QUANTIFYING AND MODELING THE MELTING THERMODYNAMICS OF CHEMICALLY MODIFIED DUPLEX DNA

by

KAREEM FAKHFAKH

B.A.Sc (Honours), The University of British Columbia, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Chemical and Biological Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2016

© Kareem Fakhfakh, 2016
ABSTRACT

Biological reagents that bind a target selectively and with high affinity are widely used as recognition molecules within diagnostic assays and as therapeutics, among other applications. By leveraging their Watson-Crick base pairing ability, short DNA oligonucleotides represent one class of such biological agents that is particularly well suited to analyzing specific elements of the human genome. Such analyses are routinely used by clinics to detect and manage disease, and those analyses are increasingly providing the richer data content and improved performance necessary for effective clinical decision-making by employing chemically modified nucleic acids. To date, the use of these unnatural nucleotides has largely been achieved empirically, but their growing use is motivating the development of new tools and guidelines that accelerate and improve their implementation in novel assays.

This thesis describes how two experimental methods may be tailored to accurately measure the melting thermodynamics of short duplex DNA containing chemical modifications – specifically locked nucleic acids (LNAs) – and then reports on a study that used those methods to measure the thermal stabilities of a large panel of DNA duplexes containing LNA substitutions in one or both strands. Those data and insights gleaned from them are used to extend a molecular thermodynamic model, the “Single Base Thermodynamic” (SBT) model[1], to enable accurate predictions of the melting thermodynamics of short B-form DNA duplexes containing i) LNA:LNA base pair and/or ii) oppositely oriented LNA:DNA base pair structures. It is the only thermodynamic model with this ability, and its value is demonstrated through its use to guide the development of a entirely new type of quantitative real-time PCR based diagnostic assay – in this case directed against clinically relevant *BRAF*V600 mutations in cancer – that improves upon commercially available assays by bettering their throughput and limit of detection.
PREFACE

A version of Chapter 2 from this dissertation has been published as:


As a first author, I drafted the initial manuscript and further expert recommendations and contributions were made by my principle supervisor Dr. Charles Haynes and laboratory post-doctoral fellow Dr. Curtis Hughesman in arriving at the final manuscript. Technical input and discussion from my laboratory manager Dr. Louise Creagh and a research assistant Vincent Kao was useful in finalizing the suggested method protocols.

A version of Chapter 3 from this dissertation has been published as:


This work was published in a special edition of A.I.Ch.E Journal honouring John Prausnitz. This work significantly extended the studied DNA sequences of Jorge Castañeda to advance the molecular thermodynamic model of Dr. Curtis Hughesman and Dr. Charles Haynes. As a first author, I designed the research together with Dr. Charles Haynes and Dr. Curtis Hughesman. I performed the experimental work and analysis as described including the developemnt of the modular Matlab® programs (Appendix A) used to analyze the results. Olivia Marais and Xinbo Cheng contributed thoughtful discussion to the work.

The work in Chapter 4 is part of a collaboration with the BC Cancer Agency (BCCA) to develop a novel high fidelity assay to identify and genotype *BRAF*V600 mutations in melanoma patients. My contribution to this work was to measure the melting thermodynamics of the PCR primers and probes under the supervision of Dr. Curtis Hughesman and Dr. Charles Haynes. I analyzed and compared the UVM derived melting thermodynamics to the model predicted melting thermodynamics using the extended model developed in Chapter 3. As with Chapter 1,
I wrote the initial manuscript draft for this chapter and it was improved with contributions from Dr. Charles Haynes.

As an objective of my thesis, I programmed in R/Shiny an online interactive version of the Single Base Thermodynamic Model of Hughesman et al.[1]. Users may input a natural or modified DNA sequence of interest and the program computes for them the predicted melting thermodynamics associated with that duplex. A summary of the program including its operation, mechanics, and code are included in B Appendix B.
TABLE OF CONTENTS

ABSTRACT .................................................................................................................................................. ii
PREFACE .................................................................................................................................................. iii
TABLE OF CONTENTS .............................................................................................................................. v
LIST OF TABLES ......................................................................................................................................... ix
LIST OF FIGURES ..................................................................................................................................... xi
LIST OF SYMBOLS .................................................................................................................................. xvii
LIST OF ABBREVIATIONS ......................................................................................................................... xviii
ACKNOWLEDGEMENTS ............................................................................................................................ xx

CHAPTER 1: INTRODUCTION ...................................................................................................................... 1

1.1 THESIS OVERVIEW ........................................................................................................................... 1

1.2 CLINICAL DETECTION OF CANCER-RELATED GENOMIC MUTATIONS ........................................... 2

1.2.1 CHALLENGES IN DETECTING GENETIC ABERRATIONS IN CANCERS .................................. 3

1.2.2 CLINICAL DIAGNOSTICS USED TO DETECT POINT MUTATIONS IN CANCER ......................... 5

1.2.2.1 IMMUNOHISTOCHEMISTRY ........................................................................................................ 5

1.2.2.2 SEQUENCING TECHNOLOGIES .................................................................................................. 9

1.2.2.3 PCR BASED DIAGNOSTIC TECHNOLOGIES ............................................................................. 11

1.2.2.3.1 POLYMERASE CHAIN REACTION (PCR) ................................................................................ 11

1.2.2.3.2 HIGH RESOLUTION MELT ANALYSIS .................................................................................... 13

1.2.2.3.3 REAL-TIME MONITORING OF PCR AMPLIFICATION BY QUANTITATIVE PCR (QPCR) ........ 14

1.3 LOCKED NUCLEIC ACID (LNA) CHEMISTRY AND PROPERTIES .................................................... 20

1.3.1 THERMODYNAMIC AND STRUCTURAL PROPERTIES OF LNA ................................................... 22

1.3.2 USE OF LNAs IN THERAPEUTICS AND DIAGNOSTICS ............................................................ 23

1.3.2.1 LNA CONTAINING THERAPEUTICS .......................................................................................... 23

1.3.2.2 USE OF LNAs IN DISEASE DIAGNOSTICS .............................................................................. 24

1.4 THESIS OBJECTIVES ......................................................................................................................... 26
1.5 MODELS FOR PREDICTING THE MELTING THERMODYNAMICS OF NATURAL AND CHEMICALLY MODIFIED DNA DUPLEXES.................................................. 27
   1.5.1 CLASSIC NEAREST NEIGHBOR THERMODYNAMIC MODELS.............. 28
   1.5.2 ADVANCED NNT MODELS .................................................................. 32
   1.5.3 CLASSIC AND ADVANCED NNT MODELS FOR PREDICTING THE STABILITY OF LNA-CONTAINING DUPLEXES............................................. 33

CHAPTER 2: CALORIMETRIC AND SPECTROSCOPIC ANALYSIS OF THE THERMAL STABILITY OF SHORT DUPLEX DNA CONTAINING SUGAR AND BASE MODIFIED NUCLEOTIDES.................................................................................37
   2.1 SYNOPSIS.................................................................................................. 37
   2.2 INTRODUCTION .......................................................................................... 37
   2.3 MELTING THERMODYNAMICS FROM DIFFERENTIAL SCANNING CALORIMETRY.............................................................................................................. 40
      2.3.1 STRENGTHS, WEAKNESSES AND WHAT CAN BE DETERMINED RELIABLY .......................................................................................................... 40
      2.3.2 SAMPLE PREPARATION ......................................................................... 44
      2.3.3 EXPERIMENTAL PROCEDURE ................................................................. 45
      2.3.4 ANALYZING A DSC THERMOGRAM ..................................................... 47
   2.4 MELTING THERMODYNAMICS FROM UV SPECTROSCOPY .................. 54
      2.4.1 STRENGTHS, WEAKNESSES AND WHAT CAN BE DETERMINED RELIABLY .......................................................................................................... 56
      2.4.2 SAMPLE PREPARATION ......................................................................... 57
      2.4.3 EXPERIMENTAL PROCEDURE ................................................................. 58
      2.4.4 ANALYSIS OF UVM DATA ...................................................................... 60
   2.5 ERRORS ANALYSIS .................................................................................... 66
   2.6 UNDERSTANDING THE CONTRIBUTION TO DUPLEX STABILITY OF SPECIFIC NUCLEOTIDE MODIFICATIONS......................................................... 67

CHAPTER 3: MOLECULAR THERMODYNAMICS OF LNA:LNA BASE PAIRS AND THE HYPERSTABILIZING EFFECT OF 5’-PROXIMAL LNA:DNA BASE PAIRS .....72
   3.1 SYNOPSIS.................................................................................................. 72
3.2 INTRODUCTION ............................................................................................................ 72
3.3 METHOD ....................................................................................................................... 77
  3.3.1 MATERIALS ........................................................................................................... 77
  3.3.2 MEASUREMENT OF MELTING THERMODYNAMICS BY UV SPECTROSCOPY .......... 78
  3.3.3 REGRESSION OF MODEL PARAMETERS ............................................................... 81
3.4 RESULTS ....................................................................................................................... 84
  3.4.1 STANDARD SBT MODEL AND ITS PERFORMANCE WHEN APPLIED TO DUPLEXES CONTAINING ANY NUMBER AND PATTERN OF LNA SUBSTITUTIONS WITHIN ONE STRAND ............................................................. 84
  3.4.2 LIBRARY DESIGN FOR ANALYZING THE EFFECT OF LNA:LNA BASE PAIRS AND OPPOSITELY ORIENTED LNA:DNA BASE PAIRS ON DUPLEX STABILITY ......................................................................................................................... 86
  3.4.3 THE INCREMENTAL STABILITY OF AN LNA:LNA BASE PAIR EXCEEDS THE SUMMED INCREMENTAL STABILITIES OF THE TWO LNA:DNA BASE PAIRS FROM WHICH IT IS DERIVED ........................................................................ 111
  3.4.4 5’-OFFSET OPPOSITELY ORIENTED LNA:DNA BASE PAIRS ALSO HYPERSTABILIZE AN LNA-SUBSTITUTED DUPLEX ................................................................................................................................. 115
  3.4.5 EXTENDING THE SBT MODEL TO PERMIT APPLICATION TO DUPLEXES CONTAINING LNA SUBSTITUTIONS ON BOTH STRANDS ......................................................................................... 117
3.5 CONCLUDING REMARKS ............................................................................................. 129

CHAPTER 4: MODEL APPLICATION TO DEVELOPMENT OF A NOVEL, REAL-TIME PCR ASSAY AGAINST BRAF V600 MUTATIONS .............................................................. 131
4.1 CLINICAL SIGNIFICANCE OF BRAF MUTATIONS IN MELANOMA .................... 131
  4.1.1 BRAF V600 MUTATIONS IN MELANOMA ............................................................ 133
  4.1.2 TARGETING BRAF KINASES FOR CANCER THERAPY ..................................... 135
  4.1.3 AVAILABLE DIAGNOSTIC ASSAYS FOR DETECTING BRAF V600 MUTATIONS ......................................................................................................................................................... 136
4.2 MATERIALS AND METHODS ..................................................................................... 137
  4.2.1 OLIGONUCLEOTIDES ........................................................................................... 137
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.2 UV MELT SPECTROSCOPY</td>
<td>137</td>
</tr>
<tr>
<td>4.2.3 ALLELE-SPECIFIC PCR WORKFLOW</td>
<td>137</td>
</tr>
<tr>
<td>4.3 RESULTS</td>
<td>138</td>
</tr>
<tr>
<td>4.3.1 BRAF V600 MUTATION-SPECIFIC REAL-TIME PCR ASSAY DESIGN</td>
<td>138</td>
</tr>
<tr>
<td>4.3.2 MODEL-BASED DESIGN OF THE LNA-SUBSTITUTED PROBES</td>
<td>143</td>
</tr>
<tr>
<td>4.3.3 ASSAY PERFORMANCE USING MODEL-DESIGNED AS PROBES</td>
<td>149</td>
</tr>
<tr>
<td>CHAPTER 5: CONCLUSIONS AND OUTLOOK</td>
<td>154</td>
</tr>
<tr>
<td>5.1 PROPOSED FUTURE DEVELOPMENTS</td>
<td>156</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>159</td>
</tr>
<tr>
<td>APPENDIX A</td>
<td>173</td>
</tr>
<tr>
<td>A.1 DESCRIPTION OF MATLAB</td>
<td>173</td>
</tr>
<tr>
<td>A.2 SUMMARY OF PARAMETER REGRESSION</td>
<td>173</td>
</tr>
<tr>
<td>A.3 MATLAB CODE</td>
<td>175</td>
</tr>
<tr>
<td>APPENDIX B</td>
<td>190</td>
</tr>
<tr>
<td>B.1 DESCRIPTION OF R / SHINY</td>
<td>190</td>
</tr>
<tr>
<td>B.2 USER INTERFACE</td>
<td>192</td>
</tr>
<tr>
<td>B.3 COMPUTATION OF MELTING THERMODYNAMICS</td>
<td>194</td>
</tr>
<tr>
<td>B.4 R CODE</td>
<td>198</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1-1 Comparison of Clinical Diagnostic Assays \(^a\) ................................................................. 8

Table 1-2 Unified nearest-neighbour parameters (\(\Delta H_0\) and \(\Delta S_0\)) for the duplexation of complementary base-pair doublets in 1M NaCl[135] ................................................................. 30

Table 1-3 Nearest-neighbour \(\Delta G_{37}\) increments (kcal mol\(^{-1}\)) for internal single mismatches next to Watson-Crick pairs in 1M NaCl \(^a\) ................................................................. 36

Table 2-1 Calorimetric and two-state thermodynamic parameters .................................................. 53

Table 2-2 Melting thermodynamics for the LNA-substituted duplex 5’- ttcata\(\text{C}gt\)-3’/5’- acgctatgaa-3’ .................................................................................................................. 62

Table 2-3 Net (incremental) enthalpic and entropic contributions of an LNA substitution \(^a\) ........ 70

Table 3-1 Melting thermodynamic data collected by UV spectroscopy for a library of complementary duplexes of either pure DNA or bearing different patterns of LNA substitutions on one strand. .................................................................................................................. 88

Table 3-2 Melting thermodynamic data collected by UV spectroscopy for the “learning set” of complementary duplexes bearing an LNA substitution within both strands. ................................. 101

Table 3-3 Melting thermodynamic data collected by UV spectroscopy for a set of duplexes having the pure-DNA sequence 5’-ggaacaagatgc-3’ and varying LNA content. .............................. 114

Table 3-4 \(\Delta\Delta G_{\text{hyper}}\) – (\(j\)) parameters and their respective standard deviations when regressed to either the “learning set” (Table 3-2) alone, or the combined “learning set” and “testing set” (Table 3-2 and Table 3-5). .................................................................................................................. 118

Table 3-5 Melting thermodynamic data collected by UV spectroscopy for the “testing set” of complementary duplexes bearing an LNA substitution on one or both strands. .......................... 120
Table 3-6  Summary of overall prediction performance of the extended SBT model (“learning set” regressed $\Delta \Delta G_{\text{hyper}} - (j) o$ parameters) when sequentially applied to the learning set, testing set, and combined set of duplex sequences reported in Table 3-2 and Table 3-5........ 126

Table 3-7 Performance of the extended SBT model when applied to a set of duplexes bearing various LNA-substitution patterns in both strands$^a$ .......................................................... 128

Table 4-1 Incidence of $\text{BRAF V600}$ Mutations and Sensitivity to Targeted Inhibitors$^a$ .......... 134

Table 4-2 Primers and probes used in reactions A and B of the $\text{BRAF V600}$ assay$^a$ ............. 145

Table 4-3 UVM derived and model predicted melting temperatures for dual-labeled $\text{BRAFV600}$ PCR hydrolysis probes$^a$ .............................................................................................................. 147

Table 4-4 Comparison of $T_m$ from UVM derived experiments and competing model predictions$^a$ ........................................................................................................................................... 148

Table 4-5 Model predicted melting temperatures for LNA containing primer-probe (pp) duplexes$^a$........................................................................................................................................... 149
LIST OF FIGURES

Figure 1-1  Molecular diagnostic assay schema (left) and their assay output for the case of a BRAF V600E mutation (right) are shown for various technologies. A) In Immunohistochemistry a sectioned and slide mounted tumour sample is stained leveraging primary (1°) antibody (Ab) – antigen interactions and a secondary (2°) enzyme-linked antibody responsible for colour localization. B) Sanger sequencing depends on the stochastic termination of primer extension by labeled nucleotides (ddNTPs) that can be separated by length to yield a colour coded sequence. C) The Illumina NGS platform ligates adapters to each DNA fragment enabling immobilization after PCR amplification. Each dNTP is color-linked and sequential rounds of single dNTP addition and 2D colour imaging yields a sequence following reassembly. D) Pyrosequencing monitors the release of light associated with the incorporation of a dispensed nucleotide. Its signal strength quantifies the number of nucleotides added and combined with the dispensation order, the sequence can be revealed................................................................. 7

Figure 1-2 A) Thermo-cycling protocol and B) PCR scheme for one amplification cycle. Reaction steps are delimited by black horizontal lines and proceed downward with time and cycle progression. First, the double stranded template present i-th cycle, [ds Template]_{ci}, (black lines) are denatured at 95°C before rapid cooling to a temperature $T_a$ at which temperature the forward and reverse primers, labeled FP and RP respectively (short blue lines) hybridize to the target allele (along with a TaqMan® probe, if added to the reaction in the case of real-time qPCR). Primers are extended by Taq polymerase, creating a complementary strand (long blue line) for each template, doubling the strands of DNA at each next (i+1-th) cycle. C) The extension step may be modified in the case of real-time qPCR by the addition of target specific
hydrolysis probe covering a SNP (red triangle) or dsDNA binding dyes (green dots) to generate a fluorescent signal commensurate with the amount of a specific target gene or total dsDNA respectively D) Amplification curves cross the threshold fluorescence at a quantification cycle, $C_q$, inversely proportional to its DNA starting quantity (i.e. low abundance samples have higher $C_q$s) and allow for comparison between to curves by computing $\Delta C_q$ [56].

**Figure 1-3** High-resolution melt (HRM) analysis monitors a sequence-specific melting profile associated with the thermally induced release of a dsDNA binding dye.

**Figure 1-4** Allele specific (AS) PCR enhances discrimination between target and non-target alleles. A) An AS dual labeled hydrolysis probe directed against a MT template (*e.g.* BRAF $V600E$) forms a t-t mismatch with the non-target WT template to destabilize the probe while it duplexes strongly with its complementary target allele. Shorter chemically modified probes are more strongly destabilized by a mismatched base pair and are thus less likely to cross-hybridize with the WT template (false positive). B) AS primers rely on a mismatch between the terminal 3’ base of the AS primer (pink) and the WT template (light grey) to disfavour the amplification of that allele. The result is a difference ($\Delta$) in $C_q$ between the MT (solid line) and WT (dotted line) alleles, $\Delta C_q (\text{MT-WT})$, which can be enhanced by adding modified bases (blue) at or near the 3’ end of the primer to improve differentiation between the two amplification curves thus improving the sensitivity of the assay.

**Figure 1-5** The structure (above) and typical conformation (below) of DNA, RNA, and LNA are shown. Typically, DNA and RNA nucleotides oscillate between C2’-endo and C3’-endo conformations and are less and more populated in the C3’-endo conformation respectively. LNA is locked in the C3’-endo conformation by a 2’-O, 4’-C-methylene bridge (red). Figure adapted from Campbell and Wengel [99].
The evaluation of equations (1-2)-(1-4) for the non-self-complementary duplex above first sums the contribution from each of nearest neighbour doublet parameters from Table 1-2 above (contained in the bolded box) as it moves left to right; counts the number the terminal base-pairs; and finally the evaluation corrects for duplex symmetry. The \( T_m \) prediction is computed for the non-self complementary duplex (\( \Delta S_{\text{sym}}=0 \)) with 5\( \mu \)M total strand concentration in 1M NaCl. \( \Delta H_{\text{DNA}} \) and \( \Delta S_{\text{DNA}} \) used in equation (1-2) are defined for the dissociation reaction rather than the duplexation reaction and thus have opposite signs.

A standard locked nucleic acid (left) can be modified by replacing its natural nucleosidic base with one of several modified base chemistries (right) to further enhance duplex stability.

Inability to match the dynamic range of \( C_p^{\text{ex}} \) values recorded in the sample and reference buffer scans leads to nonlinearity in the pre-transition baseline of the reference buffer-normalized thermogram. Generally, if \( C_p^{\text{ex}} \) values for the sample and reference buffer are close at each temperature, stable linear baselines (solid line) are observed.

UVM data for LNA-substituted duplex 5’-\( \text{gaDacagttDaag-3'/5'-ctttaactgtttc-3'} \) obtained at a \( C_T = 75 \) \( \mu \)M in buffer containing 1M NaCl, 10 mM \( \text{Na}_2\text{HPO}_4 \) (pH=7.0), and 1 mM \( \text{Na}_2\text{EDTA} \). (A) Raw \( A_{260nm}(T) \) data for sample and reference buffer, (B) reference buffer-normalized \( A_{260nm}(T) \) data, (C) \( d^2(A_{260nm})/dT^2 \) values with regions where \( d^2(A_{260nm})/dT^2 = 0 \) identified, and (D) \( \alpha(T) \) data computed using standard baseline fit.
**Figure 3-1** Melting transition measured by UV spectroscopy at 260 nm for the complementary duplex 5’-ctaacgGatgc-3’/5’-gcacctcgttag-3’ at a total strand concentration $C_T$ of 7.5 μM and in aqueous buffer (pH 7.0) containing 1 M NaCl, 10 mM Na$_2$HPO$_4$, and 1 mM Na$_2$EDTA. (A) Raw melting transition data; the slope ($m$) and intercept ($b$) of the pretransition (fully duplexed) and post-transition (fully dissociated) states are shown. (B) Pretransition and post-transition baseline normalized representation of the same melting transition data (diamonds), where $\theta$ is the fraction of strands in the single stranded state; the two-state thermodynamic model fit (solid curve) is also shown and superimposes the experimental data.

**Figure 3-2** General structure of the library of complementary duplexes created to study changes in thermal stability arising from an LNA substitution within each strand. Standard DNA nucleotides and LNA-substituted nucleotides are indicated by circles and squares, respectively. The position of each LNA on the antisense strand is indexed relative to a selected LNA (arrow) on the sense strand by the value of $j$, where $-4 \leq j \leq 3$ in this study. In this example, the duplex has an LNA:LNA base pair ($j = 0$), and oppositely oriented DNA:LNA base pairs at positions $j = 3, 2, -1, \text{ and } -3$ of the antisense strand.

**Figure 3-3** Hyperstabilization of duplexes containing a single LNA substitution on each strand defined in terms of $\Delta \Delta T_m, LNA$ (see equation (3-20) below); the value of index $j$ is as defined in Figure 3-2. A $\Delta \Delta T_m,LNA$ positive in value indicates that the specific spacing of oppositely oriented LNA:DNA base pairs hyperstabilizes the duplex relative to the sum of the incremental stability enhancements provided by each LNA:DNA base pair alone.

**Figure 3-4** $\Delta \Delta T_m$ depends linearly on $\Delta \Delta G_{hyper} - (j) o$ (whole data set derived model parameters in Table 3-4) as required by equation (3-10).
Figure 4-1 Our novel BRAF V600 mutation specific assay relies on the combined use of a set of AS primers (A) and AS probes (B). Since the 3’ terminal nucleotide (orange) of either AS primer is complementary to a distinct group of three mutant alleles that sharing a mutation at either nucleotide 1799 or 1798, that group of alleles is amplified. The specific mutant template within that amplified pool is identified by one of the multiplexed AS probes directed against that template.

Figure 4-2 A three tube assay allows for the quantification of mutant frequency and genotyping of both common and rare BRAF V600 mutations. Common forward and reverse primers first detect the total amount of all BRAF alleles in a sample (reaction A). In separate wells, two independent allele specific reactions (B and C) detect the identity and amount of each mutation type present in each tube. Comparison of output fluorescence curve from the consensus probe in reaction A (total BRAF) with the either allele specific reaction B or C (MT BRAF) allows the computation of mutant frequency.

Figure 4-3 A) An LNA (square) AS primer is combined with LNA AS probes by leveraging LNA interactions on both strands to generate fluorescence. Step 1) An LNA AS forward primer (FP) and common reverse primer (RP) bind to the template DNA (black) at the SOV (triangle) or common region respectively. Step 2) The LNA bearing primer generates an LNA bearing amplicon. Step 3a) the antisense strand serves as a template for the LNA AS primer to bind while to the LNA bearing sense strand template (blue; step 3b) a very short AS LNA modified DNA dual labeled hydrolysis probe may bind, via opposing LNA interactions, and will be hydrolyzed (step 4) upon extension of the common reverse primer. B) Schematic for LNA AS Primer Probe Technology. Both the AS primer and probe interact with the SOV to provide discrimination. The probe length, and overlapping region shared with the AS primer, \( d_p \) and \( d_O \), respectively, as well...
and LNA content must be tuned to achieve robust discrimination. C) The PCR reaction may be hindered by the formation of a stable primer-probe duplex which should be avoided, in part by keeping $dO$ low.

**Figure 4-4** Real time qPCR results for the new BRAF V600 AS PCR assay. BRAF V600E, E2, D, K, R, and M are all clearly distinguished at 10% MT frequency (A to F respectively). Both Rx A and B/ C contain a FAM reporter (blue) that are used for quantification and the AS Rx B/C also contains a different reporter dye (Hex, green; TexasRed, Red; Tye665, Magenta) to indicate the mutation genotype (Table 4-2).

**Figure 4-5** Serial dilutions in buffer (A) and in a background of WT templates (B) allow the computation of i) the PCR reaction efficiency (inset) from the slope of the $Cq$ vs starting DNA quantity curve, and ii) the limit of detection respectively. A selectivity of $10^3$ (1 MT in $10^3 = 0.1\%$) is determined from plot B for BRAF V600E by comparing the $Cq$ of the lowest distinguishable dilution of BRAF V600E from pure WT.
### LIST OF SYMBOLS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi - \pi^*$</td>
<td>Pi-Pi interaction</td>
</tr>
<tr>
<td>$\alpha(T)$</td>
<td>Fraction of strands in the duplexed state</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>Incremental change</td>
</tr>
<tr>
<td>$A_{260\text{nm}}(T)$</td>
<td>Absorbance at 260 nm; $\text{Au}$</td>
</tr>
<tr>
<td>$\Delta C_p$</td>
<td>Change in heat capacity; $\text{cal mol}^{-1} \text{K}^{-1}$</td>
</tr>
<tr>
<td>$\Delta C_p^{\text{bp}}$</td>
<td>Change in heat capacity per base pair; $\text{cal mol}^{-1} \text{K}^{-1} \text{bp}^{-1}$</td>
</tr>
<tr>
<td>$C_p^{\text{ex}}$</td>
<td>Excess heat capacity; $\text{cal K}^{-1}$ or $\text{J K}^{-1}$</td>
</tr>
<tr>
<td>$C_q$</td>
<td>Threshold cycle number- quantification cycle</td>
</tr>
<tr>
<td>$\Delta C_q(\text{MT-WT})$</td>
<td>Difference in quantification cycle of the mutant and wild-type templates</td>
</tr>
<tr>
<td>$C_T$</td>
<td>Total strand concentration; $\text{M}$</td>
</tr>
<tr>
<td>$E$</td>
<td>PCR doubling efficiency</td>
</tr>
<tr>
<td>$\Delta G$</td>
<td>Change in Gibb’s free energy change; $\text{kcal mol}^{-1}$</td>
</tr>
<tr>
<td>$\Delta H$</td>
<td>Change in enthalpy; $\text{kcal mol}^{-1}$</td>
</tr>
<tr>
<td>$K$</td>
<td>Equilibrium constant</td>
</tr>
<tr>
<td>$n$</td>
<td>Stoichiometry</td>
</tr>
<tr>
<td>$dP$</td>
<td>Difference in applied power; $\text{cal sec}^{-1}$ or $\text{J sec}^{-1}$</td>
</tr>
<tr>
<td>$\Delta S$</td>
<td>Change in entropy; $\text{cal mol}^{-1} \text{K}^{-1}$</td>
</tr>
<tr>
<td>$T_a$</td>
<td>Annealing temperature; $^\circ\text{C}$</td>
</tr>
<tr>
<td>$T_m$</td>
<td>Melting temperature; $^\circ\text{C}$</td>
</tr>
<tr>
<td>$T_{\text{ref}}$</td>
<td>Reference temperature defined as 53 $^\circ\text{C}$ in the SBT model; $^\circ\text{C}$</td>
</tr>
<tr>
<td>$T_{\text{max}}$</td>
<td>Temperature corresponding to the maximum value of $dA_{260\text{nm}}(T)/dT$; $^\circ\text{C}$</td>
</tr>
<tr>
<td>$\Delta T_{mLNA}$</td>
<td>Change in melting temperature due to LNA substitution; $^\circ\text{C}$</td>
</tr>
<tr>
<td>$T_o$</td>
<td>Temperature at a reference state; $\text{K}$</td>
</tr>
</tbody>
</table>
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Adenine</td>
</tr>
<tr>
<td>Ab</td>
<td>Antibody</td>
</tr>
<tr>
<td>ALK</td>
<td>Anaplastic lymphoma kinase</td>
</tr>
<tr>
<td>AS</td>
<td>Allele specific</td>
</tr>
<tr>
<td>BCR-ABL</td>
<td>Fusion gene produced by translocation of a segment of the Ableson murine leukemia viral oncogene (ABL) to the breakpoint cluster region (BCR) on chromosome 22</td>
</tr>
<tr>
<td>bp</td>
<td>Base-pair</td>
</tr>
<tr>
<td>BRAF</td>
<td>V-Raf murine sarcoma viral oncogene homolog B</td>
</tr>
<tr>
<td>c</td>
<td>Cytosine</td>
</tr>
<tr>
<td>CD</td>
<td>Circular dichromism</td>
</tr>
<tr>
<td>CML</td>
<td>Chronic myeloid leukemia</td>
</tr>
<tr>
<td>CGH</td>
<td>Comparative genomic hybridization</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>ddNTP</td>
<td>Dideoxynucleotide-triphosphate</td>
</tr>
<tr>
<td>dNTP</td>
<td>Di-nucleoside triphosphate</td>
</tr>
<tr>
<td>DSC</td>
<td>Differential Scanning Calorimetry</td>
</tr>
<tr>
<td>dsDNA</td>
<td>Double stranded DNA</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EGFR</td>
<td>Epidermal growth factor receptor</td>
</tr>
<tr>
<td>FFPE</td>
<td>Formalin fixed, paraffin embedded</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence In Situ Hybridization</td>
</tr>
<tr>
<td>FP</td>
<td>Forward primer</td>
</tr>
<tr>
<td>g</td>
<td>Guanine</td>
</tr>
<tr>
<td>gDNA</td>
<td>Genomic DNA</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>HRM</td>
<td>High Resolution Melt</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>IHC</td>
<td>Immunohistochemistry</td>
</tr>
<tr>
<td>KRAS</td>
<td>Kirsten rat sarcoma viral oncogene</td>
</tr>
<tr>
<td>LNA</td>
<td>Locked nucleic acid</td>
</tr>
<tr>
<td>LOD</td>
<td>Limit of Detection</td>
</tr>
<tr>
<td>MT</td>
<td>Mutant</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>NGS</td>
<td>Next Generation Sequencing</td>
</tr>
<tr>
<td>NNT</td>
<td>Nearest neighbor Thermodynamic (model)</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>qPCR</td>
<td>Real-time quantitative PCR</td>
</tr>
<tr>
<td>RFU</td>
<td>Relative fluorescence units</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>RP</td>
<td>Reverse primer</td>
</tr>
<tr>
<td>SBT</td>
<td>Single base thermodynamic (model)</td>
</tr>
<tr>
<td>siRNA</td>
<td>Small interfering RNA</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphisms</td>
</tr>
<tr>
<td>SOV</td>
<td>Site of variance</td>
</tr>
<tr>
<td>ssDNA</td>
<td>Single stranded DNA</td>
</tr>
<tr>
<td>t</td>
<td>Thymine</td>
</tr>
<tr>
<td>Taq</td>
<td><em>Thermus aquaticus</em></td>
</tr>
<tr>
<td>UVM</td>
<td>UV monitored melt</td>
</tr>
<tr>
<td>WT</td>
<td>Wild-type</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

In any endeavor, the ability to conceive and pursue ideas is paramount. I first learned by example to be inquisitive from my mother who would look up any question I had, if she did not already know it offhand. Eventually, I became a nuisance and she encouraged me to look them up myself; and I did. Graduate studies are not an extension of undergraduate courses but an entirely distinct endeavor. No more will I find in the back of the book; often there is no book, the questions are more consequential, and only the first to answer them clearly and correctly is fully credited. It is a brutal and exhilarating practice and one that is rarely practiced alone. For every author on any serious work, their family, colleagues, friends, and others have surely contributed something to its creation.

My family has been unwavering in their support of my studies, health, and wellbeing. If it were not for their support in the years during and before this work, I certainly would not have had the opportunity to begin this work or the strength to sustain it. They have my infinite gratitude. To my mother, I extend to you my heartfelt gratitude for your efforts, ethic, and patience, in setting an example for me to follow as my character was taking shape. I am also grateful to my uncle Leo – for so many things but with respect to this work– for providing me with both a humane lens through which I could interpret my findings but also an impetus to learn; he will not be forgotten.

I offer my sincere and enduring gratitude to my mentor and principal supervisor Dr. Charles Haynes for showing me the toolkit required for the successful and practical pursuit of new ideas; those skills extend beyond academia and have certainly shaped my professional future. His curiosity and insightful questioning continues to spur on new ideas and interesting discussion which challenges me to think more deeply. Finally, I would like to add that I am exceptionally grateful for his support during an extended spell of sickness; such a degree of generosity is both rare and invaluable to me. I extend special thanks to Dr. Louise Creagh for sharing with me her time and expertise in the Center for Biothermodynamics and the CHBE teaching laboratories while I learned new techniques and refined known ones.
For his continued fervor for science and infectious interest in its details, I offer my thanks to my mentor and laboratory supervisor, Dr. Curtis Hughesman. His patient mentorship and challenges have made me think more flexibly and approach to problems with creative and pragmatic freedom. I am grateful for the many interesting discussions we have shared and the ideas he helped me bring to life.

I have had the great pleasure of working with talented and insightful Ph.D colleagues in the laboratory: Dr. Eric Ouellet, Dr. Louise Lund, and Roza Bidsharhi. Collectively, they set a sterling example for me of how to conduct research and their friendship, shared experiences, and interesting discussions were the highlight of my time at UBC. Likewise, the undergraduate students that I have had the chance to work with: Olivia Marais, Xinbo Cheng, and especially Vincent Kao, their questions pushed me to improve my understanding of my field. I would also like to extend my thanks to Dean Attali for his insights, suggestions, and contributions to the development of the R/Shiny web application.

As they entertained me explaining what I do with my time, I would like to thank my friends for letting me bend their ears. For the times I forgot to thank you for taking my mind off work and helping me recharge; thank you. In particular, Sebastien Ennis, I am grateful for our comedic and insightful discussions; it appears the law and engineering have much in common. I am also grateful to Heidi Schafer for her support, interesting conversations, and time spent adventuring.

Finally, I recognize that it is not only the immediate sphere of people surrounding me that have contributed to my success. For example, friendly chats with food services staff at the Delly sandwich counter that brighten my day; the governance that make the city streets safe for me to walk or cycle to school; the opportunities I was awarded as a result of privilege; or the unceded Musqueam territory situating UBC that fosters the school’s growth and beauty. I can only do my best not to overlook these factors and their efforts, and offer them my acknowledgment and gratitude.
CHAPTER 1: INTRODUCTION

This introductory chapter provides i) an overview of the context and goals of this thesis followed by ii) a discussion of current clinical diagnostic challenges and technology developments that motivate this work along with iii) a review of recently developed thermodynamic models used to guide the design of novel clinical diagnostic assays.

1.1 THESIS OVERVIEW

The dual role of deoxyribonucleic acid (DNA) as the sequence and storage molecule for the genetic code has motivated research across the health sciences to better understand its structure and function, including its role in disease. Comprehensive efforts to map and understand the human genome [2, 3] have revealed many genes and genetic aberrations (e.g. (inherited) single nucleotide polymorphisms and (acquired) somatic mutations, translocations, copy number variations, etc.) that are causally implicated in cancer [4]. Discovery of these cancer activating genes, typically referred to as oncogenes, has prompted the development of technologies to detect them (e.g. quantification of the BCR-ABL fusion gene in genomic DNA isolated from blood is used clinically to diagnose and monitor chronic myeloid leukemia (CML) [5]). Patients may thereby be triaged by a tumour’s genetic profile and associated sensitivity to particular therapeutics (e.g. targeted kinase inhibitors). Several such predictive biomarkers are currently used to identify those patients who are likely to respond well to a particular treatment, thus enabling the setting of an effective treatment regimen and proper clinical management of the disease [5]. For example, metastatic melanoma patients harbouring a BRAF V600 coding mutation (i.e. a somatic mutation in the BRAF gene at codon 600, which, in healthy individuals, codes for the amino acid Valine (V)) produce an altered form of the BRAF protein kinase that drives the cancer phenotype; fortunately, melanomas carrying a BRAF V600 mutation are responsive to treatment with a mutant-BRAF tyrosine kinase inhibitor [6]. Other somatic (non-germ line) mutations may likewise reduce sensitivity to a standard treatment, thus requiring alternate therapies that overcome that resistance and control tumour growth. For example, the BCR-ABL fusion oncogene is formed from a somatic reciprocal translocation and is known to be causative of CML. However, an acquired T315I mutation in that oncogene renders the patient
insensitive to front-line treatment with imatinib. CML patients carrying the T315I mutation are therefore best treated with 3rd-line ABL kinase inhibitors [7].

The reliable detection and quantification of somatic mutations and the oncogenes formed from them is therefore now an essential capability of all cancer testing laboratories, and continued research into cancer genomics is expected to increase both the number and complexity of those assays [8]. Current methods used by clinics to detect oncogenic mutations are unlikely to meet that growing challenge, as many will become cost challenged within the soon-to-be enacted reimbursement limits defined by the Affordable Healthcare Act of the USA[9], while others do not provide the sensitivity and/or specificity needed to detect a cancer early in its pathogenesis — when it is generally most effectively treated. Thus, there is a need for robust diagnostic assays that may detect mutation-specific oncogenes with high sensitivity, specificity, and a sufficiently low limit of detection to allow accurate and timely clinical management of cancer. The goals of this thesis are 1) to provide the data needed to extend a recently reported molecular thermodynamic model to enable its application to the accurate prediction of melting thermodynamics (stabilities) of duplex DNA harbouring locked-nucleic-acid (LNA) substitutions in each strand, and 2) to demonstrate that the extended model can be used to design low cost, yet sensitive and specific polymerase-chain-reaction (PCR) based assays against clinically relevant mutations.

1.2 CLINICAL DETECTION OF CANCER-RELATED GENOMIC MUTATIONS

Clinics currently employ a variety of methods to detect genetic mutations prognostic of disease initiation and progression, or theranostic of proper course of treatment [10-12]. Those assays are known to facilitate early cancer detection, molecular profiling to better define prognosis, and effective tailoring of therapy to individual patients. In addition, they may be used in post-treatment monitoring of residual disease [13]. Depending on the clinical context (e.g. diagnosis or monitoring) and cancer type, one or more diagnostic technologies may be used to detect the relevant biomarkers. For example, all patients with advanced lung adenocarcinomas
are recommended to be tested, prior to therapy with epidermal growth factor receptor (EGFR) or anaplastic lymphoma kinase (ALK) inhibitors, for both mutations in EGFR and translocations that disrupt ALK [14, 15]. Testing for changes in EGFR and in ALK is currently achieved using different technologies: PCR-based assays against EGFR and fluorescence in situ hybridization (FISH) assays against ALK. This follows in part from the fact that the clinically relevant mutations in EGFR are single nucleotide polymorphisms (SNPs) within specific exons of that gene, while alterations in ALK are the result of large-scale paracentric rearrangement of human chromosome 2 [14, 15]. Additional technologies have been adopted clinically to analyze other cancer-relevant genetic aberrations, including copy number variations (e.g. SNP and comparative genomic hybridization (CGH) arrays) [16], reciprocal translocations (e.g. FISH) [17], or DNA methylation (e.g. single-molecule real-time (SMRT) sequencing) [18]. However, PCR-based assays are by far the most commonly used technology for such analyses due in part to their ability to inexpensively detect somatic point mutations and/or SNPs prognostic or theranostic of cancer [4].

1.2.1 CHALLENGES IN DETECTING GENETIC ABERRATIONS IN CANCERS

All assays against point mutations share a common need to specifically detect (and in many cases quantify) a rare allele (gene sequence) harbouring an oncogenic mutation within a specimen (genomic DNA (gDNA) isolated from a tissue biopsy or liquid biopsy (e.g. blood, urine)) containing a much higher abundance of the corresponding wild-type (WT) allele (i.e. the normal sequence of the same gene in healthy individuals or in the healthy tissue components of the specimen). In addition to the large disparity in the abundance of the related alleles, meeting this need is challenged by the fact that the minority allele may differ from the wild-type allele by as little as a single nucleotide [10]. Current DNA-based molecular diagnostics technology is therefore prone to both false positives and failed runs that compromise clinical decisions, particularly in cases where the quantity and/or quality of the specimen is low.

To fix ideas, we can consider the general requirement of a clinical assay to provide for sufficient enrichment and selective detection of the minority allele in a sample containing a $10^4$
fold or greater excess of the related WT allele [10]. A standard blood sample is ca. 7 mL and contains 10 to 100 ng/mL gDNA. That corresponds to ~30,000 copies of the human genome per 50 ng of gDNA [19], each copy of which contains a minimum of one copy of the gene sequence to be analyzed. To facilitate initial diagnosis of cancer, a clinic might therefore hope to detect in a 50 ng gDNA sample a single copy of a MT allele within a background of 30,000 copies of the WT allele, which corresponds to a mutant frequency of $1 \times 10^{-5}$ [20]. This sets the clinical criterion for unequivocal detection of a minority allele [21], including within short-lived cellular DNA fragments shed into the blood (circulating cell-free DNA), which is garnering interest as a low-invasive means to repeatedly monitor patients carrying clinically relevant gene mutations (e.g. KRAS G12D, BRAF V600E, EGFR T790M) [22]. Likewise, in solid tumour biopsies, the fraction of tissue harbouring the mutant allele can be very low [23], and the gDNA extracted from those specimens may be needed for other genetic tests. Clinics therefore generally need to detect mutations using 50 ng or less of gDNA recovered from a biopsy [24].

Sample DNA quality can affect assay performance. Formalin fixed, paraffin embedded (FFPE) patient samples are extensively used in archiving, diagnosing, and studying (including retrospective studies) solid tumour biopsies [24-26]. Unfortunately FFPE treatment and storage is known to induce DNA damage (e.g. significant DNA fragmentation) and inhibit Taq processing fidelity, thereby introducing artifacts that can bias the outcome of a molecular diagnostic assay [25].

Finally, for routine clinical application, any diagnostic technology must balance affordability (i.e. instrument and reagent costs, technician wages, instrument processing time, etc.) with the ever-increasing need for higher throughput and the richer data content required to make an informed clinical decision. No one technology satisfies all of these technical and economic criteria for all required assays. As a result, many different technologies are used in clinics to detect cancer-associated mutations.
1.2.2 CLINICAL DIAGNOSTICS USED TO DETECT POINT MUTATIONS IN CANCER

As noted, several methods have been developed to detect genetic aberrations prognostic or theranostic of cancer. Some require *a priori* knowledge of the mutation of interest and detect specific mutations at certain loci, while others may permit *de novo* discovery of driver mutations [12]. Those methods commonly used either as clinical diagnostic technologies or research tools are described below, with PCR-based technologies described in greatest depth.

1.2.2.1 IMMUNOHISTOCHEMISTRY

Widely used in clinical assays [27, 28], immunohistochemistry (IHC) leverages the exquisite specificity of antibody-antigen interactions to visualize (differentially stain) tumorous tissue within a slide-mounted section of a core (biopsy). The staining pattern is typically analyzed and graded by an experienced pathologist, yielding semi-quantitative but nevertheless clinically actionable results [29]. As illustrated in Figure 1-1A, the IHC method uses an unlabeled primary (1°) antibody to specifically bind an antigen in the sectioned tissue. Staining with a labeled secondary (2°) antibody against the 1° antibody then permits visualization of tumorous tissue under a microscope. Common labels include a fluorescent reporter (immuno-fluorescence) or an active enzyme (IHC) whose activity produces a color in the presence of a chromogenic substrate. Because it assays a gene product, IHC is not subject to the challenges associated with FFPE-induced DNA damage and extraction; the epitope (*i.e.* interacting portion of an antigen) queried by IHC generally remains intact after formalin fixation [30]. But, in general the IHC method is low throughput (~ 2 to 3 days per test) [31] and offers relatively poor sensitivity. The limit of detection (LOD) of IHC assays against *BRAF* V600E is typically 5% mutant frequency [32], but is higher (often above 15%) for more poorly expressed antigens. Moreover, acquiring reliable results by IHC testing requires a very high level of technical expertise, in part because staining quality is sensitive to variations in procedure (*e.g.* tissue preparation), standardization methods are not available, and not all antibodies are effective [30].
The general performance metrics of IHC, as well as other technologies to be described, are provided in Table 1-1.

As an example, the *BRAF V600E* mutation-specific IHC assay is widely used to guide treatment of melanoma with BRAF inhibitors [33, 34]. That assay provides modest sensitivity in detecting of *BRAF V600E* (LOD of 5%), in part because the cross-reactivity of the 1° antibody promotes false positives [31, 32].
Figure 1-1  Molecular diagnostic assay schema (left) and their assay output for the case of a BRAF V600E mutation (right) are shown for various technologies. A) In Immunohistochemistry a sectioned and slide mounted tumour sample is stained leveraging primary (1°) antibody (Ab) – antigen interactions and a secondary (2°) enzyme-linked antibody responsible for colour localization. B) Sanger sequencing depends on the stochastic termination of primer extension by labeled nucleotides (ddNTPs) that can be separated by length to yield a colour coded sequence. C) The Illumina NGS platform ligates adapters to each DNA fragment enabling immobilization after PCR amplification. Each dNTP is color-linked and sequential rounds of single dNTP addition and 2D colour imaging yields a sequence following reassembly. D) Pyrosequencing monitors the release of light associated with the incorporation of a dispensed nucleotide. Its signal strength quantifies the number of nucleotides added and combined with the dispensation order, the sequence can be revealed.
<table>
<thead>
<tr>
<th>Methodology</th>
<th>Detection of unknown mutations</th>
<th>Analytical Sensitivity / Specificity</th>
<th>Min. DNA input per test (ng)</th>
<th>LOD (% MT frequency)</th>
<th>Turnaround time</th>
<th>Cost</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Immunohistochemistry (IHC)</td>
<td>No</td>
<td>Target dependent*</td>
<td>N/A*</td>
<td>5[^32]</td>
<td>Low [^31, 32]</td>
<td>Low [^31, 32]</td>
<td></td>
</tr>
<tr>
<td>Real-time quantitative PCR (qPCR)</td>
<td>No</td>
<td>High /Average-High [^12]</td>
<td>30-125 [^31, 32, 44]</td>
<td>≤0.1-5 [^10, 32, 37, 44]</td>
<td>Low [^12, 32, 37]</td>
<td>Low [^12]</td>
<td></td>
</tr>
</tbody>
</table>

[^32] Low, 80%; average, 80-98%; high, >98%.
[^31] Low, <1 week; average, 1 week-1 month; high, >1 month.
[^33] IHC depends on protein (antibody-antigen) interactions which have variable performance; for BRAF V600E the VE1 monoclonal antobody would be recorded as High/High. Typically, 3-4μM tissue sections are mounted on a microscope slide and proteins, not DNA, serve as the basis for detection. IHC output is typically semi-quantitative (e.g. graded 0-3) and does not typically report % mutant frequency.
1.2.2.2 SEQUENCING TECHNOLOGIES

Though not (yet) used in clinics due in part to cost and sensitivity limitations, Sanger and next-generation sequencing (NGS) methods are used to analyze for known and unknown point mutations across a whole genome, a human exome (protein coding region of the genome), or specific genomic regions affected by disease [12]. Sanger sequencing [45], shown in Figure 1-1B, was integral to the completion of human genome project [46] and has found limited use in clinical genetic testing [36]. Genomic DNA is purified, fragmented and then cloned, either through the use of cloning vectors or the direct application of PCR, to replicate the single gDNA fragment into many (often billions) of copies. Sanger sequencing biochemistry is then employed, which begins with a second amplification in which each round of primer extension is terminated by the stochastic addition of a labeled non-extendible dideoxynucleotide monomer (ddNTP) then results in a mixture of amplicons of varying length whose terminal nucleotide is labeled with a reporter dye that is specific to that nucleotide. Terminal adenines, for example, might be labeled with a red reporter, while terminal guanines carry a green reporter. The ensemble of labeled amplicons produced from each fragment can be separated by length, for instance by capillary electrophoresis, with the fluorescent signal from each amplicon length determining the terminal base for that amplicon [47]. The sequence of terminal bases then provides the sequence of the associated fragment.

Next-generation sequencing (NGS) improves upon the automated Sanger sequencing method by coupling DNA sequencing with real-time imaging during primer extension to enable genomic sequence data to be collected at high throughput and then reassembled by statistical bioinformatics to uncover rare and common sequences across the whole genome [12, 48]. Of the many NGS methods and related instruments [48], most utilize the following protocol: generation of DNA sequencing libraries by clonal amplification and PCR, spatial segregation of the amplified templates in a manner that eliminates the need for gel or capillary electrophoretic separation, massively parallel sequencing by synthesis (i.e., the sequence is determined by the addition of nucleotides to
the complementary strand rather than using standard Sanger chain-termination chemistry). In the popular Illumnia [49] platform (Figure 1-1C), for example, an associated base-specific fluorescent label is imaged as each nucleotide is added and then cleaved to allow incorporation of the next base. A high resolution snapshot of the surface containing the arrayed immobilized templates reveals which nucleotide was added after each cycle [48].

In pyrosequencing technology, Figure 1-1D [50], the incorporation of one of four nucleotide monomers (adenosine $a$, thymine $t$, guanine $g$, or cytosine $c$) is accompanied by the release of pyrophosphate, which is converted via an enzymatic reaction cascade to generate an amount of light commensurate with the number of nucleotides added. The light signal is recorded and any free nucleotides in the system are degraded; a fresh set of nucleotides is then added to conduct the next cycle. A ‘pyrogram’ plots the light signal intensity against the dispensation order of $a$, $t$, $g$, and $c$ nucleotides into the system to yield the DNA sequence. In the pyrogram, somatic point mutations generate a pattern distinct from that of the background allele and can thereby be distinguished with a mutant frequency of ca. 5% [32, 38].

Finally, Ion Torrent [51] operates in a chip format by registering a change in pH associated with the release of protons when a dispensed nucleotide is incorporated. Irrespective of the instrument used, current NGS technology generally detect mutations at frequencies of ca. 5% or greater. Moreover, the complexity of the method (i.e. read length, dispensing order, peak height analysis) has led to confidence issues in mutation calling. As a result, other diagnostic technologies (see below) offering better sensitivity at lower cost are preferred by clinics for disease detection and monitoring. Sequencing’s strength in disease analysis primarily lies in the fact that previous knowledge of a particular mutation is not required, and the method may be applied to all or select regions within a genome [12]. It is therefore an ideal platform for conducting fundamental research aimed at discovering new biomarkers prognostic or theranostic of cancer [52,
As costs diminish and sensitivities improve, sequencing may also find use in the analysis of complex mutation patterns (in CML, for instance, over 80 different $BCR-ABL$ alleles have been observed, and NGS may prove useful in determining which of those alleles are eligible for imatinib treatment and which require treatment with 2$^{nd}$ or 3$^{rd}$ line tyrosine kinase inhibitors [7]) and copy number variations.

1.2.2.3 **PCR BASED DIAGNOSTIC TECHNOLOGIES**

PCR-based diagnostic technologies measure properties of an amplification product and/or the amplification reaction itself to detect and quantify a gene sequence of interest. The polymerase chain reaction and two PCR based technologies, high resolution melt analysis and quantitative real-time PCR, are described below.

1.2.2.3.1 **POLYMERASE CHAIN REACTION (PCR)**

PCR amplification of a target sequence (known as the template) within a genomic DNA sample is typically achieved by adding to that sample a pair of primers, the four nucleotide monomers ($a$, $t$, $c$, and $g$), and a heat-stable DNA polymerase (often some version of $Taq$ polymerase) [54, 55]. The pair of short ca. 20 bp single-stranded DNA-oligonucleotide primers is comprised of a forward primer (FP) and a reverse primer (RP). The FP forms a complementary duplex with the anti-sense strand, and the RP with the sense strand of the template. The sequence to be amplified is therefore bracketed by the two primers. Together, the two primers and template DNA are mixed in a buffer solution containing the four di-nucleoside triphosphate (dNTP) monomers, $Taq$ polymerase, and background salts (*e.g.* NaCl, Mg$^{2+}$) and then subjected to thermal cycling. A typical PCR thermal cycling protocol and the associated reactions occurring during different elements of a cycle are shown in Figure 1-2 panels A and B, respectively.
Figure 1-2 A) Thermo-cycling protocol and B) PCR scheme for one amplification cycle. Reaction steps are delimited by black horizontal lines and proceed downward with time and cycle progression. First, the double stranded template present i-th cycle, [ds Template]_i, (black lines) are denatured at 95°C before rapid cooling to a temperature $T_a$ at which temperature the forward and reverse primers, labeled FP and RP respectively (short blue lines) hybridize to the target allele (along with a TaqMan® probe, if added to the reaction in the case of real-time qPCR). Primers are extended by Taq polymerase, creating a complementary strand (long blue line) for each template, doubling the strands of DNA at each next (i+1-th) cycle. C) The extension step may be modified in the case of real-time qPCR by the addition of target specific hydrolysis probe covering a SNP (red triangle) or dsDNA binding dyes (green dots) to generate a fluorescent signal commensurate with the amount of a specific target gene or total dsDNA respectively D) Amplification curves cross the threshold fluorescence at a quantification cycle, $C_q$, inversely proportional to its DNA starting quantity (i.e. low abundance samples have higher $C_q$s) and allow for comparison between to curves by computing $\Delta C_q$ [56].

Each cycle is initiated with a thermal denaturation step at ca. 95 °C that serves to separate the duplex DNA into its two constituent template strands. The sample is then cooled to the annealing temperature ($T_a$), ca. 55-60°C, where the primers, present in
excess, hybridize to template strands. Each primer is then enzymatically extended by Taq polymerase at a temperature between 60 - 72 °C; during extension, dNTPs are added to the 3’ end of each primer following Watson-Crick base-pairing rules to create a new complementary strand (amplicon) for each template present at the start of the cycle [56]. Thus, following each cycle the amount of DNA amplicons is (theoretically) doubled, leading to an exponential amplification that continues until the reaction becomes inhibited, often at ca. 40 cycles. The presence of amplicons, and thus the sequence queried, may be detected during the amplification (real-time PCR). Instead or in addition, the final amplified DNA product may be analyzed either within the PCR instrument (e.g. high-resolution melt analysis), or using another technology (e.g. NGS, electrophoresis).

1.2.2.3.2 HIGH RESOLUTION MELT ANALYSIS

High-resolution melt (HRM) analysis monitors the sequence specific melting profile of a PCR product (~100 bp) or of a short duplex formed between that product and a sequence specific probe. The melting profile is detected from the decrease in signal generated from the release of a double stranded DNA (dsDNA)-binding fluorescent dye (e.g. LC Green®) during melting of the duplex into single stranded DNA (ssDNA); the fluorescence intensity (RFU) of the dye is recorded as a function of temperature to generate the melting curve, an example of which is shown in Figure 1-3 [40, 42]. Genotyping is then achieved by comparing the melting temperature, $T_m$, and/or the shape of the melting curve with a known standard. If, for example, a WT-specific probe covering the site of variation is added, its melting profile will depend on whether the amplicons are copies of the WT form of the gene (higher $T_m$) or carry a clinically relevant mutation (lower $T_m$); sequence-related WT and MT alleles may thereby be distinguished. In certain cases, the type of mutation may be identified based on the nature of the probe mismatch (e.g. a/a, a/c, or a/g) [40, 42].

However, challenges exist in reliably discerning mismatched alleles [40, 43]. As a result, clinics generally employ HRM as a secondary check of genotyping results.
obtained by another more reliable method. In this capacity, HRM has shown good performance \((i.e.\) specificity \(\sim 95\%\) and limit of detection (LOD) of \(\sim 5\%\)), and clinical value due to its closed-tube, inexpensive (unlabeled probe) format \([32, 40, 43]\). It is also easily paired with other diagnostic technologies \((e.g.\) qPCR (below) or sequencing) since duplex dissociation is reversible and, thus, the sample is not altered \([43]\).  

**Figure 1-3** High-resolution melt (HRM) analysis monitors a sequence-specific melting profile associated with the thermally induced release of a dsDNA binding dye. 

