**Kurtis M. Anderson certifies that this is the approved version of the following dissertation:**

# Dithioated Phosphates in DNA Duplex Thermodynamics and Protein-DNA Interactions

**Committee:**

David Gorenstein, Ph.D.,
Mentor

Junji Iwahara, Ph.D.,
Chair

Wlodek Bujalowski, Ph.D.

Kyung Choi, Ph.D.

B. Montgomery Pettitt, Ph.D.

Dean, Graduate School

# Dithioated Phosphates in DNA Duplex Thermodynamics and Protein-DNA Interactions

**by**

**Kurtis M. Anderson, B.S.**

## Thesis

Presented to the Faculty of the Graduate School of

The University of Texas Medical Branch

in Partial Fulfillment

of the Requirements

for the Degree of

## Doctor of Philosophy
## Biochemistry and Molecular Biology

**The University of Texas Medical Branch**

**October, 2014**

## In Dedication

…to my parents, who inspired greatness

…to my wife, who made it all possible

…and to my greyhound, that kept me smiling.

## Acknowledgements

I'd like to acknowledge the guidance and support of my mentor, Dr. David Gorenstein. After our laboratory at UTMB was destroyed by Hurricane Ike, his superb leadership kept the group intact during our transition from UTMB in Galveston to UTHSC in Houston. Despite a change of my research focus from viral infectious diseases to aptamer design, Dr. Gorenstein still provided me with enviable academic freedom to pursue my own research interests, without the burden of grant application-driven science. The result was an ambitious biophysics project that took many years to complete, but his confidence in my abilities never waivered. The success we had during the final years of the project, in terms of manuscripts and grant funding, is a great testament to Gorenstein's mentorship style and I am proud to be his last graduate student as he finishes his long career.

I am also grateful for the influence faculty scientist Dr. David Volk has had on my development as a researcher during my time in the Gorenstein lab. Dr. Volk acted as a day-to-day advisor for me in the lab, helping in supportive roles too numerous to count. I am especially grateful for the countless hours he has spent reviewing and editing my manuscripts, including this verbose dissertation. Other members of our group also deserve acknowledgement for their role in helping train me as a scientist: Drs. Varatharasa Thiviyanathan, Ganesh Lakshmana Rao Lokesh, Sai Gandham, Anoma Somasunderam, Weiguo He, and Hongyu Wang. More directly involved in the project, Dr. Miguel Elizondo-Riojas deserves my gratitude for graciously providing his computational expertise to my research questions. Additionally I'd like to thank Dr. John Ladbury of MDACC for allowing me to use his ITC instruments, as well as providing

useful discussion. Also, Dr. Xianbing Yang of the company AM Biotechnologies, who graciously allowed me to use their facilities and instruments when needed.

I would like to specifically acknowledge two people who were directly involved in my research initiatives in a mentoring capacity. The first is Dr. Richard Owczarzy of the biotech company Integrated DNA Technologies, who was a collaborator on the dithioate melting project. He was willing to lend his brilliant thermodynamics expertise to an inquisitive graduate student in Texas, and together we were able to answer some tough biophysical questions about dithioates. I am also appreciative of my committee chair Dr. Junji Iwahara for his recent interest in dithioated DNA. What initially start as a curiosity over the peculiarities of phosphorodithioates, has thankfully blossomed into a fully funded research division of his lab. We started this collaboration a few years ago, and it by all metrics has been astonishingly successful, producing manuscripts and funding from multiple national grants. This unique collaboration between his and Dr. Gorenstein's labs has brought out some inspiring work from both groups, and I am excited for future prospects of the project.

# Dithioated Phosphates in DNA Duplex Thermodynamics and Protein-DNA Interactions

Publication No._____

Kurtis M. Anderson, Ph.D.

The University of Texas Medical Branch, 2014

Supervisor: David Gorenstein

Phosphorodithioated DNA contains sulfur atoms at both non-bridging phosphoryl oxygen positions. As an achiral chemical modification of the internucleotide linkage it imparts altered biophysical properties to the phosphate moiety, but still retains normal DNA characteristics such as the ability to form duplexes with complementary strands and to be involved in protein-DNA interactions. Sparse information is currently known about the biophysical effects of phosphorodithioation in DNA, and this dissertation addresses dithioated phosphates in three fundamental aspects of DNA research: DNA synthesis, DNA duplex hybridization thermodynamics, and protein-DNA interactions.

The hybridization of two complementary DNA strands into helical duplexes is a well-studied biophysical phenomenon. Normal DNA duplex thermodynamics are predicted from its sequence using unified nearest-neighbor (NN) base pair doublet parameter sets, but currently there is a lack of information for phosphorodithioate modifications. Here a study of 40 different 11-bp duplexes containing all dithioate NNs is presented. Fitting with a single phosphorodithioate parameter revealed the general effect for dithioation: the duplex melting temperature is lowered, the transition free energy becomes less favorable ($\Delta\Delta G^\circ = +0.67$ kcal/mol), the enthalpic term becomes less favorable ($\Delta\Delta H^\circ = +3.1$

kcal/mol), and the entropic term becomes more favorable ($\Delta\Delta S° = +7.7$ cal/mol·K). Significant variation from these values was apparent and depended on sequence context, and further modeling revealed particular importance of the purine/pyrimidine identity of base 3' to the dithioate. Numerous phosphorodithioate difference-value parameters were modeled and the sets presented here are suitable for predicting thermodynamic values of short dithioated DNA duplexes from unified normal DNA NN parameter values.

Protein-DNA interactions are mediated primarily through ion pair contacts between side-chains and the phosphate backbone. The effect of site-specific phosphorodithioation at the contact point with a lysine $NH_3^+$ group was investigated for the HoxD9 homeodomain system using biophysical experiments and nuclear magnetic resonance (NMR). Binding experiments revealed a nearly three-fold increase in duplex affinity upon dithioation of a single phosphate, but isothermal titration calorimetry could not detect an enthalpic difference that would contribute to this apparent lower free energy term. NMR $^{15}N$ relaxation and hydrogen-bond scalar $^{15}N-^{31}P$ J-couplings ($^{h3}J_{NP}$) were then used to investigate the dynamics of the intermolecular ion pairs that form between a lysine $NH_3^+$ groups and the DNA phosphate or phosphorodithioate backbone. Surprisingly, order parameters and correlation times for C−N bond rotation and reorientation of the lysine $NH_3^+$ groups indicated that in general, $NH_3^+$ groups involved in intermolecular ion pairs at the protein−DNA interface are highly dynamic. There is a transition between contact ion pair (CIP) and separated ion pair (SIP) states that occurs on the sub-nanosecond time scale, which should lower the entropic costs of protein-DNA association. Phosphorodithioation was shown to increase the dynamics of the associating lysine $NH_3^+$ group, and the overall increase in entropy (i.e. reorientational + rotational) upon dithioation of the ion pair was estimated to be ~ +0.8 cal/mol·K more favorable. Together with binding data, these suggest that the affinity enhancement observed after phosphorodithioation at the location of Lys57 in HoxD9 is due primarily to entropic enhancement.

# Table of Contents

## List of Tables

## List of Figures

## List of Abbreviations

| | |
|---|---|
| 2SIP | Solvent-separated ion pairs |
| A | Adenine |
| ACN | Acetonitrile |
| ADTT | 3-Amino-1,2,4-dithiazole-5-thione |
| $\alpha$ | Fraction Unfolded |
| AMBER | MD software |
| APS | Ammonium persulfate |
| bp | basepair |
| BTTM | Benzyltriethyl-ammonium tetrathiomolybdate |
| C | Cytosine |
| CD | Circular dichroism |
| cDNA | Complementary DNA |
| CE | Capillary electrophoresis |
| CIP | Contact ion Pair |
| $c_{\mathrm{n}}$ | Heat capacity |
| $c_{\mathrm{p}}$ | Heat capacity at constant pressure |
| $C_{\mathrm{p}}$ | Heat capacity of a body at constant pressure (i.e. $mc_{\mathrm{p}}$) |
| CPG, CpG | Controlled pore glass |
| $C_{\mathrm{T}}$ | Concentration total DNA strands |
| $\delta$ | Inexact differential |
| $d$ | Exact differential |
| D | Duplex |
| DCC | N,N'-Dicyclohexylcarbodiimide |
| DCM | Dichloromethane |
| DDTT | 3H-1,2,4- dithiazole-3-thione |
| DFT | Density functional theory |
| DMSO | Dimethyl sulfoxide |
| DMT | 5'-Dimethoxyltrityl |
| DMT-OFF | DMT is taken off the oligo |
| DMT-ON | DMT is left on the oligo |
| DNA | Deoxyribonucleic acid |
| DtsNH | 1,2,4-Dithiazolidine-3,5-dione |
| $E$ | Total Energy |
| e | Elementary charge |
| EDITH | 3-Ethoxy-1,2,4-dithiazoline-5-one |
| EDTA | Ethylenediaminetetraacetic acid |
| $\epsilon$ | Dielectric constant |
| $\xi$ | Manning's charge density parameter |
| ESI-MS | Electrospray ionization - mass spectrometry |
| $F$ | Helmholtz energy (free) |
| FAM | 6-Carboxyfluorescein |
| FP | Fluorescence polarization |

| | |
|---|---|
| FRET | Fluorescence resonance energy transfer |
| $F_x$ | Force (x-direction) |
| G | Guanine |
| $G$ | Gibbs energy (free) |
| $H$ | Enthalpy |
| HISQC | Heteronuclear in-phase single quantum coherence |
| HMQC | Heteronuclear multiple quantum coherence |
| HoxD9 | Homeodomain HoxD9 |
| HSQC | Heteronuclear single quantum coherence |
| IDT | Integrated DNA Technologies (a biotech company) |
| INEPT | Insensitive nuclei enhanced by polarization transfer |
| INN | Internal nearest neighbor doublet |
| ITC | Isothermal titration calorimetry |
| K | Kelvin |
| $K$ | Kinetic Energy |
| L | Axial length of a polyelectrolyte |
| $\lambda$ | Bjerrum length |
| LC-ESI-MS | Liquid chromatrography - electrospray ionization - mass spectrometry |
| LLC | Limited license corporation |
| LNA | Locked nucleic acid oligonucleotide |
| Lys | Lysine |
| $m$ | Mass |
| MAE | Mean absolute error |
| MD | Molecular dynamics |
| MDACC | MD Anderson Cancer Center |
| MEDITH | 3-methyl-1,2,4-dithiazolin-5-one |
| MMT | 5'-Monomethoxyltrityl |
| mRNA | Messenger RNA |
| MWCO | Molecular weight cutoff |
| $n$ | Molecularity |
| NMR | Nuclear magnetic resonance |
| NN | Nearest neighbor doublet |
| $P$ | Pressure |
| PAGE | Polyacrylamide gel electrophoresis |
| PCR | Polymerase chain reaction |
| PDB | Protein data bank |
| $p_i$ | Momentum |
| PO | Normal phosphate |
| PO-DNA | Normal DNA molecule |
| PS | Monothioated phosphate |
| PS2 | Dithioated phosphate |
| PS2-DNA | Dithioated DNA molecule |
| Pur | Purine |
| Pyr | Pyrimidine |
| q, Q | Heat |
| QC | Quality control |

| | |
|---|---|
| R | Gas constant |
| RMSD | Root mean square deviation |
| RNA | Ribonucleic acid |
| $R_p$ | Monothioate enantiomer |
| RP-HPLC | Reverse phase - high pressure liquid chromatography |
| S | Entropy |
| S, $S_1$, $S_2$ | Single strand DNA |
| SIP | Solvent-shared (separated) ion pair |
| siRNA | Small interfering RNA |
| $S_p$ | Monothioate enantiomer |
| T | Thymine |
| $T$ | Temperature |
| $T_m$ | Melting temperature |
| TACC | Texas advanced computing center |
| TAMRA | 5-Carboxytetramethylrhodamine |
| TBE | Tris/borate/EDTA buffer |
| TCA | Trichloracetic acid |
| TEAA | Triethylammonium acetate |
| TEMED | Tetramethylethylenediamine |
| TETD | Tetraethylthiuram disulfide |
| $\Theta$ | Fraction Folded |
| THF | Tetrahydrofuran |
| TLC | Thin-layer chromatography |
| TsCl | p-Toluenesulfonyl chloride |
| $U$ | Internal energy |
| UTHSC | University of Texas Health Science Center |
| UTMB | University of Texas Medical Branch |
| UV | Ultraviolet |
| $V$ | Potential energy |
| $V$ | Volume |
| $w$ | Work |
| $\xi$ | Manning charge density parameter |
| $x_i$ | Position |

# DISSERTATION AIM AND SIGNIFICANCE

Briefly, the purpose of this dissertation is to determine the biophysical effects of sulfur-for-oxygen substitution in the phosphate backbone of DNA. There are two aims of research: dithioate effects on DNA duplex thermodynamics, and dithioate effects on protein-DNA interactions. The immediate significance of this research is improved knowledge of oligonucleotide chemistry, such as in areas of aptamer design. But from a broader perspective, this work provides knowledge into the general effects of sulfur-for-oxygen substitution in organic macromolecules. The fundamental thermodynamic aspect of this research and focus on intermolecular ion pairing enables these dissertation findings to have implications in areas beyond just short synthetic oligonucleotides.

# CHAPTER 1: DITHIOATED OLIGONUCLEOTIDE PRODUCTION

## 1.1) Historical Development of DNA Synthesis Strategies

The biotechnology revolution that has changed the destiny of mankind was made possible by the ability to synthesize short oligonucleotides. This was no small feat. As the molecule of life, moieties such as a sugar-phosphate backbone and numerous hydroxyl and amine functional groups give nucleic acids a rich chemical character that has kept chemists perplexed with developing synthetic methods for over a 100 years. It's important to reflect upon the accomplishments within this history in order to fully appreciate our modern capabilities in oligonucleotide synthesis. The works covered here by no means exhaustively cover the breadth of the field, but instead were selected to represent significant milestones in oligo chemistry that pertain to my work on phosphorodithioated deoxyribonucleic acid. During my graduate studies I spent a considerable amount of time synthesizing and purifying DNA, in the end producing over 200 oligonucleotide syntheses for use in my projects as well as in collaborations.

### BRIEF HISTORY OF NUCLEIC ACID CHEMISTRY

Research in the field began around 1870 when Friedrich Miescher discovered a novel material in leukocytes while trying to isolate cytoplasm-free nuclei in the lab of Felix Hoppe-Seyler. The material he found was remarkably protease resistant and elemental analysis showed it to be unusually phosphorus-rich and sulfur-poor, which were properties that differentiated it from proteins. Miescher showed that this crude material could also be isolated from the nuclei of other cell types, including salmon sperm which was easily acquired[1], and named this general substance "nuclein"[2]. His findings of a non-protein component were revolutionary and Hoppe-Seyler, who edited the journal in which the manuscript was submitted, required validation of the results before publishing—a period that took two years due to the novelty of this discovery. Albrecht Kossel, another

student of Hoppe-Seyler, continued the work of Miescher and developed a more sophisticated purification method that could further isolate the non-protein components from Miescher's nuclein preperation[3]. Later, the acidic properties of nuclein's non-protein component was described in detail by Richard Altmann[4,5], a student of Miescher who was then at Leipzig University . Interestingly, it was Altmann's published *combinatio nova* "nucleic acid" that was adopted as the wordstock that remains in place today.

During the period of 1883 to 1900, Kossel's work contributed greatly to the understanding of its composition. Using hydrolysis and chemical analysis he determined that nucleic acid consisted of five organic compounds: adenine, cytosine, guanine, thymine, and uracil[1,3,6-13]. Methods for the extraction of each of these nucleobases was improved to the point were each could be isolated individually in a moderately pure form. Thymidine purified from fish sperm was the easiest to obtain with decent yields[1,14,15] and thus became the initial target for early synthetic chemists.

While recovering from a bout of tuberculosis at the turn of the century, a chemist by the name of Pheobus Levene developed a mentorship under Kossel and became interested in the subject of nucleic acids. Known for his work on the structure of sugars, in 1905 Levene accepted a position at the newly formed Rockefeller University and embarked on a prolific career that would see him publish over 700 manuscripts in biochemistry. He discovered ribose[4,16-19], deoxyribose[20-22], and defined a nucleotide subunit as consisting of a purine or pyrimidine nucleobase, a glycosidically bonded pentose sugar, and a phosphate group[14,15,23] which were linked together in that order[14-19,24]. Levene determined that nucleic acid consisted of a string of nucleotides linked by ester bonds at the phosphates forming what was termed the phosphodiester backbone of polynucleotides[20,25]. Although the significance of his advances in nucleic acid structure was great, Levene is also remembered for his chemistry skills and was the pioneer of creating nucleic acid derivatives.

Synthetic work first requires pure nucleobase and nucleotide starting material, because producing these from scratch is impractical due to their complexity. However, purification of useful forms of these was difficult since the

stability of the phosphodiester linkage[23,26] and the instability of the glycosidic linkage contradict each other during harsh chemical workups. Levene and Medigreceanu approached the problem tactfully, finding that instead there were enzymes in organs capable of cleaving nucleic acids to nucleotides, others capable of dephosphorylating nucleotides, and some capable of hydrolyzing the glycosidic bond to liberate nucleobases.

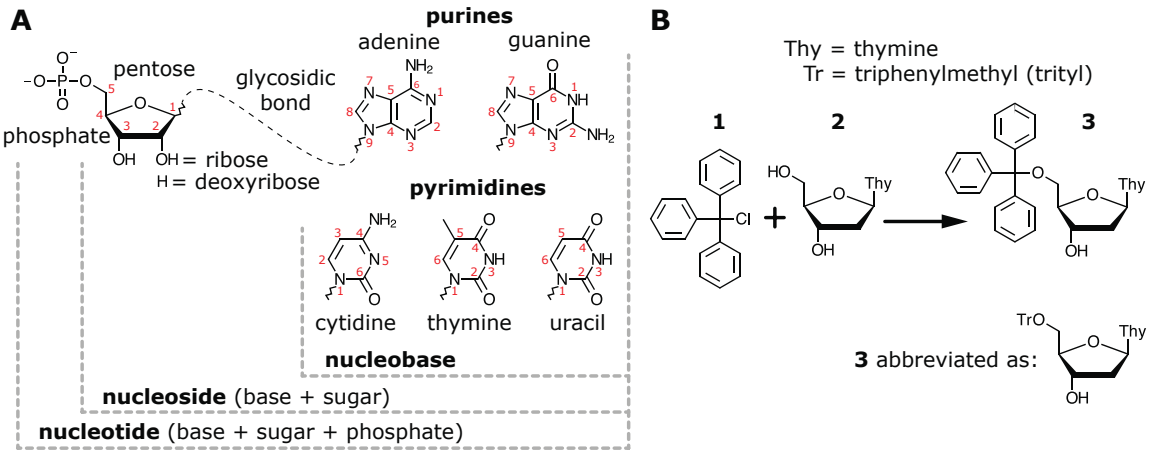



**Fig. 1.1) Levene's Contributions to Nucleic Acid Structure**

**A)** A nucleoside consists of a heterocyclic aromatic purine and pyrimidine nucleobases attached to a ribose sugar through a glycosidic bond. After phosphorylation of the 5'-OH it is termed a nucleotide. **B)** Levene develops the first method for protecting the 5'-OH by attaching a trityl moiety.

They developed an *in-vivo* animal model for the process: a dog with a gastric fistula of the stomach and a Thiry-Vella fistula of the intestinal tract (interestingly, Levene and the great psychologist Ivan Pavlov were friends). Levene's method was quite archaic; it entailed injecting 50 grams of nucleic acid into a live animal through a hole in its stomach, extracting the enzymatically processed material from a hole in its intestines, and purifying its constituents. Thankfully the technique progressed so that gastrointestinal juice could simply be harvested and then combined with nucleic acids *in vitro*, minimizing the effect on the animal[14,15,19,24,27]. Nevertheless, Levene succeeded in producing monophosphoric esters of nucleosides as well as native nucleosides that would provide the fundamental basis for early work in the chemical synthesis of oligonucleotides. The true significance of the nucleic acid derivatives he would chemically synthesize, such as 5'-O-monotrityl-thymidine[25,28],

would not be realized until decades after his passing when chemists would employ its unique 5' protecting properties in directed chemical synthesis of oligonucleotides.

## H-PHOSPHONATE/PHOSPHOTRIESTER CHEMISTRY

The birth of the chemical synthesis of oligonucleotides is largely attributed to the lab of Sir Alexander Todd who in the 1950s became the first to couple two nucleosides with a phosphodiester bond[26,29]. This feat was possible because they had improved on Levene's nucleobase production method in a way that yielded sufficient gram-scale amounts for solution-based chemical modification[27]. Instead of using crude gastrointestinal juice from dogs, Todd utilized the cattle ranching industry and purified high amounts of phosphatase from the intestinal mucosa of calves[28,30,31]. This was used to enzymatically hydrolyze large quantities of commercially available herring-sperm deoxyribonucleic acid, followed by purification via anion-exchange chromatography[26,29] into fractions containing individual nucleobases. Similar to Levene's previous work, thymidine was more easily processed and was obtained in higher yields[27,32]; therefore it became the first nucleobase to be developed for synthetic methodologies.

The chemistry developed by Todd to link nucleosides utilized acetyl and benzyl protecting groups on the 5' and 3' ends to block their reactive hydroxyl moieties, which focused reaction mechanisms on the internucleotide phosphate. It is important to note the use of multiple protecting groups, removable by different treatments strategies, because it enabled a selective and directional step-wise chemical synthetic approach for the coupling of a dinucleotide[30,31,33]. The hallmark of Todd's coupling approach was the phosphochloridate phosphate reaction center, produced by oxidizing 3'-benzyl-H-phosphonate with N-chlorosuccinimide. In this, the hydrogen of the phosphonate is replaced by a chlorine atom, which is a superior leaving group. In Todd's condensation an incoming 3'-protected monomer was added and it's 5'-OH would condense with the reactive 3'-chloridate to form a phosphotriester-linked dinucleotide. The protecting benzyl and acetyl groups on the phosphate 5'/3' ends were then removed and the final phosphodiester-linked

dinucleotide was purified by crystallization. Collectively, this chemistry is referred to as Michelson and Todd's H-phosphonate/phosphotriester coupling and was used to produce the world's first synthetically produced dinucleotide[26,34].

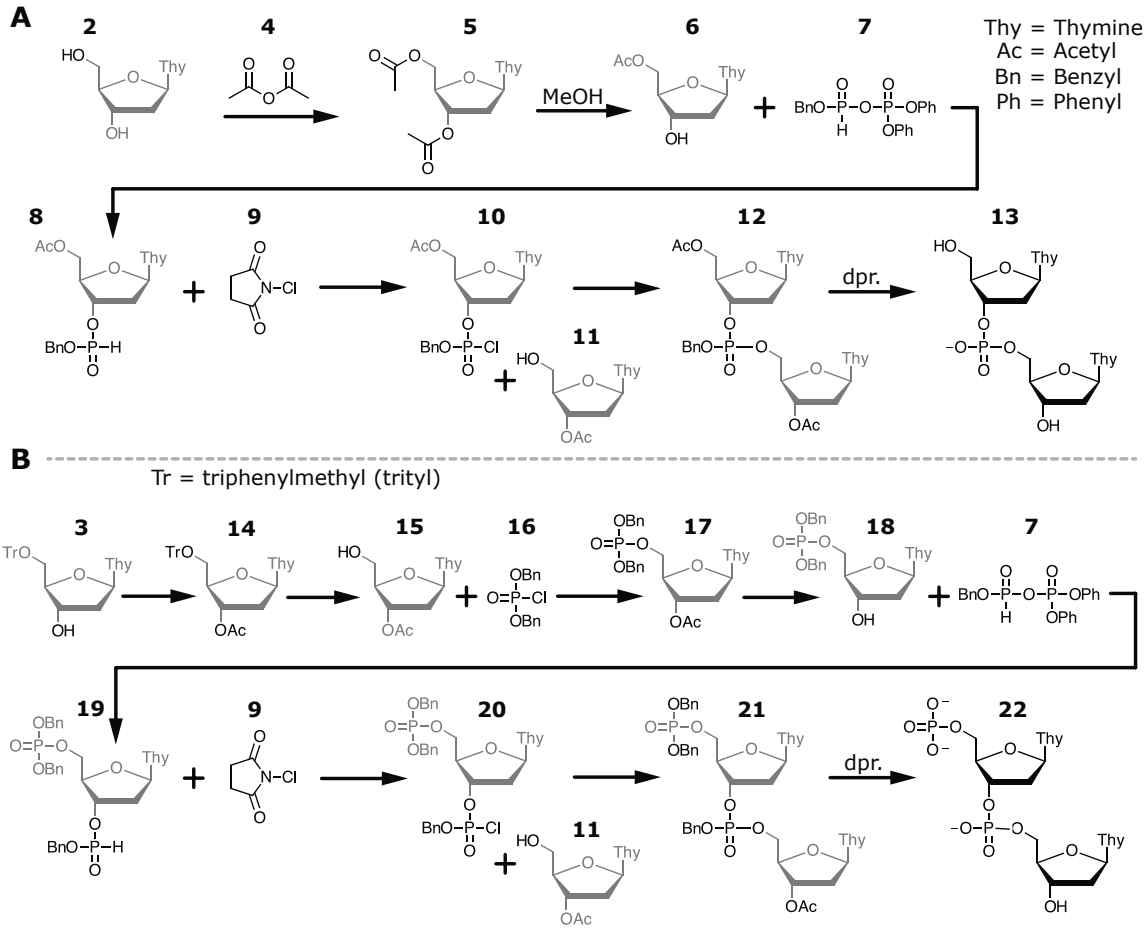



**Fig. 1.2) Michelson & Todd's H-Phosphonate/Phosphotriester Chemistry**

A dithymidine is produced in the first reported coupling of two nucleosides. **A)** Preparation of a dinucleoside phosphate: 5'-protected thymidine nucleoside is first converted to an *H-phosphonate* **8** and then into a *chloridate* **10**. Highly reactive but unstable, the chloridate couples with the 5'-OH of a second nucleoside to form a phosphotriester **12**, which is deprotected to the final product. **B)** Preparation of a true dinucleotide: a true dinucleotide has a 5'-phosphate, which is inserted through phosphorylation at the 5'-OH with **16**. The rest of the pathway is similar to **A**.

Although the dithymidinyl molecule was the first to be synthesized, Todd's group also developed similar chloridate derivatives of deoxycytidine[32,35], and the purines deoxyadenosine and deoxyguanosine[33,36] that could utilize this coupling protocol. Although groundbreaking, the significant drawback to the H-phosphonate method was that the phosphorochloridate was susceptible to hydrolysis and air oxidation and was therefore extremely unstable. In practice, to overcome this limitation one had to immediately advance to the coupling reaction without

purification of an intermediate in order to achieve reasonable yields of dinucleotide coupling. Therefore, it was not possible to scale this method to produce molecules longer than two bases, and a new type of chemistry was needed.

## PHOSPHODIESTER CHEMISTRY

The next major development in the field was the phosphodiester method of oligonucleotide synthesis developed in the lab of Gobind Khorana during the 1950s and 60s. His group is credited for great advancements in the field of oligonucleotide synthesis after the introduction of important concepts that enabled the production of molecules longer than just a few bases. The first such idea was the use of a monoesteric 5'-phosphorylated nucleotide that would be *activated* with a reagent immediately prior to coupling. Using such reagents gave much greater control over the condensation reaction and eliminated the requirement for unstable intermediates, such as Todd's chloridate. When developing their activator reagents the first compound they found success with was *p*-toluenesulfonyl chloride (TsCl)[25,34].

Soon they found dicyclohexylcarbodiimide (DCC) to be superior[35,37] and later several alternatives suitable for specific applications[36,38], but DCC would go on to become the most widely utilized activator for decades. As a result of activation-dependent coupling, the concept of producing stable monomers with long-term storage was advanced. This monomer-based approach is something that all subsequent chemistries would aim to accomplish. Khorana is also credited for advancing the practice of utilizing various groups in an "on/off" protection scheme to direct condensation to specific hydroxyls on the ribose. For instance, the activated 5'-phosphoryl nucleotide was 3'-*O*-acetyl protected so it can't self-couple whereas the incoming nucleoside was 5'-protected but had a free 3'-OH for condensation. After each addition the protection/deprotection cycle was repeated. During this period the most common 5'-protecting group was Levene's trityl moiety[25,39], which worked well for the early dithymidine molecules but was not appropriate for mixed polynucleotides. Trityl's removal involves a long acid treatment, and because the glycosidic bond of purines is more sensitive than

pyrimidines, they are incompatible with such exposure. Guanine is particularly susceptible to acid cleavage and therefore, to prevent depurination of these nucleobases, 5'-*O*-acetyl was used instead.

