

$$-\ln K_a(T_2) = -\ln K_a(T_1) + \frac{\Delta H_{ion}}{R} * \left(\frac{1}{T_2} - \frac{1}{T_1} \right) - \frac{\Delta C_p(ion)}{R} * \left[\frac{T_1}{T_2} - 1 + \ln\left(\frac{T_2}{T_1} \right) \right] \qquad (2.2)$$

Equation (2.2) is the integrated form of the van't Hoff equation (Clarke and Glew, 1966). If the enthalpy of ionization (ΔH_{ion}) and heat capacity change for buffer ionization ($\Delta C_p(ion)$) are small, then the changes in K_a with temperature are minor and pH is kept relatively constant during the course of melting experiments. For many buffers, the temperature dependence of K_a has been published (Dawson et al., 1986; Fasman, 1989). Biological buffers based on weak bases (Tris, imidazole) have large ionization enthalpies and should be avoided. Buffers based on weak acids tend to have small ionization enthalpies and their temperature corrections are relatively small. To fit these criteria suitable DNA melting buffers contain citrate, cacodylate ($As(CH_3)_2O_2^-$) or phosphate. All melting studies in this thesis were perform in phosphate buffer. Ethylendiaminetetraacetic acid (EDTA) was added to remove traces of divalent and trivalent ions, that may bind to DNA and could strikingly influence the melting process. Unfortunately, EDTA also noticeably increased UV absorbance at 268 nm but in a minor way. Sodium chloride was used to adjust the ionic strength and total Na^+ concentration.

Melting buffers contained 5 mM Na_2HPO_4 (Amresco Company), 5 mM NaH_2PO_4 (Amresco Company), 1 mM Na_2EDTA, with either 10, 40, 70 or 100 mM NaCl (Amresco Company), pH=6.80. The procedure produced solutions with the final Na^+ concentration of 25 mM, 55 mM, 85 mM, 115 mM respectively. Total Na^+ concentrations of melting buffers were analytically verified by electrical conductivity (κ) measurements. Measurements of the conductivity (κ) were made on a Accumet model 30 conductometer (Fisher Scientific) using a YSI 3417 electrode ($K=1.0\ cm^{-1}$, YSI Inc.). The meter was calibrated with a standard solution having the electric conductivity of 1 mS/cm at 25 °C (Fisher Scientific). Conductivities of buffers change rapidly with temperature. For this reason, conductivities of DNA melting buffers were determined in the temperature range from 10 to 28°C. At least 5 points were measured for each melting buffer and the electric conductivity was plotted vs temperature. These plots of κ vs temperature (T) were then least-squares fit to a straight line. For these plots correlation coefficients were always greater than 0.99. The following relationships were obtained from the measurements. For the 25 mM Na^+ melting buffer, $\kappa = (0.04691*T(°C) + 1.251)$ mS/cm. For the 55 mM Na^+ melting buffer, $\kappa = (0.09131*T(°C) + 2.907)$ mS/cm. For the 85 mM Na^+ melting buffer, $\kappa = 0.1293*T(°C) + 4.318$ mS/cm and for the 115 mM Na^+ melting buffer, $\kappa = (0.1624*T(°C) + 5.637)$ mS/cm. Melting buffers had conductivities within 5% of the values calculated by these equations. Melting buffers which did not fulfil this criteria were discarded. Measured conductivities were in a good agreement with published conductivities of sodium phosphate (Muccitelli and DiAngelo, 1994) and sodium chloride solutions (Wadi and Saxena, 1995; Dobos, 1975).

DNA stock solutions were diluted in aliquots that would give absorbance readings at 268 nm of 0.15 - 0.80. Three methods were employed to transfer samples into melting buffers. Regardless of the method essentially the same melting curves and melting temperatures were obtained.

DNA was transferred to melting buffer by Centricon-3 (Amicon) concentrators. A sample was pipetted into a sample reservoir and diluted with appropriate melting buffer to 2 ml. After centrifugation at 5000 x g on a Sorvall RC5-B ultracentrifuge, volume of the solution was reduced to about 200 µl. DNA duplexes were retained and concentrated. Under these conditions, any molecules with a molar mass less than 3000 g/mol, including the components of melting buffer, were not retained. Concentrated DNA solutions were brought up to a volume of 2 ml by adding melting buffer. Centrifugation was repeated. The procedure was repeated at least four times to remove (dilute out) any extraneous salt or small molecules in the samples. To collect the concentrated DNA solution, the retentate vial was placed over the sample reservoir, inverted and centrifuged at 800 x g for 3 minutes.

For melting experiments, DNA dumbbell samples were also equilibrated in melting buffer by equilibrium dialysis using Cellu•Sep T1 tubular membrane (MWCO 3500; Membrane Filtration Products) or Slide-A-Lyzer 2K Dialysis Cassettes (Pierce Chemical Company).

Prior to melting experiments, 1.1 ml DNA sample in melting buffer was gently filtered through a 0.45 µm nylon filter into a dry quartz cuvette. This procedure ensured that large dust particles were removed. Cuvettes with a path-length 1 cm were used. DNA solutions were degassed by bubbling with a fine stream of helium gas (AGA Gas) for at least 30 minutes. To prevent evaporation, solutions were then covered with 2 drops of mineral oil (Aldrich). To monitor temperatures, a Teflon coated temperature probe (BetaTherm Corporation) was directly immersed into the DNA solution in the sample cuvette. Teflon tape was wrapped around the top of the cuvette to seal it and prevent evaporation. The biggest systematic errors of temperature measurements came from calibration of the temperature probe. Each temperature probe was routinely checked by measurement of the temperatures of an ice/distilled water mixture (presumably 0 °C) and boiling distilled water (100 °C).

Optical melting curves were acquired as the absorbance at 268 nm (A_{268}) vs temperature. Absorbance was measured every 0.1 °C over the temperature range from 20-100 °C. Both heating and cooling melting curves were collected. Two different Hewlett-Packard 8452 diode-array single beam spectrophotometers were employed to minimize the possibility of systematic errors. The spectrophotometers were blanked to the absorbance of air. Temperature of the metallic cuvette holder was regulated with a Hewlett-Packard 89090A Peltier Temperature Controller and monitored by a thermistor probe. The system was interfaced to a 06286 or 06386 personal computer that controlled acquisition and storage of data. Cuvettes were heated and cooled at a rate of 25-30 °C/hour. At least three heating/cooling melting curves from fresh samples were measured for every dumbbell at two different concentrations (typically 0.4 and 1.2 μM).

The linear 10 base pair DNA duplexes that were studied had relatively low melting temperatures (below ~25 °C). For these samples, to measure an accurate lower baseline on their melting curves, a HP 8452 spectrophotometer and computer were placed in a cold room maintained at 10 °C. This cooler surrounding ambient temperature limited condensation on cuvette surfaces at low temperatures and allowed data collection to start at 7.4 °C.

Since the spectrophotometers were single beam construction, melting curves of buffer alone were also collected in the same sample cuvette and subtracted digitally from raw melting curves of DNA samples. This correction guaranteed that small changes of absorbance due to temperature changes of the extinction coefficient of the buffer did not contribute to DNA melting curves.

3. THEORY OF DNA DENATURATION

3.1 Introduction

Generally two distinct models are used to study DNA melting curves. To obtain the free energy of DNA denaturation, the resulting melting curve is analyzed either in terms of a concerted two-state model or a multistate zipper model. The two-state, also called all-or-none model assumes that DNA can reside solely in two states. Only the intact duplex and completely melted single strands are assumed to be significantly present during the melting transition.

The multistate model, is based on the statistical mechanical description for duplex states with varying degrees of hydrogen-bonded base pairs. It takes into account also partially melted duplexes that can be significantly populated during melting transition (Paner et al., 1992; Wartell and Benight, 1985).

For both models, relevant thermodynamic equations required to analyze the melting profiles depend on molecularity of melting reaction. The nucleic acid systems studied in this thesis have either monomolecular (DNA dumbbells) or bimolecular (linear duplexes) melting transitions. We must be careful to apply the appropriate analysis for a system of given molecularity. General equations for DNA denaturation of any molecularity have been published (Marky and Breslauer, 1987) and are applied later in this chapter.

3.2 Analysis of Raw DNA Melting Curves

From collected raw absorbance (A) vs temperature (T) curves, the total fraction of melted (broken) base pairs, θ_B, was determined using the formula,

$$\theta_B = \frac{A(T) - A_L(T)}{A_U(T) - A_L(T)} \tag{3.1}$$

where A(T), A_L(T), and A_U(T) are the absorbances of the experimental curve, the lower baseline, and the upper baseline, respectively, at a given temperature T. The upper baseline, A_U(T), is obtained by linear fitting of measured absorbances after the melting transition. Similarly, the lower baseline, A_L(T), is obtained by least-squares fitting to the melting curve at low temperatures, before onset of the melting transition (Wartell and Benight, 1985). Some dumbbells had melting temperatures above 90 °C and their melting curve lacked a sufficient number of points to accurately fit the upper baseline. For these cases the slope of the upper baseline was set to the slope of lower baseline and the y-intercept was chosen so that the upper baseline passed through the measured high temperature absorbances. There is admittedly some arbitrariness regarding this manner of choosing baselines. In general, attempts were made to make the baselines parallel. Slopes of the upper and lower baselines never differed more than 15%. Although uncertainties in the upper baselines occurred for some molecules, complete melting transitions were observed in all cases. That is, the entire sigmoidal shaped melting transition was obtained and absorbances leveled off at the highest temperatures. Accordingly, reported melting temperatures for all molecules have similar precision.

To remove instrumental noise, plots of θ_B vs T were smoothed by digital filter (Kaiser and Reed, 1977). Typically, the Nearly Equal Ripple Approximation smoothing method was used with the following parameters $\beta = 0.030$, $\delta = 0.030$ and $\lambda = 50.0$ dB. These produced a weighting window of 197 data points. Using numerical differentiation, the derivative curve,

$\left[\frac{d\theta_B}{dT}\right]$ vs T, was also calculated.

For linear duplexes, the melting temperature, T_m, is defined as the temperature at which half of the total base pairs are “melted”, i.e., $\theta_B = 0.5$. Melting temperatures were robust and changed at most approximately 0.1 °C for different reasonable choices of baselines. The maxima of $\left[\frac{d\theta_B}{dT}\right]$ vs T plots, $\left[\frac{d\theta_B}{dT}\right]_{T=T_m}$, were more affected by subjective selection of the baselines and varied as much as 15%. Thus, $\left[\frac{d\theta_B}{dT}\right]_{T=T_m}$ values have higher relative errors then T_m’s. For all samples, heating and cooling melting curves overlapped, indicating melting transitions were at equilibrium and reversible. Because the differences between heating and cooling curves were small and within the random error of experiment, melting curves and T_m’s from heating and cooling curves were averaged. Software which performed all analyses described in this chapter was written in FORTRAN77. Executable programs were generated by a Microsoft Fortran 4.0 compiler for MS-DOS.

3.3 Dumbbells

3.3.1 Some Remarkable Studies of DNA Dumbbells and Other Circular Oligomers

Nucleic acid dumbbells are molecules with unique noteworthy properties. They are closed circular loops without ends. The sequence of the circular single-strand is partially self-complementary and forms a duplex at temperatures below T_m. Thus, the intact dumbbell has a duplex stem linked on the ends by hairpin loops. When melted the dumbbell converts to a minicircle single strand. DNA dumbbells and small circular oligomers have been the subjects of several studies (Doktycz et al., 1992; Liu et al., 1996; Yamakawa et al., 1997; Abe et al., 1998; Lim et al., 1997). Some of these are described below.

Small circular DNA oligomers comprised of 26 to 74 nucleotides can serve as catalytic templates for DNA synthesis by several DNA and RNA polymerase enzymes (Liu et al., 1996). In the rolling circle mechanism, the polymerase processes the circular template multiple times and synthesizes a long complementary single-strand, whose sequence repeats according to the size of the single strand circular template. That is surprising for two reasons. First, although the polymerase synthesizes a duplex using a single-stranded circular template, the resulting duplex is probably not stable because of the tight radius of curvature. Second, the circular DNA templates have dimensions smaller than the polymerase. Even so, polymerization proceeds efficiently, and repeating sequence single strands up to 12 000 nucleotides long have been prepared (Liu et al., 1996).

Chimeric RNA/DNA dumbbells have also been investigated for their potential antisense properties. Such constructs may be used to deliver antisense DNA sequences to a cell with the intent of obstructing gene expression. Constructions of dumbbells comprised of a sense RNA sequence, complementary antisense DNA sequence and hairpin loops structures on both ends were reported (Yamakawa et al., 1997). When this construct is delivered to the cell, the antisense DNA is liberated with RNase H digestion and binds to the targeted mRNA. In this scheme, dumbbells provide several advantages over linear DNAs. They have increased nuclease resistance because they lack any free ends. In fact, a primary nuclease activity in the cell is due to exonucleases that degrade DNA from the 3'-end. Since the dumbbells do not have ends, the enzymes cannot degrade dumbbells. Moreover, dumbbells have much higher cellular uptake compared to linear antisense phosphodiester oligonucleotides (Abe et al., 1998). An *in vivo* study using an antisense chimeric dumbbell showed (Abe et al., 1998) that the dumbbell can effectively enter the cell, slowly release the antisense DNA and have a significant inhibitory effect on expression of genes of the influenza virus. Alternatively, DNA dumbbells can be used to target proteins that regulate expression of specific genes. In this application a competition exists between the actual promoter regulatory sequence and the same sequence in a dumbbell. By binding transcription factors, the dumbbell depletes the intracellular supply and the transcription factors are not available to activate gene transcription. Inhibitory effects were reported for a dumbbell with sequence containing a positive regulatory binding motif of the promotor of the Major Histocompatibility Complex (Lim et al., 1997).

In this thesis, DNA dumbbells served as a model molecular system for studying sequence dependent thermodynamics of duplex DNA. These studies are similar to those done earlier in our lab (Doktycz et al., 1992). There are several advantages to using dumbbells to study DNA thermodynamics. Dumbbells are more stable than the analogous linear duplexes containing the same sequence without end-loops. The main reason for this is the large difference in the conformational entropies of the two melted forms. The open circle of the melted dumbbell can assume only a fraction of the conformations available to the melted single strands of the linear duplex. Dumbbells have higher T_m's. Thus, it is easier to reliably measure their melting transitions, particularly in low salt environments, where linear duplexes may be unstable. Because their duplex sequences are essentially cross-linked, dumbbells more realistically simulate melting behavior of short sequences in long DNA environments. In addition, the thermodynamic parameters of loop formation can be deduced from studies of dumbbell. Because dumbbells have no ends, their melting process is monomolecular making their melting transitions and T_m independent of dumbbell concentration. If the T_m of a presumed dumbbell sample depends on DNA concentration, the dumbbell has become degraded or other process (intermolecular aggregation) has occurred. The major drawback to using DNA dumbbells is that their synthesis and preparation is cumbersome and more costly than linear duplexes.

3.3.2. The Two-state Model of Dumbbell Melting

The two-state or all-or-none model assumes that no intermediate states other than the fully intact and entirely melted dumbbells are significantly populated throughout the melting transition. For DNA dumbbells, the melting temperature T$_m$, is defined as the temperature where $\left[\frac{d\theta_B}{dT}\right]$ reaches its highest value. Consider the reversible monomolecular annealing process for dumbbells,

$$C \overset{K_D}{\rightleftarrows} D \tag{3.2}$$

Where C is the melted circle and D is the intact dumbbell. The equilibrium constant of the annealing reaction is,

$$K_D = \frac{[D]}{[C]} \tag{3.3}$$

And the total DNA concentration is given by,

$$C_T = [C] + [D] \tag{3.4}$$

where [C] and [D] are equilibrium concentrations of melted circles and intact dumbbells, respectively. Assuming an equilibrium chemical process, the total free energy, ΔG_D can be determined from the temperature dependence of the equilibrium constant,

$$\Delta G_D = \Delta H_D - T\Delta S_D = -RT\ln K_D \tag{3.5}$$

where ΔH_D and ΔS_D are the total enthalpy and entropy changes accompanying the melting transition, respectively. If both sides of equation (3.5) are divided by T and differentiated with respect to 1/T, then the familiar van't Hoff equation is obtained (Marky and Breslauer, 1987),

$$\Delta H_D = \Delta H_{VH} = -R\left[\frac{d\ \ln K_D}{d\left(\frac{1}{T}\right)}\right] = RT^2\left[\frac{d\ \ln K_D}{dT}\right] \tag{3.6}$$

To complete the calculation, the temperature dependence of K_D must be known. It is obtained from the melting curve. Because the melting curve depicts the dependence of θ_B on temperature, the relationship between K_D and the fraction of broken (θ_B) or intact (θ_{net}) base pairs must be known. The fraction of melted dumbbells is given by,

$$\theta_B = \frac{[C]}{[D] + [C]} = \frac{[C]}{C_T} \tag{3.7}$$

From equations (3.3), (3.4) and (3.7), the expression for K_D in terms of θ_B is,

$$K_D = \frac{1 - \theta_B}{\theta_B} \tag{3.8}$$

Substitution for K_D in the van't Hoff expression and performing the indicated differentiation lead to the expression for ΔH_D,

$$\Delta H_D = \Delta H_{VH} = RT^2\left[\frac{d\ \ln K_D}{dT}\right] = -4 \cdot R T_m^2 \left[\frac{d\theta_{\mathrm{B}}}{dT}\right]_{T=T_m} = 4 \cdot R T_m^2 \left[\frac{d\theta_{net}}{dT}\right]_{T=T_m} \qquad (3.9)$$

Equation (3.9) was derived for the annealing reaction. The same expression with <u>opposite sign</u> applies for the melting reaction.

The transition entropy can be found from the Gibbs equation (3.5). At the melting temperature, the concentrations of melted circles and intact dumbbells are equal, and $K_D = 1$. At $T = T_m$, equation (3.5) is modified as,

$$\Delta G_D(T_m) = \Delta H_D - T_m \Delta S_D = -RT_m \ln K_D = -RT_m \ln 1 = 0 \qquad (3.10)$$

Rearrangement and combination with (3.9) yields

$$\Delta S_D = \frac{\Delta H_D}{T_m} = -4 \cdot R T_m \left[\frac{d\theta_{\mathrm{B}}}{dT}\right]_{T=T_m} = 4 \cdot R T_m \left[\frac{d\theta_{net}}{dT}\right]_{T=T_m} \qquad (3.11)$$

After the enthalpy and entropy are known, the free energy at any temperature T can be determined from equation (3.5). This assumes that ΔH_D and ΔS_D are temperature independent over the range spanned in the melting experiment. In error analysis, the biggest error often arises from the uncertainty of $\left[\frac{d\theta_{\mathrm{B}}}{dT}\right]_{T=T_m}$. If this error is denoted $\sigma\left(\left[\frac{d\theta_{\mathrm{B}}}{dT}\right]_{T=T_m}\right)$ and the error of T_m

is denoted σ(T_m) then the following formulas can be used to calculate errors of the thermodynamic variables,

$$\sigma^2_{\Delta H_D} = \left\{ \frac{\partial \Delta H_D}{\partial \left[\frac{d\theta_\mathrm{B}}{dT} \right]_{T=T_m}} \right\}^2 \cdot \sigma\left(\left[\frac{d\theta_\mathrm{B}}{dT} \right]_{T=T_m} \right)^2 + \left(\frac{\partial \Delta H_D}{\partial T_m} \right)^2 \cdot \sigma(T_m)^2 = \tag{3.12}$$

$$= 16R^2 T_m^4 \cdot \sigma\left(\left[\frac{d\theta_\mathrm{B}}{dT} \right]_{T=T_m} \right)^2 + 64R^2 T_m^2 \left[\frac{d\theta_\mathrm{B}}{dT} \right]^2_{T=T_m} \cdot \sigma(T_m)^2$$

$$\sigma^2_{\Delta S_D} = \left\{ \frac{\partial \Delta S_D}{\partial \left[\frac{d\theta_\mathrm{B}}{dT} \right]_{T=T_m}} \right\}^2 \cdot \sigma\left(\left[\frac{d\theta_\mathrm{B}}{dT} \right]_{T=T_m} \right)^2 + \left(\frac{\partial \Delta S_D}{\partial T_m} \right)^2 \cdot \sigma(T_m)^2 = \tag{3.13}$$

$$= 16R^2 T_m^2 \cdot \sigma\left(\left[\frac{d\theta_\mathrm{B}}{dT} \right]_{T=T_m} \right)^2 + 16R^2 \left[\frac{d\theta_\mathrm{B}}{dT} \right]^2_{T=T_m} \cdot \sigma(T_m)^2$$

The melting temperature and $\left[\frac{d\theta_\mathrm{B}}{dT} \right]_{T=T_m}$ are typically uncorrelated, and their covariance is assumed to be zero.

In our analysis we assume that the average entropy change in formation of an A•T or G•C base pair is approximately the same,

$$\Delta S_{\mathrm{A \bullet T}} = \Delta S_{\mathrm{G \bullet C}} = \Delta S_{\mathrm{bp}} = (-24.85 \pm 1.74) \quad \mathrm{cal.mol^{-1}.K^{-1}} \tag{3.14}$$

Abundant support for this assumption comes from a number of calorimetric and melting studies of DNA polymers (Wartell and Benight, 1985; Delcourt and Blake, 1991; Benight et al., 1988; Klump, 1988). Assumption of a constant ΔS_{bp} value has been tested over a wide range of Na^+ concentrations from 0.020 M to 1.0 M, and found valid for DNA duplex polymers and DNA dumbbells of varying sequence composition. With the assumption of a sequence independent ΔS_{bp}, the total transition enthalpy and entropy of dumbbell duplex formation are simply,

$$\Delta H_D = \Delta S_{\mathrm{bp}} \cdot N_{\mathrm{bp}} \cdot T_m = -24.85 \cdot N_{\mathrm{bp}} \cdot T_m \qquad (3.15)$$

$$\Delta S_D = \Delta S_{\mathrm{bp}} \cdot N_{\mathrm{bp}} = -24.85 \cdot N_{\mathrm{bp}} \qquad (3.16)$$

Where N_{bp} is the number of base pairs in the dumbbell stem and T_m is the melting temperature. To calculate the error of ΔH_D, ΔS_{bp} is assumed to be exact and the error of ΔH_D is computed only from the error of T_m, $\sigma(T_m)$. Thus,

$$\sigma(\Delta H_D) = \Delta S_{\mathrm{bp}} \cdot N_{\mathrm{bp}} \cdot \sigma(T_m) = -24.85 \cdot N_{\mathrm{bp}} \cdot \sigma(T_m) \qquad (3.17)$$

In contrast to dumbbells and polymers, the entropy change upon formation of linear duplexes has been reported to be sequence dependent. The two-state model for linear duplex DNA melting is described below.

3.4 The Two-state Model of Linear DNA Duplex Melting

3.4.1 Non-self-complementary Duplexes

The thermodynamic two-state model of DNA dumbbell melting presented in Chapter 3.3 must be modified for linear DNA duplexes. The two-state model of linear duplexes assumes that intermediate states between the intact and fully melted duplexes are not significantly populated throughout the melting transition. Consider the reversible equilibrium annealing reaction of two single strands, S_1 and S_2, to form a duplex, D, with equilibrium constant, K_D (Owczarzy et al., 1997),

$$S_1 + S_2 \overset{K_D}{\rightleftarrows} D \tag{3.18}$$

K_D can be expressed in terms of the ratios of the statistical weights of the internal and external degrees of freedom of the duplex and single strands (Benight et al., 1981),

$$K_D = \frac{Z_{int}(D) \cdot Z_{ext}(D)}{Z_{int}(S_1) \cdot Z_{int}(S_2) \cdot Z_{ext}(S_1) \cdot Z_{ext}(S_2)} \tag{3.19}$$

The statistical weight ratio for the internal degrees of freedom accounts for the thermodynamic differences between the duplex and single strands for the concentration independent part of duplex formation. Sequence dependent hydrogen bonding and stacking interactions in the duplex comprise this ratio. This statistical weight ratio can be represented as,

$$\frac{Z_{int}(D)}{Z_{int}(S_1)\ Z_{int}(S_2)} = K_{duplex} \tag{3.20}$$

Where K_{duplex} is the equilibrium constant for changes in the internal degrees of freedom required for establishment of the forces (H-bonding and stacking) in duplex formation.

$$\Delta H_{duplex} - T\Delta S_{duplex} = -RT\ln K_{duplex} \tag{3.21}$$

The quantities ΔH_{duplex} and ΔS_{duplex} are the duplex melting transition enthalpy and entropy, respectively, for the sequence. These thermodynamic quantities are predicted directly from the base pair sequence using any of the available nearest-neighbor sets (Wartell and Benight, 1985; Gotoh and Tagashira, 1981; Vologodskii et al., 1984; McCampbell, 1989; Delcourt and Blake, 1991; Ornstein and Fresco, 1983; Aida, 1988; Doktycz et al., 1992; Breslauer et al., 1986; SantaLucia et al., 1996; Allawi and SantaLucia, 1997; Sugimoto et al., 1996) and are assumed to be temperature independent.

The statistical weight ratio for the external degrees of freedom contains the concentration dependence and associated factors of the external degrees of freedom involved in duplex formation (Benight and Wartell, 1983). The definition is made,

$$\beta = \frac{Z_{ext}(D)}{Z_{ext}(S_1)\ Z_{ext}(S_2)} \tag{3.22}$$

where β is the nucleation parameter. From equations (3.19), (3.20) and (3.22),

$$K_D = K_{duplex}\beta \tag{3.23}$$

The total concentration of strands, C_T, is given by

$$C_T = [S_1] + [S_2] + 2 \cdot [D] \tag{3.24}$$

At any given temperature, the net fraction of intact base pairs, θ_{net}, is expressed in terms of the external and internal degrees of freedom of the duplex and single strands, i.e.

$$\theta_{net} = \theta_{ext} \cdot \theta_{int} = 1 - \theta_B \tag{3.25}$$

Where θ_{ext} is the fraction of strands with at least one intact base pair, θ_{int} is the fraction of intact base pairs on duplex strands having at least one base pair, and θ_B is the net fraction of broken base pairs evaluated from equation (3.1). If the melting transition is two-state, and the fully intact duplex and completely dissociated single strands are the only states that are populated by every strand throughout the entire melting transition, then $\theta_{int} = 1$ and $\theta_{net} = \theta_{ext}$. Depending on whether S_1 and S_2 are self-complementary, two slightly different expressions for θ_{ext} in terms of K_D and C_T can be derived.

For the case where the strands are different ($S_1 \neq S_2$),

$$\theta_{ext} = \frac{2 \cdot [D]}{2 \cdot [D] + [S_1] + [S_2]} = \frac{2 \cdot [D]}{C_T} \tag{3.26}$$

The equilibrium constant K_D is defined as the ratio of duplex and single strands concentrations,

$$K_D = \frac{[D]}{[S_1]\ [S_2]} \tag{3.27}$$

Let us assume that $[S_1] = [S_2]$. If we substitute [D], $[S_1]$ and $[S_2]$ from equations (3.24) and (3.26), we obtain,

$$K_D = \frac{2 \cdot \theta_{ext}}{(1 - \theta_{ext})^2 C_T} \tag{3.28}$$

After rearranging, the fraction of intact duplexes is given by,

$$\theta_{ext} = \frac{1 + C_T K_D - \sqrt{1 + 2 \cdot C_T K_D}}{C_T K_D} \tag{3.29}$$

Generally, expressions for the equilibrium constant K_D and θ_{ext} depend on the molecularity of the transition and the nature of the associating species (self-complementary vs non-self-complementary sequences). We can now substitute equation (3.28) into the van't Hoff equation (3.6) and differentiate. If the melting temperature T_m, is defined as the temperature at which $\theta_{ext} = 0.5$, then an expression for ΔH_{VH} at T_m is obtained,

$$\Delta H_{VH} = 6 \cdot R T_m^2 \left[\frac{d\theta_{ext}}{dT}\right]_{T=T_m} = -6 \cdot R T_m^2 \left[\frac{d\theta_{\mathrm{B}}}{dT}\right]_{T=T_m} \tag{3.30}$$

Equation (3.30) allows determination of ΔH_{VH} from melting curves. The expression was derived for the association reaction. The same equation with opposite sign applies to melting. It should be noticed that equation (3.30) is similar to (3.9). However, the leading coefficient for linear duplex oligomers is 6 while for dumbbells the corresponding coefficient is 4. The difference arises because of the different classes of melting transition. Melting of dumbbells is monomolecular compared to the bimolecular melting transition of linear duplexes. Although equation (3.30) was derived for duplexes comprised of two non-self-complementary strands, the same equation can be used to calculate the transition enthalpy of duplexes comprised of self-complementary strands.

Because linear nucleic acid duplexes are formed from two strands, the molecularity of duplex formation is greater than one. Consequently, the melting equilibrium is concentration dependent. Thermodynamic parameters can also be evaluated from the dependence of T_m on concentration. The pertinent model analysis follows. At $T = T_m$, the fraction of intact base pairs $\theta_{ext} = 0.5$. Thus, from equation (3.28),

$$K_D(T_m) = \frac{1}{(1 - 0.5)^2 C_T} = \frac{4}{C_T} \tag{3.31}$$

Combining this result with equation (3.23), the equilibrium constant for changes in the internal degrees of freedom is,

$$K_{duplex}(T_m) = \frac{K_D}{\beta} = \frac{4}{C_T{\cdot}\beta} \tag{3.32}$$

After substitution of this expression for K_{duplex} in equation (3.21), the following is obtained,

$$\Delta H_{duplex} - T_m \Delta S_{duplex} = R T_m \ln\left[\frac{C_T}{4}\right] + R T_m \ln \beta \tag{3.33}$$

The last term on the right is the nucleation free energy,

$$\Delta G_{nuc} = -RT_m \ln \beta = \Delta H_{nuc} - T\Delta S_{nuc} \tag{3.34}$$

With these definitions and rearrangement of (3.33), an expression for $\frac{1}{T_m}$ vs $\ln\left[\frac{C_T}{4}\right]$ is obtained,

$$\frac{1}{T_m} = \frac{R}{\Delta H_{duplex} + \Delta H_{nuc}} \cdot \ln\left[\frac{C_T}{4}\right] + \frac{\Delta S_{duplex} + \Delta S_{nuc}}{\Delta H_{duplex} + \Delta H_{nuc}} \tag{3.35}$$

This expression is usually written in more compact familiar form as,

$$\frac{1}{T_m} = \frac{R}{\Delta H_D} \cdot \ln\left[\frac{C_T}{4}\right] + \frac{\Delta S_D}{\Delta H_D} \tag{3.36}$$

Where obviously, $\Delta H_D = (\Delta H_{duplex} + \Delta H_{nuc})$ and $\Delta S_D = (\Delta S_{duplex} + \Delta S_{nuc})$. Equation (3.36) is the well-known van't Hoff expression employed to evaluate ΔH_D and ΔS_D from plots of $\frac{1}{T_m}$ as a function of $\ln\left[\frac{C_T}{4}\right]$. If the melting transition is truly two-state, plots of $(1/T_m)$ versus $\ln(C_T/4)$ should yield straight lines. The slope of such a plot is $\frac{R}{\Delta H_D}$ and the intercept is $\frac{\Delta S_D}{\Delta H_D}$. Thus, the van't Hoff enthalpy can be obtained from the slope, and the intercept yields the van't Hoff entropy. The total free energy, $\Delta G_D = \Delta H_D - T\Delta S_D$, is calculated from these parameters. This analysis assumes that the plot $\frac{1}{T_m}$ vs $\ln\left[\frac{C_T}{4}\right]$ is absolutely linear, and that ΔH_D and ΔS_D are temperature independent over the range from T_m to T. It is further assumed that there is zero change in heat capacity at constant pressure for the melting transition, i.e. $\Delta C_p = 0$. It has been shown that slight curvature in van't Hoff plots due to $\Delta C_p \neq 0$ can lead to significant errors in graphically evaluated thermodynamic parameters (Chaires, 1997). On the other hand, because enthalpies and entropies derived in this graphical manner are correlated, the errors in ΔH_D and ΔS_D have a tendency to cancel out. As a result, the relative errors of ΔG_D are usually smaller than the relative errors of ΔH_D and ΔS_D. Parameters evaluated from plots of $\frac{1}{T_m}$ vs $\ln\left[\frac{C_T}{4}\right]$ are usually more accurate then those determined from $\left[\frac{d\theta_{ext}}{dT}\right]_{T=T_m}$ values by equation (3.30). If the plot of $\frac{1}{T_m}$ vs $\ln\left[\frac{C_T}{4}\right]$ is fitted in a least-squares sense, the following can be derived,

$$\frac{1}{T_m} = \frac{R}{\Delta H_D}\cdot\ln\left[\frac{C_T}{4}\right] + \frac{\Delta S_D}{\Delta H_D} = b\cdot\ln\left[\frac{C_T}{4}\right]+a \tag{3.37}$$

where the enthalpy $\Delta H_D = \frac{R}{b}$, entropy $\Delta S_D = \frac{a}{b}R$ and free energy $\Delta G_D(T) = \frac{R}{b}(1-aT)$ are given by the fitted parameters a and b and R is the gas constant. Estimates of errors in the derived thermodynamic functions are based on the variance of the y-intercept (σ_a^2), variance of the slope (σ_b^2) and their covariance (σ_{ab}) (Bevington, 1992; Snedecor and Cohran, 1980). The variance of ΔH_D is

$$\sigma^2_{\Delta H_D} = \left(\frac{\partial \Delta H_D}{\partial a}\right)^2\sigma_a^2 + \left(\frac{\partial \Delta H_D}{\partial b}\right)^2\sigma_b^2 + 2\cdot\frac{\partial \Delta H_D}{\partial a}\cdot\frac{\partial \Delta H_D}{\partial b}\sigma_{ab} = 0 + \left(-\frac{R}{b^2}\right)^2\sigma_b^2 + 0 \tag{3.38}$$

The variances of ΔS_D and ΔG_D are computed from analogous equations,

$$\sigma^2_{\Delta S_D} = \left(\frac{\partial \Delta S_D}{\partial a}\right)^2 \sigma^2_a + \left(\frac{\partial \Delta S_D}{\partial b}\right)^2 \sigma^2_b + 2\cdot\frac{\partial \Delta S_D}{\partial a}\cdot\frac{\partial \Delta S_D}{\partial b}\sigma_{ab} =$$
$$= \left(\frac{R}{b}\right)^2 \sigma^2_a + \left(\frac{-aR}{b^2}\right)^2 \sigma^2_b + 2\cdot\left(\frac{R}{b}\right)\cdot\left(\frac{-aR}{b^2}\right)\sigma_{ab} \tag{3.39}$$

$$\sigma^2_{\Delta G_D} = \left(\frac{\partial \Delta G_D}{\partial a}\right)^2 \sigma^2_a + \left(\frac{\partial \Delta G_D}{\partial b}\right)^2 \sigma^2_b + 2\cdot\frac{\partial \Delta G_D}{\partial a}\cdot\frac{\partial \Delta G_D}{\partial b}\sigma_{ab} =$$
$$= \left(\frac{-RT}{b}\right)^2 \sigma^2_a + \left(\frac{-R}{b^2} + \frac{RaT}{b^2}\right)^2 \sigma^2_b + 2\cdot\left(\frac{-RT}{b}\right)\cdot\left(\frac{-R}{b^2} + \frac{RaT}{b^2}\right)\sigma_{ab} \tag{3.40}$$

Methods to calculate variances, σ_a^2, σ_b^2 and σ_{ab} are readily available (Meyer, 1975). However, before they can be employed, the errors of the dependent (y) and independent (x) variables must be estimated. Errors of C_T are relatively low and the logarithm of C_T additionally decreases the errors with respect to the scale of the abscissa. Consequently, errors in the x-values (ln (C_T/4)) were neglected and it was assumed that all the uncertainty in a measurement originates from uncertainties in the y-values. Errors of the y-values (1/T_m) can be estimated in either of two ways, from the experimental errors of T_m determined to be 0.2 °C, or from the distances of y-values from fitted straight lines. The latter estimate was found to yield larger errors, and was applied in error analysis. The variances of $\frac{1}{T_m}$ values were evaluated from errors in fitting the slopes of $\frac{1}{T_m}$ vs $\ln\left[\frac{C_T}{4}\right]$ plots,

$$\sigma^2_{\frac{1}{T_m}} = \frac{\sum_{i=1}^{N}\left[\frac{1}{T_m(i)} - a - b\cdot\ln\left(\frac{C_T(i)}{4}\right)\right]^2}{(N-2)} \qquad (3.41)$$

where $T_m(i)$ and $C_T(i)$ are N values measured for a given duplex. From the value of $\sigma^2_{\frac{1}{T_m}}$, the variance and covariance of the slope and y-intercept were evaluated. These were in turn used to estimate errors on the graphically determined thermodynamic parameters ΔG_D, ΔH_D and ΔS_D.