### 1.2.2.3.3 REAL-TIME MONITORING OF PCR AMPLIFICATION BY QUANTITATIVE PCR (qPCR)

Quantitative real-time monitoring of the polymerase chain reaction (qPCR) \([57-59]\) is a simple, inexpensive method carried out in a closed tube format that enables detection and quantification of specific sequences within a genomic DNA specimen \([56]\). Monitoring of template amplification typically leverages one of two fluorescent dye chemistries: i) intercalating dyes \((e.g.\) SYBR® Green) that fluoresce when bound to the minor or major-groove of dsDNA or, more often in a clinical setting, ii) dual-labeled hydrolysis probes \((e.g.\) Taq Man® probes) that bind to a specific sequence within the template \([56]\). In the latter detection system, *Taq* polymerase hydrolyzes the dual labeled TaqMan® probe during complementary strand synthesis, releasing the reporter dye from the quenching agent. The associated increase in total fluorescence intensity as a function of PCR cycle number may thereby be used to detect and quantify amplification of the targeted template sequence \([57]\). Alternatively, intercalating dyes bind to any dsDNA product(s) produced during amplification. In general, they are therefore less specific, in
part because they can bind to secondary amplification products (e.g., primer-primer duplexes known as primer dimers) that do not reflect the presence of the target template [56]. For either detection method, the accumulation of fluorescent signal is recorded at each cycle (Figure 1-2C) to generate an amplification curve (Figure 1-2D) plotting relative fluorescence units (RFUs) versus cycle number that enables software-assisted quantification of the starting abundance of the target template as described below.

qPCR analysis of starting template concentration is based on determination of the quantification cycle, $C_q$, which is the cycle at which an amplification curve crosses a threshold RFU value [57]. A higher initial template concentration in the starting specimen results in a lower $C_q$ value, and *vice versa*. The absolute (or relative) quantity of a target allele/sequence in a sample or frequency of a gene (or allele) is typically determined by qPCR by measuring the difference in $C_q$ values between the unknown sample, $A$, and a control sample of known starting template concentration, $B$, and then applying the relationship

$$\frac{[\text{template}_0]_A}{[\text{template}_0]_B} = (1 + E)^{(C_q,A-C_q,B)} \quad (1-1)$$

where $[\text{template}_0]_A$ and $[\text{template}_0]_B$ are the initial concentrations of double-stranded template sequence in samples $A$ and $B$, and $C_q,A$ and $C_q,B$ are the measured quantification cycles for samples $A$ and $B$. $E$ is the PCR reaction efficiency determined from the increase in RFUs with cycle number. In theory, PCR leads to an exact doubling of dsDNA templates from cycle $i$ to cycle $i+1$ [54] corresponding to an $E = 1$; in practice, the number of amplicons generated per cycle can be somewhat less than unity [56].

Several advantages have helped establish qPCR as the most widely used method for nucleic acid quantification. For rare templates, it offers reliable amplification of starting material by up to 7-8 orders of magnitude, enabling very low (<5 copies) abundance sequences to be detected. Moreover, its closed tube format can reduce cross-
contamination, making it particularly useful in clinical settings [60]. One chooses the sequence(s) targeted in a qPCR assay. Templates specific to mutation(s) of interest are therefore generally short (~100 bp), making them less prone to fragmentation during gDNA isolation/sample processing (i.e. FFPE processing of tissue samples) [25]. Finally, qPCR assays are relatively low cost, low complexity, and versatile [61].

Many qPCR assays formats are available (e.g. wild-type blocking qPCR [62], RT-qPCR for RNA amplification and identification [63], multiplex qPCR [64, 65], etc.). Among the various techniques developed to enrich and detect minority alleles, probe or primer based allele-specific (AS) qPCR is most often used, as the method can often detect mutant frequencies down to $10^{-2}$ [10, 23, 66], with lower frequencies reported on occasion [21, 67, 68]. AS-probe based qPCR assays rely on the concept that the formation of at least one mismatched base pair between a probe and a non-target template eliminates (or at least strongly disfavors) binding of the probe to that non-complementary sequence at the PCR annealing and extension temperatures. AS-primer based qPCR assays use mismatches with non-target templates to enable selective extension by Taq polymerase of only the target sequence.

For an AS probe directed against a mutant allele bearing a single point mutation, the hydrolysis probe is therefore designed to function as shown in Figure 1-4A. It is fully complementary to the mutant allele (e.g. the mutant BRAF V600E allele) and therefore forms a mismatched base pair with WT BRAF V600 and one or more mismatched base pairs with any other mutant V600 allele. Focusing on the WT V600 allele, the $t$-$t$ mismatch formed between the probe and the WT template must be sufficiently destabilizing to ensure that the probe is not bound to the WT allele at $T_a$. Likewise the mismatch(es) formed between a BRAF V600E AS probe and any other BRAF V600 mutant template (e.g. V600K) must also discourage cross-hybridization. Regrettably, that collective ability to avoid cross-hybridization to all non-target alleles has proven difficult to realize in AS probes. The method rarely provides for reliable
detection of mutant frequencies below 1% due in part to false positives generated from probe cross-hybridization [31, 32]. Various methods to improve the performance of AS probes have therefore been proposed. One is to design a companion unlabeled probe, complementary to the WT allele, that serves to block cross-hybridization of the AS probe to the WT allele (for instance), thus improving probe specificity [44]. A second method is to chemically modify the AS probe at or near the sight of variation so as to make the probe more specific to its target allele [69].

Alternatively, an AS primer can be used to preferentially amplify a target allele. Most often this is achieved by aligning the terminal 3’ base of the primer with the site of variance (SOV) (Figure 1-4B). For example, if a primer is designed against the BRAF V600E allele, a (t-t) mismatch is formed between the 3’ terminal base of the BRAF V600E AS primer and the WT BRAF allele, which serves to inhibit extension of the cross-hybridized primer by Taq polymerase [68]. Based on this design concept, AS primers comprised of pure DNA can provide adequate discrimination to permit unequivocal detection of a point mutation; selectivities ≥ 10^{-3} are reported for certain alleles[70, 71]. But this is not a generally observed, which has motivated the use of one or more modifications (e.g. LNA substitutions or additional mismatches) at or near the 3’ end of the primer to improve its selectivity [68, 72]. Indeed, primers harbouring additional mismatches have achieved selectivities of 10^{-3} [72], while even better performance (≤ 10^{-3}) has been achieved on occasion with AS primers bearing LNA substitutions at or near the 3’ terminal base [67]. However, as with pure DNA AS primers, good selectivities are only sometimes achieved using these more advanced design concepts, pointing to the need for reliable tools for designing highly sensitive and specific AS probes and primers. In the case of LNA-substituted AS primers, those modifications must also not inhibit to any great extent the efficiency of amplification of the target template, which can be a problem [73].
Though it offers high selectivity, robustness, and relative simplicity, AS qPCR is not easily applied to all classes of cancer biomarkers. SOVs within an oncogene may be spatially separated at intervals beyond the realistic span of a PCR reaction (*i.e.* < ca. 300 bp) [74]. But many genetic variants of interest, including rare alleles, occur within an amplifiable region. For example the *BRAF* gene contains upwards of 10 clinically actionable mutations within the 3 bases constituting codon 600, as well as rarer mutations in neighbouring codons 597 and 601 that have unknown clinical significance [75, 76]. For a qPCR assay to be holistically applied to that analysis, it must function to detect all clinically actionable mutations in *BRAF* V600 [77] while using minimal DNA from a patient sample. That can be challenging [37].
Figure 1-4 Allele specific (AS) PCR enhances discrimination between target and non-target alleles. A) An AS dual labeled hydrolysis probe directed against a MT template (e.g. BRAF V600E) forms a t-t mismatch with the non-target WT template to destabilize the probe while it duplexes strongly with its complementary target allele. Shorter chemically modified probes are more strongly destabilized by a mismatched base pair and are thus less likely to cross-hybridize with the WT template (false positive). B) AS primers rely on a mismatch between the terminal 3’ base of the AS primer (pink) and the WT template (light grey) to disfavour the amplification of that allele. The result is a difference (Δ) in $C_q$ between the MT (solid line) and WT (dotted line) alleles, $\Delta C_q (MT-WT)$, which can be enhanced by adding modified bases (blue) at or near the 3’ end of the primer to improve differentiation between the two amplification curves thus improving the sensitivity of the assay.

One possible way to successfully apply qPCR to these more complex analyses is by multiplexing [60, 65]. For example, multiple AS probes, each targeting a specific allele/mutation and bearing a reporter dye with a unique fluorescence spectra, are added
to a single reaction and used i) with a single set of primers to identify clustered variants or ii) with multiple primer sets to identify variants in distinct genomic regions [65, 74]. As the PCR assay complexity increases with multiplexing, so too do the design challenges such as cross-reactivity of probes with non-target templates, interactions (e.g. self-priming) between primers sets, and differences in amplification efficiencies between reactions [74]. Further, multiplexing is limited by the ca. 4 to 5 emission wavelengths detectable by current qPCR instrumentation [65].

These limitations are motivating the continued development of qPCR assay technology and instruments. An example of the more advanced use of LNA substitutions to improve assay performance, which is the focus of this thesis, is provided by Morandi et. al. [23], who employed combinations of LNA-modified primers and LNA modified hydrolysis probes to detect and quantify *BRAF* and *KRAS* mutations. Others have employed unlabeled LNA modified blocking agents to better inhibit amplification of the non-target WT allele, thereby improving assay sensitivity for the minority allele [62, 78]. Molecular beacons, an alternate probe design, adopt a secondary structure (e.g. hairpin) that unfolds upon sequence recognition, and there is now limited evidence suggesting their performance can also be enhanced by the addition of LNAs [79]. However, each of these examples was achieved empirically. For the approach to find more routine use, new tools are needed to enable prediction of the melting behavior of short DNA duplexes that contain LNA substitutions in one or both strands.

1.3 LOCKED NUCLEIC ACID (LNA) CHEMISTRY AND PROPERTIES

Applications of synthetic analogs of deoxyribonucleotides and ribonucleotides have experienced tremendous growth in the last two decades [80, 81], due in part to their ability to enhance diagnostic and therapeutic applications of structured oligonucleotides by altering their thermodynamic and/or chemical (e.g. resistance to nucleases) stability, base-pairing specificity, antisense activity, or cellular uptake [82]. The phosphodiester
bonds of natural DNA and RNA are, by design, susceptible to hydrolysis by nucleases. Short natural oligonucleotides also tend to exhibit relatively poor plasma protein binding and transport within the cardio-vasculature, while longer oligonucleotides complex their target slowly. For oligonucleotides used as antisense agents, slow binding kinetics can be exacerbated by the tendency for their mRNA targets to be highly structured. As a result, unmodified DNA or RNA can exhibit poor pharmacokinetics and relatively rapid excretion when delivered as a parenteral therapeutic.

These limitations can be overcome through appropriate chemical modifications that serve to increase the specificity of the oligonucleotide for its target sequence, the stability of the complex/duplex formed, and/or the resistance to nuclease-catalyzed degradation. Reported chemical modifications to DNA and RNA now number in the 100s [83, 84] and include chemistries that serve to alter helix structure [85], expand the genetic alphabet beyond the natural nucleosides a, t, g and c [86, 87], or alter cellular phenotypes [88]. Unique biological functions, and improved diagnostic specificity or therapeutic potency may thereby be realized [68, 89-91].

Many of the chemical modifications reported to date alter the backbone [92], furanose sugar [93, 94], base [95], or multiple components of a nucleoside [96]. Among the most effective sugar modifications is the locked nucleic acid (LNA) [97], where a methylene bridge (Figure 1-5) is introduced between the 4’-carbon and 2’-oxygen of the furanose ring to “lock” it into a C3’-endo (RNA-like) conformation [98]. By substituting DNA or RNA nucleotides with the corresponding LNA nucleotides, both the thermal stability and the melting temperature, $T_m$, of a duplex can be increased. LNA nucleotides follow Watson–Crick base-pairing rules (i.e. a:t and c:g) and LNA-substituted oligonucleotides may be created on standard automated synthesizers and are available through commercial vendors.
The structure (above) and typical conformation (below) of DNA, RNA, and LNA are shown. Typically, DNA and RNA nucleotides oscillate between C2'-endo and C3'-endo conformations and are less and more populated in the C3'-endo conformation respectively. LNA is locked in the C3'-endo conformation by a 2'-O, 4'-C-methylene bridge (red). Figure adapted from Campbell and Wengel [99].

1.3.1 THERMODYNAMIC AND STRUCTURAL PROPERTIES OF LNA

The introduction of a single LNA substitution in one strand generally increases the $T_m$ of a short complementary DNA duplex by 1 – 8 °C [1], while somewhat larger increases in $T_m$ are seen when a single LNA substitution is made in short single-stranded DNA paired with its complementary RNA strand [100]. Enhanced stabilization results, in part, from LNA mediated pre-organization of the sugar-phosphate backbone that serves to reduce the loss in entropy accompanying the ssDNA to dsDNA transition. The duplex state therefore becomes more energetically favoured [1, 101]. LNAs may also enhance base stacking by altering the geometry of the backbone [102-104].
complex structures, such as an LNA:LNA base-pair, are less well studied but have shown hyper-stabilizing effects [97]. Conversely, duplexes containing LNA substitutions are in general less tolerant of base-pair mismatches, particularly if an LNA participates in the mismatch [105, 106]. Together, these unique LNA attributes can lead to improved binding selectivity when the LNAs are strategically substituted within an oligonucleotide or duplex.

NMR, CD spectroscopy, and X-ray crystallography have been used to investigate the structure and energetics of duplexes formed between a short (ca. 7-10 bp) LNA-containing DNA oligonucleotide and a complementary pure DNA or RNA strand [101, 103, 104, 107-111]. In a DNA duplex, LNA substitutions induce local changes in helical structure, commensurate with LNA content, towards the A-form RNA-like helix and away from the standard B-form structure of DNA. The 3’ neighbouring nucleotides adopt an LNA-like sugar pucker that increases the persistence length beyond the LNA containing base pair [101, 112]. Duplexed to RNA, LNA-substituted DNA oligonucleotides also adopt an A-type helix, the canonical geometry of natural RNA [108-110]. A fully locked strand, when duplexed with either natural DNA or RNA, therefore adopts a fully A-form helix [113]. Interestingly, however, duplexes in which both strands are fully locked show a non-natural A-type helical structure characterized by a widening of the minor groove [102-104]. And those duplexes are exceptionally stable. LNA substitutions can therefore permit tuning of the thermodynamic stability and structural properties of a duplex, and that capability can and has been exploited to create promising oligonucleotide-based therapeutics and improved diagnostic assays [98, 114, 115].

1.3.2 USE OF LNAs IN THERAPEUTICS AND DIAGNOSTICS

The discussion below outlines how the replacement of natural nucleotide(s) with LNA chemistries has allowed for the properties of oligonucleotide based therapeutics and clinical diagnostic agents to be tuned improve their performance.
1.3.2.1 LNA CONTAINING THERAPEUTICS

Oligonucleotide therapeutics may be used to modulate gene expression or to modulate the activity of a gene product. Disease-associated transcripts (mRNA), pre-mRNA or non-mRNA are targeted by an exogenous oligonucleotide to either inhibit RNA processing or to cleave or illicit cleavage of the target RNA using cellular machinery; translation is thereby inhibited [98, 116-118]. Given their enhanced stability and/or chemical resistance, oligonucleotide therapeutics bearing chemically modified nucleotides (e.g. 2'-O-methyl and 2'-O-methoxyethyl sugar and phosphorothioate backbone modifications) and/or LNA substitutions have shown clinical promise [119, 120]. These include antisense agents [121], RNAi [120], siRNA [122], anti-miRNA probes [117], LNAzymes [118, 123], and aptamers [124, 125] etc.) comprised of chimeric LNA oligonucleotides [98], where the presence of the LNA(s) (relative to the corresponding unmodified oligonucleotide) serves to improve stability, pharmacokinetics and/or target specificity [126]. LNA substitutions have also been shown to improve affinity and cellular delivery without increasing toxicity or the recruitment of enzymes responsible for duplexed RNA degradation [114, 115].

However, how best to employ LNA substitutions in therapeutic agents is not the focus of this thesis, which instead focuses on tools for improving use of LNAs in diagnostic technologies.

1.3.2.2 USE OF LNAs IN DISEASE DIAGNOSTICS

LNAs are finding increasing use in disease diagnostics. The following example helps to understand why. In qPCR, short dual-labeled hydrolysis probes (ca.≤ 25 bp) may be used to monitor amplification of a particular allele [57, 69, 105, 127]. If that probe is directed against the mutant BRAF V600E oncogene, for example, its specificity for the MT allele is largely determined by the difference (Δ) between the melting temperatures of the perfectly matched probe:MT ($T_{m,MT}$) and mismatched probe:WT
\((T_m,WT)\) duplexes. LNA substitutions increase duplex stability, permitting \(T_m,MT\) to then be held constant by reducing probe length. Shortening probe length increases the fractional contribution of a mismatched base pair to the stability of the probe:WT (non-target) duplex, lowering \(T_m,WT\) and false positives. If multiple LNAs are placed at or adjacent to the site of variance SOV, mismatch discrimination can be further enhanced depending on the LNA substitution pattern [69]. Thus, when they possess certain features (i.e. sequence, length and LNA placement) LNA bearing AS probes can be used to specifically bind and detect their target allele with good specificity [105].

LNAs have likewise been used to enhance selectivity of other diagnostic technologies, including mRNA in situ hybridization panels [128] and DNA microarrays used for miRNA expression profiling [129, 130]. To improve miRNA expression profiling, Castoldi et al. [130] micro-arrayed probes on glass to permit hybridization of samples of fluorescently labeled miRNA. The recorded pattern and intensity of fluorescence then provides a readout of miRNA expression in parallel. To improve probe specificity and assay sensitivity, Castoldi et al. adjusted the length and LNA content of each probe so that they shared a (near) common \(T_m\) when hybridized to their target miRNA [130].

LNA substitutions at or near the 3’ end of the primer have also been shown to enhance allelic discrimination [68, 73]. Some general guidelines [68, 73] for where those LNA substitutions should occur have been proposed. Specifically, three substitution patterns – L123, L25 and L45 (where L indicates an LNA substitution and each number the position of a substitution relative to the 3’ terminal (0) base) – have proven effective [67]. The length of the primer is then set to ensure that the primer-template \(T_m\) lies \(ca.\ 1–2\) °C above \(T_a\) [74].

These examples offer evidence that LNA substitutions within either probes or primers can serve to improve their specificities and sensitivities. However, those improvements are currently realized using what is largely an empirical design strategy,
which is slow, expensive, and not always productive. This points to the value of and need for developing more robust tools and associated guidelines for designing LNA-bearing AS probes and primers, including molecular thermodynamic models that accurately predict the melting thermodynamics of complementary or mismatched duplexes containing any pattern of LNA substitutions.

1.4 THESIS OBJECTIVES

The central objective of this thesis is to extend a current molecular thermodynamic model known as the single-base thermodynamic (SBT) model to allow accurate prediction of melting temperatures and thermodynamics for short complementary duplexes containing LNA substitutions in both strands. The extended model will therefore apply to duplexes containing LNA:LNA base-pairs and/or oppositely oriented LNA:DNA base pairs. The model will create the potential to design a whole new class of qPCR-based diagnostic assays in which both probes and primers contain LNA substitutions and are proximally positioned on opposite strands. Specific thesis objectives are to:

1. Use UV spectroscopy to collect, in triplicate, melting thermodynamic data for a library of short complementary DNA duplexes containing at least one LNA substitution in each strand to form a pair of oppositely oriented LNA-DNA base-pairs, an LNA:LNA base pair, or a combination thereof.

2. Use UV spectroscopy to collect, in triplicate, melting thermodynamic data for each of the three isosequential parent duplexes to those evaluated in objective 1. Those parent duplexes are created by replacing the LNA substitutions in a) the antisense strand, b) the sense strand, or c) both strands with the corresponding DNA nucleotide.

3. Extend the SBT model and then regress all new model parameters, known as incremental thermodynamic parameters, to the above data sets.
4. Validate the extended SBT model by comparing predicted melting thermodynamics to and experimental data for a set of duplexes not included in the dataset used for parameter regression.

5. Illustrate how the extended model can be used to assist LNA-substituted primer and probe design for a multiplexed qPCR based diagnostic assay against both common and rare \textit{BRAF V600} mutations.

6. Implement in R/Shiny software an online interactive version of the SBT model that computes the melting temperature and thermodynamics for a user-input duplex sequence and related solution conditions.

Chapter 2, adapted from a recently published article in \textit{Methods in Enzymology} \cite{131}, presents in detail the experimental methods and associated theories used to collect and analyze the required melting thermodynamic data. Chapter 3, a version of which was recently published in a special issue of the \textit{AIChE Journal} honouring Professor John Prausnitz, then presents the extended SBT model, as well as the data and methods used to regress the required model parameters. Finally, yet to be published work in Chapter 4 describes the application of the extended model to the proof-of-concept design of sets of novel LNA-modified probes and primers that enable multiplexed detection of all clinically relevant \textit{BRAF V600} mutations by qPCR.

I conclude this introductory chapter with a review of models for predicting the melting thermodynamics of natural and nucleotide modified DNA duplexes, including the SBT model upon which this thesis work is based.

1.5 MODELS FOR PREDICTING THE MELTING THERMODYNAMICS OF NATURAL AND CHEMICALLY MODIFIED DNA DUPLEXES

The thermal stability of complementary duplex DNA has long been known to depend on its length and \textit{g/c} base-pair content, and qualitative methods for predicting $T_m$
based on that information are available [132, 133]. However, due in part to the ever increasing range of medical and diagnostic applications of oligonucleotides, considerable attention has been given over the past 3 decades [134, 135] to establishing far more accurate models to predict the stability of a DNA duplex based on oligonucleotide length and sequence. The most widely used of those models consider the net contributions to duplex stability arising from base-pairing (i.e. hydrogen bonding) and base-stacking (i.e. $\pi - \pi^*$) interactions. Inclusion of the latter contribution leads to models that treat contributions to duplex stability at the level of neighbouring base-pairs (doublets) [135]. Called the nearest neighbour thermodynamic (NNT) models, they were developed first for pure complementary DNA duplexes [135, 136], then extended to enable prediction of melting thermodynamics for mismatched DNA duplexes [137-141] (which arise in probe:non-target duplexes for example). The growing use of LNAs has likewise motivated extension of NNT models to enable prediction of the stability of duplexes wherein one of the strands contains one or more LNAs that each participates in a complementary LNA:DNA base pair [1, 142] or a mismatched LNA:DNA base pair [105, 106].

1.5.1 CLASSIC NEAREST NEIGHBOR THERMODYNAMIC MODELS

The best known and most widely used NNT model is the “unified” NNT model of Santa Lucia, Jr. et. al. [135] for predicting melting thermodynamics for pure complementary DNA duplexes (e.g. an unlabeled pure DNA primer-template duplex). Hydrogen bonding interactions within the $m^{th}$ base pair and stacking between the $m^{th}$ and $(m+1)^{th}$ base pair are quantified in terms of a set of energetic parameters for 10 unique nearest neighbour base-pair doublets at the $m^{th}$ and $(m+1)^{th}$ positions (e.g. one such doublet is 5’-ac-3’/5’-gt-3’, which is written in the short-hand form ac/gt in Table 1-2) [143]. Though 16 different complementary doublets are possible, palindromic restrictions imposed by the anti-parallel nature of Watson-Crick base pairing limit the total number of energetically unique nearest-neighbour pairs to 10 [144, 145]. Classic NNT models
compute $T_m$ for a given duplex sequence using the fundamental thermodynamic relation describing the dissociation of a DNA duplex into its two constituent single strands ($dsDNA \Leftrightarrow ssDNA_1 + ssDNA_2$)

$$T_m = \frac{\Delta H_{\text{DNA}}}{\Delta S_{\text{DNA}} - R \ln(K)}$$  

(1-2)

where $K$ is the equilibrium constant for the dissociation reaction, $R$ is the ideal gas constant (1.987 cal mol$^{-1}$K$^{-1}$), and $\Delta H_{\text{DNA}}$ and $\Delta S_{\text{DNA}}$ are the change in enthalpy and entropy, respectively, for duplex dissociation. $\Delta H_{\text{DNA}}$ and $\Delta S_{\text{DNA}}$ are computed as

$$\Delta H_{\text{DNA}} = \sum_{i=1}^{10} n_i \Delta H_{\text{NN}_i} + \sum_{j=1}^{2} m_j \Delta H_{j}^{\text{init}}$$  

(1-3)

$$\Delta S_{\text{DNA}} = \sum_{i=1}^{10} n_i \Delta S_{\text{NN}_i} + \sum_{j=1}^{2} m_j \Delta S_{j}^{\text{init}} + \Delta S_{\text{sym}}$$  

(1-4)

with the first summation on the right side of equations (1-3) and (1-4) totaling the energetic/entropic contribution (Table 1-2) of each nearest neighbour (NN) doublet within the duplex. In equation (1-3), for example, that summation therefore totals the number, $n_i$, times the energy of denaturation, $\Delta H_{\text{NN}_i}^0$, of each of the 10 unique Watson-Crick nearest neighbour pairs present in the duplex. The second summation on the right side computes the energy needed to initiate duplex dissociation at each terminal base-pair, where $m_j$ is the number and $\Delta H_{j}^{\text{init}}$ the energetic contribution of each terminal base pair of type $j$ (i.e. $a:t$ or $g:c$) [135, 146]. Finally, NNT models also account for statistical mechanical contributions to the denaturation entropy $\Delta S_{\text{sym}}$ arising from sequence symmetry (i.e. self-complementarity) within an oligonucleotide.
Table 1-2 Unified nearest-neighbour parameters ($\Delta H^0$ and $\Delta S^0$) for the duplexation of complementary base-pair doublets in 1M NaCl\cite{135}

<table>
<thead>
<tr>
<th>Sequence</th>
<th>$\Delta H^0$ (kcal/mol)</th>
<th>$\Delta S^0$ (cal/k·mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA/TT</td>
<td>-7.9</td>
<td>-22.2</td>
</tr>
<tr>
<td>AT/TA</td>
<td>-7.2</td>
<td>-20.4</td>
</tr>
<tr>
<td>TA/AT</td>
<td>-7.2</td>
<td>-21.3</td>
</tr>
<tr>
<td>CA/GT</td>
<td>-8.5</td>
<td>-22.7</td>
</tr>
<tr>
<td>GT/CA</td>
<td>-8.4</td>
<td>-22.4</td>
</tr>
<tr>
<td>CT/GA</td>
<td>-7.8</td>
<td>-21.0</td>
</tr>
<tr>
<td>GA/CT</td>
<td>-8.2</td>
<td>-22.2</td>
</tr>
<tr>
<td>CG/GC</td>
<td>-10.6</td>
<td>-27.2</td>
</tr>
<tr>
<td>GC/CG</td>
<td>-9.8</td>
<td>-24.4</td>
</tr>
<tr>
<td>GG/CC</td>
<td>-8.0</td>
<td>-19.9</td>
</tr>
<tr>
<td>Init. w/term. G-C</td>
<td>0.1</td>
<td>-2.8</td>
</tr>
<tr>
<td>Init. w/term. A-T</td>
<td>2.3</td>
<td>4.1</td>
</tr>
<tr>
<td>Symmetry correction</td>
<td>0</td>
<td>-1.4</td>
</tr>
</tbody>
</table>

Predicting the $T_m$ of a duplex using equation (1-2) also requires a value for the equilibrium dissociation constant, $K$, which can be computed based on the concentrations of the single strands and knowledge of any duplex symmetry. For a non-self complementary strand (e.g. 5’-aaaaaaa-3’ cannot form a duplex with itself), $K$ is given by $C_T/4$ when the strands are added in equal concentration. Here, $C_T$ is the total strand concentration. If one strand is added in a greater concentration, $K = C_A - C_B/2$, where $C_A$ and $C_B$ are the more and less concentrated strands respectively \cite{143}. And finally, in the case of a self complementary duplex (i.e. $dsDNA \Leftrightarrow 2 ssDNA_1$), $K$ becomes equal to $C_T$\cite{143}. Figure 1-6 shows how equations (1-2) to (1-4) may be applied to a 6 bp non-self-complementary duplex to predict its $T_m$.  


Figure 1-6 The evaluation of equations (1-2)-(1-4) for the non-self-complementary duplex above first sums the contribution from each of nearest neighbour doublet parameters from Table 1-2 above (contained in the bolded box) as it moves left to right; counts the number the terminal base-pairs; and finally the evaluation corrects for duplex symmetry. The Tm prediction is computed for the non-self complementary duplex (ΔS_{sym}=0) with 5μM total strand concentration in 1M NaCl. ΔH_{DNA} and ΔS_{DNA} used in equation (1-2) are defined for the dissociation reaction rather than the duplexation reaction and thus have opposite signs.
1.5.2  ADVANCED NNT MODELS

Classic NNT models, including the unified model of Santa Lucia, Jr., et al. [135], assume $\Delta H_{DNA}$ and $\Delta S_{DNA}$ are temperature independent – implying that there is no change in heat capacity $\Delta C_p$ for duplex denaturation. This is a poor assumption, as shown by Hughesman et al. [136], who used calorimetry to find that duplex denaturation is accompanied by a positive heat capacity change per base pair, $\Delta C_p^{bp}$, of $42 \pm 16$ cal mol$^{-1}$ K$^{-1}$ bp$^{-1}$. That change in heat capacity is accounted for in a more advanced NNT model [136] in which $\Delta C_p$ is used to determine the temperature dependencies of $\Delta H_{DNA}$ and $\Delta S_{DNA}$ away from a reference state temperature, 53 °C, at which $\Delta H_{DNA}^0$ and $\Delta S_{DNA}^0$ are computed using equations (1-3) and (1-4), respectively. $T_m$ is therefore now computed as

$$T_m = \frac{\Delta H_{DNA}^0(T_{ref}) + \Delta C_p(T_m - T_{ref})}{\Delta S_{DNA}^0(T_{ref}) + \Delta C_p \ln(T_m/T_{ref}) - R \ln(K)}$$  \hspace{1cm} (1-5)$$

where the net change in heat capacity, $\Delta C_p$, for the helix-to-coil (dissociation) transition is given by $n_{bp} \Delta C_p^{bp}$, where $n_{bp}$ is the total number of base-pairs in the duplex and $\Delta C_p^{bp}$ is 42 cal mol$^{-1}$ K$^{-1}$ bp$^{-1}$. Here, $\Delta C_p^{bp}$ is assumed to be independent of temperature and the model (like the original NNT models) only applicable to short (< 25 bp) B-form duplexes having $T_m$s above ca. 40 °C [136]. The temperature dependencies captured in equation (1-5) improve model accuracy significantly, particularly for duplexes having a high $T_m$ [136].

Both classic and advanced NNT models, including the model of Hughesman et al. [136] have been extended to permit prediction of melting temperatures for short DNA duplexes containing a single mismatched natural DNA:DNA base pair [137-141, 147]. The energetic contribution to duplex stability of a mismatch at position $y$ within a duplex stretch 5'-x_yz-3', where $x$, $y$, and $z$ represent DNA bases, is captured in a unique set of parameters (Table 1-3 below) for the two doublets containing the mismatched base pair.
1.5.3 CLASSIC AND ADVANCED NNT MODELS FOR PREDICTING THE STABILITY OF LNA-CONTAINING DUPLICES

Both classic and advanced NNT models have been developed/extended to permit prediction of melting thermodynamics of duplexes bearing LNA substitutions in one of the two strands. In the classic NNT models, LNA substitutions are treated with a unique set of nearest neighbour parameters describing the incremental changes in the melting enthalpy, $\Delta \Delta H^o(=\Delta H^o_{\text{LNA:DNA:DNA}} - \Delta H^o_{\text{DNA:DNA}})$, and melting entropy, $\Delta \Delta S^o(=\Delta S^o_{\text{LNA:DNA:DNA}} - \Delta S^o_{\text{DNA:DNA}})$, when an LNA substitution (subscript LNA:DNA) is made within a pure DNA duplex of identical sequence (subscript DNA:DNA) [106, 142, 147].

McTigue et al.[142] and Owczarzy et al.[106] have both created models of this type, with the former capable of predicting $T_m$ values for duplexes containing isolated LNA substitutions in one strand (i.e. 5'-$x$M$^L$-$y$-3' (often called “mixmers”) where $x$ and $y$ are DNA nucleotides and $M^L$ is any one of the 4 LNA nucleotides), and the latter applicable to any pattern of LNA substitutions in one strand. Both models are accurate, predicting the difference in $T_m$ between the LNA containing and natural DNA duplex, $\Delta T_m$, with an average error of 0.4 ± 1.5 °C, which is comparable to the underlying NNT model for pure DNA duplexes [1]. However, they are highly parameterized, each requiring over 150 nearest neighbour parameters to achieve accurate predictions, and the accuracy of those predictions diminishes as the $T_m$ or LNA content of the duplex increases. Moreover, the over-parameterization in these models has obscured the physical significance of the various model parameters required. As a result, they are not well suited for further extension to allow prediction of melting thermodynamics of duplexes bearing LNA substitutions in both strands.

Hughesman et al.[1, 105] developed the Single Base Thermodynamic (SBT) model for predicting melting thermodynamics of duplexes containing any pattern of LNA substitutions in one strand by extending their temperature-dependent NNT model [136].
The SBT model, which requires ca. 1/10 the number of parameters as the model of either McTigue et al. [142] or Owczarzy et al. [106], accurately predicts $T_m$s and melting thermodynamics of a short LNA-modified oligonucleotide duplexed either to its perfect complement or a template containing a mismatched base pair. For a non-self-complementary oligonucleotide (for example), the $T_m$ value is predicted from the sequence and chemistry of the strand through the relation:

$$
T_m = \frac{\Delta H_{LNA}(T_m) + \Delta \Delta H_{f/q} + \Delta \Delta G_{DNA:MM} + \Delta \Delta G_{LNA:MM}}{\Delta S_{LNA}(T_m) + \Delta \Delta S_{f/q} + \Delta \Delta S_{salt} - R \ln \left(\frac{C}{4}\right)}
$$

(1-6)

where $\Delta H_{LNA}(T_m)$ and $\Delta S_{LNA}(T_m)$ are the enthalpy and entropy changes, respectively, for denaturation of the LNA containing duplex at $T_m$. $\Delta \Delta G_{DNA:MM}$ and $\Delta \Delta G_{LNA:MM}$ are the perturbations to duplex stability (Gibbs energy) arising from any DNA:DNA or LNA:DNA mismatches, respectively. The effect on duplex stability of any terminal fluorescent reporter dyes and quenchers is captured through their perturbations to the melting enthalpy, $\Delta \Delta H_{f/q}$, and melting entropy, $\Delta \Delta S_{f/q}$, while $\Delta \Delta S_{salt}$ accounts for the dependence of duplex stability on salt composition (i.e. Mg$^{2+}$, Na$^+$). $\Delta H_{LNA}(T_m)$ and $\Delta S_{LNA}(T_m)$ are computed as:

$$
\Delta H_{LNA}(T_m) = \Delta H^0_{DNA}(T_{ref}) + \Delta \Delta H^0_{LNA} + \Delta C_p (T_m - T_{ref})
$$

(1-7)

$$
\Delta S_{LNA}(T_m) = \Delta S^0_{DNA}(T_{ref}) + \Delta \Delta S^0_{LNA} + \Delta C_p \ln \left(\frac{T_m}{T_{ref}}\right)
$$

(1-8)

where $\Delta H^0_{DNA}(T_{ref})$, $\Delta S^0_{DNA}(T_{ref})$, and $\Delta C_p$ are as described in equation (1-5), and $\Delta \Delta H^0_{LNA}$ and $\Delta \Delta S^0_{LNA}$ are the incremental enthalpy and entropy changes, respectfully, to duplex stability computed as follows:

$$
\Delta \Delta H^0_{LNA} = \sum n_i \Delta \Delta H^0_i
$$

(1-9)
Here, \( n_i \) is the number of LNA substitutions of type \( i \), and \( \Delta \Delta H_i^0 \) and \( \Delta \Delta S_i^0 \) are the incremental enthalpy and entropy parameters for each of the four LNA-DNA base-pairs of type \( i \). The SBT model predicts \( T_m \) values for duplexes containing any number and pattern of LNA:DNA base-pairs with an accuracy of \( \pm 1.4 \, ^\circ C \) \([1]\).

Hughesman \textit{et al.} have recently extended the SBT model \([105]\) to account for DNA:DNA and LNA:DNA mismatches. Perturbations to the Gibbs energy of the duplex arising from DNA:DNA base-pair mismatches and LNA:DNA mismatches are described by \( \Delta \Delta G_{DNA:MM} \) and \( \Delta \Delta G_{LNA:MM} \), respectively, where:

\[
\Delta \Delta G_{DNA:MM} = \sum_{i=1}^{10} n_i \Delta \Delta G_{DNA}^0_i + \sum_{j=1}^{40} n_j \Delta \Delta G_{DNA:MM}^0_i
\] \hspace{1cm} (1-11)

\[
\Delta \Delta G_{LNA:MM} = \sum_{i=1}^{12} n_i \Delta \Delta G_{LNA:MM}^0_i + \Delta \Delta G_{5'NN-LNA} + \Delta \Delta G_{3'NN-LNA}
\] \hspace{1cm} (1-12)

In equation (1-11), \( \Delta \Delta G_{DNA:MM} \) is computed by first subtracting the Gibbs energy of each complementary nearest neighbour base pair \( i \) lost in the formation of a mismatch; the Gibbs energy of the nearest neighbour doublets, \( j \), containing the mismatch are then added. Participation in a mismatch of a more structurally rigid LNA nucleotide generally impacts duplex stability more punitively \([69, 105, 106]\). \( \Delta \Delta G_{LNA:MM} \) accounts for this added perturbation to the transition energy change, where the 12 \( \Delta \Delta G_{LNA:MM}^0 \) parameters describe the energetic penalty of each possible LNA:DNA mismatch relative to the corresponding isosequential pure DNA:DNA mismatch. (In contrast, those models that assume \( \Delta C_p = 0 \) use 192 parameters \( 96 \Delta \Delta H_{LNA:MM}^0 \) and \( 96 \Delta \Delta S_{LNA:MM}^0 \) and exhibit comparable accuracy.) Perturbations to the transition energy change for duplexes where an LNA is directly to the 5’ side or 3’ side of the mismatch are then accounted for.
through the two base independent parameters $\Delta \Delta G^\circ_{3'NN-LNA}$ and $\Delta \Delta G^\circ_{5'NN-LNA}$, respectively [69]. Interestingly, LNA substitutions downstream (3') of a mismatch are generally much more punitive than those upstream (5') of the same mismatch [105].

**Table 1-3** Nearest-neighbour $\Delta G^\circ_{37}$ increments (kcal mol$^{-1}$) for internal single mismatches next to Watson-Crick pairs in 1M NaCl$^a$

<table>
<thead>
<tr>
<th>Propagation sequence</th>
<th>X</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>GX/CY</td>
<td>A</td>
<td>0.17</td>
<td>0.81</td>
<td>-0.25</td>
<td>WC</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.47</td>
<td>0.79</td>
<td>WC</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>-0.52</td>
<td>WC</td>
<td>-1.11</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>WC</td>
<td>0.98</td>
<td>-0.59</td>
<td>0.45</td>
</tr>
<tr>
<td>CX/GY</td>
<td>A</td>
<td>0.43</td>
<td>0.75</td>
<td>0.03</td>
<td>WC</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.79</td>
<td>0.70</td>
<td>WC</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>0.11</td>
<td>WC</td>
<td>-0.11</td>
<td>-0.47</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>WC</td>
<td>0.40</td>
<td>-0.32</td>
<td>-0.12</td>
</tr>
<tr>
<td>AX/TY</td>
<td>A</td>
<td>0.61</td>
<td>0.88</td>
<td>0.14</td>
<td>WC</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.77</td>
<td>1.33</td>
<td>WC</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>0.02</td>
<td>WC</td>
<td>-0.13</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>WC</td>
<td>0.73</td>
<td>0.07</td>
<td>0.69</td>
</tr>
<tr>
<td>TX/AY</td>
<td>A</td>
<td>0.69</td>
<td>0.92</td>
<td>0.42</td>
<td>WC</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>1.33</td>
<td>1.05</td>
<td>WC</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>0.74</td>
<td>WC</td>
<td>0.44</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>WC</td>
<td>0.75</td>
<td>0.34</td>
<td>0.68</td>
</tr>
</tbody>
</table>

$^a$WC denotes a complementary Watson-Crick NN doublet. Values are shown without errors and can be found in the original references [137-141]. The energies are reported for the duplexation reaction ($ssDNA \rightarrow dsDNA$).

Displaying a mean error and standard deviation of -0.6 ± 2.0°C in predicting the $T_m$ of LNA loaded probes [105], the SBT model of Hughesman et al. [1] is both accurate and well suited for further extension to duplexes containing LNA substitutions in both strands that participate in either hyperstabilizing LNA:LNA base pairs [97] or oppositely oriented LNA:DNA base pairs.
CHAPTER 2: CALORIMETRIC AND SPECTROSCOPIC ANALYSIS OF THE THERMAL STABILITY OF SHORT DUPLEX DNA CONTAINING SUGAR AND BASE MODIFIED NUCLEOTIDES

2.1 SYNOPSIS

Base- and sugar-modified analogs of DNA and RNA are finding ever expanding use in medicine and biotechnology as tools to better tailor structured oligonucleotides by altering their thermal stability, nuclease resistance, base-pairing specificity, antisense activity, or cellular uptake. Proper deployment of these chemical modifications generally requires knowledge of how each affects base-pairing properties and thermal stabilities. Here, we describe in detail how differential scanning calorimetry and UV spectroscopy may be used to quantify the melting thermodynamics of short dsDNA containing chemically modified nucleosides in one or both strands. Insights are provided into why and how the presence of highly stable base pairs containing modified nucleosides can alter the nature of calorimetry or melting spectroscopy data, and how each experiment must therefore be conducted to ensure high-quality melting thermodynamics data are obtained. Strengths and weaknesses of the two methods when applied to chemically modified duplexes are also addressed.

2.2 INTRODUCTION

Applications of synthetic analogs of deoxyribonucleotides and ribonucleotides have experienced tremendous growth in the last two decades [80, 81], due in part to their ability to enhance diagnostic and therapeutic applications of structured oligonucleotides by altering their thermodynamic and/or chemical (e.g. resistance to nucleases) stability, base-pairing specificity, antisense activity, or cellular uptake [82]. The phosphodiester bonds of natural DNA and RNA are, by design, susceptible to hydrolysis by nucleases.
Short natural oligonucleotides also tend to exhibit relatively poor plasma protein binding and transport within the cardiovasculature, while longer oligos complex their target slowly. For oligonucleotides used as antisense agents, slow binding kinetics can be exacerbated by the tendency for their mRNA targets to be highly structured. As a result, unmodified DNA or RNA can exhibit poor pharmacokinetics and relatively rapid excretion when delivered as a parenteral therapeutic.

These limitations can be overcome through appropriate chemical modifications that serve to increase the specificity of the oligonucleotide for its target sequence, the stability of the complex/duplex formed, and/or the resistance to nuclease-catalyzed degradation. Reported chemical modifications to DNA and RNA now number in the 100s [83, 84] and include chemistries that serve to alter helix structure [85], expand the genetic alphabet beyond the natural nucleosides a, t, g and c, [86, 87] or alter cellular phenotypes [88]. Unique biological functions, and improved diagnostic specificity or therapeutic potency may thereby be realized [68, 89-91].

Many of the chemical modifications reported to date describe either alteration of the backbone [92], furanose sugar [93, 94], base [95], or multiple components of a nucleoside (e.g. Andersen et. al., 2013 [96]). Among the most effective sugar modifications is the locked nucleic acid (LNA) [97], where a methylene bridge has been introduced between the 4’-carbon and 2’-oxygen of the furanose sugar to “lock” it into a C3’-endo (RNA-like) conformation [98]. By substituting DNA or RNA nucleotides with the corresponding LNA nucleotides, both the thermal stability and the melting temperature, \( T_m \), of a duplex can be increased. A single LNA substitution in one strand generally increases \( T_m \) of a complementary DNA duplex (dsDNA) by 1 – 8 °C [1], while somewhat larger increases in \( T_m \) are seen when a single LNA substitution is made in single-stranded DNA (ssDNA) paired with its complementary RNA strand [100].

Novel modifications to nucleoside bases include the so-called “super-bases” 2-amino-adenine, 2′-O-(2-(methoxy)ethyl)-2-thiothymidine, and 2′-deoxy-2′-fluoro-2-thiothymidine [148, 149]. These modified bases pair more stably with their complementary
DNA base [150], and exhibit remarkably high thermal stabilities when paired with their complementary RNA base [148]. Further increases in thermal stability may be realized by introducing these base modifications into LNA nucleotides [96]. Moreover, the 2'-O-(2-(methoxy)ethyl)-2-thiothymidine modification has been shown to improve resistance to nucleases when substituted into DNA oligonucleotides [151]. The general chemistries of LNAs and LNA super-bases are provided in Figure 2-1.

![Figure 2-1](image)

**Figure 2-1** A standard locked nucleic acid (left) can be modified by replacing its natural nucleosidic base with one of several modified base chemistries (right) to further enhance duplex stability.

The capacity of nucleoside modifications to alter base-pairing properties and, as a result, duplex stabilities has motivated detailed studies of their melting thermodynamics using a wide variety of spectroscopic and calorimetric techniques. Our laboratory has experience in using two of these methods (e.g., Hughesman *et al.*, 2011 [136]), namely UV-spectroscopy monitored melting (UVM) and differential scanning calorimetry (DSC), to study the melting thermodynamics of short dsDNA containing chemically modified nucleosides in one or both strands (e.g., Fakhfakh *et. al.*, 2015 [152]). Both methods have been widely employed to study melting transitions in natural dsDNA, and descriptions of how each technique is best applied to that application are available [143, 153, 154]. Though some overlap with that prior art is unavoidable, our intent here is to detail how each method may best be applied to the study and interpretation of melting thermodynamics of short dsDNA containing chemically modified nucleotides in one or
both strands. We address the central questions of – How is each experiment conducted? What are relative the strengths and weaknesses of each approach? What thermodynamic parameters can and cannot be measured by each method? How are raw data presented and analyzed, and what assumptions must be invoked to obtain high quality values for each measurable thermodynamic parameter? And finally, how does one evaluate melting thermodynamics to properly understand the contribution to duplex stability provided by the specific nucleotide modification(s)?

2.3 MELTING THERMODYNAMICS FROM DIFFERENTIAL SCANNING CALORIMETRY

When applied to the determination of melting thermodynamics for modified (or natural) dsDNA, differential scanning calorimetry (DSC) experiments detect and characterize the helix-to-coil transition by measuring the difference in applied power dP (J sec$^{-1}$) needed to maintain the DNA sample solution and a DNA–free reference solution at the same temperature, as that temperature is linearly increased at a chosen scan rate (°C min$^{-1}$) through the melting transition (typically from ca. 1 °C to 100 °C, though larger ranges are possible). The measured differential power (dP) is divided by the scan rate (now expressed in K sec$^{-1}$) to give the excess heat capacity $C_p^{ex}$ (J K$^{-1}$), which is then plotted versus temperature ($T$) as illustrated in Figure 2-2A. The quality of that raw data, as well as corresponding $C_p^{ex}$-vs-$T$ scans for the DNA-free reference solution, determines what melting thermodynamics data may be extracted to good accuracy. For each sample, multiple up-scans at the chosen scan rate are conducted in succession to confirm reversibility in the melting transition. For a completely reversible transition, scans 2 and higher will substantially superimpose; the first scan may or may not.

2.3.1 STRENGTHS, WEAKNESSES AND WHAT CAN BE DETERMINED RELIABLY

The principle strength in determining dsDNA melting thermodynamics using DSC is that the method can provide a direct and highly accurate measure of $T_m$, as well as
the melting enthalpy ($\Delta H_{cal}$) and entropy ($\Delta S_{cal}$) at $T_m$; here, the subscript “$cal$” indicates that the $\Delta H$ and $\Delta S$ values were determined directly from melting-transition peak integration without invoking any model assumptions. If stable baselines are recorded, the heat capacity change ($\Delta C_p$) for the melting transition may be directly estimated as well. Modern DSC instruments have served to make this latter measurement much more reliable and accurate through their increased sensitivity and greatly improved baseline stability. Those instruments include the Nano DSC now sold by TA Instruments Inc., and MicroCal VP-DSC and VP-Capillary DSC now sold by Malvern Inc. – our laboratory and associated Centre for Bio-Thermodynamics (CBT) utilize the latter two instruments. Those instruments report energies in units of calories, but energies are expressed in Joules in this work in accordance with International System of Units convention.
Figure 2-2 Raw $C_p^{ex}$-vs-$T$ scans for the sample and reference buffer (A), and reference-normalized thermogram (B) for the 5’-gaDacagtDaag-3’/5’-cttaactgtttc-3’ duplex, where D is the 2-amino-adenine LNA super-base.

As detailed below, a reference and concentration normalized $C_p^{ex}$-vs-$T$ plot, known as the “thermogram”, is generated from the raw $C_p^{ex}$-vs-$T$ scans for the sample and reference buffer. It serves as the basis for determining all relevant melting thermodynamics, either through direct integration of the melting transition or through
regression to the two-state reaction model. Those melting thermodynamics include values of $T_m$, $\Delta H_{2-st}$ and $\Delta S_{2-st}$ for the helix-to-coil transition, as well as the equilibrium constant $K(T)$ for that transition, which as noted is a temperature-dependent parameter. Here, the subscript “2-st” is used to signify parameters regressed assuming a two-state transition. The ability to extract both $\Delta H_{cal}$ and $\Delta H_{2-st}$ from DSC data has led to the so-called “calorimetric criterion”, which assumes that the helix-to-coil transition obeys (and can therefore be described by) two-state melting thermodynamics if i) the transition is reversible and ii) $\Delta H_{cal}$ and $\Delta H_{2-st}$ agree to within their standard deviations (some researchers instead say to within ± 10%) (e.g. Movileanu et. al., 2002 [155]). Zhou et al. have shown that satisfying the calorimetric criterion is necessary, but not wholly sufficient, to unequivocally verify two-state melting [156]. Nevertheless, a $\Delta H_{2-st}/\Delta H_{cal}$ value near unity does appear to provide good evidence that the stability of a duplex (in the same reference solution) at temperatures away from $T_m$ can be reliably estimated using two-state thermodynamics, the total molar strand concentration $C_T$, and the experimental values of $\Delta H$, $\Delta S$, and $\Delta C_p$ collected at $T_m$. This represents the second main advantage of the DSC method.

Relative to the UVM method, however, DSC determination of melting thermodynamics of chemically modified dsDNA is challenged by a number of issues. First, the DSC experiment generally requires more material. Moreover, high purities and sequence homogeneities are required of each equimolar strand of the duplex. This is because determination of $\Delta H$ and $\Delta S$ requires the raw $C_p^{ex}$ data, which are provided in units of J K$^{-1}$, to be normalized by the total pre-transition moles of duplex $(C_T/2)$ in the sample cell volume to thereby express $C_p^{ex}$ in J mol$^{-1}$ K$^{-1}$ units. One must also assume that each strand is capable of participating in a fully complementary duplex at temperatures below and within the melting transition. If accurate melting thermodynamics are to be

---

1 The two-state melting transition model assumes that dsDNA denaturation occurs as an “all-or-none” process such that at any given $T$ the fully duplexed state and the single-strand (random-coil) state are the only states present at equilibrium.
obtained, each synthesized strand must therefore be HPLC purified to homogeneity. While this does not represent a significant problem when studying short duplexes comprised of natural DNA, chemically modified nucleotides and oligonucleotides containing them are generally more expensive, often much more expensive, and the added cost of purifying them by HPLC can be prohibitive, particularly for duplexes containing a large number of nucleotide modifications. In our view, use of the DSC method in these cases, though superior to the UVM method in principle, is therefore justified only when complete and highly accurate knowledge of melting thermodynamics, including second-derivative thermodynamic properties such as $\Delta C_p$, is required, or when one needs to be certain the duplex satisfies the calorimetric criterion.

Other more technical issues related to DSC analysis of chemically modified dsDNA are detailed below.

2.3.2 SAMPLE PREPARATION

Modified dsDNA samples analyzed by DSC are generally prepared from stock solutions, each containing one of the two required HPLC-purified single strands. Given the proprietary nature of the chemical modifications, the required strands are most often purchased from a company licensed to produce and sell them. For strands containing LNAs, those include Exiqon (Vadbaek, Denmark) and Integrated DNA Technologies (IDT, Coralville, IA). Exiqon can also reliably provide strands containing combinations of DNA nucleotides, LNAs, super-bases and/or LNA super-bases. For any strand comprised only of natural DNA, many reliable vendors are available.

Each HPLC-purified modified or natural DNA oligonucleotide is typically resuspended in aqueous buffer containing 1 M NaCl, 10 mM Na$_2$HPO$_4$ (pH 7), and 1 mM Na$_2$EDTA; this buffer is often referred to as “standard conditions” for melting analysis since much of the DNA melting thermodynamics data reported to date were collected in it [105, 157]. However, large and important data sets [135, 158], including most of the data used to regress the most widely used nearest-neighbor stability parameters for Watson-Crick base-pair doublets, have instead employed an aqueous buffer comprised of
1 M NaCl, 10 mM sodium cacodylate ((CH$_3$)$_3$AsO$_2$Na•3H$_2$O) (pH 7), and 0.5 mM Na$_2$EDTA. Both natural and modified dsDNA is in general more stable in the phosphate-buffered reference. Amalgamation of melting thermodynamics collected in the two different standard buffers therefore requires conversion of data collected in one buffer to the corresponding values in the other. Alemayehu et al. (2009) have provided the equations needed to do this for $T_m$, $\Delta G$, $\Delta H$, and $\Delta S$ [159]. Finally, for studies related to primer or probe stability, the NaCl (or possibly KCl) concentration in either reference solution is generally reduced to between 0.15 and 50 M to mimic PCR conditions.

When melting dsDNA between ca. 6 to 25 bp in length (which is the focus here), the two strands, whether complementary or mismatched, are mixed at equimolar concentration to a $C_T$ between 50 – 100 μM, with $C_T$ values at the higher end of this range employed for shorter duplexes (lower $\Delta H$). That sample is allowed to equilibrate at room temperature for 5 – 10 min to ensure dsDNA formation. A stock of the DNA-free reference buffer is also prepared and filtered using a 0.2 μm syringe filter to avoid introducing any particulate matter into the DSC (capillary) cells (we have used the syringe filters from Pall Inc. (Port Washington, NY) with good results). The DNA sample (for which the buffer used to prepare it has also been filtered) and reference buffer solutions are then degassed for 7 to 10 min under gentle stirring. In our facility, this is done by holding the sample at 80 °C for 5 – 10 min using a MicroCal Thermovac unit (Malvern Inc.).

### 2.3.3 EXPERIMENTAL PROCEDURE

Generally, a DSC experiment on a given dsDNA sample is comprised of a set of 2 or more quality thermal scans for each of three different fluid pairs described below. Each scan consists of a pre-scan system equilibration (typically at 1 °C for 10 or 15 min), a thermal up-scan in which $dP$ data are collected from 1 °C to 100 °C (or possibly a higher final temperature if larger sample pressures are employed) as the temperature is linearly increased at a fixed rate (often at 1 °C min$^{-1}$ or a lower rate), a down-scan at a rate defined by the instrument (often $ca$. 3 °C min$^{-1}$ for the instruments we employ), and a
sample re-equilibration. When a VP-DSC or VP-capillary DSC is employed, both cells are first syringe-loaded (with the instrument on and idling at 25 °C) to overflow levels with degassed nanopure water (first fluid pair). Excess liquid in each cell is then removed using a syringe with a collared needle to ensure precise and equivalent liquid volumes in the sample and reference cells. The two cells are sealed with a Teflon top, the piston in which is then actuated to set the pressure in each cell to between 26 and 30 psig to allow scans up to 100 °C while suppressing bubble formation. The programmed scanning cycle of the instrument is then activated, starting with a down-scan to the equilibration temperature, to continuously collect raw “water-vs-water” $C_p^{ex}$-vs-$T$ scans, typically overnight. Except for the first scan, which is influenced by the altered thermal history of the instrument, stable signals and substantial superimposition of the raw water-vs-water scans are expected. If they do not superimpose, it likely indicates that one or both of the cells of the calorimeter are contaminated or fouled. In that case, and at routine intervals, it is not just advisable but essential to carefully clean the cells according to guidelines specified by the manufacturer.

When satisfactory performance is observed and the instrument is either idling at 25 °C or in a down-scan at a temperature between 35 and 25 °C, the water is removed from both cells, each of which is then twice rinsed, loaded with the reference buffer, and finally pressure sealed using the same procedure as described above. If dynamic loading is used, the process should be completed before the down-scan temperature falls below 10 °C. Two substantially superimposed raw “buffer-vs-buffer” (second fluid pair) $C_p^{ex}$-vs-$T$ scans are then measured, which is usually achieved in the first two scans. Measurement of additional “buffer-vs-buffer” scans is generally not productive, as artifacts tend to appear in later scans.

With the instrument idling at 25 °C, the degassed (modified) dsDNA sample is then loaded into the sample cell. In this case, first rinsing the cell volume with the sample ensures that $C_T$ is known precisely in the subsequently loaded sample. However, the cost of the dsDNA sacrificed in the rinse step may be prohibitive. If buffer removal is
followed directly by sample loading, without conducting sample washes, a 2% sample
dilution may be assumed. As this is a crude correction/assumption, it will affect the
accuracy of the melting thermodynamics data collected. A set of 2-4 raw “sample-vs-
buffer” (third fluid pair) $C_p^{ex}$-vs-$T$ scans are then recorded. These will usually be
superimposable provided the maximum temperature of the scan is not too high.

Technical specifications and settings for the instrument used can influence data
quality as well. For example, for the VP-DSC the minimum thermal response time,
baseline noise, and baseline repeatability are 5 s, 2 μJ °C$^{-1}$ and ± 5 μJ °C$^{-1}$, respectively,
for a 1-1.5 °C min$^{-1}$ scan rate. The user may select both the filter period (data-acquisition
interval) and gain of the instrument, and for dsDNA melting studies we generally set
these at 10 sec and a null gain (passive feedback) setting, respectively, when using either
MicroCal (Malvern) instrument.