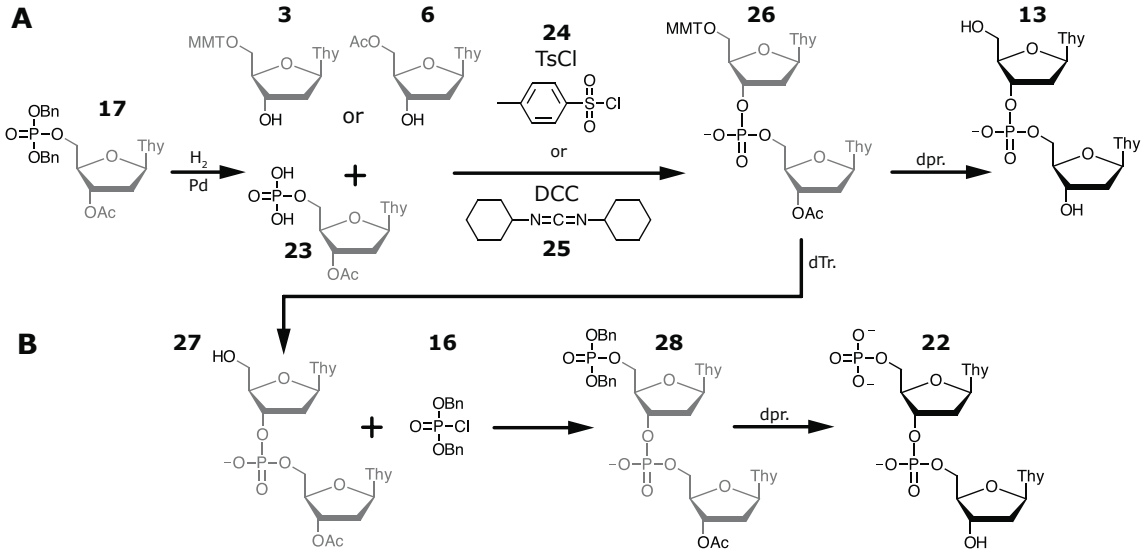



**Fig. 1.3) Khorana's Phosphodiester Chemistry**

First use of an *activator* to control condensation. **A)** Preparation of a dinucleoside-phosphate: a 5'-phosphitylated nucleoside **17** is hydrogenated to its active hydroxyl form **23**. This is mixed with a 5'-protected nucleoside and activated with either TsCl or DCC the phosphodiester bond is formed. B) Preparation of a true dinucleotide: Instead of deprotecting 26, the 5'-trityl is removed and phosphorylated with **16** as with Michelson & Todd's method.

However, trityl has one property that makes it a very special leaving group and worth improving rather than abandoning: it is one of the rare molecules with an electronic arrangement that prefers its carbocation form once cleaved, as opposed to immediately acting as a strong electrophile and causing unwanted side reactions. Recognizing this value, several groups worked on developing trityl derivatives that could be removed from DNA with gentler methods. It was known at the time that methoxy substitutions aided dissociation of hexaarylethanes to its carbocation[37,40] so this strategy was adapted to trityl (triarylmethyl) groups. Khorana utilized 5'-monomethoxyltrityl (MMT) protected nucleosides[29,38] and Hogenkamp and Oikawa[39,41,42] utilized 5'-dimethoxyltrityl (DMT) protected nucleosides to create hetero-base polynucleotides. DMT, which is removed by acid in just a few minutes rather than an hour, has become the trityl form most often used for 5'-OH protection in modern synthetic methods.

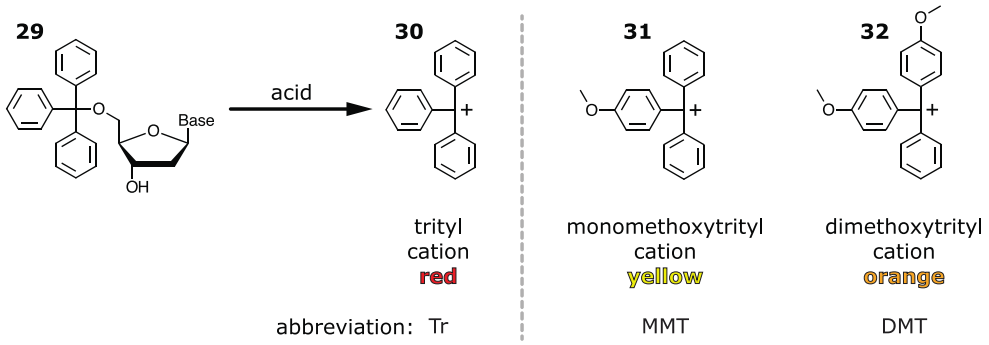



**Fig. 1.4) Development of Trityl Derivatives for 5'-OH Protection**

The 5'-OH was first protected with a trityl moiety **29**, which requires strong acid treatment to remove and is thus incompatible with the glycosidic bond of purines. Methoxy derivatives quicken the cleavage rate and reduce reaction times from hours to minutes. In the carbocation form each derivative has a unique visible color in organic solvent. Trityl **30** is orange, MMT **31** is yellow, and DMT **32** is orange. The released color then acts as a reaction progress monitor.

The term phosphodiester chemistry[40,43,44] is used to describe Khorana's method because the coupling phosphorus had two ester linkages to the 3'- and 5'-end of the nucleosides. However, the other two oxygens on the phosphorus were left unprotected and would unfortunately act as condensation sites for subsequently activated nucleobases. This resulted in oligos that were branched at the phosphorus, which had to be removed by purification, resulting in greatly reduced synthetic yield and increased processing efforts. Compared with Todd's protected (benzyl) phosphotriester approach, Khorana's phosphodiester method was seemingly a step backwards. However, the benefits of avoiding the tricky chloridate and general improvements in purifying branched nucleic acids from product by DEAE-cellulose anion exchange chromatography[29,45] led to the acceptance of the phosphodiester method.

In the 1960's, elucidation of the genetic code was of enormous biological interest, and Khorona adapted his deoxyribonucleic acid phosphodiester synthetic protocol to work with ribonucleic acids after developing suitable protection strategies[41,42,46]. He synthesized repeating di-, tri-, and tetraribonucleotides[43,44,47] and enzymatically converted them to longer polynucleotides for cell-free peptide expression[45,48-60] to support the global codon defining effort. Soon his lab turned their efforts in 1965 to the monumental task of synthesizing a complete DNA gene, choosing as a target the principle yeast alanine transfer RNA which was the first

gene sequence to be determined[46,61]. This was a monumental task because of the branching problem that plagued the method, which greatly limited the length for synthetic oligos. To compile the entire 77-base pair duplex, 15 strands of 5-16 bases were conjoined using T4 DNA ligase[47,62]. It took over five years and numerous graduate students before the group published the complete assembled gene[48-60,63-65]. When the importance of the genetic code became known, so did the realization that we needed improved chemistries capable of producing longer strands with less effort.

## MODIFIED PHOSPHOTRIESTER CHEMISTRY AND THE FIRST SOLID SUPPORTS

A peptide chemist named Robert Letsinger entered the oligonucleotide field in the 1960's and provided three major advances to the field. His first contribution was reverting back to the phosphotriester approach and adding an ester-linked protecting group, which prevented the branching that plagued Khorana's phosphodiester approach. He made a pre-protected phosphorylating group, which would couple to a 3'-OH with the help of a wide variety of activating reagents: phosphorus oxychloride, *p*-nitrophenyl phosphoro-dichloridate[61,66-69], *O*-cyanoethyl phosphoro-dichloridate, *p*-nitrophenyl phosphate plus dicyclohexylcarbodiimide, and *p*-cyanoethyl phosphate plus dicyclohexylcarbodiimide[62,70]. For the synthesis of oligonucleotides the p-cyanoethyl phosphate/DCC combination proved more practical as p-cyanoethyl phosphate was produced in high yields and could later be cleaved easily using ammonium hydroxide[63-65,71,72]. Other groups also developed similar protected phosphotriester approaches[63,66-69,73], which further improved coupling efficiency to levels in excess of 90% per step. During this period mesityl sulfonyl chloride and mesityl sulfonyl nitrotriazole became popular activators and general protecting groups for various hydroxyls were advanced[70,74,75], but chemists' preferences were often dynamic throughout 1960s and 1970s since groups were continuously optimizing reagents[64,71,72]. Given the efficiency of the modified phosphotriester approach, this chemistry was a great improvement over Khorana's phosphodiester method because eliminating branching enabled the synthesis of much longer polynucleotides and reduced the dependency on ligation.

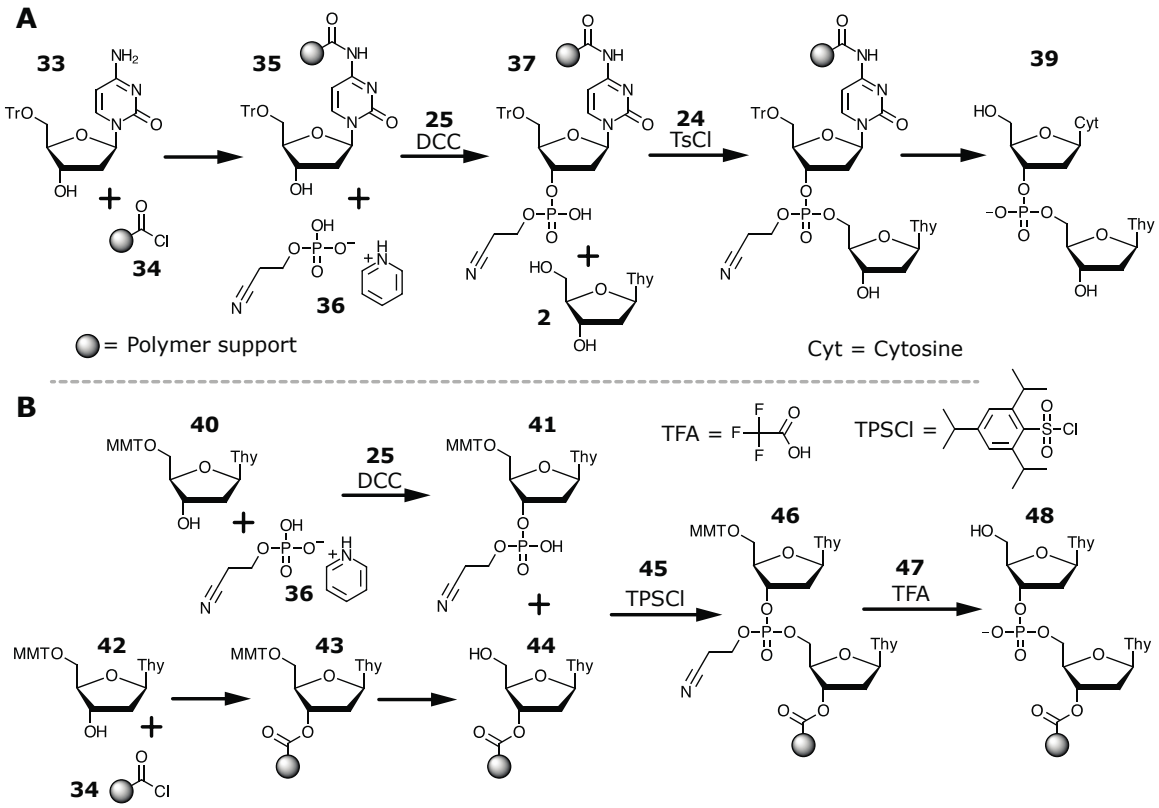



**Fig. 1.5) Letsinger's Modified Phosphotriester Synthesis on Solid Supports**

Letsinger adds a protecting group on the phosphorylating reagent, eliminating branching during coupling. The use of activating reagents to control condensation is continued, with new reagents being proposed. **A)** First solid-support synthesis of oligonucleotide is achieved by attaching a "popcorn styrene" polymer support to the amine on cytosine. **B)** The optimal attachment point for the support is determined to be the 3'-OH—in modern strategies the same design is utilized.

The second major contribution Letsinger made to the field was the development of synthesis on a solid-phase support. The value of this concept is that it would later enable machine-automated synthesis of oligonucleotides and eliminate the hands of a chemist, bringing synthetic oligo production to the masses. In Letsinger's initial report, 2'-deoxycytidine (dC) was attached through the amine at position 4 of the cytidine base[63,71,73,76,77] to an acid chloride-modified "popcorn" support made from the copolymer of styrene (88%), *p*-vinylbenzoic acid (12%), and *p*-divinylbenzene (0.2%)[74,75,78]. This amine linkage to the base was cleaved with ammonium hydroxide following synthesis, simultaneously with the cyanoethyl phosphate protecting group. Over the next few years Letsinger continued to explore solid phase synthesis and determined that the best approach was to attach the support to the 3' hydroxyl[64,79], as is done today.

## PHOSPHITE-TRIESTER CHEMISTRY (SWITCH FROM P(V) TO P(III) OXIDATION STATE)

The third and greatest advance Letsinger made in the field was the introduction of phosphite-triester chemistry in 1975, utilizing a phosphitylated nucleoside with a P(III) oxidation state instead of a P(V) state[71,76,77,80,81] . This made the phosphorus extremely reactive during coupling, which can be explained by visualizing a $S_N2$ bimolecular substitution at the phosphorus where the nucleophilic 5'-OH of the incoming nucleoside displaces the chlorine leaving group. The formation of the trigonal bipyramidal transition state involves the atoms bonded to the phosphorous center moving through the planar equatorial position. In a P(III) center two esters and a pair of electrons enter this plane, which has a much lower energetic barrier than two esters and the double-bonded oxygen making the transition in a P(V) center. As a result, condensation kinetics improved by an order of magnitude from several hours to mere minutes. After condensation, the linkage is oxidized back to the P(V) state using aqueous iodine to form the natural phosphodiester bond.

Although this new P(III) chemistry revolutionized the field of oligo synthesis seemingly overnight, its current implementation had its drawbacks. Most notable was the instability of the phosphomonochloridite[78,80,82], which like Todd's chloridate intermediate was prone to hydrolysis and air oxidation, and therefore had no shelf life. Although this required chemists to prepare the chloridite immediately prior to use, Letsinger's P(III) phosphite chemistry was simple enough to be reproduced successfully in many labs and its adaptability to solid phase methodology led to the industry's first automated DNA synthesizers being introduced in the late 1970s, which helped initiate the biotechnology revolution.

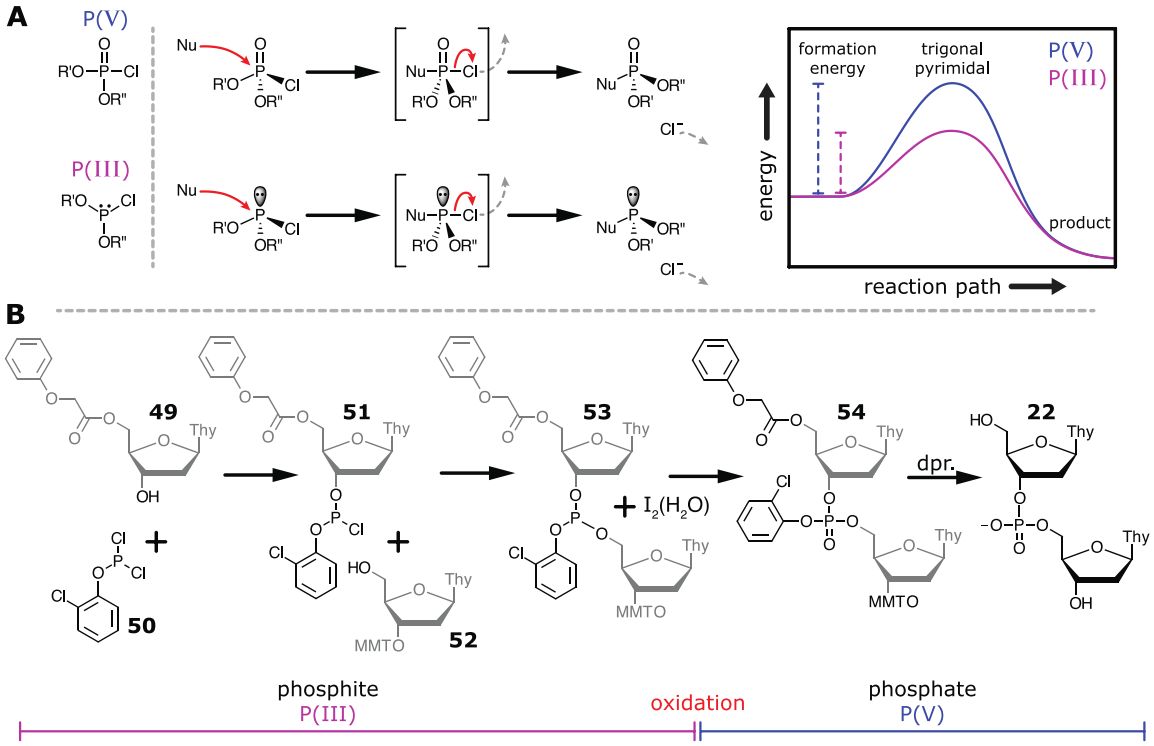



**Fig. 1.6) Phosphite Chemistry Using the P(III) Oxidation State**

Phosphite chemistry is advanced in 1975 by Letsinger and revolutionizes oligo synthesis with its rapid coupling kinetics. **A)** The kinetics are explained with an $S_N2$ reaction mechanism. P(III) has a lone electron pair instead a double-bonded oxygen, the electron pair has a lower energetic barrier to enter the equatorial planar position in the trigonal bypyrimidal transition state. B) Phosphite synthetic approach. A nucleoside is phosphitylated to a chloridite **51**, which reacts with a 5'-OH of second nucleoside **52**. The internucleotide P(III) phosphite is oxidized to the P(V) phosphate using iodine and water.

## PHOSPHORAMIDITES, P(III) AND CONTROLLED-PORE GLASS SOLID SUPPORTS

A solution to the phosphite instability issue was developed later in the lab of Marvin Caruthers, a former graduate student of Letsinger, who replaced the chlorine leaving group with an amine moiety. They initially saw a moderate improvement in stability using a *N,N*-dimethylamino group[79,80], but found that it still required rigorous anhydrous conditions, an inert $N_2$ gas atmosphere, and often yields varied and suffered from impurities due to hydrolyzed product. However, this initial result was encouraging and Caruthers and others investigated additional alkylamino substitutions, finding that both the N-morpholino[38,63,64,80,81] and diisopropylamino groups[80,82,83] were superior groups. Diisopropylamino became the moiety of choice because its precursor phosphitylating agent, chloro-*N,N*-diisopropylaminomethoxy-

phosphine has a boiling point of 35°C (20°C less than chloro-*N*-morpholino) and therefore could be distilled to complete purity with a simpler protocol[80,84-86]. With this chemistry phosphoramidites were monofunctional, stable to hydrolysis and air oxidation under normal laboratory conditions, and were now capable of being stored as dry powders for long periods.

With the high coupling efficiency of the phosphoramidite method it became apparent that more work was needed in the area of solid supports. After the first synthesis using a polystyrene[38,63,64,82,87], the field saw few other options emerge for almost 20 years[83,88]. During this period limitations of the polystyrene support were obvious: they had slow diffusion properties, they exhibited excessive swelling, and there was often irreversible absorption of organic reagents onto the polymer. Soon, two different supports were put forward that avoided these issues. First, Caruthers utilized the silica supports developed in the emerging high-performance liquid chromatography field, finding them to exhibit great performance in oligo synthesis[84-86,88]. The second support solution put forward was a subtle but major advancement, utilizing glass beads[62,67,70,82,87], which are amorphous solids of borosilicate and sodium oxide. The glass is made porous by leaching out the borates, creating a dispersed distribution of pores with uniform sizes. These voids were large enough that reagents could diffuse freely and made possible higher flow rates for machine-automation. The oligos were attached to the controlled-pore glass (CPG) at the 3'-end using a long-chain alkyl amino moiety that was cleaved in under an hour with ammonium hydroxide. When CPGs were combined with the efficient phosphoramidite method chemists were able to achieve great lengths in polynucleotide synthesis, producing a 43-nucleotide-[78,88,89] and 77-nucleotide-long[88,90] oligo, the longest chains ever produced at the time without using ligation.

By the mid-1980s oligonucleotide synthesis using phosphoramidites was effectively optimized to produce a coupling efficiency >98%, so the only major work remaining involved matching protecting group chemistries for their concerted removal after synthesis. The nucleic acid bases adenine, cytosine, and guanine each bear exocyclic amino groups that produce side products in the coupling reaction if

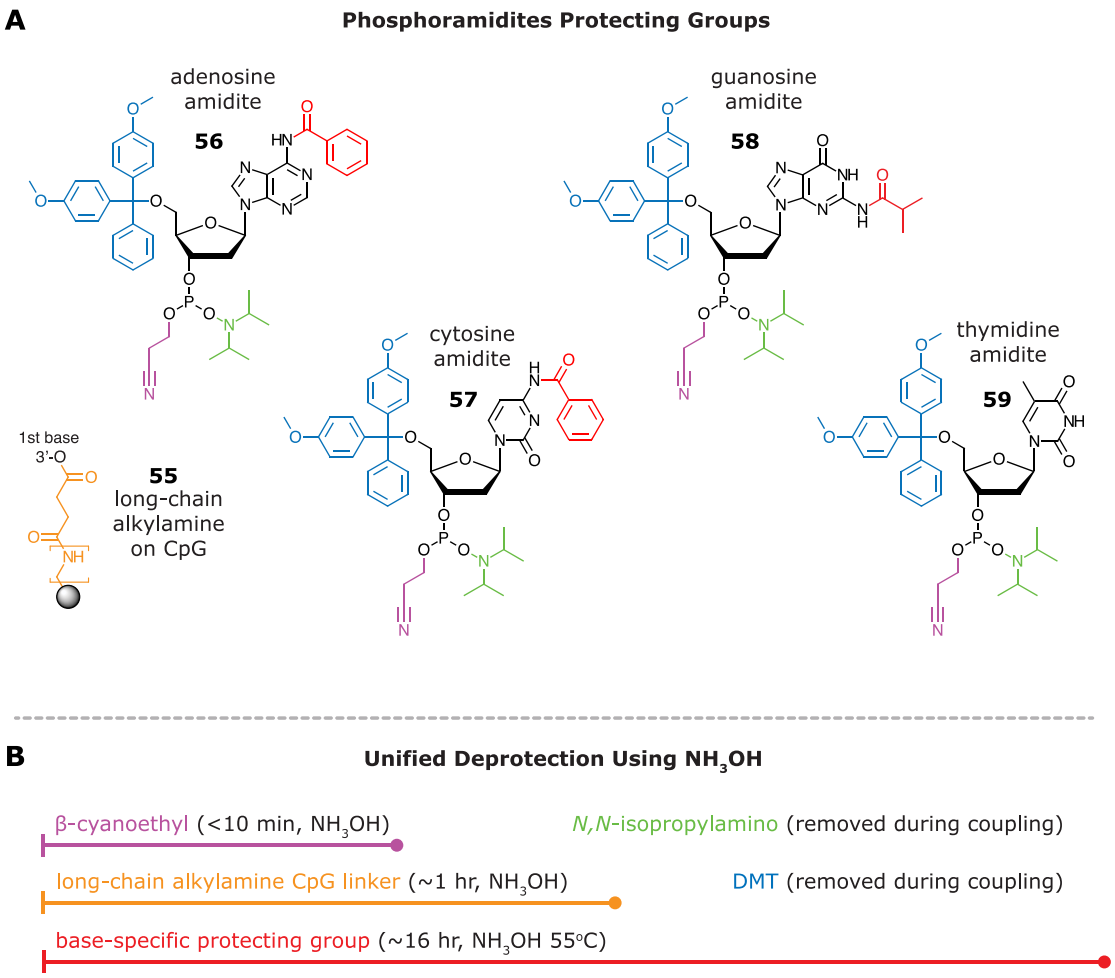



**Fig. 1.7) Phosphoramidites and their Protection Strategy**

The modern phosphoramidite is based off of the P(III) phosphite strategy. They are stable in an inert atmosphere, can be prepared in large quantities, shipped globally and stored as dry solids for many months prior to use. Only upon protonation do nucleoside phosphoramidites become reactive. **A)** Phosphoramidites for all nucleobases and their protecting groups are shown. DMT (blue), β-cyanoethyl (magenta), isopropylamino (green), base specific protecting groups (red), and the LCAA linker to the controlled-pore glass (CPG) support (orange). **B)** Removal of all protecting groups occurs with a single reagent, ammonium hydroxide, $NH_3OH$.

unprotected (e.g. Letsinger's first solid-support synthesis attached to this amine group). In contrast, thymidine has no exocyclic amine groups (another reason why the first dinucleotides were dithymidinyl structures). For A and C nucleobases the most common protecting strategy for their amine moieties is the benzoyl group, and for G an isobutyrl group is now used. All of these can be removed using aqueous ammonium hydroxide at elevated temperature (55°C) for 16 hours. To match deprotection strategies for the phosphorus and base groups there was a reversion back to the β-cyanoethyl protecting group previously used for phosphates[62,67,70,91] to

now protect the P(III) in modern phosphor-amidites[78,89,92]. Prior to cyanoethyl, groups were used that required an additional step for removal, such as Caruthers' methyl groups which were cleaved with the foul-smelling thiophenoxide[90,93]. Instead, using β-cyanoethyl groups simplifies processing because they are removed simultaneously during the ammonium hydroxide solid-support cleavage step. At this point the modern phosphoramidite had been assembled, consisting of a phosphite center and three types of protecting groups that are all removed in ammonium hydroxide. They became very popular and supplanted all other chemistries for the synthetic production of oligonucleotides. More recently, modified oligonucleotides have become popular and their sensitive chemistries have necessitated the development of even gentler deprotection strategies that avoid heating in ammonia hydroxide. Thus, new protecting groups have been developed for specific applications, such as: N,N-dimethylformadine[91,94], N-isobutyryl[92,95], methoxyacetyl, phenoxyacetyl[93,96-100], 4-isopropylphenoxy-acetyl[94,101-104], and other acetyl derivatives[95,105-108].

## 1.2) The Modern Phosphoramidite Synthetic Cycle

### INTRODUCTION

The optimization of the phosphoramidite method has made the production of synthetic oligos a relatively straightforward step-wise addition of monomers and reagents over a solid-support onto which the growing polynucleotide is attached. Synthesis proceeds in the 3'→5' direction, with one phosphite-containing nucleotide added at a time in a cycle that is repeated until completion. To efficiently manage the costly reagents and easily maintain an inert environment, a semi-autonomous DNA synthesizing machine programmed by an operator controls the volume flow of reagents. A diagram of the full synthetic cycle is shown next, and a detailed descriptions of each step's mechanism follows.