The concentration based approach to obtaining thermodynamic parameters of melting reactions is not feasible in case of a monomolecular melting process as is the case for DNA dumbbells, where melting is unimolecular and entirely independent of DNA concentration.

3.4.2 Self-complementary Linear Duplexes

If the oligomeric duplex is self-complementary then the two strands comprising the duplex are the same ($S_1 \equiv S_2 \equiv$ S) and the equation (3.26) takes the modified form,

$$\theta_{ext} = \frac{2\cdot[D]}{2\cdot[D] + [S]} = \frac{2\cdot[D]}{C_T} \qquad (3.42)$$

Analogous equations for K_D and θ_{ext} are obtained,

$$K_D = \frac{[D]}{[S]^2} = \frac{\theta_{ext}}{2(1 - \theta_{ext})^2 C_T} \qquad (3.43)$$

$$\theta_{ext} = \frac{1 + 4\cdot C_T K_D - \sqrt{1 + 8\cdot C_T K_D}}{4\cdot C_T K_D} \tag{3.44}$$

These expressions are similar to equations (3.28) and (3.29), except C_T in the former equations has been replaced by $4\cdot C_T$. Substitution of equation (3.43) in the van’t Hoff equation (3.6) and differentiation yield equation (3.30). The van’t Hoff expression used to calculate ΔH_{VH} does not depend whether the sequences are self-complementary or non-self-complementary.

In an alternative method, determination of ΔH_D, ΔS_D and ΔG_D from the concentration dependence of the melting equilibria requires slight modification. For self-complementary duplexes, the $\ln\left[\frac{C_T}{4}\right]$ is replaced by $\ln(C_T)$ in equation (3.37). Thus, for self-complementary sequences, $\frac{1}{T_m}$ is plotted vs $\ln(C_T)$, and the slope (b) and y-intercept (a) are evaluated according to,

$$\frac{1}{T_m} = \frac{R}{\Delta H_D}\cdot\ln[C_T] + \frac{\Delta S_D}{\Delta H_D} = b\cdot\ln[C_T] + a \tag{3.45}$$

Thermodynamic values, ΔH_D, ΔS_D and ΔG_D are then calculated from the slope and intercept and estimates of errors are computed according to equations (3.38), (3.39) and (3.40).

4. COMPARISONS AND TESTS OF PUBLISHED NEAREST-NEIGHBOR PARAMETERS IN PREDICTION OF DNA THERMODYNAMICS

4.1 Introduction

With the primary aim of predicting DNA stability from sequence alone, over the past 15 years a number of melting studies have been conducted to evaluate sequence dependent thermodynamic stability of DNA in terms of nearest-neighbor (n-n) base pair interactions (Wartell and Benight, 1985; Gotoh and Tagashira, 1981; Vologodskii et al., 1984; McCampbell, 1989; Delcourt and Blake, 1991; Ornstein and Fresco, 1983; Aida, 1988; Doktycz et al., 1992; Breslauer et al., 1986; SantaLucia et al., 1996; Allawi and SantaLucia, 1997; Sugimoto et al., 1996). The duplex DNA samples studied varied widely in length from long repeating copolymers and DNA restriction fragments to shorter DNA dumbbells and very short (6-16 base pairs) linear synthetic oligomers. Published studies based on the n-n model have reported thermodynamic values and procedures for calculating DNA stability from any sequence (Wartell and Benight, 1985; Delcourt and Blake, 1991; Doktycz et al., 1992; Breslauer et al., 1986; Allawi and SantaLucia, 1997; Sugimoto et al., 1996). A direct comparison of n-n parameters from different laboratories is difficult because published sets were evaluated from measurements of DNA samples in different molecular environments and solvent ionic strengths. For example, the n-n set derived from melting curves of DNA dumbbells (Doktycz et al., 1992) was evaluated in 115 mM Na^+. The n-n sets reported by Wartell and Benight (1985), Gotoh and Tagashira (1981), Vologodskii et al. (1984), McCampbell et al. (1989) and Delcourt and Blake (1991) originate from melting studies of DNA restriction fragments in ionic strengths ranging from 19.5 to 200 mM Na^+. The n-n parameters of Breslauer et al. (1986), SantaLucia et al. (1996), Allawi and SantaLucia (1997) and Sugimoto et al. (1996) were evaluated from analysis of optical melting curves of a variety of short synthetic DNA duplexes in 1 M Na^+. Breslauer et al. also included optical and calorimetric melting experiments on long synthetic repeating sequence DNA polymers in their database. The n-n set of Ornstein and Fresco (1983) was developed by fitting experimental data of repeating DNA polymers with empirical potential functions. Aida (1988) evaluated n-n stacking parameters from ab initio molecular orbital calculations in vacuum. A meaningful comparison of the reported n-n sets is complicated for several reasons. First, the n-n sets were determined and presented in different statistical thermodynamic formalisms. Furthermore, the n-n sets were evaluated from analysis of melting experiments of different types of DNA samples in different solvent ionic environments. Attempts to clarify these issues and enable direct comparisons of the different n-n sets on an equal footing have been published (Doktycz et al., 1992; Benight et al., 1995).

The results in this chapter have been previously published in Biopolymers (Owczarzy et al., 1997). The studies were conducted to address the following question. How well do n-n sequence dependent stability parameters, evaluated by melting analysis of DNA dumbbells, polymers and short duplex DNA oligomers, actually predict experimental free energies and melting temperatures of short duplex DNA oligomers? In order to compare the relative stabilities of short duplex DNA oligomers, we calculated the stability of all 10-mer sequences using 11 published sets of n-n parameters. We tested how well the n-n sets evaluated from melting studies of DNA dumbbells published over seven years ago (Doktycz et al., 1992) and those evaluated from melting studies of short duplex DNA oligomers published over 13 years ago (Breslauer et al., 1986), and those new and improved sets (SantaLucia, 1996; Allawi and

SantaLucia, 1997; Sugimoto et al., 1996), compared in overall accuracy of predicting the melting stability of short duplex DNA oligomers from their sequences.

4.2 Melting Data of Short Duplex DNA Oligomers

The databases used in the analysis consist of melting data collected from the published literature and acquired in our laboratory. Results of melting analysis of 131 unique duplex DNA oligomers were recently presented by Allawi and SantaLucia (1997). Oligomers varied in length from 4 to 16 base pairs and all experiments were conducted in 1.0 M Na^+. Although they summarized results for 131 DNAs, only 108 of these were used in their evaluation of n-n parameters. These DNAs were reported to melt in a two-state manner. The additional 23 molecules for which melting data were provided were reported to exhibit marginal two-state or non-two-state melting behavior. Close examination of their list of sequences (supplied as supplementary information) reveals that one of the 108 sequences, the 8-mer, 5'-CGATATCG-3', was included twice. Furthermore, three of the sequences, 5'-GAAGCTTC-3', 5'-GGAATTCC-3' and 5'-CGCGAATTCGCG-3', were reported to display both two-state and non-two-state behavior and their melting data were also reported twice. In addition, the data reported in their paper for the sequence 5'-CAACCAACCAAC-3' differ from the values in the cited reference (Ratmeyer et al., 1994). Consequently, Allawi and SantaLucia used 107 unique duplexes in their n-n parameter evaluations. In our analysis we utilized the melting data for the 119 molecules reported by Allawi and SantaLucia to melt in a two-state or marginally two-state manner. From this list, redundant data for the three duplicate sequences given above were averaged, and data for the sequences 5'-CGCGAATTCGCG-3' and 5-CAACCAACCAAC-3' were removed from consideration. This left melting data for a total of 114 unique duplex sequences supplied by Allawi and SantaLucia that were used in our analysis. We call this collection of melting data database A. For all sequences in this database, the enthalpy, entropy, free energy and melting temperature at known DNA concentrations were experimentally measured and reported.

Additional melting temperatures for another 136 DNA duplexes were published by Doktycz and coworkers (Doktycz et al., 1995). Their set contained UV-melting data for octamers and was also determined in 1.0 M Na^+. Although they claimed to report data for 140 molecules, for four of the molecules, 5'-GCATGGAC-3', 5'-GCCTGGAC-3', 5'-GCGTGGAC-3' and 5'-GCTTGGAC-3', data were reported twice. Thus, the total number of unique sequences actually reported was 136. In addition, melting data for the duplex sequence 5'-ACAAGCTTGCATGCCT-3' was acquired from the published literature (Sheppard and Breslow, 1996). When taken together with the 114 unique duplexes in database A, our second database, called database B, consisted of results from the published literature and contained melting temperatures for 251 different linear duplexes. Melting data in both database A and database B were measured in 1.0 M Na^+ and length of sequences ranged from 4 to 16 base pairs. This data set was used to test the predictive accuracy of different published n-n parameters. We also employed this database to evaluate the influence of different forms of the nucleation free energy on overall accuracy of predicted stabilities of short duplex DNA oligomers.

4.3 Modifications of Nearest-Neighbor Models and Their Parameters

In the n-n model, sequence dependent stability is considered in terms of n-n doublets. In duplex DNA there are 10 such unique internal nearest-neighbor doublets (INN). Listed in the

5'-3' direction, these are,

AT/AT TA/TA AA/TT AC/GT CA/TG

TC/GA CT/AG CG/CG GC/GC GG/CC

In addition, linear DNA duplexes have two ends. In principle, there are four possible types of sequence dependent end interactions. Naturally, the shorter the molecule, the higher the influence of ends will be. Therefore, a general n-n model of short oligomers should also consider end interactions. As previously described (Goldstein and Benight, 1992; Gray, 1997b), if E denotes the ends, the four end interactions (5'-3') are,

TE/EA CE/EG AE/ET GE/EC

With these added interactions, the number of possible nearest neighbor sequence dependent interactions is increased to 14. Table I summarizes parameters of several variations of the n-n model. Modifications are based on various assumptions about the ends. For instance, some postulate the use of initiation (nucleation) parameter(s) (see equation (3.34) on page 47) which includes the four end interactions, TE/EA, CE/EG, AE/ET, GE/EC. Ideally, we would like to determine the enthalpy, entropy and free energy for the 14 possible n-n parameters. Theoretically, unique values for these parameters exist, but for the following reasons, they cannot be derived from any set of melting data containing any number of molecules.

First, thermodynamic parameters evaluated from melting studies of polymers, restriction

TABLE I The Number of Possible Nearest-Neighbor Doublets and Number of Unique Parameters for Several Variations of the Nearest-Neighbor Model.

Type of molecule		Nearest-neighbor model			
		General n-n model	Allawi and SantaLucia	INN model + initiation constant	INN model
Short oligomeric duplexes	Assumption about end interactions	none	$\Delta G_{ET/AE}$ = = $\Delta G_{EA/TE}$ and $\Delta G_{EC/GE}$ = = $\Delta G_{EG/CE}$	All end interactions are the same.	All end interactions are equal zero.
	Number of nearest-neighbors	14	12	11	10
	Number of unique parameters	12	12	11	10

Type of molecule		n-n model	Type of molecule		n-n model
Polymers and circular double-stranded oligomers	Assumption about end interactions	End effects are negligible.	Dumbbells with the same ends	Assumption about end interactions	End interactions are constant.
	Number of nearest-neighbors	10		Number of nearest-neighbors	10
	Number of unique parameters	8		Number of unique parameters	9

fragments and circular double-stranded DNA molecules (plasmids) do not take into account ends, because these molecules do not have ends. Consequently, only 10 internal n-n doublets are considered and end interactions are neglected. This does not present a problem for polymer thermodynamics, where end interactions are negligible in the context of many internal n-n interactions. On the other hand, to apply such a set of parameters for very short oligomers, a nucleation parameter(s) must be included. These can be derived from melting data of short oligomers.

Furthermore, the whole procedure is complicated by the fact that every base pair must have an end or neighbor on both sides. Therefore, two mathematical constraints exist, which limit the number of unique, independent parameters that can be derived from any set of sequences. These constraints are as follows,

$$N_{\mathrm{AA/TT}} + N_{\mathrm{AT/AT}} + N_{\mathrm{AC/GT}} + N_{\mathrm{AG/CT}} + N_{\mathrm{AE/ET}} = N_{\mathrm{AA/TT}} + N_{\mathrm{TA/TA}} + N_{\mathrm{CA/TG}} + N_{\mathrm{GA/TC}} + N_{\mathrm{EA/TE}} \qquad (4.1)$$

$$N_{\mathrm{GA/TC}} + N_{\mathrm{GT/AC}} + N_{\mathrm{GC/GC}} + N_{\mathrm{GG/CC}} + N_{\mathrm{GE/EC}} = N_{\mathrm{AG/CT}} + N_{\mathrm{TG/CA}} + N_{\mathrm{CG/CG}} + N_{\mathrm{GG/CC}} + N_{\mathrm{EG/CE}} \qquad (4.2)$$

Where $\mathrm{N}_{\mathrm{AA/TT}}$ is the number of times the AA/TT n-n doublet occurs in the DNA duplex, etc. These constraints are based on the assumption that each base pair in the DNA duplex has a neighbor on the left side as well as on the right side. If the base pair does not have a neighboring base pair, then it has an end "E". Hence, the number of neighbors on the left side must be equal to the number of neighbors on the right side. Let us consider the first constraint. The left side of equation (4.1) sums all stacks where an $\begin{matrix}\mathrm{A}\\ \bullet\\ \mathrm{T}\end{matrix}$ base pair has a neighbor on the right side, ($\begin{matrix}\mathrm{A\,A}\\ \bullet\,\bullet\\ \mathrm{T\,T}\end{matrix}$, $\begin{matrix}\mathrm{A\,T}\\ \bullet\,\bullet\\ \mathrm{T\,A}\end{matrix}$, $\begin{matrix}\mathrm{A\,C}\\ \bullet\,\bullet\\ \mathrm{T\,G}\end{matrix}$, $\begin{matrix}\mathrm{A\,G}\\ \bullet\,\bullet\\ \mathrm{T\,C}\end{matrix}$, $\begin{matrix}\mathrm{A\,E}\\ \bullet\,\bullet\\ \mathrm{T\,E}\end{matrix}$). The right side of equation (4.1) sums all stacks where an $\begin{matrix}\mathrm{A}\\ \bullet\\ \mathrm{T}\end{matrix}$ base pair has a neighbor on the left side. The analogous constraint (4.2) is written for the $\begin{smallmatrix}\mathrm{G}\\ \bullet\\ \mathrm{C}\end{smallmatrix}$ base pair. In essence, these constraints decrease the number of unique parameters by two. Therefore, only 10 - 2 = 8 unique n-n parameters can be determined from polymer thermodynamics, and including ends 14 - 2 = 12 unique n-n parameters can be obtained from a set of data for short oligomers. The number of nearest-neighbors as well as the number of unique parameters for different types of molecules are summarized in Table I.

Unique n-n parameters are linear combinations of the 14 possible n-n doublets. One set of canonical linear combinations for polymers and dumbbells was published by our laboratory (Goldstein and Benight, 1992). Another set of unique parameters derived from the set of independent short sequences (ISS) for oligomers was suggested by Gray (1997a). The linearly independent linear combinations can be used to calculate sequence dependent stability of DNA. Although it may be more transparent how to calculate sequence dependent duplex thermodynamics from the original 14 n-n parameters than from their linear combinations, either method produces the same result. Values of the 14 non-unique parameters can be derived from the database of molecules by singular value decomposition (SVD) (Gray, 1997a; Goldstein and Benight, 1992). Values of the non-unique n-n parameters are useful to calculate ΔG_D of any sequence, however, these n-n parameters do not represent unique physical properties and direct comparison of their values with other published sets can be misleading. Only unique linear combinations of nearest-neighbors are suitable for direct comparison. To demonstrate the fact, we used the melting data for 114 molecules summarized in database A (see page 55 for description), and derived the free energies in different formats from the same data set by SVD. Results are shown in Table II. The first two columns contain 14 non-unique nearest-neighbors

TABLE II Values of Nearest-Neighbor Free Energies at 37 °C, ΔG_{ij}, Derived in Terms of Different Variations of the Nearest-Neighbor Model from the Same Melting Data.

Nearest-Neighbors of General Model	$\Delta G_{i,j}$ (cal/mol)	ISS model	$\Delta G_{i,j}$ (cal/mol)	Nearest-Neighbors Suggested by Allawi and SantaLucia	$\Delta G_{i,j}$ (cal/mol)
AT/AT	-740.20	EATE/EATE	1123.88	AT/AT	-862.33
TA/TA	-621.84	ETAE/ETAE	1486.50	TA/TA	-499.71
AA/TT	-1013.27	EAAE/ETTE	972.94	AA/TT	-1013.27
AC/GT	-1394.11	EACE/EGTE	524.04	AC/GT	-1471.09
CA/TG	-1431.88	ECAE/ETGE	640.22	CA/TG	-1354.90
TC/GA	-1329.80	ETCE/EGAE	710.48	TC/GA	-1284.64
CT/AG	-1253.00	ECTE/EAGE	696.97	CT/AG	-1298.15
CG/CG	-2167.56	ECGE/ECGE	-131.70	CG/CG	-2135.74
GC/GC	-2163.96	EGCE/EGCE	-191.75	GC/GC	-2195.79
GG/CC	-1829.39	EGGE/ECCE	174.65	GG/CC	-1829.39
TE/EA	932.04	EAE/ETE	1986.21	EA/TE | ET/AE	993.11
CE/EG	986.11	EGE/ECE	2004.04	EG/CE | EC/GE	1002.02
AE/ET	1054.17				
GE/EC	1017.93				

and their free energies at 37 °C. The third and fourth columns list the 12 unique parameters of the ISS model (Gray, 1997a). The last two columns consist of the nearest-neighbors suggested by Allawi and SantaLucia (1997). The last set does not distinguish between orientation of the base pair at the end. For instance, the end interaction 5'-EA-3'/5'-TE-3' is considered to have the same thermodynamic properties as 5'-ET-3'/5'-AE-3'. The n-n thermodynamic values of the Allawi and SantaLucia set can be derived from the 12 unique parameters of the ISS model by assuming that,

$$\begin{aligned} \Delta G_{\mathrm{EA/TE}} &= \Delta G_{\mathrm{ET/AE}} \\ \Delta G_{\mathrm{EC/GE}} &= \Delta G_{\mathrm{EG/CE}} \end{aligned} \tag{4.3}$$

These assumptions are arbitrary, but the real values of the free energies of 14 n-n doublets cannot be derived a priori from any set of molecules. Hence, only 12 parameters are necessary to describe oligomer thermodynamics. Additionally, it can be verified that any of three sets in Table II will predict exactly the same free energy of any sequence, therefore, all sets have the same validity. For instance, imagine duplex 5'-A-T-T-A-T-G-G-G-G-C-3'. Calculation of the total free energy from 5' to 3' based on the 14 parameter set is as follows,

$$\begin{aligned} \Delta G_T = {} & \Delta G_{\mathrm{TE/EA}} + \Delta G_{\mathrm{AT/AT}} + \Delta G_{\mathrm{AA/TT}} + \Delta G_{\mathrm{TA/TA}} + \Delta G_{\mathrm{AT/AT}} + \Delta G_{\mathrm{CA/TG}} + \Delta G_{\mathrm{GG/CC}} + \Delta G_{\mathrm{GG/CC}} + \\ & + \Delta G_{\mathrm{GG/CC}} + \Delta G_{\mathrm{GC/GC}} + \Delta G_{\mathrm{CE/EG}} = +932.04 - 740.20 - 1013.27 - 621.84 - 740.20 \\ & - 1431.88 + 3\cdot(-1829.39) - 2163.96 + 986.11 \;=\; -10{,}281.37 \ \text{cal/mol} \end{aligned}$$

Application of the ISS set yields,

$$\begin{aligned}\Delta G_T = &+\Delta G_{\mathrm{EATE/EATE}} + \Delta G_{\mathrm{EAAE/ETTE}} + \Delta G_{\mathrm{ETAE/ETAE}} + \Delta G_{\mathrm{EATE/EATE}} + \Delta G_{\mathrm{ECAE/ETGE}} + \\ &+\Delta G_{\mathrm{EGGE/ECCE}} + \Delta G_{\mathrm{EGGE/ECCE}} + \Delta G_{\mathrm{EGGE/ECCE}} + +\Delta G_{\mathrm{EGCE/EGCE}} - 4\cdot\Delta G_{\mathrm{EAE/ETE}} \\ &-4\cdot\Delta G_{\mathrm{EGE/ECE}} = +1123.88 + 972.94 + 1486.50 + 1123.88 + 640.22 + \\ &+3\cdot 174.65 - 191.75 - 4\cdot 1986.21 - 4\cdot 2004.04 \ = \ -10{,}281.37 \ \text{cal/mol}\end{aligned}$$

And parameters from the last two columns of Table II (the n-n parameters of Allawi and SantaLucia) predict the free energy,

$$\begin{aligned}\Delta G_T = &\Delta G_{\mathrm{EA/TE|ET/AE}} + \Delta G_{\mathrm{AT/AT}} + \Delta G_{\mathrm{AA/TT}} + \Delta G_{\mathrm{TA/TA}} + \Delta G_{\mathrm{AT/AT}} + \Delta G_{\mathrm{CA/TG}} + \Delta G_{\mathrm{GG/CC}} + \\ &+\Delta G_{\mathrm{GG/CC}} + \Delta G_{\mathrm{GG/CC}} + \Delta G_{\mathrm{GC/GC}} + \Delta G_{\mathrm{EG/CE|EC/GE}} = +993.11 - 862.33 - 1013.27 \\ &-499.71 - 862.33 - 1354.90 + 3\cdot(-1829.39) - 2195.79 + 1002.02 \ = \ -10{,}281.37 \ \text{cal/mol}\end{aligned}$$

Evidently, all three variations of the n-n model predict the same free energy value for the duplex sequence considered. However, the thermodynamic values of particular n-n doublets (stacks) vary and depend on assumptions about the ends. For instance, the AT/AT doublet has a free energy value of -740.20 cal/mol in the general n-n model of 14 parameters, but its free energy is -862.33 cal/mol in n-n model of Allawi and SantaLucia, despite being derived from the same database of measurements. Thermodynamic values of Allawi and SantaLucia nearest-neighbor parameters are unique only in the context of an arbitrary assumption regarding the ends (Equation 4.3). There are two n-n parameters which do not change in Table II and can be determined uniquely. These are the free energies of AA/TT and GG/CC stacks (Goldstein and Benight, 1992). They are unique, because a duplex containing only one type of these nearest-neighbors can be synthesized and measured. Therefore, thermodynamic values of these two nearest-neighbors can be measured directly and are unique.

Other hypotheses about the sequence specific end interactions in short linear duplexes were claimed in earlier studies (Breslauer et al.,1986; SantaLucia et al., 1996; Sugimoto et al., 1996). Based on assumptions regarding the ends, the number of unique n-n parameters varies. There are two more variations, the end interactions are either assumed zero or constant. Again, whether thermodynamic quantities (free energies, enthalpies and entropies) for melting each of the ten unique doublets can be uniquely determined depends on assumptions about the ends. If these end interactions are assumed to be zero (INN model), then for an appropriately chosen set of molecules, in which all 10 n-n sequences are adequately represented, 10 linearly independent equations exist. Consequently, a unique solution for each of the 10 n-n base pair stacking interactions can be obtained (subject to the initial assumption regarding the ends).

If the ends are assumed to be the same, but not equal to zero, i.e.

$$\Delta G_{\mathrm{EA/TE}} = \Delta G_{\mathrm{ET/AE}} = \Delta G_{\mathrm{EC/GE}} = \Delta G_{\mathrm{EG/CE}} = \text{constant} \tag{4.4}$$

then 11 linearly independent equations can be written, and a unique solution can be obtained for the 10 n-n sequence interactions plus the end interaction.

4.4 Influence of Ends

We investigated whether n-n sequence dependent interactions with the ends contribute significantly to the stability of short duplex DNAs. The database A of 114 DNA oligomers measured in 1.0 M Na^+ was employed. As stated above, evaluation of the 10 unique n-n sequence dependent interactions in DNA depends on how n-n interactions with the ends are treated. If sequence specific ends are considered, then there are 14 unknowns, i.e., the 10 internal n-n interactions and the four possible n-n interactions with the ends. However only 12 linearly independent equations can be written and thus only 12 linear combinations of the n-n interactions can be solved for. The ISS set contains end interactions in every linear combination. Therefore, it would be more difficult to show the influence of ends using the ISS set. The n-n set used by Goldstein and Benight (1992) consists of unique canonical linear combinations, and is more suitable for comparisons because only 4 out of 12 combinations contain the end interactions (as shown in Table III). Of these 4 linear combinations, two include contributions from only the n-n sequence dependent end interactions.

To test how sensitive evaluations of the internal n-n parameters is to the treatment of the ends, two different evaluations of the 12 linearly independent combinations were made using the same melting data. In the first case, the end interactions were set to zero and the remaining linear combinations were evaluated from fits of experimental data. The outcome is presented in the middle of Table III (without ends). In the second case, all four end n-n interactions were considered and 12 linear combinations were fitted. In the latter case, no assumptions about the end interactions were made and the n-n end interactions were explicitly fit. Results are reported on the right side of Table III (explicit ends included). The transition enthalpies, entropies and free energies at 37°C were fit in the analysis. Interestingly, the first eight linear combinations, θ_1 through θ_8, which consider only the n-n sequence dependent interactions of the base pairs in the duplex (not the ends), are in reasonable agreement (within 25% for any thermodynamic

TABLE III Results of Analysis to Determine the Influence of Ends.

Linear Combination of Nearest-Neighbor Thermodynamic Properties	Without Ends			Explicit Ends Included		
	ΔH (cal.mol^{-1})	ΔS (cal.mol^{-1}.K^{-1})	$\Delta G_{37°C}$ (cal.mol^{-1})	ΔH (cal.mol^{-1})	ΔS (cal.mol^{-1}.K^{-1})	$\Delta G_{37°C}$ (cal.mol^{-1})
$\theta_1 = \theta_{AA/TT}$	-7908	-22.97	-791.5	-7750	-21.73	-1013.3
$\theta_2 = \theta_{CC/GG}$	-8441	-22.39	-1507.3	-7952	-19.77	-1829.4
$\theta_3 = \frac{1}{2}\left[\theta_{AT/AT} + \theta_{TA/TA}\right]$	-6303	-18.63	-543.8	-6373	-18.39	-681.0
$\theta_4 = \frac{1}{2}\left[\theta_{CG/CG} + \theta_{GC/GC}\right]$	-10385	-27.90	-1740.7	-9866	-24.84	-2165.8
$\theta_5 = \frac{1}{2}\left[\theta_{AC/GT} + \theta_{CA/TG}\right]$	-8084	-22.40	-1142.7	-7797	-20.59	-1413.0
$\theta_6 = \frac{1}{2}\left[\theta_{AG/CT} + \theta_{GA/TC}\right]$	-7735	-21.60	-1027.8	-7526	-20.06	-1291.4
$\theta_7 = \frac{1}{12}\left[\theta_{AT/AT} - \theta_{TA/TA} + \theta_{CG/CG} - \theta_{GC/GC} + +2(\theta_{GA/TC} - \theta_{AG/CT})\right]$	-164	-0.45	-22.1	-78	-0.17	-23.0
$\theta_8 = \frac{1}{12}\left[\theta_{AT/AT} - \theta_{TA/TA} - \theta_{CG/CG} + \theta_{GC/GC} + +2(\theta_{CA/TG} - \theta_{AC/GT})\right]$	202	0.78	-36.1	165	0.59	-15.9
$\theta_9 = \frac{1}{2}\left[\theta_{AE/ET} + \theta_{TE/EA}\right]$	0	0.00	0.0	169	-2.79	993.1
$\theta_{10} = \frac{1}{2}\left[\theta_{CE/EG} + \theta_{GE/EC}\right]$	0	0.00	0.0	-1374	-7.70	1002.0
$\theta_{11} = \frac{1}{84}\left[\theta_{AG/CT} - \theta_{GA/TC} + \theta_{CA/TG} - \theta_{AC/GT} + +2(\theta_{CG/CG} - \theta_{GC/GC}) + 6\,(\theta_{GE/EC} - \theta_{CE/EG})\right]$	362	1.11	19.0	44	0.13	2.7
$\theta_{12} = \frac{1}{84}\left[\theta_{AG/CT} - \theta_{GA/TC} - \theta_{CA/TG} + \theta_{AC/GT} + +2(\theta_{AT/AT} - \theta_{TA/TA}) + 6\,(\theta_{TE/EA} - \theta_{AE/ET})\right]$	-49	0.12	-91.7	-6	0.01	-10.2

parameter) whether the ends are explicitly fit or assumed to be zero. Further on the right side of Table III, the linear combinations corresponding explicitly to the ends (θ_9 and θ_{10}) indicate that although their fitted ΔH and ΔS values are different, their $\Delta G_{37°C}$ values are essentially identical. The same result is obtained if a single end interaction (referred to by some as a nucleation parameter) is assumed. In this case there are 11 unknowns to be solved for from 11 linearly independent equations. However, as Table III clearly shows the fact that a single 11th parameter can be fit within the error is because the free energies of the end interactions at 37 °C are numerically close. Consequently, it appears that only one initiation parameter is required to include effects of ends. Although this assumption may be valid for free energies at 37 °C, it could depend appreciably on temperature. The end interactions of the enthalpies and entropies cannot be fit within the error with a single interaction parameter. In this case, two parameters to include the end interactions seem necessary. The transition enthalpies and entropies are parameters of principal interest, because they are used in prediction of melting temperatures. Transition free energies at different temperatures are also calculated from them. In conclusion, from our analysis of the ends, it appears that the end interactions cannot be assumed to be the same or ignored in predictions of the melting behavior of short DNA duplex oligomers.

Additional small differences between published n-n sets come from unequal treatment of self-complementary duplexes. Because self-complementary duplexes have presumably a two-fold axis of symmetry, theoretically they should require a correction in the nucleation entropy compared to non-self-complementary duplexes. That is, self-complementary duplexes have an additional entropic loss upon duplex formation. The value of the symmetry correction is, ΔS_{sym} = -1.4 cal.mol^{-1}.K^{-1} for duplex formation. The symmetry correction of the enthalpy is assumed to be zero, which yields a symmetry correction of the free energy at 37 °C, $\Delta G_{sym,37^{\circ}C}$ = +0.43 kcal.mol^{-1} for self-complementary duplexes. It should be noted that these thermodynamic values are difficult to prove experimentally because the magnitudes of the errors of the transition entropies are greater then this correction. For instance, Allawi and SantaLucia (1997) reported errors of the transition entropies of n-n doublets up to 2.6 cal.mol^{-1}.K^{-1}, nearly twice the value of the symmetry correction. Consequently, accuracy of predictions of the transition entropy is not changed significantly by symmetry correction. Free energies can be measured with higher accuracy due to the significant covariance of the enthalpy and the entropy (see Chapter 3.4.1 for details). When $\Delta G_{sym,37^{\circ}C}$, was attempted to be derived from melting data, reasonable agreement was achieved with the theoretical value of 0.43 kcal.mol^{-1} (Xia, et al., 1998). Furthermore, longer self-complementary duplexes are prone to forming hairpin structures. In that case, DNAs do not melt in a two-state process and equation (3.35) is not applicable. Validity of the two-state assumption must be confirmed experimentally for longer self-complementary duplexes.

4.5 The Singlet and Doublet Formats of the Nearest-Neighbor Model

In the n-n approximation, sequence dependent interactions in duplex DNA are considered to arise from two basic sources. These are Watson-Crick base pairing (H-bonding) and stacking interactions between n-n base pairs (stacking). Within the n-n model two types of computational methods have been employed to evaluate n-n parameters and predict DNA sequence dependent thermodynamic stability. These are referred to as the singlet and doublet formats. In the singlet format, contributions from H-bonding and stacking interactions are considered separately. In the doublet format the entire n-n interaction (H-bonding and stacking) is considered in a single parameter. Detailed descriptions of these formats and how they are equated have been provided (Doktycz et al., 1992; Benight et al., 1995). Calculated results from the singlet and doublet formats are numerically equivalent as long as the appropriate correction factor for the end base pairs is employed (Benight et al., 1995). In the comparative studies of free energies described here, calculated stability of DNA oligomers was determined using 11 different n-n sets from the published literature. As reported in Table IV, for eight of these sets, sequence dependent stability was calculated using the singlet format and the parameters previously reported for these sets in the singlet format. Three sets were calculated from published parameters in doublet format (Breslauer et al., 1986; SantaLucia et al., 1996; Sugimoto et al., 1996). In the following, use of these parameters to calculate duplex thermodynamic stability is demonstrated.

4.5.1 Calculation of Duplex Stability in the Singlet Format

Although there are 10 unique n-n base pair doublets in duplex DNA, for circular polymers or semi-infinite polymers with essentially no ends, there are only eight linearly independent combinations of n-n sequence dependent free energies, ΔG_{ij}, (Goldstein and Benight, 1992; Gray, 1997a). For dumbbells with the same type of end (same sequence and loop size), there are nine linearly independent values, eight that correspond to those of the polymers and one that accounts for the explicit type of end as presented in Table I (Doktycz et al., 1992). Nine unique parameters are derived, but only eight of these parameters are employed to calculate thermodynamics of DNA duplexes. The ninth parameter is specific for dumbbells with a given end and it should be used to calculate the thermodynamics of dumbbells with the same ends.

TABLE IV List of Experimental Conditions and Calculation Formats for 11 Published Studies of Nearest-Neighbor Parameters.

Set of Parameters	Melting Buffer	$[Na^+]$ (mM)	Calculation Format
Doktycz et al., 1992	100 mM NaCl, 5 mM Na_2HPO_4, 5 mM NaH_2PO_4, 1 mM Na_2EDTA, pH=6.8	115.0	Singlet
Delcourt and Blake, 1991	74 mM NaCl, 5 mM $NaAs(CH_3)_2O_2$, 0.2 mM NaEDTA, pH=6.85	75.0	Singlet
Wartell and Benight, 1985	0.1 M NaCl, 1mM $K_2HPO_{4,}$ 10 mM EDTA, pH=7.4	100.0	Singlet
Vologodskii et al., 1984	150 mM NaCl, 15 mM sodium citrate	200.0	Singlet
Gotoh and Tagashira, 1981	15 mM NaCl, 1.5 mM sodium citrate	19.5	Singlet
McCampbell et al., 1989	100 mM NaCl, 2 mM Na_2HPO_4, 0.01 mM EDTA, pH=6.8	102.0	Singlet
Ornstein and Fresco, 1983	Empirical potential calculations fitted to normalized melting data.	50.0	Singlet
Aida, 1988	Theoretical quantum-chemical calculation.	50.0[a]	Singlet
Breslauer et al., 1986	1 M NaCl, 10 mM Na_2HPO_4, 1 mM Na_2EDTA, pH=7	1×10^3	Doublet
SantaLucia et al., 1996	1.0 M NaCl, 10 mM $NaAs(CH_3)_2O_2$, 0.5 mM Na_2EDTA, pH=7	1×10^3	Doublet
Sugimoto et al., 1996	1 M NaCl, 10 mM Na_2HPO_4, 1 mM Na_2EDTA, pH=7	1×10^3	Doublet

[a] Aida published stacking interactions in vacuum. The same sodium concentration was used for $\Delta G_{25^\circ C}$ calculations as used by Ornstein and Fresco. Thus, both sets based on quantum chemistry calculations can be easily compared.

TABLE V Linear Combinations of Subunits and Their Deviations from Average Stacking Free Energy, δG_i, for Eight Nearest-Neighbor Sets in the Singlet Format.