2.3.4 ANALYZING A DSC THERMOGRAM

Analysis of dsDNA melting data acquired by DSC begins with subtraction of raw
$C_p^{ex}$-vs-$T$ data for the buffer-vs-buffer reference from that for each sample-vs-buffer scan.
The average of the two buffer-vs-buffer scans may be used. However, we find that pre-
and post-transition baselines generated in the resulting reference-normalized scan are
often more linear and stable if a particular scan (see below), often the one collected
immediately preceding sample loading, is used for this purpose. Each reference-
normalized $C_p^{ex}$-vs-$T$ scan is then divided by the pre-transition molar concentration of
dsDNA ($C_T/2$) and the sample-cell volume (unique to each instrument and calibrated by
the manufacturer – $V_{Cell} = 511.8$ μL in our VP-DSC) to create a set of reference and
concentration normalized $C_p^{ex}$-vs-$T$ “melting thermograms”. A representative example of
a thermogram is provided in Figure 2-2B.

As DNA duplexation is a bimolecular hybridization reaction, $T_m$ depends on $C_T$.
Thermograms displaying a melting transition with a maximum $C_p^{ex}$ between ca. 55 °C
and 70 °C tend to provide melting thermodynamics data of the highest quality, due in part
to better defined pre- and post-transition baselines. If heats of melting are sufficiently large, the value of $C_T$ should therefore be set so that $T_m$ lies within that range.

For a few representative duplexes in the full set of sequences to be studied, thermograms should be collected at several scan rates, typically between 0.3 and 2 °C min$^{-1}$. Equilibrium thermodynamic data are being determined, and their values should therefore be scan-rate independent. For short natural or modified dsDNA, this is usually the case at scan rates ≤ 1 °C min$^{-1}$, but verification is prudent.

Both $\Delta H_{cal}$ and $\Delta S_{cal}$ for the helix-to-coil transition at $T_m$ may be determined through direct integration of the thermogram:

$$\Delta H_{cal} = \int_{T_1}^{T_2} C_p^{ex}(T)\,dT$$  \hspace{1cm} (2-1)

And

$$\Delta S_{cal} = \int_{T_1}^{T_2} \left( \frac{C_p^{ex}(T)}{T} \right)\,dT$$  \hspace{1cm} (2-2)

where $T_1$ and $T_2$ are the temperatures (integration limits) defined by the intersections of the $C_p^{ex}(T)$ melting transition with the pre- and post-transition baselines, respectively, as shown in Figure 2-2B. If those two baselines are sufficiently stable, the value of $\Delta C_p$ may also be estimated as the difference in ordinate positions of the baselines (Figure 2-2B). The duplex melting temperature, $T_m$, is computed as the temperature (Figure 2-2B) bisecting the area midpoint of the melting transition bounded by $T_1$ and $T_2$.

DSC Insight #1: Although well-behaved reference-normalized baselines permitting observation and estimation of $\Delta C_p$ (such as those in Figure 2-2B and in virtually every example published by leading calorimetry experts (see, for example, Privalov and Dragan, 2007 [160]; Chalikian et al., 1999 [161]) can be generated for some systems, they cannot always be generated, as baseline stability and shape are sensitive to many
variables. One important, but poorly recognized variable is the dynamic range of the raw \( C_p^{\text{ex}} - vs - T \) data for the buffer-vs-buffer reference relative to that for the sample-vs-buffer run. As shown in Figure 2-3, stable horizontal pre- and post-transition baselines are generally observed in the reference-normalized thermogram if the range of \( C_p^{\text{ex}} \) recorded in the two scans are similar in value. When the \( C_p^{\text{ex}} \) values for the sample-vs-buffer scan are appreciably higher or lower than those recorded in the buffer-vs-buffer scan, the baselines in the reference-normalized thermogram tend to be more poorly defined.

![Graph of \( C_p^{\text{ex}} \) vs Temperature](image)

**Figure 2-3** Inability to match the dynamic range of \( C_p^{\text{ex}} \) values recorded in the sample and reference buffer scans leads to nonlinearity in the pre-transition baseline of the reference buffer-normalized thermogram. Generally, if \( C_p^{\text{ex}} \) values for the sample and reference buffer are close at each temperature, stable linear baselines (solid line) are observed.

**DSC Insight #2:** Proper assignment of pre- and post-transition baselines can be difficult, and the uncertainties associated with those assignments are often the dominant source of error in the melting thermodynamics measured. The uncertainty in the slope of either baseline (most
notably the pre-transition baseline) can typically contribute up to ±2% error in the value of $\Delta H_{cal}$ and up to ±100% error in $\Delta C_p$.

**DSC Insight #3:** Use of equations (2-1) and (2-2) require specification of the shape of the baseline, $C_p^{ex*}(T)$, through the melting transition (i.e. between $T_1$ and $T_2$). Several different methods have been proposed and used [162]. We generally compute $C_p^{ex*}(T)$ as $\alpha(T)C_{p,dsDNA}^{ex*}(T) + (1 - \alpha(T))C_{p,ssDNA}^{ex*}(T)$, a result that assumes the baseline moves from the $C_p^{ex*}(T)$ of the pure dsDNA state to that of the pure ssDNA state in proportion to the fraction $\alpha(T)$ of total strands in the dsDNA state at each $T$.

**DSC Insight #4:** Obtaining a stable pre-transition baseline can be particularly challenging for highly modified dsDNA, and peak integration may not yield reliable melting thermodynamics when the presence or quality of a pre-transition is lacking. LNA-LNA base pairs, for example, are unusually stable, increasing the potential for formation of hairpins and other stable structures in one or both single strands. These structures alter the pre-transition baseline and can thereby complicate or preclude data analysis using two-state theory. If possible, dsDNA sequences to be studied by DSC should be designed to avoid such structures. Both Exiqon Inc. (https://www.exiqon.com/oligo-tools) and Integrated DNA Technologies (https://www.idtdna.com/calc/analyzer) provide tools that predict/score the likelihood of oligonucleotides bearing modified nucleotides forming stable single-stranded structures.

**DSC Insight #5:** The width of the melting transition decreases with increasing melting enthalpy. As base pairs containing modified nucleotides (e.g. LNA-DNA base pairs) are generally more stable than their
DNA-DNA counterparts, when compared to natural dsDNA, modified dsDNA of a shorter length will exhibit the same $T_m$, but a smaller melting enthalpy. This leads to a broader melting transition, which further challenges assignment of baselines.

As noted, the thermogram can also be analyzed by fitting to the two-state reaction model. Historically, this has been done using a simplified form of the two-state model in which $\Delta C_p$ is set equal to zero [134]. For much of the DSC data collected previously this assumption is wholly justified, as early-generation calorimeters generally did not offer the sensitivity and baseline stability needed to quantify $\Delta C_p$ (with the custom-built instruments of Privalov [163] being an exception). As a result, both $\Delta H$ and $\Delta S$ were taken to be temperature-independent quantities in spite of the theoretical inconsistencies presented by that assumption [164].

In a two-state reaction, the melting enthalpy is given by the van’t Hoff enthalpy $\Delta H_{v,H}$, which may be computed from a thermogram as

$$\Delta H_{2-st} = \Delta H_{v,H} = -R \left[ \frac{d(\ln K(T))}{d(1/T)} \right]$$

(2-3)

where $R$ is the ideal gas constant 8.314 J mol$^{-1}$ K$^{-1}$. For dsDNA that melts into two non-self-complementary sequences ($i.e.$, $AB \leftrightarrow A + B$, where, for example, strand $A$ might be the sequence 5′-aaaaaaaaa-3′, which is non-self-complementary because it cannot base-pair with itself), the temperature dependence of $K$ is given by

$$K(T) = \frac{C_T}{2} \frac{(1 - \alpha(T))^2}{\alpha(T)}$$

(2-4)

and for dsDNA that melts into two self-complementary strands ($i.e.$, $A_2 \leftrightarrow 2A$) by

$$K(T) = 2C_T \frac{(1 - \alpha(T))^2}{\alpha(T)}$$

(2-5)

Here, $\alpha(T)$, the fraction of total strands in the dsDNA state at each $T$, is given by
Use of equation (2-3) to determine $\Delta H_{v.H}$ allows estimation of the ratio $\Delta H_{v.H}/\Delta H_{cal}$. As noted, a ratio near unity provides evidence a dsDNA melting transition follows two-state thermodynamics, while systems exhibiting a ratio $< 1$ are generally characterized by significant populations of intermediate states within the melting transition. Ratios $> 1$ occur infrequently and are less understood.

However, determination of $\Delta C_p$ is entirely possible using modern DSC instruments, and such measurements show that $\Delta C_p$ is not zero, but instead positive in value and to a first approximation temperature independent at temperatures above ca. 50 °C [1, 161, 165-167]. As we have shown previously[136], thermograms may then be fit to the full two-state transition model:

$$C_p^{ex}(T) = b_{post}^{cal} + m_{post}^{cal} T - \Delta C_p \alpha(T)$$

$$+ \frac{[\Delta H_{2-st} + \Delta C_p (T - T_m)]^2}{RT^2} \left( \frac{\alpha(T) - \alpha(T)^2}{1 + \alpha(T)} \right) \quad (2-7)$$

$$\alpha(T) = 1 + \frac{K(T) - \sqrt{[K(T)]^2 + 2K(T)C_T}}{C_T} \quad (2-8)$$

$$K(T) = K(\alpha = 0.5)$$

$$\quad \times \exp \left[ - \frac{\Delta H_{2-st}}{RT} \left( 1 - \frac{T}{T_m} \right) - \frac{\Delta C_p}{RT} \left( T - T_m - T \ln \left( \frac{T}{T_m} \right) \right) \right] \quad (2-9)$$

to obtain a complete description of melting thermodynamics, including an estimate of $\Delta C_p$; $m_{post}$ and $b_{post}$ are the slope and intercept of the post-transition baseline, respectively. Here we have expressed the full two-state melting model using the post-transition baseline as a reference, as it is usually stable. More reliable melting
thermodynamic data \( (T_m, \Delta H_{2-st}(T_m), K(T)) \) are therefore obtained and can then be used to compute the additional melting thermodynamics \( -\Delta G_{2-st}(T_m) = -RT_m \ln K(T_m) \) and \( \Delta S_{2-st}(T_m) = (\Delta H_{2-st}(T_m) - \Delta G_{2-st}(T_m))/T_m \). In our lab, non-linear least squares fitting of equations (2-7) - (2-9) to the melting thermogram is achieved using the Levenberg-Marquardt method [168] with local (linear least-squares) estimates used as initial guesses. A thermogram fit using this method is provided in Figure 2-2B. The fit is very good (which is required and typical), but not perfect, despite the selective (and misleading) display of only perfect data generally seen in past publications.

**Table 2-1** Calorimetric and two-state thermodynamic parameters

<table>
<thead>
<tr>
<th>Analysis Method</th>
<th>( T_m ) (°C)</th>
<th>( \Delta H^\circ ) (kcal mol(^{-1}))</th>
<th>( \Delta S^\circ ) (cal mol(^{-1}) K(^{-1}))</th>
<th>( \Delta C_p ) (cal mol(^{-1}) K(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calorimetric</td>
<td>61.1 (± 0.5)</td>
<td>412 (± 13)</td>
<td>1140 (± 34)</td>
<td>1790 (± 760)</td>
</tr>
<tr>
<td>Two-State</td>
<td>60.8 (± 0.5)</td>
<td>390 (± 12)</td>
<td>1077 (± 33)</td>
<td>1340 (± 335)</td>
</tr>
</tbody>
</table>

Determined from DSC data for the helix-to-coil transition for the duplex 5’-\textit{gaDacagt}Daag-3’/5’-ctttaactgtttc-3’, where \( D \) is the 2-amino-adenine LNA super-base. DSC data obtained at a \( C_T = 69.3 \) μM in buffer containing 1M NaCl, 10 mM Na\(_2\)HPO\(_4\) (pH=7.0), and 1 mM Na\(_2\)EDTA.

Finally, if the pre-transition baseline is not sufficiently defined to permit direct estimation of \( \Delta C_p \), the thermogram may be fit to equations (2-7) – (2-9) using a fixed value for \( \Delta C_p \). DSC-based measurements of the heat-capacity change per base-pair \( \Delta C_p^{bp} \) typically lie between ca. 125 and 300 J (mol bp\(^{-1}\) K\(^{-1}\) [161, 169]), with our group reporting and typically using a value of 175 ± 70 J (mol bp\(^{-1}\) K\(^{-1}\) [136], from which \( \Delta C_p \) can be estimated as

\[
\Delta C_p = n_{bp} \Delta C_p^{bp}
\]

(2-10)
where \( n_{bp} \) is the number of base pairs (natural or modified) in the duplex. Table 2-1 shows that melting thermodynamics obtained from two-state model fits of the thermogram shown in Figure 2-1B using these various approaches are quite comparable for good-quality DSC data – the values are generally equivalent within reported experimental errors. For DSC data of poorer quality, including those showing unstable baselines, use of equations (2-7) – (2-9) with a fixed value of \( \Delta C_p \) computed from equation (2-10) is recommended.

DSC Insight #6: Accurate DSC-based determination of melting thermodynamics for dsDNA samples is only possible on a properly functioning and calibrated instrument. Collection of “calibration” data for a duplex for which melting thermodynamics at fully specified conditions have been reported by one or more reputable laboratories (e.g., see Chiu et al., 2003[170]) should precede all studies on new sequences. Parallel studies on a sequence of interest using your instrument and within a reputable core facility providing DSC expertise (e.g., The UBC Centre for Bio-Thermodynamics) may also serve this purpose well.

2.4 MELTING THERMODYNAMICS FROM UV SPECTROSCOPY

Though next-generation DSC instruments are motivating measurements of DNA melting thermodynamics by calorimetry, spectroscopy, most notably UVM, remains the far more commonly used technique. In a standard dsDNA melting study by UVM, absorbance \( (A) \) at a single wavelength, most often 260 nm \( (A_{260nm}) \) but sometimes 268 nm (where the same extinction coefficient applies to melting of either g/c or a/t base pairs), is measured as a function of temperature. The \( A_{260nm}(T) \) data so collected, when interpreted based on the fact that the \( A_{260nm} \) of dsDNA is lower than that of the constituent single-stranded oligonucleotides, allow one to monitor the melting transition and from it the value of \( \alpha \) at each temperature. The observed change in absorbance accompanying the
melting transition (Figure 2-4A) – a phenomenon known as the hyperchromic effect – arises in large part through the loss of hydrogen bonding between paired bases and disruption of stacking interactions between neighboring base pairs [154]. The resulting variation in $A_{260nm}$ readings across a melting transition is typically little more than 0.1 absorbance units, but this is generally sufficient resolution to regress accurate melting thermodynamics at $T_m$. 
Figure 2-4 UVM data for LNA-substituted duplex 5’-gaDacagtDaag-3’/5’-ctttaactgttc-3’ obtained at a $C_T = 75 \ \mu M$ in buffer containing 1M NaCl, 10 mM Na$_2$HPO$_4$ (pH=7.0), and 1 mM Na$_2$EDTA. (A) Raw $A_{260nm}(T)$ data for sample and reference buffer, (B) reference buffer-normalized $A_{260nm}(T)$ data, (C) $d^2(A_{260nm})/dT^2$ values with regions where $d^2(A_{260nm})/dT^2 = 0$ identified, and (D) $\alpha(T)$ data computed using standard baseline fit.

### 2.4.1 STRENGTHS, WEAKNESSES AND WHAT CAN BE DETERMINED RELIABLY

As detailed below, the standard procedure for computing melting thermodynamics from raw $A_{260nm}(T)$ data requires one to assume that the $A_{260nm}$ reading at any temperature represents a linear combination of two spectral components – one specific to the DNA duplex and the other to the two single strands. Thus, the analysis assumes that the melting transition obeys the two-state reaction model. This assumption is appropriate for many natural and modified dsDNA sequences, but Haq et al. [171] have shown that it is
not necessarily true, as spectral analyses of some duplexes indicate the presence of multiple intermediate species (whose concentrations vary) within the melting envelope. Much more complex spectroscopic and data analysis methods (e.g. singular value decomposition methods) are then required to properly interpret the melting transition and extract useful thermodynamic data from it [171]. In these cases, the main advantages in analyzing dsDNA melting thermodynamics by UVM – namely i) the availability and generally low cost of the required equipment, and ii) the relative simplicity of the data analysis algorithms – are lost or at least substantially reduced. For dsDNA sequences whose melting transition can be described by two-state theory, a UVM measured melting transition may be analyzed to obtain accurate melting thermodynamics for a duplex at a given $C_T$. Ideally, one would regress that UVM data to the fundamental thermodynamic potential for the melting reaction, $\Delta G$:

$$\Delta G(T) = -RT \ln(K(T))$$

$$= \Delta H_{2-st}^0 + \Delta C_p(T - T_o)$$

$$- T \left[ \Delta S_{2-st}^0 + \Delta C_p \ln \left( \frac{T}{T_o} \right) \right]$$

(2-11)

where $K(T)$ is given by equation (2-4) or (2-5) depending on whether the dsDNA is formed from non-self-complementary or self-complementary strands, respectively and $T_o$ is the temperature at the defined reference state. This would permit one to not only determine $T_m$, $\Delta H_{2-st}$ and $\Delta S_{2-st}$, but $\Delta C_p$ as well. However, while the UVM method accurately monitors the melting transition, it does not measure the melting thermodynamics directly. As a result, only $T_m$, $\Delta G$ and first-derivative properties of $\Delta G$ (i.e., $\Delta H_{2-st}$ and $\Delta S_{2-st}$) can be determined from direct fitting of the two-state reaction model to a given UVM-derived melting transition; $\Delta C_p$ cannot[172].

### 2.4.2 SAMPLE PREPARATION

Chemically modified oligonucleotides purchased from a licensed vendor (e.g. IDT Inc.; Coralville, IA or Exiqon Inc.; Vadbæk, Denmark) are resuspended in
phosphate (1 M NaCl, 10 mM NaHPO$_4$ (pH 7), and 1 mM Na$_2$EDTA) or cacodylate (1 M NaCl, 10 mM sodium cacodylate ((CH$_3$)$_2$AsO$_2$Na•3H$_2$O) (pH 7), and 0.5 mM Na$_2$EDTA) standard buffers. $A_{260nm}$ readings for each such stock solution are then recorded at 25 °C and 80 °C to define the ssDNA concentration and, if necessary, adjust it (by buffer addition) to a desired value. The required extinction coefficient is sequence specific and may be computed using any of a number of accurate methods (see, for example, Tataurov et al., 2008 [173]). We generally employ the method available on the website of IDT Inc. (http://www.owczarzy.net/extinctionDNA.htm). Significant difference in the $A_{260nm}$ readings (or $C_T$ estimates) at 25 °C and 80 °C likely indicates formation of stable secondary structure(s) (e.g., a hairpin) in the oligonucleotide at lower temperatures, which should be avoided if at all possible for reasons defined in section 2.3. The two ssDNA solutions are then prepared and mixed at equimolar concentration as before using filtered buffer to remove insoluble materials that might distort the spectral readings. In preparing the dsDNA sample, the $C_T$ should be set so as to record $A_{260nm}$ readings for the complete melting transition (baselines included) within the linear range (ideally at the higher end of that range) of the spectrophotometer when employing either a 1 cm or lower volume optical path-length quartz cuvette. For melting of dsDNA 8 – 20 bp in length in a standard 1-cm path-length cuvette, a $C_T$ between 2 and 10 µM is often appropriate.

2.4.3 EXPERIMENTAL PROCEDURE

UVM-based monitoring of a dsDNA melting transition requires a UV/VIS spectrophoto-meter having high-quality optics, excellent signal stability in the UV/far-UV region, and a sample compartment equipped with a peltier-type thermopile (heating block) that simultaneously heats two or more cuvette-held samples at a user-programmed rate with precise temperature monitoring. Our laboratory and associated UBC Centre for Bio-Thermodynamics employ a custom 12-cell Cary 100E instrument (Agilent).

A solution of the dsDNA-free buffer is prepared and then degassed at 80 °C (Thermovac heating block) for 5 – 10 min so as to avoid formation of bubbles at higher
temperatures that might interfere with $A_{260\text{nm}}$ readings. Each cuvette to be used is cleaned, loaded with degassed reference buffer and then sealed, in our lab by tightly securing each cuvette cap using high-$T$ resistant Teflon tape (other methods to avoid evaporation of solution during heating are also used [153]). One of the caps bears a thermocouple that you have pre-calibrated to record the temperature of the sample it seals. Each sealed buffer solution is loaded into the spectrophotometer idling at 25 °C, taking careful precaution to properly align the cuvette, and the cuvette-held solutions are equilibrated until the thermocouples in the heating block and sample read 25.0 ± 0.3 °C. The user-programmed scanning cycle is then initiated, starting with a temperature down-scan, usually from 25 °C to 20 °C. In UVM, we typically avoid scanning below 20 °C, as condensation events at temperatures below ambient can mar melting transition data. Methods for solving this problem have been proposed (e.g., continuously flooding the sample-block chamber with dry nitrogen gas), but have proven unreliable in our experience. In a typical cycle, the samples are re-equilibrated at 20 °C for 10 min, then subjected to an up-scan from 20 °C to 95 °C at a fixed linear scan rate of no more than 0.8 °C min$^{-1}$, during which $A_{260\text{nm}}(T)$ data are recorded at the chosen filter period for the instrument (we collect data every 0.5 °C). Finally, the sample is subjected to a down-scan and the full cycle repeated two or more times.

**UVM Insight #1:** Compared to DSC, accurate determination of melting thermodynamics by UVM generally requires the use of a slower scan rate. This is due in part to the thermal mass and insular properties of the cuvettes, which together serve to slow the rate of sample heating. Scan rates ≤ 0.5 °C are therefore often required.

Ideally, the buffer used in UVM studies shows no absorbance at the chosen wavelength across the entire temperature range, and this is the case for the standard phosphate and cacodylate buffers most often employed. If changes in absorbance are recorded in the pure buffer scans, it suggests that the cuvette(s) is not clean or the buffer is contaminated. Data for the pure buffer scans are also used to verify that the
thermocouples in the sample heating block and the selected sample are functioning properly, and that the rate of heating is sufficiently slow to keep the block and the sample(s) in thermal equilibrium.

Cuvettes other than that containing the thermocoupled reference buffer may then be cleaned and loaded with a dsDNA sample that has been degassed at 80 °C for 5 – 10 min. Loading and sealing of the cuvettes is as described for the reference buffer. Prior to data acquisition, the sample is first heated to 95 °C for 3 minutes, then cooled to 20 °C, typically at 1 °C to 3 °C min\(^{-1}\), and held at 20 °C until the thermocouples in the heating block and sample equilibrate. The user-programmed scanning cycle is then initiated. Melting transition data for 3 up-scans are recorded and overlaid to verify that they are substantially superimposable – this confirms both proper sample handling and thermodynamic reversibility. If they do not, the scan rate is reduced and the study repeated.

Following the last scanning cycle, the sample temperature is raised to 80 °C and the \(A_{260nm}\) reading at that temperature used to check \(C_T\).

2.4.4 ANALYSIS OF UVM DATA

Buffer-normalized \(A_{260nm}\) data for each scan are plotted against \(T\) to visualize the melting transition and regress melting thermodynamics. If the reference buffer does not absorb, the raw (Figure 2-4A) and buffer-normalized (Figure 2-4B) melting transition data are essentially indistinguishable. Here we show that one may analyze buffer-normalized \(A_{260nm}(T)\) data using either a more accurate modern non-linear approach or any one of the classic methods.

**Modern Analysis Method**: Provided both the pre- and post-transition baselines are stable and linear, buffer-normalized \(A_{260nm} \text{ vs-} T\) may be used to compute \(\alpha(T)\) using the relation

\[
A_{260nm}(T) = (m_{pre}^{UVM}T + b_{pre}^{UVM})\alpha(T) + (m_{post}^{UVM}T + b_{post}^{UVM})(1 - \alpha(T))
\]  

(2-12)
where $m_{\text{pre}}^{UVM}$ and $b_{\text{pre}}^{UVM}$, and $m_{\text{post}}^{UVM}$ and $b_{\text{post}}^{UVM}$, are the slopes and intercepts of the pre- and post-transition baselines, respectively. Fitting of equations (2-8) and (2-9) to the resulting $\alpha(T)$ generally yields high-quality values of $K(T)$, $\Delta H_{2-st}^0$ and $T_m$. In equation (2-9), $K(\alpha = 0.5)$ is given by either equation (2-4) or (2-5), and the Levenberg-Marquardt algorithm is used to fit the model. As a UV-melting transition cannot be used to determine $\Delta C_p$, its value in equation (2-9) must either be computed using equation (2-10) or set equal to zero for this regression, which we find converges well if $T_{max}$ (given by the temperature at which $d(A_{260\text{nm}})/dT$ is a maximum) is used as the value of $T_m$ in the first iteration. $\Delta S_{2-st}^0$ is then computed as $(\Delta H_{2-st}^0 + RT_m \ln[K(\alpha = 0.5)])/T_m$.

Different methods have been proposed to select and fit the two required baselines. The standard approach, reflected in equation (2-12), is to compare linear least squares fits of 8–10 °C spans of the putative baseline region to identify that span having the highest $R^2$ value – this sets the slope and intercept values required in equation (2-12). If applied to the pre-transition baseline, it also sets the temperature $T_1$ defining the start of the melting transition, which is given by highest temperature within that 8–10 °C span. Setting of $T_2$, the transition end-point, and $m_{\text{post}}^{UVM}$ and $b_{\text{post}}^{UVM}$ is achieved in an analogous manner.  

**UVM Insight #2:** Other more advanced methods for fitting pre- and post transition baselines may be used. In our view, among the most useful is that developed by Owczarzy [174], who found that high-quality melting thermodynamics can often be obtained using linear baselines fit to regions where $d^2(A_{260\text{nm}})/dT^2 = 0$. An example of this approach is shown in Figure 2-4C for the data reported in Figure 2-4B. One challenge to this approach is the noise created by numerical determination of second derivatives, which can make identification of the baseline regions and the values of $T_1$ and $T_2$ challenging. A data-smoothing filter is therefore sometimes applied, which can ease assignments but also introduce error.
Melting thermodynamic parameters determined for the example sequence (Figure 2-4B) using these two different baseline fit strategies are shown in Table 2-2.

Table 2-2 Melting thermodynamics for the LNA-substituted duplex 5’- ttcataCcg-3'/5’- acggctatgaa-3’

<table>
<thead>
<tr>
<th>Baseline Fit Method</th>
<th>$T_m$ (°C)</th>
<th>$\Delta H(T_m)$ (kcal mol$^{-1}$)</th>
<th>$\Delta S(T_m)$ (cal mol$^{-1}$ K$^{-1}$)</th>
<th>$\Delta G(T_m)$ (kcal mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>65.5 (± 0.2)</td>
<td>328 (± 7)</td>
<td>878 (± 18)</td>
<td>30.6 (± 0.1)</td>
</tr>
<tr>
<td>$d(A_{260nm})^2/dT^2 = 0$</td>
<td>65.6 (± 0.3)</td>
<td>334 (± 7)</td>
<td>895 (± 19)</td>
<td>30.7 (± 0.1)</td>
</tr>
</tbody>
</table>

Determined from UVM data by regression to the two-state reaction model (eqs (2-4),(2-5),(2-8),(2-9), and (2-12)) using the standard and Owczarzy[174] baseline fitting methods. UVM data obtained at a $C_T = 75 \mu$M in buffer containing 1M NaCl, 10 mM Na$_2$HPO$_4$ (pH=7.0), and 1 mM Na$_2$EDTA.

Classic Analysis Methods: In most published analyses of dsDNA melting transition data acquired by UVM, $\Delta C_p$ is usually ignored for the reasons given in section 2.4.1. The fundamental potential used for regressing melting thermodynamics is then given by

$$\Delta G(T) = -RT \ln(K(T)) = \Delta H_{2\text{-st}}^\circ - T \Delta S_{2\text{-st}}^\circ$$  \hspace{1cm} (2-13)

where, because we have assumed $\Delta C_p = 0$, $\Delta H_{2\text{-st}}$ and $\Delta S_{2\text{-st}}$ are now temperature independent. As in equation (2-11), the superscript $^\circ$ denotes that $\Delta H_{2\text{-st}}$ and $\Delta S_{2\text{-st}}$ are defined at the standard state, which for UVM experiments is typically DNA suspended in either the standard phosphate or cacodylate buffer (pH 7) at atmospheric pressure. $K(T)$ is given by equation (2-4) for dsDNA melting into two non-self-complementary strands,
and by equation (2-5) for two self-complementary strands. \( K \) can thereby be expressed as a function of \( \alpha(T) \), and equation (2-13) rearranged to give

\[
T = \frac{\Delta H_{2-st}^0}{\Delta S_{2-st}^0 - R\ln[K(\alpha(T))]} \tag{2-14}
\]

In classic analyses of UVM data, the regression process requires one to compute \( \alpha(T) \) from the buffer-normalized UVM data. Provided the melting transition is centered at a temperature between \( ca. \) 50 °C and 70 °C, the UVM experiment will typically yield a stable post-transition baseline, making its fitting with good certainty fairly straightforward. But for chemically modified duplexes, particularly ones in which secondary structures are possible in one or both strands, fitting of the pre-transition baseline can be problematic.

Once selected, each baseline fit is extrapolated through the melting transition, permitting the lever-rule to be used to determine \( \alpha \) at each temperature within the span bounded by \( T_1 \) (where \( \alpha = 1 \)) and \( T_2 \) (where \( \alpha = 0 \)):

\[
\alpha(T) = \frac{A(T) - C(T)}{A(T) - B(T)} \tag{2-15}
\]

As shown in Figure 2-4B, \( A(T) \), \( B(T) \) and \( C(T) \) are the y-axis values at \( T \) of the extrapolated post-transition baseline, the extrapolated pre-transition baseline, and the melting transition, respectively. One may thereby plot \( \alpha(T) \) versus \( T \) for the transition (Figure 2-4D). \( T_m \) is given by the temperature within the melting transition at which \( \alpha = 0.5 \), and

\[
\Delta G(T_m) = -RT\ln[K(\alpha = 0.5)] \tag{2-16}
\]

The inverse tangent \((dT/d\alpha)\) of the melting transition shown in Figure 2-4D is then evaluated at each \( T \) to compute \((dT/d\alpha)\) values. Those data may be fit using the non-linear least squares (Levenberg-Marquardt) method to the two-state model relation derived by differentiation of equation (2-14).

63
\[
\frac{dT}{d\alpha} = -R \frac{\Delta H^0_{2-st}}{(\Delta S^0_{2-st} - R \ln[K(\alpha)])^2} \frac{(d[K(\alpha)])}{d\alpha} = \frac{-RT^2}{\Delta H^0_{2-st}} \frac{(d[K(\alpha)])}{d\alpha}
\]  

(2-17)

A value of \(\Delta H^0_{2-st}\) may thereby be regressed, and \(\Delta S^0_{2-st}\) then be determined as \(\Delta S^0_{2-st} = (\Delta H^0_{2-st} - \Delta G(T_m))/T_m\). For melting of dsDNA into either non-self-complimentary strands, where \(K(\alpha)\) is given by equation (2-4), or self-complementary strands \((K(\alpha)\) given by eq. 4B), equation (2-17) can be expressed as

\[
\frac{dT}{d\alpha} = \frac{-RT^2}{\Delta H^0_{2-st}} \frac{1 + \alpha}{\alpha(1 - \alpha)}
\]  

(2-18)

Provided the pre- and post-transition baselines are well defined and stable, the above regression approach generally represents a reliable method of determining melting thermodynamics from UVM data. In practice, the creation and use of plots such as those shown in **Figure 2-4C** and **Figure 2-4D** are not done, as the numerical methods needed to generate \(\alpha(T)\) and \(d\alpha/dT\) values from the buffer-normalized melting transition are readily available in many commercial programs, including Matlab and Excel.

**UVM Insight #3**: Equation (2-18) does not depend on \(C_T\). As a result, relative to DSC studies, the accuracy of \(\Delta H^0_{2-st}\) values determined from UVM data is less sensitive to precise knowledge of strand concentration. HPLC purified strands therefore may not be required, which is a significant advantage when studying melting of expensive duplexes comprised of chemically modified oligonucleotides.

A simplified version of this data regression algorithm may also be used, often to good effect. It is based on the fact that \(T = T_m\) when \(\alpha = 0.5\), and at that temperature equation (2-18) reduces to

\[
\Delta H^0_{2-st} = -6RT_m^2 \frac{\partial \alpha}{dT} \bigg|_{T_m}
\]  

(2-19)
Here, simple linear least squares fitting of $\alpha(T)$ data within a narrow temperature range centered about $\alpha = 0.5$ may be used to estimate $\frac{d\alpha}{dT}|_{T_m}$, which when introduced into equation (2-19) allows one to determine $\Delta H^{o}_{2-st}$.

Finally, and particularly for dsDNA sequences displaying ill-defined pre-transition (or post-transition) baselines, we note that accurate determination of $T_m$ from UVM data may not be possible. In this case, $T_m$ is usually assumed to be equal to $T_{max}$ for the buffer-normalized melting transition. Measurement of $T_{max}$ as a function of $C_T$ then allows determination of $\Delta H^{o}_{2-st}$ and $\Delta S^{o}_{2-st}$ through least squares fitting of a linear form of equation (2-14)

$$\frac{1}{T_m} = -\frac{R}{\Delta H^{o}_{2-st}} \ln[K(\alpha = 0.5)] + \frac{\Delta S^{o}_{2-st}}{\Delta H^{o}_{2-st}}$$

where $K(\alpha = 0.5)$ equals $C_T$ or $C_T/4$ for melting of self-complementary or non-self-complementary strands, respectively. The display of $1/T_m$ data as a function of $C_T$ is known as a van’t Hoff plot, from which a fairly accurate estimate of $\Delta H^{o}_{2-st}$ can be obtained from the slope.

**UVM Insight #4:** In theory, $\Delta S^{o}_{2-st}$ can be estimated from the intercept of a van’t Hoff plot. However, data extrapolation is required, and this generally leads to unacceptably large errors. One is therefore better served to compute $\Delta S^{o}_{2-st}$ as $(\Delta H^{o}_{2-st} + RT_m \ln[K(\alpha = 0.5)])/T_m$.

Curvature observed in a van’t Hoff plot can be theoretically related to a non-zero $\Delta C_p$. We find, however, that UVM data are rarely of sufficient quality to yield a reliable estimate of $\Delta C_p$ by a van’t Hoff plot analysis due to a combination of instrument noise,
uncertainties in baseline assignments, and the compensating nature of $T$ dependencies of $\Delta H$ and $\Delta S$ that result in a weak $T$ dependence of $\Delta G$.\textsuperscript{2}

Finally, though arguably the most popular method, analysis of UVM data through a van’t Hoff plot is error prone. $T_m$ and $T_{\text{max}}$ are not equal, though they generally are close in value. Additional sources of error are known [175], and together they serve to reduce the precision of melting thermodynamics obtained using this classic method.

### 2.5 ERRORS ANALYSIS

Determination of total errors in melting thermodynamic values determined from DSC and UVM experiments is done in a similar manner. For brevity, the following methodology specifically applies to UVM data, but the changes needed to apply it to DSC data are minor and obvious.

Sources of error in any thermodynamic parameter (e.g., $T_m$, $\Delta H$, $\Delta S$, etc.) determined from analysis of a single buffer-normalized UVM scan include uncertainties (standard deviations $\sigma$) in the subtracted reference-buffer scan ($\sigma_{\text{buffer}}$), the model fit ($\sigma_{\text{model}}$), and the baselines selected ($\sigma_{\text{baseline}}$). For $\Delta H$, for example, these uncertainties are typically of order $\pm 0.0003$ AU, $\pm 1.7$ kJ mol$^{-1}$, and $\pm 6.7$ kJ mol$^{-1}$, respectively. It is, however, easily shown that uncertainty in the subtracted reference-buffer scan makes a negligible contribution to the total error ($\sigma_i$) in any parameter derived from analysis of an individual UVM scan $i$. Thus

$$\sigma_i = \sqrt{\sigma_{\text{model}}^2 + \sigma_{\text{baseline}}^2}$$

\text{\textsuperscript{2} Though it cannot be done for melting of either short duplex DNA or hairpin structures with ssDNA (the $T_m$ for unimolecular melting reactions has no dependence on $C_T$), UVM data can be used to estimate $\Delta C_P$, in some cases with good accuracy, when applied to melting of triplex complexes and higher molecularity complexes [175]
the propagated error ($\sigma_j$) in the parameter (in this example $\Delta H$) determined from independent analyses of $N$ replicate scans of the same sample is then given by

$$\sigma_j = \sqrt{\sigma_{\text{instrument}}^2 + \frac{1}{N-1} \sum_{i=1}^{N} (\Delta H_i - \overline{\Delta H})^2}$$

(2-22)

where $\sigma_{\text{instrument}}$ gives the variance in instrument performance (temperature control and scan rate uncertainties, etc.), $\Delta H_i$ is the value of the parameter determined from scan $i$, and $\Delta H_j (= \overline{\Delta H})$ is the mean value computed from the $\Delta H_i$ for all scans analyzed for that sample.

If $M$ completely independent preparations of the same sample are made, the total propagated error ($\sigma_k$) in the parameter is computed as

$$\sigma_k = \sqrt{\sigma_{\text{sample}}^2 + \frac{1}{M-1} \sum_{i=1}^{M} (\Delta H_j - \Delta H_k)^2}$$

(2-23)

where $\sigma_{\text{sample}}$ gives the uncertainty in sample preparation (pipetting error, $C_T$ value, etc.) and $\Delta H_k (= \Delta H_j)$. Although the errors analysis above seeks to parse out the contributions to the overall error in a given parameter from each step in data collection and analysis, generally the error of $\Delta H$, for example, is not computed this way. Instead, the error in $\Delta H$ for a duplex is computed as the standard deviation of $\Delta H_j$ from $N$ replicate scans or $\Delta H_k$ from $M$ replicate samples.

2.6 UNDERSTANDING THE CONTRIBUTION TO DUPLEX STABILITY OF SPECIFIC NUCLEOTIDE MODIFICATIONS

As described here, DSC and UVM experiments provide melting thermodynamics for natural and modified dsDNA at the measured $T_m$ for the specified $C_T$. In addition, DSC studies have and can provide knowledge of $\Delta C_p$. However, in the vast majority of
studies reporting melting thermodynamics for these systems, the analyses ignore $\Delta C_p$ entirely. This is both regrettable and problematic. To understand why, let us consider the question of the precise contribution that replacing a single nucleotide in a natural DNA duplex with its corresponding LNA makes to the thermal stability of the duplex. As noted, the primary function of the LNA modification is to lock the furanose sugar into a C3’-endo conformation. This lowers the entropy of the nucleotide in its ssDNA state. To a first approximation then, one would expect that an LNA substitution enhances the thermal stability of dsDNA by reducing the entropy loss accompanying base-pair formation. Other factors may contribute, but one would therefore expect to observe an entropic contribution to the stability enhancement.

To rigorously determine whether this is indeed the case, one must compare measured melting thermodynamics for the un-substituted and LNA-substituted duplexes at the same temperature – the $T_m$ of the un-substituted duplex, say. As the $T_m$ of the LNA-substituted duplex ($T_{m,LNA}$) will differ from that of the un-substituted duplex ($T_{m,DNA}$), thermodynamics stipulates that knowledge of $\Delta C_p$ is required. For the melting enthalpy and entropy, for example,

$$\Delta H_{LNA}(T_{m,DNA}) = \Delta H_{LNA}(T_{m,LNA}) + \Delta C_p(T_{m,DNA} - T_{m,LNA})$$

(2-24)

and

$$\Delta S_{LNA}(T_{m,DNA}) = \Delta S_{LNA}(T_{m,LNA}) + \Delta C_p \ln \left( \frac{T_{m,DNA}}{T_{m,LNA}} \right)$$

(2-25)

One may then determine the net enthalpic and entropic contributions to thermal stability provided by the LNA substitution through the difference relations

$$\Delta \Delta H_{LNA}(T_{m,DNA}) = \Delta H_{LNA}(T_{m,DNA}) - \Delta H_{DNA}(T_{m,DNA})$$

(2-26)

and

$$\Delta \Delta S_{LNA}(T_{m,DNA}) = \Delta S_{LNA}(T_{m,DNA}) - \Delta S_{DNA}(T_{m,DNA})$$

(2-27)
Table 2-3 reports the net contribution of a single LNA substitution to the stability of a representative duplex using this analysis and melting thermodynamics data we measured and reported previously[1]. As expected, the LNA substitution results in a $\Delta \Delta S_{LNA} < 0$, indicating that the substitution does indeed thermally stabilize the duplex by reducing the entropy gain accompanying the helix-to-coil transition. This conclusion is supported by the value of $\Delta \Delta H_{LNA}$, which is statistically negligible, indicating the substitution provides no net enthalpic contribution to the observed enhancement in duplex stability.
Table 2-3 Net (incremental) enthalpic and entropic contributions of an LNA substitution

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Units</th>
<th>5’-ggeacgcttcg-3’&lt;sup&gt;b&lt;/sup&gt;</th>
<th>5’ggcaCgcttcg-3’&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_m$</td>
<td>(°C)</td>
<td>68.2 (± 0.5)</td>
<td>73.2 (± 0.5)</td>
</tr>
<tr>
<td>$\Delta G(37 , ^\circ C)$</td>
<td>(kJ mol&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>61.9 (± 1.3)</td>
<td>67.4 (± 1.3)</td>
</tr>
<tr>
<td>$\Delta H(T_m)$</td>
<td>(kJ mol&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>370 (± 11)</td>
<td>377 (± 11)</td>
</tr>
<tr>
<td>$\Delta S(T_m)$</td>
<td>(J mol&lt;sup&gt;-1&lt;/sup&gt; K&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>993 (± 29)</td>
<td>999 (± 30)</td>
</tr>
<tr>
<td>$\Delta C_p$&lt;sup&gt;a&lt;/sup&gt;</td>
<td>(J mol&lt;sup&gt;-1&lt;/sup&gt; K&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>1960 (± 490)</td>
<td>1440 (± 360)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Incremental Changes</th>
<th>$\Delta C_p = 0$ Assumption</th>
<th>$\Delta C_p &gt; 0$&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta\Delta G(37 , ^\circ C)$</td>
<td>(kJ mol&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>5.4</td>
</tr>
<tr>
<td>$\Delta\Delta H(T_m, DNA)$</td>
<td>(kJ mol&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>7.5</td>
</tr>
<tr>
<td>$\Delta\Delta S(T_m, DNA)$</td>
<td>(J mol&lt;sup&gt;-1&lt;/sup&gt; K&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>6.3</td>
</tr>
<tr>
<td>$\Delta T_m$</td>
<td>(°C)</td>
<td>5.0</td>
</tr>
</tbody>
</table>

<sup>a</sup> Duplex melting was conducted at a $C_T = 75$ μM in buffer containing 1M NaCl 10mM Na<sub>2</sub>HPO<sub>4</sub> (pH=7.0), and 1mM Na<sub>2</sub>EDTA.<sup>b</sup> Reported values are those determined from model-regression to the DSC thermogram for the specified duplex. Melting thermodynamic data are reported for the helix-to-coil transition for complementary duplexes when using a measured value for $\Delta C_p$ or assuming $\Delta C_p = 0$.<sup>c</sup> Values for $\Delta\Delta G(37 \, ^\circ C)$, $\Delta\Delta H(T_m, DNA)$, and $\Delta\Delta S(T_m, DNA)$ were computed using the measured $\Delta C_p$ for the LNA-substituted duplex.

One may then ask how this analysis would change if $\Delta C_p$ was ignored. The corresponding values of $\Delta\Delta H_{LNA}$ and $\Delta\Delta S_{LNA}$ computed based on this assumption are also reported in Table 2-3. Because $\Delta H$ increases more rapidly with temperature than does $\Delta S$, this traditional (but thermodynamically imprecise) approach will often predict, as observed here, that the net contribution of the LNA substitution to the increase in thermal stability is largely enthalpically driven. Indeed, for this particular duplex,
ΔΔ\(H_{LNA}\) is so favorable that locking the furanose sugar is predicted to **destabilize the duplex entropically**.

This single example therefore highlights the fact that the ability to accurately measure melting thermodynamics, including \(ΔC_p\), can serve to greatly improve our understanding of exactly how nucleotide modifications affect duplex stabilities. And this in turn should help to improve and extend the uses of these novel reagents.
CHAPTER 3: MOLECULAR THERMODYNAMICS OF LNA:LNA BASE PAIRS AND THE HYPERSTABILIZING EFFECT OF 5’-PROXIMAL LNA:DNA BASE PAIRS

3.1 SYNOPSIS

Locked nucleic acids (LNAs) can can greatly enhance duplex DNA stability, and are therefore creating opportunities to improve therapeutics, as well as PCR-based disease and pathogen diagnostics. Realizing the full potential of LNAs will require better understanding of their contributions to duplex stability, and the ability to predict their hybridization thermodynamics. We present melting thermodynamics data for a large set of diverse duplexes containing LNAs in one or both strands. Those data reveal that LNAs, when present on both strands, can stabilize a duplex not only through direct interaction with their base-pair partner, but also through non-local hyperstabilization effects created by LNA:LNA base pairs and/or specific patterns of oppositely oriented LNA:DNA base pairs. The data are therefore used to extend a thermodynamic model previously developed in our lab to permit accurate prediction of melting temperatures for duplexes bearing LNA substitutions within both strands using a classic group-contribution approach.

3.2 INTRODUCTION

Significant advances in non-natural nucleotide chemistry have been made over the past quarter century, motivated in part by the desire to enhance diagnostic and therapeutic applications of structured oligonucleotides by improving their thermodynamic and chemical (e.g., nuclease-resistance) stability, or by altering their cellular uptake.[176] Among the most significant and widely used nucleotide modifications[177, 178] is the locked nucleic acid (LNA),[115, 179] in which a methylene bridge has been introduced between the 4’-carbon and 2’-oxygen of the ribose sugar to “lock” it in a C3’-endo (RNA-like) conformation.[180-182] Replacement of a DNA or RNA nucleotide with its
LNA-containing oligonucleotides are finding increasing use as therapeutics, including as gene-silencing agents,[184, 185] and as DNA or RNA aptamers.[186, 187] The introduction of LNA monomers into a standard ssDNA or ssRNA aptamer can greatly improve both the thermodynamic and serum stability of the structured single-strand, without introducing immunogenicity or toxicity effects that could otherwise deter therapeutic use.[188] LNA-substituted oligonucleotides have also found widespread use in a range of diagnostic applications, including real-time PCR, digital PCR and hybridization-array based detection of either germline mutations (e.g., single nucleotide polymorphisms) or somatic mutations associated with disease,[189, 190] as well as detection of either viral or microbial pathogens within foods and drinking water sources.[191]

The growing importance of LNAs in oligonucleotide-based medicine and diagnostics has motivated careful analysis of the properties of LNA-containing ssDNA, ssRNA, dsDNA and hybrid DNA:RNA duplexes relative to their corresponding isosequential LNA-free forms.[106, 142, 192, 193] A number of chemical and structural factors have thereby been found to contribute to LNA-mediated enhancement of thermal stability. NMR studies show that substitution of an LNA monomer into complementary dsDNA induces the modified base pair and base pairs proximal to it to shift towards an A-form helix; a shift in the deoxyribose sugars towards a C3’-endo conformation is also observed.[101] An LNA substitution within the ssDNA of a DNA:RNA hybrid duplex likewise shifts the sugar toward a C3’-endo conformation.[101, 109] This conformational change has also been characterized using both resonance Raman spectroscopy[194] and
transient absorption[195, 196] to intensify certain base-stacking interactions that are important determinants of secondary structure of LNA-containing duplexes, leading, for example, to a tendency toward the more thermally stable A-form in AT-rich sequences. Motivated by these and other structural insights[182, 197, 198], thermodynamicists have conducted extensive studies aimed at quantifying the enthalpic and entropic contributions to the enhanced thermal stability of an LNA:DNA or LNA:RNA base pair.[100, 106, 129, 142, 180, 181, 192] Those studies, which typically rely on interpretation of melting thermodynamics collected by either UV spectroscopy[100, 106, 142, 180, 181, 192] or differential scanning calorimetry,[192, 199] show that the change in melting temperature ($\Delta T_m$) accompanying an LNA substitution is sequence dependent, with the dominant determinant of the degree of stabilization being the type of base substituted.[142, 192]

But for short B-form dsDNA, perhaps the greatest insights into the effect of LNA substitutions on duplex stability come from efforts to treat melting thermodynamic data with classic group-contribution modeling approaches, which have previously been used to great effect by Prausnitz[200, 201] and others[202, 203] to model thermodynamic behavior in a variety of other important liquid mixtures. The most widely used group-contribution type models developed for predicting $T_m$ and melting thermodynamics of unsubstituted and LNA-substituted dsDNA are collectively known as nearest-neighbor thermodynamic (NNT) models.[106, 129, 142, 192] Differences between available NNT models exist when applied to unsubstituted dsDNA, but collectively they predict that dsDNA melting thermodynamics depend on duplex length, as well as on the sequence and the overall $g – c$ base-pair content. They assume, in accordance with more advanced theories derived from statistical mechanics (e.g., the Poland-Sheraga potential[204, 205] and advances made to it by Fisher[206] and others[207]), that the sequence dependence arises, at least in part, from the combined contribution of base-pairing and base-stacking interactions. Both interactions are short-range in nature, permitting their contribution to the melting enthalpy $\Delta H$ to be captured through a set of nearest-neighbor enthalpy parameters for all unique $N_{3}^{+m}N_{3}^{+(m+1)}/N_{5}^{+m}N_{5}^{+(m+1)}$ base-pair doublets (tandem base
pairs) in the duplex, where \( N_{3' + m} \) is the nucleotide at the 3' + m base position on the sense strand. The melting entropy (\( \Delta S \)) is handled in an analogous manner by computing the combined contribution to \( \Delta S \) of base-pair formation and base-stacking interactions, which are both orientationally (e.g., bases stack in plane) and spatially specific; \( \Delta S \) is thus also computed as the frequency-weighted sum of nearest-neighbor contributions. NNT models provide very good predictions of \( T_m \) for dsDNA melting between ca. 45 – 65 °C.[147] In addition, good prediction of \( T_m \) values and melting thermodynamics (\( \Delta H \) and \( \Delta S \)) for duplexes melting above 65 °C may be achieved by introducing heat capacity effects into an NNT model,[136] which is consistent with experiment and theory showing that dsDNA melting involves a heat capacity change \( \Delta C_p \) of ca. 42 ± 16 cal mol\(^{-1}\) K\(^{-1}\) per base pair.[165]

NNT models that do or do not include \( \Delta C_p \) contributions have been extended to predict changes in dsDNA stability arising from LNA substitution(s), generally for the case where those substitutions occur within only one strand.[106, 142, 192] Here, we focus on models that explicitly include \( \Delta C_p \) contributions. To fix ideas, consider a short pure-DNA duplex that melts at a temperature \( T_{m,DNA} \) with a melting enthalpy and entropy of \( \Delta H_{DNA} \) and \( \Delta S_{DNA} \), respectively. Substitution of an LNA into the duplex results in an increase in melting temperature from \( T_{m,DNA} \) to \( T_{m,LNA} \). The melting enthalpy and entropy of the LNA-modified duplex are then given by

\[
\Delta H_{LNA} (T_{m,LNA}) = \Delta H_{DNA} (T_{m,DNA}) + \Delta C_p (T_{m,LNA} - T_{m,DNA}) + \Delta \Delta H_{LNA}
\]

\[
\Delta S_{LNA} (T_{m,LNA}) = \Delta S_{DNA} (T_{m,DNA}) + \Delta C_p \ln(T_{m,LNA}/T_{m,DNA}) + \Delta \Delta S_{LNA}
\]

(3-1)

where \( \Delta \Delta H_{LNA} \) accounts for the incremental change in the melting enthalpy due to intrinsic changes in base-pair and proximal duplex structure/energetics (e.g. changes in base stacking) that arise as a result of the LNA substitution and the formation of an LNA:DNA base pair. Equation (3-1) predicts that changes in \( \Delta H \) (and likewise \( \Delta S \)) may arise not only through a non-zero \( \Delta \Delta H_{LNA} \) (or by analogy \( \Delta \Delta S_{LNA} \)) value, but also simply
due to the fact that $\Delta H$ and $\Delta S$ are temperature dependent. Indeed, our recently described NNT model for LNA-containing dsDNA,[192] which accounts for the temperature dependence of $\Delta H$ and $\Delta S$ through a non-zero $\Delta C_p$, predicts that experimentally observed changes in $\Delta S$ accompanying LNA substitutions in one strand arise from both the temperature dependence of $\Delta S$ and a negative $\Delta \Delta S_{LNA}$ value that reflects restrictions to ribose sugar flexibility imposed by the methylene bridge, including the associated increase in structural preorganization of the LNA-containing single strand. Corresponding changes in $\Delta H$, however, are generally found to arise only from the temperature dependence of $\Delta H$, as $\Delta \Delta H_{LNA}$ values for dsDNA substituted with a single LNA are statistically insignificant.[192] This finding is not at odds with observed changes in structure arising from an LNA substitution; it simply indicates that structural changes induced by individual LNA substitutions (e.g., increased local A-form helix characteristics) result in a set of subprocesses (perturbation of hydrogen bond strengths, base stacking energies, etc.) that are enthalpically compensatory in nature, such that the total incremental enthalpy change, $\Delta \Delta H_{LNA}$, is either very small or athermal.

As noted, current NNT models are designed to predict melting thermodynamics of complementary or mismatched duplexes bearing LNA substitutions in one strand.[106, 142, 192] They are not intended to be applied to duplexes containing LNA substitutions within both strands, including those that form LNA:LNA base pairs. Melting studies have shown that complementary duplexes containing LNA:LNA base pairs can be exceptionally stable – a completely LNA-substituted 9-mer duplex has been shown to melt at a $T_m$ more than 60 °C higher than that of the corresponding isosequential pure-DNA duplex.[180] In principle, NNT models can be extended to permit their application to duplexes bearing LNA substitutions on both strands. However, at present, the impact on melting thermodynamics of introducing additional LNA substitutions in various positions of an opposing pure-DNA strand is not well understood. Here we address this issue, as the ability to understand and predict the thermal stability of duplex DNA bearing LNA substitutions within both strands may serve to expand applications of LNA
technology in important areas. For example, Wang et al.[79] have used LNA:LNA base pairing to improve the sensitivity and selectivity of molecular beacons, while Morandi et al.[23] have described the paired use of an LNA-substituted allele-specific primer and LNA-substituted beacon probe to detect cancer-related somatic point mutations, such as the BRAF V600E mutation associated with metastatic melanoma and colorectal cancer, with increased sensitivity and specificity. In addition, several groups have reported on the use of LNA:LNA base pairs to greatly improve the thermal and serum stability of DNA aptamers targeted for therapeutic use.[188, 208, 209]

With the ultimate aim of establishing a NNT model that can be applied to short dsDNA bearing any pattern of LNA substitutions in either or both strands, we report here melting thermodynamic data for an extensive library of complementary duplexes in which an LNA substitution is made in a pure DNA strand at different positions relative to one or more LNAs on the opposite strand. The data are used to define i) the duplex stabilization thermodynamics of an LNA:LNA base pair relative to the LNA:DNA base pairs from which it is formed, and ii) the impact on duplex stability of an LNA substitution made in one strand at a nucleotide position proximal to an existing LNA in the opposing strand. Results are then used to extend our current NNT model[192] to permit accurate prediction of $T_m$ and $\Delta T_m$ values for duplexes bearing LNA substitutions on both complementary strands.

### 3.3 METHOD

The experimental methods including descriptions of materials used, measurement of melting thermodynamics, and regression of model parameters required for this work are outlined below.

#### 3.3.1 MATERIALS

All pure DNA and LNA containing oligonucleotides were obtained from either Integrated DNA Technologies (Coralville, IA) or Exiqon Inc. (Vedbæk, Denmark). Each
strand was resuspended in buffer containing 1 M NaCl, 10 mM Na$_2$HPO$_4$, and 1 mM Na$_2$EDTA at pH = 7.0, and then quantified by UV spectrophotometry (Cary 1E; Santa Clara, CA) at 80 °C using extinction coefficients provided by the supplier.