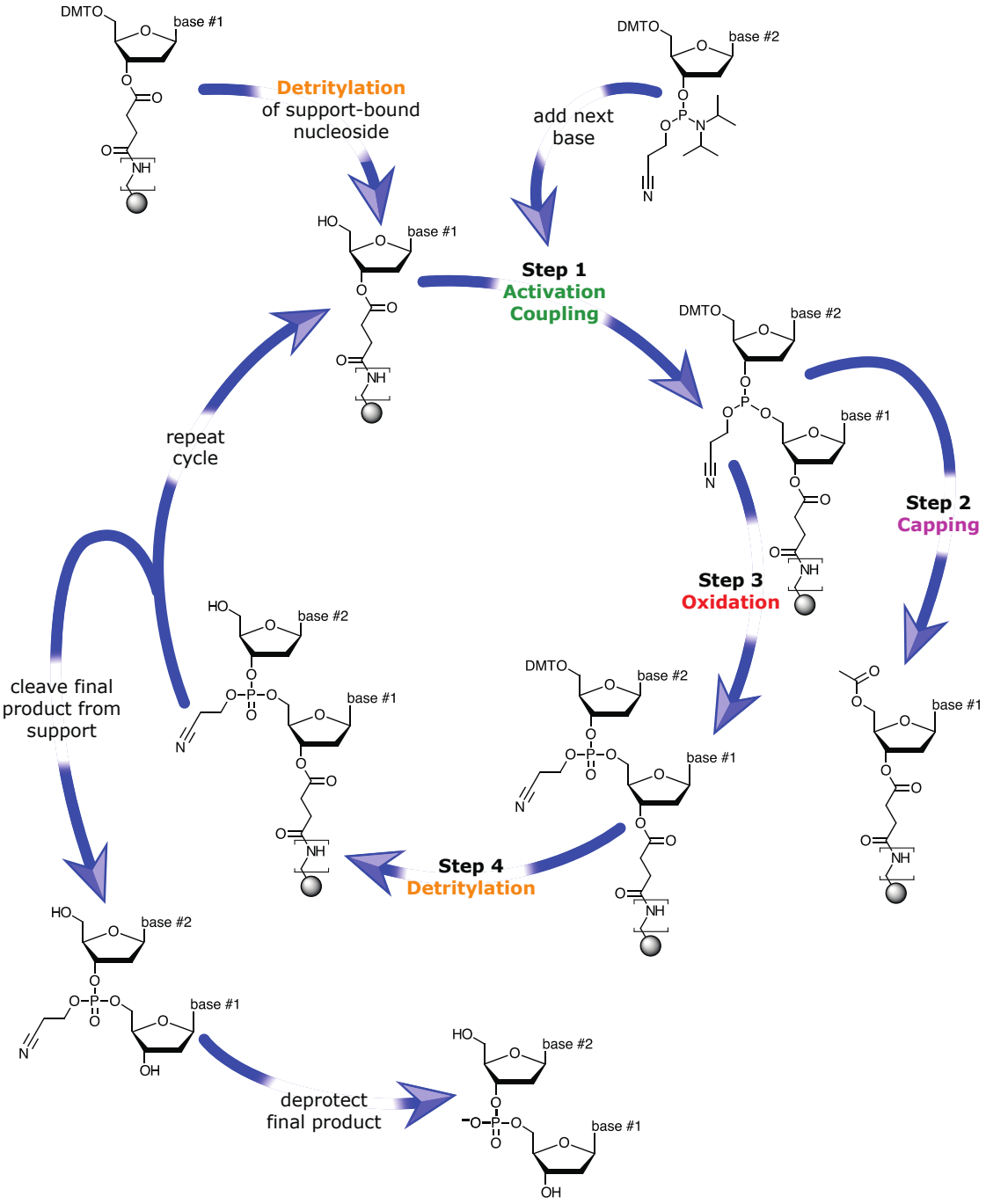



**Fig. 1.8) Synthetic Cycle Using Phosphoramidites**

## DEBLOCKING (DETRITYLATION)

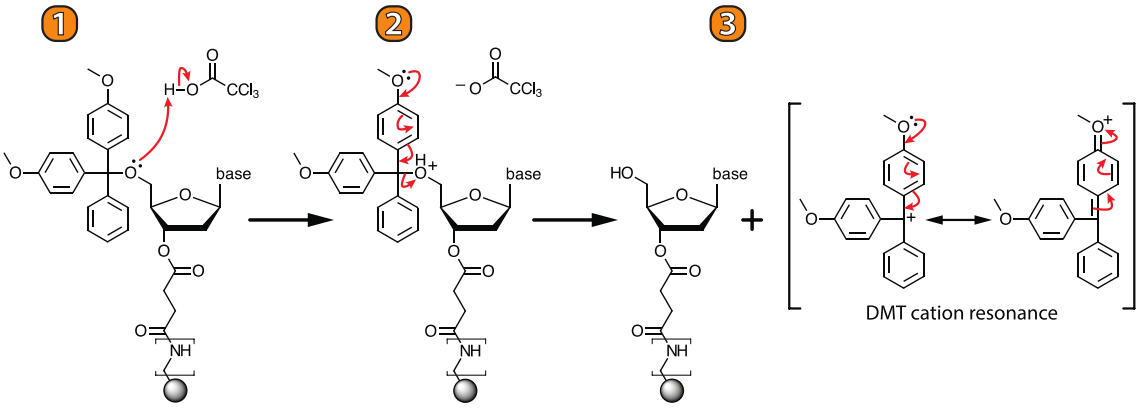



**Fig. 1.9) Acid-Catalyzed Detritylation**

Support columns are pre-packed with CPG beads having nucleosides pre-attached at their 3'-end to the beads through a long-chain alkylamine linker. Since synthesis proceeds in the 3'→5' direction, the operator selects an appropriate column (A, C, G, or T) depending on the terminal 3' nucleotide in the sequence. All reagents and monomers flow through this column and the oligonucleotide chain is assembled step by step. The first base attached to CPG contains a 5'-DMT protecting group which must be removed to reveal a free 5-OH for subsequent coupling. The mechanism of acid-catalyzed DMT cleavage is shown above. (1) In dichloromethane solvent, trichloroacetic acid first protonates the ether between the DMT and ribose ring, increasing its formal charge. (2) An electron rearrangement within the DMT follows, culminating with the transfer of an electron pair to the oxygen, which breaks the ether covalent bond. (3) The DMT protecting group is liberated to its carbocation, revealing a 5'-OH on the CPG attached base that will be the only reactive nucleophile capable of coupling in the next step. DMT makes an excellent leaving group since there is strong resonance in its carbocation form. In dichloromethane DMT has a vibrant orange color that is integrated by an optical detector as the DMT is flushed out of the reaction column to give an estimate for the efficiency of the deblocking. In practice, since deblocking is nearly ideally efficient under these conditions, the trityl signal actually reveals indirectly the success of the previous coupling step. If the synthesis is not performing well, the trityl signal provides an indication so that consumption of valuable reagents can be halted. The

deblocking is kept as short as possible, since nucleobases are susceptible to acid. After the trityl signal is integrated the reagents are flushed from the column using an acetonitrile wash, which also helps prevent premature detritylation of the incoming phosphoramidite.

### ACTIVATION AND COUPLING

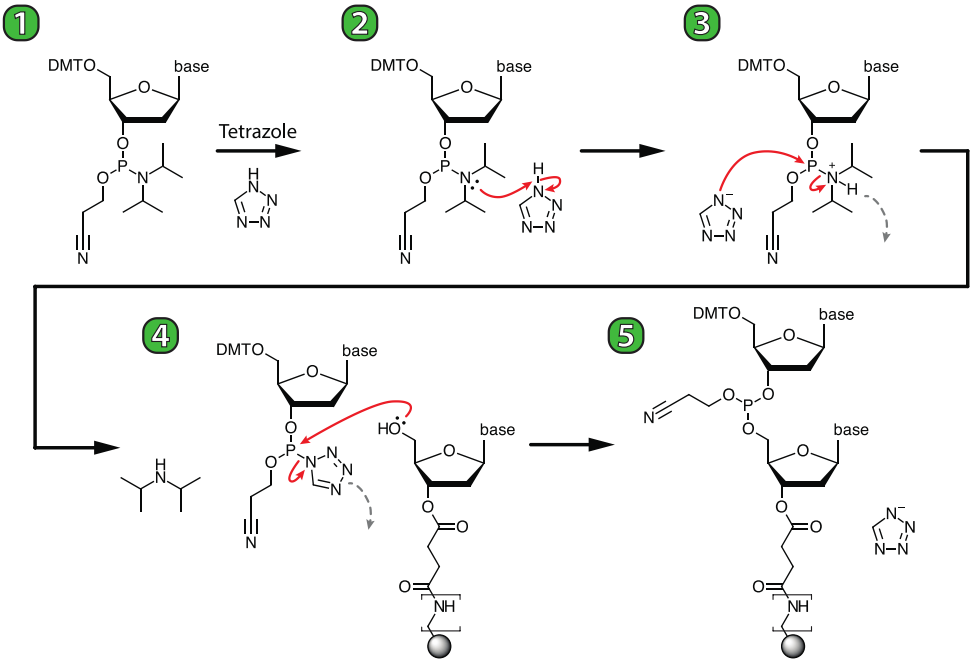



**Fig. 1.10) Phosphoramidite Tetrazole Activation and Coupling**

(1) After deblocking, a large excess of the next protected phosphoramidite is delivered to the column together with the weakly acidic tetrazole activator. They are both dissolved in acetonitrile which is an excellent polar aprotic solvent for nucleophilic displacement reactions. (2) The tetrazole activator protonates the nitrogen of the *N,N*-diisopropylamine, increasing its formal charge. (3) The *N,N*-diisopropylamine leaves the phosphite after nucleophilic displacement by the tetrazole anion produced after being deprotonated in step 2. (4) Coupling occurs when there is then a second nucleophilic displacement at the phosphite of the 5'-OH of the previously deblocked nucleoside, which displaces the tetrazole group. The true mechanism of the reaction has never been fully elucidated and there is subtle debate[96-100,109] as to whether nucleophilic action of tetrazole as I show here actually

exists. Another proposed mechanism is that only one nucleophilic displacement occurs and step 4 is skipped. Tetrazole would act solely to protonate *N,N*-diisopropylamine, which is instead displaced directly by the 5-OH of the nucleoside. (5) Regardless of the true nature of the reaction, the result of activation and coupling step is an internucleotide linkage (3'→5') of two nucleobases composed of a β-cyanoethyl-protected P(III) phosphite triester. Because the protonated phosphoramidite is so reactive the coupling process is complete within 30 seconds. To ensure coupling efficiency a molar excess of tetrazole is used over the incoming nucleobase, which in itself is kept in excess of the support-attached oligo.

## CAPPING

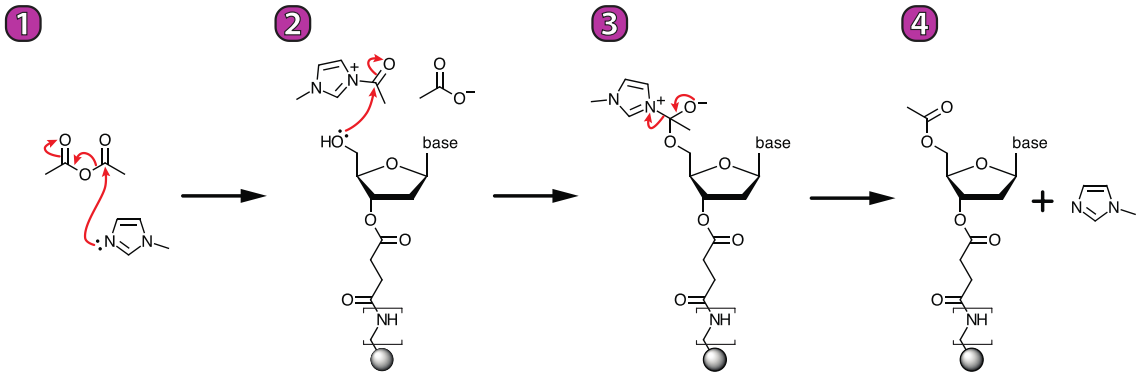



**Fig. 1.11) Capping the Unreacted Hydroxyls**

It is reasonable to expect a coupling success rate that exceeds 99%. However, in spite of this efficiency there will always be a small percentage of 5'-OH groups that do not condense with the incoming activated phosphoramidite. These hydroxyls must be rendered inactive so as to stop the propagation of deletion mutants if they were to enter the next coupling cycle. (1) Two different capping solutions are used on the synthesizer, acetic anhydride in tetrahydrofuran, and *N*-methylimidazole in tetrahydrofuran/pyridine. These two reagents are mixed immediately prior to delivery to the column, where there is a nucleophilic attack from *N*-methylimidazole to the carbonyl of the acetic anhydride, releasing an acetate anion and forming an imidazole intermediate. (2) Capping occurs with a second nucleophilic attack of the unreacted 5'-OH at the carbonyl of the methylimidazoleacetate intermediate. (3) There is a rearrangement of electrons

that allows imidazole to leave the 5'-intermediate. (4) Imidazole fully dissociates, resulting in an acetyl protecting group at the 5'-end of the nucleoside which renders it inert to subsequent reactions. The pyridine in the capping mixture helps buffer the pH during this process to maintain basicity, which prevents premature detritylation of the newly coupled nucleobase.

## OXIDATION

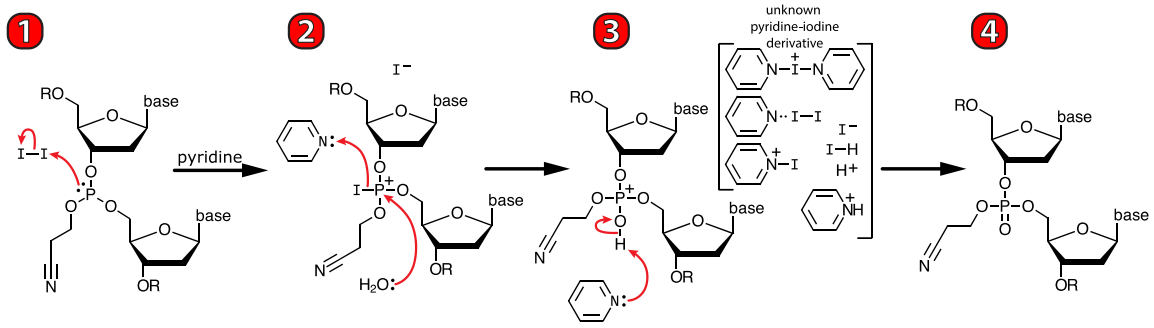



**Fig. 1.12) Oxidation of P(III) Phosphite to P(V) Phosphate**

The P(III) phosphite triester linkage must be oxidized to the P(V) state. If it isn't, it would be converted into an H-phosphonate during the next acid-catalyzed detritylation. The oxidation reagent is 0.2 M iodine in a mixture of tetrahydrofuran, water, and pyridine. (1) Iodine acts as a polarizable electrophile for the lone electron pair of the P(III) phosphite. (2) A pyridine/iodine adduct forms on the new phosphate triester, which is quickly displaced by water. (3) A second pyridine deprotonates the hydroxyl of the phosphate, reducing its formal charge. As for the remaining iodine, there are many possible pyridine-iodine-hydrogen interactions and it is unknown which of them shown here is the dominant form[101-104,110-113]. (4) After oxidation, the P(V) phosphodiester bond is formed. We have now coupled two separate nucleosides with their natural linkage, albeit one with an easily removable β-cyanoethyl protecting group.

### TERMINATION OR REPEATING THE CYCLE

The synthetic cycle is now repeated for each additional nucleotide added onto the growing oligo chain. Detritylation will remove the 5'-DMT from the 2nd coupled nucleoside and the processes described above are repeated for each new base. When finished, the oligo can be left with the final 5'-DMT remaining on or off, based on the desired purification technique. After finishing, the oligo is cleaved from the CPG using aqueous ammonium hydroxide at room temperature, which performs an ester hydrolysis of the long-chain alkylamine linker. The β-cyanoethyl groups are simultaneously cleaved during this process and form acrylonitrile byproducts. The oligo then proceeds to a longer deprotection of the heterocyclic bases involving overnight incubation at elevated temperature in the ammonium hydroxide. With the phosphoramidite strategy the incoming monomer brings with it the necessary phosphitylating group, so the synthesized oligo will have a 5'-OH devoid of a terminal phosphate. This differs from enzymatic- or PCR-based approaches to oligonucleotide synthesis, such that when using the phosphoramidite method a 24-base oligonucleotide will have a total of 23 phosphates.

## 1.3) Phosphorothioates

### INTRODUCTION

The two non-bridging phosphoryl oxygens in the phosphodiester bond are ideal locations for heteroatom substitution. Indeed, as soon as phosphite chemistry was introduced researchers quickly saw the opportunity to incorporate unique groups at the phosphorus during the P(III) to P(V) oxidation. By varying the oxidation reagent phosphate analogs such as selenophosphate, imidophosphates, and monothiophosphates were developed[105-108,111,114-117]. Of these, sulfur substitution is especially interesting because of its chemical similarity with oxygen. Being both isopolar and isoelectric, it is reasonable to assume that compared to other heteroatom substitutions, phosphorothioate moieties would have minimal effect on the overall structural integrity of DNA. Since thioated DNA performs in a

similar manner as its unmodified counterpart then thioation can be used to study topics involving backbone phosphates, and conversely with the right experimental design, DNA then becomes a model system to probe chemical differences between sulfur and oxygen.

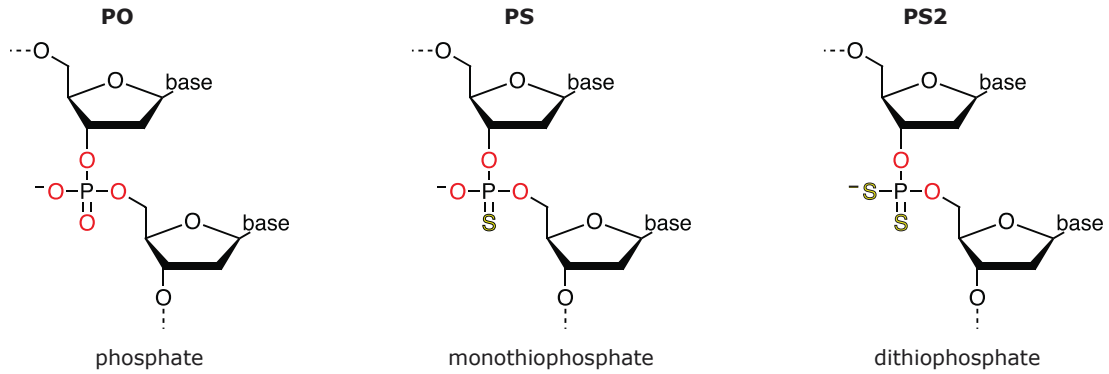



**Fig. 1.13) PO, PS, and PS2 Internucleotide linkages**

The non-bridging phosphoryl oxygens of internucleotide phosphodiester bonds can be substituted with sulfur atoms. PO – normal phosphodiester bond. PS – monothioated phosphate, PS2 – dithioated phosphate.

## EARLY MONOTHIOPHOSPHATE STUDIES

The value of monothioated phosphates (PS) as an experimental tool was recognized very early by Fritz Eckstein, who dedicated his career to probing broad concepts in biochemistry using systems containing phosphoromonothioate groups[109,118]. The first such systems were thiophosphate analogs of nucleoside di-, and tri-phosphates where sulfurs were inserted into the α, β, or γ-positions[109-113]. These were used to study phosphoryl-transfer reactions and ATP catalysis[111,114-117,119-121]. Whether inserted into an oligonucleotide chain[118,122] or as nucleoside triphosphate, the enantiomeric nature (discussed below) of a monothioated phosphorus allowed the investigation of enzyme catalysis mechanisms and stereochemistry[109,123]. Some of the many enzymes studied with monothiophosphates include: bovine ribonuclease A[119-121,124], alkaline phosphatase[122,125], adenylate cyclase[123,126,127], pancreatic ribonuclease[124,128], ribonuclease T1[125,129], DNA-dependent RNA polymerase[126,127,130], cyclic phosphor-diesterase[128,131], snake venom phosphodiesterase[129,132,133], viral reverse-transcriptase[130,134-139], and T4 ligase[118,131,133,140]. As a substrate for various

nucleases it was found that monothioation of internucleotide phosphates imparts nuclease resistance to RNA molecules[108,132,133]. During the molecular biology revolution in the early 1980s, these monothioate cleavage studies were extended to DNA restriction endonucleases, finding that deoxyribonucleic acids with monothioated phosphates also had similar trends of cleavage resistance[134-139,141,142]. With the immense amount of literature reported using thiophosphates, the usefulness of these moieties is limited only by the imagination of the investigator.

## MONOTHIOPHOSPHATE CHIRALITY

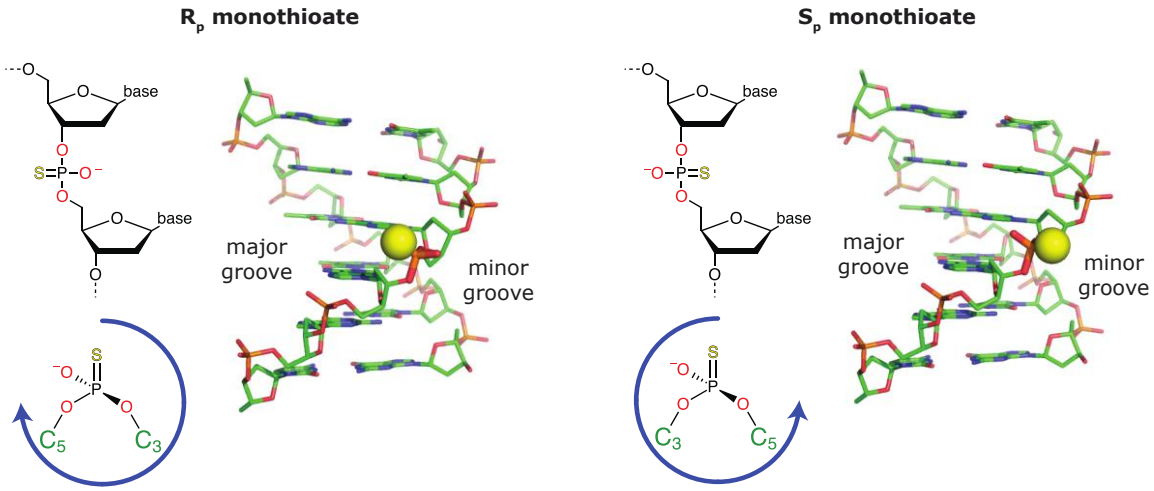



**Fig. 1.14) Monothioation of Phosphate Creates Chiral Center**

A heteroatom-substituted phosphate becomes a chiral center. In a $R_p$ monothioate the sulfur atom extends towards the major groove, whereas in a $S_p$ monothioate the sulfur faces the minor groove.

Although crude methods for introducing sulfur were developed during the phosphodiester chemical synthesis era[118,133,140,143-147], the traditional way to produce monothioated DNA is by altering the oxidation reagent of the phosphite protocol to incorporate a sulfur atom during the P(III) to P(V) step[108,148,149]. This is a relatively straightforward procedure and synthetic yields are equivalent to what is achieved with phosphoramidite synthesis of normal PO oligos. However, this is not a controlled addition, and since the sulfur can join at two possible positions a racemic mixture of $S_p$ and $R_p$ diastereomeric phosphorus centers is produced. The total number of possible enantiomers would be $2^n$, where $n$ is the number of monothios (i.e. a 100% monothioated 21-mer would yield >1,000,000 different

diastereomers). Most biological experiments can tolerate $R_p$ or $S_p$ forms but there are instances where a specific enantiomeric form is required, thus alternative methods utilizing enzymatic synthesis began to emerge. This approach relies on the stereoselectivity of polymerases to produce all-$R_p$-PS oligonucleotides[141,142,150], but yields from such methods are practical only for molecular biology experiments where concentration requirements are minimal. A chemical approach was therefore needed, and a method for stereocontrolled synthesis of PS-oligos was developed in the lab of Wojciech Stec in the 1990s. His novel chemistry incorporated an oxathiaphospholane attached at the 3'-end of the nucleotide, which would couple to the 5'-OH of an incoming nucleoside to form a monothiophosphate linkage[143-147,151].

## SULFURIZING REAGENTS AND MECHANISM

Replacing the standard oxidizing reagent (iodine and water) with a sulfurizing reagent transports a sulfur atom to the phosphorus during its P(III) to P(V) transition. The first such reagent utilized was elemental ($S_8$) sulfur[108,148,149,152], however pure sulfur has many drawbacks. It is plagued with organic solubility issues that make it unsuitable for machine use, it is poorly efficient in P(III) to P(V) conversion kinetics (~8 min), and also has a putrid smell that will permeate an entire research floor. A successor was developed, 3*H*-1,2-benzodithiol-3-one 1,1-dioxide, which became known as "Beaucage reagent"[150,153]. This reagent is soluble in acetonitrile, features rapid and efficient thioation kinetics (30 sec), and soon became the industry standard for transporting sulfur into phosphites[151,154]. Over the next few years several competing sulfurizing reagents were also suggested: tetraethylthiuram disulfide (TETD)[152,155], phenylacetyl disulfide[153,156], dibenzoyl tetrasulfide[154,157] and its derivative bis (ethoxythiocarbonyl) tetrasulfide[155,158], bis(*O,O*-diisopropoxyphosphinothioyl) disulfide (S-Tetra)[156,159], benzyltriethyl-ammonium tetrathiomolybdate (BTTM)[157,160], and bis(4-methoxybenzenesulfonyl) disulfide aryl derivatives[158,161], but none were a major improvement over Beaucage reagent in solid- phase synthesis and did not achieve widespread use beyond their respective labs. However it became apparent that there were three major limitations of Beaucage reagent. First, its long-term stability on solid-phase

synthesis machines was questionable. Second, it has poor efficiency when sulfurizing RNA phosphites due to their bulky protecting group on the 2'-OH. And lastly but most important, a by-product formed from Beaucage reagent during sulfurization, 3*H*-2,1-benzoxathiolan-3-one-1-oxide, is an oxidizing agent that will lead to a small percentage of phosphodiester bonds in the mixture. To overcome these deficiencies, Robert Hammer's lab developed two new sulfurizing reagents with a thiazoline ring structure: 1,2,4-dithiazolidine-3,5-dione (DtsNH) and 3-ethoxy-1,2,4-dithiazoline-5-one (EDITH)[159,162]. His reagents had long-term stability in acetonitrile, could be used at low concentrations (0.05 M), had rapid kinetics (30 sec), and most importantly were effective at sulfurizing RNA phosphites regardless of their additional protecting groups. Of the two, EDITH was advanced as being slightly better and became the base structure from which other groups would develop derivatives in their efforts to drive down production costs for sulfurizing reagents. These successors would be 3-methyl-1,2,4-dithiazolin-5-one (MEDITH)[160,163], 3-amino-1,2,4-dithiazole-5-thione (ADTT)[161,164], and 3H-1,2,4-dithiazole-3-thione (DDTT)[162,165]. Because of its fast kinetics, long term solubility, high performance with DNA, RNA and modified bases, DDTT has become the superior choice for sulfur transferring reagent and is now widely available commercially as "Sulfurizing Reagent II"[163,166], (Beaucage's is the first).

### PHOSPHORODITHIOATES

With increased nuclease resistance the therapeutic potential for thioated oligonucleotides became apparent, however the diastereomeric nature of monothioates makes them unsuitable for regulatory approval. Often a small molecule drug will contain one chiral center, and in the approval process the FDA requires that both enantiomers of the racemate be studied in detail. It would be logistically impossible to satisfy this requirement for a highly monothioated oligonucleotide therapeutic. Therefore an impetus was placed on developing chemistries for sulfurizing both non-bridging oxygens to produce a dithiophosphate, which is achiral about the phosphorus.