Numbers	Free energies
$N_1^{n-n} = N_{AA/TT}$	$\delta G_1^{n-n} = \delta G_{AA/TT}$
$N_2^{n-n} = N_{CC/GG}$	$\delta G_2^{n-n} = \delta G_{CC/GG}$
$N_3^{n-n} = N_{AT/AT} + N_{TA/TA}$	$\delta G_3^{n-n} = (\delta G_{AT/AT} + \delta G_{TA/TA})/2$
$N_4^{n-n} = N_{CG/CG} + N_{GC/GC}$	$\delta G_4^{n-n} = (\delta G_{CG/CG} + \delta G_{GC/GC})/2$
$N_5^{n-n} = N_{AC/GT} + N_{CA/TG}$	$\delta G_5^{n-n} = (\delta G_{AC/GT} + \delta G_{CA/TG})/2$
$N_6^{n-n} = N_{AG/CT} + N_{GA/TC}$	$\delta G_6^{n-n} = (\delta G_{AG/CT} + \delta G_{GA/TC})/2$
$N_7^{n-n} = N_{AT/AT} - N_{TA/TA} + N_{CG/CG} - N_{GC/GC} +2(N_{GA/TC} - N_{AG/CT})$	$\delta G_7^{n-n} = (\delta G_{AT/AT} - \delta G_{TA/TA} + \delta G_{CG/CG} -\delta G_{GC/GC})/12 +(\delta G_{GA/TC} -\delta G_{AG/CT})/6$
$N_8^{n-n} = N_{AT/AT} - N_{TA/TA} - N_{CG/CG} + N_{GC/GC} +2(N_{CA/TG} - N_{AC/GT})$	$\delta G_8^{n-n} = (\delta G_{AT/AT} - \delta G_{TA/TA} - \delta G_{CG/CG} +\delta G_{GC/GC})/12 +(\delta G_{CA/TG} -\delta G_{AC/GT})/6$

$\delta G_i^{a,b}$	1	2	3	4	5	6	7	8
δG_1	-190	-35	-54	-86	-99	-80	270	410
δG_2	146	155	178	184	300	-100	-30	1980
δG_3	-28	6	78	18	113	90	440	340
δG_4	-240	-153	-60	-154	-151	-75	-385	-2040
δG_5	-113	-49	-114	-46	-72	-30	-155	-365
δG_6	-6	44	97	108	57	0	-20	895
δG_7	-25	12	43	-23	-27	-18	199	-27
δG_8	-39	24	-94	-4	-6	18	-44	-117

[a] The n-n sets are as follows: column 1, Doktycz et al. (1992); column 2, Delcourt and Blake (1991); column 3, Wartell and Benight (1985); column 4, Vologodskii et al. (1984); column 5, Gotoh and Tagashira (1981); column 6, McCampbell et al. (1989); column 7, Ornstein and Fresco (1983); column 8, Aida (1988).

[b] Free energies are in cal.mol^{-1}.

The free energies for the eight linearly independent combinations, that do not involve the ends, are given in Table V. In the singlet format, the average singlet energy (H-bonding) of a base pair, $\Delta G_{\text{H-bond}}$, can take on two values, depending on whether the base pair is of the A•T or G•C type (Doktycz et al., 1992; see Chapter 5). The n-n sequence dependence is included as the deviation from average stacking energy, δG_{ij}, for each type of n-n sequence. These n-n sequence dependent values are assumed to be entirely enthalpic ($\delta G_{ij} = \delta H_{ij}$) and are summarized for eight n-n sets in Table V. The n-n sets are those determined from melting analysis of DNA dumbbells (Doktycz et al., 1992, column 1); melting studies of restriction fragments and polymers (Blake and Delcourt, 1991, column 2; Wartell and Benight, 1985, column 3; Vologodskii et al., 1984, column 4; Gotoh and Tagashira, 1981, column 5; McCampbell et al., 1989, column 6) and theoretical calculations (Ornstein and Fresco, 1983, column 7; Aida, 1988, column 8).

The values in Table V can be utilized to calculate the n-n sequence dependent stability of any duplex DNA oligomer. In the example calculations that follow, base pairs on the ends are assumed to behave just as any other base pair and any n-n dependent end interactions are assumed to be zero. In this case the n-n sequence dependent transition enthalpy of the duplex is written in terms of a hydrogen bonding component, $\Delta H_{\text{H-bond}}$, whose magnitude ranges with the %G•C of the duplex, and a n-n interaction component, ΔH_{ij}. The duplex transition enthalpy is then determined according to,

$$\Delta H_{duplex} = \Delta H_{H-bond} + \Delta H_{ij} = \Delta S_{\text{bp}}[N_{\text{A•T}} T_{\text{A•T}} + N_{\text{G•C}} T_{\text{G•C}}] + \sum_{i=1}^{8} N_i \delta G_i \tag{4.5}$$

Where $N_{\text{A•T}}$ and $N_{\text{G•C}}$ are the numbers of A•T or G•C type base pairs in the duplex sequence.

The average melting temperatures of an A•T or G•C base pair are given by $T_{A•T}$ or $T_{G•C}$. These are readily calculated as a function [Na$^+$] using the following equations derived from Frank-Kamenetskii's relationships (Frank-Kamenetskii, 1971),

$$T_{\mathrm{A\bullet T}} = 355.55 + 7.95 \cdot \ln[Na^{+}] \quad \mathrm{K} \tag{4.6}$$

$$T_{\mathrm{G\bullet C}} = 391.55 + 4.89 \cdot \ln[Na^{+}] \quad \mathrm{K} \tag{4.7}$$

The summed term on the right in equation (4.5) includes the explicit n-n sequence dependence. N_i is the number of times linear combination i (i = 1-8) occurs in the duplex sequence, and δG_i is the linear combination of deviations in the free energy from average n-n stacking free energy.

The entropy change of base pair formation, ΔS_{bp}, is assumed to be independent of sequence and Na$^+$ concentration over the range from 0.02 to 1.0 M Na$^+$, with an average value of ΔS_{bp} = -24.85 cal.mol^{-1}.K^{-1} (see Chapter 3.3.2). Thus, the total transition entropy of the duplex is simply,

$$\Delta S_{duplex} = \Delta S_{\mathrm{bp}}[N_{\mathrm{A\bullet T}} + N_{\mathrm{G\bullet C}}] \tag{4.8}$$

The duplex melting transition free energy is given by, $\Delta G_{duplex}(T) = \Delta H_{duplex} - T\Delta S_{duplex}$, which when combined with equations (4.5) and (4.8) gives,

$$\Delta G_{duplex}(T) = \Delta S_{\mathrm{bp}}[N_{\mathrm{A\bullet T}}(T_{\mathrm{A\bullet T}} - T) + N_{\mathrm{G\bullet C}}(T_{\mathrm{G\bullet C}} - T)] + \sum_{i=1}^{8} N_i \delta G_i \tag{4.9}$$

For example, consider melting the 10-mer duplex sequence 5'-A-T-T-A-T-G-G-G-G-C-3' in 115 mM Na^+. We wish to calculate the free energy of melting, ΔG_{duplex}, at 25 °C. From equations (4.6) and (4.7) in 0.115 M Na^+, $T_{A\bullet T}$ = 338.36 K (65.21 °C) and $T_{G\bullet C}$ = 380.97 K (107.82 °C). There are five A•T and five G•C base pairs, $N_{A\bullet T} = N_{G\bullet C} = 5$. The first term on the right of expression (4.9) includes the H-bond free energies and is evaluated as,

$$\begin{aligned}\Delta S_{\mathrm{bp}}[N_{\mathrm{A\bullet T}}(T_{\mathrm{A\bullet T}}-T)+N_{\mathrm{G\bullet C}}(T_{\mathrm{G\bullet C}}-T)] &= -24.85\cdot[5\cdot(338.36-298.15)+5\cdot(380.97-298.15)] \\ &= -15{,}286 \quad \mathrm{cal.mol}^{-1}\end{aligned}$$

To determine the n-n sequence dependent contributions to the stability, we must assess the numbers of the eight linear combinations (see Table V), that are present in the given duplex sequence. For the 10-mer sequence given, the numbers of these combinations are as follows (Goldstein and Benight, 1992; Doktycz, 1992):

$N_1 = N_{AA/TT} = 1$

$N_2 = N_{CC/GG} = 3$

$N_3 = N_{AT/AT} + N_{TA/TA} = 2 + 1 = 3$

$N_4 = N_{CG/CG} + N_{GC/GC} = 0 + 1 = 1$

$N_5 = N_{AC/GT} + N_{CA/TG} = 0 + 1 = 1$

$N_6 = N_{AG/CT} + N_{GA/TC} = 0 + 0 = 0$

$N_7 = N_{AT/AT} - N_{TA/TA} + N_{CG/CG} - N_{GC/GC} + 2(N_{GA/TC} - N_{AG/CT}) = 2\text{-}1 + 0 - 1 + 2(0\text{-}0) = 0$

$N_8 = N_{AT/AT} - N_{TA/TA} - N_{CG/CG} + N_{GC/GC} +2(N_{CA/TG} - N_{AC/GT}) = 2 - 1 - 0 + 1 + 2(1\text{-}0) = 4$

With these resident combinations and their corresponding values in Table V, the second term in equation (4.9) is,

$$\begin{aligned}\sum_{i=1}^{8} N_i \delta G_i &= 1\cdot\delta G_1 + 3\cdot\delta G_2 + 3\cdot\delta G_3 + 1\cdot\delta G_4 + 1\cdot\delta G_5 + 0\cdot\delta G_6 + 0\cdot\delta G_7 + 4\cdot\delta G_8 \\ &= -190 + 3(146) + 3(-28) - 240 - 113 + 4(-39) \\ &= -345 \quad \text{cal.mol}^{-1}\end{aligned}$$

Thus, from the n-n set in column 1 of Table V, the predicted value of ΔG_{duplex}(25°C) = -15,286 - 345 = -15,631 cal/mol. The calculated free energy for this sequence can be determined in an analogous manner using the other n-n sets in columns 2 thru 8 of Table V.

4.5.2. Calculation of Duplex Stability in the Doublet Format

In this format, the contributions of H-bonding and n-n stacking interactions are combined in a single parameter for each type of n-n base pair doublet. Values of the enthalpies and entropies for the 10 internal n-n doublets were evaluated by Breslauer et al. (1986), SantaLucia et al. (1996), Allawi and SantaLucia (1997) and Sugimoto et al. (1996) and reported in the doublet format. As stated earlier, in their evaluations these authors assumed that n-n interactions with the ends were either zero or a sequence independent constant. The n-n set reported by Breslauer et al. was determined from melting studies of short DNA oligomers and repeating DNA copolymers. The sets reported by SantaLucia et al., Allawi and SantaLucia, and Sugimoto et al. were determined from melting analysis of short DNA oligomers. The reported n-n doublet

TABLE VI Published Values of the Nearest-Neighbor Sequence Dependent Thermodynamic Parameters in the Doublet Format. Enthalpies (ΔH_{ij}) are in kcal.mol^{-1} and entropies (ΔS_{ij}) are in cal.mol^{-1}.K^{-1}.

Sequence (ij)	Breslauer et al.		SantaLucia et al.		Allawi and SantaLucia		Sugimoto et al.	
	ΔH_{ij}	ΔS_{ij}	ΔH_{ij}	ΔS_{ij}	ΔH_{ij}	ΔS_{ij}	ΔH_{ij}	ΔS_{ij}
AA/TT	-9.1	-24.0	-8.4	-23.6	-7.9	-22.2	-8.0	-21.9
AG/CT	-7.8	-20.8	-6.1	-16.1	-7.8	-21.0	-6.6	-16.4
AT/AT	-8.6	-23.9	-6.5	-18.8	-7.2	-20.4	-5.6	-15.2
AC/GT	-6.5	-17.3	-8.6	-23.0	-8.4	-22.4	-9.4	-25.5
GA/TC	-5.6	-13.5	-7.7	-20.3	-8.2	-22.2	-8.8	-23.5
GG/CC	-11.0	-26.6	-6.7	-15.6	-8.0	-19.9	-10.9	-28.4
GC/GC	-11.1	-26.7	-11.1	-28.4	-9.8	-24.4	-10.5	-26.4
TA/TA	-6.0	-16.9	-6.3	-18.5	-7.2	-21.3	-6.6	-18.4
TG/CA	-5.8	-12.9	-7.4	-19.3	-8.5	-22.7	-8.2	-21.0
CG/CG	-11.9	-27.8	-10.1	-25.5	-10.6	-27.2	-11.8	-29.0
Initiation if at least one G•C	0.0	-16.8	0.0	-5.9	—	—	0.6	-9.0
Initiation if only A•T	0.0	-20.1	0.0	-9.0	—	—	0.6	-9.0
Symmetry correction	0.0	-1.3	0.0	-1.4	0.0	-1.4	0.0	-1.4
5' T•A correction	—	—	0.4	0.0	—	—	—	—
A•T base pair on the end	—	—	—	—	2.3	4.1	—	—
G•C base pair on the end	—	—	—	—	0.1	-2.8	—	—

enthalpies, ΔH_{ij} and entropies, ΔS_{ij}, where ij represents one of the ten unique n-n doublet sequences, are listed in Table VI for each n-n set. Utilizing these values the calculated duplex melting enthalpy is just the sum of the individual enthalpies of the constituent n-n doublets of the duplex. That is,

$$\Delta H_{duplex} = \sum_{i,j=\mathrm{A,T,C,G}} N_{ij}\Delta H_{ij} \tag{4.10}$$

Where N_{ij} is the number of times the particular n-n doublet, ij (i,j = A,T,G,C) appears in the duplex. The sum is carried out over 10 internal n-n doublets. The duplex melting entropy is determined in an analogous manner:

$$\Delta S_{duplex} = \sum_{i,j=\mathrm{A,T,C,G}} N_{ij}\Delta S_{ij} \tag{4.11}$$

Finally, the duplex melting transition free energy is given by,

$$\Delta G_{duplex}(T) = \sum_{i,j=\mathrm{A,T,C,G}} N_{ij}(\Delta H_{ij} - T\Delta S_{ij}) \tag{4.12}$$

For example, consider the 10-mer sequence: 5'-A-T-T-A-T-G-G-G-G-C-3' and calculate the duplex melting free energy at 25 °C. The numbers of different types of n-n doublets that occur in the sequence are:

$N_{AA/TT} = 1$	$N_{AG/CT} = 0$	$N_{AT/AT} = 2$	$N_{AC/GT} = 0$	$N_{GA/TC} = 0$
$N_{GG/CC} = 3$	$N_{GC/GC} = 1$	$N_{TA/TA} = 1$	$N_{TG/CA} = 1$	$N_{CG/CG} = 0$

With these values, and the n-n parameters reported by Breslauer et al. (column 1 of Table VI) in equation (4.12), $\Delta G_{\text{duplex}}(25\,^\circ\text{C})$ is found accordingly,

$$\begin{aligned}
\Delta G_{duplex}(T) &= \sum_{i,j} N_{ij}(\Delta H_{ij} - T\Delta S_{ij}) \\
&= 1\cdot(\Delta H_{\text{AA/TT}} - T\Delta S_{\text{AA/TT}}) + 2\cdot(\Delta H_{\text{AT/AT}} - T\Delta S_{\text{AT/AT}}) + \\
&\quad 3\cdot(\Delta H_{\text{GG/CC}} - T\Delta S_{\text{GG/CC}}) + 1\cdot(\Delta H_{\text{GC/GC}} - T\Delta S_{\text{GC/GC}}) + \\
&\quad 1\cdot(\Delta H_{\text{TA/TA}} - T\Delta S_{\text{TA/TA}}) + 1\cdot(\Delta H_{\text{TG/CA}} - T\Delta S_{\text{TG/CA}}) \\
&= (-9{,}100 - 298.15\cdot(-24.0)) + 2\cdot(-8{,}600 - 298.15\cdot(-23.9)) + \\
&\quad + 3\cdot(-11{,}000 - 298.15\cdot(-26.6)) + (-11{,}100 - 298.15\cdot(-26.7)) + \\
&\quad + (-6{,}000 - 298.15\cdot(-16.9)) + (-5{,}800 - 298.15\cdot(-12.9)) \\
&= -1944 - 2948 - 9208 - 3139 - 961 - 1954 = -20{,}154 \quad \text{cal.mol}^{-1}
\end{aligned}$$

Compared to the value predicted using the singlet format and the n-n parameters evaluated from DNA dumbbells (Doktycz et al., 1992), the Breslauer et al. n-n set predicts the specific sequence to be more stable by nearly −5 kcal.mol^{-1}. Clearly, an apparent discrepancy exists between the two sets. Certainly, some of this discrepancy arises from the higher [Na$^+$] (1.0 M Na$^+$) of the Breslauer et al. n-n set where DNA is more stable and would be expected to have a lower free energy. Additional details of the predictive accuracies of these n-n sets and others are presented later.

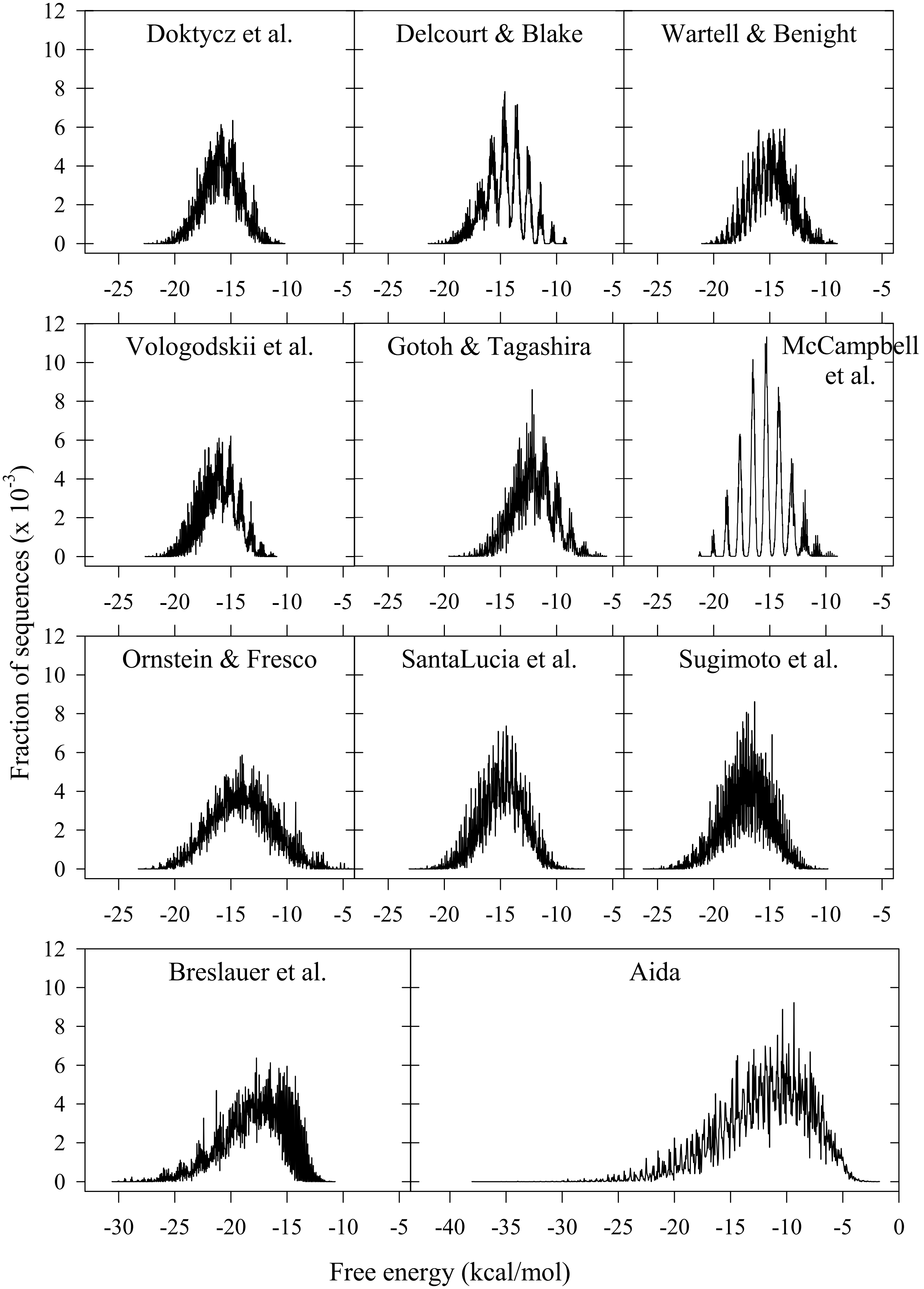
It should be mentioned at this point that in order to calculate the total stability of short DNA oligomers, that can then be compared directly to experimentally measured values, it is necessary to consider in addition to the n-n sequence dependent duplex free energy given above, the free energy of helix nucleation. The nucleation (initiation) free energy accounts for energetically unfavorable interactions between complementary single strands that must be overcome to allow duplex formation. Effects of interactions at the ends of linear duplexes and surrounding solvent are also included in the nucleation free energy. Different values and formats for nucleation parameters have been reported and employed by authors of the different n-n sets. These are summarized in the bottom rows of Table VI. As described below, in the relative statistical comparison of calculated stabilities of 10-mers that was performed, the nucleation free energies were not included in the calculated values, $\Delta G(25°C) = \Delta G_{duplex}(25°C)$. In addition, the entropy of symmetry correction for self-complementary duplexes (-1.4 cal.mol^{-1}.K^{-1}) is smaller than errors often found for other parameters and was omitted from the comparative free energy calculations.

4.6 Energy Distributions of 10-mers Predicted From 11 Nearest-Neighbor Sets

Calculated free energy distributions determined at 25 °C, $\Delta G_{duplex}(25°C)$, for all possible 10-mer duplex sequences are shown in Figure 4. These energy distributions were calculated using the 11 sets of published n-n stability parameters given in Tables V and VI. Distributions shown on Figure 4 were constructed by dividing the range between minimum and maximum $\Delta G_{duplex}(25°C)$ values into $\sqrt{524{,}800} = 724$ intervals. The fractions of sequences with calculated $\Delta G_{duplex}(25°C)$ that fall within each interval are plotted vs calculated $\Delta G_{duplex}(25°C)$.

Comparison of the distributions in Figure 4 reveals a number of interesting features of the different n-n sets. For example, the n-n sets reported by Delcourt and Blake (1991) and McCampbell et al. (1989) display polymodal distributions with nearly discrete peaks. These peaks correspond to values of $\Delta G_{duplex}(25°C)$ for the 10-mers obtained considering only the

Figure 4. Calculated free energy distributions for all DNA duplexes that are 10 base pairs long. Fraction of sequences is plotted vs. calculated free energy at 25 °C, ΔG_{duplex}(25°C). These calculations considered only the sequence dependent base pair energetics and did not include the nucleation free energy. The names of n–n sets are at the tops of graphs. All graphs have exactly the same scale on both axis for easy comparison.


Doktycz et al.
Delcourt & Blake
Wartell & Benight
Vologodskii et al.
Gotoh & Tagashira
McCampbell et al.
Ornstein & Fresco
SantaLucia et al.
Sugimoto et al.
Breslauer et al.
Aida
Fraction of sequences (x 10^{-3})
Free energy (kcal/mol)
12
10
8
6
4
2
0
-30
-25
-20
-15
-10
-5
-40
-35
0

hydrogen bonding component of base pair stability. Spikes occur at each increment of increasing %G•C base pairs in the sequences. Because distinct peaks are observed for these n-n sets, the majority of base pair stability arises from the average hydrogen bonding free energy (determined by %G•C), and deviations from the hydrogen bonding free energy due to n-n sequence dependent stacking interactions are relatively small. Distributions from other n-n sets such as those reported by Breslauer et al. (1986) and the theoretical set of Aida (1988) are non-symmetric and skewed to higher ΔG_{duplex}(25°C), revealing for these n-n sets that n-n interactions tend to stabilize a substantial number of the sequences. This stabilization occurs because smaller (more negative, stabilizing) values of free energies of n-n doublets having a G•C base pair were reported by Breslauer et al. and Aida. Also, the impact of including melting data for polymeric duplexes in evaluation of the Breslauer et al. n-n set may contribute to these observations.

4.7 Method of the Statistical Comparisons

Predictions from 11 different n-n sets were compared using a relative statistical approach. The eight sets in Table V and the sets of Breslauer et al. (1996), SantaLucia et al. (1996), and Sugimoto et al. (1996) listed in Table VI were examined. The n-n set of Allawi and SantaLucia (1997) was not used in this comparison because it was derived from the data reported by SantaLucia et al. (1996) and Sugimoto et al. (1996). Using each n-n set, the free energy, ΔG_{duplex}(25 °C), for each possible 10 base pair duplex sequence was calculated, i.e., the end parameters (including initiation parameters) were not used. Different n-n sets were then compared through comparison of the calculated ΔG_{duplex}(25 °C) values for all 10-mers. There are

4^{10} different single strands that are 10 bases long. A small fraction of these sequences is self-complementary in that two strands of the same sequence can form a duplex. Sequences were considered self-complementary when the first five bases from the 5' end were complementary to the first five bases from the 3' end. Therefore, the first 5 bases determine the sequence of the entire self-complementary 10-mer. Consequently, the number of such self-complementary 10-mers is 4^5. The remaining $4^{10} - 4^5$ non-self-complementary single strands were presumed to combine with their complement strand to form $\frac{4^{10}-4^5}{2}$ duplexes, where each duplex is comprised of two <u>different</u> single-strands. Consequently, the total number of unique 10 base pair DNA duplex sequences is $\frac{4^{10}-4^5}{2} + 4^5 = 524{,}800$. Free energies of these 10-mer sequences were calculated at 25 °C using each n-n set. Comparison of calculated values for all sequences obtained using two different n-n sets identified particular sequences whose calculated stabilities were different for the two n-n sets.

Following were the steps in the comparison process:

(1) The $\Delta G_{\text{duplex}}(25\,°\text{C})$ values for the 524,800 possible unique 10 base pair duplex sequences were calculated using each of the 11 n-n sets.

(2) For each n-n set, differences in $\Delta G_{\text{duplex}}(25\,°\text{C})$, for every possible pair of 10-mer sequences were determined. For the 524,800 possible 10 base pair duplexes, there are $\frac{524{,}800 \cdot (524{,}800 - 1)}{2} = 1.377072576 \times 10^{11}$ such pairs of sequences (i,j).

(3) Within each n-n set, these pairwise differences in free energy of duplex formation, $\Delta G_{\text{duplex}}(25\,°\text{C})$ were compared. For the rest of this section 4.7, ΔG_{duplex} (25 °C) is denoted ΔG. The free energy difference between sequence i and sequence j is defined as $\Delta\Delta G(i\text{-}j) = \Delta G(i) - \Delta G(j)$. If $\Delta\Delta G(i\text{-}j)$ is negative, duplex i is more stable than duplex j.

If the free energy difference is within the error of prediction, duplexes *i* and *j* have the same stability. Finally, if ΔΔG(*i-j*) is positive, duplex *i* is less stable than the duplex *j*. In other words, we make the following definitions. If ε is the relative error in the calculated free energy using n-n set X, then by definition, if ΔΔG(*i-j*) < 0 and $|\Delta\Delta G(i\text{-}j)| > \left| \varepsilon \cdot \frac{\Delta G(i) + \Delta G(j)}{2} \right|$, sequence *i* is *more stable* than sequence *j*. Likewise, if $|\Delta\Delta G(i\text{-}j)| \leq \left| \varepsilon \cdot \frac{\Delta G(i) + \Delta G(j)}{2} \right|$, sequence *i* has *stability equal* to sequence *j*, and the calculated stabilities of sequences *i* and *j* are equivalent within the error. If $\Delta\Delta G(i\text{-}j) > \left| \varepsilon \cdot \frac{\Delta G(i) + \Delta G(j)}{2} \right|$ sequence *i* is *less stable* than sequence *j*. For each published n-n set X (X = 1-11), ΔΔG(*i-j*) values for every possible pair of sequences were tabulated.

(4) For every possible combination of two n-n sets, calculated ΔΔG(*i-j*) values for every possible sequence pair were compared directly. Pairs of sequences for which the two n-n sets predicted the opposite order of stability were denoted, *discordant pairs*. That is, for a discordant pair of sequences *i* and *j*, either the first n-n set predicted that sequence *i* is more stable then sequence *j* and the second n-n set predicted that sequence *i* is less stable then sequence *j* or, the first n-n set predicted that sequence *i* is less stable then sequence *j* and the second n-n set predicted that sequence *i* is more stable then sequence *j*.

For each pair of n-n sets the number of *discordant pairs* was determined and tabulated. We chose to compare ΔΔG(*i-j*) values because within a given n-n set, these values are probably less sensitive to different reference duplex and single strand states that could occur in the different molecular and sodium ion environments. Free energies of duplex formation cannot be easily compared directly, because the different n-n sets were determined in different melting buffers having various ionic strengths. Thus, within a given n-n set relative differences were considered to be more suitable for direct comparisons with other n-n sets. In this way it was assumed that the relative stability of two 10-mers is independent of ionic strength.

Computer programs used to generate the sequences and to calculate $\Delta G_{duplex}(25°C)$ and the number of discordant pairs were written in ANSI FORTRAN 77, optimized for speed and thoroughly tested. They were run on Unix workstations (HP-UX 9000/735, HP-UX 900/720, IRIX Release 4.0.5 System V) and compiled with a HP-UX f77 or MIPS f77 compiler. One comparison of free energy predictions between two n-n sets took about 1000 (HP-UX) or 3000 (IRIX) CPU minutes.

4.8 Comparisons of Discordant Pairs

There are 137.7 x 10^9 different pairs of 10 base pair duplexes. In the n-n model a small number of these pairs cannot be discordant because they are comprised of precisely the same number of n-n base pair doublets, and therefore have the same predicted stability. Obviously, any pair of such sequences can never be discordant. The following sequences (5'-3') are a few examples, ACCGTCAAGA, AAGGACGTCA, AGGACGTTGA, AGTGAACGGA, ACCGTCTCAA, GAAGACACGG, GAGTTGACGG, GTGTTCGAGG, GGTTCGACAG, GGTGAACTCG. Neglecting the end interactions, the numbers of different types of n-n doublets that occur in each sequence are as follows,

$N_{AT/AT} = 0$	$N_{TA/TA} = 0$	$N_{AA/TT} = 1$	$N_{AC/GT} = 2$	$N_{CA/TG} = 1$
$N_{TC/GA} = 2$	$N_{CT/AG} = 1$	$N_{CG/CG} = 1$	$N_{GC/GC} = 0$	$N_{GG/CC} = 1$

Therefore, the predicted $\Delta G_{duplex}(25°C)$ for these sequences using any given n-n set is the same.

Because they all have the same predicted $\Delta G_{duplex}(25°C)$, no pair of them can ever be discordant. Of the 137.7 x 10^9 possible pairs of 10-mer sequences, only a small fraction of these (62,255,048 pairs) has exactly the same numbers and types of internal n-n doublets.

The numbers of discordant pairs (assuming an error, ε, in $\Delta\Delta G_{duplex}(25°C)$ of 5%), for all pairwise combinations of the 11 n-n sets in Tables V and VI, are given in Table VII. This comparison reveals the sets reported by Doktycz et al. (1992), Vologodskii et al. (1984), and Gotoh and Tagashira (1981) are identical within the assumed error. Relatively low numbers of discordant pairs are also found between these sets and those reported by Delcourt and Blake (1991), Wartell and Benight (1985) and McCampbell et al. (1989). Relatively larger numbers of discordant pairs are found for comparisons of the experimentally evaluated n-n sets with the theoretically calculated n-n sets of Ornstein and Fresco (1983) and Aida (1988). Also, relatively larger numbers of discordant pairs are found between the n-n sets reported by Breslauer et al. (1986), SantaLucia et al. (1996) and Sugimoto et al. (1996). For example, at the assumed 5% relative error, ε, comparison between the n-n set of Doktycz et al. (1992) and that of Breslauer et al. (1986), reveals 3.7 x 10^9 discordant pairs. There are 2.2 x 10^8 discordant pairs between the sets of Doktycz et al. (1992) and SantaLucia et al. (1996). In absolute terms this is a significant number of sequences. However, in relative terms, of the 1.377 x 10^{11} possible pairs, 3.7 x 10^9 discordant pairs is only 2.7% of the total, and 2.2 x 10^8 discordant pairs is only 0.2% of the total possible. Considering the differences in melting data, the duplex length distribution, and salt concentrations used to derive the n-n sets, the level of agreement is striking. Results of this global relative comparison also indicate there is quite a good agreement between the n-n sets reported by SantaLucia et al. (1996) and Sugimoto et al. (1996).

TABLE VII The Number of Discordant Pairs for Pairwise Comparison of Eleven Nearest-Neighbor Sets. A relative error of $\Delta\Delta G_{duplex}$(25°C) of 5% was assumed.

Set	Doktycz et al.	Delcourt and Blake	Wartell and Benight	Vologodskii et al.	Gotoh and Tagashira	McCampbell et al.	Ornstein and Fresco	Aida	Breslauer et al.	SantaLucia et al.	Sugimoto et al.
Doktycz et al.		$1.7\mathrm{x}10^{5}$	$6.9\mathrm{x}10^{5}$	0	0	$5.8\mathrm{x}10^{6}$	$8.3\mathrm{x}10^{8}$	$1.4\mathrm{x}10^{10}$	$3.7\mathrm{x}10^{9}$	$2.2\mathrm{x}10^{8}$	$1.1\mathrm{x}10^{8}$
Delcourt and Blake	$1.7\mathrm{x}10^{5}$		$4.5\mathrm{x}10^{7}$	0	$1.1\mathrm{x}10^{4}$	$6.1\mathrm{x}10^{4}$	$4.0\mathrm{x}10^{8}$	$1.9\mathrm{x}10^{10}$	$6.0\mathrm{x}10^{9}$	$9.6\mathrm{x}10^{8}$	$3.1\mathrm{x}10^{8}$
Wartell and Benight	$6.9\mathrm{x}10^{5}$	$4.5\mathrm{x}10^{7}$		$7.1\mathrm{x}10^{6}$	$1.3\mathrm{x}10^{8}$	$1.1\mathrm{x}10^{8}$	$2.3\mathrm{x}10^{8}$	$1.7\mathrm{x}10^{10}$	$6.0\mathrm{x}10^{9}$	$1.4\mathrm{x}10^{9}$	$1.3\mathrm{x}10^{9}$
Vologodskii et al.	0	0	$7.1\mathrm{x}10^{6}$		0	$7.5\mathrm{x}10^{5}$	$7.2\mathrm{x}10^{8}$	$1.3\mathrm{x}10^{10}$	$3.6\mathrm{x}10^{9}$	$1.8\mathrm{x}10^{8}$	$5.2\mathrm{x}10^{7}$
Gotoh and Tagashira	0	$1.1\mathrm{x}10^{4}$	$1.3\mathrm{x}10^{8}$	0		$3.7\mathrm{x}10^{7}$	$1.6\mathrm{x}10^{9}$	$1.9\mathrm{x}10^{10}$	$7.1\mathrm{x}10^{9}$	$7.2\mathrm{x}10^{8}$	$3.7\mathrm{x}10^{8}$
McCampbell et al.	$5.8\mathrm{x}10^{6}$	$6.1\mathrm{x}10^{4}$	$1.1\mathrm{x}10^{8}$	$7.5\mathrm{x}10^{5}$	$3.7\mathrm{x}10^{7}$		$8.4\mathrm{x}10^{8}$	$2.5\mathrm{x}10^{10}$	$5.9\mathrm{x}10^{9}$	$1.4\mathrm{x}10^{9}$	$4.3\mathrm{x}10^{8}$
Ornstein and Fresco	$8.3\mathrm{x}10^{8}$	$4.0\mathrm{x}10^{8}$	$2.3\mathrm{x}10^{8}$	$7.2\mathrm{x}10^{8}$	$1.6\mathrm{x}10^{9}$	$8.4\mathrm{x}10^{8}$		$2.5\mathrm{x}10^{10}$	$1.2\mathrm{x}10^{10}$	$4.3\mathrm{x}10^{9}$	$3.3\mathrm{x}10^{9}$
Aida	$1.4\mathrm{x}10^{10}$	$1.9\mathrm{x}10^{10}$	$1.7\mathrm{x}10^{10}$	$1.3\mathrm{x}10^{10}$	$1.9\mathrm{x}10^{10}$	$2.5\mathrm{x}10^{10}$	$2.5\mathrm{x}10^{10}$		$2.6\mathrm{x}10^{10}$	$1.6\mathrm{x}10^{10}$	$1.8\mathrm{x}10^{10}$
Breslauer et al.	$3.7\mathrm{x}10^{9}$	$6.0\mathrm{x}10^{9}$	$6.0\mathrm{x}10^{9}$	$3.6\mathrm{x}10^{9}$	$7.1\mathrm{x}10^{9}$	$5.9\mathrm{x}10^{9}$	$1.2\mathrm{x}10^{10}$	$2.6\mathrm{x}10^{10}$		$2.2\mathrm{x}10^{9}$	$1.8\mathrm{x}10^{9}$
SantaLucia et al.	$2.2\mathrm{x}10^{8}$	$9.6\mathrm{x}10^{8}$	$1.4\mathrm{x}10^{9}$	$1.8\mathrm{x}10^{8}$	$7.2\mathrm{x}10^{8}$	$1.4\mathrm{x}10^{9}$	$4.3\mathrm{x}10^{9}$	$1.6\mathrm{x}10^{10}$	$2.2\mathrm{x}10^{9}$		$1.5\mathrm{x}10^{5}$
Sugimoto et al.	$1.1\mathrm{x}10^{8}$	$3.1\mathrm{x}10^{8}$	$1.3\mathrm{x}10^{9}$	$5.2\mathrm{x}10^{7}$	$3.7\mathrm{x}10^{8}$	$4.3\mathrm{x}10^{8}$	$3.3\mathrm{x}10^{9}$	$1.8\mathrm{x}10^{10}$	$1.8\mathrm{x}10^{9}$	$1.5\mathrm{x}10^{5}$	

TABLE VIII Sequences of Eight DNA Duplexes in This Study.