### 3.3.2 MEASUREMENT OF MELTING THERMODYNAMICS BY UV SPECTROSCOPY

Helix-to-coil melting transition data were collected on a Cary 1E spectrophotometer equipped with a 12-cell Peltier temperature controller. All UV-monitored melt (UVM) experiments were conducted at 260 nm by scanning from 25 to 95 °C at a scan rate of 0.5 °C min$^{-1}$ to obtain absorbance ($A_{260}$) versus $T$ profiles. A cuvette with a path length of 10 mm was loaded with duplex sample to a total strand concentration $C_T$ of 5 or 7.5 µM and then sealed to prevent evaporation during heating. Buffer reference and duplex melting curves were collected in triplicate and collectively analyzed to obtain mean values and standard deviations for $T_m$, $\Delta H(T_m)$, $\Delta S(T_m)$, and $\Delta G$ (@ the reference temperature of 53 °C), hereafter referred to as $\Delta G^o$. Melting curve analysis followed a method we have previously described,[210] which assumes two-state melting thermodynamics and includes contributions from a non-zero $\Delta C_p$. Briefly, from each buffer-scan-normalized melt curve, the fraction of strands in the random-coil state ($\theta$) was determined as a function of temperature to create a fractional curve by fitting equation (3-2) to the pre-transition and post-transition baselines

$$A_{260}(T) = (m_{pre}^{UVM} T + b_{pre}^{UVM})(1 - \theta) + (m_{post}^{UVM} T + b_{post}^{UVM})\theta$$

(3-2)

where $m_{pre}$ and $b_{pre}$, and $m_{post}$ and $b_{post}$ are the slopes and intercepts of the pre and post transition baselines, respectively. A representative UVM experiment is provided in **Figure 3-1A**, within which are identified the slopes and intercepts of the two transition baselines used to normalize the melting transition data. Concordance between the normalized melting transition and two-state thermodynamics theory (equation (3-3)
below) is also demonstrated (Figure 3-1B); for all sequences studied, the experimental and two state thermodynamics agree within ±15%. $\Delta H$ and $\Delta S$ at $T_m$ were then determined by fitting equations (3-3) and (3-4) to the resulting fractional curve

$$\theta(T) = -\frac{K_d(T) + \sqrt{K_d^2(T) + 2K_d(T)C_T}}{C_T}$$ (3-3)

$$K_d(T) = \exp\left[-\left(\frac{\Delta H + \Delta C_p(T - T_m)}{RT} \right) + \frac{\Delta S + \Delta C_p \ln \left(\frac{T}{T_m}\right)}{R}\right]$$ (3-4)

where $K_d(T)$ is the temperature-dependent equilibrium constant for the melting reaction. The $\Delta C_p$ required in equations (3-3) and (3-4) was computed as $\Delta C_p = n\Delta C_p^{bp}$, where $n$ is the total number of base pairs in the duplex and $\Delta C_p^{bp}$ is the heat capacity change per base pair (bp), previously reported by Hughesman et al.[136] to be 42 cal mol$^{-1}$ K$^{-1}$ bp$^{-1}$. $T_{max}$ was used as an estimate of $T_m$ in equation (3-4) during the first solution iteration. After regressing values for $\Delta H(T_m)$ and $\Delta S(T_m)$, a new estimate of $T_m$ for non-self-complementary duplexes was then obtained from

$$T_m = \frac{\Delta H(T_m)}{\Delta S(T_m) - R\ln(C_T/4)}$$ (3-5)

where $C_T$ is the total strand concentration (M), and is divided by 4 when the two single strands are not self-complementary ($C_T/4$ is replaced by $C_T$ when the two single strands are self-complementary. The model is otherwise the same). That iteration scheme was continued until a best fit was achieved.
Figure 3-1 Melting transition measured by UV spectroscopy at 260 nm for the complementary duplex 5’-ctaacgGatgc-3’/5’-gcatacgtagt-3’ at a total strand concentration $C_T$ of 7.5 μM and in aqueous buffer (pH 7.0) containing 1 M NaCl, 10 mM Na$_2$HPO$_4$, and 1 mM Na$_2$EDTA. (A) Raw melting transition data; the slope ($m$) and intercept ($b$) of the pretransition (fully duplexed) and post-transition (fully dissociated) states are shown. (B) Pretransition and post-transition baseline normalized representation of the same melting transition data (diamonds), where $\theta$ is the fraction of strands in the single stranded state; the two-state thermodynamic model fit (solid curve) is also shown and superimposes the experimental data.
3.3.3 REGRESSION OF MODEL PARAMETERS

Levenberg-Marquardt (LM) type regression\cite{211} to melting temperature ($T_m$) data was used to fit a set of parameters needed to compute the total hyperstabilization energy $\Delta \Delta G_{\text{hyper}}^0$ of a duplex that arises from i) formation of an LNA:LNA base pair through replacement of a DNA nucleotide that is paired with an LNA within the complementary strand, and ii) substitution of an LNA into a pure-DNA strand at a position proximal to an oppositely oriented LNA:DNA base pair, or any combinations of those two substitution patterns. The parameters are used to extend our existing NNT model\cite{19}, hereafter referred to as the standard Single-Base Thermodynamic (standard SBT) model, that can be applied to duplexes containing LNA substitutions in one strand, but not both.

Regression of the resulting “extended SBT model” to the full data set was conducted using MatLab (R2007b) software. Based on an average experimental error $\sigma_i$ in $\Delta \Delta T_m$ of $\pm$ 0.8 °C, the regression scheme minimized the least squares of the residual ($\chi^2$)

$$
\chi^2 = \sum_i \left( \frac{\Delta \Delta T_m (\text{expt}) - \Delta \Delta T_m (\text{pred})}{\sigma_i} \right)^2
$$

(3-6)

by initiating the objective function ($0 = \Delta \Delta T_m (\text{expt}) - \Delta \Delta T_m (\text{pred})$) with a randomly generated set of parameter estimates. In equation (3-6), \textit{expt} and \textit{pred} indicate experimental and model predicted values, respectively, and $\Delta \Delta T_m$ is given by

$$
\Delta \Delta T_m = \Delta T_{m,LNA} - (\Delta T_{m,LNA:DNA} + \Delta T_{m,DNA:LNA})
$$

(3-7)

where, for example,

$$
\Delta T_{m,LNA} = T_{m,LNA} - T_{m,DNA}
$$

(3-8)
Here, subscript \( LNA \) specifies the duplex bearing an LNA substitution in each strand, and subscript \( DNA \) specifies the corresponding isosequential pure-DNA duplex. Subscript \( LNA:DNA \) specifies the same duplex when no LNA substitution has been made in the antisense-strand, and

\[
\Delta T_{m, LNA:DNA} = T_{m, LNA:DNA} - T_{m, DNA}
\]

In the extended SBT model, \( T_{m, LNA \ (pred)} \) for a duplex bearing an LNA in each strand is given by

\[
T_{m, LNA \ (pred)} = \frac{\Delta H_{LNA}(T_{m, LNA \ (pred)}) + \Delta \Delta G^\circ_{hyper}}{\Delta S_{LNA}(T_{m, LNA \ (pred)}) - R ln \left( \frac{C_T}{4} \right)}
\]

with \( \Delta H_{LNA} \) and \( \Delta S_{LNA} \) computed as

\[
\Delta H_{LNA} (T_{m, LNA}) = \Delta H^\circ_{DNA} + \Delta C_p (T_{m, LNA} - T_{ref}) + \Delta \Delta H^\circ_{LNA}
\]

and

\[
\Delta S_{LNA} (T_{m, LNA}) = \Delta S^\circ_{DNA} + \Delta C_p ln \left( \frac{T_{m, LNA}}{T_{ref}} \right) + \Delta \Delta S^\circ_{LNA}
\]

In both the standard (\( \Delta \Delta G^\circ_{hyper} = 0 \)) and extended SBT models, the melting enthalpy \( \Delta H^\circ_{DNA} \) and entropy \( \Delta S^\circ_{DNA} \) for the isosequential pure-DNA duplex are computed at a specific reference (\( ^\circ \)) temperature \( (T_{ref}) \) of 53 °C, from which their corresponding values at the \( T_m \) of the LNA substituted duplex can be computed using \( \Delta C_p \) as calculated from the equation described above. \( \Delta \Delta H^\circ_{LNA} \) and \( \Delta \Delta S^\circ_{LNA} \) represent the incremental change in the melting enthalpy and entropy, respectively, resulting from any pattern of LNA substitutions within one of the two strands; they are computed at \( T_{ref} \) from available standard SBT model parameters. In equation (3-10), \( R \) is the ideal gas constant, 1.987 cal mol\(^{-1}\) K\(^{-1}\).
The value of $\Delta \Delta G_{\text{hyper}}^o$, which represents the excess energetic contribution to duplex stability created by an LNA:LNA base pair and/or sets of proximal oppositely oriented LNA:DNA base pairs, is computed from the set of extended SBT model parameters $\Delta \Delta G_{\text{hyper}-(j)}^o$ as

$$\Delta \Delta G_{\text{hyper}}^o = \sum_j n_j \Delta \Delta G_{\text{hyper}-(j)}^o$$  \hspace{1cm} (3-13)

In equation (3-13), $n_j$ and $\Delta \Delta G_{\text{hyper}-(j)}^o$ then give the frequency, and energetic contribution to $\Delta \Delta G_{\text{hyper}}^o$, respectively, of each LNA:LNA base pair or set of proximal oppositely-oriented LNA:DNA base pairs of type $j$. Here, $j$ is an integer ranging from -4 to +3, and its value may be understood by considering the position of an LNA within the antisense strand of a duplex relative to a chosen LNA within the sense strand. To fix ideas, we show here by example how equation (3-13) is applied to a hypothetical duplex containing both an LNA:LNA base pair and oppositely oriented LNA:DNA base pairs

| Sense strand | 5' c a t a C g a t g 3' |
| Antisense strand | 3' g t A t G C T a c 5' |

where lower-case letters indicate standard DNA nucleotides, and capital letters the corresponding LNAs. The LNA on the sense strand is selected, and the relative position of each LNA on the anti-sense strand is noted. For the $C$ at the 5’+5 position of the sense strand, they include the $G$ with which it is paired ($j = 0$; $n_f = n_0=1$ as there is only one $C:G$ LNA-LNA base pair in the duplex), the adjacent $C$ ($j = -1, n_{-1} = 1$) on the 5’-side of the $G$ on the anti-sense strand, and the $A$ and $T$ that are in the $j = +2$ and $j = -2$ positions, respectively, of the antisense strand. When required, the same analysis would be performed for any additional LNA on the sense strand. Application of equation (3-13) for the example duplex (for which all $n_j=1$ in this case) shown above then gives
\[
\Delta \Delta G^\circ_{\text{hyper}} = \Delta \Delta G^\circ_{\text{hyper}-(2)} + \Delta \Delta G^\circ_{\text{hyper}-(0)} + \Delta \Delta G^\circ_{\text{hyper}-(1)} + \Delta \Delta G^\circ_{\text{hyper}-(0)}
\] (3-14)

The set of \( \Delta \Delta G^\circ_{\text{hyper}-(j)} \) parameters required to compute \( \Delta \Delta G^\circ_{\text{hyper}} \) were found by global iterative regression of the model to a large set (see below) of \( \Delta \Delta T_m (\text{expt}) \) data using the LM method until a minimum in \( \chi^2 \) was achieved. Each experimental \( T_m \) used to determine both \( \Delta \Delta T_m (\text{expt}) \) and model parameters represents the mean of triplicate independent runs, and is reported along with its standard deviation, which was used to compute errors in each regressed parameter.

### 3.4 RESULTS

#### 3.4.1 STANDARD SBT MODEL AND ITS PERFORMANCE WHEN APPLIED TO DUPLEXES CONTAINING ANY NUMBER AND PATTERN OF LNA SUBSTITUTIONS WITHIN ONE STRAND

The standard SBT model uses an experimental \( \Delta C_p \) value \( (42 \text{ cal mol}^{-1} \text{ K}^{-1} \text{ bp}^{-1}) \) and a set of four incremental entropy parameters \( \Delta \Delta S^\circ_i \) to compute sequence-specific melting thermodynamics and \( T_m \) for any short DNA duplex bearing any number and pattern of LNA substitutions on one of the two strands. Thermodynamic changes \( \Delta H_{\text{LNA}} \) and \( \Delta S_{\text{LNA}} \) for the helix-to-coil transition at \( T_m \) are predicted using equations (3-11) and (3-12) with \( \Delta \Delta H_{\text{LNA}}^\circ \) set equal to zero and \( \Delta \Delta S_{\text{LNA}}^\circ \) given by

\[
\Delta \Delta S_{\text{LNA}}^\circ = \sum n_i \Delta \Delta S_i^\circ
\] (3-15)

where \( n_i \) is the number (frequency) of internal LNA substitutions of type \( i \) \( (i.e. \text{ not including those at either terminus)} \), and the values of \( \Delta \Delta S_i^\circ \) are specific to the reference temperature. The value of \( T_m \) for any duplex in which one strand contains one or more
LNA substitutions is then given by equation (3-5) with $\Delta H = \Delta H_{\text{LNA}}$ and $\Delta S = \Delta S_{\text{LNA}}$. Because of the temperature dependencies of $\Delta H_{\text{LNA}}$ and $\Delta S_{\text{LNA}}$, the solution of equation (3-5) requires iteration, but convergence is rapid if an initial estimate for $T_m$ is found by setting $\Delta C_p = 0$ in equations (3-11) and (3-12).

To date, the standard SBT model has been tested on a relatively small set of complementary duplexes bearing LNA substitutions on one strand. We therefore sought to further validate model performance when applied to that class of duplexes by collecting melting thermodynamics and $T_m$ data for a large set of duplexes containing different patterns of LNA substitutions on one strand, and then comparing model predictions to the resulting data set. Those melting thermodynamics data are reported in Table 3-1, and we note that none of the sequences and associated data was used in the regression of the previously reported standard SBT model parameters.

The corresponding standard SBT model predictions for each of the pure-DNA and LNA-substituted duplexes are also reported. The results confirm that the model accurately predicts both melting thermodynamics and $T_m$, irrespective of the pattern and frequency of LNA substitutions within the one strand. For the entire set of LNA-substituted duplexes, the standard SBT model predicts $\Delta H$, $\Delta S$ and $T_m$ values with a mean error and standard deviation of $3.0 \pm 7.1$ kcal mol$^{-1}$, $9.6 \pm 21.2$ cal mol$^{-1}$ K$^{-1}$, and $-0.9 \pm 1.4$ °C, respectively. Each of these errors is essentially equivalent to the corresponding error of the underlying NNT model describing pure-DNA melting thermodynamics.

As the standard SBT model utilizes a set of $\Delta\Delta S^p_i$ parameters, each of which is specific to the base pair in which the LNA of type $i$ is present, its general accuracy indicates that the incremental contribution to duplex stability of an internal LNA:DNA base pair (relative to the DNA:DNA base pair from which it is derived) is the same irrespective of whether the LNA:DNA base pair is flanked by DNA:DNA base pairs or by one or more LNA:DNA base pairs. At present, this additive effect, which is foundational to group-contribution-type models, has only been demonstrated for the case
where all LNA:DNA base pairs are oriented in the same direction. Below, we therefore investigate cases where LNAs are present on both strands, so that LNA:LNA base pairs and/or oppositely oriented LNA:DNA base pairs are formed.

3.4.2 LIBRARY DESIGN FOR ANALYZING THE EFFECT OF LNA:LNA BASE PAIRS AND OPPOSITELY ORIENTED LNA:DNA BASE PAIRS ON DUPLEX STABILITY

Short complementary duplexes were created to study changes in thermal stability arising from an LNA substitution within each strand. Sequence information and LNA content for each duplex are provided in Table 3-2 along with $T_m$ and complete melting thermodynamic data for the duplex obtained from UV spectroscopy experiments. The general structure of the library is provided in Figure 3-2 in accordance with the $-4 \leq j \leq 3$ nomenclature described above. Each duplex having two LNA substitutions contains either a single LNA:LNA base pair ($j = 0$) or two oppositely oriented LNA:DNA base pairs. The library was designed based on pure-DNA duplexes covering a range of g/c content and base-pair sequences. From these, the set of LNA-substituted duplexes was created so as to provide complete coverage and excellent redundancy in the types of LNA:LNA and oppositely oriented LNA:DNA base pairs formed, as well as in the types of base pairs neighbouring and between each of those LNA-containing base pairs.

Most of the duplex sequences used were selected from literature as they had previously been shown to exhibit two-state melting behavior. However, four duplexes, which likewise exhibit two-state melting behavior based on repeated UV melt experiments, were designed using Exiqon’s (Vedbaek, Denmark) LNA Oligo Optimizer tool software to preclude formation of secondary structures (hairpins, homodimers) or slipped duplex structures. LNA substitutions within either terminal nearest-neighbour
base pair were excluded so as to avoid unwanted edge-effect (e.g. duplex fraying) contributions to regressed $\Delta \Delta G_{\text{hyper}-j}^o$ parameters.

**Figure 3-2** General structure of the library of complementary duplexes created to study changes in thermal stability arising from an LNA substitution within each strand. Standard DNA nucleotides and LNA-substituted nucleotides are indicated by circles and squares, respectively. The position of each LNA on the antisense strand is indexed relative to a selected LNA (arrow) on the sense strand by the value of $j$, where $-4 \leq j \leq 3$ in this study. In this example, the duplex has an LNA:LNA base pair ($j = 0$), and oppositely oriented DNA:LNA base pairs at positions $j = 3, 2, -1, -2, -3, -4, -5$ of the antisense strand.
Table 3-1 Melting thermodynamic data collected by UV spectroscopy for a library of complementary duplexes of either pure DNA or bearing different patterns of LNA substitutions on one strand.

<table>
<thead>
<tr>
<th>Duplex Name</th>
<th>Sense Strand</th>
<th>Anti-Sense Strand</th>
<th>Length (bp), Duplex Type</th>
<th>(\Delta H) (kcal mol(^{-1}))</th>
<th>(\Delta S) (cal mol(^{-1}) K(^{-1}))</th>
<th>(\Delta G^\circ) (kcal mol(^{-1}))</th>
<th>(\Delta\Delta G^\circ) (kcal mol(^{-1}))</th>
<th>(T_m) (°C)</th>
<th>(\Delta T_m) (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1(^t)</td>
<td>ttcatagecgt</td>
<td>acggctatgaa</td>
<td>11</td>
<td>76.9 ± 1.0</td>
<td>209.1 ± 3.1</td>
<td>8.72 ± 0.05</td>
<td>53.8 ± 0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(79.4)</td>
<td>(216.6)</td>
<td>(8.74)</td>
<td>(53.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-S-L3:AS</td>
<td>ttCatagcgt</td>
<td>acggetatgaa</td>
<td>LNA:DNA</td>
<td>80.8 ± 0.9</td>
<td>217.6 ± 2.7</td>
<td>9.83 ± 0.04</td>
<td>1.10 ± 0.06</td>
<td>58.3 ± 0.1</td>
<td>4.6 ± 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(81.7)</td>
<td>(219.8)</td>
<td>(9.94)</td>
<td>(58.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-S-L4:AS</td>
<td>ttcatagcgt</td>
<td>acggetatgaa</td>
<td>LNA:DNA</td>
<td>77.5 ± 1.2</td>
<td>209.2 ± 3.5</td>
<td>9.25 ± 0.03</td>
<td>0.52 ± 0.06</td>
<td>56.0 ± 0.1</td>
<td>2.2 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(80.5)</td>
<td>(218.2)</td>
<td>(9.35)</td>
<td>(56.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-S-L5:AS</td>
<td>ttcaTagcgt</td>
<td>acggetatgaa</td>
<td>LNA:DNA</td>
<td>79.5 ± 2.5</td>
<td>213.8 ± 7.3</td>
<td>9.74 ± 0.06</td>
<td>1.01 ± 0.08</td>
<td>58.0 ± 0.1</td>
<td>4.3 ± 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(81.1)</td>
<td>(219.1)</td>
<td>(9.67)</td>
<td>(57.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-S-L6:AS</td>
<td>ttcatAgcgt</td>
<td>acggetatgaa</td>
<td>LNA:DNA</td>
<td>77.9 ± 0.7</td>
<td>210.0 ± 2.2</td>
<td>9.40 ± 0.05</td>
<td>0.67 ± 0.07</td>
<td>56.6 ± 0.2</td>
<td>2.9 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(80.5)</td>
<td>(218.2)</td>
<td>(9.35)</td>
<td>(56.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-S-L7:AS</td>
<td>ttcataGcgt</td>
<td>acggetatgaa</td>
<td>LNA:DNA</td>
<td>78.1 ± 2.0</td>
<td>210.9 ± 6.1</td>
<td>9.29 ± 0.03</td>
<td>0.57 ± 0.06</td>
<td>56.2 ± 0.1</td>
<td>2.4 ± 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(81.0)</td>
<td>(219.0)</td>
<td>(9.61)</td>
<td>(57.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-S-L8:AS</td>
<td>ttcatacggt</td>
<td>acggetatgaa</td>
<td>LNA:DNA</td>
<td>80.5 ± 1.8</td>
<td>216.4 ± 5.4</td>
<td>9.93 ± 0.04</td>
<td>1.20 ± 0.06</td>
<td>58.8 ± 0.2</td>
<td>5.0 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(81.7)</td>
<td>(219.8)</td>
<td>(9.94)</td>
<td>(58.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-S:AS-L9</td>
<td>ttcatagcgt</td>
<td>acggetatGaa</td>
<td>DNA:LNA</td>
<td>76.5 ± 2.6</td>
<td>207.2 ± 8.0</td>
<td>8.92 ± 0.07</td>
<td>0.20 ± 0.09</td>
<td>54.6 ± 0.3</td>
<td>0.8 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(81.0)</td>
<td>(219.0)</td>
<td>(9.61)</td>
<td>(57.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>Length (bp), Duplex Type</td>
<td>$\Delta H$</td>
<td>$\Delta S$</td>
<td>$\Delta G^\circ$</td>
<td>$\Delta\Delta G^\circ$</td>
<td>$T_m$</td>
<td>$\Delta T_m$</td>
</tr>
<tr>
<td>------------</td>
<td>--------------</td>
<td>------------------</td>
<td>-------------------------</td>
<td>--------------------------------</td>
<td>------------------------</td>
<td>---------------------------</td>
<td>------------------------</td>
<td>------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>1-S:AS-L8</td>
<td>ttcatagccttg</td>
<td>aeggettgaag</td>
<td>DNA:LNA</td>
<td>79.3 ± 2.0</td>
<td>214.2 ± 6.1</td>
<td>9.40 ± 0.05</td>
<td>0.68 ± 0.07</td>
<td>56.6 ± 0.1</td>
<td>2.8 ± 0.3</td>
</tr>
<tr>
<td>1-S:AS-L7</td>
<td>ttcatagccttg</td>
<td>aeggettgaag</td>
<td>DNA:LNA</td>
<td>77.9 ± 1.8</td>
<td>210.9 ± 5.6</td>
<td>9.12 ± 0.03</td>
<td>0.39 ± 0.06</td>
<td>55.4 ± 0.1</td>
<td>1.7 ± 0.3</td>
</tr>
<tr>
<td>1-S:AS-L6</td>
<td>ttcatagccttg</td>
<td>aeggetaggaag</td>
<td>DNA:LNA</td>
<td>77.8 ± 0.9</td>
<td>205.7 ± 3.4</td>
<td>10.68 ± 0.01</td>
<td>1.96 ± 0.05</td>
<td>58.5 ± 0.1</td>
<td>4.8 ± 0.3</td>
</tr>
<tr>
<td>1-S:AS-L5</td>
<td>ttcatagccttg</td>
<td>aeggccaggaag</td>
<td>DNA:LNA</td>
<td>85.2 ± 0.9</td>
<td>228.9 ± 2.6</td>
<td>10.48 ± 0.07</td>
<td>1.75 ± 0.09</td>
<td>60.7 ± 0.3</td>
<td>7.0 ± 0.3</td>
</tr>
<tr>
<td>1-S:AS-L4</td>
<td>ttcatagccttg</td>
<td>aeggcctaggaag</td>
<td>DNA:LNA</td>
<td>78.2 ± 0.6</td>
<td>210.2 ± 1.7</td>
<td>9.65 ± 0.03</td>
<td>0.92 ± 0.06</td>
<td>57.7 ± 0.1</td>
<td>4.0 ± 0.2</td>
</tr>
<tr>
<td>2$^c$</td>
<td>tggctttaagtc</td>
<td>acctattgca</td>
<td>11</td>
<td>82.4 ± 2.3</td>
<td>225.6 ± 6.8</td>
<td>8.76 ± 0.03</td>
<td></td>
<td>53.8 ± 0.1</td>
<td></td>
</tr>
<tr>
<td>2-S-L7:AS</td>
<td>tggcaacagtc</td>
<td>acctattgca</td>
<td>LNA:DNA</td>
<td>76.9 ± 4.4</td>
<td>207.0 ± 13.2</td>
<td>9.38 ± 0.13</td>
<td>0.62 ± 0.13</td>
<td>56.6 ± 0.3</td>
<td>2.8 ± 0.4</td>
</tr>
<tr>
<td>2-S:AS-L3</td>
<td>tggcagctagtc</td>
<td>acTctattgca</td>
<td>DNA:LNA</td>
<td>78.6 ± 4.6</td>
<td>212.5 ± 13.6</td>
<td>9.34 ± 0.14</td>
<td>0.58 ± 0.14</td>
<td>56.3 ± 0.4</td>
<td>2.5 ± 0.4</td>
</tr>
<tr>
<td>2-S-L7:AS</td>
<td>tggcagctagtc</td>
<td>acTctattgca</td>
<td>LNA:DNA</td>
<td>84.3 ± 0.7</td>
<td>228.4 ± 2.1</td>
<td>9.81 ± 0.02</td>
<td>1.05 ± 0.04</td>
<td>58.0 ± 0.0</td>
<td>4.2 ± 0.1</td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>Length (bp), Duplex Type</td>
<td>ΔH</td>
<td>ΔS</td>
<td>ΔG°</td>
<td>ΔΔG°</td>
<td>Tm</td>
<td>ΔTm</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>-----</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>2-S:AS-L5</td>
<td>tgcggataagt</td>
<td>acttAtccgca</td>
<td>DNA:LNA</td>
<td>77.3 ± 0.4</td>
<td>209.2 ± 1.2</td>
<td>9.05 ± 0.02</td>
<td>0.29 ± 0.04</td>
<td>55.2 ± 0.1</td>
<td>1.3 ± 0.1</td>
</tr>
<tr>
<td>3b</td>
<td>ctaacggatgc</td>
<td>gcatecgttag</td>
<td>11</td>
<td>77.8 ± 1.7</td>
<td>211.8 ± 5.0</td>
<td>8.75 ± 0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-S-L4:AS</td>
<td>ctaAcggtacg</td>
<td>gcatecgttag</td>
<td>LNA:DNA</td>
<td>77.5 ± 2.0</td>
<td>209.0 ± 5.8</td>
<td>9.35 ± 0.10</td>
<td>0.60 ± 0.10</td>
<td>56.4 ± 0.3</td>
<td>2.6 ± 0.4</td>
</tr>
<tr>
<td>3-S-L5:AS</td>
<td>ctaaCggatgc</td>
<td>gcatecgttag</td>
<td>LNA:DNA</td>
<td>80.1 ± 3.0</td>
<td>215.9 ± 8.9</td>
<td>9.70 ± 0.09</td>
<td>0.95 ± 0.10</td>
<td>57.8 ± 0.2</td>
<td>4.0 ± 0.2</td>
</tr>
<tr>
<td>3-S-L6:AS</td>
<td>ctaacGgatgc</td>
<td>gcatecgttag</td>
<td>LNA:DNA</td>
<td>79.3 ± 0.6</td>
<td>213.7 ± 1.8</td>
<td>9.62 ± 0.04</td>
<td>0.87 ± 0.06</td>
<td>57.5 ± 0.1</td>
<td>3.7 ± 0.2</td>
</tr>
<tr>
<td>3-S-L7:AS</td>
<td>ctaacgGatgc</td>
<td>gcatecgttag</td>
<td>LNA:DNA</td>
<td>80.6 ± 1.0</td>
<td>216.6 ± 2.8</td>
<td>9.97 ± 0.04</td>
<td>1.22 ± 0.06</td>
<td>59.0 ± 0.1</td>
<td>5.1 ± 0.2</td>
</tr>
<tr>
<td>3-S-L8:AS</td>
<td>ctaaeggAtgc</td>
<td>gcatecgttag</td>
<td>LNA:DNA</td>
<td>77.4 ± 0.2</td>
<td>208.2 ± 0.8</td>
<td>9.45 ± 0.04</td>
<td>0.70 ± 0.05</td>
<td>56.9 ± 0.2</td>
<td>3.1 ± 0.2</td>
</tr>
<tr>
<td>3-S:AS-L9</td>
<td>ctaacggatgc</td>
<td>gcatecgtTag</td>
<td>DNA:LNA</td>
<td>85.3 ± 1.1</td>
<td>232.2 ± 3.4</td>
<td>9.58 ± 0.02</td>
<td>0.83 ± 0.04</td>
<td>57.0 ± 0.0</td>
<td>3.2 ± 0.2</td>
</tr>
</tbody>
</table>

90
<table>
<thead>
<tr>
<th>Duplex Name</th>
<th>Sense Strand</th>
<th>Anti-Sense Strand</th>
<th>Length (bp), Duplex Type</th>
<th>ΔH (kcal mol⁻¹)</th>
<th>ΔS (cal mol⁻¹ K⁻¹)</th>
<th>ΔG° (kcal mol⁻¹)</th>
<th>ΔΔG° (kcal mol⁻¹)</th>
<th>Tm (°C)</th>
<th>ΔTm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-S:AS-L8</td>
<td>catacggatgcg</td>
<td>gcaGctcgTagg</td>
<td>DNA:LNA</td>
<td>77.2 ± 1.3</td>
<td>207.4 ± 3.9</td>
<td>9.52 ± 0.02</td>
<td>0.77 ± 0.05</td>
<td>57.2 ± 0.1</td>
<td>3.4 ± 0.2</td>
</tr>
<tr>
<td>3-S:AS-L7</td>
<td>catacggatgcg</td>
<td>gcaGctcgTagg</td>
<td>DNA:LNA</td>
<td>79.8 ± 0.8</td>
<td>215.1 ± 2.4</td>
<td>9.66 ± 0.04</td>
<td>0.91 ± 0.06</td>
<td>57.7 ± 0.1</td>
<td>3.8 ± 0.2</td>
</tr>
<tr>
<td>3-S:AS-L6</td>
<td>catacggatgcg</td>
<td>gcaGctcgTagg</td>
<td>DNA:LNA</td>
<td>79.8 ± 1.3</td>
<td>213.9 ± 3.9</td>
<td>9.99 ± 0.07</td>
<td>1.24 ± 0.08</td>
<td>59.1 ± 0.2</td>
<td>5.3 ± 0.3</td>
</tr>
<tr>
<td>3-S:AS-L5</td>
<td>catacggatgcg</td>
<td>gcaGctcgTagg</td>
<td>DNA:LNA</td>
<td>79.8 ± 0.7</td>
<td>214.2 ± 2.1</td>
<td>9.95 ± 0.07</td>
<td>1.21 ± 0.08</td>
<td>59.0 ± 0.3</td>
<td>5.1 ± 0.3</td>
</tr>
<tr>
<td>3-S:AS-L4</td>
<td>catacggatgcg</td>
<td>gcaGctcgTagg</td>
<td>DNA:LNA</td>
<td>77.3 ± 1.5</td>
<td>207.5 ± 4.4</td>
<td>9.57 ± 0.04</td>
<td>0.82 ± 0.05</td>
<td>57.4 ± 0.1</td>
<td>3.6 ± 0.2</td>
</tr>
<tr>
<td>4b</td>
<td>catacgatcc</td>
<td>ggaatgttag</td>
<td>11</td>
<td>78.5 ± 2.6</td>
<td>214.0 ± 7.7</td>
<td>8.65 ± 0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-S-L9:AS</td>
<td>catacgatTcc</td>
<td>ggaatgttag</td>
<td>LNA:DNA</td>
<td>79.6 ± 1.5</td>
<td>214.4 ± 4.2</td>
<td>9.69 ± 0.06</td>
<td>1.04 ± 0.09</td>
<td>57.8 ± 0.2</td>
<td>4.4 ± 0.3</td>
</tr>
<tr>
<td>4-S-L8:AS</td>
<td>catacgatTcc</td>
<td>ggaatgttag</td>
<td>LNA:DNA</td>
<td>76.3 ± 0.2</td>
<td>205.0 ± 0.5</td>
<td>9.40 ± 0.01</td>
<td>0.75 ± 0.07</td>
<td>56.7 ± 0.1</td>
<td>3.3 ± 0.3</td>
</tr>
<tr>
<td>4-S-L7:AS</td>
<td>catacgAttcc</td>
<td>ggaatgttag</td>
<td>LNA:DNA</td>
<td>72.8 ± 1.7</td>
<td>195.6 ± 5.0</td>
<td>9.02 ± 0.06</td>
<td>0.37 ± 0.08</td>
<td>55.2 ± 0.2</td>
<td>1.7 ± 0.3</td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>Length (bp), Duplex Type</td>
<td>ΔH</td>
<td>ΔS</td>
<td>ΔG°</td>
<td>ΔΔG°</td>
<td>Tm</td>
<td>ΔTm</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>-------------------</td>
<td>--------------------------</td>
<td>-----</td>
<td>------</td>
<td>-----</td>
<td>-------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>(5’-3’)</td>
<td>(5’-3’)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-S-L6:AS</td>
<td>ctagcGatc</td>
<td>ggaatgcgtag</td>
<td>LNA:DNA</td>
<td>82.0 ± 2.4</td>
<td>220.0 ± 7.2</td>
<td>10.23 ± 0.09</td>
<td>1.57 ± 0.11</td>
<td>60.0 ± 0.2</td>
<td>6.5 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(84.7)</td>
<td>(231.2)</td>
<td>(9.25)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-S-L4:AS</td>
<td>ctaGCatc</td>
<td>ggaatgcgtag</td>
<td>LNA:DNA</td>
<td>79.4 ± 3.1</td>
<td>213.4 ± 9.3</td>
<td>9.75 ± 0.10</td>
<td>1.10 ± 0.12</td>
<td>58.1 ± 0.2</td>
<td>4.7 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(85.7)</td>
<td>(232.7)</td>
<td>(9.84)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-S:AS-L3</td>
<td>ctagcattc</td>
<td>ggAatgcgtag</td>
<td>DNA:LNA</td>
<td>79.8 ± 3.1</td>
<td>216.1 ± 9.2</td>
<td>9.35 ± 0.08</td>
<td>0.69 ± 0.10</td>
<td>56.3 ± 0.2</td>
<td>2.9 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(84.7)</td>
<td>(231.2)</td>
<td>(9.25)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-S:AS-L5</td>
<td>ctagcattc</td>
<td>ggaaTgcgtag</td>
<td>DNA:LNA</td>
<td>82.1 ± 0.4</td>
<td>222.0 ± 1.2</td>
<td>9.66 ± 0.02</td>
<td>1.00 ± 0.07</td>
<td>57.5 ± 0.1</td>
<td>4.1 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(85.2)</td>
<td>(232.0)</td>
<td>(9.57)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-S:AS-L6</td>
<td>ctagcattc</td>
<td>ggaatGcgtag</td>
<td>DNA:LNA</td>
<td>78.2 ± 2.1</td>
<td>211.6 ± 6.4</td>
<td>9.14 ± 0.06</td>
<td>0.48 ± 0.09</td>
<td>55.5 ± 0.2</td>
<td>2.1 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(85.1)</td>
<td>(231.9)</td>
<td>(9.51)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-S:AS-L8</td>
<td>ctagcattc</td>
<td>ggaatGcGtag</td>
<td>DNA:LNA</td>
<td>81.3 ± 3.4</td>
<td>220.9 ± 10.5</td>
<td>9.30 ± 0.10</td>
<td>0.65 ± 0.12</td>
<td>56.1 ± 0.4</td>
<td>2.6 ± 0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(85.1)</td>
<td>(231.9)</td>
<td>(9.51)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5’</td>
<td>ctttgGgac</td>
<td>gtcgcaaatg</td>
<td>11</td>
<td>82.4 ± 1.5</td>
<td>224.3 ± 4.4</td>
<td>9.24 ± 0.05</td>
<td></td>
<td>55.8 ± 0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(83.6)</td>
<td>(229.7)</td>
<td>(8.64)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-S-L8:AS</td>
<td>ctttgGCac</td>
<td>gtcgcaaatg</td>
<td>LNA:DNA</td>
<td>79.1 ± 1.9</td>
<td>211.0 ± 5.9</td>
<td>10.21 ± 0.06</td>
<td>0.97 ± 0.08</td>
<td>60.2 ± 0.3</td>
<td>4.4 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(85.7)</td>
<td>(232.7)</td>
<td>(9.84)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

92
<table>
<thead>
<tr>
<th>Duplex Name</th>
<th>Sense Strand</th>
<th>Anti-Sense Strand</th>
<th>Length (bp), Duplex Type</th>
<th>$\Delta H$ (kcal mol$^{-1}$)</th>
<th>$\Delta S$ (cal mol$^{-1}$ K$^{-1}$)</th>
<th>$\Delta G^\circ$ (kcal mol$^{-1}$)</th>
<th>$\Delta \Delta G^\circ$ (kcal mol$^{-1}$)</th>
<th>$T_m$ (°C)</th>
<th>$\Delta T_m$ (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-S:AS-L6</td>
<td>ctattggcgc</td>
<td>gtgcgcCaatag</td>
<td>DNA:LNA</td>
<td>81.9 ± 1.6</td>
<td>219.9 ± 4.6</td>
<td>10.11 ± 0.06</td>
<td>0.87 ± 0.07</td>
<td>59.5 ± 0.1</td>
<td>3.7 ± 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(85.7)</td>
<td>(232.7)</td>
<td>(9.84)</td>
<td></td>
<td>(58.1)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>ctaaatgcgc</td>
<td>gcggacttcag</td>
<td>11</td>
<td>82.7 ± 2.8</td>
<td>223.3 ± 8.5</td>
<td>9.80 ± 0.10</td>
<td></td>
<td>58.1 ± 0.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(87.1)</td>
<td>(237.2)</td>
<td>(9.75)</td>
<td></td>
<td>(57.6)</td>
<td></td>
</tr>
<tr>
<td>6-S-L3:AS</td>
<td>ctaaatgcgc</td>
<td>gcggacttcag</td>
<td>LNA:DNA</td>
<td>90.2 ± 1.3</td>
<td>245.2 ± 3.7</td>
<td>10.19 ± 0.05</td>
<td>0.39 ± 0.11</td>
<td>59.2 ± 0.1</td>
<td>1.1 ± 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(88.7)</td>
<td>(239.2)</td>
<td>(10.62)</td>
<td></td>
<td>(61.0)</td>
<td></td>
</tr>
<tr>
<td>6-S-L4:AS</td>
<td>ctaaatgcgc</td>
<td>gcggacttcag</td>
<td>LNA:DNA</td>
<td>96.2 ± 1.1</td>
<td>260.2 ± 3.2</td>
<td>11.26 ± 0.04</td>
<td>1.46 ± 0.11</td>
<td>62.7 ± 0.1</td>
<td>4.6 ± 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(88.2)</td>
<td>(238.6)</td>
<td>(10.36)</td>
<td></td>
<td>(60.0)</td>
<td></td>
</tr>
<tr>
<td>6-S-L5:AS</td>
<td>ctaaatgcgc</td>
<td>gcggacttcag</td>
<td>LNA:DNA</td>
<td>91.3 ± 1.3</td>
<td>244.8 ± 3.7</td>
<td>11.38 ± 0.07</td>
<td>1.58 ± 0.13</td>
<td>63.8 ± 0.2</td>
<td>5.7 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(88.2)</td>
<td>(238.6)</td>
<td>(10.36)</td>
<td></td>
<td>(60.0)</td>
<td></td>
</tr>
<tr>
<td>6-S-L6:AS</td>
<td>ctaaatgcgc</td>
<td>gcggacttcag</td>
<td>LNA:DNA</td>
<td>91.0 ± 1.3</td>
<td>244.9 ± 3.7</td>
<td>11.05 ± 0.06</td>
<td>1.25 ± 0.12</td>
<td>62.5 ± 0.1</td>
<td>4.4 ± 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(88.7)</td>
<td>(239.2)</td>
<td>(10.62)</td>
<td></td>
<td>(61.0)</td>
<td></td>
</tr>
<tr>
<td>6-S:AS-L3</td>
<td>ctaaatgcgc</td>
<td>gcggacttcag</td>
<td>DNA:LNA</td>
<td>90.5 ± 1.3</td>
<td>245.1 ± 3.7</td>
<td>10.55 ± 0.04</td>
<td>0.76 ± 0.11</td>
<td>60.6 ± 0.1</td>
<td>2.5 ± 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(88.7)</td>
<td>(239.2)</td>
<td>(10.62)</td>
<td></td>
<td>(61.0)</td>
<td></td>
</tr>
<tr>
<td>6-S:AS-L4</td>
<td>ctaaatgcgc</td>
<td>gcggacttcag</td>
<td>DNA:LNA</td>
<td>91.1 ± 1.3</td>
<td>244.9 ± 3.7</td>
<td>11.13 ± 0.07</td>
<td>1.33 ± 0.13</td>
<td>62.8 ± 0.1</td>
<td>4.7 ± 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(88.7)</td>
<td>(239.2)</td>
<td>(10.62)</td>
<td></td>
<td>(61.0)</td>
<td></td>
</tr>
<tr>
<td>6-S:AS-L5</td>
<td>ctaaatgcgc</td>
<td>gcggacttcag</td>
<td>DNA:LNA</td>
<td>91.1 ± 1.3</td>
<td>244.9 ± 3.7</td>
<td>11.14 ± 0.06</td>
<td>1.34 ± 0.12</td>
<td>62.8 ± 0.1</td>
<td>4.7 ± 0.2</td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>Length (bp), Duplex Type</td>
<td>$\Delta H$</td>
<td>$\Delta S$</td>
<td>$\Delta G°$</td>
<td>$\Delta \Delta G°$</td>
<td>$T_m$</td>
<td>$\Delta T_m$</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>-------------------</td>
<td>--------------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>----------------</td>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td></td>
<td>(5' - 3')</td>
<td>(5' - 3')</td>
<td></td>
<td>(kcal mol$^{-1}$)</td>
<td>(cal mol$^{-1}$ K$^{-1}$)</td>
<td>(kcal mol$^{-1}$)</td>
<td>(kcal mol$^{-1}$)</td>
<td>(°C)</td>
<td>(°C)</td>
</tr>
<tr>
<td>6-S:AS-L6</td>
<td>ctgaagtcgc</td>
<td>gcggacGttcag</td>
<td>DNA:LNA</td>
<td>91.5 ± 1.3</td>
<td>244.7 ± 3.7</td>
<td>11.63 ± 0.07</td>
<td>1.83 ± 0.12</td>
<td>64.7 ± 0.2</td>
<td>6.6 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(89.3)</td>
<td>(240.0)</td>
<td>(10.95)</td>
<td></td>
<td>(62.3)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>gtatcaagtct</td>
<td>agacttgtgac</td>
<td>11</td>
<td>77.5 ± 2.5</td>
<td>215.6 ± 7.8</td>
<td>7.15 ± 0.03</td>
<td></td>
<td>47.3 ± 0.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(73.4)</td>
<td>(204.7)</td>
<td>(6.61)</td>
<td></td>
<td>(44.8)</td>
<td></td>
</tr>
<tr>
<td>7-S:AS-L5</td>
<td>gtatcaagtct</td>
<td>agacTgtagac</td>
<td>DNA:LNA</td>
<td>80.9 ± 1.3</td>
<td>223.7 ± 4.0</td>
<td>7.90 ± 0.02</td>
<td>0.76 ± 0.04</td>
<td>50.5 ± 0.1</td>
<td>3.2 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(75.2)</td>
<td>(207.5)</td>
<td>(7.54)</td>
<td></td>
<td>(48.8)</td>
<td></td>
</tr>
<tr>
<td>7-S:AS-L6</td>
<td>gtatcaagtct</td>
<td>agacttgac</td>
<td>DNA:LNA</td>
<td>83.4 ± 0.5</td>
<td>230.7 ± 1.7</td>
<td>8.12 ± 0.00</td>
<td>0.97 ± 0.03</td>
<td>51.3 ± 0.0</td>
<td>4.0 ± 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(75.2)</td>
<td>(207.5)</td>
<td>(7.54)</td>
<td></td>
<td>(48.8)</td>
<td></td>
</tr>
<tr>
<td>7-S:AS-L4</td>
<td>gtatcaagtct</td>
<td>agacTgtgac</td>
<td>DNA:LNA</td>
<td>77.6 ± 1.1</td>
<td>211.4 ± 3.2</td>
<td>8.67 ± 0.02</td>
<td>1.53 ± 0.03</td>
<td>53.5 ± 0.1</td>
<td>6.2 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(75.8)</td>
<td>(208.3)</td>
<td>(7.81)</td>
<td></td>
<td>(49.9)</td>
<td></td>
</tr>
<tr>
<td>7-S-L3:AS</td>
<td>gtaAcaagtct</td>
<td>agacttgatac</td>
<td>LNA:DNA</td>
<td>83.6 ± 0.6</td>
<td>232.5 ± 1.7</td>
<td>7.73 ± 0.01</td>
<td>0.58 ± 0.03</td>
<td>49.9 ± 0.1</td>
<td>2.6 ± 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(74.6)</td>
<td>(206.6)</td>
<td>(7.22)</td>
<td></td>
<td>(47.4)</td>
<td></td>
</tr>
<tr>
<td>7-S-L4:AS</td>
<td>gtaTcaagtct</td>
<td>agacttgatac</td>
<td>LNA:DNA</td>
<td>80.9 ± 1.3</td>
<td>223.7 ± 4.0</td>
<td>7.90 ± 0.02</td>
<td>0.76 ± 0.04</td>
<td>50.5 ± 0.1</td>
<td>3.2 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(75.2)</td>
<td>(207.5)</td>
<td>(7.54)</td>
<td></td>
<td>(48.8)</td>
<td></td>
</tr>
<tr>
<td>7-S-L5:AS</td>
<td>gtaCcaagtct</td>
<td>agacttgatac</td>
<td>LNA:DNA</td>
<td>81.5 ± 4.1</td>
<td>224.7 ± 12.3</td>
<td>8.27 ± 0.07</td>
<td>1.12 ± 0.08</td>
<td>51.9 ± 0.3</td>
<td>4.6 ± 0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(75.8)</td>
<td>(208.3)</td>
<td>(7.81)</td>
<td></td>
<td>(49.9)</td>
<td></td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>Length (bp), Duplex Type</td>
<td>ΔH (kcal mol⁻¹)</td>
<td>ΔS (cal mol⁻¹ K⁻¹)</td>
<td>ΔG° (kcal mol⁻¹)</td>
<td>ΔΔG° (kcal mol⁻¹)</td>
<td>Tm (°C)</td>
<td>ΔTm (°C)</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>----------------</td>
<td>------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>7-S-L6:AS</td>
<td>gtatcAagtct</td>
<td>agactgatac</td>
<td>LNA:DNA</td>
<td>82.6 ± 2.1</td>
<td>229.8 ± 6.5</td>
<td>7.69 ± 0.00</td>
<td>0.54 ± 0.03</td>
<td>49.7 ± 0.1</td>
<td>2.4 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(74.6)</td>
<td>(206.6)</td>
<td>(7.22)</td>
<td></td>
<td>(47.4)</td>
<td></td>
</tr>
<tr>
<td>8d</td>
<td>ggacctcgaec</td>
<td>gtcgaggtcc</td>
<td>10</td>
<td>78.1 ± 3.0</td>
<td>212.2 ± 9.2</td>
<td>8.85 ± 0.17</td>
<td></td>
<td>53.2 ± 0.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(75.5)</td>
<td>(204.7)</td>
<td>(8.76)</td>
<td></td>
<td>(52.8)</td>
<td></td>
</tr>
<tr>
<td>8-S-L4:AS'</td>
<td>ggaCtcgcac</td>
<td>gtcaaggtcc</td>
<td>LNA:DNA</td>
<td>81.2 ± 4.0</td>
<td>217.0 ± 11.9</td>
<td>10.44 ± 0.11</td>
<td>1.59 ± 0.21</td>
<td>59.8 ± 0.2</td>
<td>6.6 ± 0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(77.7)</td>
<td>(207.6)</td>
<td>(9.96)</td>
<td></td>
<td>(58.0)</td>
<td></td>
</tr>
<tr>
<td>8-S-L5:AS'</td>
<td>ggacctcgcac</td>
<td>gtcagaggtcc</td>
<td>LNA:DNA</td>
<td>79.9 ± 2.2</td>
<td>214.4 ± 6.7</td>
<td>9.93 ± 0.12</td>
<td>1.09 ± 0.21</td>
<td>57.7 ± 0.5</td>
<td>4.5 ± 0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(77.7)</td>
<td>(207.6)</td>
<td>(9.96)</td>
<td></td>
<td>(58.0)</td>
<td></td>
</tr>
<tr>
<td>8-S-L7:AS'</td>
<td>ggacctcgcac</td>
<td>gtcagaggtcc</td>
<td>LNA:DNA</td>
<td>88.7 ± 1.8</td>
<td>240.9 ± 5.2</td>
<td>10.08 ± 0.09</td>
<td>1.23 ± 0.20</td>
<td>57.8 ± 0.3</td>
<td>4.6 ± 0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(77.7)</td>
<td>(207.6)</td>
<td>(9.96)</td>
<td></td>
<td>(58.0)</td>
<td></td>
</tr>
<tr>
<td>8-S:AS-L5</td>
<td>ggacctcgcac</td>
<td>gtcagaggtcc</td>
<td>DNA:LNA</td>
<td>82.2 ± 2.1</td>
<td>220.0 ± 6.3</td>
<td>10.41 ± 0.09</td>
<td>1.57 ± 0.20</td>
<td>59.6 ± 0.3</td>
<td>6.4 ± 0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(76.6)</td>
<td>(206.2)</td>
<td>(9.37)</td>
<td></td>
<td>(55.4)</td>
<td></td>
</tr>
<tr>
<td>8-S:AS-L6</td>
<td>ggacctcgcac</td>
<td>gtcagaggtcc</td>
<td>DNA:LNA</td>
<td>79.7 ± 1.7</td>
<td>212.1 ± 5.3</td>
<td>10.50 ± 0.06</td>
<td>1.65 ± 0.18</td>
<td>60.2 ± 0.3</td>
<td>7.0 ± 0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(77.1)</td>
<td>(206.8)</td>
<td>(9.63)</td>
<td></td>
<td>(56.6)</td>
<td></td>
</tr>
<tr>
<td>8-S:AS-L7</td>
<td>ggacctcgcac</td>
<td>gtcagaggtcc</td>
<td>DNA:LNA</td>
<td>77.9 ± 2.2</td>
<td>208.4 ± 6.5</td>
<td>9.95 ± 0.07</td>
<td>1.10 ± 0.19</td>
<td>57.9 ± 0.2</td>
<td>4.7 ± 0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(77.1)</td>
<td>(206.8)</td>
<td>(9.63)</td>
<td></td>
<td>(56.6)</td>
<td></td>
</tr>
<tr>
<td>8-S:AS-L67</td>
<td>ggacctcgcac</td>
<td>gtcagaggtcc</td>
<td>Tandem</td>
<td>83.7 ± 0.9</td>
<td>221.7 ± 2.4</td>
<td>11.37 ± 0.12</td>
<td>2.53 ± 0.21</td>
<td>63.6 ± 0.5</td>
<td>10.4 ± 0.9</td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>Length (bp), Duplex Type</td>
<td>$\Delta H$ (kcal mol$^{-1}$)</td>
<td>$\Delta S$ (cal mol$^{-1}$ K$^{-1}$)</td>
<td>$\Delta G^\circ$ (kcal mol$^{-1}$)</td>
<td>$\Delta \Delta G^\circ$ (kcal mol$^{-1}$)</td>
<td>$T_m$ (°C)</td>
<td>$\Delta T_m$ (°C)</td>
</tr>
<tr>
<td>------------</td>
<td>--------------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>----------------------------</td>
<td>---------------------------------</td>
<td>-------------------------------------</td>
<td>-------------------------------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>8-S-L45:AS</td>
<td>ggaCCtgcac</td>
<td>gtcgagttcc</td>
<td>Tandem</td>
<td>80.5 ± 3.8</td>
<td>213.3 ± 10.9</td>
<td>10.92 ± 0.22</td>
<td>2.07 ± 0.28</td>
<td>62.0 ± 0.5</td>
<td>8.8 ± 0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(79.9)</td>
<td>(210.5)</td>
<td>(11.15)</td>
<td>(63.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8-S-L57:AS</td>
<td>ggacCtGac</td>
<td>gtcgagttcc</td>
<td>Gapmer</td>
<td>83.0 ± 1.5</td>
<td>219.4 ± 4.4</td>
<td>11.36 ± 0.09</td>
<td>2.51 ± 0.20</td>
<td>63.6 ± 0.2</td>
<td>10.5 ± 0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(79.9)</td>
<td>(210.5)</td>
<td>(11.15)</td>
<td>(63.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>ctcgcgtgac</td>
<td>gtcategcagg</td>
<td>11</td>
<td>86.2 ± 3.8</td>
<td>233.9 ± 11.4</td>
<td>9.89 ± 0.09</td>
<td></td>
<td>57.4 ± 0.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(86.7)</td>
<td>(235.9)</td>
<td>(9.78)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-S-L3:AS</td>
<td>ccTgcgtgac</td>
<td>gtcategcagg</td>
<td>LNA:DNA</td>
<td>92.4 ± 0.8</td>
<td>249.9 ± 2.3</td>
<td>10.88 ± 0.03</td>
<td>0.99 ± 0.10</td>
<td>60.6 ± 0.1</td>
<td>3.2 ± 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(88.4)</td>
<td>(238.0)</td>
<td>(10.71)</td>
<td>(60.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-S-L4:AS</td>
<td>ctcGcgtgac</td>
<td>gtcategcagg</td>
<td>LNA:DNA</td>
<td>86.8 ± 3.8</td>
<td>234.7 ± 11.3</td>
<td>10.26 ± 0.08</td>
<td>0.37 ± 0.12</td>
<td>58.6 ± 0.1</td>
<td>1.2 ± 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(88.3)</td>
<td>(237.9)</td>
<td>(10.66)</td>
<td>(60.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-S-L5:AS</td>
<td>cctgCgtgac</td>
<td>gtcategcagg</td>
<td>LNA:DNA</td>
<td>87.6 ± 3.8</td>
<td>234.4 ± 11.3</td>
<td>11.13 ± 0.09</td>
<td>1.24 ± 0.13</td>
<td>62.1 ± 0.1</td>
<td>4.7 ± 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(88.9)</td>
<td>(238.6)</td>
<td>(10.98)</td>
<td>(61.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-S-L6:AS</td>
<td>ctcgeGtgcac</td>
<td>gtcategcagg</td>
<td>LNA:DNA</td>
<td>87.1 ± 3.8</td>
<td>234.6 ± 11.3</td>
<td>10.59 ± 0.09</td>
<td>0.70 ± 0.13</td>
<td>59.9 ± 0.0</td>
<td>2.5 ± 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(88.3)</td>
<td>(237.9)</td>
<td>(10.66)</td>
<td>(60.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-S-L7:AS</td>
<td>ctcgeGtgcac</td>
<td>gtcategcagg</td>
<td>LNA:DNA</td>
<td>87.2 ± 3.8</td>
<td>234.6 ± 11.3</td>
<td>10.65 ± 0.15</td>
<td>0.76 ± 0.18</td>
<td>60.2 ± 0.3</td>
<td>2.8 ± 0.4</td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>Length (bp), Duplex Type</td>
<td>$\Delta H$</td>
<td>$\Delta S$</td>
<td>$\Delta G^\circ$</td>
<td>$\Delta \Delta G^\circ$</td>
<td>$T_m$</td>
<td>$\Delta T_m$</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-------------</td>
<td>----------------------</td>
<td>------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td>(5'-3')</td>
<td>(5'-3')</td>
<td></td>
<td>(kcal mol$^{-1}$)</td>
<td>(cal mol$^{-1}$ K$^{-1}$)</td>
<td>(kcal mol$^{-1}$)</td>
<td>(kcal mol$^{-1}$)</td>
<td>(°C)</td>
<td>(°C)</td>
</tr>
<tr>
<td>9-S-L8:AS</td>
<td>ctgctgaTgac</td>
<td>gtctgtcagg</td>
<td>LNA:DNA</td>
<td>87.5 ± 3.8</td>
<td>234.5 ± 11.3</td>
<td>11.00 ± 0.15</td>
<td>1.11 ± 0.17</td>
<td>61.6 ± 0.2</td>
<td>4.2 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(88.4)</td>
<td>(238.0)</td>
<td>(10.71)</td>
<td></td>
<td>(60.3)</td>
<td></td>
</tr>
<tr>
<td>9-S:AS-L3</td>
<td>ctgctgaTgac</td>
<td>gtcAtgcagg</td>
<td>DNA:LNA</td>
<td>87.7 ± 3.8</td>
<td>234.4 ± 11.3</td>
<td>11.14 ± 0.11</td>
<td>1.25 ± 0.15</td>
<td>62.1 ± 0.1</td>
<td>4.7 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(88.9)</td>
<td>(238.6)</td>
<td>(10.98)</td>
<td></td>
<td>(61.4)</td>
<td></td>
</tr>
<tr>
<td>9-S:AS-L4</td>
<td>ctcggtgac</td>
<td>gtcAtgcagg</td>
<td>DNA:LNA</td>
<td>90.7 ± 2.8</td>
<td>246.2 ± 8.4</td>
<td>10.38 ± 0.10</td>
<td>0.48 ± 0.14</td>
<td>58.8 ± 0.2</td>
<td>1.4 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(87.8)</td>
<td>(237.3)</td>
<td>(10.39)</td>
<td></td>
<td>(59.1)</td>
<td></td>
</tr>
<tr>
<td>9-S:AS-L5</td>
<td>ctcggtgac</td>
<td>gtcaTgcagg</td>
<td>DNA:LNA</td>
<td>87.4 ± 3.8</td>
<td>234.5 ± 11.3</td>
<td>10.89 ± 0.06</td>
<td>1.00 ± 0.11</td>
<td>61.1 ± 0.1</td>
<td>3.7 ± 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(88.4)</td>
<td>(238.0)</td>
<td>(10.71)</td>
<td></td>
<td>(60.3)</td>
<td></td>
</tr>
<tr>
<td>9-S:AS-L6</td>
<td>ctcggtgac</td>
<td>gtcAtgcagg</td>
<td>DNA:LNA</td>
<td>94.6 ± 1.8</td>
<td>255.3 ± 5.3</td>
<td>11.26 ± 0.06</td>
<td>1.37 ± 0.11</td>
<td>61.9 ± 0.0</td>
<td>4.5 ± 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(88.9)</td>
<td>(238.6)</td>
<td>(10.98)</td>
<td></td>
<td>(61.4)</td>
<td></td>
</tr>
<tr>
<td>9-S:AS-L7</td>
<td>ctcggtgac</td>
<td>gtcAtgcagg</td>
<td>DNA:LNA</td>
<td>86.9 ± 3.8</td>
<td>234.7 ± 11.3</td>
<td>10.29 ± 0.08</td>
<td>0.40 ± 0.12</td>
<td>58.7 ± 0.1</td>
<td>1.3 ± 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(88.3)</td>
<td>(237.9)</td>
<td>(10.66)</td>
<td></td>
<td>(60.1)</td>
<td></td>
</tr>
<tr>
<td>10-S-L4:AS</td>
<td>ggaAcaagatgc</td>
<td>geatgttgccc</td>
<td>DNA:DNA</td>
<td>88.4 ± 3.4</td>
<td>242.7 ± 10.4</td>
<td>9.23 ± 0.12</td>
<td></td>
<td>54.6 ± 0.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(91.2)</td>
<td>(251.0)</td>
<td>(9.38)</td>
<td></td>
<td>(55.1)</td>
<td></td>
</tr>
</tbody>
</table>