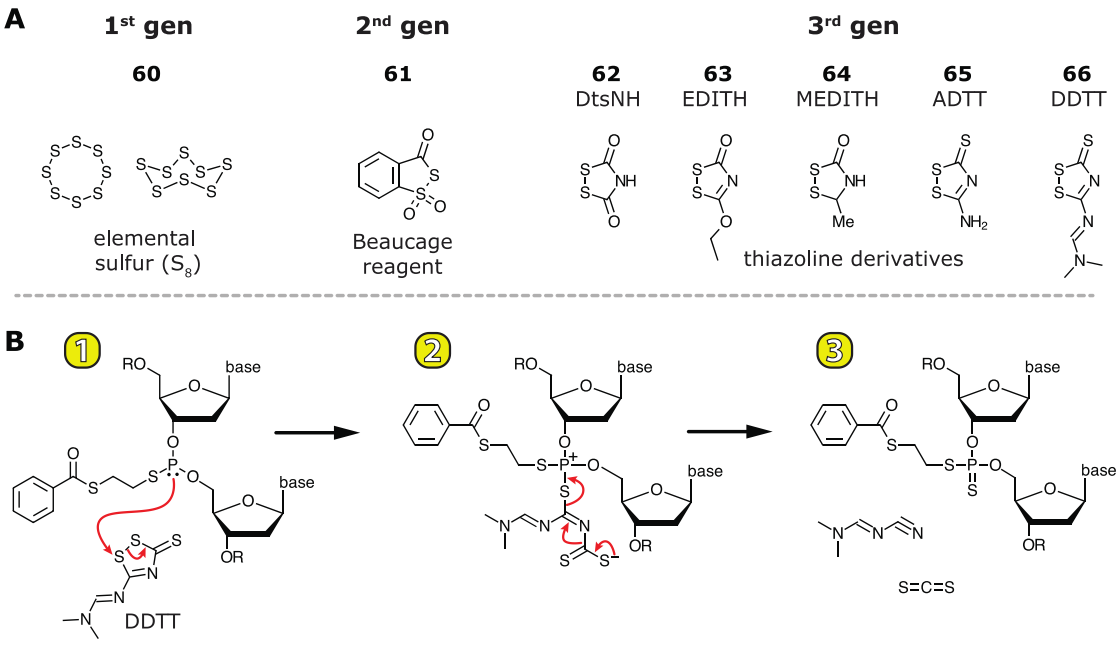



**Fig. 1.15) Sulfurizing Reagents and Mechanism of DDTT**

Sulfur is incorporated by altering the oxidation reagent for P(III) to P(V) conversion. **A)** Development of sulfurizing reagents over the last 30 years. Elemental sulfur ($S_8$) gives way to the popular Beaucage reagent. Third generation reagents feature a thiazoline ring structure, numerous derivatives are produced, resulting in the modern DDTT. **B)** The reaction mechanism of DDTT sulfurization involves nucleophilic attack and a splitting of the ring structure into two groups, transporting onto the phosphate a sulfur.

There are two ways to approach this problem, the first would be link two nucleotides together and then insert *both sulfurs afterwards*, and the second approach would be to develop an amidite *pre-primed with one sulfur atom and then sulfurize once after linkage*. The first approach was just an extension over the current practice, so it was the first to be examined. Caruthers put forward a technique[164,167] involving a phosphite linkage that was treated with hydrogen sulfide to create a H-phosphonothioate intermediate. This was subsequently sulfurized using elemental sulfur to incorporate the second atom (this period was several years prior to 2nd generation sulfurizing reagents). In a subtle improvement, it was found that replacing hydrogen sulfide with 4-chlorobenzyl mercaptan[165,168] had better yields. Although the goal of producing dithioated phosphates was achieved, this method is not amenable to automation as it requires offline-sulfurization steps. It was apparent that in order to truly advance dithioate synthesis and usage in biochemical studies development of a thiophosphoramidite was required.

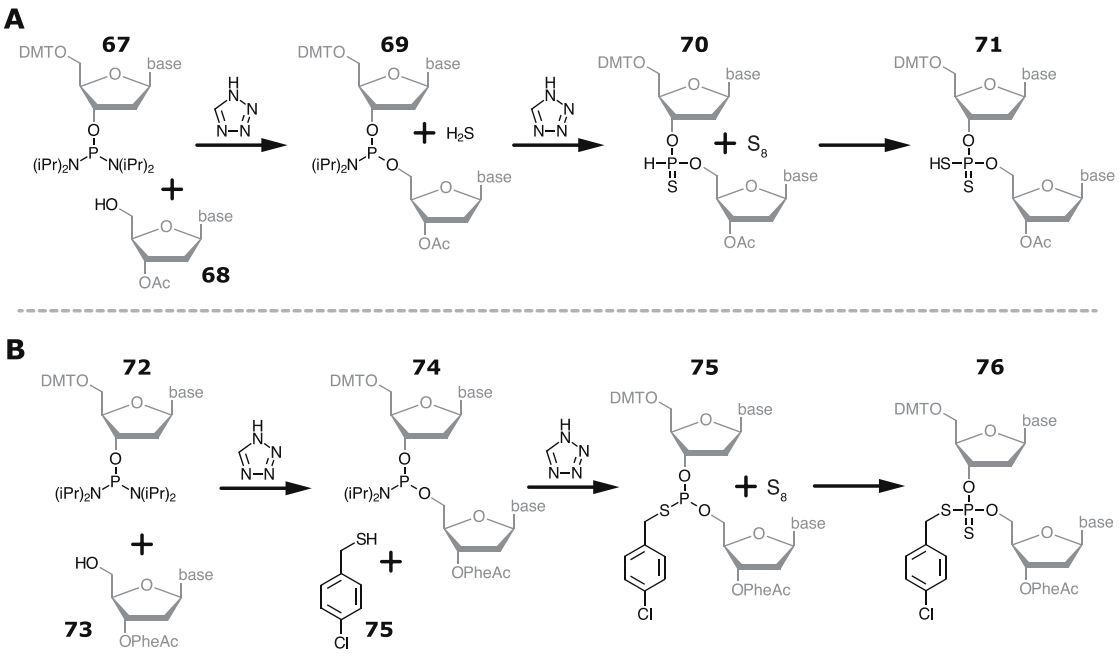



**Fig. 1.16) Dithioation After Phosphite-Bond Formation (Two Sulfurizing Steps)**

Caruthers advances a double sulfurization procedure for dithioate production. **A)** Hydrogen sulfide in the presence of activator incorporates the first sulfur, producing a P(V) H-phosphonothioate linkage **70**. The next sulfur is incorporated using a standard sulfurizing reagent, $S_8$. After, there is a trivial dehydrogenation to the dithiophosphate (omitted for brevity) **B)** 4-chlorobenzyl mercaptan **75** is adopted as an improved method for incorporation of the first sulfur.

## THIOPHOSPHORAMIDITES

During the late 1980s there was an interesting race between the lab of Caruthers and the lab of David Gorenstein towards developing the first thiophosphoramidite. Both submitted their manuscripts using different approaches to Tetrahedron Letters in the Summer of 1988, and in the Fall when published it was shown that Caruthers[166,169] had just slightly edged out Gorenstein[167] by 4 weeks with his submission. However, Gorenstein was also simultaneously preparing a patent on the process for preparing dithioated oligos using thiophosphoramidites, which was filed with the United States Patent & Trademark Office in 1989[166,168]. Caruthers' thiophosphoramidite patent was filed in 1991[169-171] as a continuation-in-part of an abandoned application having a filing date of 1989, which itself was a continuation-in-part of an abandoned application from 1988. Since the carry-over of claims is murky and the argument of prior art would require intense litigation,

the situation becomes an interesting case study for ascribing invention within the legal framework of the US patent system. Nevertheless, what is important is that now the field of oligonucleotide synthesis had access to thiophosphoramidites that could be used to produce dithioated phosphates.

The two groups utilized slightly different approaches to develop the thiophosphoramidite, although both involved a sulfur-containing phosphitylating reagent. Gorenstein created chloro-*N,N*-diisopropylaminothiomethoxyphosphine by reacting dichloro-*N,N*-diisopropylaminophosphine with sodium methoxide using aluminum trichloride and potassium iodide as catalysts. This chloro-thiophosphine **79** then phosphitylated the 3'-OH of a nucleoside to create the thiophospho-ramidite[167,172]. Caruthers developed his amidite by using S-(4-chlorobenzyl)-*N,N,N',N'*-tetraisopropylaminophosphine **81** as the phosphitylating reagent[166,173]. Gorenstein's and Caruthers' thiophosphoramidite would then couple with the 5'-OH of the previous base after activation to form a P(III) linkage containing the first sulfur atom. After a sulfurization step to oxidize up to the P(V) state, a dithiophospho-diester bond was formed.

Over time there was subtle evolution utilizing different phosphite protecting groups[162,170,171], until a consensus on the modern preparation of thiophosphoramidites was reached. The amidites commercially available today are prepared using large "one-pot" batches. Phosphitylation of the 3'-OH of a protected nucleobase is achieved with pyrrolidino phosphine **85** under tetrazole catalysis, which yields bis(pyrrolidino) phosphite **86**. This intermediate is then converted to the thiophosphoramidite by treatment with the sulfur-containing monobenzoyl-ethanedithiol **88** under tetrazole[172,174]. These thiophosphoramidites are stable and can be stored as dry powders for long periods. They are commercially available from Glen Research and AM Biotechnologies.

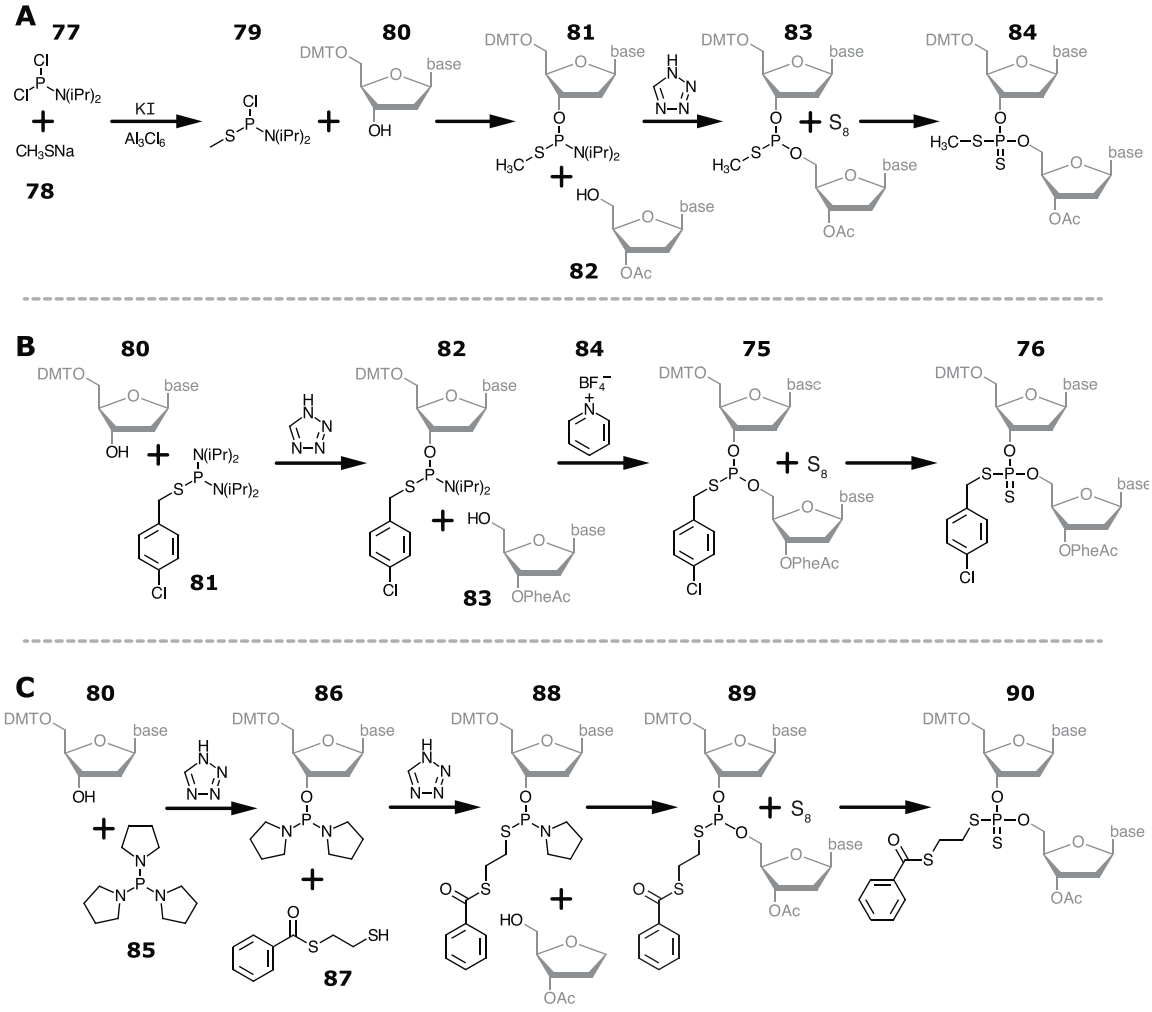



**Fig. 1.17) Three Approaches to Synthesizing Thiophosphoramidites**

**A)** Gorenstein's first published thiophosphoramidite **81** is produced by phosphitylating with chloro-*N,N*-diisopropylaminothiomethoxyphosphine **79**. **B)** Caruthers' first published thiophosphoramidite **82** is produced after phosphitylating with S-(4-chlorobenzyl)-*N,N,N′,N′*-tetraisopropylaminophosphine **81**. **C)** The work up for producing the *modern* thiophosphoramidite **88** is shown. The nucleobase is first phosphitylated with pyrrolidino phosphine, and then the sulfur atom comes from monobenzoylethanedithiol **87**. This process has been scaled and thiophosphoramidites, **88** are now commercially available.

## 1.4) Methods: Oligo Production

### METHOD: NORMAL (PO) AND DITHIOATE (PS2) DNA SYNTHESIS

Unless noted otherwise all oligonucleotides used for experiments were made on an Expedite 8909 Nucleic Acid Synthesis System (Applied Biosystem; Foster City, CA) using phosphoramidite chemistry with CPG solid-phase supports. An inert environment was maintained using pure Helium or Argon gas (Matheson Tri-Gas; Houston, TX). Organic solvents, reagents, phosphoramidites, thiophosphoramidites, and CPGs were sourced from Glen Research (Sterling, VA), Sigma-Aldrich (St. Louis, MO), and Thermo Fisher Scientific (Waltham, MA). Phosphoramidites were solubilized from dry powders into anhydrous ACN, which also served as the washing reagent. Because of reduced solubility thiophosphoramidites utilized a hybrid mixture of ACN/DCM (9:1 v/v). Deblocking was performed with 3% (w/v) TCA in DCM. The activator used was 0.45 M tetrazole sublimed in anhydrous ACN. Two capping reagents were mixed on the column; Cap A: THF and $AcO_2$ (9:1), and Cap B: 10% methylimidazole in THF/$AcO_2$ (8:1). The oxidizing reagent used was 0.02 M iodine in a mixture of THF/pyridine/$H_2O$. For dithioates, the sulfurizing reagent employed was DDTT (3H-1,2,4-dithiazole-3-thione) in pyridine/ACN (3:2).

Synthesis of dithioate DNA is very expensive since thiophosphoramidites are roughly 30 times more than the cost of their unmodified counterpart. Typically, 0.25g of thiophosphoramidite is solubilized in 3.0 mL of ACN/DCM. After priming the system (~0.5 mL), enough monomer remains to incorporate ~15 additions at the 1-µmol scale and ~5 additions at the 15-µmol scale. Synthesizing a 15-µmol scale 100% dithioated 24-bp duplex would exceed $3000 in thiophosphoramidites cost. Therefore, preparing dithioate DNA requires rigorous preparation and advanced planning, and the machine must be functioning at optimal efficiency. It is good practice to replace all other reagents with their freshest possible stock, since they are comparatively cheap. Managing the flow volumes needed for coupling is something I iteratively optimized over time.

## Table 1.1) PO Phosphoramidite Coupling Protocol (for Expedite 8909)

| Step | Reagent | Purpose | mL | sec | mL | sec |
|---|---|---|---|---|---|---|
| **1) Detritylation** | | | 1-µmol scale | | 15-µmol scale | |
| A) | 3% TCA in $MeCl_2$ | Deblock | 0.15 | – | | |
| B) | 3% TCA in $MeCl_2$ | Deblock | 0.75 | 60 | 7.5 | – |
| C) | ACN | Wash | 0.6 | – | 0.75 | – |
| D) | ACN | Wash | 0.6 | – | 6.0 | – |
| **2) Coupling** | | | | | | |
| A) | ACN | Wash | 0.08 | – | 0.6 | – |
| B) | Tetrazole | Prime activator | 0.08 | – | 0.53 | – |
| C) | Amidite + tetrazole | Add base | 0.08 | – | 0.38 | – |
| D) | Amidite + tetrazole | Add base/couple | 0.03 | 16 | 0 | 120 |
| E) | Tetrazole | Couple | 0.05 | 24 | 0.3 | 30 |
| F) | | | Repeat 2B (1µm maybe, 15µm yes 2x) | | | |
| G) | ACN | Wash | 0.6 | 56 | 1.5 | – |
| **3) Capping** | | | | | | |
| A) | ACN | Wash | 0.3 | – | 1.5 | – |
| B) | CapA + CapB (1:1) | Cap unreacted 5'-OH | 0.12 | – | 1.2 | – |
| C) | ACN | Wash | 0.9 | 15 | 0.23 | 40 |
| D) | ACN | Wash | 0.21 | – | 1.5 | – |
| **4) Oxidizing** | | | | | | |
| A) | $I_2$/pyridine/$H_2O$ | Oxidize P(III) to P(V) | 0.23 | – | 1.88 | – |
| B) | ACN | Cap unreacted 5'-OH | 0.23 | – | 1.5 | – |
| **5) Capping** | | | | | | |
| A) | CapA/CapB (1:1) | Cap resid. $H_2O$ from 4a | 0.11 | – | 0.75 | – |
| B) | ACN | Wash | 0.45 | – | 5.1 | – |

## Table 1.2) PS2 Thiophosphoramidite Coupling Protocol (for Expedite 8909)

| Step | Reagent | Purpose | mL | sec | mL | sec |
|---|---|---|---|---|---|---|
| **1) Detritylation** | | | 1-µmol scale | | 15-µmol scale | |
| A) | 3% TCA in $MeCl_2$ | Deblock | 0.15 | – | | |
| B) | 3% TCA in $MeCl_2$ | Deblock | 0.75 | 60 | 7.5 | – |
| C) | ACN | Wash | 0.6 | – | 0.75 | – |
| D) | ACN | Wash | 0.6 | – | 6.0 | – |
| **2) Coupling** | | | | | | |
| A) | ACN | Wash | 0.08 | – | 0.6 | – |
| B) | Tetrazole | Prime activator | 0.08 | – | 0.53 | – |
| C) | Thioamidite + tetrazole | Add base | 0.1 | – | 0.38 | – |
| D) | Thioamidite + tetrazole | Add base/couple | 0.03 | 136 | 0 | 360 |
| E) | Tetrazole | Couple | 0.06 | 272 | 0.3 | 180 |
| F) | ACN | Couple/Wash | 0.03 | 136 | 0.6 | – |
| G) | | | Repeat from 2B (2x total) | | | |
| H) | ACN | Wash | 0.38 | – | 1.5 | – |
| **3) Sulfurizing** | | | | | | |
| A) | DDTT in pyridine/ACD | Sulfurize | 0.3 | 5 | 1.2 | 20 |
| B) | ACN | Wash | 0.9 | 60 | 0.45 | 60 |
| C) | | | Repeat from 3B (4-8x total) | | | |
| **4) Capping** | | | | | | |
| A) | ACN | Wash | 0.3 | – | 1.5 | – |
| B) | CapA + CapB (1:1) | Cap unreacted 5'-OH | 0.12 | – | 1.12 | – |
| C) | ACN | Wash | 0.09 | 15 | 0.23 | 40 |
| D) | ACN | Wash | 0.21 | – | 1.5 | – |
| E) | CapA + CapB (1:1) | Cap unreacted 5'-OH | 0.11 | – | 0.75 | – |
| F) | ACN | Wash | 0.45 | – | 5.2 | – |

Tables 1.1 and 1.2 above show the optimized volume control protocol for programming the Expedite 8909 machine (seconds is time used to pump, where "–" equals as fast as possible). After synthesis is complete oligos are cleaved from the CPG support using concentrated $NH_3OH$, and then deprotected at 55°C for 16 hours. In attempts to increase yields for dithioated DNA, CPG cartridges can be cut open and the beads themselves deprotected for 16 hours at 55°C to ensure complete cleavage. Instead of pure $NH_3OH$ dithioates were deprotected in $NH_3OH$/EtOH (3:1) with 1mM DTT. After deprotection the liquid was diluted 4-fold with 18 MΩ $H_2O$, syringe filtered through 0.22 um, and then frozen at -80°C for 1 hour before being lyophilized to complete dryness. In this powdered form, oligos are stable for months at room temperature and years when frozen.

### METHOD: DNA PURIFICATION BY RP-HPLC

The final 5'-DMT moiety was kept on the synthesized product to act as a hydrophobic anchor for RP-HPLC purification using an AKTA Purifier FPLC/HPLC combo system. Lyophilized oligos were resuspended in running buffer composed of 100 mM triethylammonium acetate (TEAA) (pH 8.0), syringe filtered (0.22 μm), and injected onto a Hamilton PRP-1 column (10 μm, 100 Å, 7.0 x 305 mm) at a flow rate of 1-2 ml/min (~8-10 MPa). For purification the flow rate was increased to 3-4 ml/min (10-12.5 MPa) and the sample was eluted with a gradient of ACN (~0.5-1%/min v/v) during which automated fractionation is enabled. UV absorbance at 254 nm and electrical conductivity were used to monitor progress of the run. The full-length product with its hydrophobic DMT moiety eluted as the last major peak in the purification. The capped synthetic failures (n-1, n-2, etc…) lack the hydrophobic DMT moiety and elute prior to the final product. After the main peaks were collected, Tris base was added to the sample (~5 mM) before freezing at -80°C for 1 hour. The triethylammonium component of the TEAA buffer is more volatile than the acetic acid, which reduces pH as the sample is lyophilized. Tris base is non-volatile and maintains basicity to prevent depurination. After lyophilization the DMT group was removed by brief treatment with 80% acetic acid. Depurination will occur quickly, so it is important not to exceed 5 minutes with this step. After

detritylation the oligo is resuspended in TEAA running buffer and injected onto the column for a second purification.

At high oligo concentrations the cleaved DMT will gum and clog the syringe, resulting in sample loss which is unacceptable for expensive dithioate molecules. While working with fluorescent dye-containing oligos (which I could visually follow) I developed a superior detritylation protocol utilizing the synthesis deblock reagent and aqueous-organic phase separation to clear the DMT prior to syringe-filtration. Instead of 80% acetic acid, I added TCA/DCM deblock reagent to the lyophilized oligo. In this organic phase (DCM) the oligo condenses and goes to the air-liquid interface because of its polyanionic character. Adding TEAA buffer (or $H_2O$) and agitating drew both the oligo as well as the TCA into the aqueous phase, which lowers its pH causing a gentle detritylation. The hydrophobic DMT group migrates down into the DCM, eliminating it from the sample. After 5 minutes, the aqueous phase was removed by syringe and filtered before injecting onto the PRP-1 column. In addition to eliminating the gum problem, this method also significantly reduces the amount of depurination, as determined by mass spectrometry analysis.

The detritylated sample undergoes a second RP-HPLC run, where it will be the only major elution peak. Because of the lack of hydrophobic anchor, the final product oligo elutes earlier than its DMT counterpart. The appropriate fractions are then collected, treated with Tris base, frozen, and then lyophilized. Dry oligos are resuspended in TE buffer (10 mM Tris, pH 8.0, 1 mM EDTA) for quantitation and quality control analysis.

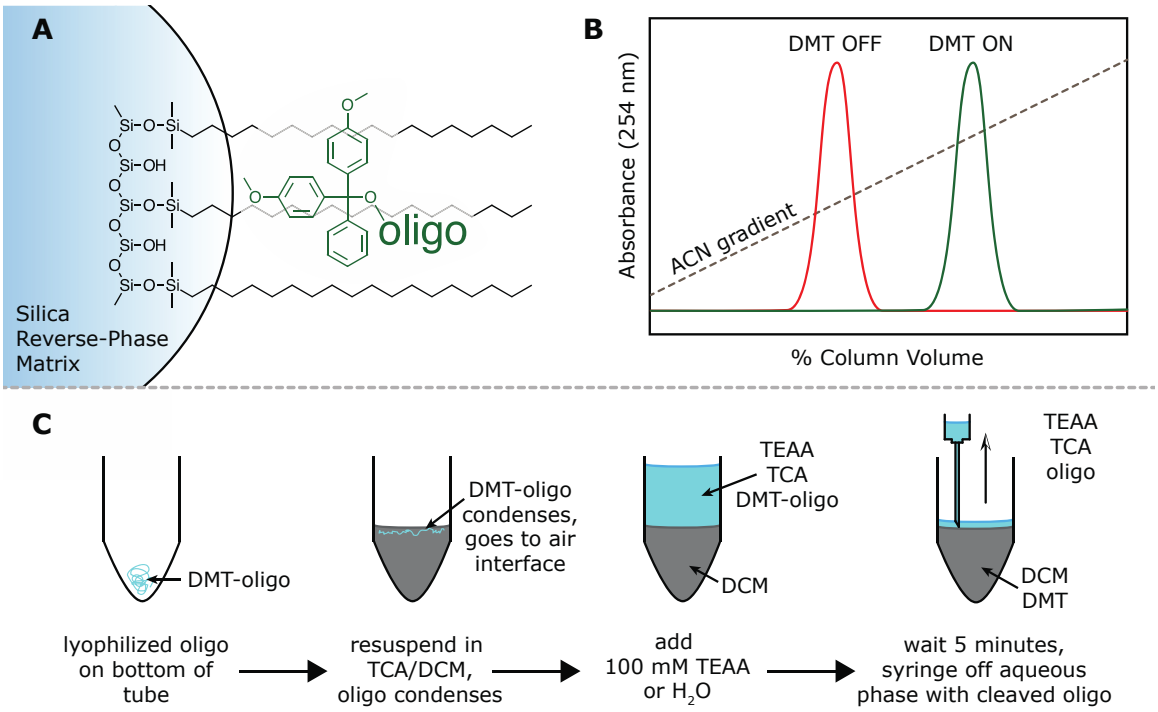



**Fig. 1.18) Ideal RP-HPLC Purification of DMT-ON and DMT-OFF Oligos**

**A)** PRP-1 RP-HPLC column is chemically similar to C18 silica. Long alkyl chains on the matrix retain the hydrophobic DMT moiety at the 5'-end of the full-length oligo. **B)** Two ideal RP-HPLC runs are overlaid. The DMT-ON full-length oligo is purified once, then detritylated and repurified, eluting earlier. **C)** Improved detritylation procedure utilizing aqueous/organic phase separation to clear cleaved DMT as well as reduce depurination.

## METHOD: LINKER CHEMISTRY OF A 5'-FLUOROPHORE (TEXAS RED)

The high-throughput thermal melts of Chapter 3 utilized a fluorophore-quencher system as an assay of DNA duplex hybridization. The system will be discussed in more detail there, but here I show the amine-crosslinking chemistry involved in creating the fluorescent molecules and the rational behind purification. Oligonucleotides were modified at the 5'-end by coupling a six-carbon TFA-protected amino phosphoramidite. After synthesis was complete the oligo was cleaved and deprotected as normal, which also removed the base-labile TFA group from the linker to reveal a primary amine ($NH_2$) suitable for crosslinking reactions. Because fluorescent moieties are often sensitive to harsh basic conditions at elevated temperatures, fluorophores are attached after the deprotection step. The oligo was buffer exchanged using Amicon Ultra 3K MWCO (EMD Millipore) spin centrifugal filters into reaction buffer (20 mM $NaHPO_4$, pH 8.0 slowly adjusted from

pH 7.0 by titration with 100 mM $NaHCO_3$). Five milligrams of the fluorescent moiety, Texas-Red-*N*-hydroxysuccinimide ester, was dissolved in anhydrous DMSO and was added to the oligo in reaction buffer at an excess molar ratio of 4:1 (dye/oligo). The coupling reaction proceeded at room temperature for four hours, and was then quenched by addition of Tris base to a concentration 20 mM and frozen at -80°C until purification. Since the fluorescent dye is large and hydrophobic, it is purified similarly to a DMT-ON oligonucleotide by RP-HPLC. There is no second "DMT-OFF" purification, but it is not needed as the fluorescent color of the molecule makes it easy to distinguish the coupled full-length product from uncoupled and failure sequences.