Duplex	Top Strand
1	5'-GTAGTAGTAG-3'
2	5'-TGAAAAAAAA-3'
3	5'-TTAATAGGGG-3'
4	5'-ACAGTGACAC-3'
5	5'-GACGTGTGAC-3'
6	5'-ATTATGGGGC-3'
7	5'-AAAAAAAGTT-3'
8	5'-CATATATATG-3'

TABLE IX Thermodynamic Values of Eight DNA Duplexes Determined Experimentally in 115 mM and 1 M Na^+. The total transition enthalpy, ΔH_D, is in $kcal.mol^{-1}$, entropy, ΔS_D, is in $cal.mol^{-1}.K^{-1}$ and the changes of free energies, $\Delta G_{25°C}$, $\Delta\Delta G_{25°C}$, are in $kcal.mol^{-1}$.

Duplex	115 mM Na^+				1 M Na^+			
	ΔH_D	ΔS_D	$\Delta G_{25°C}$	$\Delta\Delta G_{25°C}$	ΔH_D	ΔS_D	$\Delta G_{25°C}$	$\Delta\Delta G_{25°C}$
1	-86.6 ± 5	-258 ± 18	-9.6 ± 0.1	-2.7 ± 0.1	-73.9 ± 3	-211 ± 8	-11.1 ± 0.1	-1.7 ± 0.1
2	-72.4 ± 4	-220 ± 14	-6.9 ± 0.1		-62.2 ± 2	-177 ± 8	-9.3 ± 0.1	
3	-67.6 ± 3	-196 ± 10	-9.1 ± 0.1	2.7 ± 0.4	-64.9 ± 2	-182 ± 5	-10.5 ± 0.1	3.4 ± 0.6
4	-70.2 ± 7	-196 ± 23	-11.8 ± 0.4		-79.5 ± 8	-220 ± 24	-14.0 ± 0.6	
5	-82.9 ± 2	-234 ± 6	-13.0 ± 0.1	-1.3 ± 0.6	-85.6 ± 4	-237 ± 13	-14.9 ± 0.3	-3.0 ± 0.6
6	-89.3 ± 15	-260 ± 48	-11.7 ± 0.6		-62.5 ± 8	-170 ± 25	-11.9 ± 0.5	
7	-68.7 ± 7	-205 ± 24	-7.7 ± 0.1	-1.3 ± 0.2	-63.3 ± 2	-180 ± 8	-9.8 ± 0.1	-1.1 ± 0.3
8	-81.6 ± 16	-252 ± 55	-6.3 ± 0.2		-83.9 ± 13	-252 ± 42	-8.7 ± 0.3	

TABLE X Comparison of Predicted and Experimental Melting Free Energies for Eight DNA Duplexes.

		$\Delta\Delta G_{25^\circ C}(1\text{-}2)$ (kcal/mol)		$\Delta\Delta G_{25^\circ C}(3\text{-}4)$ (kcal/mol)		$\Delta\Delta G_{25^\circ C}(5\text{-}6)$ (kcal/mol)		$\Delta\Delta G_{25^\circ C}(7\text{-}8)$ (kcal/mol)	
		0.115 M Na^+	1.0 M Na^+	0.115 M Na^+	1.0 M Na^+	0.115 M Na^+	1.0 M Na^+	0.115 M Na^+	1.0 M Na^+
Experiment		-2.7 ± 0.1	-1.7 ± 0.1	2.7 ± 0.4	3.4 ± 0.6	-1.3 ± 0.6	-3.0 ± 0.6	-1.3 ± 0.2	-1.1 ± 0.3
n-n set									
Doktycz et al.	(115 mM)	-1.5	—	1.9	—	-1.6	—	0.4	—
Delcourt and Blake	(75 mM)	-3.4	—	1.8	—	-1.9	—	0.7	—
Wartell and Benight	(100 mM)	-2.1	—	2.2	—	-1.1	—	0.9	—
Vologodskii et al.	(200 mM)	-2.0	—	1.7	—	-1.6	—	0.5	—
Gotoh and Tagashira	(19.5 mM)	-2.4	—	2.5	—	-2.6	—	0.0	—
McCampbell et al.	(102 mM)	-2.4	—	1.2	—	-1.4	—	-0.2	—
Ornstein and Fresco	(50 mM)	-6.0	—	3.4	—	-2.1	—	-0.2	—
Aida	(50 mM)	-2.6	—	10.6	—	-7.7	—	3.8	—
Breslauer et al.	(1.0 M)	—	5.4	—	-3.7	—	4.1	—	-3.9
SantaLucia et al.	(1.0 M)	—	1.4	—	2.2	—	-1.7	—	-3.4
Sugimoto et al.	(1.0 M)	—	0.2	—	1.3	—	-1.2	—	-2.3

4.9 Comparison of Predictions and Experiments for Four Pairs of 10-mer Duplexes

To experimentally verify the most discordant pairs, eight linear duplexes 10 base pairs long were prepared. Their melting transitions were measured as a function of DNA concentration. Sequences of the duplexes are shown in Table VIII and their thermodynamic values are listed in Table IX. Note, in Table IX the duplexes are grouped 1-2, 3-4, 5-6 and 7-8. These pairs of duplexes were selected because their predicted $\Delta\Delta G_{25^\circ C}$ were some of the most discordant found when n-n sets determined in 75-115 mM Na^+ (Doktycz et al., 1992; Delcourt and Blake, 1991) were compared to n-n sets determined in 1 M Na^+ (Breslauer et al., 1986; SantaLucia et al., 1996; Sugimoto et al., 1996). Nearest-neighbor sets determined in low (~100 mM) Na^+ environments predict the opposite order of stability for each pair of duplexes compared to predictions based on the n-n sets determined in 1 M Na^+. This discrepancy in predictions of $\Delta\Delta G_{25^\circ C}$ for duplexes poses two questions. (1) Does the order of stability of the examined duplexes change with increasing [Na^+]? (2) Which n–n set actually predicts the observed order of stability for the pairs of duplexes? To answer these questions, melting experiments were performed on these molecules and results were compared with predictions. Their thermodynamic values were measured by UV spectroscopy as described in Chapter 2.7. Plots of $\frac{1}{T_m}$ *vs* $\ln\left[\frac{C_T}{4}\right]$ in both 115 mM and 1.0 M Na^+ environments were linear for all duplexes (not shown). Assuming a two-state melting transition, the transition enthalpy, ΔH_D, entropy, ΔS_D and free energy at 25 °C, $\Delta G_{25^\circ C}$, were calculated from the fitted slopes and intercepts of linear plots utilizing equation (3.37) (see Chapter 3.4 for details). Values for these thermodynamic parameters in the two solvent environments are summarized for each of the eight duplexes in Table IX. The standard deviations of ΔH_D, ΔS_D and $\Delta G_{25^\circ C}$ given in Table IX were estimated from errors in the fits using equations (3.38), (3.39) and (3.40). The experimentally determined differences of free energies, $\Delta\Delta G_{25^\circ C}$, for each pair, in both Na^+ environments are also listed in Table IX.

Predictions of the 11 n-n sets and the experimental results given in Table IX are compared in Table X. In the first column, the n-n set is listed and the Na^+ environment where the set was evaluated is given in parenthesis. The experimental values from Table IX are given in the first row. For the four discordant pairs of duplexes examined, experiment showed that the order of duplex stability is the same in 115 mM and 1 M Na^+ and does not change in these two different ionic strength environments. For each pair of duplexes, the calculated difference at 25 °C, $\Delta\Delta G_{25^\circ C}(i\text{-}j) = \Delta G_{25^\circ C}(i) - \Delta G_{25^\circ C}(j)$ was predicted using the n-n sets. Comparisons in Table X reveal the n-n parameters reported in Table V seem to predict the correct order of stability. That is, the predicted $\Delta\Delta G_{25^\circ C}(i\text{-}j)$ has the same sign as experimentally observed for $\Delta\Delta G_{25^\circ C}(1\text{-}2)$, $\Delta\Delta G_{25^\circ C}(3\text{-}4)$, and $\Delta\Delta G_{25^\circ C}(5\text{-}6)$. For the last pair, $\Delta\Delta G_{25^\circ C}(7\text{-}8)$, seven of the eight n-n sets in Table V (the theoretical set of Aida (1988) being the exception) predict the two duplexes to have nearly equivalent stability. The n-n sets summarized in Table VI predict the correct order of stability for $\Delta\Delta G_{25^\circ C}(7\text{-}8)$. For $\Delta\Delta G_{25^\circ C}(1\text{-}2)$, the n-n sets summarized in Table VI predict the opposite order of stability to what is experimentally observed. For the most part the magnitudes of the predicted and measured free energy differences, although close in some cases (for $\Delta\Delta G_{25^\circ C}(5\text{-}6)$), are not within the standard deviations summarized in Table IX.

At first sight comparisons in Table X would suggest for the few molecules examined that none of the n-n sets is able to accurately predict all of the observed $\Delta\Delta G_{25^\circ C}(i\text{-}j)$ values within the experimental standard deviation. Initially, this may appear somewhat disheartening.

However, we should recall that the sequences that were examined were selected deliberately because they were found to be the most discordant in the pairwise comparisons of predictions from different n-n sets. Examination of the sequences in Table IX reveals they are either repetitive or have strings of A•T base pairs. Perhaps the discordance in these sequences is due to the anomalous melting behavior that occurs in them. If so, such sequence dependent features are not currently considered by the n-n model. Also, it is likely that these molecules may not melt in a purely two state manner, in which case the disagreement underscores the deficiencies and potential problems associated with the two-state approach to melting analysis. Further investigations of these types of sequences will be required to resolve these potential issues. For the purposes of the present study, the real question is, how well do the n-n sets actually predict (on average) the experimental melting temperature, T_m? Results that address this question follow below.

4.10 Calculated Stability of DNA Oligomers

Results of the statistical comparison summarized in Table VII suggest that there is only a minor disagreement between the calculated melting stability of all possible 10 base pair sequences using the different n-n sets. However, the comparison in Table IX indicates that for certain sequences the different n-n sets cannot predict experimental results within quantitative agreement. These comparisons are interesting and provide a new vantage point from which to compare the n-n sequence dependent interactions. However, what remains is to determine which n-n sets can in general provide the most quantitatively accurate predictions of experimentally measured T_m's of duplex DNA oligomers given their sequence and concentration. General expressions for T_m in terms of the respective thermodynamic quantities and total strand concentration can be derived from equations (3.35) and (3.45). Rearrangement of equation (3.35) yields a formula for *non-self-complementary* linear duplexes,

$$T_m = \frac{\Delta H_{duplex} + \Delta H_{nuc}}{\Delta S_{duplex} + \Delta S_{nuc} + R\ln\left[\frac{C_T}{4}\right]} \tag{4.13}$$

The analogous expression for *self-complementary* duplexes has $C_T/4$ replaced by C_T. This expression and the analogous expression for self-complementary duplexes were employed to calculate T_m's of duplex sequences as a function of strand concentration. Results of these calculations were compared directly to experimental data. Calculations required input values for ΔH_{duplex}, ΔH_{nuc}, ΔS_{duplex}, ΔS_{nuc} and C_T. As demonstrated earlier, it is straight forward to employ the n-n stability parameters to calculate ΔH_{duplex}, and ΔS_{duplex} from the base pair sequence. But the nucleation parameters have been reported in several different forms. If ΔH_{nuc} is assumed to be zero, then rearrangement of equation (3.34) yields nucleation entropy, $\Delta S_{nuc} = -\Delta G_{nuc}/T = -R\mathit{ln}\beta$. Recently, non-zero values of both ΔS_{nuc} and ΔH_{nuc} have been reported (Allawi and SantaLucia, 1997; Sugimoto, 1996). In our analysis several different forms of ΔS_{nuc} and $\Delta H_{nuc} \neq 0$ were assumed and tested for their ability to improve average predictions of T_m's of short duplex DNA oligomers.

For the comparisons that follow we restricted our attention to a subset of five n-n sets. These are the set of Doktycz et al. (1992), evaluated from analysis of melting curves of short DNA dumbbells (column 1 of Table V) and the four n-n sets evaluated from melting curves of linear duplex DNA oligomers. The examined sets were those reported by Breslauer et al. (1986), SantaLucia et al. (1996), Allawi and SantaLucia (1997) and Sugimoto et al. (1996). The Breslauer et al. n-n set was determined from experiments on not only short DNA duplexes, but also from optical and calorimetric melting analysis of long synthetic repeating DNA polymers. Values for each n-n sets are given in Table VI. These five sets were employed to predict the reported experimentally measured T_m's as a function of concentration for the 251 DNA oligomers in database B (see page 55 for description of database). Oligomer lengths ranged from 4 to 16 base pairs and they were melted in 1.0 M NaCl. For these calculations the melting temperatures predicted in 115 mM Na^+ (Doktycz et al. n-n set), were scaled to 1.0 M Na^+ using the published correction (Owczarzy et al., 1997; Vallone, 1999),

$$T_m(115\text{ mM Na}^+) - T_m(1\text{ M Na}^+) = (8.91 \cdot f(G{\bullet}C) - 11.34) \quad \text{K} \qquad (4.15)$$

The other n-n sets were used directly because they predict melting temperatures in 1 M Na^+. For every duplex, a two-state melting process was assumed and T_m was calculated according to equation (4.13) and the analogous expression for self-complementary duplexes. The required sequence dependent melting enthalpies and entropies, ΔH_{duplex} and ΔS_{duplex}, were calculated directly from the base pair sequence using the n-n parameters in Tables V and VI and equations (4.5), (4.8), (4.10) and (4.11). Values of nucleation parameters, ΔH_{nuc} and ΔS_{nuc} were varied. At an arbitrary total strand concentration, C_T = 4 μM, the melting temperature of each of the 251 DNA oligomers in database B was calculated, $T_m(P)$, and compared directly with the published experimentally measured value, $T_m(M)$, that was either reported at that concentration or extrapolated to 4 μM using reported ΔH_D and ΔS_D values. The corresponding absolute value of the difference between predicted $T_m(P)$ and measured $T_m(M)$, $|\Delta T_m|$, was determined for each DNA sequence. Then the average $|\Delta T_m|$ over the entire 251 member set, $<|\Delta T_m|>_{ave}$, was determined. This procedure was repeated using each of the five n-n sets, and assuming the different forms and combinations of ΔH_{nuc} and ΔS_{nuc}. Details of these calculations and combinations were published (Owczarzy et al., 1997) and are reported in the Ph.D. thesis of Peter M. Vallone from our group (Vallone, 1999). A summary of the results is displayed in Table XI.

TABLE XI Predictions of Melting Temperatures for 251 DNA Duplexes. Average difference between predicted and experimentally measured melting temperatures, ΔT_m, is reported for several n-n sets.

Nearest Neighbor Set	$<|\Delta T_m|>_{ave}$ (K)
Doktycz et al.	19.7
Doktycz et al. including ΔH_{nuc} correction	1.6
Breslauer et al.	6.4
SantaLucia et al.	2.0
Allawi and SantaLucia	1.5
Sugimoto et al.	2.0

The predictions of $T_m(P)$ for the 251 DNAs were made using the published values for ΔH_{nuc} and ΔS_{nuc} nucleation parameters and their values for the five n-n sets examined were employed as follows, $\Delta G_{nuc}(37°C) = RT\mathit{ln}\beta = \Delta H_{nuc} - \Delta S_{nuc}\cdot(273.15 + 37)$. The n-n sets of Breslauer et al. (1986) and SantaLucia et al. (1996), assigned values of $\Delta G_{nuc}(37°C) = 6.23$ and 2.79 kcal/mol, respectively, to duplexes comprised exclusively of A•T type base pairs. Breslauer et al. (1986) and SantaLucia et al. (1996) also reported values of $\Delta G_{nuc}(37°C) = 5.21$ and 1.83 kcal/mol, respectively, for duplexes containing at least one G•C type base pair. As seen at the bottom of Table VI, these sets assume for the most part that $\Delta G_{nuc}(37°C)$ is entirely entropic, and $\Delta H_{nuc} = 0$. An exception is for the set of SantaLucia et al. (1996) that assumes a slight enthalpic penalty, $\Delta H_{nuc} = 0.4$ kcal/mol, for duplexes that contain a 5'-T•A-3' base pair at the end of duplex. For the n-n set of Sugimoto et al. (1996), a constant value of $\Delta G_{nuc}(37°C) = 3.39$ kcal/mol was assumed, comprised of an enthalpic component, $\Delta H_{nuc} = 0.6$ kcal/mol and entropic component, $-\Delta S_{nuc}\cdot(273.15 + 37)= 2.79$ kcal/mol. For the n-n set of Allawi and SantaLucia (1997), $\Delta G_{nuc}(37°C)$ depends on the identity of the base pair that terminates the duplex. For a terminal A•T base pair, $\Delta G_{nuc}(37°C) = 1.03$ kcal/mol. For a terminal G•C base pair $\Delta G_{nuc}(37°C) = 0.98$ kcal/mol. These nucleation free energies are comprised of different enthalpic and entropic contributions depending on the terminal base pair as summarized at the bottom of Table VI. For the above mentioned n-n sets (Breslauer et al., 1986; SantaLucia et al., 1996; Allawi and SantaLucia, 1997) a symmetry correction, $\Delta G_{sym}(37°C) = 0.43$ kcal/mol was also included in calculations of self-complementary sequences. For the n-n set determined from melting analysis of DNA dumbbells (Doktycz et al., 1992), in the first case, it was assumed that $\Delta H_{nuc} = 0$, $\Delta S_{nuc} = 0$ and $\Delta G_{sym}(37°C) = 0$.

Resulting $<|\Delta T_m|>_{ave}$ values are summarized in the Table XI. Only the sets of Doktycz et al. (1992), and Breslauer et al. (1986) provide relatively poor predictions of the data. The other n-n sets provide good predictions of the data. We should mention of course that the different sets of molecules used to evaluate these n-n sets are a subset of the 251 molecules predicted. We next investigated whether improved predictions could be obtained by making adjustments in ΔS_{nuc}, ΔH_{nuc}. Predictions of the Doktycz et al. n-n set improve significantly when a sequence dependent, non-zero ΔH_{nuc} is assumed (Vallone, 1999). This observation is consistent with the molecular structure of the DNA samples. Because of the loops on both ends the energetic cost of nucleating duplexes in dumbbells is minimal. Therefore, it is perhaps not surprising that when used to predict melting behavior of linear duplex oligomers, the n-n set evaluated from dumbbells requires a significant correction to account for nucleation. The evaluated nucleation enthalpy that should be used in conjunction with the dumbbell n-n parameters is (Vallone, 1999; Owczarzy et al., 1997),

$$\Delta H_{nuc} = 7945.5 - 3413.0 \cdot f(G{\bullet}C) - 201 \cdot N_{\text{bp}} \quad \text{cal.mol}^{-1} \tag{4.16}$$

Thus, as the fraction of G•C base pairs, f(G•C), and the number of base pairs, N_{bp}, increases, the enthalpic cost of duplex nucleation decreases.

Results summarized in Table XI indicate that four of the six n-n sets are able to provide reasonably accurate predictions of the melting temperatures of 251 duplex DNA sequences (~2.0 °C). The exception is the set of Breslauer et al. (1986). There are several potential sources for this observation. It should be noted that this n-n set emerged after the initial pioneering work of Breslauer and Marky on calorimetric melting of DNA oligomers (Marky and Breslauer, 1982). This n-n set is at least six years older than the other n-n sets. In addition, it was evaluated directly by calorimetric measurements and derived from melting experiments of both short oligomers and relatively much longer synthetic copolymers. The apparent lower level of accuracy for this n-n set is most likely due to the relatively lower level of measurement accuracy available at the time it was evaluated.

4.11 Conclusion

Eleven n-n sets (Wartell and Benight, 1985; Gotoh and Tagashira, 1981; Vologodskii et al., 1984; McCampbell, 1989; Delcourt and Blake, 1991; Ornstein and Fresco, 1983; Aida, 1988; Doktycz et al., 1992; Breslauer et al., 1986; SantaLucia et al., 1996; Sugimoto et al., 1996) of published n-n stability parameters were compared through differences in their predicted stabilities of all possible 10 base pair duplexes. In general, reasonable agreement among n-n sets for predictions of free energies was observed. Five of the published stability parameter sets were employed to calculate the melting stability of 251 short linear duplexes melted in solvent containing 1.0 M Na^+.

The analysis reveals that employment of the n-n set evaluated from melting analysis of DNA dumbbells (Doktycz et al., 1992) to predict melting stability of linear duplex DNAs requires a substantial unfavorable correction for the enthalpy of nucleation, ΔH_{nuc}, that depends on the %GC and length of the duplex. When this correction is employed, the dumbbell n-n set is able to provide accurate predictions of T_m's of short DNA oligomers. Its predictive accuracy is comparable to accuracy of the several new and improved n-n sets (SantaLucia, 1996; Allawi and

SantaLucia, 1997; Sugimoto, 1996) evaluated from melting data of short DNAs directly.

To summarize, given the concentration and sequence, the T_m's of duplex DNA oligomers having 20 or less duplex base pairs, in solvent ionic strength ranging from 115 mM to 1.0 M Na^+, can be calculated (on average) to within 2 °C, provided the oligomers melt in a two-state manner. However, as we found for certain 10-mer sequences, agreement between the calculated sequence dependent melting thermodynamics and parameters evaluated directly from two-state model analysis of melting data, cannot always be obtained (using any n-n set). Whatever the source of these observations, i.e., peculiar melting behavior of the particular sequences examined and/or deviations of their melting transitions from a strictly two-state process, they serve to underscore shortcomings of the two-state model and raise the following concern. For any sequence, unless two-state melting is independently verified, uncertainties due to potential deviations from two-state behavior will always surround predictions of melting stability based on the two-state model.

5. EVALUATION OF NEXT-NEAREST-NEIGHBOR PARAMETERS

5.1 Introduction

Sequence dependent stability of duplex DNA has been quantitatively evaluated in terms of the nearest-neighbor (n-n) model. Investigations described in Chapter 4 reveal that the current n-n sets of parameters predict T_m's in 1 M Na^+ within 2 °C. Such an error is close to limitations of the n-n model. It was reported (Sugimoto et al., 1994) that different duplex sequences containing the same number of nearest-neighbors can have differences in T_m's up to 5.6 °C. For instance, the DNA duplexes TCTATAGA and TAGATCTA in 1 M Na^+ have melting temperatures of 27.3 and 32.9 °C, respectively, at the same total single strand concentration of C_T = 0.1 mM (Sugimoto et al., 1994). Thermodynamic properties of such duplexes are not distinguishable in the n-n model, because the calculation procedure of the n-n model is based solely on the numbers of nearest-neighbors. Thus, the n-n model predicts the same T_m for such duplexes and more accurate prediction cannot be obtained with the n-n model. Logical extension to possibly account for these observations would be to formulate a next-nearest-neighbor (n-n-n) model that also takes into account interactions in next-nearest-neighbor base pairs (triplets).

Several studies have suggested that DNA structure and structural deformations of duplex DNA are sequence dependent (Dickerson and Chiu, 1997). The sequence dependence of sodium and magnesium ion binding to DNA was also observed in high resolution (1.4 Å) structures of a DNA dodecamer (Shui et al., 1998). It was suggested that the structural deformations of duplex DNA are sequence dependent and may be a function of sequence dependent interactions of sodium ions. Thus, such sequence dependent responses to cation binding could have a structural basis and persist over long distances. Early melting studies conducted on a series of DNA hairpins formed from $d(TA)_N$ (N = 8 - 22) showed that electrostatic interactions in DNA duplexes are salt dependent (Elson, 1970). They demonstrated that electrostatic interactions that extend beyond nearest-neighbor base pairs must be taken into account to accurately describe stability of DNA duplexes in moderate to low salt environments (3 - 60 mM Na^+). The question is whether these interactions are strictly sequence dependent and over which distances they persist.

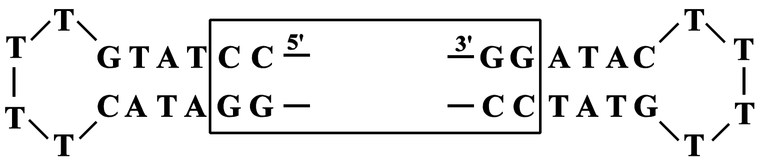
Several years ago, our laboratory published results of melting studies of DNA dumbbells conducted as a function of $[Na^+]$ in the range from 25 to 115 mM. From these results the nearest-neighbor (n-n) sequence dependent interactions in duplex DNA were evaluated. Furthermore these studies revealed that the nearest neighbor model could not accurately represent the melting behavior of the dumbbells in 25 and 55 mM Na^+. Inability of the n-n model to accurately fit melting data in 25 mM Na^+ suggested that interactions beyond nearest-neighbors could be significant in this relatively low sodium ion environment. However, at that time our database was not large enough, not enough sequences were represented to rigorously test this hypothesis that longer range sequence dependent interactions in low salt exist.

In this Chapter we report the expansion of the melting database on DNA dumbbells to include sequences of all the next-nearest-neighbor (n-n-n) sequence dependent interactions and evaluate these by analysis of melting curves in 25 mM Na^+. We employed optical melting analysis on DNA dumbbells to evaluate n-n-n sequence dependent interactions in DNA duplexes. The dumbbells have short duplex regions 14 to 20 base pairs long crosslinked on both ends by T_4 single strand loops. Dumbbell molecules offer several advantages to the study of sequence dependent effects in short DNA sequences. Because of the intramolecular structure of the dumbbells, melting of the duplex stem is independent of concentration (at the concentrations where melting experiments are performed). In addition, melting of dumbbell molecules does not seem to suffer from the anomalies (end effects, sequence dependent nucleation and melting entropy, etc.) that are generally associated with melting short linear duplexes oligomers. Without the concentration dependence and anomalies associated with melting of short, linear duplex DNA oligomers, we believe that dumbbells provide a more realistic mimic of short DNA sequences in a long sequence environment. Methods employed to exploit optical melting curve data in order to evaluate nearest-neighbor (n-n) interactions in DNA have been presented earlier (Doktycz et al., 1992; Goldstein and Benight, 1992). In this Chapter, we evaluate the singlet, doublet and triplet sequence dependent interactions in duplex DNA at total sodium ion concentrations of 25, 55, 85, 115 mM.

5.2 DNA Molecules

Twenty-two DNA dumbbell molecules were prepared for these studies. Procedures for dumbbell synthesis and purification are reported in Chapter 2. Data gathered on these molecules was combined with results from an earlier melting study of 17 similar dumbbells (Doktycz et al., 1992). Taken together results of melting experiments on 39 different DNA dumbbells comprise our optical melting database. Sequences of these molecules are shown in Figure 5. Inspection of Figure 5 readily reveals that the DNA dumbbells share several common features. Dumbbell end-loops consist of four thymidines. Lengths of the duplex stems of the dumbbells vary from

14 to 20 base pairs. Dumbbell stems contain the same sequence, 5'-G-T-A-T-C-C-3', on both ends, and have variable central sequences. For the 22 new dumbbells (dumbbells 18-39 on



1	-ATAT-
2	-TATA-
3	-TTAA-
4	-AAAA-
5	-ATAC-
6	-ATCA-
7	-ACAC-
8	-CACA-
9	-GAGA-
10	-CTCT-
11	-GCAC-
12	-CGGA-
13	-CCCC-
14	-GCGC-
15	-CGCG-
16	-AC-
17	-ACACAC-
18	-TAATTA-
19	-TTTAAA-
20	-AATATT-
21	-CTTAAG-
22	-TCTAGA-
23	-TTCAAG-
24	-AGTACT-
25	-TGATCA-
26	-AAGCTT-
27	-GAATTC-
28	-AAGGTTCC-
29	-GGTAAC-
30	-GTAC-
31	-GACT-
32	-GATC-
33	-CTCGAG-
34	-CCTGGT-
35	-GGATCC-
36	-GAGCCT-
37	-ACCGGT-
38	-GTCGAC-
39	-GCATGC-

Figure 5. Structure and central sequences of 39 DNA dumbbells prepared and used in evaluation of next-nearest-neighbor parameters. Dumbbells 1-17 were reported by Doktycz et al. (1992). Dumbbells 18-39 were prepared and measured in this study.

Figure 5) four to eight base pairs in the center are unique for each molecule. These unique central sequences were chosen to augment those sequences from our earlier studies (dumbbells 1-17 of Figure 5) so that all possible n-n-n (triplet) sequences were represented.

5.3 Sequence Specific Thermodynamic Parameters in DNA Duplexes

Methods used to determine nearest-neighbor parameters from UV-melting curves and modifications to account for the differences between linear DNA oligomers and DNA dumbbells have been presented in previous work (Doktycz et al., 1992) and in Chapters 3 and 4. Our approach to evaluating sequence specific interactions of increasing order in DNA, i.e. singlet, doublet (nearest-neighbor), triplet (next-nearest-neighbor), etc., involves considering first singlet, then doublets and then triplet interactions and applying procedures that have been described previously (Doktycz et al., 1992; Goldstein and Benight, 1992). In this method each level of interaction is evaluated sequentially. The method is developed in its entirety here starting from singlet (single base pair) expanding to doublet (nearest-neighbor base pair) sequence dependent interactions and further expansion to include triplet (next-nearest-neighbor base pair) sequence specific interactions. Pertinent thermodynamic parameters for each level of interaction are evaluated in a sequential manner as described below.

Considering singlet, doublet, and triplet interactions the total transition free energy change upon formation of dumbbell *p* having $N_{bp}(p)$ base pairs in the stem is written as,

$$\Delta G_D(p) \ = \ \Delta G_{loop} \ + \sum_{i=1}^{N_{\mathrm{bp}}(p)} \Delta G_i(p) \ + \sum_{i,j=\mathrm{A,T,C,G}} N_{ij}(p)\cdot\delta G_{ij}^{\,n-n} \ + \\ + \sum_{i,j,k=\mathrm{A,T,C,G}} N_{ijk}(p)\cdot\delta G_{ijk}^{\,n-n-n} \tag{5.1}$$

Descriptions of each of these terms and the processes involved in their evaluation follow.

The present treatment considers all of the n-n doublets in the duplex sequences. That is, if the dumbbell stem contains 16 base pairs, then 15 n-n doublets are considered. This usage of the sequence differs from what was done in the previous analysis in our laboratory (Doktycz et al., 1992). There, only the unique central sequence and adjoining base pairs were considered. For example, if the central sequence was four base pairs in length, five n-n were considered. The remaining doublets common to the end sequence of every dumbbell were assumed to behave identically and thus were weighted the same for every molecule. As will be seen, within the errors, either treatment yields comparable results revealing effects of dumbbell ends on the n-n sequence dependent stability are fairly localized.

5.4 Evaluation of Sequence Specific Singlet Interactions

The first order of interactions to be considered are the singlet interactions. These sequence dependent interactions are evaluated first. In terms of singlet interactions only, the total free energy of formation of dumbbell *p*, ΔG_D(*p, singlet*) is given by,

$$\Delta G_D(p, singlet) \ = \ \Delta G_{loop} \ + \sum_{i=1}^{N_{\mathrm{bp}}(p)} \Delta G_i(p) \tag{5.2}$$

The first term on the right, ΔG_{loop}, includes all energetic contributions from the end-loops to the free energy of dumbbell formation. For the molecules in our database whose melting data will be analyzed, the end-loops and six base pairs on both ends of the stem are the same. Therefore, we assumed the influence of these identical loops and end sequences on the melting is the same for all dumbbells. The second term on the right hand side of equation (5.2), $\sum_{i=1}^{N_{\text{bp}}(p)} \Delta G_i(p)$, is the free energy change in creating hydrogen-bonded base pairs in the duplex stem. This term includes contributions from the forming of hydrogen bonds and *average* n-n stacking interactions. Consequently, the total singlet free energy of the p^{th} DNA dumbbell at temperature T, depends only on the number of A•T or G•C type base pairs in the stem. That is,

$$\Delta G_{loop} + \sum_{i=1}^{N_{\text{bp}}(p)} \Delta G_i(p) = N_{\text{A•T}}(p) \cdot \Delta G_{\text{A•T}} + N_{\text{G•C}}(p) \cdot \Delta G_{\text{G•C}} \tag{5.3}$$

Where $N_{\text{A•T}}(p)$ and $N_{\text{G•C}}(p)$ are the numbers of A•T and G•C base pairs, respectively, in the duplex stem of dumbbell p and $\Delta G_{\text{A•T}}$ and $\Delta G_{\text{G•C}}$, are the average free energies of melting A•T or G•C type base pairs in the DNA dumbbells. Since the influence of ΔG_{loop} is assumed constant for all dumbbells and absorbed in the values of $\Delta G_{\text{A•T}}$ and $\Delta G_{\text{G•C}}$, these singlet interactions depend only on the base pair type. At temperature T they have the form,

$$\Delta G_{\text{G•C}}(T) = \Delta S_{\text{bp}}(T_{\text{G•C}} - T) \tag{5.4}$$

$$\Delta G_{\text{A•T}}(T) = \Delta S_{\text{bp}}(T_{\text{A•T}} - T) \tag{5.5}$$

Where $T_{A•T}$ and $T_{G•C}$ are the average melting temperatures of A•T and G•C base pairs in the dumbbell stems.

The entropy change in base pair formation in equations (5.4) and (5.5) is assumed to be independent of base pair type and constant at $\Delta S_{bp} = -24.85$ cal.mol^{-1}.K^{-1} (Doktycz et al., 1992; Delcourt and Blake, 1991). Validity of this assumption for melting base pairs in DNA dumbbells is provided later in Results. Invoking this assumption the singlet enthalpies can be estimated directly from the average melting temperatures and are given by,

$$\Delta H_{\mathrm{G•C}} = \Delta S_{\mathrm{bp}} T_{\mathrm{G•C}} \tag{5.6}$$

$$\Delta H_{\mathrm{A•T}} = \Delta S_{\mathrm{bp}} T_{\mathrm{A•T}} \tag{5.7}$$

With the definitions in equations (5.4) and (5.5) and a constant value for ΔS_{bp}, equation (5.3) can be written,

$$N_{\mathrm{A•T}}(p) \cdot \Delta G_{\mathrm{A•T}} + N_{\mathrm{G•C}}(p) \cdot \Delta G_{\mathrm{G•C}} = \Delta S_{\mathrm{bp}} \cdot N_{\mathrm{A•T}}(p) \cdot (T_{\mathrm{A•T}} - T) + \Delta S_{\mathrm{bp}} \cdot N_{\mathrm{G•C}}(p) \cdot (T_{\mathrm{G•C}} - T) \tag{5.8}$$

In a similar manner the singlet transition enthalpy of the p^{th} dumbbell is given by,

$$\Delta H_D(p, singlet) = \Delta H_{loop} + \sum_{i=1}^{N_{\mathrm{bp}}(p)} \Delta H_i(p) \tag{5.9}$$

with,

$$\begin{aligned} \Delta H_{loop} + \sum_{i=1}^{N_{\mathrm{bp}}(p)} \Delta H_i(p) &= N_{\mathrm{A{\bullet}T}}(p) \cdot \Delta H_{\mathrm{A{\bullet}T}} + N_{\mathrm{G{\bullet}C}}(p) \cdot \Delta H_{\mathrm{G{\bullet}C}} \\ &= \Delta S_{\mathrm{bp}} \cdot N_{\mathrm{A{\bullet}T}}(p) \cdot T_{\mathrm{A{\bullet}T}} + \Delta S_{\mathrm{bp}} \cdot N_{\mathrm{G{\bullet}C}}(p) \cdot T_{\mathrm{G{\bullet}C}} \end{aligned} \quad (5.10)$$

The values of $N_{A\bullet T}$ and $N_{G\bullet C}$ are determined from the length and sequence composition of the dumbbell stem and ΔS_{bp} is assumed constant. Values of $T_{A\bullet T}$ and $T_{G\bullet C}$ are determined by extrapolation from linear least-squares fits of plots of the melting temperatures of the dumbbells in the database versus the fraction of G•C type base pairs in each dumbbell stem, f(G•C). The value of $T_{A\bullet T}$ is obtained as the melting temperature at f(G•C) = 0. Similarly, the value of $T_{G\bullet C}$ is determined from the fitted line where f(G•C) = 1. By virtue of the manner in which they are evaluated the average singlet melting temperatures, $T_{A\bullet T}$ and $T_{G\bullet C}$, are specific for the dumbbells of our database and include the average effects of end-loops on average base pair stability. Slight changes in their values might be expected if additional data from dumbbells with different ends or loop sequences were also included in the database. To apply our thermodynamic parameters for linear duplex DNAs or polymers, $T_{A\bullet T}$ and $T_{G\bullet C}$ are calculated from Frank-Kamenetskii's relationships (equations (4.6 and 4.7)).