97
<table>
<thead>
<tr>
<th>Duplex Name</th>
<th>Sense Strand</th>
<th>Anti-Sense Strand</th>
<th>Length (bp)</th>
<th>ΔH (kcal mol⁻¹)</th>
<th>ΔS (cal mol⁻¹ K⁻¹)</th>
<th>ΔG° (kcal mol⁻¹)</th>
<th>ΔΔG° (kcal mol⁻¹)</th>
<th>Tm (°C)</th>
<th>ΔTm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-S-L6:AS³</td>
<td>ggaacAagatgc</td>
<td>gcatcgttcc</td>
<td>LNA:DNA</td>
<td>89.0 ± 1.9</td>
<td>243.3 ± 5.8</td>
<td>9.62 ± 0.06</td>
<td>0.39 ± 0.13</td>
<td>56.0 ± 0.2</td>
<td>1.4 ± 0.5</td>
</tr>
<tr>
<td>10-S-L7:AS³</td>
<td>ggaacaAgatgc</td>
<td>gcatcgttcc</td>
<td>LNA:DNA</td>
<td>94.4 ± 3.5</td>
<td>258.1 ± 10.5</td>
<td>10.24 ± 0.09</td>
<td>1.01 ± 0.15</td>
<td>58.1 ± 0.2</td>
<td>3.5 ± 0.5</td>
</tr>
<tr>
<td>10-S-L9:AS³</td>
<td>ggaacaagAtgc</td>
<td>gcatcgttcc</td>
<td>LNA:DNA</td>
<td>91.8 ± 2.0</td>
<td>251.0 ± 6.1</td>
<td>9.96 ± 0.08</td>
<td>0.73 ± 0.14</td>
<td>57.2 ± 0.2</td>
<td>2.6 ± 0.5</td>
</tr>
<tr>
<td>10-S:AS-L5</td>
<td>ggaacaagatgc</td>
<td>gcatCtgttcc</td>
<td>DNA:LNA</td>
<td>93.6 ± 2.5</td>
<td>254.3 ± 7.3</td>
<td>10.62 ± 0.10</td>
<td>1.39 ± 0.15</td>
<td>59.6 ± 0.2</td>
<td>5.0 ± 0.5</td>
</tr>
<tr>
<td>10-S:AS-L6</td>
<td>ggaacaagatgc</td>
<td>gcateTgttcc</td>
<td>DNA:LNA</td>
<td>91.3 ± 1.5</td>
<td>249.5 ± 4.4</td>
<td>9.96 ± 0.05</td>
<td>0.73 ± 0.13</td>
<td>57.2 ± 0.1</td>
<td>2.6 ± 0.5</td>
</tr>
<tr>
<td>10-S:AS-L7</td>
<td>ggaacaagatgc</td>
<td>gcateTgttcc</td>
<td>DNA:LNA</td>
<td>89.1 ± 1.5</td>
<td>241.3 ± 4.5</td>
<td>10.37 ± 0.05</td>
<td>1.14 ± 0.13</td>
<td>58.9 ± 0.1</td>
<td>4.3 ± 0.5</td>
</tr>
<tr>
<td>10-S:AS-L8</td>
<td>ggaacaagatgc</td>
<td>gcateTgttcc</td>
<td>DNA:LNA</td>
<td>98.5 ± 1.9</td>
<td>270.3 ± 5.6</td>
<td>10.33 ± 0.04</td>
<td>1.10 ± 0.12</td>
<td>58.2 ± 0.1</td>
<td>3.6 ± 0.5</td>
</tr>
<tr>
<td>10-S-L46:AS</td>
<td>ggaAcAagatgc</td>
<td>gcatcgttcc</td>
<td>Gapmer</td>
<td>92.3 ± 1.8</td>
<td>249.9 ± 5.3</td>
<td>10.75 ± 0.08</td>
<td>1.52 ± 0.14</td>
<td>60.1 ± 0.2</td>
<td>5.6 ± 0.5</td>
</tr>
<tr>
<td>10-S-L67:AS</td>
<td>ggaacAAgatgc</td>
<td>gcatcgttcc</td>
<td>Tandem</td>
<td>90.8 ± 0.5</td>
<td>246.1 ± 1.5</td>
<td>10.53 ± 0.01</td>
<td>1.30 ± 0.12</td>
<td>59.4 ± 0.0</td>
<td>4.9 ± 0.5</td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>Length (bp), Duplex Type</td>
<td>ΔH (kcal mol⁻¹)</td>
<td>ΔS (cal mol⁻¹ K⁻¹)</td>
<td>ΔG° (kcal mol⁻¹)</td>
<td>ΔΔG° (kcal mol⁻¹)</td>
<td>Tₘ (°C)</td>
<td>ΔTₘ (°C)</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>-----------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>10-S-L79:AS</td>
<td>ggaacaAgAtgc</td>
<td>gcacttgttce</td>
<td>Gapmer</td>
<td>93.3 ± 0.9</td>
<td>251.8 ± 2.5</td>
<td>11.10 ± 0.06</td>
<td>1.87 ± 0.13</td>
<td>61.4 ± 0.2</td>
<td>6.8 ± 0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(93.5)</td>
<td>(254.0)</td>
<td>(10.60)</td>
<td>(59.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-S-L:AS-L67</td>
<td>ggaacaagatgc</td>
<td>gcacTTgttce</td>
<td>Tandem</td>
<td>89.8 ± 2.2</td>
<td>241.4 ± 6.3</td>
<td>11.05 ± 0.12</td>
<td>1.82 ± 0.17</td>
<td>61.6 ± 0.3</td>
<td>7.0 ± 0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(94.6)</td>
<td>(255.5)</td>
<td>(11.23)</td>
<td>(61.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>a cgaccagagttacag</td>
<td>ctgtaacetgtgtgt</td>
<td>16</td>
<td>117.2 ± 2.4</td>
<td>317.3 ± 6.9</td>
<td>13.54 ± 0.11</td>
<td>67.3 ± 0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(130.3)</td>
<td>(357.4)</td>
<td>(13.61)</td>
<td>(65.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-S:AS-L4</td>
<td>a cgaccagagttacag</td>
<td>ctgTaacetgtgtgt</td>
<td>DNA:LNA</td>
<td>122.2 ± 1.9</td>
<td>329.1 ± 5.5</td>
<td>14.60 ± 0.11</td>
<td>1.06 ± 0.15</td>
<td>70.1 ± 0.1</td>
<td>2.7 ± 0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(132.0)</td>
<td>(359.5)</td>
<td>(14.55)</td>
<td>(68.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-S:AS-L5</td>
<td>a cgaccagagttacag</td>
<td>ctgtAacetgtgtgt</td>
<td>DNA:LNA</td>
<td>120.0 ± 0.7</td>
<td>324.2 ± 2.1</td>
<td>14.05 ± 0.02</td>
<td>0.52 ± 0.11</td>
<td>68.6 ± 0.1</td>
<td>1.3 ± 0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(131.4)</td>
<td>(358.8)</td>
<td>(14.23)</td>
<td>(67.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-S:AS-L6</td>
<td>a cgaccagagttacag</td>
<td>ctgtaAcctgtgtgt</td>
<td>DNA:LNA</td>
<td>113.7 ± 1.1</td>
<td>304.4 ± 3.3</td>
<td>14.08 ± 0.05</td>
<td>0.55 ± 0.12</td>
<td>69.8 ± 0.0</td>
<td>2.4 ± 0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(131.4)</td>
<td>(358.8)</td>
<td>(14.23)</td>
<td>(67.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-S:AS-L9</td>
<td>a cgaccagagttacag</td>
<td>ctgtaacctCgtgtgt</td>
<td>DNA:LNA</td>
<td>118.8 ± 1.3</td>
<td>318.4 ± 3.6</td>
<td>14.66 ± 0.14</td>
<td>1.12 ± 0.18</td>
<td>70.9 ± 0.2</td>
<td>3.5 ± 0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(132.5)</td>
<td>(360.1)</td>
<td>(14.83)</td>
<td>(69.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-S:AS-L11</td>
<td>a cgaccagagttacag</td>
<td>ctgtaacctGgtgtgt</td>
<td>DNA:LNA</td>
<td>119.7 ± 0.9</td>
<td>320.8 ± 2.7</td>
<td>14.78 ± 0.03</td>
<td>1.24 ± 0.11</td>
<td>71.1 ± 0.1</td>
<td>3.8 ± 0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(131.9)</td>
<td>(359.4)</td>
<td>(14.50)</td>
<td>(68.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-S:AS-L12</td>
<td>a cgaccagagttacag</td>
<td>ctgtaacctGtgtgt</td>
<td>DNA:LNA</td>
<td>114.4 ± 1.5</td>
<td>306.0 ± 4.4</td>
<td>14.27 ± 0.07</td>
<td>0.73 ± 0.13</td>
<td>70.3 ± 0.0</td>
<td>3.0 ± 0.0</td>
</tr>
</tbody>
</table>

99
<table>
<thead>
<tr>
<th>Duplex Name</th>
<th>Sense Strand</th>
<th>Anti-Sense Strand</th>
<th>Length (bp), (5'-3')</th>
<th>(5'-3')</th>
<th>Duplex Type</th>
<th>ΔH (kcal mol(^{-1}))</th>
<th>ΔS (cal mol(^{-1}) K(^{-1}))</th>
<th>ΔG° (kcal mol(^{-1}))</th>
<th>ΔΔG° (kcal mol(^{-1}))</th>
<th>T_m (°C)</th>
<th>ΔT_m (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11-S-L3:AS</td>
<td>acGaccagagttacag</td>
<td>ctgtaacttggtcgt</td>
<td>LNA:DNA</td>
<td>116.2 ± 2.2</td>
<td>313.6 ± 6.4</td>
<td>13.69 ± 0.10</td>
<td>0.15 ± 0.15</td>
<td>68.0 ± 0.0</td>
<td>0.7 ± 0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(131.9)</td>
<td>(359.4)</td>
<td>(14.50)</td>
<td></td>
<td></td>
<td>(68.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-S-L4:AS</td>
<td>acgAccagagttacag</td>
<td>ctgtaacttggtcgt</td>
<td>LNA:DNA</td>
<td>115.8 ± 0.0</td>
<td>311.1 ± 0.0</td>
<td>14.12 ± 0.01</td>
<td>0.58 ± 0.11</td>
<td>69.5 ± 0.0</td>
<td>2.2 ± 0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(131.4)</td>
<td>(358.8)</td>
<td>(14.23)</td>
<td></td>
<td></td>
<td>(67.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-S-L5:AS</td>
<td>acgaCagagttacag</td>
<td>ctgtaacttggtcgt</td>
<td>LNA:DNA</td>
<td>113.0 ± 0.6</td>
<td>301.1 ± 1.7</td>
<td>14.50 ± 0.02</td>
<td>0.96 ± 0.11</td>
<td>71.4 ± 0.1</td>
<td>4.0 ± 0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(132.5)</td>
<td>(360.1)</td>
<td>(14.83)</td>
<td></td>
<td></td>
<td>(69.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-S-L6:AS</td>
<td>acgcCagagttacag</td>
<td>ctgtaacttggtcgt</td>
<td>LNA:DNA</td>
<td>113.3 ± 2.3</td>
<td>302.7 ± 6.6</td>
<td>14.25 ± 0.12</td>
<td>0.71 ± 0.16</td>
<td>70.4 ± 0.0</td>
<td>3.1 ± 0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(132.5)</td>
<td>(360.1)</td>
<td>(14.83)</td>
<td></td>
<td></td>
<td>(69.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-S-L9:AS</td>
<td>acgaccagAgttacag</td>
<td>ctgtaacttggtcgt</td>
<td>LNA:DNA</td>
<td>112.0 ± 1.5</td>
<td>298.0 ± 4.3</td>
<td>14.48 ± 0.07</td>
<td>0.94 ± 0.13</td>
<td>71.5 ± 0.0</td>
<td>4.2 ± 0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(131.4)</td>
<td>(358.8)</td>
<td>(14.23)</td>
<td></td>
<td></td>
<td>(67.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-S-L10:AS</td>
<td>acgaccagaGttacag</td>
<td>ctgtaacttggtcgt</td>
<td>LNA:DNA</td>
<td>110.2 ± 1.8</td>
<td>294.3 ± 5.1</td>
<td>13.96 ± 0.10</td>
<td>0.42 ± 0.15</td>
<td>69.9 ± 0.0</td>
<td>2.6 ± 0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(131.9)</td>
<td>(359.4)</td>
<td>(14.50)</td>
<td></td>
<td></td>
<td>(68.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-S-L11:AS</td>
<td>acgaccagagTtacag</td>
<td>ctgtaacttggtcgt</td>
<td>LNA:DNA</td>
<td>111.6 ± 1.7</td>
<td>298.2 ± 4.9</td>
<td>14.00 ± 0.09</td>
<td>0.46 ± 0.14</td>
<td>69.8 ± 0.1</td>
<td>2.5 ± 0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(132.0)</td>
<td>(359.5)</td>
<td>(14.55)</td>
<td></td>
<td></td>
<td>(68.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-S-L12:AS</td>
<td>acgaccagagTcacag</td>
<td>ctgtaacttggtcgt</td>
<td>LNA:DNA</td>
<td>113.1 ± 1.1</td>
<td>302.3 ± 3.1</td>
<td>14.23 ± 0.05</td>
<td>0.69 ± 0.12</td>
<td>70.4 ± 0.0</td>
<td>3.0 ± 0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(132.0)</td>
<td>(359.5)</td>
<td>(14.55)</td>
<td></td>
<td></td>
<td>(68.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Values reported in parentheses are the corresponding melting thermodynamics predicted by the standard SBT model. Duplex names, for example 1-S-L3:AS, indicate the number (name) of the isosequential pure-DNA duplex (e.g. 1), the LNA content of the sense (S) strand (L3, which indicates an LNA at the 5'±3 nucleotide position), and the LNA content of the anti-sense (AS) strand (the AS strand has no LNA substitution in the 1-S-L3:AS duplex). The reported UVM data (ΔH(Tm), ΔS(Tm), ΔG°, and Tm) are for the helix-to-coil transition and were determined as reported in Materials and Methods. The reported errors refer to the standard deviation of the triplicate runs. All duplex samples were resuspended in buffer containing 1M NaCl, 10mM Na2HPO4, and 1mM Na2EDTA (pH=7.0). Pure DNA or LNA bearing duplexes isosequential to sequences 1 - 7 or 8-11 were resuspended to a CΤ of 7.5 μM or 5 μM, respectively. Duplex sequences marked by superscripts b, c, and d were previously studied by Hughesman et al.[1], Hughesman et al.[136], and McTigue et al.[142], respectively.

Table 3-2 Melting thermodynamic data collected by UV spectroscopy for the “learning set” of complementary duplexes bearing an LNA substitution within both strands.

<table>
<thead>
<tr>
<th>Duplex Name</th>
<th>Sense Strand</th>
<th>Anti-Sense Strand</th>
<th>ΔH (kcal mol⁻¹)</th>
<th>ΔS (cal mol⁻¹ K⁻¹)</th>
<th>ΔG° (kcal mol⁻¹)</th>
<th>ΔΔG° (kcal mol⁻¹)</th>
<th>Tm (°C)</th>
<th>ΔTm (°C)</th>
<th>ΔΔTm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-S-L5:AS-L7</td>
<td>ttcTaGcgcgt</td>
<td>aeggCtAtgaa</td>
<td>75.4 ± 1.3</td>
<td>199.1 ± 4.0</td>
<td>10.41 ± 0.05</td>
<td>1.68 ± 0.07</td>
<td>61.5 ± 0.1</td>
<td>7.7 ± 0.3</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(10.50)</td>
<td>(61.1)</td>
<td>(7.4)</td>
<td>(0.9)</td>
<td></td>
</tr>
<tr>
<td>1-S-L6:AS-L6</td>
<td>ttcAgGcgcgt</td>
<td>aeggcTAtgaa</td>
<td>77.4 ± 0.6</td>
<td>204.1 ± 1.9</td>
<td>10.78 ± 0.01</td>
<td>2.05 ± 0.05</td>
<td>63.0 ± 0.2</td>
<td>9.2 ± 0.3</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(10.50)</td>
<td>(61.1)</td>
<td>(7.4)</td>
<td>(0.9)</td>
<td></td>
</tr>
<tr>
<td>1-S-L7:AS-L5</td>
<td>ttGatGcgcgt</td>
<td>aeggCtAtgaa</td>
<td>73.6 ± 1.2</td>
<td>192.7 ± 3.5</td>
<td>10.64 ± 0.07</td>
<td>1.92 ± 0.09</td>
<td>62.9 ± 0.2</td>
<td>9.1 ± 0.3</td>
<td>-0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(11.03)</td>
<td>(63.2)</td>
<td>(9.5)</td>
<td>(0.9)</td>
<td></td>
</tr>
<tr>
<td>1-S-L8:AS-L4</td>
<td>ttcagGcgcgt</td>
<td>aegGtAtgaa</td>
<td>79.0 ± 1.9</td>
<td>208.1 ± 5.7</td>
<td>11.03 ± 0.09</td>
<td>2.30 ± 0.10</td>
<td>63.9 ± 0.1</td>
<td>10.2 ± 0.3</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(11.03)</td>
<td>(63.2)</td>
<td>(9.5)</td>
<td>(0.9)</td>
<td></td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>ΔH (kcal mol⁻¹)</td>
<td>ΔS (cal mol⁻¹ K⁻¹)</td>
<td>ΔG° (kcal mol⁻¹)</td>
<td>ΔΔG°</td>
<td>Tm (°C)</td>
<td>ΔTm (°C)</td>
<td>ΔΔTm (°C)</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>------------------</td>
<td>----------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>-------</td>
<td>--------</td>
<td>---------</td>
<td>----------</td>
</tr>
<tr>
<td>1-S-L4:AS-L8</td>
<td>ttcAtagccgt</td>
<td>acggctaTgaa</td>
<td>85.7 ± 0.4</td>
<td>230.5 ± 1.1</td>
<td>10.44 ± 0.02</td>
<td>1.71 ± 0.05</td>
<td>60.5 ± 0.0</td>
<td>6.8 ± 0.2</td>
<td>1.7</td>
</tr>
<tr>
<td>1-S-L6:AS-L5</td>
<td>ttcatAgccgt</td>
<td>acggCtatgaa</td>
<td>72.9 ± 2.1</td>
<td>189.7 ± 6.2</td>
<td>10.88 ± 0.11</td>
<td>2.16 ± 0.12</td>
<td>64.2 ± 0.2</td>
<td>10.5 ± 0.3</td>
<td>0.6</td>
</tr>
<tr>
<td>1-S-L7:AS-L4</td>
<td>ttcatatGcgt</td>
<td>acgGctatgaa</td>
<td>71.9 ± 1.8</td>
<td>188.2 ± 5.3</td>
<td>10.47 ± 0.09</td>
<td>1.75 ± 0.10</td>
<td>62.3 ± 0.2</td>
<td>8.5 ± 0.3</td>
<td>2.1</td>
</tr>
<tr>
<td>1-S-L5:AS-L6</td>
<td>ttcaTagccgt</td>
<td>acggeTaTgaa</td>
<td>81.8 ± 1.9</td>
<td>215.3 ± 5.5</td>
<td>11.48 ± 0.10</td>
<td>2.75 ± 0.12</td>
<td>65.6 ± 0.2</td>
<td>11.8 ± 0.3</td>
<td>2.8</td>
</tr>
<tr>
<td>1-S-L4:AS-L7</td>
<td>ttcAtagccgt</td>
<td>acggctAtgaa</td>
<td>79.1 ± 0.3</td>
<td>211.6 ± 1.0</td>
<td>10.04 ± 0.02</td>
<td>1.31 ± 0.05</td>
<td>59.4 ± 0.1</td>
<td>5.6 ± 0.2</td>
<td>1.7</td>
</tr>
<tr>
<td>1-S-L3:AS-L8</td>
<td>ttcAtagccgt</td>
<td>acggctAtgaa</td>
<td>81.7 ± 1.0</td>
<td>215.9 ± 3.0</td>
<td>11.15 ± 0.05</td>
<td>2.43 ± 0.07</td>
<td>64.1 ± 0.1</td>
<td>10.4 ± 0.2</td>
<td>2.9</td>
</tr>
<tr>
<td>1-S-L4:AS-L5</td>
<td>ttcAtagccgt</td>
<td>acggctAtgaa</td>
<td>86.6 ± 1.7</td>
<td>228.7 ± 4.9</td>
<td>11.90 ± 0.08</td>
<td>3.18 ± 0.09</td>
<td>66.7 ± 0.1</td>
<td>12.9 ± 0.2</td>
<td>1.7</td>
</tr>
<tr>
<td>1-S-L4:AS-L6</td>
<td>ttcAtagccgt</td>
<td>acggctAtgaa</td>
<td>79.5 ± 0.9</td>
<td>210.7 ± 2.6</td>
<td>10.73 ± 0.07</td>
<td>2.00 ± 0.09</td>
<td>62.5 ± 0.2</td>
<td>8.7 ± 0.3</td>
<td>1.7</td>
</tr>
<tr>
<td>1-S-L6:AS-L4</td>
<td>ttcatAgccgt</td>
<td>acgGctatgaa</td>
<td>78.5 ± 0.6</td>
<td>206.9 ± 1.8</td>
<td>10.91 ± 0.02</td>
<td>2.19 ± 0.06</td>
<td>63.5 ± 0.1</td>
<td>9.7 ± 0.2</td>
<td>2.9</td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>ΔH (kcal mol⁻¹)</td>
<td>ΔS (cal mol⁻¹ K⁻¹)</td>
<td>ΔG° (kcal mol⁻¹)</td>
<td>ΔΔG° (kcal mol⁻¹)</td>
<td>Tm (°C)</td>
<td>ΔTm (°C)</td>
<td>ΔΔTm (°C)</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>------------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>------------------</td>
<td>--------</td>
<td>---------</td>
<td>----------</td>
</tr>
<tr>
<td>1-S-L3:AS-L7</td>
<td>ttCatagccgt</td>
<td>aeggcTatgaa</td>
<td>87.0 ± 0.3</td>
<td>232.2 ± 1.0</td>
<td>11.17 ± 0.03</td>
<td>2.45 ± 0.06</td>
<td>63.5 ± 0.1</td>
<td>9.7 ± 0.2</td>
<td>3.4</td>
</tr>
<tr>
<td>1-S-L3:AS-L6</td>
<td>ttCatagccgt</td>
<td>aeggcTatgaa</td>
<td>80.7 ± 1.0</td>
<td>212.6 ± 2.8</td>
<td>11.25 ± 0.06</td>
<td>2.53 ± 0.08</td>
<td>64.7 ± 0.1</td>
<td>11.0 ± 0.2</td>
<td>1.6</td>
</tr>
<tr>
<td>1-S-L5:AS-L4</td>
<td>ttcaTagccgt</td>
<td>aegGctatgaa</td>
<td>87.6 ± 3.9</td>
<td>233.4 ± 11.4</td>
<td>11.40 ± 0.22</td>
<td>2.68 ± 0.22</td>
<td>64.3 ± 0.4</td>
<td>10.6 ± 0.4</td>
<td>2.4</td>
</tr>
<tr>
<td>1-S-L4:AS-L5</td>
<td>ttcAtagccgt</td>
<td>aeggCtatgaa</td>
<td>77.2 ± 2.8</td>
<td>204.0 ± 8.3</td>
<td>10.60 ± 0.14</td>
<td>1.87 ± 0.15</td>
<td>62.1 ± 0.3</td>
<td>8.4 ± 0.4</td>
<td>-0.8</td>
</tr>
<tr>
<td>1-S-L4:AS-L4</td>
<td>ttcAtagccgt</td>
<td>aeggCtatgaa</td>
<td>72.6 ± 3.9</td>
<td>192.3 ± 11.3</td>
<td>9.87 ± 0.18</td>
<td>1.14 ± 0.19</td>
<td>59.2 ± 0.2</td>
<td>5.4 ± 0.3</td>
<td>-0.8</td>
</tr>
<tr>
<td>1-S-L3:AS-L5</td>
<td>ttCatagccgt</td>
<td>aeggCtatgaa</td>
<td>88.2 ± 0.3</td>
<td>234.2 ± 1.0</td>
<td>11.74 ± 0.01</td>
<td>3.01 ± 0.05</td>
<td>65.7 ± 0.0</td>
<td>11.9 ± 0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>1-S-L4:AS-L9</td>
<td>ttcAtagccgt</td>
<td>aeggctatGaa</td>
<td>77.5 ± 2.2</td>
<td>208.1 ± 6.5</td>
<td>9.59 ± 0.07</td>
<td>0.86 ± 0.09</td>
<td>57.5 ± 0.2</td>
<td>3.7 ± 0.3</td>
<td>0.7</td>
</tr>
<tr>
<td>1-S-L5:AS-L8</td>
<td>ttcaTagccgt</td>
<td>aeggctaTgaa</td>
<td>78.9 ± 1.8</td>
<td>210.8 ± 5.2</td>
<td>10.10 ± 0.07</td>
<td>1.37 ± 0.09</td>
<td>59.7 ± 0.1</td>
<td>5.9 ± 0.3</td>
<td>-1.2</td>
</tr>
<tr>
<td>1-S-L6:AS-L7</td>
<td>ttcatAgeccgt</td>
<td>aeggetAtgaa</td>
<td>75.9 ± 1.1</td>
<td>201.1 ± 2.8</td>
<td>10.27 ± 0.05</td>
<td>1.54 ± 0.07</td>
<td>59.4 ± 0.1</td>
<td>5.6 ± 0.3</td>
<td>1.1</td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>ΔH (kcal mol⁻¹)</td>
<td>ΔS (cal mol⁻¹ K⁻¹)</td>
<td>ΔG° (kcal mol⁻¹)</td>
<td>ΔΔG (kcal mol⁻¹)</td>
<td>Tm (°C)</td>
<td>ΔTm (°C)</td>
<td>ΔΔTm (°C)</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>---------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>1-S-L7:AS-L6</td>
<td>ttcataGcggt</td>
<td>acggcTatgaa</td>
<td>72.7 ± 2.3</td>
<td>191.5 ± 6.8</td>
<td>10.22 ± 0.09</td>
<td>1.49 ± 0.10</td>
<td>60.9 ± 0.2</td>
<td>7.1 ± 0.3</td>
<td>-0.1</td>
</tr>
<tr>
<td>1-S-L8:AS-L5</td>
<td>ttcataCggt</td>
<td>acggCtatgaa</td>
<td>76.1 ± 0.3</td>
<td>199.6 ± 1.0</td>
<td>10.91 ± 0.02</td>
<td>2.18 ± 0.05</td>
<td>63.8 ± 0.0</td>
<td>10 ± 0.2</td>
<td>-1.9</td>
</tr>
<tr>
<td>1-S-L6:AS-L8</td>
<td>ttcatAgccgt</td>
<td>acggctaTgaa</td>
<td>77.7 ± 2.3</td>
<td>207.8 ± 6.9</td>
<td>9.87 ± 0.08</td>
<td>1.15 ± 0.09</td>
<td>58.8 ± 0.2</td>
<td>5 ± 0.3</td>
<td>-0.7</td>
</tr>
<tr>
<td>1-S-L7:AS-L7</td>
<td>ttcataGcggt</td>
<td>acggetAgtgaa</td>
<td>85.4 ± 1.5</td>
<td>231.3 ± 4.4</td>
<td>9.98 ± 0.06</td>
<td>1.25 ± 0.08</td>
<td>58.6 ± 0.1</td>
<td>4.9 ± 0.3</td>
<td>0.8</td>
</tr>
<tr>
<td>1-S-L8:AS-L6</td>
<td>ttcataCggt</td>
<td>acggeCtatgaa</td>
<td>85.5 ± 2.6</td>
<td>227.8 ± 7.7</td>
<td>11.16 ± 0.11</td>
<td>2.43 ± 0.12</td>
<td>63.6 ± 0.1</td>
<td>9.8 ± 0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>1-S-L7:AS-L8</td>
<td>ttcataGcggt</td>
<td>acggetaTgaa</td>
<td>86.7 ± 3.8</td>
<td>234.7 ± 11.3</td>
<td>10.15 ± 0.09</td>
<td>1.42 ± 0.10</td>
<td>59.2 ± 0.1</td>
<td>5.5 ± 0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>1-S-L8:AS-L7</td>
<td>ttcataCggt</td>
<td>acggetAgtgaa</td>
<td>85.3 ± 0.9</td>
<td>229.3 ± 2.7</td>
<td>10.52 ± 0.04</td>
<td>1.80 ± 0.07</td>
<td>60.9 ± 0.1</td>
<td>7.2 ± 0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>2-S-L7:AS-L3</td>
<td>tgcggaTaagt</td>
<td>acTtatecgea</td>
<td>83.4 ± 2.5</td>
<td>222.3 ± 7.4</td>
<td>10.88 ± 0.10</td>
<td>2.12 ± 0.11</td>
<td>62.6 ± 0.2</td>
<td>8.8 ± 0.2</td>
<td>3.5</td>
</tr>
<tr>
<td>2-S-L9:AS-L5</td>
<td>tgcggataAgt</td>
<td>aettAtecega</td>
<td>79.7 ± 3.7</td>
<td>213.9 ± 10.7</td>
<td>9.94 ± 0.15</td>
<td>1.18 ± 0.15</td>
<td>58.9 ± 0.4</td>
<td>5.1 ± 0.4</td>
<td>-0.5</td>
</tr>
<tr>
<td>Duplex Name</td>
<td><em>Sense</em></td>
<td><em>Anti-Sense</em></td>
<td>ΔH</td>
<td>ΔS</td>
<td>ΔG°</td>
<td>ΔΔG&lt;sup&gt;‡&lt;/sup&gt;</td>
<td>T&lt;sub&gt;m&lt;/sub&gt;</td>
<td>ΔT&lt;sub&gt;m&lt;/sub&gt;</td>
<td>ΔΔT&lt;sub&gt;m&lt;/sub&gt;</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>--------------</td>
<td>---------------</td>
<td>---------------</td>
<td>--------------</td>
<td>----------------</td>
<td>------------</td>
<td>-------------</td>
<td>----------------</td>
</tr>
<tr>
<td>3-S-L4:AS-L8</td>
<td>ctaAcggatgc</td>
<td>gcatccgTtag</td>
<td>82.0 ± 3.7</td>
<td>221.1 ± 11.0</td>
<td>9.87 ± 0.13</td>
<td>1.12 ± 0.14</td>
<td>58.4 ± 0.3</td>
<td>4.6 ± 0.4</td>
<td>-1.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(10.40)</td>
<td>(60.2)</td>
<td>(7.0)</td>
<td>(0.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-S-L5:AS-L7</td>
<td>ctaaCggatgc</td>
<td>gcatccGttag</td>
<td>78.8 ± 1.6</td>
<td>209.1 ± 4.6</td>
<td>10.50 ± 0.06</td>
<td>1.75 ± 0.08</td>
<td>61.5 ± 0.1</td>
<td>7.7 ± 0.2</td>
<td>-0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(10.93)</td>
<td>(62.3)</td>
<td>(9.0)</td>
<td>(0.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-S-L6:AS-L6</td>
<td>ctaacGgatgc</td>
<td>gcateCgtag</td>
<td>77.1 ± 0.4</td>
<td>203.0 ± 1.1</td>
<td>10.84 ± 0.03</td>
<td>2.09 ± 0.05</td>
<td>63.3 ± 0.1</td>
<td>9.5 ± 0.2</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(10.93)</td>
<td>(62.3)</td>
<td>(9.0)</td>
<td>(0.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-S-L7:AS-L5</td>
<td>ctaacgGatgc</td>
<td>gcatCcgtag</td>
<td>75.9 ± 1.7</td>
<td>199.1 ± 5.1</td>
<td>10.89 ± 0.12</td>
<td>2.14 ± 0.13</td>
<td>63.8 ± 0.6</td>
<td>9.9 ± 0.6</td>
<td>-0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(10.93)</td>
<td>(62.3)</td>
<td>(9.0)</td>
<td>(0.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-S-L8:AS-L4</td>
<td>ctaacggAtgc</td>
<td>gcaTcctgtag</td>
<td>78.9 ± 1.1</td>
<td>211.1 ± 3.4</td>
<td>10.01 ± 0.03</td>
<td>1.26 ± 0.05</td>
<td>59.3 ± 0.1</td>
<td>5.4 ± 0.2</td>
<td>-1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(10.40)</td>
<td>(60.2)</td>
<td>(7.0)</td>
<td>(0.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-S-L4:AS-L7</td>
<td>ctaAeggatgc</td>
<td>gcatccgTtag</td>
<td>83.8 ± 3.4</td>
<td>224.6 ± 9.5</td>
<td>10.50 ± 0.27</td>
<td>1.75 ± 0.27</td>
<td>60.9 ± 0.8</td>
<td>7.1 ± 0.8</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(10.67)</td>
<td>(61.2)</td>
<td>(7.9)</td>
<td>(2.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-S-L5:AS-L6</td>
<td>ctaaCggatgc</td>
<td>gcatCcgtag</td>
<td>76.0 ± 0.7</td>
<td>197.4 ± 2.0</td>
<td>11.48 ± 0.03</td>
<td>2.73 ± 0.05</td>
<td>66.7 ± 0.0</td>
<td>12.8 ± 0.2</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(11.58)</td>
<td>(64.8)</td>
<td>(11.5)</td>
<td>(2.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-S-L6:AS-L5</td>
<td>ctaacGgatgc</td>
<td>gcatCgttag</td>
<td>75.1 ± 1.9</td>
<td>195.8 ± 5.3</td>
<td>11.17 ± 0.11</td>
<td>2.42 ± 0.11</td>
<td>65.3 ± 0.2</td>
<td>11.4 ± 0.3</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(11.26)</td>
<td>(63.5)</td>
<td>(10.2)</td>
<td>(2.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-S-L7:AS-L4</td>
<td>ctaacgGatgc</td>
<td>gcaTcctgtag</td>
<td>74.1 ± 2.7</td>
<td>193.5 ± 7.7</td>
<td>10.90 ± 0.14</td>
<td>2.15 ± 0.15</td>
<td>64.1 ± 0.3</td>
<td>10.3 ± 0.3</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(10.98)</td>
<td>(62.5)</td>
<td>(9.2)</td>
<td>(2.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>$\Delta H$ (kcal mol$^{-1}$)</td>
<td>$\Delta S$ (cal mol$^{-1}$ K$^{-1}$)</td>
<td>$\Delta G^\circ$ (kcal mol$^{-1}$)</td>
<td>$\Delta \Delta G^\ddagger$ (kcal mol$^{-1}$)</td>
<td>$T_m$ (°C)</td>
<td>$\Delta T_m$ (°C)</td>
<td>$\Delta \Delta T_m$ (°C)</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>----------------</td>
<td>-----------------------------</td>
<td>---------------------------------</td>
<td>---------------------------------</td>
<td>---------------------------</td>
<td>---------</td>
<td>----------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>3-S-L4:AS-L6</td>
<td>ctaAcggatgc</td>
<td>gccatCgttag</td>
<td>80.2 ± 0.8</td>
<td>211.8 ± 2.4</td>
<td>11.03 ± 0.05</td>
<td>2.28 ± 0.06</td>
<td>63.8 ± 0.1</td>
<td>9.9 ± 0.2</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(11.09)</td>
<td></td>
<td>(62.9)</td>
<td></td>
<td>(9.6)</td>
<td></td>
<td>(2.5)</td>
</tr>
<tr>
<td>3-S-L5:AS-L6</td>
<td>ctaAcggatgc</td>
<td>gccatCgttag</td>
<td>75.3 ± 2.6</td>
<td>195.9 ± 7.6</td>
<td>11.28 ± 0.14</td>
<td>2.54 ± 0.15</td>
<td>65.8 ± 0.2</td>
<td>12 ± 0.3</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(11.68)</td>
<td></td>
<td>(65.2)</td>
<td></td>
<td>(11.9)</td>
<td></td>
<td>(2.4)</td>
</tr>
<tr>
<td>3-S-L6:AS-L4</td>
<td>ctaacGgatgc</td>
<td>gccatCgttag</td>
<td>78.4 ± 1.3</td>
<td>206.5 ± 3.6</td>
<td>10.96 ± 0.07</td>
<td>2.21 ± 0.08</td>
<td>63.7 ± 0.2</td>
<td>9.9 ± 0.2</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(11.08)</td>
<td></td>
<td>(62.9)</td>
<td></td>
<td>(9.6)</td>
<td></td>
<td>(2.5)</td>
</tr>
<tr>
<td>3-S-L4:AS-L5</td>
<td>ctaAcggatgc</td>
<td>gccatCgttag</td>
<td>82.1 ± 5.1</td>
<td>216.2 ± 17.6</td>
<td>11.53 ± 0.54</td>
<td>2.78 ± 0.54</td>
<td>62.4 ± 0.7</td>
<td>8.6 ± 0.7</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(10.79)</td>
<td></td>
<td>(61.8)</td>
<td></td>
<td>(8.5)</td>
<td></td>
<td>(1.3)</td>
</tr>
<tr>
<td>3-S-L5:AS-L4</td>
<td>ctaacGgatgc</td>
<td>gccatCgttag</td>
<td>80.0 ± 0.4</td>
<td>211.0 ± 1.1</td>
<td>11.11 ± 0.02</td>
<td>2.37 ± 0.05</td>
<td>64.2 ± 0.9</td>
<td>10.3 ± 0.2</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(11.11)</td>
<td></td>
<td>(63.0)</td>
<td></td>
<td>(9.8)</td>
<td></td>
<td>(1.3)</td>
</tr>
<tr>
<td>3-S-L5:AS-L8</td>
<td>ctaacGgatgc</td>
<td>gccatCgttag</td>
<td>76.1 ± 0.3</td>
<td>202.1 ± 8.8</td>
<td>10.17 ± 0.12</td>
<td>1.42 ± 0.13</td>
<td>60.3 ± 0.3</td>
<td>6.4 ± 0.3</td>
<td>-0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(10.70)</td>
<td></td>
<td>(61.5)</td>
<td></td>
<td>(8.2)</td>
<td></td>
<td>(-0.2)</td>
</tr>
<tr>
<td>3-S-L6:AS-L7</td>
<td>ctaacGgatgc</td>
<td>gccatCgttag</td>
<td>77.7 ± 1.7</td>
<td>206.3 ± 5.1</td>
<td>10.38 ± 0.08</td>
<td>1.63 ± 0.09</td>
<td>61.1 ± 0.2</td>
<td>7.2 ± 0.3</td>
<td>-0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(10.32)</td>
<td></td>
<td>(61.5)</td>
<td></td>
<td>(6.7)</td>
<td></td>
<td>(-0.2)</td>
</tr>
<tr>
<td>3-S-L7:AS-L6</td>
<td>ctaacGgatgc</td>
<td>gccatCgttag</td>
<td>74.1 ± 2.7</td>
<td>193.5 ± 7.7</td>
<td>10.90 ± 0.14</td>
<td>2.15 ± 0.15</td>
<td>64.1 ± 0.3</td>
<td>10.3 ± 0.3</td>
<td>-0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(10.65)</td>
<td></td>
<td>(61.2)</td>
<td></td>
<td>(8.0)</td>
<td></td>
<td>(-0.2)</td>
</tr>
<tr>
<td>3-S-L8:AS-L5</td>
<td>ctaacGgatgc</td>
<td>gccatCgttag</td>
<td>76.5 ± 1.5</td>
<td>203.0 ± 4.5</td>
<td>10.31 ± 0.06</td>
<td>1.56 ± 0.07</td>
<td>60.9 ± 0.1</td>
<td>7.0 ± 0.2</td>
<td>-1.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(10.39)</td>
<td></td>
<td>(60.2)</td>
<td></td>
<td>(6.9)</td>
<td></td>
<td>(-0.2)</td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>ΔH (kcal mol⁻¹)</td>
<td>ΔS (cal mol⁻¹ K⁻¹)</td>
<td>ΔG° (kcal mol⁻¹)</td>
<td>ΔΔG⁻¹ (kcal mol⁻¹)</td>
<td>Tm (°C)</td>
<td>ΔTm (°C)</td>
<td>ΔΔTm (°C)</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>------------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>-------------------</td>
<td>--------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>3-S-L4:AS-L9</td>
<td>ctaAcgatgc</td>
<td>gcacgctTag</td>
<td>82.5 ± 1.6</td>
<td>222.0 ± 4.6</td>
<td>10.06 ± 0.05</td>
<td>1.31 ± 0.07</td>
<td>59.2 ± 0.1</td>
<td>5.4 ± 0.2</td>
<td>-0.4</td>
</tr>
<tr>
<td>3-S-L6:AS-L8</td>
<td>ctaacGatgc</td>
<td>gcacgcTag</td>
<td>82.2 ± 0.8</td>
<td>220.8 ± 2.5</td>
<td>10.20 ± 0.02</td>
<td>1.45 ± 0.05</td>
<td>59.8 ± 0.0</td>
<td>6.0 ± 0.2</td>
<td>-1.0</td>
</tr>
<tr>
<td>3-S-L8:AS-L6</td>
<td>ctaacggAtgc</td>
<td>gcacgcGTag</td>
<td>78.1 ± 1.2</td>
<td>207.4 ± 3.9</td>
<td>10.44 ± 0.04</td>
<td>1.69 ± 0.06</td>
<td>60.8 ± 0.1</td>
<td>7.0 ± 0.2</td>
<td>-1.3</td>
</tr>
<tr>
<td>3-S-L6:AS-L7</td>
<td>ctaacgGatgc</td>
<td>gcacgcGTag</td>
<td>79.7 ± 0.7</td>
<td>212.8 ± 2.2</td>
<td>10.28 ± 0.02</td>
<td>1.53 ± 0.05</td>
<td>60.8 ± 0.6</td>
<td>7.0 ± 0.7</td>
<td>-1.9</td>
</tr>
<tr>
<td>3-S-L8:AS-L7</td>
<td>ctaacggAtgc</td>
<td>gcacgcGTag</td>
<td>78.1 ± 1.9</td>
<td>209.5 ± 5.8</td>
<td>9.74 ± 0.07</td>
<td>0.99 ± 0.08</td>
<td>58.1 ± 0.2</td>
<td>4.3 ± 0.2</td>
<td>-2.6</td>
</tr>
<tr>
<td>3-S-L6:AS-L8</td>
<td>ctaacgGatgc</td>
<td>gcacgcGTag</td>
<td>77.7 ± 2.2</td>
<td>206.2 ± 6.3</td>
<td>10.43 ± 0.10</td>
<td>1.68 ± 0.11</td>
<td>61.3 ± 0.2</td>
<td>7.5 ± 0.3</td>
<td>-1.0</td>
</tr>
<tr>
<td>4-S-L6:AS-L6</td>
<td>ctacgcattcc</td>
<td>ggaatGcTag</td>
<td>79.3 ± 1.8</td>
<td>209.3 ± 5.3</td>
<td>10.97 ± 0.06</td>
<td>2.31 ± 0.09</td>
<td>63.6 ± 0.0</td>
<td>10.2 ± 0.3</td>
<td>1.6</td>
</tr>
<tr>
<td>4-S-L8:AS-L3</td>
<td>ctacgcatTcc</td>
<td>ggaatgcGTag</td>
<td>87.6 ± 3.8</td>
<td>234.5 ± 11.3</td>
<td>11.01 ± 0.13</td>
<td>2.36 ± 0.15</td>
<td>62.7 ± 0.1</td>
<td>9.3 ± 0.3</td>
<td>2.0</td>
</tr>
<tr>
<td>4-S-L4:AS-L8</td>
<td>ctaCgcattcc</td>
<td>ggaatgcGTag</td>
<td>84.1 ± 0.4</td>
<td>223.8 ± 1.2</td>
<td>10.99 ± 0.01</td>
<td>2.34 ± 0.07</td>
<td>63.1 ± 0.0</td>
<td>9.6 ± 0.3</td>
<td>2.3</td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand (5'-3')</td>
<td>Anti-Sense Strand (5'-3')</td>
<td>ΔH (kcal mol⁻¹)</td>
<td>ΔS (cal mol⁻¹ K⁻¹)</td>
<td>ΔG° (kcal mol⁻¹)</td>
<td>ΔΔG° (kcal mol⁻¹)</td>
<td>Tₘ (°C)</td>
<td>ΔTₘ (°C)</td>
<td>ΔΔTₘ (°C)</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------------</td>
<td>--------------------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>------------------</td>
<td>--------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>4-S-L7:AS-L5</td>
<td>ctaegcAttcc</td>
<td>ggaaTgcgtag</td>
<td>79.8 ± 1.4</td>
<td>211.7 ± 4.0</td>
<td>10.67 ± 0.07</td>
<td>2.01 ± 0.09</td>
<td>62.2 ± 0.1</td>
<td>8.7 ± 0.3</td>
<td>2.9</td>
</tr>
<tr>
<td>4-S-L8:AS-L3</td>
<td>ctaegcaTtcc</td>
<td>ggaAtgcgtag</td>
<td>82.4 ± 1.5</td>
<td>219.4 ± 4.4</td>
<td>10.75 ± 0.07</td>
<td>2.10 ± 0.10</td>
<td>62.2 ± 0.1</td>
<td>8.8 ± 0.3</td>
<td>2.6</td>
</tr>
<tr>
<td>4-S-L7:AS-L3</td>
<td>ctaegcAttcc</td>
<td>ggaAtgcgtag</td>
<td>77.0 ± 1.8</td>
<td>204.0 ± 5.4</td>
<td>10.41 ± 0.07</td>
<td>1.76 ± 0.09</td>
<td>61.3 ± 0.1</td>
<td>7.9 ± 0.3</td>
<td>3.3</td>
</tr>
<tr>
<td>4-S-L4:AS-L6</td>
<td>ctaGgattcc</td>
<td>ggaatGcgtag</td>
<td>81.4 ± 0.4</td>
<td>215.1 ± 1.2</td>
<td>11.13 ± 0.01</td>
<td>2.48 ± 0.07</td>
<td>64.1 ± 0.0</td>
<td>10.6 ± 0.3</td>
<td>3.9</td>
</tr>
<tr>
<td>4-S-L4:AS-L5</td>
<td>ctaGgccttcc</td>
<td>ggaatTgcgtag</td>
<td>85.8 ± 0.4</td>
<td>227.7 ± 1.0</td>
<td>11.46 ± 0.04</td>
<td>2.80 ± 0.07</td>
<td>64.8 ± 0.1</td>
<td>11.4 ± 0.3</td>
<td>2.7</td>
</tr>
<tr>
<td>4-S-L9:AS-L5</td>
<td>ctacgcTcTcc</td>
<td>ggaatTgcgtag</td>
<td>76.8 ± 4.3</td>
<td>204.0 ± 12.5</td>
<td>10.24 ± 0.18</td>
<td>1.59 ± 0.19</td>
<td>60.5 ± 0.4</td>
<td>7.1 ± 0.5</td>
<td>-1.4</td>
</tr>
<tr>
<td>4-S-L6:AS-L8</td>
<td>ctacgCattcc</td>
<td>ggaatgeGtag</td>
<td>81.0 ± 1.7</td>
<td>214.8 ± 4.1</td>
<td>10.91 ± 0.06</td>
<td>2.25 ± 0.09</td>
<td>61.1 ± 0.6</td>
<td>7.7 ± 0.7</td>
<td>-1.5</td>
</tr>
<tr>
<td>5-S-L8:AS-L6</td>
<td>cttatggCgac</td>
<td>gtcgcCaatag</td>
<td>82.1 ± 3.0</td>
<td>218.3 ± 6.3</td>
<td>10.88 ± 0.26</td>
<td>1.64 ± 0.26</td>
<td>62.8 ± 0.9</td>
<td>7.0 ± 0.9</td>
<td>-1.0</td>
</tr>
<tr>
<td>6-S-L6:AS-L6</td>
<td>ctgaatGtcgc</td>
<td>gcggacTtcag</td>
<td>82.9 ± 0.6</td>
<td>216.6 ± 1.8</td>
<td>12.12 ± 0.04</td>
<td>2.33 ± 0.11</td>
<td>68.4 ± 0.0</td>
<td>10.3 ± 0.2</td>
<td>-0.7</td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>$\Delta H$ (kcal mol$^{-1}$)</td>
<td>$\Delta S$ (cal mol$^{-1}$ K$^{-1}$)</td>
<td>$\Delta G^\circ$ (kcal mol$^{-1}$)</td>
<td>$\Delta \Delta G^\ddagger$ (kcal mol$^{-1}$)</td>
<td>$T_m$ (°C)</td>
<td>$\Delta T_m$ (°C)</td>
<td>$\Delta \Delta T_m$ (°C)</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>---------------------------------</td>
<td>---------------------------------</td>
<td>-------------------</td>
<td>-----</td>
<td>----------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>6-S-L6:AS-L4</td>
<td>ctgaaGtccgc</td>
<td>gcgGaacttcag</td>
<td>84.2 ± 0.9</td>
<td>220.7 ± 2.5</td>
<td>12.07 ± 0.05</td>
<td>2.27 ± 0.11</td>
<td>67.9 ± 0.0</td>
<td>9.8 ± 0.2</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>(5'-3')</td>
<td></td>
<td>(12.04)</td>
<td></td>
<td></td>
<td></td>
<td>(66.4)</td>
<td>(8.9)</td>
<td>(0.8)</td>
</tr>
<tr>
<td>6-S-L6:AS-L3</td>
<td>ctgaaGtccgc</td>
<td>gcGgaacttcag</td>
<td>85.9 ± 0.8</td>
<td>226.9 ± 2.2</td>
<td>11.76 ± 0.03</td>
<td>1.96 ± 0.11</td>
<td>66.2 ± 0.1</td>
<td>8.1 ± 0.2</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>(5'-3')</td>
<td></td>
<td>(12.14)</td>
<td></td>
<td></td>
<td></td>
<td>(66.7)</td>
<td>(9.2)</td>
<td>(2.4)</td>
</tr>
<tr>
<td>6-S-L4:AS-L5</td>
<td>ctgAagtcgc</td>
<td>gcggActtcag</td>
<td>87.3 ± 0.3</td>
<td>229.9 ± 1.0</td>
<td>12.18 ± 0.02</td>
<td>2.39 ± 0.11</td>
<td>67.8 ± 0.0</td>
<td>9.7 ± 0.2</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>(5'-3')</td>
<td></td>
<td>(11.84)</td>
<td></td>
<td></td>
<td></td>
<td>(65.6)</td>
<td>(8.1)</td>
<td>(1.3)</td>
</tr>
<tr>
<td>6-S-L5:AS-L4</td>
<td>ctgAagtcgc</td>
<td>gcgGacttcag</td>
<td>85.3 ± 2.2</td>
<td>223.1 ± 6.4</td>
<td>12.39 ± 0.14</td>
<td>2.59 ± 0.17</td>
<td>69.1 ± 0.1</td>
<td>11.0 ± 0.2</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>(5'-3')</td>
<td></td>
<td>(11.58)</td>
<td></td>
<td></td>
<td></td>
<td>(64.6)</td>
<td>(6.1)</td>
<td>(1.3)</td>
</tr>
<tr>
<td>6-S-L3:AS-L6</td>
<td>ctGaagtccgc</td>
<td>gcggacTtcag</td>
<td>92.0 ± 1.3</td>
<td>244.6 ± 3.7</td>
<td>12.12 ± 0.06</td>
<td>2.32 ± 0.12</td>
<td>66.7 ± 0.1</td>
<td>8.6 ± 0.2</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>(5'-3')</td>
<td></td>
<td>(12.17)</td>
<td></td>
<td></td>
<td></td>
<td>(66.9)</td>
<td>(9.4)</td>
<td>(1.3)</td>
</tr>
<tr>
<td>7-S-L3:AS-L6</td>
<td>gtaAtcaagtet</td>
<td>agactTgatac</td>
<td>85.0 ± 0.9</td>
<td>232.9 ± 2.8</td>
<td>9.01 ± 0.03</td>
<td>1.86 ± 0.04</td>
<td>54.8 ± 0.1</td>
<td>7.5 ± 0.3</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>(5'-3')</td>
<td></td>
<td>(8.50)</td>
<td></td>
<td></td>
<td></td>
<td>(52.8)</td>
<td>(8.1)</td>
<td>(1.5)</td>
</tr>
<tr>
<td>7-S-L4:AS-L5</td>
<td>gtaTcaagctt</td>
<td>agacTgatac</td>
<td>82.6 ± 1.0</td>
<td>225.4 ± 3.0</td>
<td>9.06 ± 0.02</td>
<td>1.92 ± 0.03</td>
<td>55.1 ± 0.1</td>
<td>7.8 ± 0.2</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>(5'-3')</td>
<td></td>
<td>(8.81)</td>
<td></td>
<td></td>
<td></td>
<td>(54.2)</td>
<td>(9.5)</td>
<td>(1.5)</td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>ΔH (kcal mol⁻¹)</td>
<td>ΔS (cal mol⁻¹ K⁻¹)</td>
<td>ΔG° (kcal mol⁻¹)</td>
<td>ΔΔG° (kcal mol⁻¹)</td>
<td>Tm (°C)</td>
<td>ΔTm (°C)</td>
<td>ΔΔTm (°C)</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>------------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>------------------</td>
<td>--------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>7-S-L5:AS-L4</td>
<td>gtatCaagtct</td>
<td>agaCttgatac</td>
<td>80.8 ± 1.0</td>
<td>216.7 ± 3.1</td>
<td>10.07 ± 0.04</td>
<td>2.92 ± 0.04</td>
<td>59.4 ± 0.1</td>
<td>12.1 ± 0.2</td>
<td>1.3 (9.36)</td>
</tr>
</tbody>
</table>

These data were globally fit to the extended SBT model using the Levenberg-Marquardt method to regress the required set of ΔΔG°₂_hyper⁻⁻(f) parameters. Values reported in parenthesis are the corresponding melting thermodynamics predicted by the extended SBT model. The reported UVM data (ΔH(Tm), ΔS(Tm), ΔG°, and Tm) are for the helix-to-coil transition and were determined as reported in Materials and Methods. The reported errors refer to the standard deviation of triplicate runs. All duplex samples were resuspended in buffer containing 1 M NaCl, 10 mM Na₂HPO₄, and 1 mM Na₂EDTA (pH=7.0) to a C_T = 7.5 μM.
3.4.3 THE INCREMENTAL STABILITY OF AN LNA:LNA BASE PAIR EXCEEDS THE SUMMED INCREMENTAL STABILITIES OF THE TWO LNA:DNA BASE PAIRS FROM WHICH IT IS DERIVED

Our library of duplexes containing an LNA substitution within each strand (Table 3-2) includes 15 duplexes where the LNAs align to form a single LNA:LNA base pair (i.e., the $j = 0$ case). For each, the library also includes (see Table 3-1) two single-LNA-substituted duplexes of the same sequence, each of which bears one of the two parent LNA:DNA base pairs from which the LNA:LNA base pair can form via a second LNA substitution. The library likewise includes the unsubstituted pure-DNA isosequence, which completes a set of four duplexes of the same sequence but of varying LNA content at the $j = 0$ position. Melting thermodynamic data were collected for the 15 such sets of sequence-related duplexes so as to parse out the incremental contribution of an LNA:LNA base pair to duplex stability, including its relation to the stabilities of the two duplexes bearing one of the parent LNA:DNA base pairs. Taking the Gibbs energy change for melting of the isosequential pure-DNA duplex $\Delta G_{\text{DNA}}^0$ at $T_{\text{ref}}$ ($53^\circ \text{C}$) as a reference, the incremental enhancement to duplex stability ($\Delta \Delta G_{\text{LNA:LNA} \rightarrow 0}^0$) provided by the replacement of an internal DNA:DNA base pair with its corresponding LNA:LNA base pair is given at $T_{\text{ref}}$ by

$$\Delta \Delta G_{\text{LNA:LNA} \rightarrow 0}^0 = \Delta G_{\text{LNA}}^0 - \Delta G_{\text{DNA}}^0$$

(3-16)

As the standard SBT model accounts for the energetics of LNA:DNA base pairs, it can be most easily extended to permit prediction of melting thermodynamics for duplexes bearing an LNA:LNA base pair by expressing $\Delta \Delta G_{\text{LNA:LNA} \rightarrow 0}^0$ as

$$\Delta \Delta G_{\text{LNA:LNA} \rightarrow 0}^0 = \Delta \Delta G_{\text{LNA:DNA}}^0 + \Delta \Delta G_{\text{DNA:LNA}}^0 + \Delta \Delta G_{\text{hyper} \rightarrow (0)}^0$$

(3-17)
where

$$\Delta \Delta G_{\text{LNA:DNA}}^o = \Delta G_{\text{LNA:DNA}}^o(T_{\text{ref}}) - \Delta G_{\text{DNA}}^o(T_{\text{ref}}) \tag{3-18}$$

and

$$\Delta \Delta G_{\text{DNA:LNA}}^o = \Delta G_{\text{DNA:LNA}}^o(T_{\text{ref}}) - \Delta G_{\text{DNA}}^o(T_{\text{ref}}) \tag{3-19}$$

In equation (3-17), $\Delta G_{\text{LNA:DNA}}^o$ and $\Delta G_{\text{DNA:LNA}}^o$ are defined at $T_{\text{ref}}$ and are computed using the standard SBT model and the relation $\Delta G^o = \Delta H^o - T_{\text{ref}} \Delta S^o$; $\Delta \Delta G_{\text{LNA:DNA}}^o$ and $\Delta \Delta G_{\text{DNA:LNA}}^o$ represent the change in the Gibbs energy of melting (relative to that of the isosequential pure-DNA duplex) due to a single LNA substitution in the sense or anti-sense strand, respectively. $\Delta \Delta G_{\text{hyper-}(0)}^o$ then quantifies excess contributions to the stability enhancement provided by the LNA:LNA base pair that are not captured by the sum of the incremental stabilities of the parent LNA:DNA base pairs from which it is derived.