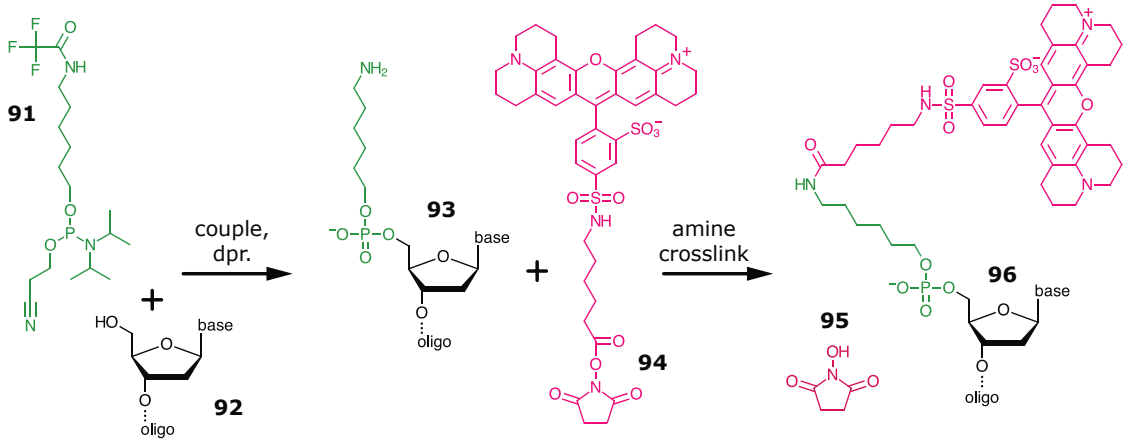



**Fig. 1.19) 5'-Amine Crosslinking Chemistry for Coupling Texas Red Fluorophore**

Texas Red is attached to the 5'-end of oligonucleotides via NHS ester-amine crosslinking. During the synthesis a terminal 6-carbon linker with a protect amine (green) is attached using the phosphoramidite method to the 5'-OH of the last base. The oligo is cleaved, deprotected to reveal a primary amine, and then coupled with Texas-Red-NHS ester (pink) under slightly alkaline conditions. Because of the fluorophore's hydrophobic properties the product is purified via RP-HPLC similarly to a DMT-ON oligo.

## METHOD: DNA PURIFICATION BY DENATURING PAGE

Adding dithiophosphates increases the hydrophobicity of oligos such that eventually the DMT moiety on the 5'-end will no longer sufficiently separate full-length product from failure sequences during RP-HPLC. Although anion-exchange chromatography has been suggested as an alternative column purification method[173,174], during my studies I opted instead for large-scale PAGE purification of heavily dithioated oligonucleotides. Oligonucleotides are well suited for electrophoretic movement due to their polyanionic character and so they separate

based on size when moving through the acrylamide matrix. For these purifications, the DMT is not needed and were therefore removed ahead of time on the Expedite system before deprotection.

The resolving power of electrophoresis is dependent on the length of the PAGE gel and the voltage applied so the purification of full-length oligos from n-1 deletion failures requires massive gels. For my studies I utilized custom glass plates 0.6 m tall by 0.35 m wide (Moliterno Inc.; Pepperell, MA). There are two types of glass currently available, soda-lime and borosilicate. They are easily distinguishable by looking at an edge, which will appear opaque green with soda-lime and crystal-clear with borosilicate. Besides appearance, they differ in their thermal expansion coefficients and tensile strength. Borosilicate performs excellently under high heats (e.g. Pyrex is made from borosilicate), but is extremely fragile when handling. Soda-lime on the other hand, is much more durable but will easily crack under the intense heat of high-voltage electrophoresis. I often ran my gels between 300-600 volts, and after several cracked soda-lime plates I decided that borosilicate was the superior choice, despite its fragility.

Denaturing PAGE gels were prepared by polymerizing a solution consisting of 7 M urea, 15-20% acylamide (acryl/bis 37.5:1), and 1x TBE (Tris/Borate/EDTA) with 0.07% v/v TEMED and 0.07% v/v APS. Two plates were separated by plastic edge spacers and sealed with electrophoresis tape (3M; St. Paul, MN). Approximately 500-700 mL of polymerizing acrylamide solution was required to cast each gel. Following solidification of the acrylamide gel the plates were clipped to a custom-made vertical gel apparatus with top and bottom reservoirs in which 1x TBE electrolyte solution was poured. The lyophilized oligo was resuspended in formamide and two dyes were added to monitor electrophoretic progress towards the cathode. In the case of a 20-30 nucleotide- long oligo xylene cyanol dye will migrate behind, and bromophenol blue migrates ahead.

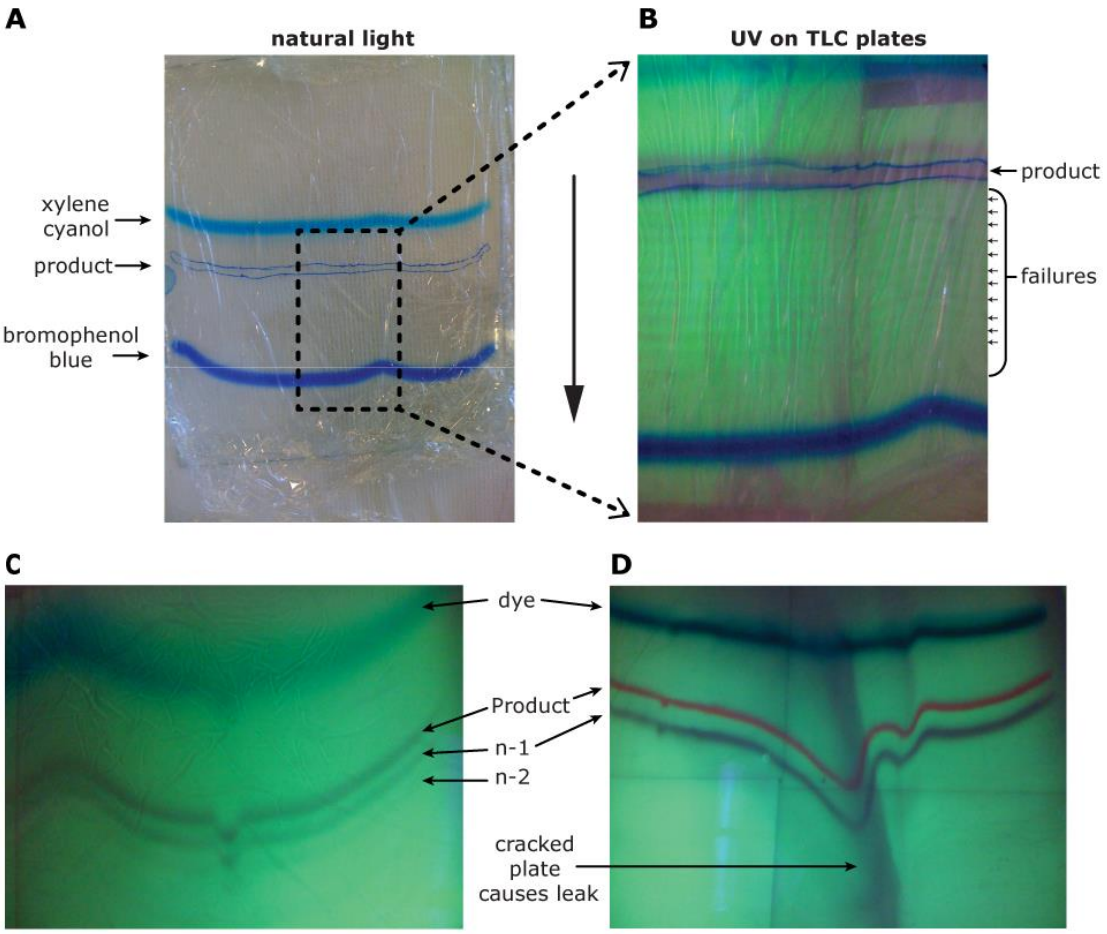



**Fig. 1.20) Examples of Denaturing PAGE Purification**

Large (35 cm x 60 cm) PAGE purification separates DNA based on size. A) Appearance under normal lighting. DNA migrates in between two visible dyes used for tracking. B) DNA shadowing by TLC plates and UV light shows a large major band for the full-length product. Below this but barely visible are striated bands representing n-1, n-2, etc… capped failures, which migrate faster than the full-length product. C) Another purification by PAGE shows the failures in better detail. Gel imperfections, such as cast air bubbles, cause distortions to the bands that propagate through the entire gel run (i.e. the bump in middle). D) Fluorescent oligos are coupled after synthesis, which can be inefficient. The pink final product is shown, with the uncoupled DNA migrating below. This gel also demonstrates the cracking of a soda-lime glass plates that occurs using 300-600 volts, causing a smearing of the oligo. Product was still recoverable from unsmeared portions

It took approximately 16-20 hours when running between 300-600 volts for the bromophenol blue to migrate to the bottom of the gel. At this point the gel was removed from the apparatus and placed on clear polyvinyl chloride wrap. To visualize the oligo bands, DNA shadowing was used with underlying TLC plates and a handheld UV lamp. The largest slow-migrating band was excised as the full-length final product and the oligo was electroeluted using a Whatman Elutrap System (GE Healthcare; Little Chalfont, UK).

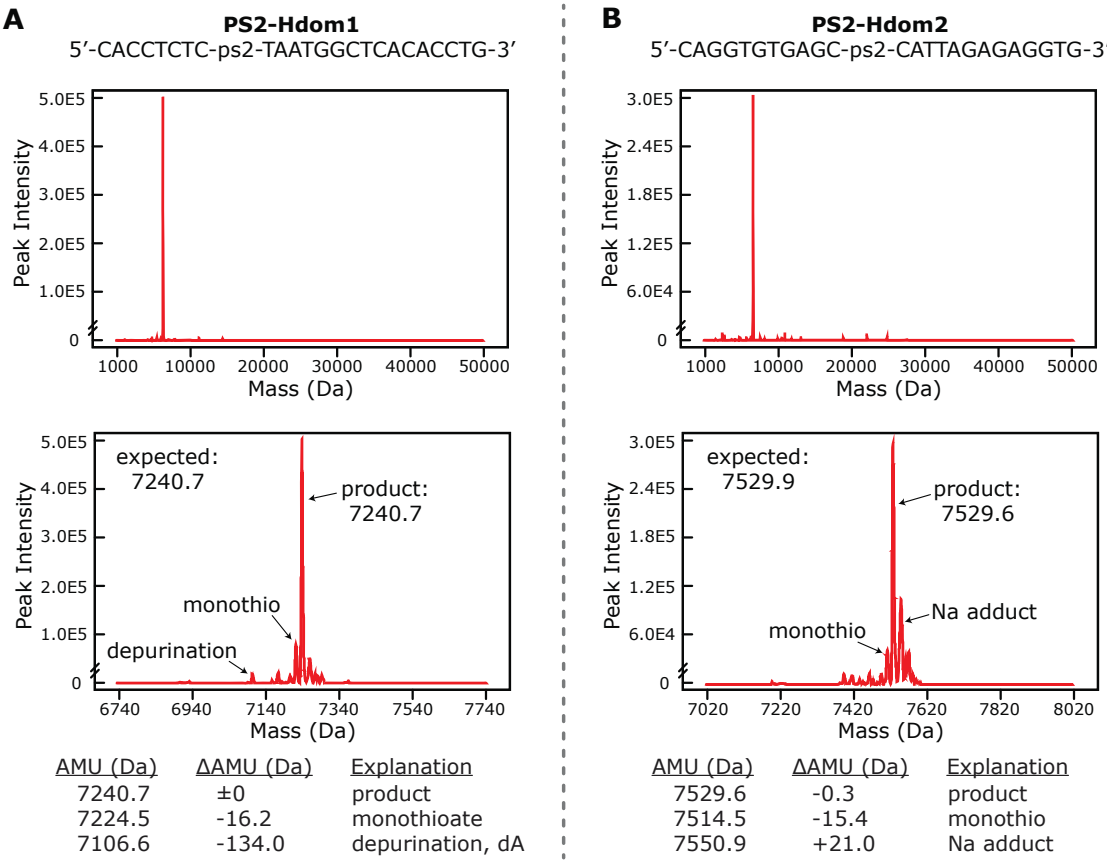
## METHOD: QC BY MASS SPEC AND CAPILLARY ELECTROPHORESIS

After purification samples of the oligonucleotides were sent to Novatia, LLC (Monmouth Junction, NJ) for LC-ESI-MS analysis on a Novatia Oligo HTCS system. The combined liquid chromatography and mass spectrometry protocol gives an insightful measure of the quality of the synthesized oligo. During the last few years we began using Integrated DNA Technologies (Coralville, IA) for a combined LC-ESI-MS and CE analysis. IDT also uses Novatia's Oligo HTCS system and CE is performed on a Beckman P/ACE MDQ system (Beckman Coulter; Fullerton, CA). In both cases raw data were deconvoluted in-house at the companies and then delivered via webserver for interpretation. To pass quality control an oligo was required to show 95% purity via CE and have a mass accuracy within 2 daltons.

## 1.5) Results: Synthesis and Quality Control of Oligos from All Studies

### Result: Synthesis and Characterization of a Homeodomain-Recognition Sequence Containing a Single Dithioate Substitution

Two 24-base oligonucleotides corresponding to the recognition sequence for HoxD9 homeodomain were synthesized at the 15-µmol scale. Each had a single dithioate substitution and was purified by RP-HPLC. QC by LC-ESI-MS shows pure product with minimal contamination from the monothioated species.




| AMU (Da) | ΔAMU (Da) | Explanation |
|---|---|---|
| 7240.7 | ±0 | product |
| 7224.5 | -16.2 | monothioate |
| 7106.6 | -134.0 | depurination, dA |

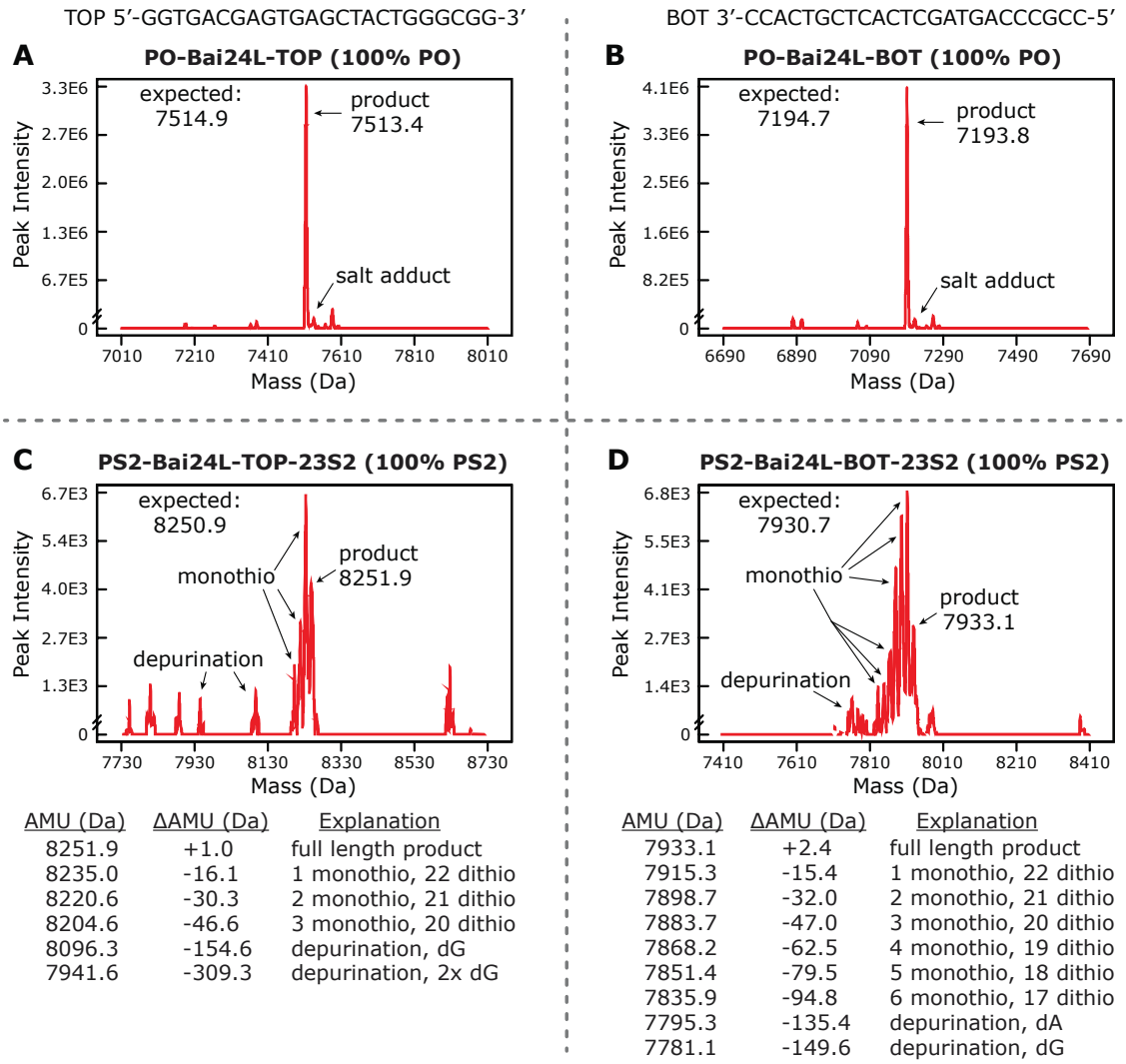
| AMU (Da) | ΔAMU (Da) | Explanation |
|---|---|---|
| 7529.6 | -0.3 | product |
| 7514.5 | -15.4 | monothio |
| 7550.9 | +21.0 | Na adduct |

**Fig. 1.21) ESI-MS of 24-Base Homodeomains With a Single PS2 (Ch. 3)**

Mass spectrometry QC shows that dithioates can be produced with high quality at the 15-µmol scale, a quantity which enables structural methods such as X-ray crystallo-graphy or NMR. **A)** The Hdom1 sequence shows a single major peak in the broad spectrum and when zoomed reveals only minor percentage of monothioate species (sulfurization was not successful). **B)** Hdom2 was produced in high quality as well, with one major peak in the broad spectrum. Upon zooming a sodium adduct is observed, which is a common phenomenon in ESI-MS.

## RESULT: SYNTHESIS AND CHARACTERIZATION OF A 24-BASEPAIR 100% DITHIOATED DUPLEX

To compare biophysical properties of dithioated and normal DNA I prepared two 24 base-pair duplexes, one with normal phosphodiester backbone and the other with both strands completely dithioated. To date, these two PS2 strands are the longest full phosphorodithioate oligonucleotides to be synthesized and purified. A total of four 1-μmol synthesis each were produced and the batches were pooled before purifying via denaturing PAGE. The mass spectrometry results for the PO oligos came back as expected, with one single peak representing the correct full-length PO product. The PS2 mass spectrometry results were very interesting since they provided insight into the efficiency of the sulfurization procedure. Previous reports with DDTT reported >98% efficiency in sulfurizing a thymidine 10-mer oligonucleotide to its monothio version (PO → PS)[162], however, my data suggests that for sulfurizing a monothio oligonucleotide to its dithioate version (PS → PS2) the efficiency falls to ~95%. As shown in Fig. 1.15 DDTT is a high MW sulfurizing reagent, so a possible explanation for the observed reduction in efficiency could be steric hindrance of DDTT by the sulfur atom in a PS oligo, which is larger than the oxygen in a PO atom. In addition to steric hindrance, a second explanation could be the different electron density distribution between a phosphite and a thiophosphate that disrupts DDTT's ideal reaction mechanism. And another explanation could be oxidative replacement during the deprotection step. Several intervals of 16 AMUs were observed, allowing us to identify them as varying degrees of monothioation. With 23 dithioate linkages in this 24-base oligonucleotide, it is reasonable to expect there to be some deviations from perfect dithioation throughout the entire length. Nevertheless, this molecule is of extremely high quality and suitable for biophysical studies.




| AMU (Da) | ΔAMU (Da) | Explanation |
|---|---|---|
| 8251.9 | +1.0 | full length product |
| 8235.0 | -16.1 | 1 monothio, 22 dithio |
| 8220.6 | -30.3 | 2 monothio, 21 dithio |
| 8204.6 | -46.6 | 3 monothio, 20 dithio |
| 8096.3 | -154.6 | depurination, dG |
| 7941.6 | -309.3 | depurination, 2x dG |

| AMU (Da) | ΔAMU (Da) | Explanation |
|---|---|---|
| 7933.1 | +2.4 | full length product |
| 7915.3 | -15.4 | 1 monothio, 22 dithio |
| 7898.7 | -32.0 | 2 monothio, 21 dithio |
| 7883.7 | -47.0 | 3 monothio, 20 dithio |
| 7868.2 | -62.5 | 4 monothio, 19 dithio |
| 7851.4 | -79.5 | 5 monothio, 18 dithio |
| 7835.9 | -94.8 | 6 monothio, 17 dithio |
| 7795.3 | -135.4 | depurination, dA |
| 7781.1 | -149.6 | depurination, dG |

**Fig. 1.22) ESI-MS of a 100% PS2 24-Base Pair Duplex and its PO Equivalent (Ch.2)**

Two duplexes were synthesized, one normal PO and the other containing 100% dithioated linkages. This is the largest complete PS2 duplex ever produced. **A)** and **B)** top and bottom strands of the PO duplex show single major peaks corresponding to their full length product. **C)** The 100% PS2 top strand shows the major product peak and then full-length product with varying degrees of monothioation. **D)** 100% PS2 bottom strand shows a peak for the correct product, the degree of monothioation is slightly higher. From these our sulfurization efficiency was estimated to be approximately 94%

## RESULT: SYNTHESIS AND CHARACTERIZATION OF 40-MODIFIED OLIGONUCLEOTIDES FOR PROBING THE THERMODYNAMICS OF DITHIOATE DNA-DUPLEX STABILITY

A set of 40 oligonucleotides was designed for probing dithioate effects on thermal stability of the DNA duplex. The design of these 12-base fluorophore-containing oligos is covered in detail in Chapter 3, but here their synthesis and characterization are discussed. Each 12-base sequence was synthesized in both its normal PO form and then its PS2 form containing two dithioate substitutions out of 11-internucleotide linkages. Shown below is a representative example of a PO-5'-Texas Red purification via HPLC and then QC by LC-ESI-MS.

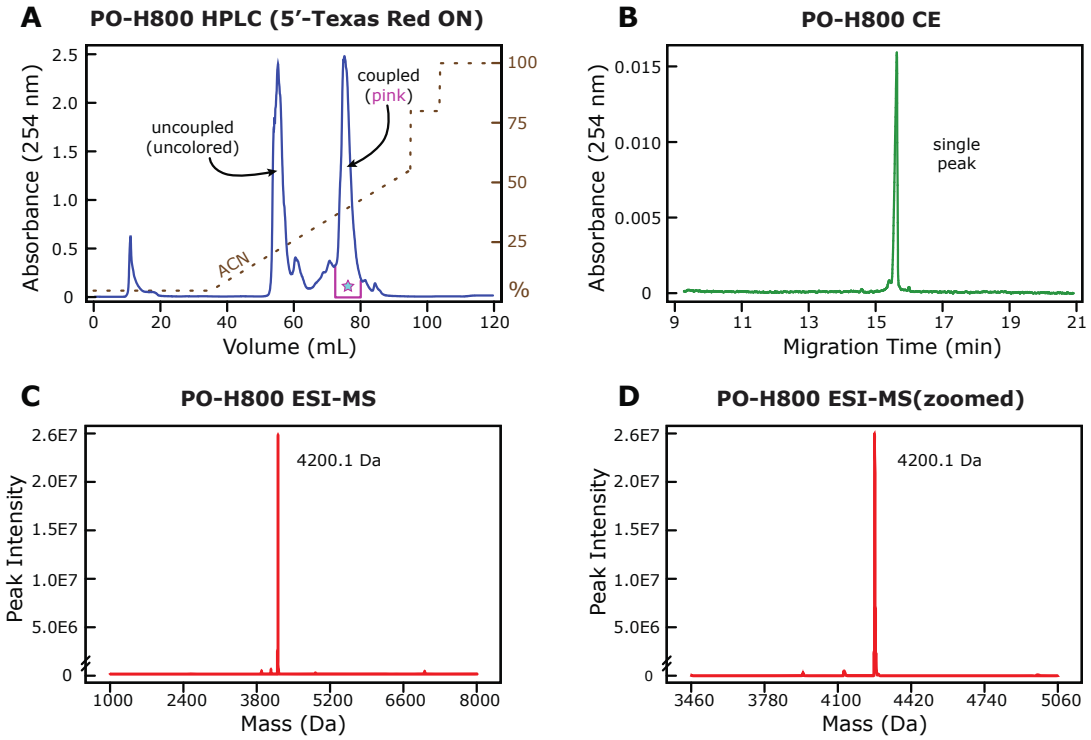



**Fig. 1.23) Purification and QC of a PO-5'-Texas Red Fluorophore Oligo**

**A)** HPLC purification of PO-H800 shows a normal chromatogram for a Texas Red oligonucleotide. The second major peak is bright pink in color, representing the full-length coupled product. The previous peak of failure sequences and uncoupled oligos is uncolored. **B)** The product peak CE run shows one species. **C)** and **D)** Both the ESI-MS spectra show well behaved, single peak corresponding to the correct MW for full-length product.

Next is show what the purification of a 3'-Iowa Black RQ oligo looks like. This molecule is purified by a 5'-DMT moiety, which is then removed using the modified detritylation protocol discussed earlier, before a second purification run.

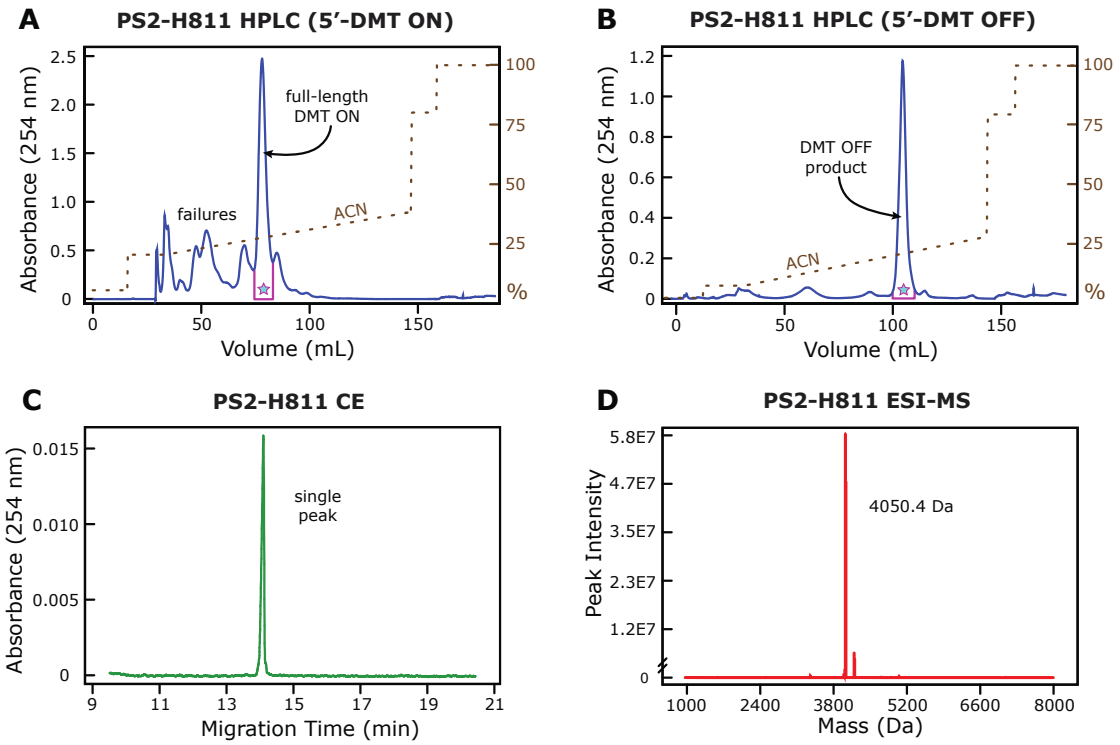



**Fig. 1.24) Purification of and QC of a PS2-3'-Iowa Black Quencher Oligo**

**A)** HPLC purification of DMT-ON oligo. Since the dye is on the 3'-end, all peaks and fractions contain are blue in color. The last major peak is collected as the full-length product and detritylated using the modified protocol described earlier. **B)** The DMT-OFF purification shows one major final product peak. **C)** CE shows a single species. **D)** ESI-MS shows one major product peak corresponding to the correct length, but also a small extraneous peak, which was a trend with dithioates.

During characterization there were significant differences in the performance of the PO and PS2 versions of certain sequences in ESI-MS analysis. In about ten instances the PO oligo behaved ideally whereas its PS2 version showed the tendency to decompose and fragment during ionization, resulting in numerous extraneous peaks registering on the detector. At first I thought this was contamination due to synthetic failures, which was surprising given how cleanly the HPLC was performing and that the CE peaks showed only a single PS2 species. So I resynthesized a second batch of PS2 oligos (Batch#B) that included an additional purification run on the HPLC. However, the fragmentation was still evident which led me to believe that this is just a peculiar behavior of dithioates in ESI-MS.