5.5 Evaluation of Sequence Specific Nearest-Neighbor Doublet Interactions

Following singlets, the next higher order of sequence dependent interactions are the doublet or n-n interactions. Consider dumbbell *p* with $N_{bp}(p)$ base pairs and T_4 end-loops. In terms of doublets, when the p^{th} dumbbell melts the total free energy change is comprised of the following terms,

$$\Delta G_D(p) = \Delta G_D(p, singlet) + \sum_{i,j=\mathrm{A,T,C,G}} N_{ij}(p)\cdot\delta G_{ij}^{n-n} \tag{5.11}$$

Where ΔG_D(*p, singlet*) is given by equation (5.2) and the second term on the right hand side, $\sum_{i,j=\mathrm{A,T,C,G}} N_{ij}(p)\cdot\delta G_{ij}^{n-n}$ comprises the contributions from n-n sequence dependent interactions. This term includes all deviations, δG_{ij}^{n-n}, from singlet interactions due to sequence dependent n-n effects and depends on the identities of the particular n-n base pair doublets present in the duplex stem of dumbbell *p*. The coefficient, $N_{ij}(p)$, is the number of times the particular n-n doublet (*ij* = A,T,G,C) appears in the stem of dumbbell *p*. Since the *p*th dumbbell stem has $N_{bp}(p)$ base pairs, it necessarily contains $N_{bp}(p)$ - 1 nearest-neighbor doublets. That is, $\sum_{i,j=\mathrm{A,T,C,G}} N_{ij}(p) = N_{\mathrm{bp}}(p) - 1$. The sum in equation (5.11) is over the 10 possible n-n doublets, i.e., (5'-3'): AT/AT, TA/TA, AA/TT, AC/GT, CA/TG, TC/GA, CT/AG, CG/CG, GC/GC, GG/CC. Because the entropy of base pair melting is assumed to be constant, doublet free energies, δG_{ij}^{n-n}, are composed of only enthalpic components. Hence, for all possible doublets, *ij*, it is assumed,

$$\delta G_{ij}^{n-n} = \delta H_{ij}^{n-n} \tag{5.12}$$

Thus, if nearest-neighbor doublet interactions are included, the total transition enthalpy is given by,

$$\begin{aligned}\Delta H_D(p) &= \Delta H_D(p, singlet) + \Delta H_D(p, doublet) \\ &= \Delta H_D(p, singlet) + \sum_{i,j=\mathrm{A,T,C,G}} N_{ij}(p)\cdot\delta H_{ij}^{n-n}\end{aligned} \tag{5.13}$$

where the $\delta H_{ij}^{n\text{-}n}$ are the n-n sequence dependent deviations from the singlet transition enthalpy. Contributions from n-n interactions were determined by fitting the residuals obtained after subtraction of the singlet enthalpies from the total transition enthalpy. After singlet fitting, the residual of the transition enthalpy for dumbbell *p* is,

$$\begin{aligned} \Delta H_{res}(p) &= \Delta H_D(p) - \Delta H_D(p, singlet) \\ &= \Delta H_D(p) - [N_{\mathrm{A\bullet T}}(p)\cdot\Delta H_{\mathrm{A\bullet T}} + N_{\mathrm{G\bullet C}}(p)\cdot\Delta H_{\mathrm{G\bullet C}}] \end{aligned} \tag{5.14}$$

For the p^{th} dumbbell, the combination of equations (5.13) and (5.14) leads to the following linear equation,

$$\sum_{i,j=\mathrm{A,T,C,G}} N_{ij}(p)\cdot\delta H_{ij}^{n-n} = \Delta H_{res}(p) \tag{5.15}$$

The n-n sequence specific interactions are determined by minimizing the difference between the left and right hand sides of equation (5.15) for all dumbbells of the database. For dumbbell *p* this difference is,

$$\sum_{i,j=\mathrm{A,T,C,G}} N_{ij}(p)\cdot\delta H_{ij}^{n-n} - \left\{\Delta H_D(p) - [N_{\mathrm{A\bullet T}}(p)\cdot\Delta H_{\mathrm{A\bullet T}} + N_{\mathrm{G\bullet C}}(p)\cdot\Delta H_{\mathrm{G\bullet C}}]\right\} \tag{5.16}$$

The term in brackets is the residual, $\Delta H_{res}(p)$. Our database contains melting data for 39 molecules (*p* = 1, 39) with different unique central sequence (two to eight base pairs) in the duplex stem. The unique sequence of each stem sequence provides a unique linear equation of the type in equation (5.16). For the 39 molecules in the database, the resulting system of linear equations is,

$$\begin{bmatrix} N_{\mathrm{AA}}(1) & N_{\mathrm{AT}}(1) & \ldots & N_{\mathrm{GG}}(1) \\ N_{\mathrm{AA}}(2) & N_{\mathrm{AT}}(2) & \ldots & N_{\mathrm{GG}}(2) \\ \ldots & \ldots & \ldots & \ldots \\ N_{\mathrm{AA}}(39) & N_{\mathrm{AT}}(39) & \ldots & N_{\mathrm{GG}}(39) \end{bmatrix} * \begin{bmatrix} \delta H_{\mathrm{AA}}^{n-n} \\ \delta H_{\mathrm{AT}}^{n-n} \\ \ldots \\ \delta H_{\mathrm{GG}}^{n-n} \end{bmatrix} = \begin{bmatrix} \Delta H_{res}(1) \\ \Delta H_{res}(2) \\ \ldots \\ \Delta H_{res}(39) \end{bmatrix}$$

Or in the abbreviated matrix notation,

$$\mathbf{M}^{\mathbf{n-n}} * \boldsymbol{\delta}\mathbf{H}^{\mathbf{n-n}} = \boldsymbol{\Delta}\mathbf{H}_{\mathbf{res}} \tag{5.17}$$

The matrix on the left, $\mathbf{M}^{\mathbf{n-n}}$, is constructed according to the stem sequences of the 39 dumbbells. The $N_{AA}(1)$ is the number of occurrences of the AA/TT doublet in the 1st dumbbell, $N_{AT}(2)$ is the number of occurrences of the AT/AT doublet in the 2th dumbbell, etc. Each element of the residuals matrix, $\boldsymbol{\Delta}\mathbf{H}_{\mathbf{res}}$, is determined according to equation (5.14) using the values of $T_{A\bullet T}$ and $T_{G\bullet C}$ determined from singlet fitting, definitions in equations (5.6) and (5.7), constant value of $\Delta S_{bp} = -24.85$ cal.mol^{-1}.K^{-1} and $\Delta H_D(p) = N_{bp}(p) \cdot T_m(p) \cdot \Delta S_{bp}$. The ten unknowns evaluated in the fit, $\boldsymbol{\delta}\mathbf{H}^{\mathbf{n-n}}$, are the deviations from singlet enthalpy due to n-n sequence effects. Fitting of these values is performed by minimizing χ^2,

$$\chi^2 = | \ [\mathbf{M}^{\mathbf{n-n}} * \boldsymbol{\delta}\mathbf{H}^{\mathbf{n-n}} - \boldsymbol{\Delta}\mathbf{H}_{\mathbf{res}}] * \boldsymbol{\sigma}_{\boldsymbol{\Delta H_D}}^{\mathbf{-1}} \ |^2 \tag{5.18}$$

where $\boldsymbol{\sigma}_{\Delta H_D}$ is the diagonal error matrix. Elements of $\boldsymbol{\sigma}_{\Delta H_D}^{-1}$ are reciprocal values of the standard errors of $\Delta H_D(p)$. The errors of $\Delta H_D(p)$ were determined from errors in T_m using equation (3.17). In our analysis, the errors of melting temperatures, $\sigma(T_m)$, were assumed to be the same for all dumbbells (0.2 °C). The errors of $\Delta H_D(p)$ are weighting parameters in χ^2 minimization.

For the n-n doublet interactions our database generates 39 linear equations in only 10 unknowns, δH_{ij}^{n-n}. Since the system (5.17) contains more equations than unknowns, it is overdetermined. The solution for the unknowns that comprise matrix $\mathbf{\delta H}^{\mathbf{n-n}}$ is unique only if matrix $\mathbf{M}^{\mathbf{n-n}}$ is not rank deficient, i.e., it has no zero singular values. For evaluation of n-n effects, if the rank of matrix $\mathbf{M}^{\mathbf{n-n}}$ is 10, then in principle the 10 unknowns can be solved for.

Although there are 10 n-n sequence dependent interactions, unless the ends are ignored, 10 linearly independent equations for the 10 n-n interactions cannot be written. If explicit sequence dependent n-n interactions at the ends are considered there are 14 possible interactions. However, because of constraints on these interactions (equations (4.1) and (4.2)) only 12 linearly independent equations can be written (Goldstein and Benight, 1992; Gray, 1997a). For infinite length or circular DNAs there are only eight linearly independent equations. However, if the molecules considered all have the same non-symmetric ends, as in the case for our dumbbells, the rank of matrix $\mathbf{M}^{\mathbf{n-n}}$ is 9 and there are only nine sequence specific linear combinations of the n-n doublets that can be evaluated that are unique. The nine linearly independent combinations of n-n sequence dependent interactions we have chosen to evaluate from melting curves of dumbbells as a function of solvent ionic strength are listed in Table XII. A number of numerical methods are available for solving the system of linear equations based on the methods of maximum likelihood. As will be described below we employed singular value decomposition

(SVD) to solve our problem (Press et al., 1989).

Even if the unknowns are linearly dependent, matrix $\mathbf{M}^{\mathbf{n-n}}$ is singular with rank less than the number of unknowns, and a unique solution for the unknowns does not exist, a useful solution can be obtained by SVD. In this case the derived δH_{ij}^{n-n} values are not unique, and therefore cannot be meaningfully interpreted in terms of n-n sequence dependent thermodynamic properties for individual n-n doublets (Goldstein and Benight, 1992; Gray, 1997a). Yet these non-unique values can be appropriately summed to predict thermodynamic values of unique sequences, and such a prediction is unique. Alternatively, the unique linearly independent linear combinations of the δH_{ij}^{n-n} can be used to calculate the sequence dependent transition enthalpy.

In this thesis we employ the same nine linear combinations designed earlier (Doktycz et al., 1992). Linear combinations of the numbers of different types of n-n doublets, N_c^{n-n} (c = 1, 9), and linear combinations of the n-n dependent enthalpies, δH_c^{n-n} are listed in Table XII. The method for deriving the set of linear combinations of transition enthalpies, δH_c^{n-n}, from the specified numbers, N_c^{n-n}, is explained in the Appendix. The first eight linear combinations for n-n doublets are general and can be used to calculate the sequence dependent thermodynamics of any duplex DNA oligomers or duplex polymer. However, the ninth combination is specific for dumbbells in our database. Strictly speaking this linear combination should be used only for dumbbells with ends identical to the ends of the dumbbells in the database. In evaluating the n-n doublet interactions from our database of 39 molecules, the number of degrees of freedom is ν = 39-9 = 30.

TABLE XII The Set of Unique Linear Combinations of Nearest-Neighbor Subunits and Corresponding Linear Combinations of Their Enthalpies. The first eight combinations are employed in calculation of any DNA duplex sequence. The last, ninth combination is specific for dumbbells in this study because of the ends.

Numbers ($N_c^{\text{n-n}}$)	Enthalpies ($\delta H_c^{\text{n-n}}$)
$N_1^{\text{n-n}}= N_{\text{AA/TT}}$	$\delta H_1^{\text{n-n}}= \delta H_{\text{AA/TT}}$
$N_2^{\text{n-n}}= N_{\text{CC/GG}}$	$\delta H_2^{\text{n-n}}= \delta H_{\text{CC/GG}}$
$N_3^{\text{n-n}}= N_{\text{AT/AT}} + N_{\text{TA/TA}}$	$\delta H_3^{\text{n-n}}= (\delta H_{\text{AT/AT}} + \delta H_{\text{TA/TA}})/2$
$N_4^{\text{n-n}}= N_{\text{CG/CG}} + N_{\text{GC/GC}}$	$\delta H_4^{\text{n-n}}= (\delta H_{\text{CG/CG}} + \delta H_{\text{GC/GC}})/2$
$N_5^{\text{n-n}}= N_{\text{AC/GT}} + N_{\text{CA/TG}}$	$\delta H_5^{\text{n-n}}= (\delta H_{\text{AC/GT}} + \delta H_{\text{CA/TG}})/2$
$N_6^{\text{n-n}}= N_{\text{AG/CT}} + N_{\text{GA/TC}}$	$\delta H_6^{\text{n-n}}= (\delta H_{\text{AG/CT}} + \delta H_{\text{GA/TC}})/2$
$N_7^{\text{n-n}}= N_{\text{AT/AT}} - N_{\text{TA/TA}} + N_{\text{CG/CG}} - N_{\text{GC/GC}} +2(N_{\text{GA/TC}} - N_{\text{AG/CT}})$	$\delta H_7^{\text{n-n}}= (\delta H_{\text{AT/AT}} - \delta H_{\text{TA/TA}} + \delta H_{\text{CG/CG}} - \delta H_{\text{GC/GC}})/12 +(\delta H_{\text{GA/TC}} -\delta H_{\text{AG/CT}})/6$
$N_8^{\text{n-n}}= N_{\text{AT/AT}} - N_{\text{TA/TA}} - N_{\text{CG/CG}} + N_{\text{GC/GC}} +2(N_{\text{CA/TG}} - N_{\text{AC/GT}})$	$\delta H_8^{\text{n-n}}= (\delta H_{\text{AT/AT}} - \delta H_{\text{TA/TA}} - \delta H_{\text{CG/CG}} + \delta H_{\text{GC/GC}})/12 +(\delta H_{\text{CA/TG}} -\delta H_{\text{AC/GT}})/6$
$N_9^{\text{n-n}}= N_{\text{AG/CT}} - N_{\text{GA/TC}} + N_{\text{CA/TG}} - N_{\text{AC/GT}} +2(N_{\text{CG/CG}} - N_{\text{GC/GC}})$	$\delta H_9^{\text{n-n}}= (\delta H_{\text{AG/CT}} - \delta H_{\text{GA/TC}} + \delta H_{\text{CA/TG}} - \delta H_{\text{AC/GT}})/12 +(\delta H_{\text{CG/CG}} - \delta H_{\text{GC/GC}})/2$

TABLE XIII The Set of Unique Linear Combinations of Next-Nearest-Neighbor Subunits and Corresponding Linear Combinations of Their Enthalpies.

Numbers (N_c^{n-n-n})		Enthalpies (δH_c^{n-n-n})	
$N_1^{n-n-n} =$	$N_{AAA/TTT}$	$\delta H_1^{n-n-n} =$	$\delta H_{AAA/TTT}$
$N_2^{n-n-n} =$	$N_{ATA/TAT}$	$\delta H_2^{n-n-n} =$	$\delta H_{ATA/TAT}$
$N_3^{n-n-n} =$	$N_{GGG/CCC}$	$\delta H_3^{n-n-n} =$	$\delta H_{GGG/CCC}$
$N_4^{n-n-n} =$	$N_{GCG/CGC}$	$\delta H_4^{n-n-n} =$	$\delta H_{GCG/CGC}$
$N_5^{n-n-n} =$	$N_{ACA/TGT} + N_{GTG/CAC}$	$\delta H_5^{n-n-n} =$	$(\delta H_{ACA/TGT} + \delta H_{GTG/CAC})/2$
$N_6^{n-n-n} =$	$N_{AGA/TCT} + N_{GAG/CTC}$	$\delta H_6^{n-n-n} =$	$(\delta H_{AGA/TCT} + \delta H_{GAG/CTC})/2$
$N_7^{n-n-n} =$	$N_{TAA/TTA} + N_{AAT/ATT}$	$\delta H_7^{n-n-n} =$	$(\delta H_{TAA/TTA} + \delta H_{AAT/ATT})/2$
$N_8^{n-n-n} =$	$N_{GGC/GCC} + N_{CGG/CCG}$	$\delta H_8^{n-n-n} =$	$(\delta H_{GGC/GCC} + \delta H_{CGG/CCG})/2$
$N_9^{n-n-n} =$	$N_{AAC/GTT} + N_{TTG/CAA} + N_{TGG/CCA} + N_{ACC/GGT}$	$\delta H_9^{n-n-n} =$	$(\delta H_{AAC/GTT} + \delta H_{TTG/CAA} + \delta H_{TGG/CCA} + \delta H_{ACC/GGT})/4$
$N_{10}^{n-n-n} =$	$N_{AAG/CTT} + N_{GAA/TTC} + N_{GGA/TCC} + N_{AGG/CCT}$	$\delta H_{10}^{n-n-n} =$	$(\delta H_{AAG/CTT} + \delta H_{GAA/TTC} + \delta H_{GGA/TCC} + \delta H_{AGG/CCT})/4$
$N_{11}^{n-n-n} =$	$N_{GTA/TAC} + N_{ATG/CAT} + N_{GCA/TGC} + N_{ACG/CGT}$	$\delta H_{11}^{n-n-n} =$	$(\delta H_{GTA/TAC} + \delta H_{ATG/CAT} + \delta H_{GCA/TGC} + \delta H_{ACG/CGT})/4$
$N_{12}^{n-n-n} =$	$N_{ATC/GAT} + N_{TAG/CTA} + N_{AGC/GCT} + N_{TCG/CGA}$	$\delta H_{12}^{n-n-n} =$	$(\delta H_{ATC/GAT} + \delta H_{TAG/CTA} + \delta H_{AGC/GCT} + \delta H_{TCG/CGA})/4$
$N_{13}^{n-n-n} =$	$N_{TGA/TCA} + N_{AGT/ACT} + N_{GAC/GTC} + N_{CAG/CTG}$	$\delta H_{13}^{n-n-n} =$	$(\delta H_{TGA/TCA} + \delta H_{AGT/ACT} + \delta H_{GAC/GTC} + \delta H_{CAG/CTG})/4$
$N_{14}^{n-n-n} =$	$N_{ACC/GGT} + N_{TGG/CCA} - N_{TTG/CAA} - N_{AAC/GTT} + N_{GTA/TAC} + N_{ATG/CAT} - N_{GCA/TGC} - N_{ACG/CGT}$	$\delta H_{14}^{n-n-n} =$	$(\delta H_{ACC/GGT} + \delta H_{TGG/CCA} - \delta H_{TTG/CAA} - \delta H_{AAC/GTT} + \delta H_{GTA/TAC} + \delta H_{ATG/CAT} - \delta H_{GCA/TGC} - \delta H_{ACG/CGT})/8$
$N_{15}^{n-n-n} =$	$N_{AGG/CCT} + N_{GGA/TCC} - N_{GAA/TTC} - N_{AAG/CTT} + N_{ATC/GAT} + N_{TAG/CTA} - N_{AGC/GCT} - N_{TCG/CGA}$	$\delta H_{15}^{n-n-n} =$	$(\delta H_{AGG/CCT} + \delta H_{GGA/TCC} - \delta H_{GAA/TTC} - \delta H_{AAG/CTT} + \delta H_{ATC/GAT} + \delta H_{TAG/CTA} - \delta H_{AGC/GCT} - \delta H_{TCG/CGA})/8$
$N_{16}^{n-n-n} =$	$N_{GTG/CAC} - N_{ACA/TGT} - N_{AGA/TCT} + N_{GAG/CTC} + N_{TGA/TCA} + N_{AGT/ACT} - N_{GAC/GTC} - N_{CAG/CTG}$	$\delta H_{16}^{n-n-n} =$	$(\delta H_{GTG/CAC} - \delta H_{ACA/TGT} - \delta H_{AGA/TCT} + \delta H_{GAG/CTC} + \delta H_{TGA/TCA} + \delta H_{AGT/ACT} - \delta H_{GAC/GTC} - \delta H_{CAG/CTG})/8$

TABLE XIII (continued)

Numbers ($N_c^{n\text{-}n\text{-}n}$)	Enthalpies ($\delta H_c^{n\text{-}n\text{-}n}$)
$N_{17}^{n\text{-}n\text{-}n} = N_{AGA/TCT} - N_{GAG/CTC} - N_{GTG/CAC} + N_{ACA/TGT} + N_{TTG/CAA} + N_{AAC/GTT} - N_{ACC/GGT} - N_{TGG/CCA} + N_{GAA/TTC} + N_{AAG/CTT} - N_{AGG/CCT} - N_{GGA/TCC} + N_{GTA/TAC} + N_{ATG/CAT} + N_{TGA/TCA} + N_{AGT/ACT} + N_{ATC/GAT} + N_{TAG/CTA} - N_{GAC/GTC} - N_{CAG/CTG} - N_{GCA/TGC} - N_{ACG/CGT} - N_{AGC/GCT} - N_{TCG/CGA}$	$\delta H_{17}^{n\text{-}n\text{-}n} = (\delta H_{AGA/TCT} - \delta H_{GAG/CTC} - \delta H_{GTG/CAC} + \delta H_{ACA/TGT} + \delta H_{TTG/CAA} + \delta H_{AAC/GTT} - \delta H_{ACC/GGT} - \delta H_{TGG/CCA} + \delta H_{GAA/TTC} + \delta H_{AAG/CTT} - \delta H_{AGG/CCT} - \delta H_{GGA/TCC} + \delta H_{GTA/TAC} + \delta H_{ATG/CAT} + \delta H_{TGA/TCA} + \delta H_{AGT/ACT} + \delta H_{ATC/GAT} + \delta H_{TAG/CTA} - \delta H_{GAC/GTC} - \delta H_{CAG/CTG} - \delta H_{GCA/TGC} - \delta H_{ACG/CGT} - \delta H_{AGC/GCT} - \delta H_{TCG/CGA})/24$
$N_{18}^{n\text{-}n\text{-}n} = 2*(N_{ACA/TGT} - N_{GTG/CAC} + N_{GAG/CTC} - N_{AGA/TCT}) + N_{TTG/CAA} + N_{AAC/GTT} - N_{ACC/GGT} - N_{TGG/CCA} - N_{GAA/TTC} - N_{AAG/CTT} + N_{AGG/CCT} + N_{GGA/TCC} + N_{GTA/TAC} + N_{ATG/CAT} - N_{ATC/GAT} - N_{TAG/CTA} - N_{GCA/TGC} - N_{ACG/CGT} + N_{AGC/GCT} + N_{TCG/CGA}$	$\delta H_{18}^{n\text{-}n\text{-}n} = (\delta H_{ACA/TGT} - \delta H_{GTG/CAC} + \delta H_{GAG/CTC} - \delta H_{AGA/TCT})/16 + (\delta H_{TTG/CAA} + \delta H_{AAC/GTT} - \delta H_{ACC/GGT} - \delta H_{TGG/CCA} - \delta H_{GAA/TTC} - \delta H_{AAG/CTT} + \delta H_{AGG/CCT} + \delta H_{GGA/TCC} + \delta H_{GTA/TAC} + \delta H_{ATG/CAT} - \delta H_{ATC/GAT} - \delta H_{TAG/CTA} - \delta H_{GCA/TGC} - \delta H_{ACG/CGT} + \delta H_{AGC/GCT} + \delta H_{TCG/CGA})/32$
$N_{19}^{n\text{-}n\text{-}n} = N_{ATG/CAT} - N_{GTA/TAC} - N_{ACG/CGT} + N_{GCA/TGC}$	$\delta H_{19}^{n\text{-}n\text{-}n} = (\delta H_{ATG/CAT} - \delta H_{GTA/TAC} - \delta H_{ACG/CGT} + \delta H_{GCA/TGC})/4$
$N_{20}^{n\text{-}n\text{-}n} = N_{ATC/GAT} - N_{TAG/CTA} - N_{AGC/GCT} + N_{TCG/CGA}$	$\delta H_{20}^{n\text{-}n\text{-}n} = (\delta H_{ATC/GAT} - \delta H_{TAG/CTA} - \delta H_{AGC/GCT} + \delta H_{TCG/CGA})/4$
$N_{21}^{n\text{-}n\text{-}n} = N_{GTA/TAC} - N_{ATG/CAT} - N_{TGA/TCA} + N_{AGT/ACT} - N_{GAC/GTC} + N_{CAG/CTG} + N_{GCA/TGC} - N_{ACG/CGT}$	$\delta H_{21}^{n\text{-}n\text{-}n} = (\delta H_{GTA/TAC} - \delta H_{ATG/CAT} - \delta H_{TGA/TCA} + \delta H_{AGT/ACT} - \delta H_{GAC/GTC} + \delta H_{CAG/CTG} + \delta H_{GCA/TGC} - \delta H_{ACG/CGT})/8$
$N_{22}^{n\text{-}n\text{-}n} = N_{TGA/TCA} - N_{AGT/ACT} + N_{ATC/GAT} - N_{TAG/CTA} - N_{GAC/GTC} + N_{CAG/CTG} + N_{AGC/GCT} - N_{TCG/CGA}$	$\delta H_{22}^{n\text{-}n\text{-}n} = (\delta H_{TGA/TCA} - \delta H_{AGT/ACT} + \delta H_{ATC/GAT} - \delta H_{TAG/CTA} - \delta H_{GAC/GTC} + \delta H_{CAG/CTG} + \delta H_{AGC/GCT} - \delta H_{TCG/CGA})/8$
$N_{23}^{n\text{-}n\text{-}n} = N_{AAT/ATT} - N_{TAA/TTA} + N_{CGG/CCG} - N_{GGC/GCC} + N_{GAA/TTC} - N_{AAG/CTT} - N_{AGG/CCT} + N_{GGA/TCC}$	$\delta H_{23}^{n\text{-}n\text{-}n} = (\delta H_{AAT/ATT} - \delta H_{TAA/TTA} + \delta H_{CGG/CCG} - \delta H_{GGC/GCC} + \delta H_{GAA/TTC} - \delta H_{AAG/CTT} - \delta H_{AGG/CCT} + \delta H_{GGA/TCC})/8$
$N_{24}^{n\text{-}n\text{-}n} = N_{AAT/ATT} - N_{TAA/TTA} - N_{CGG/CCG} + N_{GGC/GCC} + N_{TTG/CAA} - N_{AAC/GTT} - N_{ACC/GGT} + N_{TGG/CCA}$	$\delta H_{24}^{n\text{-}n\text{-}n} = (\delta H_{AAT/ATT} - \delta H_{TAA/TTA} - \delta H_{CGG/CCG} + \delta H_{GGC/GCC} + \delta H_{TTG/CAA} - \delta H_{AAC/GTT} - \delta H_{ACC/GGT} + \delta H_{TGG/CCA})/8$

TABLE XIII (continued)

Numbers (N_c^{n-n-n})	Enthalpies (δH_c^{n-n-n})
$N_{25}^{n-n-n} = 2*(N_{TTG/CAA} -N_{AAC/GTT} +N_{ACC/GGT}$ $-N_{TGG/CCA}) +N_{AAT/ATT} -N_{TAA/TTA}$ $+N_{CGG/CCG} -N_{GGC/GCC} -N_{GAA/TTC}$ $+N_{AAG/CTT} +N_{AGG/CCT} -N_{GGA/TCC}$ $+N_{GTA/TAC} -N_{ATG/CAT} +N_{TGA/TCA} -N_{AGT/ACT}$ $+N_{GAC/GTC} -N_{CAG/CTG} +N_{GCA/TGC} -N_{ACG/CGT}$	$\delta H_{25}^{n-n-n} = (\delta H_{TTG/CAA} -\delta H_{AAC/GTT} +\delta H_{ACC/GGT}$ $-\delta H_{TGG/CCA})/16 +(\delta H_{AAT/ATT} -\delta H_{TAA/TTA}$ $+\delta H_{CGG/CCG} -\delta H_{GGC/GCC} -\delta H_{GAA/TTC} +\delta H_{AAG/CTT}$ $+\delta H_{AGG/CCT} -\delta H_{GGA/TCC} +\delta H_{GTA/TAC} -\delta H_{ATG/CAT}$ $+\delta H_{TGA/TCA} -\delta H_{AGT/ACT} +\delta H_{GAC/GTC} -\delta H_{CAG/CTG}$ $+\delta H_{GCA/TGC} -\delta H_{ACG/CGT})/32$
$N_{26}^{n-n-n} = 2*(N_{GAA/TTC} -N_{AAG/CTT} +N_{AGG/CCT}$ $-N_{GGA/TCC}) +N_{AAT/ATT} -N_{TAA/TTA} -N_{CGG/CCG}$ $+N_{GGC/GCC} -N_{TTG/CAA} +N_{AAC/GTT}$ $+N_{ACC/GGT} -N_{TGG/CCA} +N_{TGA/TCA} -N_{AGT/ACT}$ $-N_{ATC/GAT} +N_{TAG/CTA} -N_{GAC/GTC} +N_{CAG/CTG}$ $-N_{AGC/GCT} +N_{TCG/CGA}$	$\delta H_{26}^{n-n-n} = (\delta H_{GAA/TTC} -\delta H_{AAG/CTT} +\delta H_{AGG/CCT}$ $-\delta H_{GGA/TCC})/16 +(\delta H_{AAT/ATT} -\delta H_{TAA/TTA}$ $-\delta H_{CGG/CCG} +\delta H_{GGC/GCC} -\delta H_{TTG/CAA} +\delta H_{AAC/GTT}$ $+\delta H_{ACC/GGT} -\delta H_{TGG/CCA} +\delta H_{TGA/TCA} -\delta H_{AGT/ACT}$ $-\delta H_{ATC/GAT} +\delta H_{TAG/CTA} -\delta H_{GAC/GTC} +\delta H_{CAG/CTG}$ $-\delta H_{AGC/GCT} +\delta H_{TCG/CGA})/32$
$N_{27}^{n-n-n} = N_{ATC/GAT} -N_{AGC/GCT} +N_{TAG/CTA}$ $-N_{TGG/CCA} -N_{TTG/CAA} -N_{TCG/CGA}$ $+2*[N_{AGG/CCT} -N_{AGA/TCT} +N_{GAG/CTC}$ $-N_{GAA/TTC} +2*(N_{AAG/CTT} +N_{ATG/CAT}$ $-N_{GGA/TCC} -N_{GCA/TGC}) +5*(N_{ACA/TGT}$ $-N_{GTG/CAC})] +3*(N_{AAT/ATT} -N_{TAA/TTA}$ $+N_{CGG/CCG} -N_{GGC/GCC}) +5*(N_{CAG/CTG}$ $-N_{TGA/TCA}) +11*(N_{ACC/GGT} -N_{AAC/GTT})$ $+13*(N_{AGT/ACT} -N_{GAC/GTC}) +14*(N_{ACG/CGT}$ $-N_{GAC/GTC})$	$\delta H_{27}^{n-n-n} = \{\delta H_{ATC/GAT} -\delta H_{AGC/GCT} +\delta H_{TAG/CTA}$ $-\delta H_{TGG/CCA} -\delta H_{TTG/CAA} -\delta H_{TCG/CGA}$ $+2*[\delta H_{AGG/CCT} -\delta H_{AGA/TCT} +\delta H_{GAG/CTC}$ $-\delta H_{GAA/TTC} +2*(\delta H_{AAG/CTT} +\delta H_{ATG/CAT}$ $-\delta H_{GGA/TCC} -\delta H_{GCA/TGC}) +5*(\delta H_{ACA/TGT}$ $-\delta H_{GTG/CAC})] +3*(\delta H_{AAT/ATT} -\delta H_{TAA/TTA}$ $+\delta H_{CGG/CCG} -\delta H_{GGC/GCC}) +5*(\delta H_{CAG/CTG}$ $-\delta H_{TGA/TCA}) +11*(\delta H_{ACC/GGT} -\delta H_{AAC/GTT})$ $+13*(\delta H_{AGT/ACT} -\delta H_{GAC/GTC}) +14*(\delta H_{ACG/CGT}$ $-\delta H_{GAC/GTC}) \}/1344$

5.6 Evaluation of Next-Nearest-Neighbor Sequence Specific Interactions

To evaluate the next nearest neighbor (n-n-n) sequence specific interactions the total free energy of dumbbell p is partitioned into the sum of three terms,

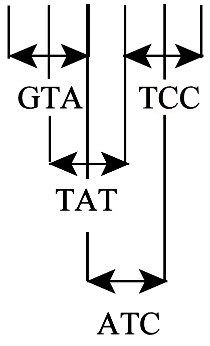
$$\Delta G_D(p) \;=\; \Delta G_D(p, singlet) \;+\; \Delta G_D(p, doublet) \;+ \sum_{i,j,k=\mathrm{A,T,C,G}} N_{ijk}(p)\cdot\delta G_{ijk}^{n-n-n} \qquad (5.19)$$

The first two terms for the singlet and doublet contributions are evaluated as described above. The last term accounts for n-n-n triplet interactions. The sum is calculated over the 32 possible triplet sequences in duplex DNA. These are (5'-3'): AAA/TTT, AAG/CTT, AAT/ATT, AAC/GTT, AGA/TCT, AGG/CCT, AGT/ACT, AGC/GCT, ATA/TAT, ATG/CAT, ATC/GAT, ACA/TGT, ACG/CGT, ACC/GGT, GAA/TTC, GAG/CTC, GAC/GTC, GGA/TCC, GGG/CCC, GGC/GCC, GTA/TAC, GTG/CAC, GCA/TGC, GCG/CGC, TAA/TTA, TAG/CTA, TGA/TCA, TGG/CCA, TTG/CAA, TCG/CGA, CAG/CTG and CGG/CCG. Thus, if the number of duplex base pairs in the stem of dumbbell p is $\mathrm{N_{bp}}(p)$, there are $\mathrm{N_{bp}}(p) - 2$ triplets in the duplex. That is,

$$\sum_{i,j,k=\mathrm{A,T,C,G}} N_{ijk}(p) \;=\; N_{\mathrm{bp}}(p) - 2 \qquad (5.20)$$

The last term in equation (5.19), $\sum_{i,j,k=\mathrm{A,T,C,G}} N_{ijk}(p)\cdot\delta G_{ijk}^{n-n-n}$, is the sum of deviations from the average doublet free energy over all possible sequence dependent triplet interaction free energies due to the particular n-n-n triplets resident in the duplex stem of dumbbell p. To illustrate the calculation procedure for n-n-n interactions, consider the 24th dumbbell in our database (depicted in Figure 5) having the stem sequence:

- 5'-G-T-A-T-C-C-A-G-T-A-C-T-G-G-A-T-A-C-3' -
- 3'-C-A-T-A-G-G-T-C-A-T-G-A-C-C-T-A-T-G-5' -



The dumbbell stem is 18 base pairs long. The first four triplets from the left side are marked, GTA/TAC, TAT/ATA, ATC/GAT, TCC/GGA. The next triplets are CCA/TGG, CAG/CTG, AGT/ACT, GTA/TAC, TAC/GTA, ACT/AGT, CTG/CAG, TGG/CCA, GGA/TCC, GAT/ATC, ATA/TAT and TAC/GTA. To calculate the n-n-n dependent free energy for this dumbbell, the 16 triplet free energies, δG_{ijk}^{n-n-n} for the triplets in the sequence (listed above) are added up. Of 16 triplets, some are present more then once. For example, the GTA/TAC triplet is the same as the TAC/GTA triplet and is present four times ($N_{GTA/TAC}(24) = 4$). The numbers of the 32 possible triplets in the 24th dumbbell are,

$N_{AAA/TTT}(24) = 0$ $N_{AAG/CTT}(24) = 0$ $N_{AAT/ATT}(24) = 0$ $N_{AAC/GTT}(24) = 0$

$N_{AGA/TCT}(24) = 0$ $N_{AGG/CCT}(24) = 0$ $N_{AGT/ACT}(24) = 2$ $N_{AGC/GCT}(24) = 0$

$N_{ATA/TAT}(24) = 2$ $N_{ATG/CAT}(24) = 0$ $N_{ATC/GAT}(24) = 2$ $N_{ACA/TGT}(24) = 0$

$N_{ACG/CGT}(24) = 0$ $N_{ACC/GGT}(24) = 0$ $N_{GAA/TTC}(24) = 0$ $N_{GAG/CTC}(24) = 0$

$N_{GAC/GTC}(24) = 0$	$N_{GGA/TCC}(24) = 2$	$N_{GGG/CCC}(24) = 0$	$N_{GGC/GCC}(24) = 0$
$N_{GTA/TAC}(24) = 4$	$N_{GTG/CAC}(24) = 0$	$N_{GCA/TGC}(24) = 0$	$N_{GCG/CGC}(24) = 0$
$N_{TAA/TTA}(24) = 0$	$N_{TAG/CTA}(24) = 0$	$N_{TGA/TCA}(24) = 0$	$N_{TGG/CCA}(24) = 2$
$N_{TTG/CAA}(24) = 0$	$N_{TCG/CGA}(24) = 0$	$N_{CAG/CTG}(24) = 2$	$N_{CGG/CCG}(24) = 0$

It should be noted that the consecutive next-nearest-neighbor subunits partially overlay each other and share the same doublet. This treatment is possible because the shorter range, nearest-neighbor and singlet interactions were subtracted from the total free energy in previous steps. Because each triplet consists of two successive n-n doublets, the n-n-n interactions are evaluated as a deviation from the average n-n interactions. For instance, triplet 5'-GTA-3' can be considered as the combination of 5'-GT-3' and 5'-TA-3' stacks. Thus, the n-n-n interaction in 5'-GTA-3' is a deviation from the average of the 5'-GT-3' and 5'-TA-3' n-n interactions. This approach to the n-n-n model is an extension of the singlet formalism of the n-n model (see Chapter 4.5).