Analyzing the melting thermodynamics data for the learning set, we find that $\Delta \Delta G_{\text{hyper-}(0)}^o = 0.20 \pm 0.35$ kcal mol$^{-1}$, and that its value, within experimental error, is not sensitive to either the type of LNA:LNA base pair formed ($A:T$ or $C:G$) or the type of DNA:DNA base pairs neighboring it. Thus, though its error is somewhat larger than desired, $\Delta \Delta G_{\text{hyper-}(0)}^o$ is small, positive and constant in value, indicating that the incremental stability of an LNA:LNA base pair exceeds that predicted by the summed incremental stabilities of the parent LNA:DNA base pairs.

A representative example of this effect is provided in Table 3-3, which reports melting thermodynamic data for a set of duplexes having the pure-DNA 12-mer sequence 5’-ggaacaagatgc-3’ and varying LNA content. This pure-DNA duplex, hereafter identified as duplex 10, melts at 54.6 °C (@ $C_T = 5$ µM). From it, we created three duplexes containing a single LNA substitution in the sense strand; when melted at the
same $C_T$, each has an enhanced stability relative to the isosequential pure-DNA duplex as reflected in the positive $\Delta T_m$ value reported. Likewise, the 3 duplexes having a single LNA in the anti-sense strand all exhibit a positive $\Delta T_m$. Now consider the duplex 10-S-L6:AS-L7 containing a single LNA:LNA base pair that is formed from strands S-L6 (where “S” indicates the sense strand, and L6 indicates an LNA substitution at the 5’+6 position of that strand) and AS-L7 (where “AS” indicates the antisense strand). The $T_m$ of this duplex is 62.3 °C, giving a $\Delta T_m$ of 7.7 °C relative to the pure-DNA duplex. Application of equation (3-7) then yields a $\Delta \Delta T_m$ of ~2 °C (due to the fact that the sum $\Delta T_m,10-S-L6:AS + \Delta T_m,10-S:AS-L7 = 5.7$ °C), illustrating the hyperstabilizing effect of an LNA:LNA base pair.
Table 3-3 Melting thermodynamic data collected by UV spectroscopy for a set of duplexes having the pure-DNA sequence 5’-ggaacaagatgc-3’ and varying LNA content.

<table>
<thead>
<tr>
<th>DNA Duplex</th>
<th>$T_m$(°C)</th>
<th>$\Delta T_m$(°C)</th>
<th>$\Delta T_m$(°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-S-L7:AS</td>
<td>58.1</td>
<td>3.5</td>
<td>10-S-L6:AS</td>
</tr>
<tr>
<td>10-S-L9:AS</td>
<td>57.2</td>
<td>2.6</td>
<td>10-S:AS-L6</td>
</tr>
<tr>
<td>10-S:AS-L5</td>
<td>59.6</td>
<td>5.0</td>
<td>10-S:AS-L7</td>
</tr>
</tbody>
</table>

DNA Duplexes with Single LNA Substitutions on a Single Strand

<table>
<thead>
<tr>
<th>DNA Duplex</th>
<th>$T_m$(°C)</th>
<th>$\Delta T_m$(°C)</th>
<th>$\Delta T_m$(°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-S-L7:AS</td>
<td>61.4</td>
<td>6.8</td>
<td>10-S:AS-L6</td>
</tr>
<tr>
<td>10-S-L9:AS</td>
<td>64.5</td>
<td>10.0</td>
<td>10-S:AS-L7</td>
</tr>
<tr>
<td>10-S-L7:AS-L5</td>
<td>61.9</td>
<td>7.3</td>
<td>10-S-L6:AS-L6</td>
</tr>
</tbody>
</table>

DNA Duplexes with Multiple LNA Substitutions on Opposing Strands

<table>
<thead>
<tr>
<th>DNA Duplex</th>
<th>$T_m$(°C)</th>
<th>$\Delta T_m$(°C)</th>
<th>$\Delta T_m$(°C)</th>
</tr>
</thead>
</table>

114
3.4.4 5’-OFFSET OPPOSITELY ORIENTED LNA:DNA BASE PAIRS ALSO HYPERSTABILIZE AN LNA-SUBSTITUTED DUPLEX

Inclusion in the library of duplexes in which an LNA substitution on the sense strand is either 5’ or 3’ offset relative to an LNA on the anti-sense strand permitted measurement of $\Delta T_m$ and $\Delta \Delta T_m$ values for duplexes bearing oppositely oriented LNA:DNA base pairs of type $j$. **Figure 3-3** shows that hyperstabilization of duplexes containing a single LNA substitution on each strand occurs for certain arrangements of oppositely oriented LNA:DNA base pairs, such that $\Delta T_{m,LNA} = T_{m,LNA} - T_{m,DNA}$ exceeds $\Delta T_{m,LNA:DNA} + \Delta T_{m,DNA:LNA}$ by an amount $\Delta \Delta T_{m,LNA}$ given by equation (3-7). In particular, statistically significant hyperstabilization of duplexes is observed when the LNA on the anti-sense strand lies in a position $-3 \leq j \leq -1$. The average magnitude of the observed hyperstabilization depends on the separation distance between the two oppositely oriented LNA:DNA base pairs, with a maximum stability enhancement observed when the base pairs are 5’ offset and either directly adjacent to one another ($j = -1$), as illustrated by the $\Delta \Delta T_m$ values for duplexes 10-S-L7:AS-L5 and 10-S-L6:AS-L6 (**Table 3-3**), or separated by a single DNA:DNA base pair ($j = -2$).
Figure 3-3 Hyperstabilization of duplexes containing a single LNA substitution on each strand defined in terms of $\Delta \Delta T_{m,LNA}$ (see equation (3-20) below); the value of index $j$ is as defined in Figure 3-2. A $\Delta \Delta T_{m,LNA}$ positive in value indicates that the specific spacing of oppositely oriented LNA:DNA base pairs hyperstabilizes the duplex relative to the sum of the incremental stability enhancements provided by each LNA:DNA base pair alone.

Interestingly, we find conversely that 3’ off-setting of oppositely oriented LNA-DNA base pairs is slightly destabilizing (relative to the summed incremental stabilities of the parent LNA:DNA base pairs) for the specific case of $j = 2$.

Our proposed extension (i.e. equation (3-13)) to the SBT model assumes that the total hyperstabilization of a duplex bearing multiple LNA substitutions in each strand may be computed as a frequency weighted sum of $\Delta \Delta G_{hyper-(j)}^0$; in turn, we will later show that $\Delta \Delta T_{m,j}$ for a duplex bearing a pair of LNAs on each strand of type $j$ orientation is linearly proportional to $\Delta \Delta G_{hyper-(j)}^0$. As a result, the measured total $\Delta \Delta T_m$ for a duplex bearing multiple LNAs on each strand is equal to the sum of the frequency weighted $\Delta \Delta T_{m,j}$.
\[ \Delta \Delta T_m = \sum n_j \Delta \Delta T_{m,j} \]  

(3-20)

This is illustrated for both the 10-S-L79:AS-L5 duplex, which bears oppositely oriented LNA:DNA base pairs of type \( j = 1 \) and \( j = -1 \), and the 10-S-L6:AS-L67 duplex, which bears a single \( j = 0 \) and a single \( j = -1 \) type LNA-substitution pattern. For duplex 10-S-L79:AS-L5, \( \Delta \Delta T_m = 1.2 \) °C, which matches the sum of \( \Delta \Delta T_{m,j} \) (\( = 1.5 \) °C + (-0.3 °C)). Similar support for equation (3-20) is provided by the 10-S-L6:AS-L67 duplex.

### 3.4.5 EXTENDING THE SBT MODEL TO PERMIT APPLICATION TO DUPLEXES CONTAINING LNA SUBSTITUTIONS ON BOTH STRANDS

The illustrative results reported above indicate that the standard SBT model may be extended as defined in equations (3-10) to (3-13) to model duplexes containing combinations of LNA:LNA base pairs and offset LNA:DNA base pairs. Extension of the model requires a set of \( \Delta \Delta G_{\text{hyper-}(j)}^0 \) parameters. Those parameters were determined by globally fitting equations (3-10) to (3-13) to the collective melting thermodynamic data for the “learning set” (Table 3-2). The regressed \( \Delta \Delta G_{\text{hyper-}(j)}^0 \) are reported in Table 3-4 along with their respective standard deviations. In accordance with Figure 3-3 statistically significant \( \Delta \Delta G_{\text{hyper-}(j)}^0 \) parameters were obtained for LNA substitution patterns on opposite strands corresponding to \( j = -3 \) to -1, as well as to \( j = 2 \) (small destabilizing effect). As observed with the \( \Delta \Delta G_{\text{hyper-}(0)}^0 \) parameter, we find that the value of each \( \Delta \Delta G_{\text{hyper-}(j)}^0 \) is insensitive to both the types of LNAs substituted and the types of neighboring base pairs, at least within the error of our experimental data. In addition, and more importantly, we indeed find that the total hyperstabilization of a duplex containing multiple substitution patterns in each strand is given by the linear addition of frequency-weighted \( \Delta \Delta G_{\text{hyper-}(j)}^0 \), allowing the classic group-contribution method embodied in equation (3-13) to be applied to duplexes having dual-strand patterns of LNA substitutions. Two exceptions are noted. The first is that no
hyperstabilization effect is applied to any LNA substitutions at either terminus, as both our lab[192] and others[212, 213] have demonstrated that LNA substitutions made to either the 5’ or 3’ termini provide little to no additional stability to the duplex. The second applies to internal tandem (directly adjacent) LNA:LNA base pairs, which have not been investigated here and for which the extended SBT model therefore cannot be applied. Preliminary melting data acquired in our laboratory suggest that the tandem LNA:LNA motif strongly hyperstabilizes a duplex, but the precise magnitude and nature of this effect remains an open question that is currently under study.

**Table 3-4** $\Delta \Delta G_{hyper-\,(j)}^0$ parameters and their respective standard deviations when regressed to either the “learning set” (Table 3-2) alone, or the combined “learning set” and “testing set” (Table 3-2 and Table 3-5).

<table>
<thead>
<tr>
<th>LNA Substitution Pattern</th>
<th>$\Delta \Delta G_{hyper-,(j)}^0$ (kcal mol$^{-1}$)</th>
<th>Learning Set</th>
<th>Whole Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$j$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-4</td>
<td>0.01 ± 0.19</td>
<td>0.10 ± 0.25</td>
<td></td>
</tr>
<tr>
<td>-3</td>
<td>0.32 ± 0.25</td>
<td>0.35 ± 0.22</td>
<td></td>
</tr>
<tr>
<td>-2</td>
<td>0.67 ± 0.25</td>
<td>0.64 ± 0.19</td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td>0.54 ± 0.25</td>
<td>0.54 ± 0.20</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.20 ± 0.35</td>
<td>0.22 ± 0.32</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-0.11 ± 0.23</td>
<td>-0.06 ± 0.24</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.22 ± 0.21</td>
<td>-0.23 ± 0.19</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.18 ± 0.36</td>
<td>-0.12 ± 0.35</td>
<td></td>
</tr>
</tbody>
</table>
Table 3-5 reports melting thermodynamics collected by UV spectroscopy for a “testing set” of complementary duplexes bearing various levels and patterns of LNA substitutions. In general, users of models designed to predict melting thermodynamics of either pure-DNA duplexes or LNA-substituted duplexes are primarily interested in the prediction of $T_m$, as well as $\Delta T_m$ and $\Delta\Delta T_m$ to a lesser extent, because that information is essential to the design of primers, probes, and gene-silencing agents. Thus, for each duplex within the testing set, $T_m$, $\Delta T_m$ and $\Delta\Delta T_m$ values predicted by the extended SBT model are also reported (in parentheses) in Table 3-5.

Table 3-6 then summarizes the overall melting-temperature prediction performance of the extended SBT model when sequentially applied to the learning set, testing set, and full set of duplex sequences reported in Table 3-2 and Table 3-5. As one might expect, excellent performance is observed when the model is applied to the “learning set” sequences whose melting thermodynamics were collectively used to regress extended model parameters. The NNT model,[136] upon which both the standard and extended SBT model are based, predicts $T_m$ values for pure-DNA duplexes with a mean error ± standard deviation of –0.2 ± 1.4 °C. The fact that a comparable overall error is found for the extended SBT model in predicting $T_m$ values for duplexes containing various LNA substitution patterns suggests that the proposed group-contribution approach to modeling the stabilizing and hyperstabilizing effects of LNAs is not only sound, but highly accurate. This is particularly true of the hyperstabilization effects, which were the central concern in this work; indeed, the overall error in predicted $\Delta\Delta T_m$ values (see equation (3-7)) is 0.0 ± 1.0 °C when applied to the learning set.
Table 3-5 Melting thermodynamic data collected by UV spectroscopy for the “testing set” of complementary duplexes bearing an LNA substitution on one or both strands.

<table>
<thead>
<tr>
<th>Duplex Name</th>
<th>Sense Strand</th>
<th>Anti-Sense Strand</th>
<th>ΔH [kcal mol⁻¹]</th>
<th>ΔS [cal mol⁻¹ K⁻¹]</th>
<th>ΔG° [kcal mol⁻¹]</th>
<th>ΔΔG° [kcal mol⁻¹]</th>
<th>Tm [°C]</th>
<th>ΔTm [°C]</th>
<th>ΔΔTm [°C]</th>
</tr>
</thead>
<tbody>
<tr>
<td>8-S-L4:AS-L7</td>
<td>ggaCctcgac</td>
<td>gtcgagGtcc</td>
<td>79.5 ± 2.1</td>
<td>207.7 ± 5.9</td>
<td>11.70 ± 0.14</td>
<td>2.85 ± 0.22</td>
<td>65.7 ± 0.3</td>
<td>12.6 ± 0.8</td>
<td>1.2</td>
</tr>
<tr>
<td>8-S-L5:AS-L6</td>
<td>ggacCtcgac</td>
<td>gtcgaGgtcc</td>
<td>75.4 ± 1.0</td>
<td>195.1 ± 2.7</td>
<td>11.61 ± 0.07</td>
<td>2.77 ± 0.19</td>
<td>66.1 ± 0.2</td>
<td>12.9 ± 0.8</td>
<td>1.4</td>
</tr>
<tr>
<td>8-S-L5:AS-L5</td>
<td>ggacCtcgac</td>
<td>gtcgAggtcc</td>
<td>77.8 ± 2.5</td>
<td>202.2 ± 7.3</td>
<td>11.71 ± 0.15</td>
<td>2.86 ± 0.23</td>
<td>66.1 ± 0.2</td>
<td>13 ± 0.8</td>
<td>2.0</td>
</tr>
<tr>
<td>8-S-L4:AS-L6</td>
<td>ggaCctcgac</td>
<td>gtcgaGgtcc</td>
<td>76.7 ± 4.2</td>
<td>197.8 ± 12.0</td>
<td>12.00 ± 0.25</td>
<td>3.15 ± 0.30</td>
<td>67.8 ± 0.4</td>
<td>14.6 ± 0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>8-S-L7:AS-L5</td>
<td>ggacCtcgac</td>
<td>gtcgaGgtcc</td>
<td>84.0 ± 0.8</td>
<td>221.6 ± 2.5</td>
<td>11.66 ± 0.03</td>
<td>2.81 ± 0.18</td>
<td>64.8 ± 0.2</td>
<td>11.6 ± 0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>8-S-L5:AS-L7</td>
<td>ggacCtcgac</td>
<td>gtcgagGtcc</td>
<td>79.2 ± 4.1</td>
<td>210.0 ± 11.9</td>
<td>10.71 ± 0.16</td>
<td>1.87 ± 0.24</td>
<td>61.2 ± 0.3</td>
<td>8.0 ± 0.8</td>
<td>-1.2</td>
</tr>
<tr>
<td>9-S-L7:AS-L4</td>
<td>cctgcgAtgac</td>
<td>gtcAtcgccagg</td>
<td>88.7 ± 1.3</td>
<td>236.4 ± 3.6</td>
<td>11.55 ± 0.08</td>
<td>1.66 ± 0.12</td>
<td>63.7 ± 0.2</td>
<td>6.3 ± 0.3</td>
<td>2.1</td>
</tr>
<tr>
<td>9-S-L6:AS-L5</td>
<td>cctgcGatgac</td>
<td>gtcatcgccagg</td>
<td>85.5 ± 0.8</td>
<td>225.1 ± 2.2</td>
<td>11.92 ± 0.05</td>
<td>2.03 ± 0.10</td>
<td>65.8 ± 0.1</td>
<td>8.4 ± 0.2</td>
<td>2.1</td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>$\Delta H$ (kcal mol$^{-1}$)</td>
<td>$\Delta S$ (cal mol$^{-1}$ K$^{-1}$)</td>
<td>$\Delta G^\circ$ (kcal mol$^{-1}$)</td>
<td>$\Delta G^\circ$ (kcal mol$^{-1}$)</td>
<td>$T_m$ (°C)</td>
<td>$\Delta T_m$ (°C)</td>
<td>$\Delta T_m$ (°C)</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>------------------</td>
<td>-----------------------------</td>
<td>---------------------------------</td>
<td>-------------------------------</td>
<td>-------------------------------</td>
<td>----------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>9-S-L8:AS-L3</td>
<td>cctgcgaTgac</td>
<td>gtCatcgcaagg</td>
<td>96.5 ± 1.8</td>
<td>254.7 ± 5.3</td>
<td>13.21 ± 0.06</td>
<td>3.31 ± 0.11</td>
<td>69.3 ± 0.1</td>
<td>11.8 ± 0.2</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(12.45)</td>
<td>(67.0)</td>
<td>(10.3)</td>
<td>(2.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-S-L4:AS-L7</td>
<td>cctGcgcgatgc</td>
<td>gtcatGCcagg</td>
<td>85.3 ± 0.4</td>
<td>227.3 ± 1.2</td>
<td>11.14 ± 0.04</td>
<td>1.25 ± 0.10</td>
<td>62.4 ± 0.1</td>
<td>5.0 ± 0.3</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(12.07)</td>
<td>(65.5)</td>
<td>(8.8)</td>
<td>(2.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-S-L5:AS-L6</td>
<td>cctgCgatgc</td>
<td>gtcatCgcagg</td>
<td>85.8 ± 0.8</td>
<td>223.6 ± 2.4</td>
<td>12.66 ± 0.05</td>
<td>2.77 ± 0.11</td>
<td>69.1 ± 0.1</td>
<td>11.7 ± 0.3</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(12.73)</td>
<td>(68.0)</td>
<td>(11.3)</td>
<td>(2.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-S-L3:AS-L7</td>
<td>ccTgcgatgc</td>
<td>gtcateGcagg</td>
<td>88.2 ± 3.8</td>
<td>234.3 ± 11.3</td>
<td>11.67 ± 0.11</td>
<td>1.78 ± 0.14</td>
<td>64.3 ± 0.1</td>
<td>6.9 ± 0.2</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(12.22)</td>
<td>(66.1)</td>
<td>(9.4)</td>
<td>(2.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-S-L4:AS-L6</td>
<td>ccTGcgcgatgc</td>
<td>gtcatCgcagg</td>
<td>88.9 ± 1.0</td>
<td>235.7 ± 3.0</td>
<td>11.94 ± 0.05</td>
<td>2.05 ± 0.10</td>
<td>65.3 ± 0.1</td>
<td>7.9 ± 0.2</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(12.50)</td>
<td>(67.1)</td>
<td>(10.4)</td>
<td>(2.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-S-L5:AS-L5</td>
<td>cctgCgatgc</td>
<td>gtcATcgcagg</td>
<td>88.7 ± 0.2</td>
<td>233.1 ± 0.7</td>
<td>12.52 ± 0.02</td>
<td>2.62 ± 0.09</td>
<td>67.8 ± 0.0</td>
<td>10.4 ± 0.2</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(12.55)</td>
<td>(67.3)</td>
<td>(10.7)</td>
<td>(2.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-S-L6:AS-L4</td>
<td>cctgcGatgc</td>
<td>gtcATcgcaagg</td>
<td>95.2 ± 1.8</td>
<td>255.1 ± 5.3</td>
<td>11.86 ± 0.07</td>
<td>1.97 ± 0.11</td>
<td>64.1 ± 0.0</td>
<td>6.7 ± 0.2</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(11.91)</td>
<td>(64.8)</td>
<td>(8.2)</td>
<td>(2.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-S-L5:AS-L4</td>
<td>cctgCgatgc</td>
<td>gtcATcgcaagg</td>
<td>89.1 ± 0.8</td>
<td>236.4 ± 2.3</td>
<td>11.87 ± 0.04</td>
<td>1.98 ± 0.10</td>
<td>65.0 ± 0.1</td>
<td>7.6 ± 0.2</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(11.94)</td>
<td>(65.0)</td>
<td>(8.4)</td>
<td>(1.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-S-L4:AS-L5</td>
<td>cctGcgcgatgc</td>
<td>gtcATcgcaagg</td>
<td>95.0 ± 1.8</td>
<td>255.2 ± 5.3</td>
<td>11.69 ± 0.05</td>
<td>1.80 ± 0.10</td>
<td>63.5 ± 0.1</td>
<td>6.1 ± 0.2</td>
<td>1.1</td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>ΔH (kcal mol⁻¹)</td>
<td>ΔS (cal mol⁻¹ K⁻¹)</td>
<td>ΔG° (kcal mol⁻¹)</td>
<td>ΔΔG° (kcal mol⁻¹)</td>
<td>Tm (°C)</td>
<td>ΔTm (°C)</td>
<td>ΔΔTm (°C)</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>------------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>--------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>9-S-L6:AS-L3</td>
<td>ccTgcgatgac</td>
<td>gtCategcagg</td>
<td>95.9 ± 1.8</td>
<td>254.9 ± 5.3</td>
<td>12.66 ± 0.08</td>
<td>2.77 ± 0.12</td>
<td>67.1 ± 0.0</td>
<td>9.7 ± 0.2</td>
<td>2.5</td>
</tr>
<tr>
<td>9-S-L3:AS-L6</td>
<td>ccTgcgatgac</td>
<td>gtCatesgcagg</td>
<td>88.5 ± 3.8</td>
<td>234.1 ± 11.3</td>
<td>12.06 ± 0.14</td>
<td>2.17 ± 0.17</td>
<td>65.9 ± 0.0</td>
<td>8.5 ± 0.2</td>
<td>0.8</td>
</tr>
<tr>
<td>10-S-L7:AS-L6</td>
<td>ggaacaAgatgc</td>
<td>gcateTgttcc</td>
<td>91.2 ± 4.4</td>
<td>245.9 ± 13.0</td>
<td>10.91 ± 0.16</td>
<td>1.68 ± 0.20</td>
<td>60.9 ± 0.3</td>
<td>6.3 ± 0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>10-S-L6:AS-L7</td>
<td>ggaacaAagatgc</td>
<td>gcateTgttcc</td>
<td>89.8 ± 3.4</td>
<td>240.8 ± 10.1</td>
<td>11.22 ± 0.12</td>
<td>1.99 ± 0.17</td>
<td>62.3 ± 0.1</td>
<td>7.7 ± 0.5</td>
<td>1.9</td>
</tr>
<tr>
<td>10-S-L7:AS-L5</td>
<td>ggaacaAgatgc</td>
<td>gcateCtbtgccc</td>
<td>94.9 ± 1.5</td>
<td>253.9 ± 4.3</td>
<td>11.95 ± 0.07</td>
<td>2.72 ± 0.14</td>
<td>64.5 ± 0.1</td>
<td>10.0 ± 0.5</td>
<td>1.5</td>
</tr>
<tr>
<td>10-S-L6:AS-L6</td>
<td>ggaacaAagatgc</td>
<td>gcateTgttcc</td>
<td>93.6 ± 2.3</td>
<td>253.1 ± 6.7</td>
<td>11.00 ± 0.08</td>
<td>1.77 ± 0.14</td>
<td>61.0 ± 0.1</td>
<td>6.4 ± 0.5</td>
<td>2.3</td>
</tr>
<tr>
<td>10-S-L4:AS-L8</td>
<td>ggaAcaagatgc</td>
<td>gcatectGtccc</td>
<td>97.1 ± 1.2</td>
<td>262.6 ± 3.6</td>
<td>11.42 ± 0.04</td>
<td>2.19 ± 0.12</td>
<td>62.2 ± 0.1</td>
<td>7.7 ± 0.5</td>
<td>2.3</td>
</tr>
<tr>
<td>10-S-L4:AS-L7</td>
<td>ggaAcaagatgc</td>
<td>gcateTgttcc</td>
<td>91.1 ± 1.5</td>
<td>244.0 ± 4.3</td>
<td>11.44 ± 0.06</td>
<td>2.21 ± 0.13</td>
<td>63.0 ± 0.1</td>
<td>8.4 ± 0.5</td>
<td>2.3</td>
</tr>
<tr>
<td>10-S-L6:AS-L5</td>
<td>ggaacAagatgc</td>
<td>gcateCtbtgccc</td>
<td>90.2 ± 3.1</td>
<td>241.3 ± 9.2</td>
<td>11.46 ± 0.13</td>
<td>2.23 ± 0.17</td>
<td>63.2 ± 0.1</td>
<td>8.6 ± 0.5</td>
<td>2.2</td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>ΔH</td>
<td>ΔS</td>
<td>ΔG°</td>
<td>ΔΔG°</td>
<td>Tm</td>
<td>ΔTm</td>
<td>ΔΔTm</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------</td>
<td>-------------------</td>
<td>---------</td>
<td>----------</td>
<td>---------</td>
<td>--------</td>
<td>---------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>10-S-L4:AS-L6</td>
<td>ggaAcaagatgc</td>
<td>gcacttGtgttcc</td>
<td>89.9 ± 1.9</td>
<td>242.8 ± 5.6</td>
<td>10.71 ± 0.05</td>
<td>1.48 ± 0.13</td>
<td>60.2 ± 0.0</td>
<td>5.6 ± 0.5</td>
<td>1.2</td>
</tr>
<tr>
<td>10-S-L4:AS-L5</td>
<td>ggaAcaagatgc</td>
<td>gcacttGtgttcc</td>
<td>94.5 ± 1.2</td>
<td>254.3 ± 3.7</td>
<td>11.47 ± 0.05</td>
<td>2.24 ± 0.13</td>
<td>62.7 ± 0.1</td>
<td>8.2 ± 0.5</td>
<td>1.4</td>
</tr>
<tr>
<td>10-S-L6:AS-L8</td>
<td>ggaAcAagatgc</td>
<td>gcacttGtgttcc</td>
<td>91.5 ± 4.9</td>
<td>246.9 ± 14.3</td>
<td>10.91 ± 0.18</td>
<td>1.68 ± 0.22</td>
<td>60.8 ± 0.3</td>
<td>6.3 ± 0.5</td>
<td>1.2</td>
</tr>
<tr>
<td>10-S-L7:AS-L7</td>
<td>ggaacaAgatgc</td>
<td>gcacttGtgttcc</td>
<td>93.6 ± 1.7</td>
<td>251.3 ± 5.2</td>
<td>11.55 ± 0.06</td>
<td>2.32 ± 0.13</td>
<td>63.1 ± 0.1</td>
<td>8.6 ± 0.5</td>
<td>0.7</td>
</tr>
<tr>
<td>10-S-L9:AS-L5</td>
<td>ggaacaagAtgc</td>
<td>gcacttGtgttcc</td>
<td>87.3 ± 3.5</td>
<td>233.4 ± 10.3</td>
<td>11.06 ± 0.12</td>
<td>1.83 ± 0.17</td>
<td>61.9 ± 0.1</td>
<td>7.3 ± 0.5</td>
<td>-0.3</td>
</tr>
<tr>
<td>10-S-L9:AS-L6</td>
<td>ggaacaagAtgc</td>
<td>gcacttGtgttcc</td>
<td>92.2 ± 1.6</td>
<td>250.8 ± 4.6</td>
<td>10.43 ± 0.07</td>
<td>1.20 ± 0.14</td>
<td>58.9 ± 0.2</td>
<td>4.4 ± 0.5</td>
<td>-0.9</td>
</tr>
<tr>
<td>10-S-L7:AS-L8</td>
<td>ggaacaAgatgc</td>
<td>gcacttGtgttcc</td>
<td>96.9 ± 1.8</td>
<td>263.6 ± 5.5</td>
<td>10.94 ± 0.05</td>
<td>1.71 ± 0.13</td>
<td>60.5 ± 0.1</td>
<td>5.9 ± 0.5</td>
<td>-1.2</td>
</tr>
<tr>
<td>10-S-L9:AS-L7</td>
<td>ggaacaagAtgc</td>
<td>gcacttGtgttcc</td>
<td>93.3 ± 1.3</td>
<td>251.3 ± 3.7</td>
<td>11.26 ± 0.06</td>
<td>2.03 ± 0.13</td>
<td>62.0 ± 0.1</td>
<td>7.5 ± 0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>11-S-L12:AS-L4</td>
<td>aegaccagagcTacag</td>
<td>ctgTaacttGtgttgc</td>
<td>115.1 ± 1.8</td>
<td>304.2 ± 5.1</td>
<td>15.41 ± 0.10</td>
<td>1.87 ± 0.15</td>
<td>74.3 ± 0.0</td>
<td>6.9 ± 0.1</td>
<td>1.2</td>
</tr>
<tr>
<td>Duplex Name</td>
<td>Sense Strand</td>
<td>Anti-Sense Strand</td>
<td>$\Delta H$ (kcal mol$^{-1}$)</td>
<td>$\Delta S$ (cal mol$^{-1}$ K$^{-1}$)</td>
<td>$\Delta G^\circ$ (kcal mol$^{-1}$)</td>
<td>$\Delta G^\circ$ (kcal mol$^{-1}$)</td>
<td>$T_m$ (°C)</td>
<td>$\Delta T_m$ (°C)</td>
<td>$\Delta T_m$ (°C)</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>-------------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>11-S-L11:AS-L4</td>
<td>aegaccagTtaacagt</td>
<td>ctgtaactctggtcgt</td>
<td>113.8 ± 0.9</td>
<td>301.3 ± 2.7</td>
<td>15.13 ± 0.05</td>
<td>1.60 ± 0.12</td>
<td>73.5 ± 0.1</td>
<td>6.2 ± 0.1</td>
<td>1.0</td>
</tr>
<tr>
<td>11-S-L9:AS-L6</td>
<td>acgaccAggttaacagt</td>
<td>ctgtAactctggtcgt</td>
<td>110.6 ± 0.5</td>
<td>290.5 ± 1.3</td>
<td>15.36 ± 0.04</td>
<td>1.82 ± 0.11</td>
<td>75.2 ± 0.1</td>
<td>7.8 ± 0.1</td>
<td>1.2</td>
</tr>
<tr>
<td>11-S-L3:AS-L12</td>
<td>acGaccaggttaacagt</td>
<td>ctgtaactctggtcgt</td>
<td>114.7 ± 0.6</td>
<td>304.8 ± 1.7</td>
<td>14.90 ± 0.03</td>
<td>1.36 ± 0.11</td>
<td>72.5 ± 0.1</td>
<td>5.2 ± 0.1</td>
<td>1.5</td>
</tr>
<tr>
<td>11-S-L6:AS-L9</td>
<td>acgaccCaggttaacagt</td>
<td>ctgtaactCtggtcgt</td>
<td>112.1 ± 0.7</td>
<td>294.6 ± 2.0</td>
<td>15.53 ± 0.03</td>
<td>2.00 ± 0.11</td>
<td>75.5 ± 0.1</td>
<td>8.1 ± 0.1</td>
<td>1.5</td>
</tr>
<tr>
<td>11-S-L9:AS-L4</td>
<td>acgaccAggttaacagt</td>
<td>ctgtAactctggtcgt</td>
<td>110.6 ± 0.5</td>
<td>291.2 ± 1.6</td>
<td>15.19 ± 0.03</td>
<td>1.65 ± 0.11</td>
<td>74.5 ± 0.1</td>
<td>7.1 ± 0.1</td>
<td>1.7</td>
</tr>
<tr>
<td>11-S-L3:AS-L11</td>
<td>acGaccaggttaacagt</td>
<td>ctgtaactctggtcgt</td>
<td>117.7 ± 1.5</td>
<td>312.8 ± 4.2</td>
<td>15.30 ± 0.08</td>
<td>1.77 ± 0.14</td>
<td>73.3 ± 0.0</td>
<td>6.0 ± 0.1</td>
<td>1.6</td>
</tr>
<tr>
<td>11-S-L5:AS-L9</td>
<td>acgacCaggttaacagt</td>
<td>ctgtaactCtggtcgt</td>
<td>112.2 ± 1.3</td>
<td>294.4 ± 3.8</td>
<td>15.62 ± 0.09</td>
<td>2.08 ± 0.14</td>
<td>75.8 ± 0.1</td>
<td>8.4 ± 0.1</td>
<td>0.9</td>
</tr>
</tbody>
</table>
Values reported in parenthesis are the corresponding melting thermodynamics predicted by the extended SBT model. The reported UVM data ($\Delta H(T_m)$, $\Delta S(T_m)$, $\Delta G^\circ$, and $T_m$) are for the helix-to-coil transition and were determined as reported in Materials and Methods. The reported errors refer to the standard deviation of triplicate runs. All duplex samples were resuspended in buffer containing 1 M NaCl, 10 mM Na$_2$HPO$_4$, and 1 mM Na$_2$EDTA (pH=7.0) to a $C_T = 5 \mu$M.
Extended SBT model performance remains very good when applied to the prediction of $T_m$, $\Delta T_m$ and $\Delta \Delta T_m$ values for all “testing set” duplexes containing LNAs in both strands, with the model again showing a near zero mean error in $\Delta \Delta T_m$. Of greatest importance here is the fact that through use of the group-contribution approach embodied in equation (3-13), the extended SBT model accurately captures all hyperstabilization effects, which can be significant. For the 10-S-L6:AS-L67 duplex (Table 3-1), for example, hyperstabilization effects contribute 3.7 °C (31%) of the 12.1 °C increase in duplex $T_m$ arising from the LNA substitutions.

**Table 3-6** Summary of overall prediction performance of the extended SBT model (“learning set” regressed $\Delta \Delta G^\circ_{\text{hyper}-(j)}$ parameters) when sequentially applied to the learning set, testing set, and combined set of duplex sequences reported in Table 3-2 and Table 3-5.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>$T_m$ (°C)</th>
<th>$\Delta T_m$ (°C)</th>
<th>$\Delta \Delta T_m$ (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning</td>
<td>-0.7 ± 1.3</td>
<td>-0.3 ± 1.3</td>
<td>0.0 ± 1.0</td>
</tr>
<tr>
<td>Testing</td>
<td>-0.9 ± 1.9</td>
<td>-0.4 ± 1.8</td>
<td>-0.2 ± 0.7</td>
</tr>
<tr>
<td>Overall</td>
<td>-0.8 ± 1.5</td>
<td>-0.3 ± 1.5</td>
<td>-0.1 ± 0.9</td>
</tr>
</tbody>
</table>

The stability of the hyperstabilization parameters was then evaluated through their re-regression to the entire data set (*i.e.* “learning set” + “testing set”), and the resulting whole-dataset derived parameters are also reported in Table 3-4. For those parameters that are statistically significant (-3 ≤ $j$ ≤ -1, $j = 2$), the two sets of regressed values are essentially identical, indicating a lack of end-effect or other error sources in the “learning set” that might serve to bias parameter estimates. Moreover, for either data set, $\Delta \Delta T_m$ is found to depend linearly on $\Delta \Delta G^\circ_{\text{hyper}}$, as required by equation (3-10) and as shown in Figure 3-4 for the whole-dataset derived parameters.
Figure 3-4  $\Delta \Delta T_m$ depends linearly on $\Delta \Delta G^o_{hyper-(f)}$ (whole data set derived model parameters in Table 3-4) as required by equation (3-10).

As a final test, we applied the extended SBT model (w/ whole-dataset derived parameters) to a small set of duplexes bearing more complex LNA-substitution patterns in both strands, with a comparison of predicted and experimental melting thermodynamics provided in Table 3-7. The results show that the extended SBT model predicts well melting thermodynamics for duplexes bearing more complex combinations of both LNA:LNA and oppositely oriented LNA:DNA base pairs, exhibiting errors of $-1.3 \pm 1.5 \degree C$, $-0.8 \pm 1.4 \degree C$, and $-0.3 \pm 0.7 \degree C$ for $T_m$, $\Delta T_m$ and $\Delta \Delta T_m$, respectively. At present, no other model is designed or able to make such predictions. As an extreme test of the model, we also used the model to predict $\Delta \Delta T_m$ for a highly substituted duplex, of sequence S-L257:AS-L358 using our nomenclature, which Koshkin et al.[97] reported melts at a $T_m$ 34 °C higher than that for the corresponding isosequential pure-DNA duplex. The hyperstabilization effect, quantified as $\Delta \Delta T_m$, accounts for ca. 23% of that dramatic increase in $T_m$ and the extended SBT model predicts a $\Delta \Delta T_m$ of 7.6 °C, which is in reasonably good agreement with the measured values of 8 °C.
<table>
<thead>
<tr>
<th>Duplex Name</th>
<th>Sense Strand</th>
<th>Anti-Sense Strand</th>
<th>$T_m$ (°C)</th>
<th>$\Delta T_m$ (°C)</th>
<th>$\Delta\Delta T_m$ (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8-S-L57:AS-L5</td>
<td>ggacCtCgac</td>
<td>gtcgAggtcc</td>
<td>72.5 ± 0.4</td>
<td>19.3 ± 0.9</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(67.8)</td>
<td>(15.0)</td>
<td>(2.0)</td>
</tr>
<tr>
<td>8-S-L5:AS-L67</td>
<td>ggacCtCtagac</td>
<td>gtcgagGGtcc</td>
<td>68.9 ± 0.3</td>
<td>15.7 ± 0.8</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(66.2)</td>
<td>(13.4)</td>
<td>(0.7)</td>
</tr>
<tr>
<td>8-S-L4:AS-L67</td>
<td>ggacCtCtagac</td>
<td>gtcgagGGtcc</td>
<td>72.4 ± 0.3</td>
<td>19.2 ± 0.8</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(68.7)</td>
<td>(15.9)</td>
<td>(3.2)</td>
</tr>
<tr>
<td>8-S-L45:AS-L6</td>
<td>ggacCtCgac</td>
<td>gtcgagGGtcc</td>
<td>72.5 ± 1.0</td>
<td>19.3 ± 1.2</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(70.1)</td>
<td>(17.3)</td>
<td>(3.2)</td>
</tr>
<tr>
<td>8-S-L45:AS-L7</td>
<td>ggacCtCtagac</td>
<td>gtcgagGGtcc</td>
<td>67.5 ± 0.4</td>
<td>14.4 ± 0.9</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(67.6)</td>
<td>(14.8)</td>
<td>(0.7)</td>
</tr>
<tr>
<td>8-S-L57:AS-L67</td>
<td>ggacCtCgac</td>
<td>gtcgagGGtcc</td>
<td>74.8 ± 0.5</td>
<td>21.6 ± 0.9</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(70.0)</td>
<td>(17.2)</td>
<td>(-0.7)</td>
</tr>
<tr>
<td>10-S-L7:AS-L67</td>
<td>ggacaaAagtgc</td>
<td>gcacTTggtcc</td>
<td>66.5 ± 0.1</td>
<td>11.9 ± 0.5</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(64.5)</td>
<td>(9.5)</td>
<td>(0.6)</td>
</tr>
<tr>
<td>10-S-L6:AS-L67</td>
<td>ggacaaAagtgc</td>
<td>gcacTTggtcc</td>
<td>66.7 ± 0.0</td>
<td>12.1 ± 0.5</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(66.7)</td>
<td>(11.6)</td>
<td>(2.7)</td>
</tr>
<tr>
<td>10-S-L6:AS-L67</td>
<td>ggacAAgatgc</td>
<td>gcacTTggtcc</td>
<td>66.3 ± 0.1</td>
<td>11.7 ± 0.5</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(65.5)</td>
<td>(10.5)</td>
<td>(2.7)</td>
</tr>
<tr>
<td>10-S-L6:AS-L7</td>
<td>ggacAAgatgc</td>
<td>gcacTTggtcc</td>
<td>65.3 ± 0.1</td>
<td>10.8 ± 0.5</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(63.4)</td>
<td>(8.3)</td>
<td>(0.6)</td>
</tr>
<tr>
<td>10-S-L46:AS-L8</td>
<td>ggacAAgatgc</td>
<td>gcacTTggtcc</td>
<td>65.7 ± 0.1</td>
<td>11.1 ± 0.5</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(64.3)</td>
<td>(9.3)</td>
<td>(1.7)</td>
</tr>
<tr>
<td>10-S-L79:AS-L5</td>
<td>ggacaaAgtgc</td>
<td>gcacCttggtcc</td>
<td>67.6 ± 0.3</td>
<td>13.0 ± 0.5</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(65.5)</td>
<td>(10.5)</td>
<td>(1.7)</td>
</tr>
<tr>
<td>10-S-L79:AS-L67</td>
<td>ggacaaAgtgc</td>
<td>gcacTTggtcc</td>
<td>70.1 ± 0.2</td>
<td>15.5 ± 0.5</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(65.5)</td>
<td>(10.5)</td>
<td>(-0.7)</td>
</tr>
<tr>
<td>10-S-L79:AS-L8</td>
<td>ggacaaAgtgc</td>
<td>gcacTTggtcc</td>
<td>64.5 ± 0.0</td>
<td>9.9 ± 0.5</td>
<td>-0.5</td>
</tr>
</tbody>
</table>

Table 3-7 Performance of the extended SBT model when applied to a set of duplexes bearing various LNA-substitution patterns in both strands.
Melting thermodynamic data reported above were collected by UV spectroscopy at a $C_T = 5 \mu M$ in buffer containing 1 M NaCl, 10 mM Na$_2$HPO$_4$, and 1 mM Na$_2$EDTA (pH=7.0).

### 3.5 CONCLUDING REMARKS

The unique stabilization properties of locked nucleic acids are creating exciting opportunities to generate highly specific therapeutic aptamers exhibiting improved stability and half-life, as well as more potent cell-internalization characteristics, to improve real-time and digital PCR based detection of rare, clinically important gene mutations, and to create next-generation molecular beacons and related diagnostic probes that show unequivocal specificity for their target. However, fully exploiting LNA technology in these areas will require model-based methods for predicting the hybridization thermodynamics of duplexes bearing complex LNA substitution patterns. The nearest-neighbor SBT model we previously reported accurately predicts melting thermodynamics for both complementary[192] and mismatched[214] DNA duplexes containing any pattern of LNA substitutions in one strand.

Here, we have shown that a group-contribution modeling approach can be used to extend the SBT model to achieve accurate predictions of melting temperatures for short complementary duplex DNA containing various patterns of LNA substitutions within both strands. It accounts for both the intrinsic stability of an LNA-containing base pair, and hyperstabilization effects created by certain patterns of proximal LNA-containing base pairs. As the hyperstabilization effects are generally subtle, experiment does not
permit their associated enthalpic and entropic contributions to be defined in a statistically significant manner. However, for specific patterns of LNA:LNA and oppositely oriented LNA:DNA base pairs, the contribution of hyperstabilization effects can be and therefore is described at the Gibbs energy level. The resulting extended SBT model thereby achieves good prediction of $T_m$ and $\Delta T_m$ for duplexes bearing LNAs in both strands; since $\Delta \Delta T_m$ depends linearly on $\Delta \Delta G_{\text{hyper}}^0$, $\Delta \Delta G_{\text{LNA}}^0$ values are predicted with equivalent accuracy. $\Delta H$ and $\Delta S$ are, of course, not treated by the model, which is not of significant concern as models of this type are almost exclusively used by biologists and biotechnologists to predict $T_m$. It is important to note that all melting thermodynamics reported in this work were collected in buffer containing 1 M NaCl, 10mM Na$_2$HPO$_4$, and 1 mM Na$_2$EDTA at pH=7.0. All model predictions reported apply to that standard condition as well. However, excellent, highly accurate methods for converting thermodynamics to other solution conditions are available (see reference [210] for a review of these methods).

The development of the extended SBT model required the collection of complete melting thermodynamics data for a large set of pure-DNA and LNA-substituted duplexes redundantly covering all possible LNA:LNA base pairs and oppositely oriented LNA:DNA base pairs. Such data, particularly for the latter class of duplexes, have rarely appeared in the literature,[150] making this a valuable contribution that may serve to better understand the molecular nature of the stabilizing properties of LNAs. Evidence for this is provided by our use of the data set to uncover the general ability of specific patterns of LNA substitutions in opposite strands to hyperstabilize a duplex. At present, the model does not apply to all possible substitution patterns, as it does not address certain motifs such as tandem LNA:LNA base pairs and their hyperstabilizing effect. The study of those motifs is underway, but the results and model presented here constitute a novel and reliable approach to predicting $T_m$ and $\Delta T_m$ values for many duplexes for which effective therapeutic or diagnostic use depends on precise LNA substitution patterns and associated thermal stabilities.
CHAPTER 4: MODEL APPLICATION TO DEVELOPMENT OF A NOVEL, REAL-TIME PCR ASSAY AGAINST BRAF V600 MUTATIONS

Skin cancer is the most commonly diagnosed of all cancers. Melanoma is among the most aggressive, accounting for 80% of skin cancer deaths despite representing only ca. 2–4% (73,870 US, 2015) of all cases [215]. Globally, ca. 200,000 new cases of melanoma are diagnosed per annum and incidence rates are increasing [216]. Cutaneous melanomas are commonly diagnosed in the early stages of disease and are curable in most cases through surgical removal [217]; the 5-year survival rate of these early stage localized melanomas is ~98%. However, the rate of survival decreases dramatically to 63% and 16% for patients with local and distant metastases, respectively [216]. For later stage melanomas (i.e. unresectable stage III, metastatic stage IV, and recurrent melanoma), immunotherapy (e.g. ipilimumab), chemotherapy (e.g. dacarzabine), and targeted therapy are possible treatment options [218]. Notably, targeted therapies have shown remarkable response rates (~50%) and improved progression-free survival [219, 220]. However, only patients with certain genetic mutations are eligible for such treatment, thus motivating the need for patient specific tumour genotyping assays.

4.1 CLINICAL SIGNIFICANCE OF BRAF MUTATIONS IN MELANOMA

Protein phosphorylation is an important cellular mechanism that enables the transmission of a chemical signal within the cell. Those signalling pathways regulate basic cellular functions (i.e. cell survival, genome maintenance, cell fate), and abnormal protein phosphorylation within them can cause pathway dysregulation that, in certain cases, results in cancer [4, 221]. Somatic (acquired) mutations within the v-raf murine sarcoma viral oncogene homolog B (BRAF) protein kinase or its upstream effector protein kinase, KRAS (Kirsten rat sarcoma viral oncogene homolog), are found in many
cancers (e.g. colon, thyroid, lung, skin, etc.), where they can cause the dysregulated and sustained activation of the mitogen activated protein kinase (MAPK) pathway (RAS-RAF-MEK-ERK) that is normally responsible for the regulation of cell survival and proliferation [4, 6, 222]. Somatic mutations in protein kinases are among the most common mutations in cancer and have spurred development of a class of targeted therapeutics known as mutant-kinase inhibitors [221, 223]. As of 2013, twenty one kinase inhibitors have been approved for treating various cancers, with hundreds more in development [8]. Since only a fraction of patients with a particular cancer generally harbour a mutated protein kinase, the cancerous lesion must first be analyzed to determine if genome copies present within it encode a mutant kinase and, if so, that the patient is eligible for treatment with a kinase inhibitor approved for that indication [4]. This distinction is made because many different somatic mutations can occur in a (proto-)oncogene, but only a subset of those is typically eligible for or responsive to treatment with a targeted kinase inhibitor [4]. A clinically useful diagnostic assay must therefore determine the presence and genotype of a particular mutation, and preferably its mutation frequency as this often provides a useful metric of disease progression and thus, patient prognosis. As described in Chapter 1, several different platforms (next-generation sequencing, immunohistochemistry, PCR) are available for tumour genotyping, with each having particular features and benefits [12]. Here, I explore the potential to develop a new allele specific quantitative PCR (AS-qPCR) assay against \(BRAF\ V600\) mutations by using the molecular thermodynamic model described in Chapter 3 to optimize the sequences and LNA substitution patterns in the required primers and probes. I note that the assay described includes a number of other technical features that were developed by other members of the laboratory (Curtis Hughesman and Roza Bidshahri). Here, I therefore focus on the probe and primer design components of that assay, which I led and which required the use of the model described in Chapter 3 to account for the LNA:LNA base-pairs and proximal oppositely oriented LNA:DNA base pairs formed.
4.1.1  

BRAF V600 MUTATIONS IN MELANOMA

Although several RAF kinases (e.g. ARAF and CRAF) have been discovered and implicated in cancer, altered forms of the BRAF kinase are the most common; they are observed in ca. 8% of all cancers and in ca. 50% of melanomas [224]. In the BRAF gene, mutations most often occur within the kinase domain, either within or proximal to the activation loop (A-loop) located near the 600th amino acid (valine) encoded in exon 15 (87-99%) of the BRAF gene or, to a far lesser extent, the phosphate binding loop (P-loop) located at residues 464-469 encoded in exon 11 (1-7%) [6, 75, 76, 225]. Of the mutations observed in BRAF V600, a variant defined by a single point mutation (c.1799 t>a) resulting in an amino acid change from valine (V) to glutamic acid (E), known as the BRAF V600E mutation, accounts for ca. 80% of all BRAF mutations [226]. However, other BRAF V600 mutations, such as V600K (c.1798_1799 gt>aa), V600R (c.1798_1799gr>ag), and V600M (c.1798 g>a), also occur and are oncogenic (Table 4-1), as are certain especially rare mutations within V600 [76, 227]. The BRAF V600E mutation deregulates BRAF kinase activity though growth-factor-independent phosphorylation of downstream kinases (MEK and ERK) of the MAPK pathway, leading to uncontrolled cell proliferation in many cancers [6]. As a result, melanomas harbouring the BRAF V600E mutation are currently treated with BRAF V600E inhibitors. While the clinical significance of BRAF P-loop mutations is currently unknown, all BRAF V600 mutated cancers respond to mutant BRAF inhibitors [6].
Table 4-1 Incidence of *BRAF* V600 Mutations and Sensitivity to Targeted Inhibitors

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1795 1796 1797</td>
<td>g t g</td>
<td>T g a a</td>
<td>T a a g</td>
<td>T g a g</td>
<td>T g g a</td>
<td>T g a t</td>
<td>T a t g</td>
</tr>
<tr>
<td></td>
<td>Threonine</td>
<td>Valine (V)</td>
<td>Glutamic Acid (E)</td>
<td>Lysine</td>
<td>Arginine (R)</td>
<td>Aspartate (D)</td>
<td>Methionine (M)</td>
</tr>
<tr>
<td>1798 1799 1800</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1801 1802 1803</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mutant Frequency (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Cancers [228]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melanoma [75, 229-231]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inhibitor Sensitivity b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dabrafenib, D; Vemurafenib, V; Trametanib, T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Incidence rates are reported as the frequency of a particular variant within *BRAF* V600 mutations, excluding non-*BRAF*V600 mutations. The listed inhibitors represent FDA approved therapies for particular mutants while those in brackets are for therapies for which little data is available but where some effectiveness was observed. Sequences are written 5’ to 3’ and bolded letters indicate those natural DNA bases that were mutated to yield a mutated amino acid residue. The mutation frequencies in melanoma compare the meta-analysis of Kim et al. [231] with a conflated group of highly cited studies [75, 229, 230] and a recent large scale NGS study [225].
4.1.2 TARGETING BRAF KINASES FOR CANCER THERAPY

Two BRAF inhibitors – vemurafenib [77] and dabrafenib [238, 239] – as well as an inhibitor targeting a downstream MEK protein kinase – trametinib [240] – have been FDA approved for the treatment of BRAF V600E-positive metastatic melanoma, while multiple new inhibitors are in clinical development [219, 241]. Mutant BRAF inhibitors are remarkably effective in treating melanomas harbouring a BRAF V600E mutation, improving response rates and overall survival compared to standard chemotherapy (i.e. dacarbazine) [241]. Moreover, such treatments are generally well tolerated – both mutant-BRAF targeted drugs have manageable side effects. Mutant BRAF targeted therapies are also effective against rare mutations. Dabrafenib is additionally approved for V600K positive melanomas [239] and recent evidence suggests that V600K-positive melanomas also respond positively to verumafenib [77, 236, 242] and that at least one of either inhibitor may be effective against melanomas harbouring the more rare BRAF V600R, V600M, V600D (c.1799_1800tg>at), or V600E2 (c.1799_1800tg>aa) mutations [234, 235, 237]. However, treatment of melanomas harbouring wild-type (native) BRAF accelerates progression of pre-malignant lesions to cancer by the paradoxical activation of the MAPK pathway. Thus, mutant BRAF inhibitors should not be prescribed to patients negative for a BRAF V600 mutation [241, 243].

Improvement in patient outcomes has been observed in recent clinical trials combining RAF and MEK kinase inhibitors, prompting FDA approval of this therapeutic regimen for advanced melanomas [244, 245]. Parallel targeting of BRAF and MEK protein kinases in BRAF V600E or K positive melanomas with a combination of dabrafenib and trametinib can prolong survival by delaying the onset of inhibitor resistance [241] observed for monotherapy and attenuating the dabrafenib-induced (re)activation of the MAPK pathway [240, 244]. Clinical trials involving other combinations of BRAF and MEK inhibitors are ongoing [239]. That effort is motivated in
part by the fact that vemurafenib-resistant tumours can exhibit drug-dependent proliferation [6]. Nevertheless, moving forward, all clinical actions related to melanoma therapy require assaying for BRAF V600 coding mutations, including both rare and common BRAF V600 mutations that determine eligibility for treatment with specific BRAF and/or MEK inhibitors.

4.1.3 AVAILABLE DIAGNOSTIC ASSAYS FOR DETECTING BRAF V600 MUTATIONS

Currently, two commercial assays, the Cobas 4800 BRAF V600 mutation assay (Roche) [66] and the THxID™-BRAF assay (Biomérieux) [246], are FDA approved for BRAF V600E and for BRAF V600E and K testing, respectfully. The former is a real-time PCR assay utilizing an allele-specific (AS) probe, while the latter uses amplification of refractory mutation system (ARMS) PCR, a form of AS-PCR that relies on AS primers to discriminate between alleles. In the ARMS AS-PCR system, additional mismatches are introduced near the 3’ end of a standard pure-DNA AS primer (see Chapter 1) to discourage amplification of non-target alleles [72]. The two PCR based assays detect BRAF V600E or BRAF V600E and V600K mutations, respectively, to a mutant frequency (MF) detection limit of 5% (i.e. one mutant allele within a background of 20 wild-type alleles) [44, 66, 246, 247]. Thus, neither test provides for detection of V600-positive melanomas early in their oncogenesis (i.e. at an MF < 5%). Moreover, both assays exhibit cross reactivity to other BRAF V600 mutations: the Cobas assay cross-amplifies V600K and V600D (c.1799_1800tg>at) mutations, while the THxID assay cross-amplifies V600D mutations owing to the base-pair shared between the single and tandem mutations at c.1799 [44, 248]. The output for each assay is only qualitative (i.e. yes/no) when detecting mutant allele frequencies at or above 5%, which mitigates the ability of clinicians to fully define the cancer stage using the test. A need therefore exists for a more highly sensitive, quantitative, and truly allele specific test to enable more targeted treatment of patients with rare or common BRAF mutated cancers.
4.2 MATERIALS AND METHODS

The oligonucleotides, UV melt spectroscopy experiments, and allele specific PCR assay workflow required for this work are described below.

4.2.1 OLIGONUCLEOTIDES

Plasmid DNA harboring a segment of the human BRAF gene containing either the wild-type V600, V600E, V600E2, V600D, V600K, V600R, or V600M allele genotype was synthesized by Integrated DNA Technologies (IDT, Coralville, IA). Each plasmid mini-gene was restriction enzyme digested by ClaI (NEB, Ipswich, MA) yielding a linear DNA fragment. All pure DNA and LNA containing primers and probes were synthesized by either IDT or Exiqon Inc. (Hercules, CA).

4.2.2 UV MELT SPECTROSCOPY

Melting thermodynamics were collected by thermal-ramp UV spectroscopy for BRAF related duplexes in a buffer solution mimicking PCR solution conditions. The method used, including data analyses, is described in Chapter 2. Each equimolar sample containing a labeled or unlabeled probe and its complementary or mismatched target template oligonucleotide was re-suspended to a total concentration of 2 μM in a solution containing Taq standard reaction buffer (NEB) (at 1X: 50 mM KCl, 1.5mM Mg2+, pH 8.3). Additional magnesium was added to 3 mM total Mg2+. The samples were melted on a 12-cell peltier-controlled Cary 1E (Agilent, Santa Clara, CA) UV-Vis spectrophotometer described previously [249].

4.2.3 ALLELE-SPECIFIC PCR WORKFLOW

Allele-specific qPCR experiments used to study and confirm the specificity of AS probes designed using the model were conducted on a 96-well BioRad (Hercules, CA) CFX-96 real-time thermocycler. The qPCR thermocycling protocol consisted of 3 min at
95 °C, followed by 50 cycles consisting of a denaturation step at 95 °C for 15 seconds, and an annealing/extension step at 60 °C for 30 seconds. For the reactions monitored by SYBR green, a melt curve for PCR products was carried out to check for non-specific amplification by increasing the temperature from 50 °C to 95 °C at ca. 0.5 °C min⁻¹. PCR efficiency was determined by plotting the $C_q$ values for a series dilution of template in buffer (1xTE pH 8, IDT) against the starting template concentration of each reaction. The resulting slope was used to compute the efficiency as described in Chapter 1.