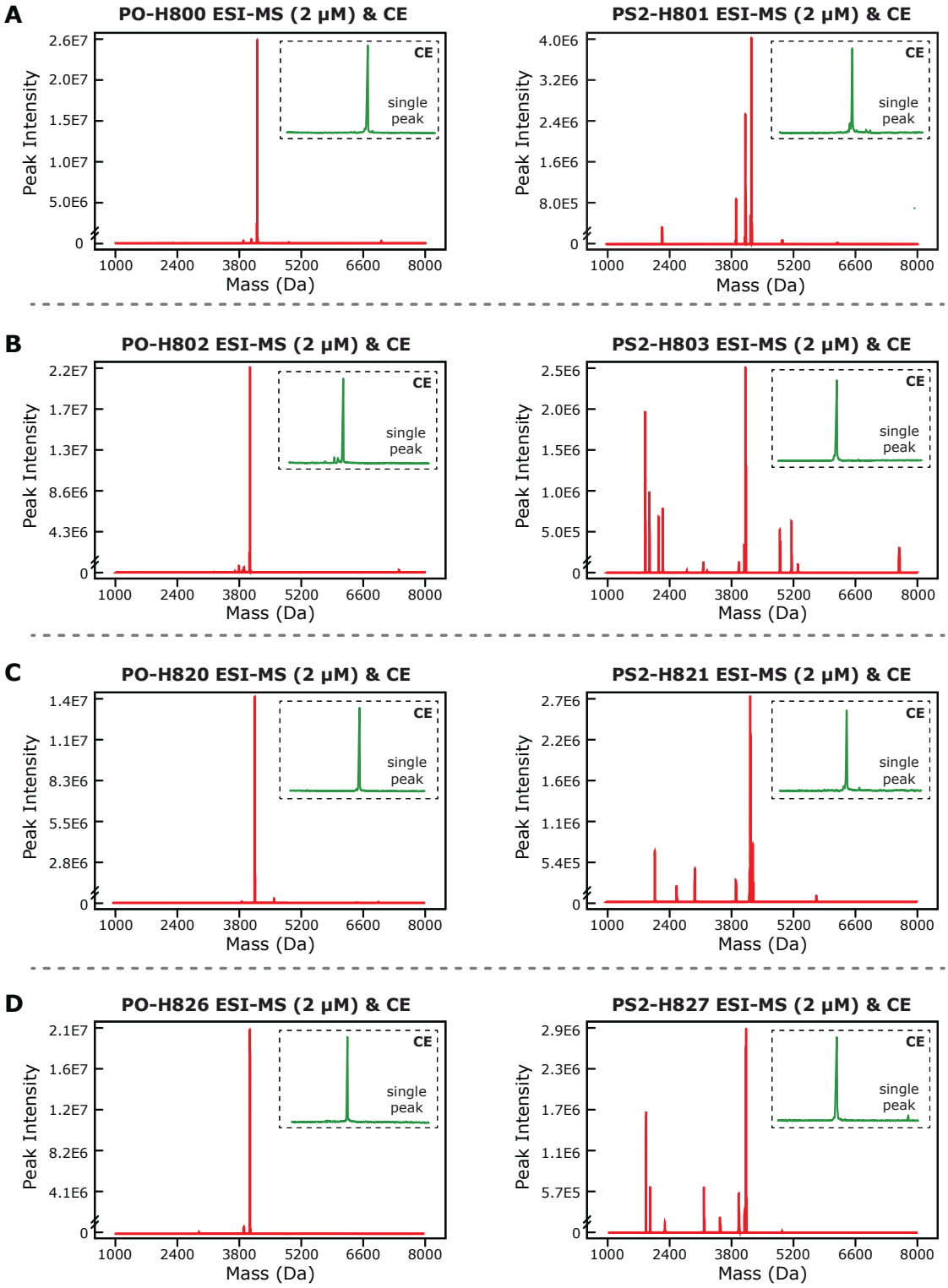



**Fig. 1.25) PS2 Oligos Often Fragment During Ionization in Mass Spectrometry**

**A)**, **B)**, **C)**, **D)** are each a different sequence (see Table 1.3). The normal PO version is on the left, and the PS2 version (2 dithioates) is on the right. Despite the appearance of a single species in CE, the PS2 oligos show extraneous peaks that cannot be deduced (e.g. depurination, unremoved protecting groups, etc…)

Fortunately the fragmentation behavior appeared concentration dependent, in that 2 μM, 4 μM, and 6 μM injections would yield different rates of fragmentation. Therefore, we attempted different concentrations of ESI-MS injections until a single major peak spectra was obtained. To pass QC, a molecule was accepted as pure when CE showed >95% purity and the cleanest ESI-MS spectra showed a single, major peak corresponding within 2 Da of the oligo's correct MW.

**Table 1.3) ESI-MS QC of the H800 Oligonucleotide Series (Ch. 2)**

| ID | Sequence Content 5'→3' | Mod | MW | ESI-MS | ΔAMU |
|---|---|---|---|---|---|
| H800 | /5TexRd-XN/CGAAACCAAGC | PO | 4200.2 | 4200.1 | -0.1 |
| H801 | /5TexRd-XN/CGAA-ps2-ACCA-ps2-AGC | A-ps2-A | 4264.5 | 4264.0 | -0.5 |
| H802 | GCTTGGTTTCG/3IAbRQSp/ | PO | 4032.8 | 4033.1 | +0.3 |
| H803 | GCT-ps2-TGGT-ps2-TTCG/3IAbRQSp/ | T-ps2-T | 4097.0 | 4096.3 | -0.7 |
| H804 | /5TexRd-XN/CGAACTCACGC | PO | 4167.2 | 4167.3 | +0.1 |
| H805 | /5TexRd-XN/CGAA-ps2-CTCA-ps2-CGC | A-ps2-C | 4231.4 | 4231.7 | +0.3 |
| H806 | GCGTGAGTTCG/3IAbRQSp/ | PO | 4066.8 | 4066.8 | +0.0 |
| H807 | GCG-ps2-TGAG-ps2-TTCG/3IAbRQSp/ | G-ps2-T | 4131.1 | 4130.7 | -0.4 |
| H808 | /5TexRd-XN/CGAAGTCAGGC | PO | 4247.3 | 4247.5 | +0.2 |
| H809 | /5TexRd-XN/CGAA-ps2-GTCA-ps2-GGC | A-ps2-G | 4311.5 | 4312.7 | +1.2 |
| H810 | GCCTGACTTCG/3IAbRQSp/ | PO | 3986.8 | 3986.8 | +0.0 |
| H811 | GCC-ps2-TGAC-ps2-TTCG/3IAbRQSp/ | C-ps2-T | 4051.0 | 4050.4 | -0.6 |
| H812 | /5TexRd-XN/CGAATCTATGG | PO | 4237.3 | 4237.5 | +0.2 |
| H813 | /5TexRd-XN/CGAA-ps2-TCTA-ps2-TGG | A-ps2-T | 4301.5 | 4300.8 | -0.7 |
| H814 | CCATAGATTCG/3IAbRQSp/ | PO | 3994.8 | 3994.8 | +0.0 |
| H815 | CCA-ps2-TAGA-ps2-TTCG/3IAbRQSp/ | A-ps2-T | 4059.0 | 4058.3 | -0.7 |
| H816 | /5TexRd-XN/CGACATCCAGC | PO | 4167.2 | 4167.6 | +0.4 |
| H817 | /5TexRd-XN/CGAC-ps2-ATCC-ps2-AGC | C-ps2-A | 4231.4 | 4230.3 | -1.1 |
| H818 | GCTGGATGTCG/3IAbRQSp/ | PO | 4066.8 | 4067.0 | +0.2 |
| H819 | GCT-ps2-GGAT-ps2-GTCG/3IAbRQSp/ | T-ps2-G | 4131.1 | 4130.3 | -0.8 |
| H820 | /5TexRd-XN/CACCGAGCGTC | PO | 4183.2 | 4183.5 | +0.3 |
| H821 | /5TexRd-XN/CACC-ps2-GAGC-ps2-GTC | C-ps2-G | 4247.4 | 4246.9 | -0.5 |
| H822 | GACGCTCGGTG/3IAbRQSp/ | PO | 4051.8 | 4051.8 | +0.0 |
| H823 | GAC-ps2-GCTC-ps2-GGTG/3IAbRQSp/ | C-ps2-G | 4116.0 | 4115.7 | -0.3 |
| H824 | /5TexRd-XN/AGACCTTCCAG | PO | 4182.2 | 4182.1 | -0.1 |
| H825 | /5TexRd-XN/AGAC-ps2-CTTC-ps2-CAG | C-ps2-C | 4246.5 | 4245.9 | -0.6 |
| H826 | CTGGAAGGTCT/3IAbRQSp/ | PO | 4050.8 | 4050.9 | +0.1 |
| H827 | CTG-ps2-GAAG-ps2-GTCT/3IAbRQSp/ | G-ps2-G | 4115.1 | 4115.2 | +0.1 |
| H828 | /5TexRd-XN/ACGGACTGAGC | PO | 4247.3 | 4247.4 | +0.1 |
| H829 | /5TexRd-XN/ACGG-ps2-ACTG-ps2-AGC | G-ps2-A | 4311.5 | 4311.0 | -0.5 |
| H830 | GCTCAGTCCGT/3IAbRQSp/ | PO | 3986.8 | 3986.9 | +0.1 |
| H831 | GCT-ps2-CAGT-ps2-CCGT/3IAbRQSp/ | T-ps2-C | 4051.0 | 4050.4 | -0.6 |
| H832 | /5TexRd-XN/TCTGCGAGCCG | PO | 4214.2 | 4214.0 | -0.2 |
| H833 | /5TexRd-XN/TCTG-ps2-CGAG-ps2-CCG | G-ps2-C | 4278.5 | 4278.7 | +0.2 |
| H834 | CGGCTCGCAGA/3IAbRQSp/ | PO | 4020.8 | 4021.1 | +0.3 |
| H835 | CGG-ps2-CTCG-ps2-CAGA/3IAbRQSp/ | G-ps2-C | 4085.0 | 4084.6 | -0.4 |
| H836 | /5TexRd-XN/CGCTAGATACG | PO | 4222.2 | 4222.3 | +0.1 |
| H837 | /5TexRd-XN/CGCT-ps2-AGAT-ps2-ACG | T-ps2-A | 4286.5 | 4286.1 | -0.4 |
| H838 | CGTATCTAGCG/3IAbRQSp/ | PO | 4010.8 | 4010.8 | +0.0 |
| H839 | CGT-ps2-ATCT-ps2-AGCG/3IAbRQSp/ | T-ps2-A | 4075.0 | 4074.8 | -0.2 |

## 1.6) Original Scientific Contributions

### IMPROVED GENERAL KNOWLEDGE OF DITHIOATE DNA PRODUCTION

- Optimized dithioate synthesis protocols for the common Expedite 8909 DNA synthesizer are presented (Table 1.1, Table 1.2).
- The large quantity of PS2-modified oligos synthesized (Table 1.3) and QC's revealed a peculiar fragmentation propensity for PS2 oligos in ESI-MS (Fig. 1.24).
- As with any chemically modified oligo, PS2-DNA is still rare, and the 24-bp double stranded 100% dithioated duplex synthesized in Chapter 1 appears to be the largest dithioated duplex to pass mass spec quality control (Fig. 1.22).

# CHAPTER 2: DITHIOATED PHOSPHATES IN DNA DUPLEX THERMODYNAMICS

## 2.1) Reviewing the Principles of Thermodynamics

The introduction to thermodynamics provided here is intended for first-year graduate students having no experience in the field. Major influence on the organization of this section comes from the texts of Levine[175] and Devoe[176], but any thermochemistry or physical chemistry text would make a suitable starting point for readers to engage these topics further.

"A theory is the more impressive the greater the simplicity of its premises, the more different kinds of things it relates, and the more extended its area of applicability. Therefore the deep impression that classical thermodynamics made upon me. It is the only physical theory of universal content which I am convinced will never be overthrown, within the framework of applicability of its basic concepts."

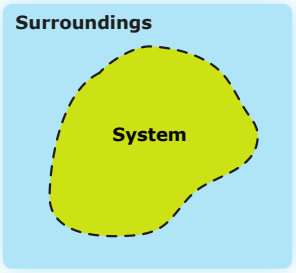
-Albert Einstein

### SYSTEMS, SURROUNDINGS, AND STATE CHANGES

Coming from the Greek terms for "heat" and "power", thermodynamics is a branch of physical chemistry that studies the relation of temperature and heat to energy and work produced when a system changes state. The term system has broad implications; any macroscopic region of the universe that is experimentally observable can be considered a system. It is an object in the universe, and one that can be defined by a boundary, which is the real or imaginary closed surface separating the volume occupied by the thermodynamic system from its surroundings. These surroundings are composed of everything else in the universe beyond the boundary that is capable of interacting with the system. As a microcosm within the surrounding universe, a system is described by physical properties called state quantities, which as functions of the state of the system are synonymously named state functions and state variables. These so-called equations of state are differentials capable of describing the system in terms of state quantities that are

dependent on the change of another, while the rest of the state variables are held constant. In other words, a "state" is defined by a specific set of state quantities, and when one quantity changes then the system has a different state. In biophysical thermodynamics the state quantities studied are mass, energy, work, entropy, volume, and particle number, which are all extensive thermodynamic properties. The field also investigates intensive properties of temperature and pressure. Determination of thermodynamic properties as extensive or intensive depends on whether their values change with the size of the system. For instance, cutting in half a system containing an ideal gas straight down the middle would reduce the extensive property of mass, yet the intensive property of pressure would stay the same. In addition to the fundamental biophysical thermodynamic state variables mentioned, some aspects of this research also use state quantities carried over from chemical thermodynamics, such as chemical potential and chemical activity, both intensive properties.

The central concept that is critical for understanding thermodynamics across all research subfields, including biophysical thermodynamics, is recognizing that the equations describe what state quantities must be transferred from the surroundings to the system, or vice versa, in order for the system to reach a second state from some initial starting state. To induce this state change; matter, heat, and work are transferred between the system and the surroundings. In this overview, because of applicability to my research, I'll cover in more detail the thermodynamics of closed systems. These are systems capable of exchanging energy in the form of heat, $Q$, or work, $W$, but not mass or matter. Alternative system types include isolated systems in which nothing is exchanged and open systems in which everything can be exchanged, as well as a few mechanical subtypes.




| System Type | Exchangeable? | |
|---|---|---|
| | Matter | Heat or Work |
| **Open** | yes | yes |
| **Closed** | no | yes |
| **Isolated** | no | no |

**Fig. 2.1) Types of Thermodynamic Systems**

**Left**: Diagram of system and surroundings. **Right**: Types of thermodynamic systems and the exchangeability of matter heat and work

## CLASSICAL THERMODYNAMICS IS A MACROSCOPIC SCIENCE

Classical thermodynamics is a macroscopic science and this scale is important because it means that the microscopic details of the system are not explicitly considered. Since biophysical thermodynamics is a solution analytical method, individual molecular motions are neglected and the system is instead interpreted as a homogenous, uniform distribution of molecules having bulk properties. The term homogenous has a specific thermodynamic definition, it means that every intensive thermodynamic property is constant throughout the entire system, and it also means that the system is composed of one uniform phase. Heterogeneous systems on the other hand, have multiple phases within the system, each having their own thermodynamic properties. But these types of systems will not be considered in this dissertation introduction. This dissertation focuses exclusively on equilibrium thermodynamics, which in an isolated system means that its macroscopic properties remain constant with time, or that time doesn't alter its phase space.

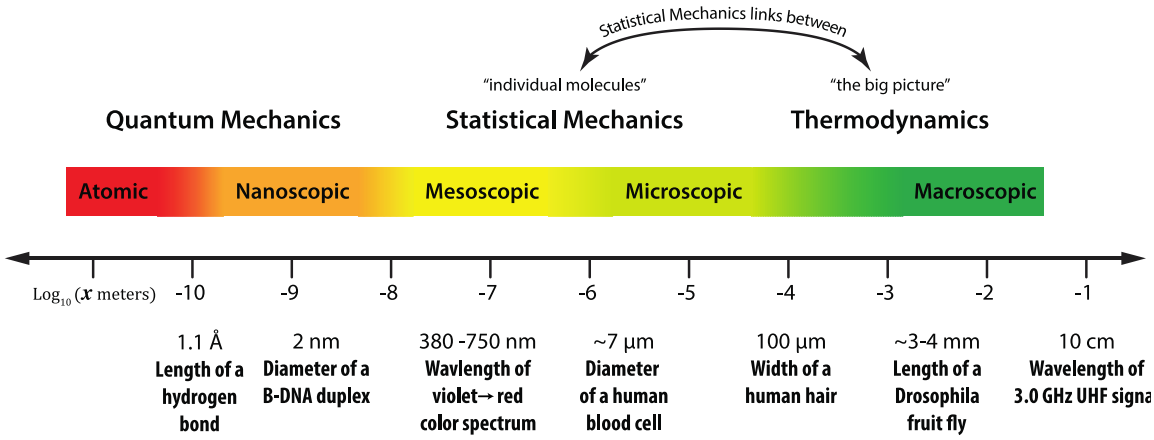



**Fig. 2.2) The Scale of Classical Thermodynamic Inquiry is Macroscopic**

In practice, statistical mechanics functions as the bridge between macroscopic observations from thermodynamics and molecular motions at the sub-microscopic level

Some of the laws in thermodynamics are applicable only for isolated systems, but empirical biophysics technically cannot involve perfectly isolated systems because experimental systems are inherently exchanging state quantities with the surroundings. In a pragmatic sense there is no such thing as an isolated system if it exists within a universe, since it will always exchange energy, material, or heat, in at least infinitesimal quantities with its surroundings. There is, for example, no such thing as a wall that remains perfectly adiabatic over an infinite time period. However in practice this disparity between theoretical and real system types is overcome by assuming that we can treat a closed system as an isolated system mathematically if its macroscopic properties (1) remain independent of time over an arbitrary time period appropriate for the experiment, and (2) remain unchanged upon removal from contact with the surroundings. In other words, if we can't detect a change in the system within a finite time period or if the surroundings are removed, then we accept the assumption that a closed system can be considered an isolated system mathematically. So it is our experimental limit of detection that allows us to apply thermodynamic laws to closed systems, even though the laws were intended for isolated systems. If a closed system satisfies the first but not second requirement, it is instead considered to be in a steady state and not equilibrium state.

The equilibrium concept is further subdivided into (1) mechanical equilibrium, in which no unbalanced forces act on the system and there is no

acceleration, and (2) material equilibrium, in which the concentration of chemical species remains constant with time and there are no net chemical reactions, and then (3) thermal equilibrium, in which there is no change in properties of the system if the separating boundary is a thermally conducting, nonadiabatic wall. For thermodynamic equilibrium, all three of these equilibrium subtypes must be true.

### WORK AND REVERSIBLE PROCESSES

As can be seen from the state quantities transferred during a state change, two very important concepts in thermodynamics are work and energy. Work in thermodynamics is developed in a similar way to classical mechanics. Meaning that the infinitesimal amount of work, d$w$, caused by a force in the $x$ direction is determined from the product of this force and displacement in the $x$ direction, d$x$, such that:

$$\mathrm{d}w \equiv F_x \mathrm{d}x \tag{2.1}$$

This shows that when $F_x$ and the displacement d$x$ are in the same direction, positive work is done on the system by the surroundings. When $F_x$ and d$x$ are in opposite directions, d$w$ is negative. This helps explain why the unit of measurement for thermodynamic work is the Joule ($1\ \mathrm{J} = 1\ \mathrm{N\ m} = 1\ \mathrm{kg\ m^2/s^2}$). There are several types of thermodynamic work, and they appear as conjugate pairs.

**Table 2.1) Two Conjugate Pairs Encompass Each Type of Thermodynamic Work**

| Type of Thermodynamic Work | Conjugate Pair |
|---|---|
| Mechanical | Pressure and Volume |
| Electromagnetic | Magnetic Flux Density and Magnetization |
| Chemical | Mole Fraction and Chemical Potential |

The foundations of thermodynamics were developed in the 19$^{\mathrm{th}}$ century, as society was embracing the industrial revolution, so it has its roots in the theory of heat engines. The concept "work" in classical thermodynamic focused on mechanical work performed where the surroundings change the system's volume. If the system contracts then d$w > 0$, and if the system expands then d$w < 0$, which means that the work done on the system is negative, or the system is performing

work on the surroundings. This type of expansion/contraction work is referred to informally as "P-V work", since volume displacement is determined from pressure $P$ and volume $V$ changes. The basis for the standard state equations in biophysical thermodynamics assumes that only P-V mechanical work is done on the system by surroundings, and other types of work such as electrical work are ignored. Of course, these are considered explicitly in other types of thermodynamics—but in these cases, then PV work is assumed to be zero. The formula for infinitesimal changes in the work done on a thermodynamic system can then be written as:

$$\mathrm{d}w_{rev} = -P\mathrm{d}V \tag{2.2}$$

which is valid for reversible processes. This is a very important term in thermodynamics, a "reversible process" is one where the system is always infinitesimally close to equilibrium, and reversing the process will restore both system and surroundings to their initial states. Upon returning to initial conditions their is no dissipation of energy, or production of entropy in the surroundings (entropy is discussed shortly). The negative term indicates the direction of work; that energy flows from the system to the surroundings, which demonstrates that traditionally energy changes are determined from the perspective of the system. The work performed on the system during the entire process would be the sum of the infinitesimal amounts of work along each step, and can be expressed according to French physicist Emile Clapeyron's 1934 formula as a definite integral:

$$w_{\mathrm{rev}} = -\int_0^1 P\mathrm{d}V, \tag{2.3}$$

where 0 is initial state and 1 is the final state. Work performed on the system by the surroundings is path dependent, producing different amount of work according to which specific process path was used to arrive at the second state, regardless of whether the two paths have identical initial and final P-V values. Take a look at the integral of P-V work shown below:

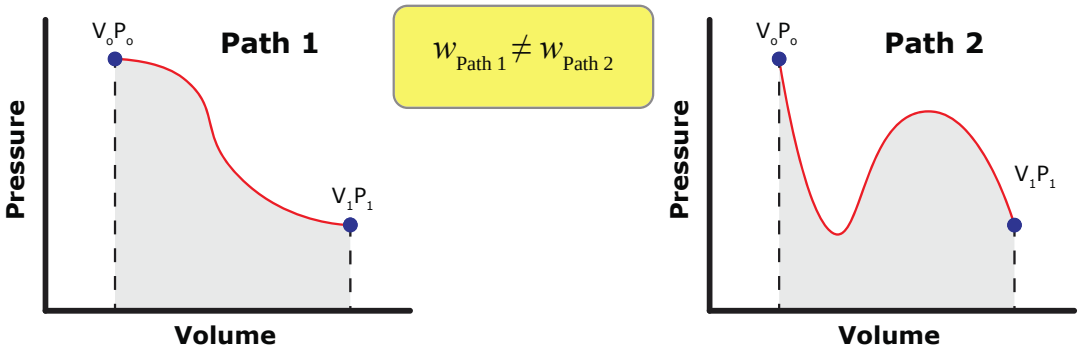



**Fig. 2.3) Thermodynamic Work is Path Dependent**

Despite identical initial and final pressure-volume values, the amount of work performed over the two paths is not identical because the computed integral is different. P-V work is therefore said to be path-dependent.

## THE ZEROETH LAW OF THERMODYNAMICS: TEMPERATURE, HEAT, AND HEAT CAPACITY

Virtually every physical property of a material object or substance is affected by temperature, but temperature is a rather abstract concept to grasp. Its development stems from the situation of equilibrium between two different systems capable of interaction. It was discussed previously that equilibrium entails three separate components: mechanical, chemical, and thermal equilibrium. When two systems performing P-V work are in mechanical equilibrium, they have a common pressure. If that were so, then what would be common between two systems if they are in thermal equilibrium? This is where the concept of temperature arises. By definition, two systems in thermal equilibrium must have the same temperature, whereas two systems not in thermal equilibrium have different temperatures. This definition provides the foundation for the Zeroeth Law of Thermodynamics, which establishes thermal equilibrium as an equivalence relationship. It asserts that if two systems are each found to be in thermal equilibrium with a third system, then they are also in thermal equilibrium with each other.

Experimentally, temperature cannot be measured directly. Instead, properties that are dependent on temperature in a uniform and consistent manner are used to infer temperature. A thermometer, for instance, relies on P-V expansion of thermal fluid such as mercury or ethanol, which expands in a predictable manner according to temperature. Other physical properties can also be used, such as

electrical resistance thermometers as well as optical pyrometers that measure the glow intensity of extremely hot metals.

It is known from observation that if two objects have different temperatures $T_1$ and $T_2$, where $T_2 > T_1$ then they will reach thermal equilibrium if they are placed in contact with each other. To arrive at their common intermediate temperature, we say that "heat", $q$, has flowed from the hotter object to the cooler object. The amount of heat is dependent on the masses of the objects, $m_1$ and $m_2$, and the specific composition of each object, which determines its specific heat capacity, $c_1$ and $c_2$ which are constants that are empirically determined for every material. The heat that flows is calculated as:

$$q \equiv m_2 c_2 (T_2 - T_1) = m_1 c_1 (T_1 - T_2) \tag{2.4}$$

The specific heat capacity $c_n$ of an object is a function of both temperature and pressure, but isobaric processes are most common in biophysical thermodynamics because pressure is often assumed to be constant under normal laboratory conditions. Thus specific heat capacity at constant pressure, $c_\mathrm{P}$, is used instead. A look at the equation shows an inverse identity, so the infinitesimal amounts of heat flowed into an objects is determined as:

$$\mathrm{d}q_\mathrm{P} = mc_\mathrm{P}\mathrm{d}T \tag{2.5}$$

The quantity $mc_\mathrm{P}$ is the heat capacity at constant temperature, which is denoted with a capital letter, as $C_\mathrm{P}$. It follows that:

$$\mathrm{d}q_\mathrm{P} = C_\mathrm{P}\mathrm{d}T \tag{2.6}$$

After rearrangement, the formal definition for heat capacity is apparent:

$$C_\mathrm{P} = \frac{\mathrm{d}q_P}{\mathrm{d}T} \tag{2.7}$$

It should be noted that in addition to heat flow through direct contact, radiative processes are also capable of transferring heat between two bodies that are not physically in contact. In this type of heat transfer, electromagnetic waves are emitted by one body and then absorbed by the second body. Electromagnetic radiation is capable of traveling through the medium separating the two bodies but can also travel through a vacuum, if appropriate.

## THE FIRST LAW OF THERMODYNAMICS: INTERNAL ENERGY

From classical mechanics, the motion of a body through space and the presence of fields on the body together compose the body's macroscopic kinetic energy $K$ and potential energy $V$, respectively. The sum of these, along with the internal energy $U$ of the body, comprise the body's total energy $E$:

$$E \equiv K + V + U \tag{2.8}$$

Internal energy is energy at the molecular level, but classical thermodynamics is macroscopic and does not explicitly consider individual molecules, instead approaching the sum of their energies from the system point of view (e.g. the whole pot). Conceptually however, it is easier to comprehend internal energy by appreciating that its components encompass many molecular energies. There are several contributors to the internal energy of a system: translational kinetic energy of its molecules, rotational, vibrational, and electronic energies; the relativistic energy $mc^2$ of the system molecules' electrons and the nuclei; as well as short-range intermolecular interaction forces. In classical thermodynamics, all of these energies are incorporated into the macroscopic system's internal energy $U$ term.