As for n-n interactions the n-n-n free energies, δG_{ijk}^{n-n-n}, are assumed to be composed entirely of enthalpic components, and for all possible triplets, *ijk*, it is presumed that,

$$\delta G_{ijk}^{n-n-n} = \delta H_{ijk}^{n-n-n} \tag{5.21}$$

Consequently, the total transition enthalpy for melting the of p^{th} dumbbell is given by,

$$\Delta H_D(p) = \Delta H_D(p, singlet) + \Delta H_D(p, doublet) + \sum_{i,j,k=\mathrm{A,T,C,G}} N_{ijk}(p) \cdot \delta H_{ijk}^{n-n-n} \tag{5.22}$$

The last term, $\sum\limits_{i,j,k=\mathrm{A,T,C,G}} N_{ijk}(p)\cdot\delta H_{ijk}^{n-n-n}$, contains the enthalpic contributions from n-n-n triplet interactions. It is the sum over deviations of the transition enthalpies for the n-n-n triplets present in the particular duplex of dumbbell *p*, from the average.

The n-n-n interactions were evaluated from the residual enthalpies obtained after singlet and doublet fitting. After subtraction of the singlet and doublet enthalpies from the total transition enthalpy, $\Delta H_D(p)$ the residual enthalpy of the p^{th} dumbbell is,

$$\Delta H_{res}^{n-n}(p) = \Delta H_D(p) - [N_{\mathrm{A\bullet T}}(p)\cdot\Delta H_{\mathrm{A\bullet T}} + N_{\mathrm{G\bullet C}}(p)\cdot\Delta H_{\mathrm{G\bullet C}}] - \sum_{i,j=\mathrm{A,T,C,G}} N_{ij}(p)\cdot\delta H_{ij}^{n-n} \qquad (5.23)$$

Thus, the system of linear equations in the n-n-n triplet interactions, generated from the 39 dumbbells of the database, is,

$$\begin{bmatrix} N_{\mathrm{AAA}}(1) & N_{\mathrm{AAT}}(1) & \ldots & N_{\mathrm{GGG}}(1) \\ N_{\mathrm{AAA}}(2) & N_{\mathrm{AAT}}(2) & \ldots & N_{\mathrm{GGG}}(2) \\ \ldots & \ldots & \ldots & \ldots \\ N_{\mathrm{AAA}}(39) & N_{\mathrm{AAT}}(39) & \ldots & N_{\mathrm{GGG}}(39) \end{bmatrix} * \begin{bmatrix} \delta H_{\mathrm{AAA}}^{n-n-n} \\ \delta H_{\mathrm{AAT}}^{n-n-n} \\ \ldots \\ \delta H_{\mathrm{GGG}}^{n-n-n} \end{bmatrix} = \begin{bmatrix} \Delta H_{res}^{n-n}(1) \\ \Delta H_{res}^{n-n}(2) \\ \ldots \\ \Delta H_{res}^{n-n}(39) \end{bmatrix} \qquad (5.24)$$

Where $N_{\mathrm{AAA}}(p)$ is the number of times the AAA/TTT triplet is present, $N_{\mathrm{AAT}}(p)$ is the number of times the AAT/ATT triplet is present, etc., in the duplex stem of dumbbell *p*. The unknowns $\delta H_{\mathrm{ijk}}^{\mathrm{n-n-n}}$, are the n-n-n dependent deviations that the solution provides. The right hand side contains the residual enthalpies, $\Delta H_{\mathrm{res}}^{\mathrm{n-n}}(p)$, after the singlet and n-n fit, which are calculated for each dumbbell (p = 1, 39) according to equation (5.23). In abbreviated form equation (5.24) can be written,

$$\mathbf{M}^{\mathbf{n-n-n}} * \boldsymbol{\delta}\mathbf{H}^{\mathbf{n-n-n}} = \boldsymbol{\Delta}\mathbf{H}^{\mathbf{n-n}}_{\mathbf{res}} \qquad (5.25)$$

Since there are 32 triplets in DNA duplex, the system of linear equations in equation (5.25) has 32 unknowns in the column matrix $\boldsymbol{\delta}\mathbf{H}^{\mathbf{n-n-n}}$. Equation (5.25) is solved in a least-squares sense by minimizing χ^2,

$$\chi^2 = | \ [\mathbf{M}^{\mathbf{n-n-n}} * \boldsymbol{\delta}\mathbf{H}^{\mathbf{n-n-n}} - \boldsymbol{\Delta}\mathbf{H}^{\mathbf{n-n}}_{\mathbf{res}}] * \boldsymbol{\sigma}^{-1}_{\Delta H_D} \ |^2 \qquad (5.26)$$

Ideally, we would like to solve the system of linear equations in equation (5.24) for the 32 unknown δH^{n-n-n}_{ijk} values. However, the matrix $\mathbf{M}^{\mathbf{n-n-n}}$ is singular and rank deficient, and 32 unknowns cannot be uniquely solved for. This is because for any DNA duplex with ends, E, there are six constraints on the numbers of n-n-n triplet interactions that include the ends (Goldstein, R.F., personal communication). The corresponding constraint equations are given in Table XIV. These six constraint equations reduce by six the possible number of linearly independent equations that can be written for the n-n-n triplets and therefore reduce by six the number of unknowns that can be evaluated. For example, polymeric DNAs can contain up to 32 different n-n-n triplets in their sequence, but only 26 unique n-n-n dependent parameters can be evaluated. The set of linearly independent combinations of n-n-n sequences we chose to evaluate are listed in Table XIII. The method to generate the n-n-n linearly independent combinations δH_c^{n-n-n} from the linear combinations of numbers, N_c^{n-n-n}, is given in the Appendix. Twenty six linear combinations of triplets correspond to unique linear combinations in any

TABLE XIV Six Constraints of the Numbers of Triplets Including Ends, E, in DNA Duplexes. N_{ijk} denotes the number of the given triplet *ijk* in DNA duplex.

Constraints
$N_{AAT/ATT} - N_{TAA/TTA} - N_{CAA/TTG} + N_{AAC/GTT} - N_{GAA/TTC} + N_{AAG/CTT} + N_{AAE/ETT} - N_{EAA/TTE} = 0$
$N_{CCG/CGG} - N_{GCC/GGC} - N_{ACC/GGT} + N_{CCA/TGG} + N_{AGG/CCT} - N_{GGA/TCC} + N_{CCE/EGG} - N_{ECC/GGE} = 0$
$N_{CAA/TTG} - N_{AAC/GTT} + N_{ACC/GGT} - N_{CCA/TGG} - N_{GTA/TAC} - N_{TCA/TGA} + N_{ACT/AGT} + N_{CAT/ATG} -$ $-N_{GAC/GTC} - N_{GCA/TGC} + N_{ACG/CGT} + N_{CAG/CTG} + N_{CAE/ETG} - N_{ECA/TGE} + N_{ACE/EGT} - N_{EAC/GTE} = 0$
$N_{GAA/TTC} - N_{AAG/CTT} + N_{AGG/CCT} - N_{GGA/TCC} - N_{TCA/TGA} + N_{ATC/GAT} + N_{ACT/AGT} - N_{CTA/TAG} +$ $+N_{GAC/GTC} + N_{AGC/GCT} - N_{CGA/TCG} - N_{CAG/CTG} + N_{GAE/ETC} - N_{EGA/TCE} + N_{AGE/ECT} - N_{EAG/CTE} = 0$
$- N_{GAA/TTC} + 2 *N_{AGA/TCT} - N_{AAG/CTT} + N_{AGG/CCT} - 2*N_{GAG/CTC} + N_{GGA/TCC} + N_{TCA/TGA} -$ $-N_{ATC/GAT} + N_{ACT/AGT} - N_{CTA/TAG} - N_{GAC/GTC} + N_{AGC/GCT} + N_{CGA/TCG} - N_{CAG/CTG} - N_{GAE/ETC} +$ $+N_{EGA/TCE} + N_{AGE/ECT} - N_{EAG/CTE} = 0$
$- N_{CAA/TTG} + 2*N_{ACA/TGT} - N_{AAC/GTT} + N_{ACC/GGT} - 2*N_{CAC/GTG} + N_{CCA/TGG} - N_{GTA/TAC} + N_{TCA/TGA} +$ $N_{ACT/AGT} - N_{CAT/ATG} - N_{GAC/GTC} + N_{GCA/TGC} + N_{ACG/CGT} - N_{CAG/CTG} - N_{CAE/ETG} + N_{ECA/TGE} +$ $+ N_{ACE/EGT} - N_{EAC/GTE} = 0$

duplex. Linear combination number 27 is specific for the identical ends of the dumbbells in the database. Strictly speaking this linearly independent combination is only useful for the dumbbells in the database or dumbbells with the same end and end-loop sequences as those in the database. The number of degrees of freedom in fits of the n-n-n triplet combinations is $\nu = 39 - 27 = 12$.

In order to solve the above systems of linear equations for the n-n doublet interactions (equation (5.18)) and n-n-n triplet interactions (equation (5.26)), as mentioned above we employed singular value decomposition, SVD. This method is a powerful and rigorous analytical tool, with a number of practical advantages over other common methods. It can detect situations where there is insufficient information (insufficiently representative set of DNA sequences) to allow a fit with a given model. SVD can also diagnose ill-conditioned linear systems or matrices that are not actually singular, but numerically close to being singular. Such matrices are singular within the rounding error of the computer algorithm. In such cases, other methods (LU decomposition or Gaussian elimination) fail to give satisfactory results (Press et al., 1989). SVD allows elimination of combinations of equations that are so corrupted by roundoff error as to be useless. SVD also indicates if the unknowns are linearly dependent or independent.

In this work two algorithms of SVD were utilized and programed in FORTRAN77. The first algorithm was published in Numerical Recipes (Press et al., 1989). Its routines are fast and as simple as possible. Consequently, the algorithm is less robust and may encounter problems with computer instabilities in the case of a few specific matrixes. The second algorithm from Linear Algebra Package (LAPACK) was more robust and was used to check results generated by the first algorithm (Anderson et al., 1994). The LAPACK version 2.0 library was downloaded from the Netlib software depository at http://www.netlib.org. The Basic Linear Algebra Subprograms (BLAS) that are closely connected and called by LAPACK subroutines were supplied and optimized as part of a f77 compiler (Hewlett-Packard). Calculations were run on an HP Unix workstation (Model 735/125) with HP-UX 9000/735 operating system. Programs were compiled with an HP-UX f77 compiler. Both algorithms produced results that were identical within the rounding error of computer (five significant figures).

5.7 The Goodness of Fit and Value of Q.

In the aforementioned analytical procedures, the significance of χ^2 and the quality of all fits was judged by their Q values. Assuming that the errors in T_m are distributed normally, Q can be calculated as the incomplete gamma function (Press et al., 1989), viz.

$$Q(\frac{\nu}{2},\frac{\chi^2}{2}) = \frac{\int_{\frac{\chi^2}{2}}^{\infty} e^{-t} t^{\frac{\nu}{2}-1} dt}{\int_{0}^{\infty} e^{-t} t^{\frac{\nu}{2}-1} dt} \tag{5.27}$$

Where ν is the number of degrees of freedom in the fit. The number of degrees of freedom, ν, is equal to the difference between the number of linear equations (39 in our case) and the number of linear independent unknowns to be solved for. The value of Q is the probability that a fit as poor or poorer (greater χ^2) could be obtained by random chance alone. The better the fit, the higher Q. If Q is very small then it is improbable that discrepancies between the database and model fits arise merely from random fluctuations (errors) alone. In which case the model is judged to be inadequate and should be rejected based on the low probability that random errors would yield a value of χ^2 as large as that observed. Small Q values can originate not only from use of an inadequate model, but also from an underestimate of errors. In total, our measurements indicate that the average T_m error was about 0.19 °C regardless of Na^+ concentration. For our analysis we assumed that all melting temperatures have an error of 0.2 °C. Calculating Q according to equation (5.27) presumes that T_m errors follow a normal distribution. Since three to eight experiments were performed for each sample in each Na^+ environment, we were able to obtain reproducible T_m values (within the error stated above), but the data sample size was too small to accurately determine the distribution of errors. When it is not possible to rigorously prove that errors in measurements follow a normal distribution the value of Q for acceptance of the fit is often set to a relatively low value (Xia, 1998). Following this precedent, we consider Q values less than 0.01 to convey an inadequate, unacceptable fit with the model. Fits are deemed reasonable if $0.1 > Q > 0.01$ and the model is marginal. Fits having $Q > 0.1$ are considered to be excellent, the fitting function is a good approximation of measured values within the error of measurements and the model should not be rejected.

Evaluations of the integrals in (5.27) are tedious. The probability Q was computed by two algorithms. One can be found in Numerical Recipes (Press et al., 1989), the other was downloaded from the Netlib software depository at http://www.netlib.org (DiDonato and Morris, 1987) (The employed algorithm 654 is located at the TOMS library.). Both algorithms produced Q values within the roundoff errors of the computer (five significant figures).

5.8 Investigations of the Sequence Dependence of the Entropy of Base Pair Melting

Our analytical approach relies on the assumption that the entropy change in melting a base pair, ΔS_{bp}, is essentially independent of sequence. That is, the entropy of melting A•T or G•C base pairs is the same, i.e. $\Delta S_{A\bullet T} = \Delta S_{G\bullet C} \equiv \Delta S_{bp}$. We employed the optical melting database to verify this assumption and investigate the potential sequence dependence of ΔS_{bp}. Melting data collected in 115 mM Na^{+} for the 39 dumbbells of the database provided values for the peak heights on the differential melting curves, $\left[\frac{d\theta_B}{dT}\right]_{T=T_m}$, and T_m's. A method to determine the transition enthalpy and entropy of DNA dumbbells from these values is explained in Chapter 3.3.2. Averages of these values for each molecule determined from several experiments were employed to obtain estimates of the transition enthalpy, ΔH_D, and entropy, ΔS_D, according to equations (3.9) and (3.11). Errors of the enthalpies and entropies were estimated using equations (3.12) and (3.13). These estimates were made assuming melting transitions of the 39 molecules of the database are two-state. Our studies of dumbbells with different size end-loops and 16 base pair stems indicate dumbbells with this duplex length and T_4 end-loops melt in a two-state process. For the present analysis we assumed the 39 dumbbells of the optical melting database with duplex stems having 14 to 20 base pairs and T_4 end-loops also melt in a two-state manner. We assessed whether the two-state values of ΔS_D for the dumbbells could be accurately fitted with a single sequence independent functional form for the loop entropy of melting, that only depends on the number of nucleotides in the melted circle. The graphically evaluated two-state entropy values from equation (3.11) were fit using the functional form that has been employed in several previous studies of dumbbells with variable end-loop and circle size (Paner et al., 1992; Paner et al., 1996). The form of the sequence independent circle loop entropy function is, $F_{circ}(N_p) = L(N_p)/(N_p+1)^{1.5}$, where N_p is the total number of bases in the melted minicircle. Thus, for each dumbbell, p = 1, 39, the following equations were generated,

$$\Delta S_D(p) = -4 \cdot R \cdot T_m(p) \cdot \left[\frac{d\theta_B}{dT}\right]_{\max} = R \cdot \ln\left[\frac{L(N_p)}{(N_p+1)^{1.5}}\right] \tag{5.28}$$

$$\Delta S_{Loop}(p) = R \cdot \ln\left[\frac{L(N_p)}{(N_p+1)^{1.5}}\right] \tag{5.29}$$

Where $\Delta S_D(p)$ is the two-state transition entropy of the p^{th} dumbbell comprised of N_p nucleotides. Using SVD, the resulting 39 equations, one for each dumbbell, were solved simultaneously by least-squares fitting $L(N_p)$ within the limits of the experimental errors involved in determining T_m and $\left[\frac{\mathrm{d}\theta_B}{\mathrm{d}T}\right]_{T=T_m}$. The goodness of the fits was ascertained from the Q value (described above). A value of Q ($\geq$0.01) was taken to mean that the residuals between the experimentally determined (two-state) $\Delta S_D(p)$ values and those calculated from the fitted values of $L(N_p)$ were within experimental error and there is no significant length or sequence dependence of the entropy of base pair melting. Alternatively if Q was very small ($\leq$0.01), the sequence independent loop entropy function would be deemed unable to fit all of the data (within experimental error) suggesting that the entropy of base pair melting may depend significantly on sequence composition.

5.9 Optical Melting Curves of DNA Dumbbells

Melting transitions of DNA dumbbells were investigated by UV spectroscopy. Procedures are described in Section 2.7. The average melting curves for the 22 new dumbbells of the present study (dumbbells 18-39 of Figure 5) determined from at least three experiments in 25 mM Na^+ are depicted in Figure 6. The fraction of broken base pairs, θ_B, is plotted vs temperature. Similar curves were also obtained in 55 mM, 85 mM and 115 mM Na^+ (not shown). Since the heating and cooling melting curves had precisely the same shape and overlapped within the error, they were averaged together. The sigmoidal melting curves in Figure 6 span a wide range of temperatures due to the different central sequences of the dumbbells and display clear differences due to length, sequence and overall %GC of the dumbbell stems. The curves in Figure 6 indicate that the melting transitions of the DNA dumbbells are essentially monophasic, although for a few molecules (dumbbells 19, 21, 22, 23 and 37 of Figure 5) melting curves exhibit significant pre-transitions shoulders. Since heating and cooling curves overlapped for all DNA samples, the melting transitions were reversible and assumed to occur at equilibrium. The temperature spread between the melting transitions for the most stable and least stable molecules is about 13 °C in 25 mM Na^+, and is about the same in the other Na^+ environments.

Values of the melting temperature, T_m, were determined from the temperature at the maximum peak height, $\left[\frac{d\theta_B}{dT}\right]_{T=T_m}$, on plots of the derivative ($d\theta_B/dT$) versus temperature. These plots (not shown) were obtained from the curves in Figure 6 and other melting curves measured in 55 mM, 85 mM and 115 mM Na^+. The T_m values obtained for the 22 dumbbells in each salt environment are summarized in Table XV. For all molecules in all salt environments, T_m was independent of DNA dumbbell concentration over the range from 0.4 to 1.2 μM. This is

Figure 6. Average melting curves of 22 dumbbells measured in 25 mM Na^+. The fraction of broken base pairs, θ_B, is plotted vs temperature for dumbbells 18-39 of Figure 5.

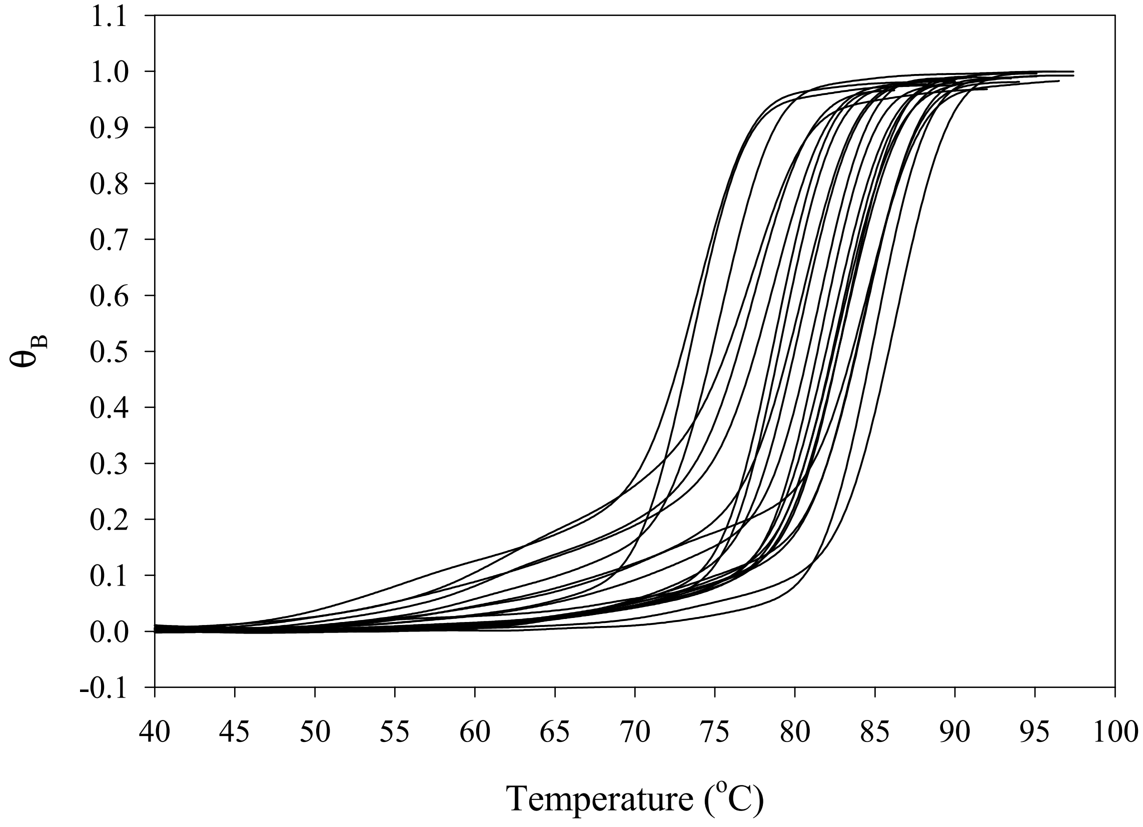

1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
θ_B
40
45
50
55
60
65
70
75
80
85
90
95
100
Temperature (^{o}C)

TABLE XV Average Melting Temperatures of 22 DNA Dumbbells in 25, 55, 85 and 115 mM Na^+. Melting temperatures of dumbbells 1-17 were reported earlier (Doktycz et al., 1992).

Central Sequence	*f*(G•C)	25 mM Na^+ T_m (°C)	55 mM Na^+ T_m (°C)	85 mM Na^+ T_m (°C)	115 mM Na^+ T_m (°C)
18. TAATTA	0.333	73.57	77.35	79.55	81.32
19. TTTAAA	0.333	73.80	77.43	79.83	81.32
20. AATATT	0.333	75.35	78.62	80.82	82.38
21. CTTAAG	0.444	77.19	81.22	83.42	84.82
22. TCTAGA	0.444	77.47	81.24	83.04	84.72
23. TTCAAG	0.444	78.60	82.57	84.53	86.03
24. AGTACT	0.444	78.76	82.66	84.80	86.10
25. TGATCA	0.444	79.11	82.95	84.96	86.47
26. AAGCTT	0.444	80.37	84.17	86.20	87.60
27. GAATTC	0.444	80.40	84.08	86.12	87.70
28. AAGGTTCC	0.500	81.58	85.22	87.45	88.99
29. GGTAAC	0.500	81.62	85.18	87.14	88.79
30. GTAC	0.500	82.24	85.65	87.42	88.73
31. GACT	0.500	82.80	86.22	87.95	89.55
32. GATC	0.500	83.03	86.25	88.35	89.67
33. CTCGAG	0.556	82.69	86.58	88.51	90.03
34. CCTGGT	0.556	83.10	86.70	88.83	90.25
35. GGATCC	0.556	84.25	87.90	89.85	91.33
36. GAGCCT	0.556	84.41	88.31	90.19	91.68
37. ACCGGT	0.556	84.46	87.93	89.83	91.20
38. GTCGAC	0.556	84.94	88.69	90.68	92.25
39. GCATGC	0.556	86.22	89.77	91.90	93.20

TABLE XVI Average Maxima of Derivative θ_B With Respect to Temperature Determined in 25, 55, 85 and 115 mM Na$^+$.

Central Sequence	$\left[\frac{d\theta_B}{dT}\right]_{T=T_m}$			
	25 mM Na$^+$	55 mM Na$^+$	85 mM Na$^+$	115 mM Na$^+$
18. TAATTA	0.13378	0.14085	0.13585	0.14044
19. TTTAAA	0.11141	0.11218	0.11170	0.11534
20. AATATT	0.12472	0.12456	0.13046	0.12263
21. CTTAAG	0.09217	0.09401	0.09485	0.09174
22. TCTAGA	0.10491	0.10895	0.10583	0.10675
23. TTCAAG	0.10905	0.10959	0.10909	0.10970
24. AGTACT	0.13942	0.14107	0.13832	0.14275
25. TGATCA	0.13759	0.13673	0.14049	0.14375
26. AAGCTT	0.12821	0.12981	0.13232	0.13535
27. GAATTC	0.11618	0.12551	0.12417	0.11938
28. AAGGTTCC	0.12413	0.14373	0.12771	0.14218
29. GGTAAC	0.14013	0.13604	0.13836	0.13122
30. GTAC	0.12593	0.12215	0.12269	0.12521
31. GACT	0.12014	0.12337	0.12387	0.12407
32. GATC	0.12554	0.12695	0.12641	0.12679
33. CTCGAG	0.13326	0.13328	0.13470	0.13424
34. CCTGGT	0.13500	0.13424	0.13774	0.14029
35. GGATCC	0.12826	0.12721	0.13032	0.13200
36. GAGCCT	0.12875	0.13600	0.13521	0.13671
37. ACCGGT	0.11175	0.11420	0.10797	0.12795
38. GTCGAC	0.14140	0.13882	0.13849	0.14762
39. GCATGC	0.13421	0.13123	0.13018	0.14255

consistent with our expectations of a unimolecular melting transition. The average standard deviation of melting temperatures for all dumbbells did not vary significantly in different Na^+ environments. The average standard deviation of T_m for all dumbbells was 0.17, 0.18, 0.18 and 0.23 °C in 25, 55, 85 and 115 mM Na^+, respectively and taken to be 0.20 °C regardless of Na^+ environment. At least three independent melting experiments (and as many as eight in some cases) were conducted for each of the 22 dumbbell at the four Na^+ concentrations. The data in Table XV summarizes average melting data from 890 heating and cooling melting curves. The fraction of G•C type base pairs, *f*(G•C), in the duplex stem of each dumbbell is also given in Table XV.

Examination and comparison of the values in Table XV reveals a number of interesting aspects of the sequence dependent melting behavior of the dumbbells. For instance, Table XV reports in 115 mM Na^+ that the T_m's of dumbbells 22 and 27 of Figure 5 differ by ΔT_m = 87.70 - 84.72 = 2.98 °C, nearly 15 times more than the error in T_m. This is interesting because these dumbbells have the same length duplex stems with the same fraction of G•C base pairs. This observation implies that the differences in T_m values must be attributed to n-n and perhaps longer range sequence dependent interactions in the duplex stems of these dumbbells. A similar statement can be made for the difference between the T_m's of dumbbells 28 and 32 in 25 mM Na^+. Dumbbell 28 has a 20 base pair duplex stem while the duplex stem of dumbbell 32 is only 16 base pairs. For the stem sequences of both dumbbells *f*(G•C) is 0.5. Yet, the T_m of dumbbell 32 is higher than that of dumbbell 28 by 83.03 - 81.58 = 1.45 °C. One would expect the opposite effect, i.e. an increase in T_m for a longer duplex stem. Again this suggests the existence of significant n-n and possibly longer range effects in the sequences of these dumbbells.

Although the absolute values of T_m's decrease with decreasing Na^+ environment, the difference between the highest and lowest T_m is essentially independent of Na^+ concentration. In addition, with few exceptions, the relative stabilities of the dumbbells, i.e. order of melting is the same and did not change in the different Na^+ environments. The exceptions are dumbbells 32 and 33 with the central sequences, 5'-GATC-3' and 5'-CTCGAG-3', respectively. In 25 mM Na^+, dumbbell 32 has T_m higher than dumbbell 33 by $\Delta T_m = 0.34$ °C. However, in 115 mM Na^+, the T_m of dumbbell 32 is lower than that of dumbbell 33 by 0.36 °C.

5.10 The Sequence Independent Entropy of Base Pair Melting

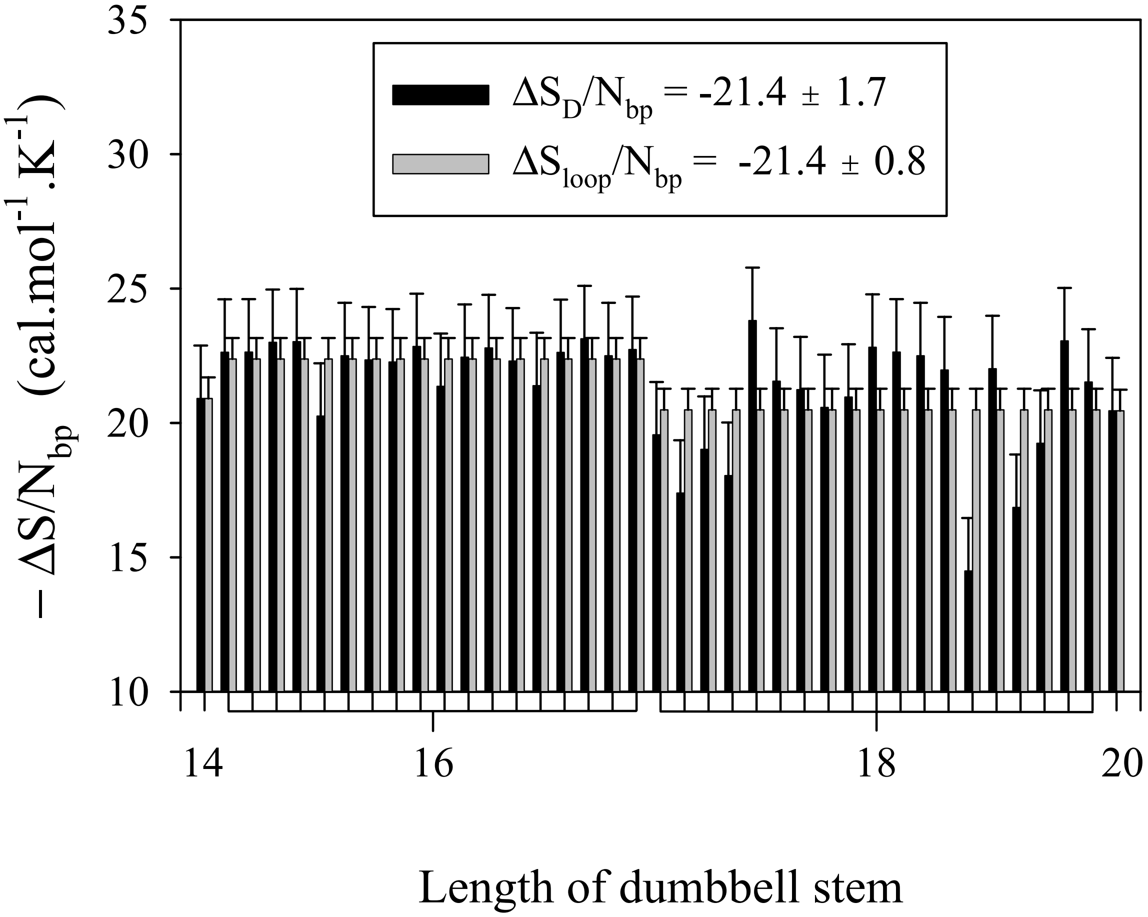
Values of the transition entropies for the 39 dumbbells were determined using equation (3.11) assuming a two-state melting transition. From the values of T_m in Table XV, and $\left[\frac{d\theta_B}{dT}\right]_{T=T_m}$ in Table XVI, the two-state transition entropy, $\Delta S_D(p)$, was evaluated for each dumbbell. The average $\left[\frac{d\theta_B}{dT}\right]_{T=T_m}$ determined from melting experiments of the 22 dumbbells, conducted in the four salt environments, are summarized in Table XVI. Values of the derivative curve peak heights were somewhat sensitive to the choice of linear baselines before and after the transition as described in Chapter 3.2. As a result, errors of $\left[\frac{d\theta_B}{dT}\right]_{T=T_m}$ are relatively large. An average standard deviation of $\left[\frac{d\theta_B}{dT}\right]_{T=T_m}$ is 0.01 at all Na^+ concentrations. Furthermore, in contrast to T_m values, the presence of small impurities (less than 10%) of nicked dumbbell in dumbbell samples could significantly influence the measured values of $\left[\frac{d\theta_B}{dT}\right]_{T=T_m}$. The same impurities only mildly affected the T_m (≤ 0.2 °C), while corresponding values of $\left[\frac{d\theta_B}{dT}\right]_{T=T_m}$ varied from 0.092 to 0.148. Errors of $\Delta S_D(p)$ were estimated using equation (3.13).

The average graphically determined values of $\Delta S_D(p)$ in 115 mM Na^+ for each dumbbell,

$p = 1, 39$ where then divided by the number of base pairs in the dumbbell stem and converted to average per base pair values, $\Delta S_D(p)/N_{bp}(p)$. For every dumbbell, these values were nearly equivalent. Assuming an error of 0.01 for $\left[\frac{d\theta_B}{dT}\right]_{T=T_m}$ and 0.2 °C for T_m, the average value over all $\Delta S_D(p)/N_{bp}(p)$ ($p = 1, 39$) values was $\Delta S_D/N_{bp} = -21.37 \pm 1.70$ cal.mol^{-1}.K^{-1}.

Values of the two-state entropies $\Delta S_D(p)$ in 115 mM Na$^+$ were also fit using the sequence independent functional form in equation (5.28) varying only the weighting parameter $L(N_p)$ to fit the values. The values of $L(N_p) = 2.225 \times 10^{-62}$, 1.372×10^{-76}, 8.783×10^{-79}, 1.227×10^{-87} for $N_p = 36, 40, 44, 48$ provided an acceptable fit with $Q = 0.086$, $\chi^2 = 45.7$. The average value of the loop entropy per base pair obtained from fitting with the loop entropy function was $\Delta S_{loop}/N_{bp} = -21.4 \pm 0.8$ cal.mol^{-1}.K^{-1}. This error was estimated from the average of errors associated with the $L(N_p)$ values for $N_p = 36, 40, 44, 48$. Results of the analysis of these data are shown in Figure 7, where the entropies of formation per base pair determined from the two-state analysis of melting curve parameters, $\Delta S_D/N_{bp}$, and that determined by fitting the data with the sequence independent loop entropy function dependent only on the number of nucleotides comprising the dumbbells, $\Delta S_{loop}/N_{bp}$, are compared. In Figure 7, values of the entropy of base pair formation, $\Delta S/N_{bp} = \Delta S_D/N_{bp}$ (dark bars) or $\Delta S_{loop}/N_{bp}$ (light bars) are plotted in histogram form versus dumbbell stem length. The averages of these values are in exact agreement, $\Delta S/N_{bp} = -21.37$ cal.mol^{-1}.K^{-1}. For a few of the molecules, dumbbells 19, 21, 22, 23 with central sequences, 5'-TTTAAA-3', 5'-CTTAAG-3, 5'-TCTAGA-3', 5'-TTCAAG-3', agreement is not within the error. This probably arises because these sequences cause the dumbbells to deviate substantially from two-state melting behavior, or there is a length dependence of the melting entropy of these sequences.

Figure 7. Entropies of annealing per base pair (dark bars), $\Delta S_D/N_{bp}$, determined from T_m's and peak heights of derivative melting curves are displayed for different lengths of the dumbbell stem. Loop entropies per base pair (light bars), $\Delta S_{loop}/N_{bp}$, were obtained from fits with the loop function.


35
30
25
20
15
10
− ΔS/N_{bp} (cal.mol^{-1}.K^{-1})
$\Delta S_D/N_{bp}$ = -21.4 ± 1.7
$\Delta S_{loop}/N_{bp}$ = -21.4 ± 0.8
14
16
18
20
Length of dumbbell stem

In spite of these deviations, the “reasonable” fit obtained with a constant entropy suggests that the entropies per base pair, ΔS_{bp}, are constant within the error of measurement. Admittedly, errors in transition enthalpies and entropies determined from $\left[\frac{d\theta_B}{dT}\right]_{T=T_m}$ are relatively large. Therefore, it is not possible to ascertain whether there exists subtle sequence dependent entropic differences smaller than the errors. Within the errors the entropy of melting per base pair appears to be constant. The apparently sequence independent entropy value evaluated from the two-state analysis is -21.37 ± 1.70 $cal.mol^{-1}.K^{-1}$. This is slightly smaller than the value used in determination of n-n and n-n-n interactions, ΔS_{bp} = -24.85 ± 1.74 $cal.mol^{-1}.K^{-1}$ as evaluated from optical and calorimetric measurements of long restriction fragments (Delcourt and Blake, 1991). To justify this result based on the two-state analysis we explicitly measured the melting entropy of three dumbbells of the database (dumbbells 25, 33, 38 of Figure 5) that all have the same length (18 base pairs) duplex stem (data not shown).