4.3 RESULTS

The following results outline the design strategy for the $BRAF$ $V600$ mutation specific real-time PCR assay, how the thermodynamic model developed in Chapter 3 was used to guide the allele specific LNA probe design, and finally the performance of the $BRAF$ $V600$ mutation specific assay directed against plasmids containing the target mutations.

4.3.1 $BRAF$ $V600$ MUTATION-SPECIFIC REAL-TIME PCR ASSAY DESIGN

A new assay for detecting and discriminating between common and rare $BRAF$ $V600$ mutations was developed by other members of the Haynes’ lab. That assay, whose basic design is shown in Figure 4-1 and Figure 4-2, requires the use of AS primers and AS probes that both harbour LNA substitutions, creating the potential to form LNA:LNA and proximal oppositely oriented LNA:DNA base pairs between those entities. To enable creation of the assay, I therefore was tasked with applying the model described in Chapter 3 to the design of the required LNA-substituted probes. Below I provide a brief description of the assay to provide the context needed to understand the probe designs I provided.

Clinically relevant $BRAF$ $V600$ mutations include $V600E$, $V600E2$, $V600D$, $V600K$, $V600R$, and $V600M$, and the intent was therefore to develop a new assay capable
of detecting (and discriminating between) each of these mutations, as well as quantifying the mutant frequency. The somatic point mutations in BRAF associated with each of these clinically relevant mutant BRAFs are provided in Figure 4-1A. Those data show that BRAF V600E, E2, and D share a common mutation at nucleotide 1799 (c.1799t>a), while BRAF V600K, R, and M share a common mutation at nucleotide 1798 (c.1798g>a). For each of these common mutations, an LNA-substituted AS forward primer was created to selectively amplify the three variants that share that mutation, but not amplify wild-type (WT) BRAF or any other mutant BRAF alleles. A common forward primer designed to detect and quantify the total copies of BRAF was also created. The assay then partitions gDNA from a patient tumour biopsy into three separate reaction wells – A, B and C (Figure 4-2). Real-time PCR (qPCR) is performed in each well using one of the forward primers, a common reverse primer and a common “consensus” probe that targets a highly conserved region of BRAF away from the V600 site of variance (SOV) but within the amplified template. Well A contains the set of common primers and the consensus probe needed to detect and quantify the total concentration of BRAF in the sample.
Figure 4-1 Our novel BRAF V600 mutation specific assay relies on the combined use of a set of AS primers (A) and AS probes (B). Since the 3’ terminal nucleotide (orange) of either AS primer is complementary to a distinct group of three mutant alleles that sharing a mutation at either nucleotide 1799 or 1798, that group of alleles is amplified. The specific mutant template within that amplified pool is identified by one of the multiplexed AS probes directed against that template.
Figure 4-2 A three tube assay allows for the quantification of mutant frequency and genotyping of both common and rare BRAF V600 mutations. Common forward and reverse primers first detect the total amount of all BRAF alleles in a sample (reaction A). In separate wells, two independent allele specific reactions (B and C) detect the identity and amount of each mutation type present in each tube. Comparison of output fluorescence curve from the consensus probe in reaction A (total BRAF) with the either allele specific reaction B or C (MT BRAF) allows the computation of mutant frequency.

Each of the two remaining wells (B and C) contains one of the AS forward primers, the common reverse primer, and the consensus probe to allow direct quantitative comparison with the control reaction conducted in well A. As each AS forward primer
selectively amplifies three different mutant \textit{BRAF V600} alleles, a corresponding set of three AS probes, each labeled with a unique reporter dye and specific to one of the three mutations, is also added to the well to enable discriminated detection of those mutations. As will be shown, achieving that level of discrimination is enabled through the introduction of LNA substitutions in the AS probe designs. Moreover, each AS probe is designed to hybridize to the template strand created by extension of the LNA-substituted AS forward primer. As a result, both LNA:LNA and oppositely oriented LNA:DNA base pairs may be formed between the AS forward primer extension product and each of the three AS probes within the well (Figure 4-3A). Indeed, each AS probe must be designed so that the formation of those modified base-pairs is required for probe hybridization (and thus hydrolysis) at the extension temperature $T_a$ used, as illustrated in panes 3b and 4 of Figure 4-3A.

Through this strategy, \textit{BRAF V600E}, \textit{E2}, and \textit{D} are detected and quantified in well B, while \textit{BRAF V600K}, \textit{R}, and \textit{M} are likewise analyzed in well C. Importantly, the combination of AS primers and probes allows reduction of the number of reactions in the assay from 7 separate wells (\textit{i.e.} one for each mutant allele and a control reaction) down to 3, thus reducing input DNA load and assay costs, while increasing throughput [37]. This level of assay densification is possible because current qPCR instrumentation allows for detection of up to five different reporter dyes, and thus 5 unique targets [65], although parallel targeting (multiplexing) introduces significant probe fidelity concerns (see below) and significant challenges in assay design (\textit{e.g.} probe-probe interactions) [74].

LNA substitution patterns in the 3’ end of primers that serve to greatly improve their allele specificity have previously been reported by Hughesman \textit{et al.} [67, 68] and were used to design the two required AS forward primers. Below I described how the set of LNA-substituted AS probes were then designed using the new model presented in Chapter 3 of this thesis.
4.3.2 MODEL-BASED DESIGN OF THE LNA-SUBSTITUTED PROBES

LNA-modifications have been shown to increase the allele specificity of probes, and this has bolstered their use in qPCR based technologies, among other applications.
LNA substitutions increase mismatch discrimination in part because base pairs formed with them increase duplex stability, allowing a probe of shorter length, when duplexed to its target (fully complementary) template, to reach the required $T_m$ for detecting amplification of the target allele by qPCR. As a result of the shorter probe length, mismatched base pairs formed with an LNA-substituted probe have a proportionately higher destabilizing effect on duplex stability. Moreover, due to their more rigid structure, LNAs present in a probe at the SOV are more intolerant to a mismatched base at that position [69, 105]. Consider, for example, a dual labeled hydrolysis probe designed to be complementary to the MT $BRAF \ V600E$ allele. That probe will form a mismatched duplex with any non-target $BRAF$ allele, such as WT $BRAF$ for which a single base-pair mismatch is formed at the SOV at nucleotide 1799. If that mismatched duplex is sufficiently stable at the qPCR annealing temperature, $T_a$, a false-positive signal will be recorded. Pure DNA AS probes often suffer from such cross-reactivity. However, the mismatch discrimination of an AS probe can be improved by LNA modifications placed in the probe at or near the SOV that serve to keep the $T_m$ of the duplex formed with the perfectly matched target allele ($T_{m,TA}$) above $T_a$, while lowering that ($T_{m,NT}$) formed with any mismatched non-target (NT) $BRAF \ V600$ allele [69, 105] to below $T_a$.

Table 4-2 reports the sequences of the common forward (BF FP164) and reverse (BF RP215) primers, the two LNA substituted AS forward primers, and the FAM-labeled $BRAF$ consensus probe (BF P16L5) that have been created by others in the lab for use in the 3-well $BRAF \ V600$ mutation-specific qPCR assay. The AS primer BF FP L123 used in well B selectively amplifies $BRAF \ V600E$, $E2$ and $D$ by aligning its 3’ terminal nucleotide with the SOV at nucleotide 1799 (c.1799t>a) of $BRAF$, and by including LNA substitutions at nucleotide positions 3’ – 1, 3’ – 2 and 3’ – 3 of the primer. Similarly, the AS primer BF FP L245 used in well C selectively amplifies $BRAF \ V600K$, $R$ and $M$ by aligning its 3’ terminal nucleotide with the SOV at nucleotide 1798 (c.1798g>a) of
BRAF, and by including LNA substitutions at nucleotide positions 3’–2, 3’–4 and 3’–5 of the primer.

Table 4-2 Primers and probes used in reactions A and B of the BRAF V600 assay

<table>
<thead>
<tr>
<th>Reaction A</th>
<th>Primer / Probe</th>
<th>Sequence (5'–3')</th>
<th>Length (bp)</th>
<th>Fluorophore</th>
<th>Quencher</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF FP164</td>
<td>ggtgatttttgcctagctaca</td>
<td>21</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BF RP215</td>
<td>agcctcaatcattacctca</td>
<td>21</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BF P16L5</td>
<td>tCcCaTcagTtTgaac</td>
<td>16</td>
<td>FAM</td>
<td>IBFQ</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reaction B</th>
<th>Primer / Probe</th>
<th>Sequence (5'–3')</th>
<th>Fluorophore</th>
<th>Quencher</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF FP L123</td>
<td>taggtgatttttgtagcCTaCAGa</td>
<td>25</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BF RP215</td>
<td>agcctcaatcattacctca</td>
<td>21</td>
<td>HEX</td>
<td>BHQ1</td>
</tr>
<tr>
<td>BF E 13L2</td>
<td>agattCTctgta</td>
<td>13</td>
<td>Texas Red</td>
<td>BHQ2</td>
</tr>
<tr>
<td>BF E2 13L3</td>
<td>agattTTctgta</td>
<td>13</td>
<td>Texas Red</td>
<td>BHQ2</td>
</tr>
<tr>
<td>BF D 13L3</td>
<td>agattTATctgta</td>
<td>13</td>
<td>TYE665</td>
<td>BHQ2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reaction C</th>
<th>Primer / Probe</th>
<th>Sequence (5'–3')</th>
<th>Fluorophore</th>
<th>Quencher</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF FP L245</td>
<td>ggtgatttttgtagCTaCaa</td>
<td>22</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BF RP215</td>
<td>agcctcaatcattacctca</td>
<td>21</td>
<td>HEX</td>
<td>BHQ1</td>
</tr>
<tr>
<td>BF K 13L2</td>
<td>atttCTtgtagc</td>
<td>13</td>
<td>Texas Red</td>
<td>BHQ2</td>
</tr>
<tr>
<td>BF R 13L2</td>
<td>atttCTtgtagc</td>
<td>13</td>
<td>Texas Red</td>
<td>BHQ2</td>
</tr>
<tr>
<td>BF M 13L3</td>
<td>atttCATtgtagc</td>
<td>13</td>
<td>TYE665</td>
<td>BHQ2</td>
</tr>
</tbody>
</table>

LNA and DNA bases are written in uppercase and lowercase respectively. Bolded letters represent those bases that are mutated from the BRAF WT sequence. FAM, HEX, TEX and TYE665 refer to 6’-carboxyfluorescein, Hexachlorofluorescein, Texas Red® and Tye™665 fluorescent dyes respectively. BHQ refers to a ‘Black Hole’ Quencher dye.

Like the AS probes to be designed, each AS primer exploits a matched versus a mismatched base pair at the SOV to distinguish between alleles. If the AS-forward-primer/AS-probe pairs are designed to be complementary to opposite strands of BRAF, the LNA-containing strand of amplicons created through extension of the AS forward primer may then be complementary to an AS probe (Figure 4-3A, pane 3b), which must also contain LNA substitutions to ensure that $T_{m,TA} > T_a > T_{m,NT}$. This creates the
potential for primer-probe hybridization within the overlap region \( d_o \) (Figure 4-3B) that may hinder the PCR reaction if improperly designed (Figure 4-3C).

The results reported in Chapter 3 show that hyperstabilizing structures (e.g. LNA:LNA base pairs) may be leveraged, in principle, to design short AS probes with very high specificity for their target. The model presented in Chapter 3 was therefore used to define the sequence and LNA-substitution pattern for each of the required set of six short AS probes (Figure 4-1B) so that each exhibits 1) a \( T_{m,TA} \) that is greater than \( T_a \), and 2) fails to form at stable duplex with the forward AS primer at \( T_a \) (a condition met if the probe-primer (PP) \( T_m, T_{m,PP} \), is less than \( T_a \)). For those model designed probes, UVM experiments were then conducted to determine if the added requirement that \( T_{m,NT} < T_a \) is satisfied, since the model described in Chapter 3 does not predict \( T_m \) values for duplexes containing one or more mismatched base pairs.

Table 4-3 reports the measured and model-predicted \( T_{m,TA} \) for each AS probe used in the assay. UVM-measured \( T_{m,NT} \) values are also reported in Table 4-3, all of which fall below \( T_a \). Designed in silico using the model, each of the six AS probes bind their target with a (stability) \( T_{m,TA} \) above \( T_a \) (60 °C), and bind non-target templates with stabilities, indicated by \( T_{m,NT} \) values, below \( T_a \), as required. Previous work in the Haynes’ lab has shown that AS probes must also display a \( T_{m,TA} - T_{m,NT} \geq 7 \) °C to be truly specific for their target alleles [251]. Each probe meets this added criterion for the three reactions occurring in the well in which they are present, with most of the AS probes designed using the model exhibiting \( T_{m,TA} - T_{m,NT} \) values greater than 10 °C.
Table 4-3 UVM derived and model predicted melting temperatures for dual-labeled *BRAF V600* PCR hydrolysis probes

<table>
<thead>
<tr>
<th>Reaction B</th>
<th>Template</th>
<th>Probe</th>
<th>BF-E-13L2</th>
<th>BF-E2-13L3</th>
<th>BF-D 13L3&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BRAF V600E</strong></td>
<td>67.3 ± 0.2 (64.4)</td>
<td>57.2 ± 0.1</td>
<td>55.3 ± 0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BRAF V600E2</strong></td>
<td>46.0 ± 0.0</td>
<td>65.3 ± 0.2 (62.6)</td>
<td>52.7 ± 0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BRAF V600D</strong></td>
<td>46.4 ± 0.1</td>
<td>52.8 ± 0.1</td>
<td>69.0 ± 0.2 (64.5)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reaction C</th>
<th>Template</th>
<th>Probe</th>
<th>BF-K-13L2</th>
<th>BF-R-13L2</th>
<th>BF-M 13L3&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BRAF V600K</strong></td>
<td>61.5 ± 0.3 (62.5)</td>
<td>45.0 ± 0.1</td>
<td>52.3 ± 0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BRAF V600R</strong></td>
<td>55.1 ± 0.1</td>
<td>64.4 ± 0.1 (63.0)</td>
<td>57.3 ± 0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BRAF V600M</strong></td>
<td>52.1 ± 0.2</td>
<td>47.6 ± 0.2</td>
<td>69.2 ± 0.1 (67.7)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>UVM data was collected under buffer conditions mimicking PCR relevant probe concentrations and the values above are corrected to 0.2 μM. Error bars represent standard deviations computed from triplicate scans of a single sample. The extended SBT model predictions for T<sub>m</sub> (in brackets) were computed for the perfectly matched targets (bold) using the parameters derived in chapter 3 while those for the mismatch containing non-target duplexes were computed assuming all LNA hyperstabilities are lost upon mismatch formation. <sup>b</sup>Thermodynamic parameters were not available for Tye665 reporter dyes so those probes were treated as unlabeled probes.

The mean model error and standard deviation for predicted T<sub>m,TA</sub> values was -2.0 ± 1.9 °C as shown in Table 4-4. Thus, although the extended SBT model developed in Chapter 3 slightly under-predicts T<sub>m,TA</sub>, Table 4-4 shows that correcting for the hyperstabilizing effects of proximal opposing LNA substitutions greatly improves model accuracy compared to models that do not have such corrections. Potential reasons for the reduced accuracy of the extended SBT model when applied to these duplexes include their higher LNA content compared to the duplexes used to regress the (current) hyperstability parameters, as well as known inaccuracies in the correlations used to account for the presence of a reporter dye and quencher on each probe, and used to convert T<sub>m</sub> values predicted at UVM solution conditions to the solution conditions used.
in qPCR. Nevertheless, as I will soon show, the model is sufficiently accurate to allow the in silico design of highly effective AS probes for the assay.

**Table 4-4** Comparison of $T_m$ values from UVM derived experiments and competing model predictions\(^a\)

<table>
<thead>
<tr>
<th>BRAF V600 Probe Duplexed With Its Target</th>
<th>$T_m,\text{expt}$ (°C)</th>
<th>$T_m,\text{pred. (model)} - T_m,\text{expt}$ (°C)</th>
<th>$\Delta C_P$ corrected Santalucia NN Model(^{[135,136]})</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF-E-13L2</td>
<td>67.3</td>
<td>-2.9</td>
<td>-11.6</td>
</tr>
<tr>
<td>BF-E2-13L3</td>
<td>65.3</td>
<td>-2.7</td>
<td>-13.4</td>
</tr>
<tr>
<td>BF-D 13L3</td>
<td>69.0</td>
<td>-4.5</td>
<td>-15.6</td>
</tr>
<tr>
<td>BF-K-13L2</td>
<td>61.5</td>
<td>1.0</td>
<td>-1.7</td>
</tr>
<tr>
<td>BF-R-13L2</td>
<td>64.4</td>
<td>-1.4</td>
<td>-5.0</td>
</tr>
<tr>
<td>BF-M 13L3</td>
<td>69.2</td>
<td>-1.5</td>
<td>-5.5</td>
</tr>
</tbody>
</table>

Mean error ± SD -2.0 ± 1.9  -8.8 ± 5.5  -31.4 ± 4.8

\(^a\)UVM data was collected under buffer conditions mimicking PCR relevant probe concentrations and the values above are corrected to 0.2 μM. Model predictions are also computed for $C_T = 0.2 \mu M$ assuming equimolar addition of complementary strands.

Finally, **Table 4-5** shows for each AS probe the associated model-predicted $T_{m,PP}$ values, all of which fall below $T_a$. Together, these results demonstrate the ability to use the model, in combination with a limited set of experiments verifying specificity, to design a putative set of LNA-substituted AS probes for the assay.
Table 4-5 Model predicted melting temperatures for LNA containing primer-probe (pp) duplexes

<table>
<thead>
<tr>
<th>Template</th>
<th>Probe</th>
<th>( T_{m,pp} ) (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reaction B</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRAF V600E</td>
<td>BF-E-13L2</td>
<td>25.1</td>
</tr>
<tr>
<td>BRAF V600E2</td>
<td>BF-E2-13L3</td>
<td>25.1</td>
</tr>
<tr>
<td>BRAF V600D</td>
<td>BF-D 13L3</td>
<td>25.1</td>
</tr>
<tr>
<td><strong>Reaction C</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRAF V600K</td>
<td>BF-K-13L2</td>
<td>42.5</td>
</tr>
<tr>
<td>BRAF V600R</td>
<td>BF-R-13L2</td>
<td>42.5</td>
</tr>
<tr>
<td>BRAF V600M</td>
<td>BF-M 13L3</td>
<td>42.5</td>
</tr>
</tbody>
</table>

\(^a\) Model predictions were computed assuming equimolar addition of primers and probes to a total \( C_T \) equal to 400 nM (200nM primer + 200nM probe) in PCR solution conditions (i.e. containing 50 mM Na\(^+\) and 3 mM Mg\(^{2+}\)).

4.3.3 ASSAY PERFORMANCE USING MODEL-DESIGNED AS PROBES

The performance of the multiplexed mutant-specific \( BRAF \) V600 qPCR assay in detecting each of the six most common and clinically relevant \( V600 \) mutant alleles is demonstrated in Figure 4-4. Each 10\% mutant-frequency reference sample assayed is comprised of the specified mutant allele in a background of WT allele. The result of the assay applied to each sample is comprised of the combined output from the three reaction wells, with positive identification of a \( BRAF \) V600 mutant allele provided by the observance of three real-time amplification curves – 1) amplification of total copies of \( BRAF \) (WT plus all mutants) in well A that is monitored in real time through hydrolysis of the FAM-labeled consensus probe and that shows the smallest quantitation cycle \( C_q \) (often between 22 and 24), 2) allele-specific amplification of \( BRAF \) in either one of the AS reaction wells (B or C) that is monitored through hydrolysis of the consensus probe and that shows a larger \( C_q \), and 3) allele-specific amplification of \( BRAF \) in either one of the AS reaction wells (B or C) that is monitored through hydrolysis of a HEX, Texas Red, or TYE665-labeled AS probe and therefore shows a lower end-point fluorescence.
For each *BRAF* mutant allele tested, the assay registers the expected trio of amplification curves collectively indicative of a mutant-positive result. Moreover, from the well in which AS amplification curves were recorded and the recorded spectral properties of dye released from the AS probe, the assay provides an unambiguous call on which mutant allele (*BRAF V600E, E2, D, K, R, or M*) is present.

![Figure 4-4](image)

**Figure 4-4** Real time qPCR results for the new *BRAF V600* AS PCR assay. *BRAF V600E, E2, D, K, R, and M* are all clearly distinguished at 10% MT frequency (A to F respectively). Both Rx A and B/C contain a FAM reporter (blue) that are used for quantification and the AS Rx B/C also contains a different reporter dye (Hex, green; TexasRed, Red; Tye665, Magenta) to indicate the mutation genotype (Table 4-2).

Amplification curves generated from a serial dilution of *BRAF V600E* in buffer were used to compute the amplification efficiency of the AS primer used in well B in the presence of the AS probes (**Figure 4-5A**). The measured $C_q$ value for each dilution was plotted against the logarithm of the starting template quantity, and the amplification efficiency $E$ then computed from the slope of the curve; an $E = 100.0\%$ was recorded, indicating a doubling of each DNA template at each cycle. This maximum possible
amplification efficiency suggests that the primer is not being inhibited by the LNA-substituted probes through primer-probe duplexation within region $d_o$ (see Figure 4-3). Finally, in Figure 4-5B, a set of amplification curves are reported for $BRAF V600E$ DNA serially diluted in a background of WT DNA from 100% $BRAF V600E$ to 0% (WT only). The lowest mutant frequency that was significantly and reliably distinguishable from the pure WT template was 0.1% $BRAF V600E$ (1 MT allele in 1000 WT alleles), a selectivity of $10^3$. This limit of detection (LOD) represents an improvement of greater than one order of magnitude over the currently approved FDA companion diagnostics for $BRAF V600E$. 
Figure 4-5 Serial dilutions in buffer (A) and in a background of WT templates (B) allow the computation of i) the PCR reaction efficiency (inset) from the slope of the $C_q$ vs starting DNA quantity curve, and ii) the limit of detection respectively. A selectivity of $10^3$ (1 MT in $10^3 = 0.1\%$) is determined from plot B for $BRAF\ V600E$ by comparing the $C_q$ of the lowest distinguishable dilution of $BRAF\ V600E$ from pure WT.

In summary, these preliminary data show at a proof of concept level that AS probes designed with the new molecular thermodynamic model presented in Chapter 2 may be combined with AS primers to detect and distinguish with high selectivity the six $BRAF\ V600$ mutations that are prognostic of melanoma patient sensitivity to targeted BRAF inhibitors. The performance of the LNA AS primers and probes in the assay depends on the hyperstabilities provided by the interaction of LNA substitutions on opposing strands. This is accounted for in the model, which enables the design of AS
probes that are extremely short and thereby provide superior discrimination. This in turn allows multiplexed detection of multiple variants, which reduces the cost of the test and limits the amount of DNA required. Indeed, even in its current proof-of-concept form, the assay outperforms both FDA approved tests by detecting and quantifying each of the six clinically relevant *BRAF V600* mutations below a 5% mutant frequency.
CHAPTER 5: CONCLUSIONS AND OUTLOOK

Advances in DNA analysis technologies, including next generation sequencing (NGS) platforms for identifying genetic aberrations or polymorphisms (e.g. somatic point mutations), have provided troves of genetic information that are serving to connect disease with the human genome. The catalogue of somatic mutations in cancer (COSMIC)[228], for example, compiles findings from genetic studies to reveal those mutations prognostic of cancer risk and progression. Knowledge of these mutations is likewise guiding the identification of genetic biomarkers that correlate with drug sensitivity. The \textit{BRAF V600E} mutation is one example of a burgeoning number of cases where molecular markers or patterns of markers are used to detect and monitor cancers or to classify patients and their cancers by prognosis and sensitivity to treatments. New regulations imposed by the Affordable Care Act are exerting significant pressure on clinical diagnostic technologies, including those that detect established biomarkers in genomic DNA of patients, to operate at reduced cost, while also providing suitable throughputs and turn-around times. New concepts and whole new assay development strategies will be required to overcome these challenges. This includes the use of non-natural nucleotide chemistries. Conformationally restricted locked nucleic acid (LNA) chemistries replacing natural DNA nucleotides have become widely used in diagnostic and therapeutic technologies due to their flexible and tailored enhancement of specificity and binding affinity, as well as their favourable \textit{in vivo} properties that were integral to the development of LNA-based therapeutics such as Miravirsen (Roche – formerly Santaris Pharma, Copenhagen DK), a therapeutic effective in reducing Hepatitis C viral loads in humans (phase II trial) [119].

However, the design of oligonucleotide agents incorporating LNA substitutions remains challenging due to the paucity of useful design tools. Clear instructions dictating the number and pattern of LNA substitutions are needed to replace simple and often inexact heuristics currently used. That need spurred the recent development of a collection of molecular thermodynamic models for predicting the stability of LNA-
incorporating duplexes whose collective purpose is to determine how $T_m$ and melting thermodynamics ($\Delta H(T_m), \Delta S(T_m), \Delta C_p^o$, and $\Delta G(T_m)$) vary with duplex length and LNA substitution pattern. Application of these nearest-neighbour molecular thermodynamic models depends on the availability of the required model parameters, which are regressed from libraries of thermodynamic data collected from diverse sequences of varying length to ensure universal application. Chapter 2 provides updated and more exacting protocols for applying two techniques, UV-spectroscopy monitored melt (UVM) experiments and differential scanning calorimetry (DSC), to the measurement of $T_m$, $\Delta H(T_m)$, $\Delta S(T_m)$, and $\Delta G(T_m)$ data for short chemically modified duplex DNA. These protocols account for the unique challenges presented by duplexes containing chemically modified nucleotides, and describe effective ways to address those challenges in the experimental design.

Many diagnostic technologies probing genomic DNA rely on its selective base-pairing ability. For example, probes may be designed to selectively hybridize to a target allele with duplex stability that exceeds the stability of a duplex formed between the probe and a nearly identical non-target allele in which a mismatch is formed at one or more base positions. Thermodynamic models may be used to design a probe to maximize this difference in stabilities – a key driver of assay performance – and thereby identify the most promising candidate probe sequences. Although current models designed for this purpose are capable of accurately predicting the $T_m$ of short DNA duplexes containing LNAs in a single strand[105], they remain limited in some important respects. The stabilities of PCR-inhibiting structures formed between two probes or a primer-probe pair that both contain LNAs cannot currently be predicted. Moreover, the hyperstabilities of LNA:LNA and proximal oppositely oriented LNA:DNA base pairs have not been exploited in diagnostic assays due in large part to an inability to predict the stability of duplexes containing LNAs in both strands. To address this gap, Chapter 3 presents a set of parameters required to accurately predict – within the framework of an existing nearest neighbour thermodynamic model previously developed in the Haynes lab – the stability
of short B-form duplexes containing LNAs in both strands by using a group-contribution approach. The large set of UVM-derived thermodynamic data collected and used to regress those parameters shows that an asymmetric hyper-stabilization is observed for 3’ offset oppositely oriented LNA:DNA base pairs, but not when those pairs are offset to the 5’ side of one another. Additionally, the data show that LNA-LNA base pairs are highly stabilizing. Although the origins of the hyper-stabilization phenomena were not addressed, the model extension created through that discovery enables accurate prediction of the $T_m$ of any short complementary DNA duplex containing LNAs in both strands; it is the only model to do so.

To demonstrate its utility, that model was applied to the design of a set of allele-specific (AS) probes to be used in a novel real-time PCR assay designed to detect and distinguish the clinically relevant somatic mutations that occur in the $BRAF$ V600 codon. The assay combines LNA AS primers and LNA AS dual-labeled hydrolysis probes to detect and distinguish six $BRAF$ V600 mutations in parallel with high selectivity. Critical elements of the $BRAF$ V600 mutant-specific assay are presented at the proof-of-concept level in Chapter 4, including i) a description of how the extended thermodynamic model, developed in Chapter 3, was used to design the LNA AS probes used in the assay and ii) real-time qPCR data showing how the assay selectively detects and distinguishes each mutant gene in a background of wild-type $BRAF$. The parallel detection of several mutations with high selectivity provides a means to reduce the quantity of DNA needed to obtain a reliable result. The performance of the qPCR assay (e.g. simplicity, low cost, closed-tube reaction), made possible by use of the molecular thermodynamic model reported in Chapter 3, therefore illustrates the potential value in exploiting more complex LNA substitutions in routine clinical testing and monitoring of disease.

5.1 PROPOSED FUTURE DEVELOPMENTS

The new parameters and extended model reported in Chapter 3 describing hyper-stabilizing effects of LNA:LNA base-pairs and proximal oppositely oriented LNA:DNA
base pairs can be used to guide the development of new advanced diagnostic assays. This is demonstrated in Chapter 4 through the development of a novel BRAF V600 mutation assay. In that application, the model shows good accuracy for duplexes containing 3 or fewer LNA substitutions, but performs somewhat more poorly for duplexes containing higher LNA content and/or certain substitution patterns, such as multiple neighboring LNA:LNA base pairs. Future work can address these limitations by conducting further UVM studies designed to better understand why LNAs exhibit hyper-stabilizing behavior, and to refine (or generate new) parameters for complex and or highly substituted LNA substitution patterns. Only one structural study has been reported for a palindromic DNA duplex containing LNAs in a $j=\pm1$ (5’ offset – non hyperstabilizing) orientation, and its helical structure resembled the A-form geometry of RNA-RNA duplexes, albeit with subtle differences that were not fully explained [252]. Future structural studies on both 5’ and 3’ offset LNA patterns may therefore reveal the origins of the hyperstabilizing effects first reported in this thesis and its associated publications.

Further extension of the molecular thermodynamic model reported in Chapter 3 to enable the accurate prediction of $T_m$ values for short mismatched DNA duplexes containing LNAs on both strands would be of very significant value. Such a model could be applied to the in silico design of any combination of primers, probes, wild-type blockers, or other oligonucleotides containing LNAs. In addition, the prospects of RNA based duplexes containing oppositely oriented LNAs remains an open question and may, based on the enhancements of LNAs in single stranded RNA[107], be more pronounced than effects reported in Chapter 3 for DNA.

Single-strand oligonucleotide-based diagnostic and therapeutic technologies may also benefit from LNA substitutions. The parameters in Chapter 3 apply to duplexes formed from two non-self complementary strands, and are therefore not directly applicable to single-stranded structures. However, LNA substitutions can enhance the performance of therapeutic aptamers [125] and molecular beacon probes [23], for
example, where the LNAs may on occasion appear in duplexed stem portion of the structure to form hyper-stabilized base pairs. Reliable methods for designing those molecules and predicting their melting thermodynamics are not available. Their development using the approaches described in this work could perhaps enable stem regions to be shortened and/or stabilized by opposing LNA interactions, potentially leading to new therapeutic aptamers or useful diagnostic agents.

Finally, as demonstrated with the proof-of-concept BRAF V600 assay described in Chapter 4, new and useful assays may be realized through the use of interacting chimeric LNA strands. The strategy employed could be used to create assays against other commonly mutated genes (e.g. KRAS, EGFR). And those assays could be tailored for analyzing genomic DNA recovered from solid tumour biopsies, or within circulating DNA fragments shed by tumour cells into the blood (ctDNA) as a means to reduce the invasiveness of testing. The detection and quantification of biomarkers within ctDNA is hindered, in part, by their short size (~160-200 bp) and low abundance (< 1% of WT DNA) [22], making the design of highly efficient and specific probes and primers particularly challenging. The LNA primer-probe design concepts introduced in this thesis can potentially allow for the creation of very short amplicons, which may enable new and effective ctDNA assays. Overall, then, the results presented herein could further the clinical management of cancers through LNA-enhanced assay selectivity and modified assay design.


Dabbs DJ. *Diagnostic immunohistochemistry*: Elsevier Health Sciences, 2013.


Reed GH, Kent JO, Wittwer CT. High-resolution dna melting analysis for simple and efficient molecular diagnostics. 2007.


Wittwer CT, Reed GH, Gundry CN, Vandersteen JG, Pryor RJ. High-resolution genotyping by amplicon melting analysis using lcgreen. *Clinical chemistry*. 2003; **49**: 853-860.


Hughesman C. Molecular thermodynamics of the stability of natural, sugar and base-modified DNA duplexes and its application to the design of probes and primers for sensitive detection of somatic point mutations. 2012.


Hughesman CB, Turner RF, Haynes C. Correcting for heat capacity and 5'-ta type terminal nearest neighbors improves prediction of dna melting temperatures using nearest-neighbor thermodynamic models. *Biochemistry*. 2011; **50**: 2642-2649.


McTigue PM, Peterson RJ, Kahn JD. Sequence-dependent thermodynamic parameters for locked nucleic acid (Ina)-dna duplex formation. *Biochemistry*. 2004; **43**: 5388-5405.


Gray DM. Derivation of nearest-neighbor properties from data on nucleic acid oligomers. I. Simple sets of independent sequences and the influence of absent nearest neighbors. *Biopolymers*. 1997; **42**: 783-793.

Gray DM. Derivation of nearest-neighbor properties from data on nucleic acid oligomers. II. Thermodynamic parameters of dna-Rna hybrids and dna duplexes. *Biopolymers*. 1997; **42**: 795-810.


Chaires JB. Possible origin of differences between van't hoff and calorimetric enthalpy estimates. *Biophys Chem*. 1997; **64**: 15-23.


Situ BCI. Cancer facts. 2015.


Chisholm CL, Garlick R, Anekella B. Controls to detect limit of detection for braf v600e as an actionable mutation of interest by nextgen sequencing and qpcr. Cancer Research. 2015; 75: 3394-3394.

Inc. B. Thxid™-braf. In: -USA B, ed. 164614 C - En edn. <div data-canvas-width="141.756" style="left: 626.8px; top: 1291.43px; font-size: 15px; font-family: sans-serif; transform: scaleX(1.03472);">100 Rodolphe Street <div data-canvas-width="252.79500000000002" style="left: 626.8px; top: 1308.64px; font-size: 15px; font-family: sans-serif; transform: scaleX(1.03605);">Durham, North Carolina 27712 - USA 2013; 1-23.


Egli M, Minasov G, Teplova M, Kumar R, Wengel J. X-ray crystal structure of a locked nucleic acid (lna) duplex composed of a palindromic 10-mer dna strand containing one lna thymine monomercoordinates and structure factors have been deposited in the protein data bank (pdb code 1i5w). Chemical Communications. 2001: 651-652.


Appendix A

MATLAB CODE FOR SINGLE BASE THERMODYNAMIC (SBT) MODEL PARAMETER REGRESSION

A.1 Description of Matlab

MATLAB 2007b® (Mathworks, Natick, MA) software was used to regress the set of thermodynamic parameters describing, at the Gibb’s energy level, the incremental change in DNA duplex stability resulting from LNA substitutions in both strands described in Chapter 3. Matlab is a commercial computing package adept at matrix and data handling that is commonly used in engineering and scientific applications including for example, signal processing. Matlab users can generate M files using an upper level programming language based on C whose tasks are facilitated by pre-loaded modular M files – for root finding, solving ordinary differential equations, or signal processing (using extensible toolkits) for example. Matlab was chosen for its matrix handling ability and pre-loaded file packages which are necessary for non-linear regression, especially when handling large datasets of melting thermodynamic data.

A.2 Summary of Parameter Regression

New thermodynamic parameters describing the change in DNA duplex stability resulting from LNA substitutions in both strands were regressed using a Levenberg Marquardt (LM) scheme within the context of the SBT model by minimizing the sum of squared errors (SSE, $\chi^2$) of the difference between the model predicted and experimentally observed incremental change in melting temperature ($\Delta \Delta T_m$) for a library of LNA loaded duplexes. The Matlab files that were critical to this regression were Levmarq.M and ddtnfuncLNAModel.M which carried out the LM scheme and predicted the incremental melting temperature of a duplex respectively and are briefly described below.

Briefly, LM regression carried out by Levmarq.M implements the LM method which combines the fastest descent and Gauss Newton root finding methods. A factor $\lambda$ controls the
relative contribution of either root finding method to the refinement of parameter set estimate. At each iteration, as shown in Figure A-1, a trial parameter set is computed and if that parameter set reduces $\chi^2$ the trial is accepted, else $\lambda$ is modified to favor the refined Gauss Newton method and a new trial parameter set is generated. To generate a trial, the Jacobian matrix $J_i$ containing all the partial derivatives of a function – in this case the equation (3-10) computing $T_m$ with respect to the number of parameters, one for each LNA pattern studied – was computed using a central difference approximation for each partial derivative element. A trial parameter set $a^*$ is computed from the current parameter set $a$ as below:

$$a^* = a + \Delta a = a + (\alpha'^{-1} \beta)$$  \hspace{1cm} (A-1)

$$\alpha' = J^T W J + \lambda \cdot \text{diag}(J^T W J) = \frac{\partial^2 \chi^2}{\partial a^2} (1 + I \lambda) \cdot \frac{1}{\sigma^2}$$  \hspace{1cm} (A-2)

$$\beta = J^T W (T_m - T_{m,(pred)}) = \frac{\partial \chi^2}{\partial a} \frac{(T_m - T_{m,(pred)})}{\sigma^2}$$  \hspace{1cm} (A-3)

$$J_i = \frac{T_{m,i,(pred)}(x, a_i + dt) - T_{m,i,(pred)}(x, a_i - dt)}{2 \cdot dt}$$  \hspace{1cm} (A-4)

where $\alpha'$ is the Hessian matrix, $W$ is the weighted error matrix with diagonal error elements $1/\sigma$, $I$ is the identity matrix, and $\beta$ results from terms derived from the partial derivative of $\chi^2$. The scheme iterates until a stopping criterion (e.g. $\chi^2(a^*) - \chi^2(a) < \epsilon$) is met and $a$ is returned to the user.

`ddtmfuncLNAModel.M` is an extended version of an M file used to compute the incremental change melting temperature ($\Delta \Delta T_m$) of a given duplex. This file computes $T_m$ for four sequence-related duplexes: the DNA duplex containing LNA substitutions in both strands and the three parent duplexes in which the LNA modifications in either or both parent strands are replaced with their natural pure DNA analog nucleotide(s). To compute $T_m$ from a sequence of characters representing DNA and LNA bases, `ddtmfuncLNAModel.M` relies on several helper files to evaluate equations (1-6) to (1-8). For instance, stacking matrices summing the number
and type of nearest-neighbour (NN) DNA doublets, LNA bases in one strand, and number of distinct LNA pattern geometries (j in equation (3-13)) in a duplex are carried out by dnaSeq.M, lnaSeq.M, and lnaStack.M respectively whose code is shown below. A Matlab root finding routine, fzero.M, then solves the non-linear implicit equation (3-10) to compute $T_m$ for each of the four duplexes. The initial guess of $T_m$ for each duplex is computed assuming $\Delta C_p$ is equal to zero. Together Levmarq.M and ddtmfuncLNAModel.M are used to compute and improve upon an initial parameter set trial until a new and refined parameter set is regressed.

**Figure A-1** A Levenberg Marquardt scheme for non-linear parameter regression. An initial guess for a parameter set $\hat{p}$ is input, a variable factor $\lambda$ is initiated, and its associated sum of squared errors $\chi^2(\hat{p})$ is computed. A new trial parameter set is generated based on a blending of the fastest descent method and Gauss-Newton method and its $\chi^2$ has decreased, the parameter set is accepted, else a new trial is computed. The iteration continues until a stopping criterion is met and the parameters are returned.

### A.3 Matlab Code

*levMarq.M*

```matlab
function [params SD] = LevMarq(data,InitGuess,func)
```

175
% the goal of this function is to return an iterative parameter regression
% using the Levenberg Marquardt scheme that combines the Newton Raphson
% method and the fastest descent

% Function must be a variable input function defined in an m file
% data should be input in the form [ Xi, Yi, Sigmai]

%% DEFINE CONSTANTS

% set initial parameters
lambda=1E-3; % this is our initial guess for the stepping size
NU=9; % this will be used to change the order of lambda later
a=InitGuess; % is our initial guess of the parameter set
maxIT=500; % this is the maximum number of iterations
count=1; % the first iteration;
tol=1E-6; % set a tolerance for comparison of chi squared iterations
tmerr=0.8;

%% DEFINE DATA

Ydata=data{length(data)}; % a vector: Comes in as cell array
Xdata=data(1:length(data)-1); % contains duplexes and Ks
Sigma=tmerr*ones(length(Ydata),1); % this can be manipulated to be a vector
when data is provided
W=1/(tmerr)^2*eye(length(Ydata)); % ERROR WEIGHTED Matrix

%% L_M METHOD STEPS

% calculate the SSE of the first guess
chiSquared=sum(((Ydata-func(Xdata,a))./Sigma).^2)

while chiSquared>tol && count <maxIT% if chi squared reaches zero we can stop
    J=numJacobian(func,Xdata,a);
    alpha=J'*W*J +lambda*(diag(diag(J'*W*J),0));
    beta=J'*W*(Ydata-func(Xdata,a));
    astep=alpha\betaa;

    aFuture=a+astep;

    chiSquaredFuture=sum(((Ydata-func(Xdata,aFuture))./Sigma).^2)
% Need some criteria to stop if chi squared won't reach our tolerance
   if abs(chiSquaredFuture-chiSquared)<1E-10 % chi Squared is greater than
   the tolerance but is not changing anymore
       break
   end

   if chiSquaredFuture<chiSquared % if our guess improved
       %update a
       a=aFuture;
       %reduce lambda
       lambda=lambda/NU;
       chiSquared=chiSquaredFuture
       display('good guess')
       a
   
   else % our guess did not improve
       % no need to update a
       % increase lambda though:
       lambda=lambda*NU;
       display('bad guess')
   end
   count=count+1; % count the number of iterations
end
% out of while loop, we have solved for parameters
params=a;
% Now that we have the parameters we need to calculate the parameter SDs

J=numJacobian(func,Xdata,a);
alpha=J'*W*J +lambda*(diag(diag(J'*W*J),0));
V=alpha^-1; % the co variance matrix

% Want the asymtotic standard parameters: measure of how unexplained
% variablity in the data propegates to variability in the parameters
SD=sqrt(diag(V));

end

ddtmfuncLNAModel.M
function ddTM = ddtmfuncKCMModel(duplexesK, params, varargin)

% Calculates Tm given sequence information, ct, a ddH parameterset, and
% optional parameter inputs:

% K is ct/4 for non self complementary and ct for self complementary
% Delta CP would be varargin 1
% Tref would be Varargin 2

% params has the form of a two 1x4 stacked column vectors for ddH and ddS
% respectively
% DEFINE DATA
% data comes in as a 1x2 horz cell array of {sequences and K}
sequences = duplexesK{1}; % gathers sequences from the cell array
K = duplexesK{2}; % now this is a vector from a passed cell array
% dCP = duplexesK{3}; % collect cp data

% DEFINE Constants
R = 1.987; % cal/molK
cp = 42; % cal/molbp
Tref = 53; % reference temperature

% Calculate Santalucia parameters ddH and ddS
% Santa Lucia dataset:
SLparams = [
    -7.9  0.2 -22.2  0.8
    -7.2  0.7 -20.4  2.4
    -7.2  0.9 -21.3  2.4
    -8.5  0.6 -22.7  2
    -8.4  0.5 -22.4  2
    -7.8  0.6 -21.0  2
    -8.2  0.6 -22.2  1.7
    -10.6 0.6 -27.2  2.6
    -9.8  0.4 -24.4  2
    -8.0  0.9 -19.9  1.8
    0.1  1.1 -2.8   0.2
    2.3  1.3  4.1  0.2
    0.0  0  -1.4   0];

% Data in the form of dHDNA SD dSDNA SD for the list as specified in
% dnaStack

% PARAM vector = Stacking Matrix * Param List
S0=dnaStack(sequences); % Collect the DNA Stacking Matrix
dHDNA=S0*(SLparams(:,1)); % These are the DNA enthalpies
dSDNA=S0*(SLparams(:,3));% These are the DNA entropies

TMDNA=zeros(length(sequences(:,1)),1);

for i=1:length(sequences(:,1))
    dCP=length(sequences{i})*cp; % uses the num of bp * cp = cal/molK
    fd=@(tm) -tm + (dHDNA(i)+dCP/1000*(tm-tref))/(-dSDNA(i)-
    R*log(K(i))+dCP*log((tm+273.15)/(tref+273.15))*1000-273.15);
    TMDNA(i)=fzero(fd,tref);
end

%% Calculate LNA parameters

SBTparams=[
    0,-1.9
    0,-2.9
    0,-2.7
    0,-3.7];

S1=lnaStack(sequences(:,1));% collect LNA stacking matrix of Oligo 1
S2=lnaStack(sequences(:,2));% collect LNA stackking matrix for oligo 2
ddHLNA1=S1*SBTparams(:,1);
ddHLNA2=S2*SBTparams(:,1);
ddSLNA1=S1*SBTparams(:,2);
ddSLNA2=S2*SBTparams(:,2);

ddHLNA=ddHLNA1+ddHLNA2; % LNA enthalpies (always 0)
ddSLNA=ddSLNA1+ddSLNA2; % contributions from both oligos
TMLNA1=zeros(length(sequences),1);
TMLNA2=zeros(length(sequences),1);

for i=1:length(sequences)
    dCP=length(sequences{i})*cp; % uses the num of bp * cp = cal/molK
    f1=@(tm) -tm + (dHDNA(i)+ddHLNA(i)+dCP/1000*(tm-tref))/(-
    dSDNA(i)+ddSLNA(i)-R*log(K(i))+dCP*log((tm+273.15)/(tref+273.15))*1000-273.15;
    f2=@(tm) -tm + (dHDNA(i)+ddHLNA2(i)+dCP/1000*(tm-tref))/(-
    dSDNA(i)+ddSLNA2(i)-R*log(K(i))+dCP*log((tm+273.15)/(tref+273.15))*1000-273.15;
TMLNA1(i)=fzero(f1,TMDNA(i));
TMLNA2(i)=fzero(f2,TMDNA(i));
end

%% Calculate ddG Orientation Prameter
S3=seqLNA(sequences); % returns the stacking matrix for 8 Orientation parameters
ddGLNAO=S3*params; %returns the value for orientation

%% Root Finding for Tm
TM=zeros(length(sequences),1); % make an empty tm vector

%fill in the tm vector
for i=1:length(sequences) % for every sequence calculate the TM
    dCP=length(sequences{i})*cp; % uses the num of bp * cp = cal/molK
    % Function to find root of implicit equation is :
    g=@(tm) -tm + (-dHDNA(i)+ddHLNA(i)+dCP/1000*(tm-tref)+ddGLNAO(i))/(-dSDNA(i)+ddSLNA(i)-R*log(K(i))+dCP*log((tm+273.15)/(tref+273.15)))*1000-273.15;

    %check root finding function with TMDNA as initial guess
    %fplot(g,[40 80])
    %pause
    TM(i)=fzero(g,TMDNA(i)); % find the solution to t with 53 as an initial guess
end

%% Calculate ddTM
ddTM=(TM-TMDNA)-((TMLNA2-TMDNA)+(TMLNA1-TMDNA));
function orientStack = seqLNA(duplexes)

% This function returns a vector of LNAs of length = to the rows(duplexes)
% or number of duplexes from a set of forward and reverse sequences in the
% input argument matrix "duplexes" which has the form [ forward sequence
% reverse sequence ] for i rows written 5'-3' always.

% LIMITATIONS
% this is only good for single LNA interactions where an LNA has a positino
% on both sides of the duplex

% Example:

duplexes ={
    'ttcatAgccgt'  'acggCtatgaa'
    'ctaacGgatgc'  'gcaTccgtag'
    'ctgaaGtccgc'  'gcGgacttcag'
    'ttcAtagccgt'  'acgGcta
tgaa'
    'ctaacggAtgc'  'gcatccGttag'}; % these brackets are called cell arrays and
work similar square brackets

% Example Returns: N x 11 stacking matrix as follows
% POSITION -4 -3 -2 -1 0 1 2 3
%             %
%            0 0 0 1 0 0 0 0 0
%            0 0 0 0 0 0 0 0 0
%            0 0 0 0 0 0 0 0 0
%            0 0 0 0 0 0 0 1 0
%            0 0 0 0 0 0 0 0 0

%% INPUT ERROR CHECKS

% 1. just checks to make sure some oligos arent missing
if length(duplexes(:,1))~= length(duplexes(:,2))
    error(' Missing at least one of oligo 1 or one of oligo 2')
end

% 2. Needs to check to make sure all the oligos are complementary, if not tell
% you which ones aren't complementary

for i=1:length(duplexes(:,1)) % for every row (duplex)
    oligol=duplexes{i,1}; % get oligol !!
```matlab
seqcomp=cell(1,length(oligo1)); % start with an empty complement string of the same length as oligo 1

% gets the 5'-3' complement of this string
for m =1:length(oligo1)
    base=oligo1(m); % gets the base starting at the 5' (left) side

    % fills in the duplex from right (3') to left (5')
    if strcmpi(base,'a')
        seqcomp{length(oligo1)+1-m}='t';
    elseif strcmpi(base,'t')
        seqcomp{length(oligo1)+1-m}='a';
    elseif strcmpi(base,'g')
        seqcomp{length(oligo1)+1-m}='c';
    elseif strcmpi(base,'c')
        seqcomp{length(oligo1)+1-m}='g';
    end
end

% the seqcomp is written 5' to 3'
seqcomp=transpose(char(seqcomp));% squish the cell array into an array then transpose it

% get oligo 2 (for subsequent comparison)
oligo2=duplexes{i,2};

if strcmpi(oligo2,seqcomp) % checks for oligo 1 and 2 complementary
    % don't need to do anything if complementary
else % if not complementary
    display('duplex',i,'is not complementary')
    error('FIX IT')
end
end

% Start creating the stacking matrix
% we now know that we have a complementary set of matched oligos
numDuplexes=length(duplexes(:,1)); % returns the num rows in the first column

% makes an empty Nx11 stacking matrix with the 11 orientations as described above
orientStack= zeros(numDuplexes,8);

for j=1:numDuplexes % for every duplex (or row in OrientStack)
```

182
% Find the LNA position for each oligo by
% 1. get the sequences
  oligo1=duplexes{j,1}; % Get oligo 1
  oligo2=duplexes{j,2}; % Get oligo 2

% 2. asks if the sequence has LNAs and gets position
  isCaps1=isstrprop(oligo1,'upper'); % finds caps, returns vector of 0s and
  ls: 1 for caps 0 for non caps
  isCaps2 =isstrprop(oligo2,'upper');

  LNApos1 =find(isCaps1); % returns CAPS indexes the isLNA[nth, mth,
  oth...] letters in duplexes are LNAs
  LNApos2 =find(isCaps2); % returns LNA index as just above

% gets a numerical definition for the orientation type based on the
% distance from the 5’ end of the first oligo compared to the distance
% off from the 3’ end of the second oligo

  orientNum=(LNApos1-1)-(length(oligo2)-LNApos2);
% 3. Assigns an orientation value and fills in the stacking matrix

  if orientNum==-4 % a G3F
    orientStack(j,1)=orientStack(j,1)+1;
  elseif orientNum==-3 % a G2F
    orientStack(j,2)=orientStack(j,2)+1;
  elseif orientNum==-2 % a G1F
    orientStack(j,3)=orientStack(j,3)+1;
  elseif orientNum==-1 % a KF
    orientStack(j,4)=orientStack(j,4)+1;
  elseif orientNum==0 % a KF
    orientStack(j,5)=orientStack(j,5)+1;
  elseif orientNum==1 % a KR
    orientStack(j,6)=orientStack(j,6)+1;
  elseif orientNum==2 % a G1R
    orientStack(j,7)=orientStack(j,7)+1;
  elseif orientNum==3 % a G2R
    orientStack(j,8)=orientStack(j,8)+1;
  end
end
function S = lnaStack (duplexes)

% this method returns the stacking matrix for a series of duplexes having
% some LNAs A T G or C.

% NOTE: Capital Letters are LNAs

% Example:

% duplexes ={
%   'ggaacAagatgc';
%   'gcatcttGttCc';
%   'ggAAcaAgatgc';
%   'gcatctTgttcc';
%   'ggaacaagAtgc';
%   'gcatCttgttcc'};

% Make an empty (A T G C) 4 x N  Stacking matrix
S=zeros(length(duplexes(:,1)),4);

% Fill in the stacking matrix by comparing each string to find Capital
% Letters which represent LNAs
for i=1:length(S(:,1))  % for all the rows of S
    seq=duplexes(i);  % GETS the sequence of interest

    % asks if the sequence has LNAs
    isCaps=isstrprop(seq,'upper');  % finds caps, returns vector of 0s and 1s:
    l for caps 0 for non caps
    LNApos =find(isCaps);  % says that the isLNA[nth, mth, oth...] letters in
duplexes are LNAs

    %Fll up the first row with LNAs
    for k=1:length(LNApos)  % for the number of LNA in the sequence, what
        letter is it?  --> Fill up the vectors in LNA

            if duplexes{i}(LNApos(k))=='A'
                S(i,1)=S(i,1)+1;  % add one point to the A spot
            elseif duplexes{i}(LNApos(k))=='T'

            end

        end

    end

end
function dnaStack = dnaStack (duplexes)
% this function returns the stacking matrix S for a set of duplexes written
% as [oligo(i)1     oligo(i) 2
%     oligo(i+1)1   oligo(i+1) 2] and so on. We have this type of
% information already in the thermodynamic collection excel
%
% Rows of the stacking matrix will be the number of NNs of the type below
% for each sequence
% [aa/tt at/ta ta/at ca/gt gt/ca ct/ga ga/ct cg/cg gc/cg cc
% DGINIT]
%
% EXAMPLE SET OF DUMPLES BELOW
%
%duplexes= {
%'ttcatagccgt'  'acggctatgaa'
%'tgccgataagt'  'acttatccgca'
%'ctaabgatgc'  'gcctctggtag'
%'ctagcattcc'  'ggaatgctgtag'
%'ctaaatgccgc'  'gcctctgccag'
%'gtatcaagtc'  'agatcctgatac'
%'ggacctcgac'  'gtctgaggtcc'
%'cctgcatgac'  'gtcatcgcag'
%'ggaacaagatgc' 'gcctctgccag'
%'acgaccagagttacag' 'ctgtaactctggtcgt'}
%
% Make a matrix with i rows (same as duplexes) and 11 columns for the
% parameters
dnaStack=zeros(length(duplexes(:,1)),13); % Empty Stacking matrix

%% Fill in the stacking matrix for the NN params (first 10 columns)

for i=1:length(dnaStack(:,1)) % for each sequence
  seq=duplexes{i,1}; % takes oligo 1 from the row i
  for j=1:length(seq)-1 % our frame is doublets
    doubleti=seq(j:j+1); % this is a doublet string frame (Starts with first two letters)

    if strcmpi(doubleti,'aa')|strcmpi(doubleti,'tt')
      dnaStack(i,1)=dnaStack(i,1)+1; % add one more of this double type
    end
    % similarly:
    if strcmpi(doubleti,'at')|strcmpi(doubleti,'ta')
      dnaStack(i,2)=dnaStack(i,2)+1; % add one more of this double type
    end
    if strcmpi(doubleti,'ga')|strcmpi(doubleti,'cg')
      dnaStack(i,7)=dnaStack(i,7)+1; % add one more of this double type
    end
    if strcmpi(doubleti,'gc')|strcmpi(doubleti,'cg')
      dnaStack(i,8)=dnaStack(i,8)+1; % add one more of this double type
    end

    if strcmpi(doubleti,'ca')|strcmpi(doubleti,'tg')
      dnaStack(i,4)=dnaStack(i,4)+1; % add one more of this double type
    end
    if strcmpi(doubleti,'ct')|strcmpi(doubleti,'ag')
      dnaStack(i,6)=dnaStack(i,6)+1; % add one more of this double type
    end
    if strcmpi(doubleti,'ga')|strcmpi(doubleti,'tc')
      dnaStack(i,7)=dnaStack(i,7)+1; % add one more of this double type
    end
    if strcmpi(doubleti,'cg')|strcmpi(doubleti,'gc')
      dnaStack(i,8)=dnaStack(i,8)+1; % add one more of this double type
    end
if strcmpi(doubleti,'gg')|strcmpi(doubleti,'cc')
    dnaStack(i,10)=dnaStack(i,10)+1; % add one more of this double type
end
end
end

%% Fill In Initiation Parameters

for k=1:length(dnaStack(:,1))
    seq=duplexes{k,1}; % takes oligo 1 from the row i

    %Check the 5' end for terminal BPs
    term5=seq(1); % gets the terminal 5' base
    if strcmpi(term5,'a')|strcmpi(term5,'t')
        dnaStack(k,12)=dnaStack(k,12)+1; % is this a terminal a-t bp?
    end

    if strcmpi(term5,'c')|strcmpi(term5,'g')
        dnaStack(k,11)=dnaStack(k,11)+1; % is this a c-g bp??
    end

    term3=seq(length(seq)); % gets the terminal 3' base
    if strcmpi(term3,'a')|strcmpi(term3,'t')
        dnaStack(k,12)=dnaStack(k,12)+1; % is this a terminal a-t bp?
    end

    if strcmpi(term3,'c')|strcmpi(term3,'g')
        dnaStack(k,11)=dnaStack(k,11)+1; % is this a c-g bp??
    end
end

%% Symmetry Parameter

for l=1:length(dnaStack(:,1))
    seq=duplexes{l,1}; % get oligol
    seqcomp=cell(1,length(seq)); % start with an empty complement string

    % gets the 5'-3' complement of this string
    for m=1:length(seq)
        base=seq(m); % gets the base starting at the 5' side
        if strcmpi(base,'a')
seqcomp{length(seq)+1-m}='t';
elseif strcmpi(base,'t')
    seqcomp{length(seq)+1-m}='a';
elseif strcmpi(base,'g')
    seqcomp{length(seq)+1-m}='c';
elseif strcmpi(base,'c')
    seqcomp{length(seq)+1-m}='g';
end
end

seqcomp=transpose(char(seqcomp));% squish the cll array into an array
then transpose it

if strcmpi(seq,seqcomp) % checks for self complementary
    dnaStack(l,13)=1; % if self complementary it sets this to 1
end
end
end
end

function J = numJacobian(func,xdata,a)
% This function evaluates the numerical jacobian Ndatapoints x M parameters
% ex: J (i,j)= dFunci/dxj
% func is a function handle,
% data is input as y,x
% a is the parameter list in the form of a column vector of the parameters
% define constants
da=1E-5; % this is our step size for our parameters

% Make the Jacobian matrix
J=zeros(length(xdata{1}(:,1)),length(a)); % make an NxM matrix for J

for i =1:length(J(1,:)) % for every column in J
% every row we need to step only the parameter corresponding to that row
aFstep=a;
aRstep=a;
aFstep(i)=aFstep(i)+da; % step only this parameter
aRstep(i)=aRstep(i)-da;

% Using a central differencing approximation
J(:,i) = (func(xdata,aFstep)-func(xdata,aRstep))./(2*da);
end

% J
% pause

End
Appendix B

ONLINE INTERACTIVE MODEL FOR PREDICTING DNA MELTING THERMODYNAMICS.