Determination of total energy $E$ in thermodynamic systems is then further simplified by assuming that at the macroscopic level the system is at rest, kinetically, and there is an absence of external fields acting on the body (e.g. $K = 0$ and $V = 0$). This assumption means that total energy is equivalent to the internal energy:

$$E = U \tag{2.9}$$

As with many thermodynamic properties, internal energy cannot be directly measured. Instead, changes of internal energy $\Delta U$ needed to acquire a different state of the system by the transition from some arbitrary standard state are measured. The result of a state transition is the transfer of energy into, or out of the system. The system internal energy is a thermodynamic potential that is determined as the sum of these energy transfers:

$$\Delta U = \sum_i E_i \tag{2.10}$$

where $\Delta U$ denotes the energy difference between the two states and $E_i$are the various energies transferred in the steps between the two states. A positive $\Delta U$ indicates that heat $q$ is added to the system or work $w$ is done on the system by the surroundings during the transition, whereas a negative $\Delta U$ denotes the converse and the system acts on the surroundings. For a closed thermodynamic system that can exchange only heat, $q$, and work, $w$, with its surroundings, a relationship of the change in internal energy to these parameters is obtained from the first law of thermodynamics expressed here in infinitesimal terms using differentials:

$$\mathrm{d}U = \delta q + \delta w \tag{2.11}$$

This is analogous to the physical law of conservation of energy, meaning that the amount of internal energy change will equal the amount of heat and work energy provided to or by the system. The internal energy is expressed using the exact differential symbol, d, because it is a state function. A state function quantitatively describes properties of an equilibrium state of a thermodynamic system, and is independent of the manner or path the system followed to arrive there. In contrast, heat and work are mechanical terms expressed using inexact differentials, $\delta$, because they are process quantities that are dependent on path taken between the two states, as mentioned in the previous figure. This symbolism of exact and inexact differentials used to express the relation of internal energy to heat and work was put forward by German mathematician Carl Gottfried Neumann in 1875[177] and the formalism remains in place today.

### THE SECOND LAW OF THERMODYNAMICS: ENTROPY

So far it's been shown that the zeroeth law of thermodynamics leads to the development of temperature, and from the first law of thermodynamics the concept of internal energy is created. In turn, the second law of thermodynamics provides the foundation for the development of entropy, $S$. Entropy was first described by Rudolf Clausius in the 1850s and 1860s as an explanation for lost work observed after the conservation of energy principles set forth by James Joule in the 1840s[178].

Clausius described entropy as the transformation-content of a thermodynamic system during a change of state. It was his method of quantifying energy loss in heat engines due to friction and dissipation. Clausius provided a mathematical description for the entropy during a state change:

$$\Delta S = \int \frac{Q_{\text{rev}}}{T} \tag{2.12}$$

which is the uniform thermodynamic temperature of a closed system dividing an incremental reversible transfer of heat in that system. Since this definition arose from macroscopic experimental observations, it is also referred to as experimental entropy, $\Delta S_{\text{E}}$.

Entropy is the foundation for the second law of thermodynamics, which states that in an isolated system, entropy never decreases because isolated systems always evolve towards thermodynamic equilibrium at which entropy is maximized. This can be written formulaically as:

$$\Delta S_{\text{sys}} + \Delta S_{\text{surr}} = \Delta S_{\text{univ}} \geq 0 \tag{2.13}$$

In this situation $\Delta S_{\text{univ}}$ isn't the entire universe, *per se*, but rather the isolated system that would contain the closed system and its surroundings capable of interacting with it, that is $\Delta S_{\text{univ}} = \Delta S_{\text{sys}} + \Delta S_{\text{surr}}$. This arrangement is analogous to a sphere, which represents the part of the universe capable of interacting with the system, whose interior shares an inner permeable boundary with the system it contains. The sphere is contained within a box, and the space between the sphere and the box is the portion of the universe incapable of interacting with the system. Whether this $\Delta S_{\text{univ}}$ box applies to the entire universe, in the cosmological sense, is currently being debated and is discussed briefly later. The second law of thermodynamics combined with the definition for entropy is together a very profound relation because it provides insight into the true reality of reversible vs. irreversible processes. Let me clarify this statement by recapitulating my previous statements on reversible processes. If a process induces a state change in which the surroundings act upon a system, then $\Delta S_{\text{sys}} = -\Delta S_{\text{surr}}$. If we take the definition for entropy applied for reversible processes as infinitesimal heat transfers:

$$\mathrm{d}S = \frac{\delta \mathrm{q}_{\mathrm{rev}}}{T} \tag{2.14}$$

we can combine this differential with the second law and to show that at a common temperature,

$$\mathrm{d}S_{\mathrm{univ}} = \mathrm{d}S_{\mathrm{sys}} + \mathrm{d}S_{\mathrm{surr}} = \frac{\delta \mathrm{q}_{\mathrm{rev}}}{T_{\mathrm{sys}}} + \frac{-\delta \mathrm{q}_{\mathrm{rev}}}{T_{\mathrm{surr}}} = \frac{\delta \mathrm{q}_{\mathrm{rev}}}{T_{\mathrm{sys}}} - \frac{\delta \mathrm{q}_{\mathrm{rev}}}{T_{\mathrm{sys}}} = 0 \tag{2.15}$$

So for a reversible process, $\mathrm{d}S_{\mathrm{univ}} = 0$. It can be shown through line integrals (not covered here, see any modern physical chemistry textbook) that for an irreversible process, $\mathrm{d}S_{\mathrm{univ}} > 0$. This comes full circle back to my brief statement in a previous section that for a reversible process, returning to the initial state of the system does not result in the production of entropy.

The second law of thermodynamics has a rich history; numerous statements of definition have been put forward over the past 200 years. It's origins are traced to the French engineer Sadi Carnot (the son of Lazare Carnot) in his theorization of ideal heat engines, and was later revised by German physicist Rudolf Clausius, and then further by William Thomson (Lord Kelvin), Max Planck, and Willard Gibbs. The second law and the concept entropy might be one of the most important equations in physics because it provides an explanation for the one-way direction of "time's arrow", a term coined by British astrophysicist Arthur Eddington[179]. Time progresses forward, and so does entropy; they never go backwards. After the recent discovery that the universe is increasing its expansion rate[180-182] following the cosmological "big bang" model, rather than slowing into a future "big crunch" from gravitational attraction, it is remarkable that the cosmological one-way arrow agrees with the macroscopic thermodynamic one-way arrow. The universe is not currently in an equilibrium state and entropy will increase until it reaches its maximum. It has even been argued that the thermodynamic arrow of time and the second law underlies the cosmological arrow, and logically provides the foundation for the expansion of the universe and Albert Einstein's cosmological constant $\Omega_{\Lambda}$[181-183].

## CLASSICAL THERMODYNAMICS: INTERNAL ENERGY

Using Clausius' definition for experimental entropy (Eq. 2.14), we can rearrange into $\delta q = T\mathrm{d}S$ and substitute back into the first law of thermodynamics (Eq. 2.11), arriving at:

$$\mathrm{d}U = T\mathrm{d}S + \delta w \tag{2.16}$$

Assuming that only pressure-volume work is performed and using the system as the point of reference, this equation becomes:

$$\mathrm{d}U = T\mathrm{d}S - P\mathrm{d}V \tag{2.17}$$

This shows that internal energy is comprised of two components, one that changes with respect to $S$ alone and one that is dependent on $V$ alone. This is written formally as:

$$\mathrm{d}U(S,V) = T\mathrm{d}S - P\mathrm{d}V \tag{2.18}$$

This is the fundamental thermodynamic relation and represents an important concept in thermodynamics: that potentials change with respect to one thermal natural variable (temperature or entropy), and one mechanical natural variable (pressure or volume). The change in energy of a system can always be written as the product of an intensive and an extensive parameter. This takes the form of:

$$\Delta U = \sum_i X_i \Delta Y_i \tag{2.19}$$

where $X_i$ is an intensive quantity, such as pressure or temperature, and $Y_i$ is an extensive quantity, such as volume. Shown first here with the case of internal energy, entropy and volume are used. From partial differentiation it follows:

$$\mathrm{d}U = \left(\frac{\partial U}{\partial S}\right)_{\mathrm{V}} \mathrm{d}S + \left(\frac{\partial U}{\partial V}\right)_S \mathrm{d}V \tag{2.20}$$

By comparison with $\mathrm{d}U(S,V) = T\mathrm{d}S - P\mathrm{d}V$ we see that:

$$\left(\frac{\partial U}{\partial S}\right)_V = T \tag{2.21}$$

$$\left(\frac{\partial U}{\partial V}\right)_S = -P \tag{2.22}$$

This partial derivative formulation of state functions, $T$ and $-P$, is an important tool in thermodynamic analysis because comparison to similarly treated thermodynamic potential state functions allows their linkage through common variables. This is the mechanism from which Maxwell relations are derived, named after physicist James Clerk Maxwell who first showed the equality of the second derivatives of the four thermodynamic potentials in the nineteenth century.

### CLASSICAL THERMODYNAMICS: ENTHALPY

The next thermodynamic potential to discuss is enthalpy, $H$. This is defined mathematically as:

$$H = U + PV \tag{2.23}$$

which is the internal energy plus pressure-volume expansion. This can be developed by going back to the total differential relation $\mathrm{d}U = T\mathrm{d}S - P\mathrm{d}V$ (Eq. 2.17) and adding $\mathrm{d}(PV)$ expansion to both sides of the expression, arriving at:

$$\mathrm{d}U + \mathrm{d}(PV) = T\mathrm{d}S - P\mathrm{d}V + \mathrm{d}(PV) \tag{2.24}$$

Which reduces to:

$$\mathrm{d}(U + PV) = T\mathrm{d}S + V\mathrm{d}P \tag{2.25}$$

Similar to how the internal energy was interpreted above, we see that the enthalpy change between states is dependent on the $S$, $P$, and $T$ of the system, which are all state function, so this makes $\mathrm{d}H$ a state functions as well. That is:

$$\mathrm{d}H(S,P) = T\mathrm{d}S + V\mathrm{d}P \tag{2.26}$$

and through partial differentiation yields:

$$\mathrm{d}H = \left(\frac{\partial H}{\partial S}\right)_P \mathrm{d}S + \left(\frac{\partial H}{\partial P}\right)_S \mathrm{d}P \tag{2.27}$$

Again, we see alternative thermodynamic definitions for the intensive property of temperature and the extensive property of volume, $\left(\frac{\partial H}{\partial S}\right)_P = T$ and $\left(\frac{\partial H}{\partial P}\right)_S = V$.

## THE EQUIVALENCY OF ENTHALPY AND OBSERVABLE HEAT

At this point it's necessary to highlight an important aspect of enthalpy in regards to experimental thermodynamics, which is the relation of enthalpy to observable heat. From the definition of enthalpy, $H = U + PV$ (Eq. 2.23), the total differential of the transition enthalpy can be expanded by the product rule into:

$$\mathrm{d}H = \mathrm{d}U + P\mathrm{d}V + V\mathrm{d}P \tag{2.28}$$

Substituting the internal energy state function, $\mathrm{d}U = T\mathrm{d}S - P\mathrm{d}V$ (Eq. 2.17), it follows:

$$\mathrm{d}H = T\mathrm{d}S - P\mathrm{d}V + P\mathrm{d}V + V\mathrm{d}P \tag{2.29}$$

The second law of thermodynamics, $\delta q = T\mathrm{d}S$, can be substituted, producing after simplification:

$$\mathrm{d}H = \delta q + V\mathrm{d}P \tag{2.30}$$

This expression reveals the equivalency of enthalpy and heat, $\delta q$, when the pressure is kept constant in a closed system, $\mathrm{d}P = 0$, that is:

$$\mathrm{d}H = \delta q + \cancel{V\mathrm{d}P} \tag{2.31}$$

$$\mathrm{d}H = \delta q \tag{2.32}$$

The equivalency of the state function enthalpy with the mechanical process quantity heat, is a critical concept in thermodynamics because heat transfer can be experimentally observed. Most experiments are run under isobaric conditions ($\mathrm{d}P = 0$), therefore enthalpy often serves as the bridge linking empirical heat with equations of state.

## CLASSICAL THERMODYNAMICS: GIBBS FREE ENERGY AND HELMHOLTZ FREE ENERGY

So far it was shown that the thermodynamic potential *internal energy* can be derived directly from the first and second law of thermodynamics, and its derived differential reveals a dependency on the natural variables *entropy and volume*. Then an expression for the thermodynamic potential *enthalpy* is developed from internal energy by adding pressure-volume expansion terms, and its derived differential shows a dependency on *entropy and pressure*. Next we will consider

two other thermodynamic potentials that are dependent on a different thermal variable, temperature instead of entropy. This represents a movement towards describing useful work obtainable from a closed thermodynamic system.

American engineer Willard Gibbs put forward the concept of "available energy", or "free energy" in 1873 to describe maximum amount of work that can be extracted from a closed thermodynamic system. This Gibbs energy, $G$, is defined mathematically as the internal energy of the system plus pressure-volume forces minus thermal entropy:

$$G = U + PV - TS \tag{2.33}$$

Since enthalpy can be defined as $H = U + PV$ (Eq. 2.23), it can substituted in, and then after rewriting in total differential form the Gibbs energy becomes:

$$\mathrm{d}G = \mathrm{d}H - T\mathrm{d}S - S\mathrm{d}T. \tag{2.34}$$

It was earlier shown that $\mathrm{d}H = \mathrm{d}U + P\mathrm{d}V + V\mathrm{d}P$ (Eq. 2.28), so this can be substituted in as well:

$$\mathrm{d}G = \mathrm{d}U + P\mathrm{d}V + V\mathrm{d}P - T\mathrm{d}S - S\mathrm{d}T \tag{2.35}$$

Earlier it was shown that $\mathrm{d}U = T\mathrm{d}S - P\mathrm{d}V$ (Eq. 2.18), and we can combine again and reduce:

$$\mathrm{d}G = T\mathrm{d}S - P\mathrm{d}V + P\mathrm{d}V + V\mathrm{d}P - T\mathrm{d}S - S\mathrm{d}T \tag{2.36}$$

$$\mathrm{d}G = V\mathrm{d}P - S\mathrm{d}T \tag{2.37}$$

$$\mathrm{d}G(P,T) = V\mathrm{d}P - S\mathrm{d}T \tag{2.38}$$

$$\mathrm{d}G = \left(\frac{\partial G}{\partial P}\right)_T \mathrm{d}P + \left(\frac{\partial G}{\partial T}\right)_P \mathrm{d}T \tag{2.39}$$

We see that $\left(\frac{\partial G}{\partial P}\right)_T = V$ and $\left(\frac{\partial G}{\partial T}\right)_P = -S$. This second value, this particular definition of entropy, is quite important because it serves as the mechanism for the Gibbs-Helmholtz equation that focuses on the variation of Gibbs energy with temperature.

This $G$ is termed "free energy", and is a thermodynamic potential that is equal to the internal energy of the system, or some form of it, minus the amount of energy that cannot be used to perform work. This is where the term "free" comes from; it is energy that is available for work. The energy that cannot be used for thermodynamic work is attributed to entropic loss. This entropy is a state function

that measures the progression towards thermodynamic equilibrium, which is defined as the state with maximum entropy. Once a closed system is in thermodynamic equilibrium, there is no net flow of energy, and thus no more work can be performed, and entropy is maximized.

Another definition for free energy is Helmholtz free energy $F$, named after the German physicist Hermann von Helmholtz who first described it in late 1800s. Helmholtz free energy can be developed mathematically from internal energy through the product rule, (*e.g.* $\mathrm{d}(TS) = T\mathrm{d}S + S\mathrm{d}T$).

$$\mathrm{d}U = T\mathrm{d}S - P\mathrm{d}V \tag{2.40}$$

$$\mathrm{d}U = \mathrm{d}(TS) - S\mathrm{d}T - P\mathrm{d}V \tag{2.41}$$

$$\mathrm{d}(U - TS) = -S\mathrm{d}T - P\mathrm{d}V \tag{2.42}$$

$$\mathrm{dF}(T, \mathrm{V}) = -S\mathrm{d}T - P\mathrm{d}V \tag{2.43}$$

$$\mathrm{d}F = \left(\frac{\partial F}{\partial T}\right)_V \mathrm{d}T + \left(\frac{\partial F}{\partial V}\right)_T \mathrm{d}V. \tag{2.44}$$

The definition of Helmholtz energy, $F$, becomes clear (Eq. 2.42 left): it is the internal energy of the thermodynamic system minus unusable entropic energy:

$$F = U - TS \tag{2.45}$$

### THE VAN'T HOFF EQUATION

The van't Hoff Equation is named after Dutch chemist Jacobus Henricus van't Hoff who in the mid 1880s proposed a relation between the change in the equilibrium constant of a chemical equilibrium to the change in temperature given the standard enthalpy change for the process.

Before reaching the end goal, first a critical thermodynamic relation must be discussed:

$$\Delta G = -RT \ln K_{\mathrm{eq}} \tag{2.46}$$

This relation is attributed supposedly to van't Hoff's work, however since he wrote mostly about chemical affinities it's likely this form of the relation was never actually penned by him. Nevertheless, he pioneered the development of expressions for free energy in terms of equilibrium constants and receives recognition. An

expression from this time period that is most similar the now ubiquitous Eq. 2.46 was actually developed by German physical chemist Walther Nernst using Helmholtz free energy. Inspired by van't Hoff's work, Nernst provided a robust derivation of a relation in 1893[184], showing that showed that:

$$\Delta \mathrm{F} = RT \ln K_{\mathrm{eq}} \tag{2.47}$$

Where $K_{\mathrm{eq}}$ is the equilibrium constant as defined by the law of mass action. Because of the assumed identity:

$$\mathrm{F} = -\Delta G \tag{2.48}$$

the Gibbs energy can be expressed as:

$$\Delta \mathrm{G} = -RT \ln K_{\mathrm{eq}} \tag{2.46}$$

This relation is one of the principle thermodynamic relations. Combining it with the Gibbs definition for free energy ($\Delta G = \Delta H - T\Delta S$) we have:

$$\Delta H - T\Delta S = -RT \ln K_{\mathrm{eq}} \tag{2.49}$$

$$\ln K_{\mathrm{eq}} = \frac{\Delta H}{-RT} - \frac{T\Delta S}{-RT} \tag{2.50}$$

$$-\ln K_{\mathrm{eq}} = \frac{\Delta H}{RT} - \frac{\Delta S}{R} \tag{2.51}$$

This last equation is the van't Hoff equation and is one of the principle tools used by thermodynamicists to determine the enthalpy and entropy of a chemical reaction. The method for this involves measuring the equilibrium constant of the reaction as a function of the change in temperature.

### Experimental Approaches for Measuring Thermodynamics

The thermodynamics of DNA hybridization can be investigated using either direct or indirect experimental designs. Both have different advantages and limitations but most noteworthy is that the direct approach measures actual heat using thermocalorimeters and is model independent, while the indirect approach uses state equation relations to equilibrium measurements and is model dependent. There are obvious benefits to directly measuring thermodynamic values for biological reactions, but high sample requirements and sparse instrument availability has encumbered progress compared to the wealth of DNA

thermodynamic information published using indirect methodologies, which employs widely available spectrophotometers. Nevertheless, a significant amount of literature has been written comparing thermodynamic values determined from the two techniques[185].

## MOLECULAR INTERPRETATION OF THERMODYNAMICS: PHASE SPACE AND CELL SIZE

So far the focus has been on macroscopic thermodynamics, which is also considered as "classical" thermodynamics because of the time period in which its concepts were first formulated and experiments first performed. Properties such as $T$, $P$, $V$, $U$, *et cetera* could be probed directly or indirectly at the time, and it was pragmatic to define thermodynamic systems using such parameters. However, with the discovery of the atom in 1890s-1920s it was apparent that macroscopic thermodynamics were indeed sensitive to mechanics at the atomic scale. There was great interest in developing a molecular explanation for thermodynamics, but ordinary mechanics cannot practically approach this issue. These methods best treat the behavior of one, two, or at most a few bodies; intractable for scaling to the total number of system molecules, and information on a single motion is not very useful considering the large collection of individual motions contained within a system. Interest instead shines on the *average* individual motions, for it is this average that produces the system's macroscopically observed thermodynamic properties.

The field of statistical thermodynamics was developed to address molecular interpretations of classical thermodynamic state functions using statistical mechanics approaches. Initial development of this field is credited to German physicist Ludwig Boltzmann who applied phase space theory and probability in his statistical description of ideal gases in the 1870's. The thought behind this movement is that the macroscopic state of the system can be defined by a distribution of microstates accessible through normal thermal fluctuations. During these fluctuations, within a system energy is constantly being redistributed as particles (molecules, atoms) continuously collide, relocate, and alter individual quanta. The specific spatial arrangement of each molecule and its quanta in one

instant of time is a microstate; it can thought of as a snapshot or instantaneous photograph of every vector particle in a system. More formally, in dynamics the state of a single particle is specified by its momentum $\vec{p}_i$ and position $\vec{x}_i$. Once the initial momentum and position of a single particle is known, then by classical mechanics the complete motion of the particle can be specified. For a molecule in our thermodynamic system the combined position and momentum is referred to as its phase space. It has in turn, 6 degrees of freedom, which are its three internal momentum degrees and three translational position degrees of freedom. The element of its phase space is:

$$\mathrm{d}\Gamma = \mathrm{d}p_x\, \mathrm{d}p_y\, \mathrm{d}p_z \;\cdot\; \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z \tag{2.52}$$

$$\mathrm{d}\Gamma = \prod_{i=1}^{N_S} d\vec{p}_i\, d\vec{x}_i \tag{2.53}$$

The state of the system can therefore be represented as a point in $6N_S$ dimensional $\Gamma$ phase space, and the behavior of particles in the system is described by the motion of this point through this phase space. The small cube $\Delta\Gamma = \prod_i \Delta\vec{p}_i\, \Delta\vec{x}_i$ represents an element of the $6N_S$ $\Gamma$ phase space that is called a cell. The cells within the phase space represent microstates of the system. Cell size is the resolution of phase space boundaries between microstates. The upper cell size limit is chosen so that properties of the system do not change substantially if the point transits from one end of the cell to the other. If this occurs, the cell size is reduced until each cell represents its own precise description of the system state. Exceedingly small cell sizes, however, approach the Heisenberg's uncertainty limit from quantum mechanics:

$$\Delta\vec{p}_i\, \Delta\vec{x}_i \geq \frac{\hbar}{2} \tag{2.54}$$

This means that $\vec{p}_i$ and $\vec{x}_i$ can only be specified within:

$$\Delta\vec{p}_i\, \Delta\vec{x}_i \geq \hbar^3 \tag{2.55}$$

and a six dimensional cube of sides $\hbar^{\frac{1}{2}}$ represents the lower limit for cell size. In statistical mechanics the cell size utilized is:

$$\hbar^{3N_S} \tag{2.56}$$

The number of states in volume dΓ is:

$$\frac{d\Gamma}{\hbar^{3N_S}} \tag{2.57}$$

Historically, Boltzmann first utilized a cell size of $\hbar^3$ but this was heavily criticized for being too small. The chance that a cell of this phase space size captures the particle behavior of molecules is very low, and this produces statistical fluctuations in the cell occupation levels (although you can work around this by grouping individual cells together). The cell size used now in statistical thermodynamics is larger and scaled to the system $\hbar^{3N_S}$, but its dimensions should be accepted as arbitrary.

Now that we have decided an arbitrary cell size, we can ask ourselves what we expect the occupancy level of our cell to be. That is, how often is our phase space occupied? And in turn, what is the probability of the microstate represented by our phase space cell? The basic assumption of statistical mechanics is the "equal *a priori* probability of all regions of phase space". That is, each region or cell of phase space is intrinsically equally likely to be occupied and we have no reason to choose or favor one region over another. This means that each energy state of the system is equally likely to be occupied. Only afterwards, when external conditions are used to weight our statistical arguments do certain regions of phase space become more heavily occupied than others. But initially, these equal *a priori* probability conditions are imposed.

## MOLECULAR INTERPRETATION OF THERMODYNAMICS: COMBINATORIAL DISTRIBUTION

Another important foundation in the field put forward by Boltzmann was his combinatorial method of numerating gas particles in a system.

| | |
|---|---|
| State # | $1, 2, 3 \ldots s \ldots r$ |
| Energy | $\epsilon_1, \epsilon_2, \epsilon_3 \ldots \epsilon_s \ldots \epsilon_r$ |
| Number | $n_1, n_2, n_3 \ldots n_s \ldots n_r$ |

This allows the internal energy of the system and the number of microstates to be restricted to:

$$U = \sum_{s=1}^{r} n_s \, \epsilon_s \tag{2.58}$$

$$N = \sum_{s=1}^{r} n_s \tag{2.59}$$

Because of the assumption of equal *a priori* probability of phase space we know that each cell or state is equally likely to be occupied. Boltzmann's work with gases showed that the probability of seeing a particular distribution of cells is proportional to the number of ways the distribution can be formed. Meaning that the number of permutations of the particles among the states which leads to the same distribution is proportional to the probability of observing this distribution. Permutability is a measure of the probability. The number of ways the distribution enumerated in the table can be calculated by:

$$W = {}^{N!}/_{\prod_i n_i!} = \frac{N!}{n_1! \, n_2! \, n_3! \ldots n_s! \ldots n_r!} \tag{2.60}$$

Statistical mechanics instead seeks to determine the distribution over the possible momentum and position values, and with this distribution in statistical thermodynamics we can evaluate average properties such as energy, pressure and specific heat

## BOLTZMANN'S CONSTANT: THE BRIDGE BETWEEN MACROSCOPIC AND MICROSCOPIC PHYSICS

The Boltzmann constant is a physical constant that bridges between macroscopic and microscopic physics, functioning by relating energy at the individual particle level with temperature. It is defined as:

$$k_{\mathrm{B}} = \frac{R}{N_{\mathrm{A}}} \tag{2.61}$$

where $R$ is the appropriately selected gas constant and $N_{\mathrm{A}}$ is the Avogadro constant ($6.022E23$ mol$^{-1}$). The gas constant's development is complex, culminating from four gas laws presented over two centuries by Robert Boyle ~1662 ($PV = k$), Jacques Charles ~1787-1802 ($V/T = k$), Joseph Louis Gay-Lussac ~1809 ($P/T = k$), and Amedeo Avogadro ~1811 ($V/n = a$). A user would select the gas constant having unit dimensions appropriate for the field of study when the Boltzmann

constant is applied in physics. The work by Italian scientist Avogadro in 1811[186] discussing the proportionality of gas volume and the number of atoms was especially important for the development of the Boltzmann constant for thermodynamics. Substituting in unit dimensions it is shown that the Boltzmann constant and entropy $S \equiv \frac{q}{T}$ have the same dimension, being units of Joules per Kelvin, J/K:

$$k_{\mathrm{B}}(\text{units}) = \frac{R}{N_{\mathrm{A}}} = \frac{{}^{\mathrm{J}}/_{\mathrm{mol\cdot K}}}{\mathrm{mol}^{-1}} = \frac{\mathrm{J}}{\mathrm{K}}.$$

To see how the Boltzmann constant bridges between macroscopic and microscopic physics we start with the ideal gas law:

$$PV = n\mathrm{R}T \tag{2.62}$$

where $n$ is number of moles of gas and $P$, $V$, $R$, and $T$ are defined previously. After substituting in the Boltzmann constant, the ideal gas law then simplifies, as first introduced in 1900 by German physicist Max Planck, into:

$$PV = Nk_{\mathrm{B}}T \tag{2.63}$$

where $N$ is the number of molecules of gas. By looking at the equation the function of the Boltzmann constant as the bridge becomes apparent; the left hand side is the macroscopic thermodynamic properties of pressure and volume and it is equated to the right hand side which describes particles in the system at a given temperature and the effect the Boltzmann physical constant.

### THERMODYNAMIC ENSEMBLES: MICROCANONICAL, CANONICAL, AND GRAND CANONICAL

In 1902 Gibbs first put forward the concept of an ensemble in a book entitled *Elementary Principles in Statistical Mechanics*[187]. Thermodynamic ensembles are statistical ensembles representing the probability distribution of microscopic states of the system. Ensembles were developed as a mathematical idealization of an infinitely large number of imaginary copies of a system each having their own constraints (constant parameters). In the context of thermodynamics, the copies of the system are constrained by macroscopic properties such as fixed volume, energy, number of particles, and chemical potential. Which particular properties are fixed

determines the type of ensemble that is applicable for statistical mechanic interpretation of thermodynamics. Any number of microstates can satisfy the fixed parameters, and the entire collection of these satisfying microstates is the "ensemble". There are three main types of statistical ensembles in thermodynamics: microcanonical, canonical, and grand canonical. These ensembles are appropriate for describing isolated, closed, and open systems, respectively. However, since they are all special cases of the general maximum entropy principle, state function quantities determined from all three approaches will be identical; the method chosen is a matter of preference based on the aim of the calculation and the properties of the system.