For dumbbells 25, 33 and 38, the calorimetrically measured transition entropies were -424.4, -421.1 and -470.0 $cal.mol^{-1}.K^{-1}$, respectively and agreed within 10%. The dumbbells have N_{bp} = 18 base pairs in their duplex stems, thus the average entropy per base pair for all three molecules is $\Delta S_C/N_{bp}$ = -24.36 ± 1.24 $cal.mol^{-1}.K^{-1}$. This value is consistent with the value employed in fitting and that obtained from melting studies of long DNA restriction fragments, and again larger by approximately 3 $cal.mol^{-1}.K^{-1}$ then the value obtained assuming a two-state analysis. Since DSC measures the entropy directly with higher accuracy, without dependence on the two-state assumption, we used the consensus value of ΔS_{bp} = -24.85 $cal.mol^{-1}.K^{-1}$ to evaluate n-n and n-n-n parameters.

The influence of the magnitude of the constant ΔS_{bp} value on the evaluation of n-n and n-n-n parameters was also investigated. Using the smaller value determined from two-state analysis and fitting with ΔS_{bp} = -21.37 cal.mol^{-1}.K^{-1}, instead of -24.85 cal.mol^{-1}.K^{-1}, fits of the singlet and n-n parameters were repeated. As expected, the singlet enthalpies changed accordingly, but the relative orders and magnitudes of n-n and n-n-n interactions were unchanged (not shown). Similar analyses were also performed in the lower Na^{+} environments using the melting data of the 22 new dumbbells (18-39 of Figure 5) and results supported the assumption of a constant entropy ΔS_{bp} there as well.

The entropy value we employ (-24.85 cal.mol^{-1}.K^{-1}) is very close to that obtained from early theoretical estimates (-26.2 cal.mol^{-1}.K^{-1}) based on purely configurational arguments (DeVoe and Tinoco, 1962). As the above evidence suggests ΔS_{bp} appears to be independent of sequence in the melting of DNA dumbbells, just as found for long DNAs. In fact, the only melting studies that reported a sequence dependence to the base pair melting entropy are those of short DNA and RNA duplex oligomers and RNA/DNA hybrid duplexes, where a very definite sequence dependence to the entropy of base pair melting was found (Allawi and SantaLucia, 1997; Breslauer et al., 1986; Sugimoto et al., 1996). The origins of this discrepancy are not presently known but probably arise from the effects of sequence dependent nucleation in short linear duplex DNAs.

5.11 Evaluation of the Singlet Interactions

The plot of the melting temperatures of dumbbells vs. fraction of G•C base pairs in stem, *f*(G•C), in the four Na^{+} environments is shown on Figure 8. The least-squares fitted lines for experiments conducted in each melting buffer are also depicted. The average melting

Figure 8. Melting temperatures of 39 dumbbells plotted versus the fraction of G•C base pairs in the dumbbell stems. Plotted melting temperatures were determined in the following Na^+ concentrations, 25 mM (bottom solid line), 55 mM (second dash line from bottom), 85 mM (second dash-dot line from top), 115 mM (top dash-dot-dot line). The straight lines are results of least squares fits to all T_m's at each Na^+ concentration.

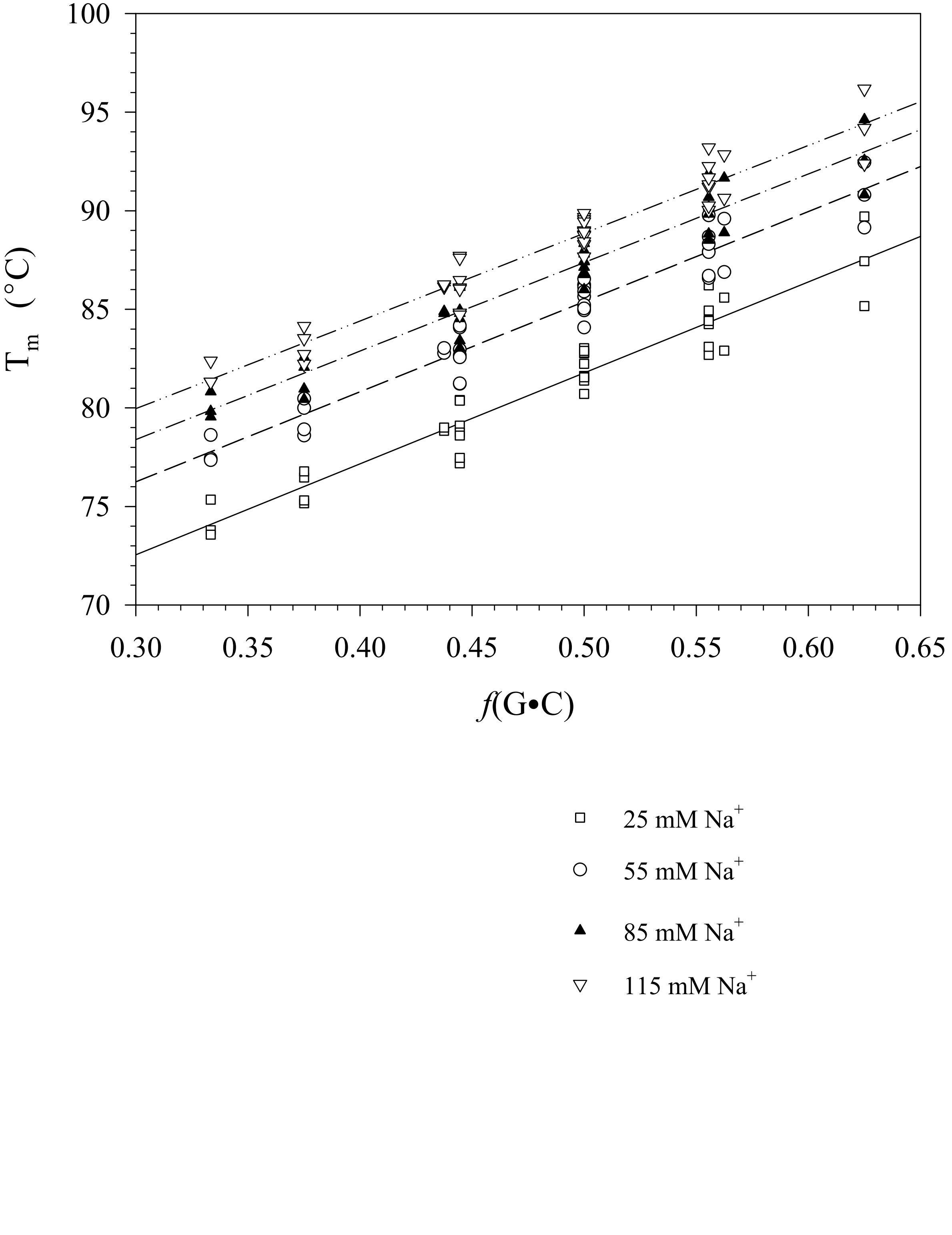

100
95
90
85
80
75
70
T_m (°C)
0.30
0.35
0.40
0.45
0.50
0.55
0.60
0.65
f(G•C)
□ 25 mM Na^+
○ 55 mM Na^+
▲ 85 mM Na^+
▽ 115 mM Na^+

TABLE XVII Values of $T_{A•T}$, $T_{G•C}$ and Singlet Enthalpies in 25, 55, 85 and 115 mM Na^+. Results of least-squares fits shown on Figure 8 are also reported.

$[Na^+]$	25 mM	55 mM	85 mM	115 mM
Equation of Fitted Straight Line	T_m (°C) = 46.125·f(G•C) + 58.714	T_m (°C) = 45.707·f(G•C) + 62.544	T_m (°C) = 44.891·f(G•C) + 64.932	T_m (°C) = 44.503·f(G•C) + 66.615
$T_{A•T}$ (°C)	58.71	62.54	64.93	66.62
$T_{G•C}$ (°C)	104.84	108.25	109.82	111.12
$\Delta H_{A•T}$ (cal.mol^{-1})	-8246.83	-8342.00	-8401.35	-8443.17
$\Delta H_{G•C}$ (cal.mol^{-1})	-9393.03	-9477.83	-9516.90	-9549.07
r^2	0.92	0.94	0.93	0.93
χ^2	1110.92	873.60	955.95	835.59
Q	4.3 x 10^{-209}	2.2 x 10^{-159}	1.4 x 10^{-176}	1.8 x 10^{-151}

temperatures of A•T and G•C base pairs in the dumbbells, $T_{A\bullet T}$ and $T_{G\bullet C}$, were determined from these fits. $T_{A\bullet T}$ is the average melting temperature of a dumbbell when *f*(G•C) approaches zero. Similarly, $T_{G\bullet C}$ is the melting temperature read from the fitted straight line when *f*(G•C) equals one.

Results of singlet fitting are summarized in Table XVII where the average enthalpies of formation for A•T and G•C base pairs in dumbbells, as well as equations of the fitted straight lines in each Na^+ environment shown in Figure 8 are given. The poor fits with the singlet model defined by the absurdly low values of Q and linear correlation coefficients, r^2, indicate that the two singlet interactions, $\Delta H_{A\bullet T}$, $\Delta H_{G\bullet C}$, are unable to account for the observed sequence dependent stability of the dumbbells.

5.12 Evaluation of the Nearest-Neighbor (Doublet) Sequence Dependent Interactions

Values of the ten non-unique n-n dependent deviations from the singlet transition enthalpy, $\delta H_{ij}^{n\text{-}n}$, are listed in Table XVIII. Also given in Table XVIII are values for the nine unique linear combinations (see Table XII) of the n-n dependent deviations from the singlet enthalpy, $\delta H_c^{n\text{-}n}$. It should be noted that δH_{ij}^{n-n} and $\delta H_c^{n\text{-}n}$ are reported for duplex formation. The same values with opposite sign should be used to calculate melting enthalpies for dumbbell denaturation. Examination of Tables XII and XVIII reveals that the $\delta H_c^{n\text{-}n}$ values corresponding to the first six linear combinations (c = 1, 6) $\delta H_1^{n\text{-}n}$, $\delta H_2^{n\text{-}n}$, $\delta H_3^{n\text{-}n}$, $\delta H_4^{n\text{-}n}$, $\delta H_5^{n\text{-}n}$, $\delta H_6^{n\text{-}n}$ containing mixtures of the doublet repeats, and the ninth combination $\delta H_9^{n\text{-}n}$, corresponding to the unique ends of the dumbbells in the database, change significantly when the ionic strength is increased from 25 mM to 55 mM Na^+. The 85 mM values are different again but similar to the

TABLE XVIII Enthalpies, δH_{ij}^{n-n}, of 10 Non-unique Nearest-Neighbor Stacking Interactions and Enthalpies, $\delta H_c^{n\text{-}n}$, of 9 Unique Linear Combinations of Nearest-Neighbors as a Function of Ionic Strength. Enthalpies are in cal.mol^{-1}. Standard deviations of values of unique linear combinations, $\sigma(\delta H_c^{n\text{-}n})$, are reported in the rightmost column.

δH_{ij}^{n-n}	25 mM	55 mM	85 mM	115 mM
$\delta H_{AT/AT}^{n-n}$	-39.3	-56.5	-76.4	-80.2
$\delta H_{TA/TA}^{n-n}$	124.0	105.2	92.0	75.2
$\delta H_{AA/TT}^{n-n}$	-97.1	-116.1	-155.5	-155.6
$\delta H_{AC/GT}^{n-n}$	-338.5	-125.0	-3.2	50.3
$\delta H_{CA/TG}^{n-n}$	210.8	-58.9	-214.8	-240.0
$\delta H_{GA/TC}^{n-n}$	-366.2	-163.9	-41.1	-0.4
$\delta H_{AG/CT}^{n-n}$	509.6	225.6	84.4	20.1
$\delta H_{CG/CG}^{n-n}$	567.6	81.1	-214.6	-313.1
$\delta H_{GC/GC}^{n-n}$	-882.6	-408.6	-188.6	-95.4
$\delta H_{GG/CC}^{n-n}$	261.8	206.8	188.9	172.0

$\delta H_c^{n\text{-}n}$	25 mM	55 mM	85 mM	115 mM	$\sigma(\delta H_c^{n\text{-}n})$
$\delta H_1^{n\text{-}n}$	-97.1	-116.1	-155.5	-155.6	16.0
$\delta H_2^{n\text{-}n}$	261.8	206.8	188.9	172.0	16.6
$\delta H_3^{n\text{-}n}$	42.3	24.3	7.8	-2.5	19.0
$\delta H_4^{n\text{-}n}$	-157.5	-163.7	-201.6	-204.3	17.8
$\delta H_5^{n\text{-}n}$	-63.8	-91.9	-109.0	-94.8	14.2
$\delta H_6^{n\text{-}n}$	71.7	30.9	21.7	9.9	13.7
$\delta H_7^{n\text{-}n}$	-38.7	-37.6	-37.1	-34.5	5.0
$\delta H_8^{n\text{-}n}$	-42.9	-43.3	-47.1	-43.2	4.9
$\delta H_9^{n\text{-}n}$	360.5	119.6	-11.5	-58.8	101.0

TABLE XIX Enthalpies, δH_{ijk}^{n-n-n}, of 32 Non-unique Next-Nearest-Neighbor (Triplet) Interactions and Enthalpies, δH_c^{n-n-n}, of 27 Unique Linear Combinations of Next-Nearest-Neighbors in 25 and 55 mM Na^+. Enthalpies are in cal.mol^{-1}. The $\sigma(\delta H_c^{n-n-n})$ denotes the standard deviations of enthalpic values of unique linear combinations.

δH_{ijk}^{n-n-n}	25 mM	55 mM	δH_{ijk}^{n-n-n}	25 mM	55 mM
$\delta H_{AAA/TTT}^{n-n-n}$	62.19	31.27	$\delta H_{GAC/GTC}^{n-n-n}$	-167.34	-188.53
$\delta H_{AAG/CTT}^{n-n-n}$	71.24	64.29	$\delta H_{GGA/TCC}^{n-n-n}$	14.64	-4.09
$\delta H_{AAT/ATT}^{n-n-n}$	67.71	58.81	$\delta H_{GGG/CCC}^{n-n-n}$	55.19	29.27
$\delta H_{AAC/GTT}^{n-n-n}$	-101.65	-85.97	$\delta H_{GGC/GCC}^{n-n-n}$	-117.68	-165.15
$\delta H_{AGA/TCT}^{n-n-n}$	-0.31	76.43	$\delta H_{GTA/TAC}^{n-n-n}$	-184.12	-208.76
$\delta H_{AGG/CCT}^{n-n-n}$	26.26	33.86	$\delta H_{GTG/CAC}^{n-n-n}$	-92.41	-135.12
$\delta H_{AGT/ACT}^{n-n-n}$	205.37	152.54	$\delta H_{GCA/TGC}^{n-n-n}$	13.57	-39.60
$\delta H_{AGC/GCT}^{n-n-n}$	18.12	-61.83	$\delta H_{GCG/CGC}^{n-n-n}$	21.76	70.93
$\delta H_{ATA/TAT}^{n-n-n}$	41.88	63.17	$\delta H_{TAA/TTA}^{n-n-n}$	17.72	12.30
$\delta H_{ATG/CAT}^{n-n-n}$	153.50	79.82	$\delta H_{TAG/CTA}^{n-n-n}$	-32.79	-56.47
$\delta H_{ATC/GAT}^{n-n-n}$	20.16	57.40	$\delta H_{TGA/TCA}^{n-n-n}$	30.30	-80.81
$\delta H_{ACA/TGT}^{n-n-n}$	126.48	187.18	$\delta H_{TGG/CCA}^{n-n-n}$	-38.56	-11.63
$\delta H_{ACG/CGT}^{n-n-n}$	315.60	295.04	$\delta H_{TTG/CAA}^{n-n-n}$	-25.31	-4.06
$\delta H_{ACC/GGT}^{n-n-n}$	112.66	195.91	$\delta H_{TCG/CGA}^{n-n-n}$	-31.93	-15.11
$\delta H_{GAA/TTC}^{n-n-n}$	44.89	28.88	$\delta H_{CAG/CTG}^{n-n-n}$	96.01	114.50
$\delta H_{GAG/CTC}^{n-n-n}$	114.98	78.67	$\delta H_{CGG/CCG}^{n-n-n}$	21.92	4.44

TABLE XIX (continued)

$\delta H_c^{n\text{-}n\text{-}n}$	25 mM	55 mM	$\sigma(\delta H_c^{n\text{-}n\text{-}n})$
$\delta H_1^{n\text{-}n\text{-}n}$	62.2	31.3	43.2
$\delta H_2^{n\text{-}n\text{-}n}$	41.9	63.2	46.7
$\delta H_3^{n\text{-}n\text{-}n}$	55.2	29.3	25.8
$\delta H_4^{n\text{-}n\text{-}n}$	21.8	70.9	29.6
$\delta H_5^{n\text{-}n\text{-}n}$	17.0	26.0	24.1
$\delta H_6^{n\text{-}n\text{-}n}$	57.3	77.5	28.7
$\delta H_7^{n\text{-}n\text{-}n}$	42.7	35.6	24.0
$\delta H_8^{n\text{-}n\text{-}n}$	-47.9	-80.4	63.2
$\delta H_9^{n\text{-}n\text{-}n}$	-13.2	23.6	23.2
$\delta H_{10}^{n\text{-}n\text{-}n}$	39.3	30.7	22.5
$\delta H_{11}^{n\text{-}n\text{-}n}$	74.6	31.6	21.0
$\delta H_{12}^{n\text{-}n\text{-}n}$	-6.6	-19.0	25.4
$\delta H_{13}^{n\text{-}n\text{-}n}$	41.1	-0.6	25.7
$\delta H_{14}^{n\text{-}n\text{-}n}$	-19.8	-13.8	21.3
$\delta H_{15}^{n\text{-}n\text{-}n}$	-9.3	1.8	13.6
$\delta H_{16}^{n\text{-}n\text{-}n}$	25.4	-21.8	26.4
$\delta H_{17}^{n\text{-}n\text{-}n}$	-3.1	-2.2	10.9
$\delta H_{18}^{n\text{-}n\text{-}n}$	1.0	-4.7	9.4
$\delta H_{19}^{n\text{-}n\text{-}n}$	8.9	-11.5	15.1
$\delta H_{20}^{n\text{-}n\text{-}n}$	0.7	40.1	21.4
$\delta H_{21}^{n\text{-}n\text{-}n}$	-25.2	-10.9	15.2
$\delta H_{22}^{n\text{-}n\text{-}n}$	23.9	17.1	15.5
$\delta H_{23}^{n\text{-}n\text{-}n}$	19.0	17.8	19.1
$\delta H_{24}^{n\text{-}n\text{-}n}$	-20.6	-31.1	22.0
$\delta H_{25}^{n\text{-}n\text{-}n}$	-12.4	-9.1	13.3
$\delta H_{26}^{n\text{-}n\text{-}n}$	-1.8	0.3	6.4
$\delta H_{27}^{n\text{-}n\text{-}n}$	13.6	15.1	9.9

TABLE XX Goodness-of-fit Described by Q and χ^2 as a Function of [Na^+].

[Na^+]	Fit to Nearest-Neighbors (doublets)		Fit to Next-Nearest-Neighbors (triplets)	
	χ^2	Q	χ^2	Q
25 mM	76.4	6.41×10^{-6}	25.1	0.0144
55 mM	42.3	0.0671	7.7	0.8065
85 mM	35.5	0.2264	7.3	0.8352
115 mM	36.7	0.1873	6.7	0.8783

115 mM values. The values titrate with solvent ionic strength and decrease (become more stable) as the salt is increased. The two remaining combinations, $\delta H_7^{n\text{-}n}$ and $\delta H_8^{n\text{-}n}$, that contain mixtures of the n-n doublet repeats, remain relatively constant at all ionic strengths. The standard deviations in the reported values, $\sigma(\delta H_c^{n\text{-}n})$, are given in the far right column and determine the statistical relevance of differences in Table XVIII. If the changes in values for a particular combination with ionic strength are greater than the corresponding $\sigma(\delta H_c^{n\text{-}n})$, then the changes are deemed to be statistically significant.

Comparison of values of unique linear combinations in Table XVIII with those obtained from our original database containing only the first 17 dumbbells of Figure 5 (Doktycz et al., 1992) (not shown) reveals the thermodynamic values of all n-n linear combinations, $\delta H_c^{n\text{-}n}$, are in agreement within the errors found previously. Augmentation of the original database of 17 molecules with the additional 22 new dumbbells of this study did not change the resulting n-n values significantly. Furthermore, the standard deviations, $\sigma(\delta H_c^{n\text{-}n})$, for the new set of values in Table XVIII are 25-50% smaller. An exception is the standard deviation of the ninth end interaction, $\delta H_9^{n\text{-}n}$, perhaps because it represents the end interactions specific only for the dumbbells with the same ends as these studied here. This result also confirms that considering the entire duplex stem sequence or the central unique core sequence yields the same result.

Statistical parameters that determine the quality of fit in each Na^+ environment are shown on the left side of Table XX. The Q values for the n-n (doublet) fits reveal excellent fits are obtained in 85 mM (Q = 0.22) and 115 mM Na^+ (Q = 0.19), indicating the n-n model is sufficient to describe (within the error of measurements) the melting thermodynamics of the dumbbells. The values of Q at lower sodium ion concentrations show limitations of the nearest-neighbor model. The marginal Q value at 55 mM Na^+ (0.07) and the unacceptable Q value at 25 mM (6.4×10^{-6}) suggest the n-n model is inadequate to explain sequence dependent melting behavior in these salt environments, and that interactions beyond nearest-neighbors should be considered for DNA duplex solutions with [Na^+] less than 55 mM.

This analysis indicates the n-n approximation is able to account accurately for sequence dependent stability of DNA in 85 and 115 mM Na^+. However, at 25 and 55 mM Na^+ the n-n model was unable to account for the measured free energies within experimental error, suggesting that longer range sequence dependent interactions may be significant at the lower salt concentrations.

5.13 Evaluation of the Next-Nearest-Neighbor (Triplet) Interactions

Thirty-two non-unique values of deviations of transition enthalpies from the average of singlet and doublet enthalpies, δH_{ijk}^{n-n-n}, and the 27 unique linear combinations, $\delta H_c^{n\text{-}n\text{-}n}$, for n-n-n interactions in 25 and 55 mM Na^+ are listed in Table XIX. For each of the 27 linear combinations of the n-n-n triplets the corresponding standard deviation $\sigma(\delta H_c^{n\text{-}n\text{-}n})$ is given in the right most column. If the magnitude of a particular $\delta H_c^{n\text{-}n\text{-}n}$ is larger than its corresponding $\sigma(\delta H_c^{n\text{-}n\text{-}n})$ then that linear combination is deemed to be significant. In 25 mM Na^+ 10 of the 27 unique linear combinations of n-n-n base pair triplets are larger than their standard deviations. These are combinations, $\delta H_1^{n\text{-}n\text{-}n}$, $\delta H_3^{n\text{-}n\text{-}n}$, $\delta H_6^{n\text{-}n\text{-}n}$, $\delta H_7^{n\text{-}n\text{-}n}$, $\delta H_{10}^{n\text{-}n\text{-}n}$, $\delta H_{11}^{n\text{-}n\text{-}n}$, $\delta H_{13}^{n\text{-}n\text{-}n}$, $\delta H_{21}^{n\text{-}n\text{-}n}$, $\delta H_{22}^{n\text{-}n\text{-}n}$, $\delta H_{27}^{n\text{-}n\text{-}n}$. As Table XX reveals, if the melting data are fit with the n-n-n model, the fits in low salt are considerably improved with ($Q > 0.01$) in 25 mM Na^+ and $Q = 0.81$ in 55 mM Na^+. Examination of the triplet sequence combinations in Table XIX and their definitions in

Table XIII reveals the largest $\delta H_c^{n\text{-}n\text{-}n}$ values are for combinations of the triplet sequences, AAA/TTT, GGG/CCC, AGA/TCT, GAG/CTC, GTA/TAC, ATG/CAT, GCA/TGC, ACG/CGT, TAA/TTA and AAT/ATT.

Throughout this analysis in deriving the n-n and n-n-n values rigorous statistical criteria with careful consideration of the errors involved has been maintained. Thus, although the magnitudes of the resulting $\delta H_c^{n\text{-}n\text{-}n}$ values in Table XIX are relatively small, they were obtained using robust statistical criteria and consideration of errors and therefore deemed to be significant. Recall, these are only the deviations from the average of singlet and doublet parameters and would be expected to be relatively small in comparison.

5.14 An Independent Test of the Evaluated Parameters

Invoking the n-n-n model improves fits of the melting data in 25 and 55 mM Na^+ of the 39 dumbbells in database. We wished to determine how well our n-n-n model parameters could predict the behavior of dumbbells with identical ends and end-loops but different stem sequences that were not part of the database. For this test melting data were collected in 25 mM Na^+ for two independent dumbbell molecules. These DNA dumbbells (I and II) have slightly longer stem sequences (34 base pairs) than those in the database used to evaluate sequence dependent parameters. Dumbbells I and II have the duplex stem sequences shown in Table XXI, linked on the ends by T_4 loops. The melting temperatures obtained from melting experiments and those predicted using the n-n and n-n-n parameters are also given. The experimentally measured T_m's of dumbbells I and II in 25 mM Na^+ are compared with T_m's calculated using the n-n parameters in Table XVIII or the n-n-n parameters in Table XIX. For example, consider dumbbell I with

TABLE XXI Predicted and Measured Melting Temperatures of Two DNA Dumbbells which Were Not Employed in Evaluation of Nearest-Neighbor and Next-Nearest-Neighbor Thermodynamic Parameters. Sequences of the dumbbell stems are given. Stem sequences are linked by T_4 end-loops.

	Sequence of dumbbell stem (5' to 3')	Measured T_m (°C)	Predicted T_m (°C) (n-n set)	Predicted T_m (°C) (n-n-n set)
I.	GTA TAG GAT CCA CTG TGT TAC AAG GAT CCA ATA C	75.05	77.49	77.31
II.	GTA TAG GAT CCA AGA GAG AGA GAG GAT CCA ATA C	74.90	77.78	77.19

the stem sequence 5'-GTATAGGATCCACTGTGTTACAAGGATCCAATAC-3' and calculate the enthalpy of annealing reaction in 25 mM Na$^+$. As shown in equations (5.6), (5.7) and (5.10), the singlet enthalpy is determined from the numbers of A•T and G•C base pairs ($N_{A\bullet T}$, $N_{G\bullet C}$) and singlet base pair enthalpies in Table XVII,

$$\begin{aligned}\Delta H(\mathrm{I}, \mathit{singlet}) &= N_{\mathrm{A\bullet T}}\cdot H_{\mathrm{A\bullet T}} + N_{\mathrm{G\bullet C}}\cdot H_{\mathrm{G\bullet C}} = \\ &= 20\cdot(-8246.83) + 14\cdot(-9393.03) = -296{,}439.02\ \text{cal.mol}^{-1}\end{aligned}$$

The doublet enthalpy is calculated using the n-n values. The numbers of n-n doublets that are present in the sequence are $N_{AT/AT} = 4$, $N_{TA/TA} = 4$, $N_{AA/TT} = 3$, $N_{AC/GT} = 6$, $N_{CA/TG} = 5$, $N_{TC/GA} = 4$, $N_{CT/AG} = 3$, $N_{CG/CG} = 0$, $N_{GC/GC} = 0$, and $N_{GG/CC} = 4$. Considering these values and the n-n parameters reported in Table XVIII, the doublet enthalpy is then,

$$\begin{aligned}\Delta H(\mathrm{I}, \mathit{doublet}) &= \sum_{i,j=\mathrm{A,T,C,G}} N_{ij}(\mathrm{I})\cdot\delta H_{ij}^{n-n} = 4\cdot(-39.3) + 4\cdot 124.0 + 3\cdot(-97.1) + \\ &+ 6\cdot(-338.5) + 5\cdot(210.8) + 4\cdot(-366.2) + 3\cdot(509.6) + 4\cdot(261.8) = 181.7\ \text{cal.mol}^{-1}\end{aligned}$$

Twenty five triplets reside in the duplex stem of dumbbell I. The numbers of different triplets are $N_{AAA/TTT} = 0$, $N_{AAG/CTT} = 1$, $N_{AAT/ATT} = 1$, $N_{AAC/GTT} = 1$, $N_{AGA/TCT} = 0$, $N_{AGG/CCT} = 2$, $N_{AGT/ACT} = 1$, $N_{AGC/GCT} = 0$, $N_{ATA/TAT} = 3$, $N_{ATG/CAT} = 0$, $N_{ATC/GAT} = 4$, $N_{ACA/TGT} = 3$, $N_{ACG/CGT} = 0$, $N_{ACC/GGT} = 0$, $N_{GAA/TTC} = 0$, $N_{GAG/CTC} = 0$, $N_{GAC/GTC} = 0$, $N_{GGA/TCC} = 4$, $N_{GGG/CCC} = 0$, $N_{GGC/GCC} = 0$, $N_{GTA/TAC} = 3$, $N_{GTG/CAC} = 2$, $N_{GCA/TGC} = 0$, $N_{GCG/CGC} = 0$, $N_{TAA/TTA} = 1$, $N_{TAG/CTA} = 1$, $N_{TGA/TCA} = 0$, $N_{TGG/CCA} = 2$, $N_{TTG/CAA} = 2$, $N_{TCG/CGA} = 0$, $N_{CAG/CTG} = 1$, and $N_{CGG/CCG} = 0$. When these values are combined with n-n-n parameters (shown in Table XIX), the triplet enthalpy is calculated,

$$\Delta H(\mathrm{I}, \mathit{triplet}) = \sum_{i,j,k=\mathrm{A,T,C,G}} N_{ijk}(\mathrm{I})\cdot\delta H_{ijk}^{n-n-n} = 1\cdot(71.24) + 1\cdot(67.71) + 1\cdot(-101.65) +$$

$$+ 2\cdot(26.26) + 1\cdot(205.37) + 3\cdot(41.88) + 4\cdot(20.16) + 3\cdot(126.48) +$$

$$+ 4\cdot(14.64) + 3\cdot(-184.12) + 2\cdot(-92.41) + 1\cdot(17.72) + 1\cdot(-32.79) +$$

$$+ 2\cdot(-38.56) + 2\cdot(-25.31) + 1\cdot(96.01) = 155.49 \ \text{cal.mol}^{-1}$$

The total enthalpy of forming the fully intact dumbbell I is the sum of singlet, doublet and triplet enthalpies,

$$\Delta H_D(\mathrm{I}) = \Delta H(\mathrm{I}, \mathit{singlet}) + \Delta H(\mathrm{I}, \mathit{doublet}) + \Delta H(\mathrm{I}, \mathit{triplet})$$

$$= -296{,}439.02 + 181.7 + 155.49 = -296{,}101.83 \ \text{cal.mol}^{-1}$$

Because the entropy per base is considered to be constant,

$\Delta S_D = N_{\mathrm{bp}}\cdot\Delta S_{\mathrm{bp}} = 34\cdot(-24.85) = -844.9$ cal.mol^{-1}.K^{-1}. The predicted melting temperature is obtained by rearranging equation (3.11),

$$T_m = \frac{\Delta H_D}{\Delta S_D} = \frac{-296{,}101.83}{-844.9} = 350.46 \ \mathrm{K} = 77.31 \ ^\circ\mathrm{C}$$

Using the n-n parameters, the predicted T_m's are 2.4 and 2.9 °C higher than the experimental T_m's. The n-n-n prediction is slightly better with T_m's 2.2 and 2.3 °C higher than the experiments. Obviously, these values are extremely sensitive to the value of ΔS_D used in calculation. With only slight increases (using 25.00 instead of 24.85) the calculated T_m's are in agreement (within error) with the experimentally determined values. This comparison serves to indicate that the n-n-n model improvement of predictions of the melting temperatures for the two molecules reported in Table XXI, is minor but in the right direction. The prediction of the T_m of dumbbell II is affected more by the n-n-n parameters than that of dumbbell I. This is because the stem sequence of dumbbell II contains a string of AG/CT n-n doublets in the center of the dumbbell stem. This sequence is comprised of the AGA and GAG n-n-n triplet sequences. As Table XIII shows for these sequences, $\delta H_6^{n\text{-}n}$ is the average of the AGA/TCT and GAG/CTC triplet interactions, and has significant magnitude. Consequently the n-n-n interactions affect the predicted stability of dumbbell II more than dumbbell I. With an error in T_m of 0.2 °C, the improvement of the T_m prediction is only significant for dumbbell II.

5.15 Predictions of Sequence Dependent Thermodynamics of Linear DNA Duplexes

In order to use the n-n parameters evaluated from dumbbell melting analysis to predict the behavior of short linear duplex oligomers having solvent exposed ends, a significant correction that accounts for the nucleation enthalpy is required. Using a nucleation enthalpy, ΔH_{nuc} in conjunction with the n-n parameters evaluated here improves significantly predictions of melting temperatures of linear DNAs. This nucleation enthalpy correction as originally published (Owczarzy et al., 1997) and reported in Chapter 4 (equation 4.16) was derived using our earlier n-n set evaluated from 17 dumbbells in 115 mM Na^+ (Doktycz et al., 1992). A slightly different correction results if the newest set of n-n parameters evaluated from the set of 39 dumbbells is employed.

To evaluate this correction the available data on 251 linear DNA duplexes was used (see page 55 for description of database B). Reported melting temperatures of molecules correspond to a total single strand concentration of 4 μM in 1 M Na^+ solvent. A correction factor for scaling

T_m in 115 mM and 1.0 M Na^+ was also reported earlier (equation 4.15). Using this factor the melting temperatures were scaled from 1.0 M to 115 mM Na^+. These experimentally determined melting temperatures were compared with the predicted melting temperatures using the newer n-n parameters in Table XVIII. From the difference of the predicted and measured melting temperatures, $\Delta T_m(r)$, the nucleation enthalpy was calculated for each duplex r according to,

$$\Delta H_{nuc}(r) = \Delta T_m(r) \cdot \left(\Delta S(r) + R \cdot \ln\left[\frac{4x10^{-6}}{\alpha}\right]\right) \tag{5.30}$$

where $\alpha = 1$ for self-complementary duplexes and $\alpha = 4$ for non-self-complementary duplexes. For an $N_{bp}(r)$ base pair oligomeric duplex, the entropy $\Delta S(r)$ was calculated as $N_{bp}(r) \cdot \Delta S_{bp}$. The nucleation enthalpies of all duplexes (r = 1 to 251) were least-squares fit to three parameters, H_1, H_2 and H_2, viz.

$$\begin{bmatrix} 1 & f(G\bullet C)(1) & N_{bp}(1) \\ 1 & f(G\bullet C)(2) & N_{bp}(2) \\ \ldots & \ldots & \ldots \\ 1 & f(G\bullet C)(251) & N_{bp}(251) \end{bmatrix} * \begin{bmatrix} H_1 \\ H_2 \\ H_3 \end{bmatrix} = \begin{bmatrix} \Delta H_{nuc}(1) \\ \Delta H_{nuc}(2) \\ \ldots \\ \Delta H_{nuc}(251) \end{bmatrix} \tag{5.31}$$

It was assumed that H_{nuc} depends linearly on the fraction of G•C base pairs, f(G•C), and on the number of base pairs in the duplex, N_{bp}. The solution of equation (5.31) was obtained by SVD by minimizing the difference between the right and left sides. The newer nucleation enthalpy of duplex formation that should be used in conjunction with new n-n parameters is,

$$\Delta H_{nuc} = 7654.71 - 3469.93 \cdot f(\mathrm{G{\bullet}C}) - 186.51 \cdot N_{\mathrm{bp}} \quad \mathrm{cal.mol}^{-1} \tag{5.32}$$

Values of the numerical coefficients in equation (5.32) are very similar to those found earlier (see equation 4.16, page 96). Evidently, the correction factor for the nucleation enthalpy evaluated for the new n-n parameters is not significantly different from that recommended for use in conjunction with our earlier n-n set.