Model predictions for the melting temperature, $T_m$, of short B-form DNA duplexes are often used by biologists, biotechnologists, or researchers to design oligonucleotide primers and probes used in PCR and PCR based technologies. Since the end users of such models often lack computer programming expertise, an online version of our previously reported molecular thermodynamic model was developed to predict $T_m$s of short B-form DNA oligonucleotide duplexes based on user-input DNA sequences and solution conditions. The model interface now joins other notable model interfaces (e.g. IDT, http://biophysics.idtdna.com/) in accurately predicting DNA melting thermodynamics.

The single base thermodynamic (SBT) model developed in our lab is capable of accurately predicting the melting thermodynamics, including $T_m$, of short (typically 8-22 base-pairs) B-form DNA duplexes. The model can provide accurate predictions for pure DNA duplexes, DNA duplexes containing any pattern of one or more locked nucleic acid (LNA) substitutions in a single strand, as well as duplexes containing one or two independent DNA:DNA or LNA:DNA mismatches[1, 105, 136]. For use in PCR, the model contains inputs accounting for PCR relevant solution conditions (i.e. low Na$^+$, high Mg$^{2+}$) and oligonucleotide labeling (i.e. fluorophores and quenchers used in probes) that may alter its prediction of duplex stability[253]).

B.1 Description of R / Shiny

Open source R Studio (V.3.2.1) software and R Shiny online interface was selected to implement the model. The R Project was developed at Bell Laboratories by John Chambers as an open source statistical computing and data analysis language which, by design, eases data storage, plotting, and analysis. A distinguishing feature of R is its extensibility whereby users
can develop computing packages and make them available to other users[254]. For example, to recognize and convert a set of characters representing DNA or LNA bases into a quantified number of parameterized base-pair doublets, modular functions from both the basic R package and “stringr” text handling package were used.

In a recent advancement, R Shiny software now provides a link between typical website coding languages (hypertext mark-up language, HTML/ cascading style sheet format, CSS) and R. Specifically, the R Shiny software allows the developer to program, in R programming language, the constituent functions of a working web page: a user interface function and a server function, which are translated to HTML and CSS in the background. Typically, both functions are written in R studio and the user’s web browser runs the program files in tandem within the R Shiny platform.

Within the user interface function, the developer may choose from a list of standard but adaptable inputs (e.g. check boxes, drop numerical input boxes) and input data dependent outputs (e.g. tables, plots) called widgets. Although the code for the widgets is written sequentially in the script, it is typically written in a nested fashion allowing for the widgets to appear, to the user, inset within a set of flexible panels.

To generate the outputs (e.g. thermodynamic data), the server function component calls user-defined R files to execute functions based on data input by the user on the interface. In some instances, these functions may be used to update a list of inputs available to the user (e.g. possible mismatched base substitutions). This dissertation uses the R shiny interface and the R studio code to enable biotechnologists and researchers to predict, using the SBT model, the melting temperatures of pure DNA and LNA modified duplexes. The model application package is divided into two files: an app.R file containing the user interface and server code and a helpers.R file containing all the functions necessary to support the prediction of the melting thermodynamics of a user-defined DNA duplex. The discussion below summarizes the workflow of the user interface and the treatment of inputs and outputs within the computation of melting thermodynamics.
B.2 User Interface

Figure-B1 shows a simplified version of the online model application user interface. Ultimately within the application, the function *tmpred.R* within *helpers.R* seeks to evaluate the variables in the SBT model defined by equations (1-6) to (1-8) above by using the sequence and solution information input by the user. The minimum user input consists of a DNA sequence (9.0), written 5’ to 3’, an oligonucleotide concentration (1.0), and a Na⁺ concentration (5.0) and the minimum output of the model is $T_m$, $\Delta H(T_m)$, $\Delta S(T_m)$ (11.0). The user may also input a target concentration (2.0), choose to add a fluorophore (3.0) or quencher dye (4.0), or change the solution conditions (*i.e.* concentration of mono-valent salts like Na⁺ and K⁺ (5.0), divalent salts like Mg²⁺ (6.0), and di-nucleoside triphosphates (dNTPs) (7.0)) from their default values. Pressing the “Analyze” button (8.0) activates the model and the melting thermodynamics are only fully updated when it is pressed again after making changes to the inputs above. The remainder of section B.2 describes in detail the syntax for entering the information in boxes 1.0, 2.0, and 9.0 to 10.12.
Figure B-1 A simplified version of the user interface: essential manipulation input objects (boxed, listed 1.0 through 10.1 in bold) are used to compute output data objects (boxed, listed 11.0 in bold). Primer and target concentrations, 5’ and 3’ modifications, salt conditions, and the sequence inputs are used to compute melting thermodynamics output. Mismatches can be introduced by checking the “Mismatch?” box and selecting a mismatched base.

The oligonucleotide (1.0) and target (2.0) concentrations have slightly different meanings. The oligonucleotide (1.0) refers to both a probe or primer or a self complementary duplex when the target concentration (2.0) is set to 0; in this case the equilibrium constant, \( K \), in the SBT model equation above is equal to the total strand concentration \( C_T/2 \) assuming a pseudo-first order reaction unless the duplex is self complementary in which case \( K = C_T \) where for both cases \( C_T \) is equal to the input in box 1.0. If 2.0 is not zero, \( K \) is equal to \( C_A - C_B / 2 \) where subscripts A and B are the concentrations of the oligonucleotide (1.0) or target (2.0) depending on their value where \( C_A > C_B \). If the concentrations A and B are equal, \( K \) is set to \( (C_A + C_B = C_T)/4 \). Self complementary duplexes can be input as equimolar. For example if for a 50 \( \mu \)M total duplex is
input as 25 μM oligonucleotide (1.0) and 25μM target strands (2.0). The program protocol will check for self complementarity and apply the conditions above.

The sequence input (9.0) syntax allows upper or lower case DNA bases representing adenine $a$, thymine $t$, guanine $g$, and cytosine $c$, or LNA analogs denoted by a single “+” followed by a base (e.g. $+c$ or $+C$). Currently, the only chemical modifications supported at this time are LNAs. During sequence input, the option to select and input mismatches (10.0) is available as a check box. When checked, the conditional panel (dashed rectangle in Figure B-1) appears with the sequence printed opposite a series of drop down select boxes with their default values, one for each base in the input sequence, set to that of the complementary base (10.1-10.12). Selecting a DNA base other than the default will constitute a mismatched DNA base and will change the computation of $T_m$ but will also trigger the computation of a $\Delta T_m$, the difference in $T_m$ between the mismatched and perfectly complementary duplex. Before and after the “Analyze” button (8.0) is pressed, $\Delta H_{f/q}$, $\Delta S_{f/q}$ and $\Delta S_{salt}$ are considered constant.

B.3 Computation of Melting Thermodynamics

Once the “Analyze” button (8.0) is pressed, the melting thermodynamics are computed. A function `tmpred.R` takes in the sense strand oligonucleotide (9.0) and the resulting mismatched oligonucleotide defined by the series of bases in the conditional panel; if the mismatch checkbox (10.0) is not selected, a function `compgen.R` simply computes the complement to the sense strand oligonucleotide. The inputs (1.0) through (7.0) are used to evaluate $\Delta H_{f/q}$, $\Delta S_{f/q}$, $\Delta S_{salt}$, and $K$ within `tmpred.R`. In `tmpred.R`, $H_{DNA}^{0}$ and $S_{DNA}^{0}$ are first computed by assuming a perfectly complementary pure DNA duplex (based on the input sequence (9.0)) then by scanning the duplex left to right (5’ to 3’ on the sense strand) in `dnaStack.R`, each nearest neighbour (NN) doublet is counted and the number of each unique doublet is multiplied its corresponding enthalpy and entropy parameter[135]. The number of each LNA type (i.e. $+a$, $+c$, $+t$, $+g$) is also counted, by `lnaStack.R`, and multiplied by each LNAs incremental contribution to incremental entropy [1] to compute a total $\Delta S_{LNA}$ parameter. $\Delta C_p$ is simply computed by $\Delta C_p = 42 * n_{bp}$ where $n_{bp}$ is the number of base pairs in the duplex [136].
Computing the mismatch parameters, $\Delta \Delta G_{DNA-MM}$ and $\Delta \Delta G_{LNA-MM}$, requires neighbouring DNA and LNA sequence context as well sizeable parameter matrices, so both are computed together within the same function `dnamm.R` as shown in the computation routine Figure B-2 below. `dnamm.R` initializes four variables named `ddG_{DNA}`, `ddG_{LNA}`, `ddG_{LNA-3}`, and `ddG_{LNA-5}` corresponding to pure DNA mismatches, directly opposing LNA mismatches, and any mismatch having an LNA at its upstream 3’ or 5’ downstream neighbour on the sense strand respectively. Any time a mismatch of these four classes is observed, their value is updated with the addition of its respective parameter. Following the routine in shown in Figure B-2B, the four incremental Gibbs energy variables are updated either by one of the two functions, `ddGLNA1.R` or `ddG37dna.R`, for collecting the mismatch parameters for LNA (`ddG_{LNA}`) and DNA (`ddG_{DNA}`) respectively or by updating the parameters `ddG_{LNA-3}` and `ddG_{LNA-5}` within `dnamm.R`. The function scans the duplex (Figure B-2A) from 5’ to 3’ (of the sense strand) until it reaches the length of the duplex and the cumulative values of each incremental Gibbs variable is returned to `tmpred.R`.

Finally, once all the parameters have been computed, a root finding algorithm is carried out to evaluate the non-linear SBT model function above to compute $T_m$. The function `uniroot.R` within the package “rootsolve” gets the root of the objective function, defined as the SBT model equation minus $T_m$, through a Newton-Raphson method given an initial guess defined by the SBT model equation (1-6) to (1-8) with $\Delta C_P$ set to 0. An output table summarizing the melting thermodynamics of the input duplex are passed to the `server.R` code and rendered before being output and shown to the user (11.0).

Two screenshots of the “Model” and “About” tabs showing the model user interface and information background information are shown in Figure B-3 and Figure B-4 respectively. The online model is available for use at the hyperlink: [https://masckareem.shinyapps.io/sbtmodel/](https://masckareem.shinyapps.io/sbtmodel/).
**Figure B-2** A) The function *dnamm.R* scans nearest neighbour (NN) doublets to compute the incremental changes in duplex stability. B) Each NN doublet and its previously scanned doublet are analyzed by looking at each base-pair (5’- left and 5’+1-right) in the doublet for LNA:DNA or pure DNA mismatches. Since LNA mismatch parameters are incremental parameters (ref), both *ddGLNA1.R* and *ddG37dna.R* functions will be called to compute the mismatch parameters for the LNA containing mismatch and the isosequential pure DNA mismatch. To avoid double counting LNA mismatches, the previous (i-1) right base-pair is checked for LNA mismatch when the current (i) left mismatch is being evaluated. LNAs in the non-mismatched base-pair are checked for neighbouring LNAs which contribute to the mismatch destabilization of the duplex. The function scans until all NN doublets have been observed.
Figure B-3 Screen shot of the online user interface “Model” tab containing the functional aspect of the model.
**Figure B-4** Screenshot of the online user interface “About” tab containing background information of the model.

### B.4 R Code

```r
app.R
```

```r
library(shiny)
library(data.table)
library(rootSolve)
library(stringr)
source("helpers.R")
```
#Sys.umask() # change permission settings

```
css <- "
#sequence {
  width: 500px;
}
#mismatches {
  white-space: nowrap;
  overflow-x: auto;
  overflow-y: visible;
}
#mismatches .sequence-location {
  white-space: normal;
  display: inline-block;
  width: 34px;
  margin-right: 5px;
  text-align: center;
}
#mismatches .sequence-connector {
  margin-bottom: 5px;
}
#mismatches .form-control {
  text-align:center;
  height: 31px;
  padding: 0;
  overflow-y:visible;
}
.selectize-control.single .selectize-input:after {
  right: 4px !important;
}
.selectize-control.single .selectize-input {
  padding-left: 4px !important;
}
#mismatches .selectize-dropdown {
  float: left;
  position: static;
  margin-bottom: 12px;
"
Welcome to the Single Base Thermodynamic (SBT) model interface. This model predicts the melting thermodynamics for complementary or mismatched DNA duplexes containing one or more LNA substitutions. Changes to solution conditions and modifications such as fluorophores and quencher dyes are also supported. (Scroll down for details.).
#p("DNA concentrations"),
fluidRow( # in the first row of the side panel
column(width=6,
    numericInput("oligoconc",paste0("[oligo]\"U03BC","M"),value=0.25,min=0)
),
column(width=6,offset=0,
    numericInput("targetconc",paste0("[target]\"U03BC","M"), value=0,min=0)
), # ROW 1 FOR OLIGO AND TARGET
#p("Probe Label Options"),
fluidRow(  
column(width=6,
    selectInput("reporter", "Fluorophore", choices=list("N/A","FAM","HEX","TET","TAMARA","Texas Red","Cy3","Cy5"))
),
column(width=6,offset=0,
    selectInput("quencher","Quencher", choices=list("N/A","Dab","IBFQ","QSY7","BHQ1","BHQ2"))
)  )  # ROW 2 for REPORTER AND QUENcher DYE

#p("Salt Conditions"),
numericInput("monosalt","[Na+,K+]mM",value=50,min=0),
numericInput("magnesium", "[Mg2+]mM", value=0,min=0),
numericInput("dntps","[dNTPs]mM", value=0,min=0),

actionButton("start", "Analyze",icon=NULL)

#) # end of well panel  
), #end of control buttons panel (sidebar panel)

mainPanel( width=8,
verticalLayout(  
fluidRow( #top=NULL,left=280, width=1200,
    wellPanel(  
        textInput("oligo1","Sequence (5' to 3')", value=""), # sequence niput
    )  
)  
)  
)
checkboxInput("mismatchCHK", "Mismatch Input?", value=FALSE), # mismatch checkbox button

conditionalPanel(width="100%",
condition = "input.mismatchCHK", # ARG1 IS MISMATCH BOX CHECKED?
uiOutput("mismatches") # this will be a data frame or a set of lists
), # end of conditional Panel 1

conditionalPanel(width="100%",
condition = "input.oligo1==\'diagints\'",
textInput("oligox","input sequence (5' to 3')",value=""),
numericInput("diagonal","Net diagonal interactions (ddGkcal/mol)",value="0")
)

#end of conditional panel 2
# end of well panel
), # end of sequence business (fluidrow)

fluidRow(
  wellPanel(
    p(strong("Thermodynamic Parameters for the Helix-to-Coil Transition")),
    tableOutput("thermotable")
  ), # end of well Panel
#end of Thermodynamic output table panel
)
#end of vertical layout
# end of main panel
), # end of sidebaylayout

fluidRow(
  wellPanel(
    p("Please follow the guide below to get started:")
    ,br()
    ,p("1. Insert your primer, probe, or oligonucleotide sequence (sequence should be written 5’ to 3’ containing pure DNA bases (e.g. a,t,g, or c) or locked nucelic acid (LNA) bases denoted by ‘+’ (e.g. +c).")
    ,p("2. Set your oligonucleotide (oligo) probe or primer concentration (e.g. 0.25 uM). ")
)
3. Set your target concentration: A) For PCR experiments, primers and probes where the target is a template of varying concentration, the target can be left equal to 0 and the primer or probe concentration will be considered in excess. B) For oligonucleotide experiments the target may be specified. If the oligo to be studied is self-complementary, target may be left at 0 or specified as [oligo]/2.

4. Select fluorophores and quencher dyes, if applicable.

5. Select salt conditions (e.g. for PCR, [Na+] = 50 mM, [Mg2+] = 3 mM, [dNTPs] = 0.8 mM).

6. Add mismatches: Select mismatches from the dropdown list once the check box is selected.

Once the settings above have been input, press analyze to predict the helix-to-coil melting thermodynamics for the new sequence.

About

The Haynes' lab, based at the University of British Columbia's Michael Smith Laboratories, has focused on the development of highly accurate and comprehensive models for predicting the melting temperatures and thermodynamics of short natural and chemically modified duplexes by combining statistical mechanics with thermodynamic data libraries – derived from differential scanning calorimetry (DSC) and UV-spectroscopy (UVM). Our nearest neighbour type Single Base Thermodynamic (SBT) model [1-3], shown in its extended form below, can be applied to pure-DNA duplexes or to duplexes in which one strand contains any number or pattern of LNA substitutions. Additionally, the model accounts perturbations to duplex stability arising from DNA:DNA and LNA:DNA mismatches, introduction of common reporter and quencher dyes, and changes in solution conditions, thus making it suitable for the design of dual-labeled hydrolysis probes used in PCR and dPCR. Lastly, recent work focused on understanding and quantifying perturbations to duplex stability arising from LNA substitutions in both strands but was not included in this implementation of the SBT model.
p(paste0("Figure 1. The melting temperature, Tm,MT, of a probe-mutant (MT) target allele duplex, and the difference (",","U0394",")) in Tm relative to the probe-wild type (WT) non-target allele duplex, ",","U0394","Tm,(MT-WT), can be computed using the SBT model above to ensure probe specificity for the target allele.

1 Hughesman CB, Turner RF, Haynes C. Correcting for heat capacity and 5'-ta type terminal nearest neighbors improves prediction of dna melting temperatures using nearest-neighbor thermodynamic models. Biochemistry. 2011; 50: 2642-2649

2 Hughesman CB, Turner RF, Haynes CA. Role of the heat capacity change in understanding and modeling melting thermodynamics of complementary duplexes containing standard and nucleobase-modified lna. Biochemistry. 2011; 50: 5354-5368


#SERVER RULES

#1.#save outputs as output$ELEMENTNAME (e.g. output$hist) = #code
#2. output should be rendered (whole list for most objects)
#2.1 renderPlot({
#code
#code
#hist(rnorm(input$num))
#code
#code based on other inputs!
#
#})

#3.input$elementname (eg. slider 1 is input$num)
server=function(input,output){

output$mismatches= renderUI({# this must be the conditional input for mismatches
  if(identical(input$oligo1,'diagints')){
    oligo1=input$oligox
  } else {
    oligo1=input$oligo1
  }

  sequence <- toupper(oligo1)
  sequence =toupper(seqcheck(sequence)) # check sequence
  sequence= isolate_LNA(sequence) # expand the bad boy and concatenate LNAs

  lapply(1:length(sequence), function(x) {
    div(class = "sequence-location",
    div(sequence[x], class = "sequence-base"),
    div("|", class = "sequence-connector"),
    selectInput(paste0("base_", x),
    label=NULL,multiple=FALSE,selected=toupper(compgen(sequence[x])), selectize=TRUE,
    choices= getComplementList(compgen(sequence[x]))# options=list(maxItems = 1)
  )
  }) # end of making the mismatch inputs

# make this part reactive: dependent on pressing the action button

oligo1 = eventReactive(input$start, {
  if(input$oligo1=='diagints'){
    oligo1=input$oligox
  } else {

  })} // end of server function
oligo1 = input$oligo1
}

oligo2 = eventReactive(input$start, {

if(input$mismatchCHK) {
    # build oligo2 from mismatch input boxes (comes in as 3' to 5')
    oligo2 = NULL
    for(i in 1:nchar(seqcheck(oligo1()))) {
        oligo2 = c(oligo2, input[[paste0("base_", i)]]))
    }

    # squish, turn to lower case, and reverse oligo 2 to read 5' to 3'
    oligo2 = paste(oligo2, collapse = "")
    oligo2 = tolower(oligo2)
    oligo2 = reverse_chars(oligo2)
}
else {
    oligo2 = compgen(seqcheck(oligo1()))
}

}

# end of event reactive for oligo2

output$thermotable = renderTable(

# CODE TO GET THE THERMOTABLE

# Get oligo1 and 2 check input for unexpected expressions

oligo1 = oligo1()
# print(compgen(oligo1))

206
oligo1 = seqcheck(oligo1)
oligo2 = oligo2()
# print(c("oligo1 is ", oligo1))
# print(oligo2)

# put together the salt matrix [Na/K+, Mg2+, dNTPs] and fluorescent and diagonal interactions
salt = c(input$monosalt, input$magnesium, input$dntps)
fq = c(input$reporter, input$quencher)
if(is.null(input$diagonal)){
  diagonal = 0
} else {
  diagonal = input$diagonal
}
# The computation of the rest depends on the analyze button
thermtable = tmpred(oligo1, oligo2, salt, fq, input$oligoconc, input$targetconc, diagonal)

# end of render table

} # end of set server

##########################################################################
####
shinyApp(ui=ui, server=server) # combine both

Helpers.R

compgen = function(oligo1, ...) {

# This function returns the PURE DNA complement of any single strand (sequence) handled as a string "atgctgctgctc" written 5'-3'

#OLIGO1 MUST BE WRITTEN 5'-3'

#error checking example

#oligo1="atgcgcat"
#print(length(oligo1)) Check
#print(oligo1)

# remove LNAS from sequence RETURNS ONLY THE PURE DNA COMPLEMENT

oligo1=gsub(pattern="\d+", "", oligo1)
oligo1=tolower(oligo1)
#print(oligo1) #check

#expand the sequence

sequence=substring(oligo1, 1:nchar(oligo1), 1:nchar(oligo1)) # turns the string into a character vector eg. [1] "a" "t" g" "g" "g" "c"

#print(sequence) CHECK
complement=c(rep(0,length(sequence))) # make empty vector the same length as complement

## FILL THE VECTOR IN BACKWARDS SO WE DONT NEED TO FLIP IT AFTER
#print(length(sequence)) CHECK

for (i in 1:length(sequence)) {

  #print(i) CHECK
  base=sequence[i]

  if (identical(base,"a"){
    complement[length(complement)-i+1]="t"
  }
  else if (identical (base,"t"){
    complement[length(complement)-i+1]="a"
  }

  else if (identical(base,"g"){
    complement[length(complement)-i+1]="c"
  }
else if (identical(base,"c")) {
    complement[length(complement)-i+1]="g"
}

} # end of for loop

## now need to reverse the orders of the letters

## collapse into a single string

# print(complement)# CHECK
complement=paste(complement,sep="", collapse="")
# print(complement)# CHECK
# print(oligo1)

return(complement)

}

DNAstack =function(oligo1) {

# returns a stacking matrix of DNA NNs
#goal is to compute, from a sequence, (e.g. atgccattgc) a stacking matrix [n duplexes, 13] e.g. [0 1 2 0 0 0 0 1 0 1 2 1 1]

# where 13 is the number of 10 NN params and 2 initiation parameters

#seqs=read.csv("teststrings.csv", header=TRUE) # get a set of test strings 17x3 (column 1 and 2 are seqs and 3 is data)

#oligo1=seqs[4,2] # get sequence

#oligo1="cta+++Agc+gac" # example oligo

#DATA TREATMENT (REMOVE CAPITALS AND + (LNAS))

#print(oligo1) #CHECK

oligo1=gsub(pattern="\\", replacement="",x=oligo1) # remove all + values

#print(oligo1) #CHECK

oligo1=tolower(oligo1) # convert all caps to lower case

#print(oligo1) #CHECK

#Initiate DNA Stack

DNAstack=c(rep(0,13)) # make a 0 by 13 row vector (Will fill it later on)
sequence = substring(oligo1, 1:nchar(oligo1), 1:nchar(oligo1)) # turns the string into a character vector eg. [1] "a" "t" "g" "g" "g" "c"

####### We are now focused on the first 10 values in the output vector DNAstack (NN parameter frequencies)

for (i in 1:length(sequence)-1){

  nn = sequence[i:(i+1)]

  if (identical(nn,c("a","a")) | identical(nn,c("t","t"))) {

    DNAstack[1] = DNAstack[1] + 1 } ## NN 1: aa or tt

  else if (identical(nn,c("a","t"))) {


  else if (identical(nn,c("t","a"))) {


  else if (identical(nn,c("c","a")) | identical(nn,c("t","g"))) {


  else if (identical(nn,c("a","c")) | identical(nn,c("t","g"))) {


  else if (identical(nn,c("a","g")) | identical(nn,c("t","c"))) {


  else if (identical(nn,c("t","a")) | identical(nn,c("g","c"))) {


  else if (identical(nn,c("a","t")) | identical(nn,c("g","c"))) {

    DNAstack[8] = DNAstack[8] + 1 } ## NN 8: at or gc

  else if (identical(nn,c("t","g")) | identical(nn,c("a","c"))) {


  else if (identical(nn,c("c","t")) | identical(nn,c("g","a"))) {

    DNAstack[10] = DNAstack[10] + 1 } ## NN 10: ct or ac

else if (identical(nn, c("g", "t")) | identical(nn, c("a", "c"))) {


else if (identical(nn, c("c", "t")) | identical(nn, c("a", "g"))) {


else if (identical(nn, c("g", "a")) | identical(nn, c("t", "c"))) {


else if (identical(nn, c("c", "g"))) {


else if (identical(nn, c("g", "c"))) {

else if (identical(nn,c("g","g")) | identical(nn,c("c","c"))){


} #end for loop (DNAstck[1:10]is done)

##### NOW we must add the initiation parameters (just check the ends for as and ts or gs and cs)

#check the 5' end

if (identical(sequence[1],"a") | identical (sequence[1], "t")){

    DNAstack[12]=DNAstack[12]+1}

else {


}

# Check the 3' end
if (identical(sequence[length(sequence)],"a") | identical (sequence[length(sequence)], "t")){

}

else {


}

###### Just need to check for self-complementarity using my complement generator function
self comp

# if it is self complementary, the answer is 1

if(identical(compgen(oligo1),oligo1)) {

    DNAstack[13]=1

}

###### Return DNA Stack

### 215
LNAstack=function (oligo1) {

  # This function returns the 1x4 stacking matrix for LNAs [A T G C]
  
  # data treatment
  oligo1=tolower(oligo1) # make everything lower case
  
  # expand the duplex into a character vector
  sequence=substring(oligo1,1:nchar(oligo1), 1:nchar(oligo1)) # turns the string into a character vector eg. [1] "c" "t" "a" "A+" "G" "C" "g" "a" "c"

  #print(sequence)
# initialize LNA stack
LNAstack=c(rep(0,4))  # empty LNA stacking matrix i.e. [0 0 0 0]

# find indices of "+" in char vector (sequence)
indicies=grep(pattern="\+", sequence) # in example returns 4 and 8 (2 LNAs) as [4 8]

# print(indicies) check

# Fill LNA stack matrix

for (i in 1:length(indicies)){

  # print(sequence[indicies[i]+1])

  if (identical(sequence[indicies[i]+1],"a")){ # we have an LNA a
    if(indicies[i]==1 | indicies[i] == length(sequence)-1){
      LNAstack[1]=LNAstack[1]+0.5  # If its a terminal LNA, make it only worth 1/2
    } else{ LNAstack[1]=LNAstack[1]+1 # if its an internal LNA make it worth 1
    }
  }

else if (identical(sequence[indicies[i]+1],"t")){ # we have an LNA t
  if(indicies[i]==1 | indicies[i] == length(sequence)-1){
    LNAstack[2]=LNAstack[2]+0.5  # If its a terminal LNA, make it only worth 1/2
  }
}
} else{ LNAstack[2]=LNAstack[2]+1 # if its an internal LNA makit it worth 1
}
}

else if (identical(sequence[indicies[i]+1],"g")) { # we have an LNA g
if(indicies[i]==1 | indicies[i] == length(sequence)-1){
    LNAstack[3]=LNAstack[3]+0.5 # If its a terminal LNA, make it only worth 1/2
} else{ LNAstack[3]=LNAstack[3]+1 # if its an internal LNA makit it worth 1
}
}

else if (identical(sequence[indicies[i]+1],"c")) { # we have an LNA c
if(indicies[i]==1 | indicies[i] == length(sequence)-1){
    LNAstack[4]=LNAstack[4]+0.5 # If its a terminal LNA, make it only worth 1/2
} else{ LNAstack[4]=LNAstack[4]+1 # if its an internal LNA makit it worth 1
}
}

}

} #end for loop
# Return LNA stack

return(LNAstack)

}

ddGLNA1calc=function(oligo1,oligo2){

  #this function returnd the ddG associated with LNA MISMATCHES AT THE SITE OF VARIANCE:
  # ------------------- X -------------------
  # --------------------- L -------------------

  #oligo1=c("c","+c") # written 5' to 3' Example inputs should give:
  #oligo2=c("g","t") # written 3' to 5

  #get data for lna parameters

  lnaparams=read.csv("./data/lnammparams.csv", header=TRUE, sep="",
  )

#print(lnaparams)

#initially set ddGLNAparam=a
ddGLNAparam=0

#print(lnaparams) check

if(mismatch(oligo1[1],oligo2[1])>1){ #we have an LNA mismatch at bp 1
    #ONLY OLIGO 1 CAN CONTAIN LNAs so we should get the indices of it

    #get index1 (Type of LNA)

    if(oligo1[1]=="+a"){
        ind1=1
    } else if (oligo1[1]=="+t"){
        ind1=2
    } else if (oligo1[1]=="+g"){
        ind1=3
    } else if (oligo1[1]=="+c"){
        ind1=4
    }
}
# get Index 2 (type of mismatched DNA basepair)

if(oligo2[1]="a"){
    ind2=1
} else if (oligo2[1]="t"){
    ind2=2
} else if (oligo2[1]="g"){
    ind2=3
} else if (oligo2[1]="c"){
    ind2=4
}

# print(ind1)
# print(ind2)

mmparamrow=(ind1-1)*4+ind2

ddGLNAparam=lnaparams[mmparamrow,1]

} # check 2nd basepair for mismatch (above)
if(mismatch(oligo1[2],oligo2[2])>1){
    #print("right")
    #get index1 (Type of LNA)
    if(oligo1[2]=="+a"){
        ind1=1
    } else if (oligo1[2]=="+t"){
        ind1=2
    } else if (oligo1[2]=="+g"){
        ind1=3
    } else if (oligo1[2]=="+c"){
        ind1=4
    }

    #get Index 2 (type of mismatched DNA basepair)
    if(oligo2[2]=="a"){
        ind2=1
    } else if (oligo2[2]=="t"){
        ind2=2
    }
} else if (oligo2[2]=="g"){
    ind2=3
} else if (oligo2[2]=="c"){
    ind2=4
}

#print(ind1)
#print(ind2)

mmparamrow=(ind1-1)*4+ind2

ddGLNAparam=lnaparams[mmparamrow,1]

} # check 1st basepair for mismatch (above)

#print(ddGLNAparam)
return(ddGLNAparam)

} # Calls params
dyes=function(reporter, quencher) {

# reporter = "FAM"
# quencher = "BHQ1"

# get the relevant parameters in two files when a reporter string and a quencher string are given

reporterParams = read.csv("./data/reporters.csv", sep = ",", header = TRUE) # these are saved as data.frames

quencherParams = read.csv("./data/quenchers.csv", sep = ",", header = TRUE) # these are saved as data.frames

# print(reporterParams) # check
# print(quencherParams) # check

# get your parameters

# fluorophore

for (i in 1:length(reporterParams[,1])) {

#print(as.character(reporterParams[,1])) check

if(identical(reporter,as.character(reporterParams[,1]))) {  # if we match the string index

    #print("found reporter") check

    ddHrep=reporterParams[,2]
    ddSrep=reporterParams[,3]

}


#quencher

for (i in 1:length(quencherParams[,1])){

    if(identical(quencher,as.character(quencherParams[,1]))) {  # if we match the string index

        ddHquench=quencherParams[,2]
        ddSqquench=quencherParams[,3]

    }
}

225
return(c(ddHrep, ddSrep, ddHquench, ddSquench))

} # calls params

dnamm =function(oligo1, oligo2){

# this function returns the incremental gibbs energy for a dna mismatch
# OLIGO 1 and OLIGO 2 Must be handed in 5' to 3' with LNAs only allowed in the sense strand (OLIGO1)

#DATA NEEDED

    dnammparams=read.csv("dnammparams.csv",header=TRUE, sep=",") # get the mismatch paamters in kcal/mol

#example duplex shoudl give (1.58, 0, +0.54, 0)
#oligo1="caaggt+tggaa" #written 5' to 3'
#oligo2="ttccatcttgt" #perfect match is "ttcaaccttg"

#Inititate all parameters at 0 then update in loops

   ddG37dna=0
   ddGLNA=0
   ddGLNA3=0
   ddGLNA5=0
# SEQUENCE CONDITIONING START

# expand both oligos into character vectors
oligo1=substring(oligo1,1:nchar(oligo1), 1:nchar(oligo1)) # can have LNAs (+)
oligo2=substring(oligo2,1:nchar(oligo2),1:nchar(oligo2)) # cannot have LNAs

# need to group oligo1 into bases where + is conjoined to its base
# so the length of the two vectors is the same and we can work on it

oligo1mod=NULL # create NULL (empty) vector

# initial loop params
i=1

while (i <= length(oligo1)){

if(oligo1[i]=="+"){  
    plusLNA=c(oligo1[i],oligo1[i+1])
    oligo1mod=c(oligo1mod,paste(plusLNA,sep="",collapse=""))

    i=i+2
}
else {
    oligo1mod=c(oligo1mod,oligo1[i])
    i=i+1
}

} #CONCATENATE "+", "c" to "+c" when applicable so we can scan better

#print("made it to check 1")
#print(oligo1)
#print(oligo1mod)
#print(oligo2)

# SEQUENCE CONDITIONING END

# now the oligos are the same length, we can scan for mismatches

for (i in 1:(length(oligo1mod)-1)){ # scan in doublets (n-1 of them in a duplex of length n)

    nnoligo1=c(oligo1mod[i],oligo1mod[i+1]) #get the first two bases in the sense strand (e.g. "a" "c")
    nnoligo2=c(oligo2[length(oligo2)+1-i],oligo2[length(oligo2)+1-(i+1)]) #(e.g if comp "t" "g")
if(i>1){
    nnoligo1prev=c(oligo1mod[i-1],oligo1mod[i]) #get the first two bases in the sense strand (e.g. "a" "c")
    nnoligo2prev=c(oligo2[length(oligo2)+1-i-1],oligo2[length(oligo2)+1-(i+1)-1]) #(e.g if comp "t" "g")
} # need to know the previous doublet to make sure we don't double count LNAs

for (j in 1:2){ #check each bp within a mismatched doublet

    #mismatch function gives: 2 for LNA mm, 1 for DNA mm, and 0 for PM

    #print("made it to check 2- after mod for loop")

    print(j)
    #print(mismatch(nnoligo1[j],nnoligo2[j]))

    if (mismatch(nnoligo1[j],nnoligo2[j])>0){    # we have a DNA mismatch   #### NEED TO DO THIS FOR LNA MISMATCH TOO DON"T FORGET

        #pretend that the doublet not in question is a perfect match
        if (j==1){ # mismatch is in left BP
            print(j+50)
            #condition the base pair neighbour in case of tandem mismatch to pretend the second mismatch (right bp) doesn't exist right now
nnoligo2perfect=c(nnoligo2[1],compgen(nnoligo1[2]))  # make the RIGHT BP a perfect match (OFTEN THIS WILL =NNOLIGO2) for single MM

#now we have on the sense strand: nnoligo1 "x" "y"
# and on the perfect nnoligo2perfect have "mm" "ycomp"
#just need to get the mismatch parameter now

# now we need to get ddG based on this new parameter

#print(nnoligo1)
#print(nnoligo2)
#print(nnoligo2perfect)

ddG37dna=ddG37dna+ddGdnacalc(nnoligo1,nnoligo2perfect) # add ddG37dna for DNA if it is needed
#print( ddGdnacalc(nnoligo1,nnoligo2perfect))
#print("mismatch in bp 1 above")
#print(ddG37dna)

} else if (j==2) { #mismatch is in right hand bp

print(j+50)

#condition the base pair neighbour in case of tandem mismatch to pretend the second mismatch (LEFT bp) doesn't exist right now
nnoligo2perfect = c(compgen(nnoligo1[1]), nnoligo2[2])  # make the LEFT BP a perfect match (OFTEN THIS WILL =NNOLIGO2) for single MM

#now we have on the sense strand: nnoligo1 "x" "y"
# and on the perfect nnoligo2perfect have "xcomp" "mm"
#just need to get the mismatch parameter now

#print(nnoligo1)
#print(nnoligo2)
#print(nnoligo2perfect)

ddG37dna = ddG37dna + ddGdnacalc(nnoligo1, nnoligo2perfect)  # add ddG37dna for DNA if it is needed
#print(ddG37dna)

}  
}  # compute ddG37dna for DNA mismatch (have to do this for LNA as well)

#print("made it to check 3 after DNA mm")
print(j+10)

if (mismatch(nnoligo1[j], nnoligo2[j]) > 1) {  # in the case of an LNA mismatch, compute ddGLNA

#pretend that the doublet not in question is a perfect match

231
if (j==1){ # LNA mismatch is in left BP

    # NEED TO MAKE SURE WE DON'T DOUBLE COUNT LNA MisMATCHES by checking last doublet for right hand LNA mismatch
    if(mismatch(nnoligo1prev[2],nnoligo2prev[2])<=1){

        # condition the base pair neighbour in case of tandem mismatch to pretend the second mismatch (right bp) doesn't exist right now
        nnoligo2perfect=c(nnoligo2[1],compgen(nnoligo1[2])) # make the RIGHT BP a perfect match (OFTEN THIS WILL =NNOLIGO2) for single MM

        # now we have on the sense strand: nnoligo1 "x" "y"
        # and on the perfect nnoligo2perfect have "mm" "ycomp"
        # just need to get the mismatch parameter now

        # now we need to get ddG based on this new parameter

        #print(nnoligo1)
        #print(nnoligo2)
        #print(nnoligo2perfect)
ddGLNA=ddGLNA+ddGLNA1calc(nnoligo1,nnoligo2perfect) \# add ddG37LNA for LNA if it is needed

}\# If the previous RH MM of the last doublet was not an LNA mismatch the LNA can be counted

}

else if (j==2){ \#mismatch is in right hand bp

    \#condition the base pair neighbour in case of tandem mismatch to pretend the second mismatch (LEFT bp) doesn't exist right now

    nnoligo2perfect=c(compgen(nnoligo1[1]),nnoligo2[2]) \# make the LEFT BP a perfect match (OFTEN THIS WILL =NNOLIGO2) for single MM

    \#now we have on the sense strand: nnoligo1 "x" "y"
    \# and on the perfect nnoligo2perfect have "xcomp" "mm"
    \#just need to get the mismatch parameter now

    \#print(nnoligo1)
    \#print(nnoligo2)
    \#print(nnoligo2perfect)

    ddGLNA=ddGLNA+ddGLNA1calc(nnoligo1,nnoligo2perfect) \# add ddG37dna for DNA if it is needed
compute ddGLNA for LNA mismatch

# check for LNAS in the neighbouring mismatches

if(mismatch(nnoligo1[j],nnoligo2[j])>0) { # if there's a mismatch of any kind (LNA or DNA) in this basepair, check the neighbour for LNAs

if(j==2) {print(nchar(nnoligo1[1])+90)} # 2 is for LNA (WE NEED TO TAKE ACTION) 1 is for DNA (NO 5' or 3; LNA MM Parameter is needed)

if(j==1) { # 2 is for LNA (WE NEED TO TAKE ACTION) 1 is for DNA (NO 5' or 3; LNA MM Parameter is needed)

if (j==1 & nchar(nnoligo1[2])>1) { # if there's a mismatch in the left hand bp and there's an LNA in the next base over (ON OLIGO1) 3' SIDE

print("5Ld3/5dxd3")

# print(nchar(nnoligo1[2]))

ddGLNA3=ddGLNA3+-0.55 #kcal/mol # doesn't matter what the mismatch type is (LNA or DNA), add the parameter if an LNA is 3' of it

}

else if (j==2 & nchar(nnoligo1[1])>1) { # NOW looking at the case where the mismatch is in the right hand base. need to see if we add 5' param or not
```python
print("5Ld3/5dxd3")

if(nchar(nnoligo1[2])>1){  # we have the B-M situation (nnoligo1[2] is the M where M is defined as +m)
    ddGLNA5=ddGLNA5+0 # don't apply the parameter
}

if(nchar(nnoligo1[2])==1){  # we have a B-m situation
    ddGLNA5=ddGLNA5 +0.17 # apply the parameter
}
```

}  # outer loop doublet

# print(ddG37dna) # check the amount of DDG added in each doublet

}  # end of scanning for loop
#check all parameters

print(paste0("ddG37dna is ",ddG37dna))
print(paste0("ddGLNA is ", ddGLNA))
print(paste0("ddGLNA3 is ", ddGLNA3))
print(paste0("ddGLNA5 is ", ddGLNA5))

params=c(-ddG37dna,ddGLNA,ddGLNA3,ddGLNA5)

# print(params)
return(params)

}

ddGdnacalc=function(oligo1,oligo2){

# this function takes in two pairs of vectors: a sense strand vector (e.g. 5' ("a", "t")) and and anti sense mismatch vector (e.g. 3' "a","a")

# needs to get ddG from the 16x4 matrix of MMs and PMS

######################################## the NN DOUBLET coming in here can ONLY HAVE ONE MISMATCH  ####################################
# this uses 3 indicies: for (CX/GY) where the logic is the group is in brackets, base is X and mismatch is Y

# Index 1: group
# Index 2: base
# Index 3: Mismatch

#print("check 0")

#oligo1=c("c","c") # written 5' to 3' Example inputs should give 2.65 ddG37
#oligo2=c("a","g") # written 3' to 5'

# Load dna mismatch parameters
dnammparams=read.csv("./data/dnammparams.csv", header=TRUE, sep="","")
#print(dnammparams)

#REMOVE ALL LNAS HERE

oligo1=gsub(pattern="\\+","",oligo1)
#print(oligo1)

#print("check 1")
ddG37param=0 # start parameter at 0 (NO mismatch)

# check for right hand or left hand mismatch

if (mismatch(oligo1[2],oligo2[2])==1){ # DNA mismatch in right bp

    #compute index 1 (group)

    if(oligo1[1]=="a"){
        ind1=3
    } else if (oligo1[1]=="t"){
        ind1=4
    } else if (oligo1[1]=="g"){
        ind1=1
    } else if (oligo1[1]=="c"){
        ind1=2
    }

    #compute index 2 (X)

    if(oligo1[2]=="a"){
        ind2=1
    } else if (oligo1[2]=="t"){
        ind2=2
    }
ind2=4
} else if (oligo1[2]=="g"){
    ind2=3
} else if (oligo1[2]=="c"){
    ind2=2
}

#compute index 3 (Y)
if(oligo2[2]=="a"){
    ind3=1
} else if (oligo2[2]=="t"){
    ind3=4
} else if (oligo2[2]=="g"){
    ind3=3
} else if (oligo2[2]=="c"){
    ind3=2
}

#compute perfect match index 3 (different Y now because PM) # checks the base complement of oligo1 [2]
if(compgen(oligo1[2])=="a"){
    ind3pm=1
} else if (compgen(oligo1[2])=='t') {
    ind3pm=4
} else if (compgen(oligo1[2])=='g') {
    ind3pm=3
} else if (compgen(oligo1[2])=='c') {
    ind3pm=2
}

# retrieve which parameter it is
paramrow=(ind1-1)*4+ind2

dG37mmparam=dnampparams[paramrow,ind3] # get the mismatch parameter from the table
dG37pmparam=dnampparams[paramrow,ind3pm] # get the perfect match parameter
ddG37param=dG37mmparam-dG37pmparam # incremental ddG37

} # for right side base-pair dna mismatch (value of 1)

else if (mismatch(oligo1[1],oligo2[1])==1) { # DNA mismatch in left side bp

    # compute index 1 (group)
if (oligo2[2] == "a") {
    ind1 = 3
} else if (oligo2[2] == "t") {
    ind1 = 4
} else if (oligo2[2] == "g") {
    ind1 = 1
} else if (oligo2[2] == "c") {
    ind1 = 2
}

# compute index 2 (X)
# print(oligo2[1])

if (oligo2[1] == "a") {
    ind2 = 1
} else if (oligo2[1] == "t") {
    ind2 = 4
} else if (oligo2[1] == "g") {
    ind2 = 3
} else if (oligo2[1] == "c") {
    ind2 = 2
}
# compute index 3 (Y) mismatch

if(oligo1[1]="a"){
    ind3=1
} else if (oligo1[1]="t"){
    ind3=4
} else if (oligo1[1]="g"){
    ind3=3
} else if (oligo1[1]="c"){
    ind3=2
}

# compute perfect match index 3 (different Y now because PM) # checks the base complement of oligo2[1]

if(compgen(oligo1[1])="a"){
    ind3pm=1
} else if (compgen(oligo1[1])="t"){
    ind3pm=4
} else if (compgen(oligo1[1])="g"){
    ind3pm=3
} else if (compgen(oligo1[1])="c"){
ind3pm=2
}

# print(ind1)
# print(ind2)
# print(ind3)

# retrieve which parameter it is
mmparamrow=(ind1-1)*4+ind2 # for mismatch only
pmparamrow=(ind1-1)*4+ind3pm

dG37mmparam=dnammparams[mmparamrow,ind3] # get the mismatch parameter from the table
dG37pmparam=dnammparams[pmparamrow,ind3] # get the perfect match parameter
ddG37param=dG37mmparam-dG37pmparam # incremental ddG37

# print(dG37mmparam)
# print(dG37pmparam)

} # for left side basepair dna mismatch (value of 1)
getComplementList=function (base) {

  #return a list of choices that include all except the complementary base
  #input can be +A, +a, a, or A
  #base="c"

  #need to condition LNA bases to get the actual base
  if (nchar(base)>1) {
    base=substring(base,2)
  }

  #turn to upper case
  base=toupper(base)

  #generate list

  if (base=="A") {

  }

}
compList=list("A","T","G","C")

}
else if (base=="T"){
    compList=list("T","A","G","C")
}
else if (base=="G"){
    compList=list("G","A","T","C")
}
else if (base=="C"){
    compList=list("C","A","T","G")
}

return(compList)

}

isolate_LNA=function(oligo){

    #expand oligo

    oligo1=substring(oligo,1:nchar(oligo),1:nchar(oligo)) # expand it into a vector
# initial loop params

i=1

oligo1mod= NULL

while (i <= length(oligo1)){

if(oligo1[i]=="+"){
    plusLNA=c(oligo1[i],oligo1[i+1])
    oligo1mod=c(oligo1mod,paste(plusLNA,sep="",collapse=""))
    i=i+2
}

else {
    oligo1mod=c(oligo1mod,oligo1[i])
    i=i+1
}

} # CONCATENATE "+" , "c" to "+c" when applicable so we can scan better
mismatch=function (base1,base2) {

# find a mismatch of a function

# if LNA is handed in wrong way flip the bases (ONLY one should have the LNA)

if(nchar(base2)>nchar(base1)) {

  holder=base1
  base1=base2
  base2=holder

}

# the bases should come in here in all lower case with LNAs having a + only

#check for LNA or DNA mismatch and tell which one it is

LNAMM=NULL
DNAMM=NULL
# CAN ONLY BE ONE OR THE OTHER

if (nchar(base1)>1) { # we are dealing with an LNA

  # modify base 1 to get rid of the + in front meaning LNA
  base1L=substring(base1,2)
  # print(base1L)

  # get complement for base 1
  if (base1L=="a"){
      comp1="t"
  } else if (base1L=="t"){
      comp1="a"
  } else if (base1L=="g"){
      comp1="c"
  } else if (base1L=="c"){
      comp1="g"
  }

  if(identical(comp1,base2)) { # are the bases complementary?
      LNAMM=FALSE
DNAMM=FALSE
} else {
    LNAMM=TRUE
    DNAMM=FALSE
}

} # end of LNA mismatch check

if (nchar(base1)==1) {

    # get complement for base 1
    if (base1=="a") {
        comp1="t"
    } else if (base1=="t") {
        comp1="a"
    } else if (base1=="g") {
        comp1="c"
    } else if (base1=="c") {
        comp1="g"
    }

}
comp1 ="g"

if(identical(comp1,base2)){ # are the bases complementary?
    DNAMM=FALSE
    LNAMM=FALSE
} else {
    DNAMM=TRUE
    LNAMM=FALSE
}

} # end of DNA mismatch check

############ Output a way to check if an LNA or DNA msiamtch has happened

#print(LNAMM)
#print(DNAMM)

if (LNAMM){
    MMOUT=2
}
} else if (DNAMM) {
    MMOUT = 1
} else {
    MMOUT = 0
}

#return MMout (2 for LNA mm, 1 for DNA mm, and 0 for PM)
return(MMOUT)

reverse_chars=function(string) {

    string_split = strsplit(as.character(string), split = "")
    reversed_split = string_split[[1]][nchar(string):1]
    paste(reversed_split, collapse = "")
}

seqcheck=function(oligo1) {

    # this function is to ensure that only the proper inputs get passed on to the tm pred program.
    # IT ALSO CONDITIONS THE OLIGO TO LOWER CASE
# oligo1 = "+a+a+gc" Example oligo

############### Preliminary Conditioning ###############

# remove spaces
oligo1 = gsub(" ", ",", oligo1)

# put it to lower case
oligo1 = tolower(oligo1)

############### ANY OF THE FOLLOWING SEQUENCE MISTAKES WILL GIVE AN ERROR ###############
nchar(oligo1)

# 0. Make sure there is an input
if (nchar(oligo1) == 0 | is.na(oligo1)) {
  stop("Please enter a sequence written 5' to 3'")
}
# print("passed check 0")

# 1. NON Watson CRICK or + checaters (we have some characters)
nonwc=grep(pattern="^[atgc+ATGC]",oligo1,value=TRUE)

if(length(nonwc)>0){ # the oligo contains non W/C or + characters

stop("Only a,t,g,c bases and + are allowed")
}

#print("passed check 1")

#2. CHECK THAT + is always followed by a letter (a,t,g,c)

oligo1exp=substring(oligo1,1:nchar(oligo1),1:nchar(oligo1))

for (i in 1:(length(oligo1exp))){

if(oligo1exp[i]=="+"){# found a +

# if the next element isn't a letter, give error

#print(oligo1exp[i])

#print(oligo1exp[i+1])

  nextletter=grep(pattern="^[atgcATGC]",oligo1exp[i+1], value=TRUE) # vector with non atgc elements
if(length(nextletter)>0){
  stop("a + must be followed by an a,t,g, or c base")
}

if(is.na(oligo1exp[i+1])){
  stop("a + must be followed by an a,t,g, or c base")
}

# end of for loop

#print("passed check 2")

#######################################################################

#3. Check that the duplex isn't too short (4 bp is shortest allowed _AS PER IDT BIOPHYSICS_)

oligo1short=gsub("\+", "]", oligo1) # get rid of all +s

if(nchar(oligo1short)<4 & nchar(oligo1short)>0){
stop ("sequence is too short")

} #print("passed check 3")

return(oligo1)

} # this function returns the melting temperature of a given duplex knowing:

tmpred=function(oligo1,oligo2,salt,fq,oligoconc,targetconc,diagonal) {

#1. The LNAs and sequence
#2. The salt conditions
#3. The fluorophores and quenchers
#print(oligo1)
#print(oligo2)

#oligo1="cta+Agc+gac" example oligo

#Define Constants
R=1.987 #cal / mol K
Tref=53 +273.15# K
nbp=nchar(gsub(pattern="\+","",oligo1))

################Add In diagonal ddG ###################

ddGdiag=diagonal # get ddG diagonal in kcal/mol

################get pureDNA enthalpy and entropy###################

#keep Santalucia [pure DNA params]

slparams= read.csv("./data/slparams.csv", header=TRUE,sep="," # get teh santalucia parameters

#get stacking matrix for pureDNA
DNAstk=DNASTack(oligo1)

#USE 1997 SL NN Parameters
dHdna=DNAstk %*% slparams[,2] # kcal/mol :matrix multiply first row with stacking matrix
dSdna=DNAstk %*% slparams[,3] # cal/mol K

# print(dSdna)

# print(dHdna) # CHECK

################################get LNA incremental enthalpy and entropy################################

# keep SBT params here

sbtparams=c(-2.3, -3.2, -2.5, -4.8) # these are only ddS params for A, T, G, and C respectively

# get stacking matrix for LNA
LNAstk=LNAstack(oligo1)
# print(LNAstk)

ddHlna=0
ddSlna=LNAstk%*%sbtparams # check to make sure matrix multiplication is working
# print(ddSlna)
get salt condition entropy

salt is given in a vector as: [Na/K+, Mg2+,dNTPs]

if (salt[3]>salt[2]){
  stop("Please input an [Mg2+] that is at least 100% of [dNTPs]")
}

naeq=salt[1]/1000+0.12*sqrt(salt[2]-salt[3])

ddSsalt=0.847*(nbp-1)*log10(naeq)

get fluorophores and quencer thermodynamics

reporter=fq[1] # fron inputs
quencher=fq[2]
fqvec=dyes(reporter,quencher)

ddHfq=sum(fqvec[1],fqvec[3]) #kcal/mol
ddSfq=sum(fqvec[2],fqvec[4]) #cal/molK
get incremental gibbs change for mismatches

mmparms=dnamm(oligo1,oligo2) # THIS DOES ddG for LNA and DNA MISMATCHES and RETURNS A VECTOR

ddGMMs=sum(mmparms) # whats the sum effect of all of these in kcal/mol

compute K and CP

# ct #must be in M units (MULTIPLY BY 1e-6)

#need to copmute K based on self complmentarity: if selfcomp, K=ct, else=ct/4
if (targetconc==0){
    K=oligoconc*10^-6/2

    if(compgen(oligo1)==oligo1){ # if you have a self complementary probe and are inputting data this way
        warning("Self complementary probe/oligo, K=[oligo] (M)")
        K=oligoconc*10^-6
    }
}

# we are meant to do a probe style equation CT/2

# if not we can go into this loop
else if(DNAstk[13]>0){
    K=oligoconc*10^-6
}
else {

if(oligoconc==targetconc){ # for non-probe, non-self comp, equimolar solutions
    K=(oligoconc+targetconc)*10^-6/4
} # for equimolar solutions
else{ #
    if (oligoconc>targetconc){ # for non-probe, non-self comp, solutions where [oligo]>[target]

        K=oligoconc*10^-6-targetconc*10^-6/2
    }
    else { # for non-probe, non-self comp, solutions where [target]>[oligo]

        K=targetconc*10^-6-oligoconc*10^-6/2
    }

} # for non probe, non-equimolar solutions
} # for non-self complementary duplexes

delCP=42*nbp# cal/mol K ---- get CP/bp *length oligo (minus the + signs)

#print(delCP) #Check

# predict Tm

# make initial guess ignoring deltacP

Tminit=(-dHdna*1000+ddGdiag+ddHfq*1000+ddGMMs*1000)/(-dSdna+ddSlna-ddSSalt+ddSfq -R*log(K))

# eliminate cp if T is below 40, set cp to 0
if (Tminit -273.15<40){
    delCP=0
}

# do the same for the perfect match

Tm_complementary_init=(-dHdna*1000+ddHfq*1000)/(-dSdna+ddSlna-ddSSalt+ddSfq -R*log(K))

g=function(Tm){ # this will be the function to set to zero in order to find Tm
    Tm-(-dHdna+ddGdiag+ddHfq+delCP/1000*(Tm-Tref)+ddGMMs)/(-dSdna+ddSlna-ddSSalt+ddSfq -R*log(K)+delCP*log(Tm/Tref))*1000
}

h=function(Tm){ # this will be the function to set to zero in order to find Tm
    Tm-(-dHdna+ddGdiag+ddHfq+delCP/1000*(Tm-Tref))/(dSdna+ddSlna-ddSSalt+ddSfq -R*log(K)+delCP*log(Tm/Tref))*1000
}
```r
Tm = multiroot(g, Tm_init) # get the root and turn to degC units (whole bunch of other crap in there) for the true duplex

Tm_complementary = multiroot(h, Tm_complementary_init) # get the root and other crap for the perfectly matched duplex

#################
Tmout = round(Tm[[1]] - 273.15, 1)
dHout = -dHdna[1, 1] + ddHfq + delCP/1000*(Tmout+273.15-273.15)
dSout = (-dSdna + ddSlna - ddSsalt + ddSfq + delCP*log((Tmout+273.15)/273.15))
Ttmpmout = round(Tm_complementary[[1]] - 273.15, 1)
deltatm = Tmout - Ttmpmout

# Make headers

tmheader = paste0("Tm", " (°C)")
deltatmheader = paste0("ΔTm", " (°C)")
deltahheader = paste0("ΔH", " (kcal/mol)")
deltasheader = paste0("ΔS", " (cal/mol)")
lengthheader = "bp"

# print(dHdna)
# print(dSdna)
# print(ddSlna)
# print(LNAsstk)
```
#print(oligo1)

thermotable=data.frame(tmheader=Tmout,deltatmheader=deltatm,
deltahheader=dHout,deltasheader=dSout,lengthheader=nchar(gsub("\+","",oligo1))))

names(thermotable)=c(tmheader,deltatmheader,deltahheader,deltasheader,lengthheader)

name.width <- max(sapply(names(thermotable), nchar))

row.names(thermotable)=c(oligo1)

format(thermotable, width=name.width, justify = "centre")

#print(deltatmheader)

return(thermotable)

} # calls params