To demonstrate use of the n-n set and nucleation enthalpy, consider the duplex oligomer with the sequence 5'-ATT ATG GGG C-3'. It is assumed the transition enthalpy in 115 mM Na^+ consists of only singlet and doublet components, and the nucleation enthalpy correction,

$$\Delta H_D = \Delta H(singlet) + \Delta H(doublet) + \Delta H_{nuc} \tag{5.33}$$

The average melting temperatures of A•T or G•C base pairs are found from the Frank-Kamenetskii equations (4.6) and (4.7). Singlet enthalpies are determined from them and from the number of Watson-Crick A•T and G•C base pairs,

$$\Delta H(singlet) = T_{\mathrm{A{\bullet}T}} \cdot \Delta S_{\mathrm{bp}} \cdot N_{\mathrm{A{\bullet}T}} + T_{\mathrm{G{\bullet}C}} \cdot \Delta S_{\mathrm{bp}} \cdot N_{\mathrm{G{\bullet}C}} \tag{5.34}$$

Assuming $[\mathrm{Na}^+] = 0.115$ mM and $\Delta S_{\mathrm{bp}} = -24.85$ $\mathrm{cal.mol}^{-1}.\mathrm{K}^{-1}$, the singlet enthalpy is,

$$\Delta H(singlet) = 338.36 \cdot (-24.85) \cdot 5 + 380.97 \cdot (-24.85) \cdot 5 = -89{,}376.75 \ \ \mathrm{cal.mol}^{-1}$$

The n-n enthalpy is obtained from the numbers of different types of doublets in the sequence. These are $N_{AT/AT} = 2$, $N_{TA/TA} = 1$, $N_{AA/TT} = 1$, $N_{AC/GT} = 0$, $N_{CA/TG} = 1$, $N_{TC/GA} = 0$, $N_{CT/AG} = 0$, $N_{CG/CG} = 0$, $N_{GC/GC} = 1$, and $N_{GG/CC} = 3$. Consequently, using n-n values of the enthalpies in Table XVIII yields,

$$\Delta H(doublet) \;=\; \sum_{i,j=\mathrm{A,T,C,G}} N_{ij}(\mathrm{I})\cdot\delta H_{ij}^{n-n} \;=\; 2\cdot(-80.2) + 1\cdot(75.2) + 1\cdot(-155.6) + $$

$$+ 1\cdot(-240.0) + 1\cdot(-95.4) + 3\cdot(172.0) \;=\; -60.2 \;\; \mathrm{cal.mol}^{-1}$$

The third component of the transition enthalpy is the nucleation enthalpy correction required when using the dumbbell parameters to predict the stability of linear DNA duplex oligomers. The length of our duplex is $N_{bp} = 10$ and $f(\mathrm{G{\bullet}C}) = 0.5$. Therefore, the nucleation correction calculated from equation (5.32) is,

$$\Delta H_{nuc} \;=\; 7654.71 - 3469.93\cdot 0.5 - 186.51\cdot 10 \;=\; 4054.65 \quad \mathrm{cal.mol}^{-1}$$

and the total transition entropy is estimated to be,

$$\Delta S_D \;=\; N_{\mathrm{bp}}\cdot\Delta S_{\mathrm{bp}} \;=\; 10\cdot(-24.85) \;=\; -248.5 \quad \mathrm{cal.mol}^{-1}.\mathrm{K}^{-1}$$

Assuming a two-state melting process and a total DNA concentration of $C_T = 4$ μM, the melting temperature is predicted using equation (4.13),

$$T_m = \frac{\Delta H(singlet) + \Delta H(doublet) + \Delta H_{nuc}}{\Delta S_D + R \cdot \ln\left[\frac{C_T}{4}\right]} = \frac{-89{,}376.75 - 60.2 + 4054.65}{-248.5 + 1.9865 \cdot \ln\left[\frac{4x\,10^{-6}}{4}\right]} =$$

$$= 309.42 \quad \mathrm{K} = 36.3\ ^\circ\mathrm{C}$$

This predicted melting temperature is within one degree of the experimentally measured T_m = 37.3 °C (Owczarzy et al., 1997).

5.16 Conclusion

Using 39 DNA dumbbell molecules containing 36 to 48 bases, nearest-neighbor and next-nearest-neighbor stability parameters were evaluated. Dependence of these parameters on the sodium ion concentration was also determined. Within the errors of measurements, nearest-neighbor parameters agreed with those determined in an earlier study (Doktycz et al., 1992). The nearest-neighbor model was able to describe the thermodynamics of dumbbells in 85 and 115 mM Na^+. However, at lower sodium ion concentrations (25 and 55 mM), next-nearest-neighbor interactions were found to be significant. To test the evaluated thermodynamic parameters, melting curves of two additional dumbbells were measured in 25 mM Na^+. If the n-n-n (triplet) interactions are employed, melting temperatures of these two DNA dumbbells are predicted more accurately, although the improvement of accuracy is minor. To apply the newer n-n and n-n-n parameters to predict melting temperatures of DNA duplex oligomers, a correction of nucleation entropy is required. The nucleation correction was evaluated from melting data of DNA duplex oligomers.

APPENDIX

In this Appendix the procedure of obtaining unique linear combinations of physical properties (enthalpies in our case) from unique linear combinations of subunits is described. Linear combinations of subunits and deviations of the transition enthalpies, $\delta H_c^{n\text{-}n}$ and $\delta H_c^{n\text{-}n\text{-}n}$ are reported in Tables XII and XIII. Analogous procedures are employed to derive linear combinations of the n-n and n-n-n transition parameters. For this development we exploit the fact that the stability calculated using non-unique n-n values or unique linear combinations of n-n parameters must be the same. In general for the case of n-n interactions it must be true that,

$$\sum_{i,j=\mathrm{A,C,T,G}} N_{ij}\cdot\delta H_{ij}^{n-n} = \sum_{c=1}^{9} N_c^{n-n}\cdot\delta H_c^{n-n} \qquad \text{(A.1)}$$

In matrix form,

$$[\,N_{\mathrm{AA}} \;\dots\; N_{\mathrm{GG}}\,] * \begin{bmatrix} \delta H_{\mathrm{AA}}^{n-n} \\ \dots \\ \delta H_{\mathrm{GG}}^{n-n} \end{bmatrix} = [\,N_1^{n-n} \;\dots\; N_9^{n-n}\,] * \begin{bmatrix} \delta H_1^{n-n} \\ \dots \\ \delta H_9^{n-n} \end{bmatrix} \qquad \text{(A.2)}$$

The row vector on the left side contains the number of the 10 possible n-n doublets, and the column vector consists of the n-n dependent deviations from the average enthalpy. The row vector on the right side of equation (A.2) contains the numbers of the nine unique linear combinations of the numbers of the n-n doublets. The column vector on the right contains the linear combinations of the n-n enthalpies. The number of n-n doublets and their linear combinations are related according to,

$$[\,N_{\mathrm{AA}}\ \ldots\ N_{\mathrm{GG}}\,]\ *\ \mathbf{A}\ =\ [\,N_1^{n-n}\ \ldots\ N_9^{n-n}\,] \tag{A.3}$$

where **A** is the 10 x 9 transformation matrix,

$$\begin{array}{rcc}
\mathbf{A} = & \left[\begin{array}{ccccccccc}
0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & -1 & -1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & -2 & -1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 2 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 2 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 1 & -2 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & -1 & 2 \\
0 & 0 & 0 & 1 & 0 & 0 & -1 & 1 & -2 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right] & \begin{array}{l}
\mathrm{N_{AT/AT}} \\
\mathrm{N_{TA/TA}} \\
\mathrm{N_{AA/TT}} \\
\mathrm{N_{AC/GT}} \\
\mathrm{N_{CA/TG}} \\
\mathrm{N_{GA/TC}} \\
\mathrm{N_{AG/CT}} \\
\mathrm{N_{CG/CG}} \\
\mathrm{N_{GC/GC}} \\
\mathrm{N_{GG/CC}}
\end{array} \\
\text{Linear combinations, } \mathrm{N_c^{n-n}}: & \begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array} &
\end{array}$$

Each column of **A** corresponds to a particular n-n linear combination while each row corresponds to a particular n-n doublet. In other words, each column of matrix **A** contains coefficients which describe how many times a n-n doublet appears in a particular linear combination. For example, $\mathrm{N_1^{n\text{-}n}} = \mathrm{N_{AA/TT}}$, therefore, the first column of **A** contains 1 in the row describing the numbers of AA/TT doublets and zero values for the numbers of the other n-n doublets. If matrix **A** is not singular, then there is one and only one inverse matrix $\mathbf{A}^{-1}$ such that,

$$\mathbf{A} * \mathbf{A}^{-1} = \mathbf{I} \tag{A.4}$$

Multiplication of any matrix by the unit matrix, **I**, of the same order results in the original matrix,

$$[\, N_{\mathrm{AA}} \;\; \ldots \;\; N_{\mathrm{GG}} \,] * \begin{bmatrix} 1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 \end{bmatrix} * \begin{bmatrix} \delta H_{\mathrm{AA}}^{n-n} \\ \ldots \\ \delta H_{\mathrm{GG}}^{n-n} \end{bmatrix} = [\, N_1^{n-n} \;\; \ldots \;\; N_9^{n-n} \,] * \begin{bmatrix} \delta H_1^{n-n} \\ \ldots \\ \delta H_9^{n-n} \end{bmatrix} \tag{A.5}$$

The unit matrix, **I**, is 10 x 10 with all off diagonal elements of zero and diagonal elements of one. After substitution for the unit matrix **I**, equation (A.5) becomes,

$$[\, N_{\mathrm{AA}} \;\; \ldots \;\; N_{\mathrm{GG}} \,] * \mathbf{A} * \mathbf{A}^{-1} * \begin{bmatrix} \delta H_{\mathrm{AA}}^{n-n} \\ \ldots \\ \delta H_{\mathrm{GG}}^{n-n} \end{bmatrix} = [\, N_1^{n-n} \;\; \ldots \;\; N_9^{n-n} \,] * \begin{bmatrix} \delta H_1^{n-n} \\ \ldots \\ \delta H_9^{n-n} \end{bmatrix} \tag{A.6}$$

Combining equations (A.3) and (A.6) yields,

$$[\, N_1^{n-n} \;\; \ldots \;\; N_9^{n-n} \,] * \mathbf{A}^{-1} * \begin{bmatrix} \delta H_{\mathrm{AA}}^{n-n} \\ \ldots \\ \delta H_{\mathrm{GG}}^{n-n} \end{bmatrix} = [\, N_1^{n-n} \;\; \ldots \;\; N_9^{n-n} \,] * \begin{bmatrix} \delta H_1^{n-n} \\ \ldots \\ \delta H_9^{n-n} \end{bmatrix} \tag{A.7}$$

and,

$$\mathbf{A}^{-1} * \begin{bmatrix} \delta H_{\mathrm{AA}}^{n-n} \\ \dots \\ \delta H_{\mathrm{GG}}^{n-n} \end{bmatrix} = \begin{bmatrix} \delta H_{1}^{n-n} \\ \dots \\ \delta H_{9}^{n-n} \end{bmatrix} \tag{A.8}$$

The 9 x 10 inverse matrix $\mathbf{A}^{-1}$ contains the coefficients of the linear combinations of the n-n enthalpies, $\delta \mathrm{H}_{\mathrm{ij}}^{\mathrm{n-n}}$. To determine $\mathbf{A}^{-1}$ we employed SVD (Press et al., 1989). For the n-n model, our dumbbells and chosen set of n-n linear combinations the inverse matrix $\mathbf{A}^{-1}$ was determined to be,

$$\mathbf{A}^{-1} = \begin{array}{cccccccccc|c}
 & & & & & & & & & & \delta \mathrm{H}_{\mathrm{c}}^{\mathrm{n-n}} \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 \\
\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 4 \\
0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 5 \\
0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 6 \\
\frac{1}{12} & -\frac{1}{12} & 0 & 0 & 0 & \frac{1}{6} & -\frac{1}{6} & \frac{1}{12} & -\frac{1}{12} & 0 & 7 \\
\frac{1}{12} & -\frac{1}{12} & 0 & -\frac{1}{6} & \frac{1}{6} & 0 & 0 & -\frac{1}{12} & \frac{1}{12} & 0 & 8 \\
0 & 0 & 0 & -\frac{1}{12} & \frac{1}{12} & -\frac{1}{12} & \frac{1}{12} & \frac{1}{6} & -\frac{1}{6} & 0 & 9 \\
\mathrm{AT} & \mathrm{TA} & \mathrm{AA} & \mathrm{AC} & \mathrm{CA} & \mathrm{GA} & \mathrm{AG} & \mathrm{CG} & \mathrm{GC} & \mathrm{GG} &
\end{array}$$

The coefficients for the linear combinations of the n-n transition enthalpies are listed in each row. The corresponding number of the n-n combinations of the enthalpies (in Table XII) is indicated in the column at the right. The n-n doublet corresponding to each column is given below the matrix.

To derive the combinations of n-n-n dependent deviations from the average transition enthalpy, $\delta H_c^{\text{n-n-n}}$, from the linear combinations of the number of n-n-n subunits, $N_c^{\text{n-n-n}}$, an analogous procedure is applied. No matter if the n-n-n interactions are calculated from the non-unique transition enthalpies of next-nearest-neighbors (triplets) or from their unique linear combinations, the n-n-n sequence dependent stability must be the same, i.e.,

$$\sum_{i,j,k=\mathrm{A,C,T,G}} N_{ijk} \cdot \delta H_{ijk}^{n-n-n} = \sum_{c=1}^{27} N_c^{n-n-n} \cdot \delta H_c^{n-n-n} \tag{A.9}$$

There are 32 triplets and 27 unique linearly independent linear combinations of the n-n-n triplets. Therefore, the 32 x 27 transformation matrix $\mathbf{A_{n\text{-}n\text{-}n}}$ exists,

$$[\,N_{\mathrm{AAA}} \;\ldots\; N_{\mathrm{GGG}}\,] \;*\; \mathbf{A_{n-n-n}} \;=\; [\,N_1^{n-n-n} \;\ldots\; N_{27}^{n-n-n}\,] \tag{A.10}$$

Its inverse matrix, $\mathbf{A_{n-n-n}^{-1}}$, connects the n-n-n enthalpies with their linear combinations of n-n-n subunits.

$$\mathbf{A}_{\mathbf{n-n-n}}^{\mathbf{-1}} * \begin{bmatrix} \delta H_{\mathrm{AAA}}^{n-n-n} \\ \ldots \\ \delta H_{\mathrm{GGG}}^{n-n-n} \end{bmatrix} = \begin{bmatrix} \delta H_{1}^{n-n-n} \\ \ldots \\ \delta H_{27}^{n-n-n} \end{bmatrix} \qquad (A.11)$$

From the $\mathbf{A}_{\mathbf{n-n-n}}^{\mathbf{-1}}$, the coefficients for the linear combinations of transition enthalpies of $\delta\mathrm{H}_{\mathrm{ijk}}^{\mathrm{n-n-n}}$, reported in the second column of Table XIII, were obtained.

CITED LITERATURE

Abe, T., Takai, K., Nakada, S., Tomoyuki, Y., Takaku, H.: Specific inhibition of influenza virus RNA polymerase and nucleoprotein gene expression by circular dumbbell RNA/DNA chimeric oligonucleotides containing antisense phosphodiester oligonucleotides. FEBS Letters 425:91-96, 1998.

Abrams, E.S., Murdaugh, S.E., Lerman, L.S.: Intramolecular DNA melting between stable helical segments: melting theory and metastable states. Nucleic Acids Research 23:2775-2783, 1995.

Aida, M.: An ab initio molecular orbital study on the sequence-dependency of DNA conformation: An evaluation of intra- and inter-strand stacking interaction energy. J. Theor. Biol. 130:327-335, 1988.

Allawi, H.T., SantaLucia, J., Jr.: Thermodynamics and NMR of internal G•T mismatches in DNA. Biochemistry 36:10581-10594, 1997.

Anderson E., Bai Z., Bischof, C., Demmel, J., Dongarra J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A., Ostrouchov, S., Sorensen, D.: LAPACK User's Guide - Release 2.0. SIAM, Philadelphia, PA, 1994.

Ashley, G.W., Kushlan, D.M.: Chemical synthesis of oligodeoxynucleotide dumbbells. Biochemistry 30:2927-2933, 1991.

Avery, O.T., MacLeod, C.M., McCarthy, C.: Studies on the chemical nature of the substance inducing transformation of pneumococcal types. J. Exp. Med. 79:137-158, 1944.

Benight, A.S., Gallo, F.J., Paner, T.M., Bishop, K.D., Faldasz, B.D., Lane, M.J.: Sequence context and DNA reactivity: Application to sequence-specific cleavage of DNA. Advances in Biophysical Chemistry 5:1-55, 1995.

Benight, A.S., Schurr, J.M., Flynn, P.F., Reid, B.R.: Melting of a self-complementary DNA minicircle. J. Mol. Biol. 200:377-399, 1988.

Benight, A.S., Wartell, R.M., Howell, D.K.: Theory agrees with experimental denaturation of short DNA restriction fragments. Nature 289:203-205, 1981.

Benight, A.S., Wartell, R.M.: Influence of base-pair changes and cooperativity parameters on the melting curves of short DNAs. Biopolymers 22:1409-1425, 1983.

Berners-Price, S.J., Corazza, A., Guo, Z., Barnham, K.J., Sadler, P.J., Ohyama, Y., Leng, M., Locker, D.: Structural transitions of a GC-platinated DNA duplex induced by pH, temperature, and box A of high-mobility-group protein 1. Eur. J. Biochem. 243:782-791, 1997.

Bevington, P.R., Robinson, D.K.: Data Reduction and Error Analysis for the Physical Sciences. New York, McGraw-Hill, Inc., 1992.

Blake, R.D., Hydorn, T.G.: Spectral analysis for base composition of DNA undergoing melting. J. Biochem. Biophys. Meth. 11:307-316, 1985.

Breslauer, K.J., Frank, R., Blöcker, H., Marky, L.A.: Predicting DNA duplex stability from the base sequence. Proc. Natl. Acad. Sci. 83:3746-3750, 1986.

Cantor and Warshaw: Oligonucleotide interactions. III. Circular dichroism studies of the conformation of deoxyoligonucleotides. Biopolymers 9:1059-1077, 1970.

Caruthers, M.H.: Gene synthesis machines: DNA chemistry and its uses. Science 230:281-285, 1985.

Caruthers, M.H., Beaton, G., Wu, J.V., Wiesler, W.: Chemical synthesis of deoxyoligonucleotides and deoxyoligonucleotide analogs. In: Methods in Enzymology eds. D.M.J. Lilley, J.E. Dahlberg, pp. 3-20. San Diego, Academic Press, 1992.

Chaires, J.B.: Possible origin of differences between van't Hoff and calorimetric enthalpy estimates. Biophys. Chem. 64:15-23, 1997.

Chrambach, A., Jovin, T.M.: Selected buffer systems for moving boundary electrophoresis on gels at various pH values, presented in a simplified manner. Electrophoresis 4:190-204, 1983.

Clark, E.C.W., Glew, D.N.: Evaluation of thermodynamic functions from equilibrium constants. Trans. Faraday Soc. 62:539-547, 1966.

Davis, T.M., McFail-Isom, L., Keane E., Williams L.D.: Melting of a DNA hairpin without hyperchromism. Biochemistry 37:6975-6978, 1998.

Dawson, R.M.C., Elliott, D.C., Elliott, W.H., Jones, K.M.: Data for Biochemical Research. pp. 422-432, Oxford, Clarendon Press, 1986.

Delcourt, S.G., Blake, R.D.: Stacking energies in DNA. J. Biol. Chem. 266:15160-15169, 1991.

DeVoe, H., Tinoco, I. Jr.: The stability of helical polynucleotides: Base contributions. J. Mol. Biol. 4:500-517, 1962

Dickerson, R.E., Chiu, T.K.: Helix bending as a factor in protein/DNA recognition. Biopolymers 44:361-403, 1997.

DiDonato, A.R.,. Morris, A.H. Jr.: ALGORITHM 654: FORTRAN subroutines for computing the incomplete gamma function ratios and their inverse. ACM Trans. Math. Softw. 13:318-319, 1987.

Dobos, D.: Handbook of Electrochemical Data. Amsterdam, Elsevier Scientific Publishing Company, 1975.

Doktycz, M.J., Goldstein, R.F., Paner, T.M., Gallo, F.J., Benight, A.S.: Studies of DNA Dumbbells I: Melting curves of 17 DNA dumbbells with the different duplex stem sequences linked by T_4 end-loops: Evaluation of the nearest-neighbor stacking interactions in DNA. Biopolymers 32:849-864, 1992.

Doktycz, M.J.: Discontinuous electrophoresis of DNA: Adjusting DNA mobility by trailing ion net mobility. Analytical Biochemistry 213:400-406, 1993.

Doktycz, M.J., Morris, M.D., Dormady, S.J., Beattie, K.L., Jacobson, K.B.: Optical melting of 128 octamer DNA duplexes. J. Biol. Chem. 270:8439-8445, 1995.

Dove, W.F., Davidson, N.: Cation effects on the denaturation of DNA. J.Mol. Biol. 5:467-478, 1962.

Drmanac, S., Kita, D., Labat, I., Hauser, B., Schmidt, C., Burczak, J., Drmanac, R.: Accurate sequencing by hybridization for DNA diagnostic and individual genomics. Nature Biotechnology 16:54-58, 1998.

Eichhorn, G.L., Shin, Y.A.: Interaction of metal ions with polynucleotides and related compounds. XII. The relative effect of various metal ions on DNA helicity. J. Am. Chem. Sci. 90:7323-7328, 1968.

Elson, E.L.: Helix formation by d(TA) oligomers. III. Electrostatic effects. J.Mol. Biol. 54:401-415, 1970.

Fareed, G.C., Wilt, E.M., Richardson, C.C.: Enzymatic breakage and joining of deoxyribonucleic acid. J. Biol. Chem. 246:925-932, 1971.

Fasman, G.D.: Handbook of Biochemistry and Molecular Biology. Volume I, p. 589, CRC Press, 1975.

Fasman, G.D.: Practical Handbook of Biochemistry and Molecular Biology. pp. 536-552, Boca Raton, Florida, CRC Press, 1989.

Frank-Kamenetskii, M.D.: Simplification of the empirical relationship between melting temperature of DNA, its GC content and concentration of sodium ions in solution. Biopolymers 10:2623-2624, 1971.

Gmelin, E., Sarge, St.M.: Calibration of differential scanning calorimeters. Pure and Appl. Chem. 67:1789-1800, 1995.

Goldstein, R.F., Benight, A.S.: How many numbers are required to specify sequence-dependent properties of polynucleotides? Biopolymers 32:1679-1693, 1992.

Good, N.E., Winget, G.D., Winter, W., Connolly, T.N., Izawa, S., Singh, R.M.M.: Hydrogen ion buffers for biological research. Biochemistry 5:467-477, 1966.

Gotoh, O., Tagashira, Y.: Stabilities of nearest-neighbor doublets in double-helical DNA determined by fitting calculated melting profiles to observed profiles. Biopolymers 20:1033-1042, 1981.

Gray, D. M., Tinoco, I.: A new approach to the study of sequence-dependent properties of polynucleotides. Biopolymers 9:223-244, 1970.

Gray, D.M.: Derivation of nearest-neighbor properties from data on nucleic acid oligomers. I. Simple sets of independent sequences and influence of absent nearest neighbors. Biopolymers 42: 783-793, 1997a.

Gray, D.M.: 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, 1997b.

Gruenwedel, D.W., Chi-Hsia, H.: Salt effects on the denaturation of DNA. Biopolymers 7:557-570, 1969.

Gruenwedel, D.W., Chi-Hsia, H., Lu, D.S.: The effects of aqueous neutral-salt solutions on the melting temperatures of deoxyribonucleic acids. Biopolymers 10:47-68, 1971.

Harada, K., Orgel, L.E.: Unexpected substrate specificity of T4 DNA ligase revealed by in vitro selection. Nucleic Acids Research 21:2287-2291, 1993.

Ippel, J.H., Lanzotti, V., Galeone, A., Mayol, L., Boogaart Van den, J.E., Pikkemaat, J.A., Altona, C.: Thermodynamics of melting of circular dumbbell d<pCGC-TT-GCG-TT> Biopolymers 36:701-710, 1995.

Ivanov, I.G., AbouHaidar, M.G.: Thermal stability of oligodeoxynucleotide duplexes bound to nitrocellulose filters. Analytical Biochemistry 232:249-251, 1995.

Jovin, T.M: Multiphasic zone electrophoresis. II. Design of integrated discontinuous buffer systems for analytical and preparative fractionation. Biochemistry 12:879-898, 1973.

Kaiser, J.F., Reed, W.A.: Data smoothing using low-pass digital filters. Rev. Sci. Instrum. 48:1447-1457, 1977.

Klump, H.H.: Conformational transitions in nucleic acids. In: Studies in Modern Thermodynamics 8: Biochemical Thermodynamics, eds M.N. Jones, 2nd ed., Amsterdam, Elsevier, pp. 100-144, 1988.

Kunitsyn, A., Kochetkova, S., Timofeev, E., Florentiev, V.: Partial thermodynamic parameters for prediction stability and washing behavior of DNA duplexes immobilized on gel matrix. J. Biomol. Struc. Dyn. 14:239-244, 1996.

Ladbury, J.E., Chowdhry, B.Z.: Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. Chemistry and Biology 3:791-801, 1996.

Landick, R., Maguire, D., Lutter, L.C.: Optimization of polyacrylamide gel electrophoresis conditions used for sequencing mixed oligodeoxyribonucleotides. DNA 3:413-419, 1984.

Lim, C.S., Jabrane-Ferrat, N., Fontes, J.D., Okamoto, H., Garovoy, M.R., Peterlin, B.M., Hunt, C.A.: Sequence-independent inhibition of RNA transcription by DNA dumbbells and other decoys. Nucleic Acids Research 25:575-581, 1997.

Lipshutz, R.J., Morris, D., Chee, M, Hubbell, E., Kozal, M.J., Shah, N., Shen, N., Yang, R., Fodor, S.P.A.: Using oligonucleotide probe arrays to access genetic diversity. Biotechniques 19:442-447, 1995.

Liu, D., Daubendiek, S.L., Zillman, M.A., Ryan, K., Kool, E.T.: Rolling circle DNA synthesis: Small circular oligonucleotides as efficient templates for DNA polymerases. J. Am. Chem. Soc. 118:1587-1594, 1996.

Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., Brown, E.L.: Expression monitoring by hybridization to high-density oligonucleotide arrays. Nature Biotechnology 14:1675-1680, 1996.

Maniatis, T., Jeffrey, A., van deSande, H.: Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis. Biochemistry 14:3787-3794, 1975.

Marky, L.A., Breslauer, K.J.: Calorimetric determination of base-stacking enthalpies in double-helical DNA molecules. Biopolymers, 21:2185-2194, 1982.

Marky, L.A., Breslauer, K.J.: Calculating thermodynamic data for transition of any molecularity from equilibrium melting curves. Biopolymers 26:1601-1620, 1987.

Marmur, J., Doty, P.: Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. J. Mol. Biol. 5:109-118, 1962.

Maurer, H.R.: Disc Electrophoresis and Related Techniques of Polyacrylamide Gel Electrophoresis. Berlin, Germany, Walter de Gruyter, 1971.

McCampbell, C.M., Wartell, R.M., Plaskon, R.R: Inverted repeat sequences can influence the melting transitions of linear DNAs. Biopolymers 28:1745-1758, 1989.

McGall, G., Labadie, J., Brock, P., Wallraff, G., Nguyen, T., Hinsberg, W.: Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. Proc. Natl. Acad. Sci. USA 93:13555-13560, 1996.

Meyer, S.L.: Data Analysis for Scientists and Engineers., pp. 71-75, New York, John Wiley & Sons, 1975.

Muccitelli, J.A., DiAngelo, N.A.: Electrolytic conductivity of aqueous solutions of potassium and sodium phosphates to 325°C. J. Chem. Eng. Data 39:131-133, 1994.

Mukerji, I., Sokolov, L., Mihailescu, M.R.: A UV resonance Raman investigation of poly(rI): Evidence for cation-dependent structural perturbation. Biopolymers 46:475-487, 1998.

Neiderweis, M., Lederer, T., Hillen, W.: Matrix effects suggest an important influence of DNA-polyacrylamide interactions on the electrophoretic mobility of DNA. J. Biol. Chem. 269:10156-10162, 1994.

Nelson, N.C., Cheikh, A.B., Matsuda, E., Becker, M.M.: Simultaneous detection of multiple nucleic acid targets in homogeneous format. Biochemistry 35:8429-8438, 1996.

Ornstein, R.L., Fresco, J.R.: Correlation of T_m and sequence of DNA duplexes with ΔH computed by an improved empirical potential method. Biopolymers 22:1979-2000, 1983.

Owczarzy, R., Vallone, P.M., Gallo, F.J., Paner, T.M., Lane, M.J., Benight, A.S.: Predicting sequence-dependent melting stability of short duplex DNA oligomers. Biopolymers 44:217-239, 1997.

Owen, R.J., Hill, L.R., Lapage, S.P.: Determination of DNA base compositions from melting profiles in dilute buffers. Biopolymers 7:503-516, 1969.

Panasenko, S.M., Cameron, J.R., Davis, R.W., Lehman, I.R.: Five hundredfold overproduction of DNA ligase after induction of a hybrid lambda lysogen constructed in vitro. Science 196:188-189, 1977.

Panasenko, S.M., Alazard, R.J., Lehman, I.R.: A simple, three-step procedure for the large scale purification of DNA ligase from a hybrid λ lysogen constructed in vitro. J. Biol. Chem. 253:4590-4592, 1978.

Paner, T.M., Amaratunga, M., Benight, A.S.: Studies of DNA dumbbells III: Theoretical analysis of optical melting curves of dumbbells with a sixteen base-pair duplex stem and T_n end-loops (n = 2, 3, 4, 6, 8, 10, 14). Biopolymers 32:881-892, 1992.

Paner, T.M., Riccelli, P.V., Owczarzy, R., Benight, A.S.: Studies of DNA dumbbells VI: Analysis of optical melting curves of dumbbells with a sixteen-base pair duplex stem and end-loops of variable size and sequence. Biopolymers 39:779-793, 1996.

Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P., Fodor, S.P.A.: Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc. Natl. Acad. Sci. USA 91:5022-5026, 1994.

Pheiffer, B.H., Zimmerman, S.B.: Polymer-stimulated ligation: enhanced blunt- or cohesive-end ligation of DNA or deoxyribooligonucleotides by T4 DNA ligase in polymer solutions. Nucleic Acids Research 11:7853-7871, 1983.

Pirrung, M.C., Fallon, L., McGall, G.: Proofing of photolithographic DNA synthesis with 3',5'-Dimethoxybenzoinyloxycarbonyl-protected Deoxynucleoside phosphoramidites. J. Org. Chem. 63:241-246, 1998.

Pohl, F.M., Thomae, R., Karst, A.: Temperature dependence of the activity of DNA-modifying enzymes: Endonucleases and DNA ligase. Eur. J. Biochem. 123:141-152, 1982.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T.: Numerical Recipes. The Art of Scientific Computing (FORTRAN Version). Cambridge, Cambridge University Press, 1989.

Raae, A.J., Kleppe, R.K., Kleppe, K.: Kinetics and effect of salts and polyamines on T4 polynucleotide ligase. Eur. J. Biochem. 60:437-443, 1975.

Ratmeyer, L., Vinayak, R., Zhong, Y.Y., Zon, G., Wilson, W.D.: Sequence specific thermodynamic and structural properties for DNA•RNA duplexes. Biochemistry 33:5298-5304, 1994.

Record, M.T., Jr., Mazur, S.J., Melançon, P., Roe, J.-H., Shaner, S.L., Unger, L.: Double helical DNA: Conformations, physical properties, and interactions with ligands. Ann. Rev. Biochem. 50:997-1024, 1981.

Riccelli, P.V., Vallone, P.M., Kashin, I., Schildkraut, I., Faldasz, B.D., Lane, M.J., Benight, A.S.: Sequence-dependent stability of flanking DNA modulates binding of BamHI restriction endonuclease to its cognate recognition sequence. Submitted to Biophysical Journal, 1998.

SantaLucia, J. Jr., Allawi, H.T., Seneviratne, P.A.: Improved nearest-neighbor parameters for predicting DNA duplex stability. Biochemistry 35:3555-3562, 1996.

Sauer, K.: Biochemical Spectroscopy: Methods in Enzymology. Vol. 246, San Diego, California, Academic Press, 1995.

Schildkraut, C., Lifson, S.: Dependence of the melting temperature of DNA on salt concentration. Biopolymers 3:195-208, 1965.

Sheppard, T.L., Breslow, R.C: Selective binding of RNA, but not DNA, by complementary 2'-5'-linked DNA. J. Am. Chem. Soc. 118:9810-9811, 1996.

Shui, X., McFail-Isom, L., Hu, G.G., Williams, L.D.: The B-DNA dodecamer at high resolution reveals a spine of water on sodium. Biochemistry 37:8341-8355, 1998.

Snedecor, G.W., Cochran, W.G.: Statistical Methods. Ames, Iowa, U.S.A., The Iowa State University Press, 1980.

Sokolova, N.I., Ashirbekova, D.T., Dolinnaya, N.G., Shabarova, Z.A.: Chemical reactions within DNA duplexes. Cyanogen bromide as an effective oligodeoxyribonucleotide coupling agent. FEBS Lett. 232:153-155, 1988.

Sturtevant, J.M.: Biochemical applications of differential scanning calorimetry. Ann. Rev. Phys. Chem. 38:463-488, 1987.

Sugimoto, N., Honda, K., Sasaki, M.: Application of the thermodynamic parameters of DNA stability prediction to double-helix formation of deoxyribooligonucleotides. Nucleosides & Nucleotides 13:1311-1317, 1994.

Sugimoto, N., Nakano, S., Yoneyama, M., Honda, K.: Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. Nucleic Acids Research 24:4501-4505, 1996.

Sugino, A., Goodman, H.M., Heyneker, H.L., Shine, J., Boyer, H.W., Cozzarelli, N.R.: Interaction of Bacteriophage T4 RNA and DNA ligases in joining of duplex DNA at base-paired ends. J. Biol. Chem. 252:3987-3994, 1977.

Thomas, R.: Recherches sur la dénaturation des acides desoxyribonucléiques. Biochimica et Biophysica Acta 14:231-240, 1954.

Vallone, P.: Ph.D. thesis, University of Illinois, Chicago, 1999.

Vologodskii, A.V., Amirikyan, B.R., Lyubchenko, Y.L., Frank-Kamenetskii, M.D.: Allowance for heterogeneous stacking in the DNA helix-coil transition theory. J. Biomol. Struct. Dynam. 2:131-148, 1984.

Vamosi, G., Clegg, R.M.: The helix-coil transition of DNA duplexes and hairpins observed by multiple fluorescence parameters. Biochemistry 37:14300-14316, 1998.

Wada, A., Yabuki, S., Husimi, Y.: Fine structure in the thermal denaturation of DNA: High temperature-resolution spectrophotometric studies. CRC Critical Revievs in Biochemistry 9:87-144, 1980.

Wadi, R.K., Saxena, P.: Molar conductivity of alkali halides in ethanolamine and water + ethanolamine at 298.15 K. Indian J. Chem. 34A:273-277, 1995.

Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M.S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T.J., Lipshutz, R., Chee, M., Lander, E.S.: Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280:1077-1082, 1998.

Wang, S., Kool, E.T.: Circular RNA oligonucleotides. Synthesis, nucleic acid binding properties, and a comparison with circular DNAs. Nucleic Acids Research 22:2326-2333, 1994.

Wartell, R.M., Benight, A.S.: Thermal denaturation of DNA molecules: A comparison of theory with experiment. Phys. Rep. 126:67-107, 1985.

Wartell R.M., Hosseini S., Powell S., Zhu J.: Detecting single base substitutions, mismatches and bulges in DNA by temperature gradient gel electrophoresis and related methods. J. Chromatography 806:169-185, 1998.

Watson, J.D., Crick, F.H.C.: Molecular structure of nucleic acids. Nature 171:737-738 , 1953.

Weiss, B.: Endonuclease II of Escherichia coli is Exonuclease III. J. Biol. Chem. 251:1896-1901, 1976.

Wemmer, D.E., Benight, A.S.: Preparation and melting of single strand circular DNA loops. Nucleic Acids Research 13:8611-8621, 1985.

Widlak, P., Bykov, V.J., Hemminki, K.: Formation of UV-photoadducts during DNA purification. Mutation Research 347:117-119, 1995.

Xia, T., SantaLucia, J. Jr., Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C., Turner, D.H.: Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. Biochemistry 37:14719-14735, 1998.

Yamakawa, H., Ishibashi, T., Abe, T., Hatta, T., Takai, K., Takaku, H.: Anti-influenza activities of nicked and circular dumbbell RNA/DNA chimeric oligonucleotides. Nucleosides and Nucleotides 16:1713-1716, 1997.

Zhu J., Wartell R.M.: The relative stabilities of base pair stacking interactions and single mismatches in long RNA measured by temperature gradient gel electrophoresis. Biochemistry 36:15326-15335, 1997.

Zimmerman, S.B., Pheiffer, B.: Macromolecular crowding allows blunt-end ligation by DNA ligases from rat liver or Escherichia coli. Proc. Natl. Acad. Sci. 80:5852-5856, 1983.

Zsolnai, A., Orbán L., Chrambach, A.: Agarose electrophoresis of DNA in discontinuous buffers, using a horizontal slab apparatus and a buffer system with improved properties. Electrophoresis 14:179-184, 1993.