**Table 2.2) Three of the Primary Ensembles in Statistical Thermodynamics**

| *Ensemble Name* | *Nickname* | *What is constrained?* | *Applicable System* | *Partition Function* |
|---|---|---|---|---|
| **Microcanonical** | "NVE" ensemble | • Number of particles<br>• Volume<br>• Energy | Isolated | $\mathrm{Q}(N, V, E)$ |
| **Canonical** | "NVT" ensemble | • Number of particles<br>• Volume<br>• Temperature | Closed | $Q(N, V, T)$ |
| **Grand Canonical** | "μVT" ensemble | • Chemical potential<br>• Volume<br>• Temperature | Open | $Q(\mu, V, T)$ |

Ensembles in statistical thermodynamics can be interpreted as being regions of phase space with dimensions of macroscopic thermodynamic properties, in which specific dimensions are constrained. For instance, in the canonical ensemble the number of particles in the system, and the system's volume and temperature are constrained. The three most commonly utilized thermodynamic ensembles are shown here.

## MICROSTATE PROBABILITY: BOLTZMANN'S FACTOR AND PARTITION FUNCTIONS

I'll start with Boltzmann distribution, which is a term in statistical mechanics that expresses the probability distribution of a state with energy $E$ relative to the probability of a state containing a theoretical energy of zero. It has the form:

$$X(\text{state}) \propto e^{-\frac{E}{k_{\mathrm{B}}T}} \tag{2.64}$$

Since $\frac{1}{k_B T}$ appears so frequently in thermodynamics, it is given the shorthand notation of "thermodynamic beta", $\beta$. It can be rewritten as:

$$X(\text{state}) \propto e^{-\beta E} \tag{2.65}$$

The origin of the exponential may not be intuitively apparent upon first impression, but since the exponential formulation is so ubiquitous throughout statistical thermodynamics it is worthwhile to explain its implementation. If one imagines that there are three states of energy $j$, $k$, and $l$, then the ratio of the number of states ($a$) with energy $j$, $k$, or $l$ ($a_j, a_k, a_l$) can be written as some arbitrary function of the states, that is:

$$\frac{a_j}{a_k} = f\left(E_j - E_k\right), \quad \frac{a_j}{a_l} = f\left(E_j - E_l\right), \quad \text{and} \quad \frac{a_k}{a_l} = f(E_k - E_l) \tag{2.66}$$

Then:

$$\frac{a_j}{a_k} = \frac{a_j}{a_l} \bullet \frac{a_k}{a_l} = f\left(E_j - E_l\right) \bullet f(E_k - E_l) \tag{2.67}$$

It follows that:

$$f\left(E_j - E_k\right) = f\left(E_j - E_l\right) \bullet f(E_k - E_l) \tag{2.68}$$

To make the equation simpler to work, one identifies $\left(E_j - E_l\right) = x$ and $(E_k - E_l) = y$. It is apparent then that the formula can be rewritten as:

$$f(x + y) = f(x) \bullet f(y) \tag{2.69}$$

This is where the connection with exponentials is made. As a property of exponential functions it is true that the product of two exponentials is equal to exponential of the sum, $e^{x+y} = e^x \bullet e^y$. So given this identity one can rewrite this ratio using the exponential function. The general form of the equation for the number of states having energy $i$ is:

$$a_i = C e^{-\beta E_i} \tag{2.70}$$

where $C$ is an arbitrary constant stemming from the proportionality from probability as mentioned above (it soon cancels out). This quantity in exponential form is referred to as the Boltzmann factor.

Now since we are discussing probabilities it is then necessary to normalize the Boltzmann factor to a value of one by dividing over by all possible microstates $N$, where $N = \sum_n a_n$, so that the Boltzmann factor becomes a Boltzmann probability.

The normalized Boltzmann probability of the number of microstates having the energy $i$ would be:

$$p_i = \frac{a_i}{N} \tag{2.71}$$

We can combine this with equation (2.70) and simplify into:

$$p_i = \frac{a_i}{N} = \frac{Ce^{-\beta E_i}}{C\sum_n e^{-\beta E_n}} = \frac{e^{-\beta E_i}}{\sum_n e^{-\beta E_n}} \tag{2.72}$$

The quantity in the final denominator appears so frequently in thermodynamics that it is given a special name, the partition function $Q$. The partition function is simply what is obtained when all of the Boltzmann factors for a system are added up. The partition function acts as the normalization factor for the probability distribution of microstates. It is given a special notation according to what ensemble the system corresponds to (what parameters are fixed) and in this example has the form:

$$Q(N,V,\beta) = \sum_n e^{-\beta E_{\mathrm{n}}(N,V)} \tag{2.73}$$

However, $\beta$ isn't directly a thermodynamic parameter, but since thermodynamic beta is:

$$\beta = \frac{1}{k_B T} \tag{2.74}$$

we can rewrite the partition function as:

$$Q(N,V,T) = \sum_n e^{-\frac{E_n(N,V)}{k_B T}} \tag{2.75}$$

Now since $Q$ is written in terms of constant $N$, $V$, and $T$, when referenced to the thermodynamic ensembles described in Table 2.2 we see that we have formulated the partition function corresponding to the "canonical ensemble".

## STATISTICAL MECHANICS DEFINITION OF ENTROPY: BOLTZMANN ENTROPY AND GIBBS ENTROPY

Boltzmann's statistical mechanics definition for entropy is a probability equation relating entropy to the quantity $W$ discussed earlier, which is an enumeration of microstates corresponding to a specific macrostate. Remember that:

$$W = {}^{N!}/_{\prod_i n_i!} = \frac{N!}{n_1!\, n_2!\, n_3! \ldots n_s! \ldots n_r!} \tag{2.60}$$

Boltzmann defines entropy as:

$$S = k_{\mathrm{B}} \ln W \tag{2.76}$$

where $k_{\mathrm{B}}$ is the Boltzmann constant and $W$ is the enumeration of microstates. This is one of the most important equations in all of physics. This definition for entropy developed by Boltzmann is applicable for microstates of the universe as a whole, where the equal *a priori* probability assumption means that each microstate is equally probably. In thermodynamics however, the entire universe is not considered but instead just the particular system of interest and the regions of the universe capable of interacting with it. In such instances, microstates in thermodynamic systems are not equally probable. For example, if the temperature is kept constant then high-energy states are less probable than low-energy states. Willard Gibbs took Boltzmann's equation and generalized it for thermodynamic systems[188]:

$$S = -k_{\mathrm{B}} \sum p_i \ln p_i \tag{2.77}$$

This is referred to as the Gibbs statistical mechanical definition of entropy, and is differentiated from Boltzmann entropy because it considers the ensemble of states of the system as a whole, rather than equally probably single particle states.

## 2.2) Introduction to Nucleic Acid Thermodynamics

Oligonucleotides have a rich and diverse chemistry with many moieties capable of forming bonds that stabilize complex secondary structures. These range from unimolecular interactions such as hairpins and stem-loops, to bimolecular interactions such as A, B, and Z-form DNA duplexes, and to even more complex triple and higher-order structures. This dissertation focuses on the hybridization of single strand oligos into the common B-DNA duplex, and in particular DNA strands containing dithiophosphate modified backbones.

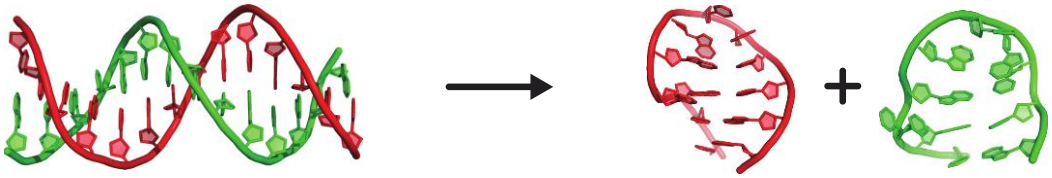



**Fig. 2.4) Helix-to-Coil Transition of a DNA Duplex**

Graphical image demonstrating the transition of a double-stranded DNA duplex into two independent single strands.

The interplay of energetic forces governing the association and dissociation processes of the double helix is of great interest to the scientific community because all active biological functions of nucleic acids, such as replication and transcription, require the formation and breakage of base pairs in duplex DNA. An interesting aspect of the hybridization phenomenon is the numerous energetic barriers strands must overcome in order to associate. For example, the many internucleotide phosphate linkages impart significant polyanionic charge density to single strand oligos that creates a repulsive electrostatic term that increases sharply as the negatively charged strands approach each other, which inhibits association. However, in solution this electrostatic repulsion is ameliorated by the presence of cations. These positively charged counter-ions screen the repulsive phosphate anionic charges to enable strands to associate. In the cell many mechanisms such as ion pumps and gradient channels perform this task, sequestering specific ions to regions where nucleic acids are localized.

Although the task of measuring intracellular ion concentrations is difficult, developments in the fields of electron microprobe analysis and ion-specific fluorescence dyes have improved our understanding of cation distribution within cells. One such study using microprobe analysis investigated $Na^+$, $K^+$, and $Cl^-$ concentration differences between the nucleus and cytoplasm of oocytes from the common toad *Bufo bufo*. These are large, easy to collect cells, and isolating these two fractions for analysis is relatively straightforward. In this particular *Bufo bufo* microprobe study it was shown that a significant difference in potassium concentration existed between the nucleus, which has a high density of nucleic acids, and the cytoplasm, which has little nucleic acid content[189]. As an ion with negative charge, it is reasonable to also expect the concentration of chloride to be lower in the nucleus because of the density of anionic phosphates from DNA. Indeed the authors found a small difference, 98.8 mM compared to 91.3 mM, but this result

**Table 2.3) Distribution of Ions Between the Cytoplasm and Nucleus (Common Toad)**

| Ion (mM) | Cytoplasm | Nucleus |
|---|---|---|
| Na+ | 10.9 ± 1.9 | 10.4 ± 1.8 |
| K+ | 70.2** ± 3.2 | 266.4** ± 22.8 |
| Cl- | 98.8 ± 11.0 | 91.3 ± 9.0 |

** Indicates statistical significance

Data from Dick et al., 1978

was not statistically significant given the sample size. In a supporting study they also showed how incub-ation for 5 hours in a lithium-doped buffer would cause sodium to be displaced from the nucleus fraction (dropping from 10.4 to $2.4 \pm 0.7$ mM), but the potassium concentration remained unchanged versus the control. This is interesting because it suggests a preference for potassium cations over sodium in the stabilization of nucleic acid in nucleosomes and chromosomes within the nucleus.

When the presence of salt ions electrostatically enable strands to be in close proximity, non-covalent molecular contacts between interacting complementary strands spontaneously drive the association reaction. Hydrogen bonding between base pairs might initially be considered the driving force for association. However, the actual driving force is instead the favorable stacking of base pairs at the center

of the helix, driven by van der Waals and pi-pi orbital interactions between bases[190]. The hydrophobicity of the planar nucleobases also contributes to the stacking of nucleobases in the core of the duplex, but this mechanism is less prominent than in protein folding because nucleobases contain comparatively more heteroatoms than amino acid side chains and thus are less hydrophobic[191]. During the hybridization process several hydrogen bonds form between base pairs, three between G ⋮ C pairs, and two between A:T pairs. In terms of association forces, these contacts should be viewed as a secondary layer of interaction that supplements base stacking, and one that emphasizes sequence complementary between the strands.

The biophysical processes underlying the formation of the DNA duplex have been studied for many years, pioneered by early research in the 1950s from the lab of Paul Doty utilizing viscosity vs. temperature measurements for the thermal denaturation of bulk genomic DNA[192]. Although this dissertation focuses exclusively on thermal denaturation of nucleic acid, it should be noted that other methods such as pH and chemical denaturants such as urea can also be used for this purpose. As discussed in Chapter 1, during this time period of nucleic acid research the ability to synthesize polynucleotides of moderate length for biophysical experiments was not yet perfected; therefore the samples used in DNA melting studies were prepared from cellular extracts. Using this type of material it was shown that absorbance of ultraviolet light at the wavelength of 260 nm could be used instead of viscosity measurements and was a superior assay for monitoring the extent of nucleic acid denaturation[193]. With the spectroscopic technique it is assumed that the change in the extinction coefficient is proportional to the extent of denaturation. This means that as DNA denatures as a function of temperature the absorbance of UV light increases, and this phenomenon is termed hyperchromism. This effect is reversible, such that the UV absorbance near 260 nm decreases as DNA transitions from the random coil back to the duplex state, which is described as hypochromism. The idea that unstacked nucleobases exhibit hyperchromism is supported by the fact that a similar observation is made when polymer DNA is digested to its constituent mononucleotides by ribonucleases[194]. This chromicity effect ranges in magnitude from 20-40% depending on the length of polynucleotide studied[195,196]. A concept

central to all thermal denaturation studies is that of the melting temperature, $T_m$, which is taken as the midpoint of the melting transition. There are various methods to determine $T_m$ such as $\theta$ values and derivative approaches, but these will be discussed in further detail later

Since A:T and G ⋮ C base pairs involve a different number of hydrogen bonds, it can be expected that they contribute differently to the stability of the DNA double helix. It was showed by Julius Marmur and Doty that the $T_m$ is linearly related to the DNA base composition; a higher G ⋮ C content confers a higher thermal stability[197]. This enabled melting temperature information to be an indirect means to determine base composition from uncharacterized samples[198,199]. An improvement to this approach was made by observing additional wavelengths and using the calculated ratio of absorbance at 260 nm (A260) with A282 as well as A270 to determine base composition more accurately[200,201]. These multiple wavelengths have specific uses: the 260 nm wavelength is ideal for observing A:T base pairs, and 282 nm is better suited for G ⋮ C base pairs. At 270 nm the increase in hyperchromicity for unwinding A:T and G ⋮ C base pairs is approximately identical[202], and this evolved into the modern approach of absorbance at 268 nm as the ideal wavelength for spectroscopic analysis of DNA thermal denaturation[196].

### DEFINITION OF NEAREST-NEIGHBOR INTERACTIONS

As spectroscopic instrumentation improved, high resolution DNA melting experiments showed that thermodynamic stability depends not only on G ⋮ C content but also on sequence context[203]. This means that a given base pair (e.g. A:T) will have different thermodynamic stabilities depending on the identity of adjacent base pairs. With the ability synthesize precisely defined sequences it became evident that two duplexes with equal A:T content but unique sequence context will have different melting temperatures, which means one cannot assume all G ⋮ C or A:T base pairs to be identical in energetic contribution. For instance, a poly(dA)-poly(dT) duplex is more stable than a poly[d(AT)]-poly[d(AT)] duplex, even though the sequence composition is identical[204]. This finding led to expansion of the single

base pair model to the nearest-neighbor model (NN). In such an approximation, base pair interactions for the purposes of thermodynamic interaction parameters are considered as doublet base pairs where influences from adjacent base pairs are considered. The logic is that the thermodynamic properties enthalpy, entropy, and free energy ($\Delta H$, $\Delta S$, $\Delta G$, respectively) of a polynucleotide sequence can be expressed as the sum of the elementary nearest neighbor doublets. Another way to look at the NN approximation is that it considers the melting unit from the perspective of the internucleotide phosphate rather than individual base pairs. The fact that stacking of adjacent nucleobases is the major driving force of DNA stability increases the validity of this approach because the NN model emphasizes stacking whereas the single base pair approach emphasizes hydrogen bonding across the base pair. Beginning at the end of a 10-base pair duplex and processing forward (neglect the external ends—they are discussed later), 9 internal NN doublets (INN), each with their own thermodynamic contribution, would compose the overall energetics of the molecule. Given that there are four possible nucleobases (A, C, G, and T), in a two-base doublet model there are $4^2 = 16$ possible base pair combinations. However because of symmetry arising from the 5'→3' directionality of the helix, several combinations are repeated. For example, AC/GT, which is written in the form 5'→3'/5'→3', is identical to GT/AC. Because of these redundancies the number of INN pairs needed to describe duplex sequence context drops from 16 to 10.

**Table 2.4) Internal Nearest Neighbor Doublets**

| | | | |
|---|---|---|---|
| **AA/TT** | **AC/GT** | **AG/CT** | **AT/AT** |
| **CA/TG** | **CC/GG** | **CG/CG** | CT/AG |
| **GA/TC** | **GC/GC** | GG/CC | GT/AC |
| **TA/TA** | TC/GA | TG/CA | TT/AA |

The 16 possible Internal Nearest Neighbor (INN) doublets are written in a 5'→3'/5'→3' format. Because of symmetry arising from the 5'→3' directionality of the duplex there are redundancies (gray). Thus, only 10 unique INNs are needed to describe duplex sequence context (bold).

With the advent of PCR the need to predict thermodynamic stability and melting temperature of short DNA duplexes became obvious since a trial-and-error

experimental determination of primer melting temperatures is impractical due to financial and time costs of polynucleotide synthesis. Many laboratories[190,196,204-210] in the 1980's and 1990's pursued the extraction of thermodynamic parameters from experimental measurements that could describe the 10 INN doublets, with the ultimate goal of predicting thermodynamic behavior of unrelated sequences. This topicwas also investigated theoretically[211]. However, directly comparing developed NN parameters from different laboratories was difficult because each group utilized different oligo designs, buffer compositions, and most importantly, varying ion strengths. In one set of experiments DNA dumbbells were used[207], and in others DNA restriction fragments were used in ionic strengths ranging from 19.5 to 200 mM $Na^+$[190,196,205,212]. In a trend towards a unified parameter set[213] under standardized conditions with short synthetic DNA duplexes, 1M $Na^+$ was used in several of the studies[204,209,210,214], and the importance of this particular salt concentration will be explained over the next few sections.

## DNA Polyelectrolyte Theory: Phosphate Neutralization by Condensed and Screening Counterions

As mentioned earlier, the local cation concentration at the surface of DNA is very high due to electrostatic attraction from the polyanionic phosphate backbone. A conceptual framework for the interaction of monovalent $Na^+$ and $K^+$ with DNA can be elucidated from polyelectrolyte theory put forward by Gerald S. Manning[215,216]. In this model the polyelectrolyte chain (i.e. DNA) is substituted by a continuous line charge bearing $N_P$ charge groups (i.e. the number of phosphates) with density $\beta$, given by:

$$\beta = z_p e/b \tag{2.78}$$

where $z_p$is the valence of the charge groups (*i.e.* that of a phosphate, -1), $e$ is the elementary charge for a proton or electron ($1.60217657 \times 10^{-19}$ coulombs), and $b$ is the axial charge spacing (*i.e.* $b = L/N_P$, where $L$ is the axial length of the polyelectrolyte).

Charge density $\beta$ is invariable and not an adjustable parameter. Its value can be calculated for DNA using structural measurements obtained from

crystallographic or NMR observations In polyelectrolyte theory interactions between two or more polyions (*e.g.* DNA molecules) are neglected and the dielectric constant $\epsilon$ of the solution is taken as that of pure bulk continuum solvent (*i.e.* concentrations of $H^+$ and $OH^-$ from water are neglected). Monovalent counterions, $M^+$, interact with the polyelectrolyte in two ways, which is determined by Manning's linear charge-density parameter $\xi$:

$$\xi = \frac{e^2}{\epsilon k_B T b} \tag{2.79}$$

where $k_B$ is Boltzmann's constant, $T$ is the absolute temperature in Kelvin, $e$ and $\epsilon$ are defined previously. Recognize that the quantity:

$$\frac{e^2}{\epsilon k_B T} = \lambda_B \tag{2.80}$$

is the Bjerrum length $\lambda_B$, which is the distance between two elementary charges at which the electrostatic interaction is comparable in magnitude to thermal fluctuations, given the dielectric constant of the medium. So in essence, Manning's charge density parameter $\xi$ is simply the Bjerrum length divided by the axial phosphate charge spacing:

$$\xi = \frac{\lambda_B}{b} \tag{2.81}$$

Counterion Condensation (CC) theory states that if $\xi$ is less than unity ($\xi < 1$) then in sufficiently dilute solutions counterions will interact with the polyion (DNA) according to the Debye-Hückel approximation for a 2-charge system. However, if $\xi$ exceeds unity ($\xi > 1$) then counterions will "condense" on the polyion to lower the net value of the charge-density parameter $\xi$ to unity. The critical nature of the unity point ($\xi = 1$) had been discussed by Imai and Onishi in their application of Poisson-Boltzmann equation to rod-like charges and the first appearance of the term "counterion condensation" is attributed to them[217]. The particular word choice of "condensed" is significant because the ions are not site-bound, *per se*, but rather closely associated, so to avoid confusion the new terminology was adopted by Manning. The Bjerrum length in bulk water, which has a dielectric constant of 78.5 at 25°C is approximately 7.135 Å. Plugging this value

into equation (2.81) it becomes apparent that the maximum axial phosphate spacing so that $\xi > 1$ and counterion condensation occurs is 7.135 Å.

The actual spacing between phosphates is much lower, about 1.7 Å for B-form DNA double helix determined from structural data, and the coil can be assigned a value of 4.3 Å, which was estimated from experimental observations[218]. It should be noted that there is incredible debate as to the true phosphate spacing in a "coiled but linear" single strand DNA that is applicable for linear rod polyelectrolyte theory[219], but here this specific value is accepted merely for illustrative purposes since it has an often-cited literature reference. Taking these spacing lengths, the corresponding $\xi$ values in aqueous solution at 25°C are $\xi_h = 4.2$ for the helix and $\xi_c = 1.7$ for the coil[220]. On a polyion containing $N$ charges, the fraction of a phosphate charge neutralized by condensed counterions, $\theta_{M^+}$, is given as[215]

$$\theta_{M^+} = \frac{n_{M^+}}{N} = (1 - \xi^{-1}) \tag{2.82}$$

Using this calculation the fraction $\theta_{M^+} = 1 - (1/4.2) = 0.76$ of a counterion charge is condensed per charged phosphate in a double stranded DNA helix. Similarly, $\theta_{M^+} = 1 - (1/1.7) = 0.41$ of a counterion charge neutralizes a phosphate charge on a fully extended single strand "coiled but linear" through condensation. Since a lower fraction of the phosphate charge is neutralized by a condensed counterion in the strand form, this implies that counterions are released from the phosphate upon DNA denaturation. An experimental observation of the $\theta_{M^+}$ value for duplex DNA was made using NMR sodium-23 titration measurements[221]. It was shown that the value of $\theta_{M^+}$ remains constant, independent of NaCl concentration with a value of $\theta_{M^+} = 0.75 \pm 0.1$. This salt titration provided strong evidence that a constant charge fraction of a DNA phosphate is neutralized by counterions through delocalized association in a manner independent of counterion concentration, which is in excellent agreement with Manning's theoretical predictions[221,222].

Even after partial neutralization phosphates by counterion condensation, that is, $\xi$ has reached unity, both forms the nucleic acid are still highly charged polyelectrolytes and electrostatically attract additional ions. These were initially described by Manning as "uncondensed mobile ions"[215] and were treated with a

Debye-Huckel approximation. The concept of a "thermodynamically bound" fraction of a counterion charge was put forward by M. Thomas Record[220,223,224] as an explanation for this additional neutralization accomplished by a secondary layer of charge-screening counterions. The total thermodynamic extent of counterion binding was described using the parameter $\psi$, and is composed of condensed counterion $\psi_C$ and screening counterion $\psi_S$. That is:

$$\psi = \psi_C + \psi_S \tag{2.83}$$

In this nomenclature, Record's $\psi_C$ term is equivalent to Manning's $\theta_{M^+}$ term:

$$\psi_C = \theta_{M^+} \tag{2.84}$$

Using Manning's $\xi$ parameter, Record's $\psi_C$ and $\psi_S$ can be calculated for any polyelectrolyte of known axial charge density using:

$$\psi_C = 1 - \xi^{-1} \tag{2.85}$$

$$\psi_S = (2\xi)^{-1} \tag{2.86}$$

A quick calculation shows that:

$$\psi_C + \psi_S = 1 - (2\xi)^{-1} \tag{2.87}$$

So a simpler equation to calculate the thermodynamically bound fraction of a counterion's charge can be used instead:

$$\psi = 1 - (2\xi)^{-1} \tag{2.88}$$

This means that the screening effect of a solution counterion on the interactions of the residual phosphate charges is thermodynamically equivalent to the binding of an additional fraction $(2\xi)^{-1}$ of a counterion charge per phosphate. Combining equations (2.83, 2.85, and 2.86) the theoretical value of $\psi$ can be calculated for both the DNA duplex and single stranded coil using Manning's charge-density parameters of $\xi_h = 4.2$ for the helix and $\xi_c = 1.7$ for the coil[220]. Using these values, for double helical DNA $\psi = 0.88$ and for single stranded DNA $\psi = 0.70$. The results are summarized in Table 2.5.

**Table 2.5) The Theoretical Fraction of a Phosphate Charge Neutralized by Counterions: DNA Duplex vs. DNA Single Strand**

| Parameter | Description | DNA Double Helix | DNA Single Strand |
|---|---|---|---|
| $\xi$ | Manning's charge-density parameter | 4.2 | 1.7 |
| $\psi_C$ | Fraction of a phosphate charge neutralized by "Condensed" counterions | 0.76 | 0.41 |
| $\psi_S$ | Fraction of a phosphate charge neutralized by "Screening" counterions | 0.12 | 0.29 |
| $\psi$ | Total fraction of a phosphate charge neutralized by counterions | 0.88 | 0.70 |

These calculations reveal that a smaller fraction of a counterion charges are consumed by neutralizing DNA phosphates in a single strand vs. a duplex's phosphates. Since they aren't interacting with phosphates, counterions are more capable of electrostatically interacting with other charges in the solution, such as the solvent or other ions. This can be interpreted as counterions having a greater entropy when DNA is in the single strand state vs. the duplex state.

The implications of these calculations must be emphasized to explain in general the thermodynamic consequences of DNA denaturation. These values show that a smaller fraction of a counterion's charge is thermodynamically bound to a phosphate of a DNA duplex compared to what is bound to a phosphate of a DNA single strand. In other words, when a DNA duplex denatures into single strands, a greater fraction of a solution counterion's charge is capable of interacting with something other than DNA phosphates. Perhaps this could the solvent, or maybe other ions, the identity is not important but it can be assumed that its new interacting partner will be more mobile than the large DNA macromolecule. If one assumes a relation between mobility, thermodynamic disorder, and entropy, then the entropic term for counterions increase when DNA is in the single strand state.

### SALT EFFECTS ON THE STABILITY OF DNA ARE ENTROPIC IN ORIGIN

It is well known that the stability of DNA increases as a function of moderate salt concentrations, and Manning developed a theoretical description that decomposed the thermodynamic change in free energy into it's enthalpic and entropic terms[216]. The rigorous derivation of equations is beyond the scope of this chapter but those details can be found in his well-known "Application of Limiting

Laws" series, specifically Section 1 where the helix-to-coil transition in monovalent salts is discussed[216]. To briefly summarize the final equations: the free energy difference *per mole base* between helix and coil states ($\Delta g^{\circ} = (G_{\mathrm{c}}^{\circ} - G_{\mathrm{h}}^{\circ})/N_{\mathrm{P}}$), at a specific temperature, $T$, and counterion concentration, $M_{\mathrm{A}}^{+}$, from an "AB" salt, is a function of salt concentration and can be expressed as:

$$\Delta g(T, M_{\mathrm{A}}^{+}) = \Delta\mu^{\circ}(T) + RT\alpha(T) + 1.15\,\eta RT\,\log M_{\mathrm{A}}^{+} \tag{2.89}$$

where $\Delta\mu^{\circ}(T)$ is the chemical potential difference per base between the two states (e.g. $\Delta\mu^{\circ} = \mu_{\mathrm{c}}^{\circ} - \mu_{\mathrm{h}}^{\circ}$), $\eta = \xi_{\mathrm{c}}^{-1} - \xi_{\mathrm{h}}^{-1}$, and $\alpha$ is a term representing the temperature dependence of the chemical potential of the counterions, which is described further in Manning's text[216]. Expressions for the enthalpy change, $\Delta h$, and entropy change, $\Delta s$, are obtained from the temperature derivatives of $\Delta g$:

$$\Delta h(T, M_{\mathrm{A}}^{+}) = \Delta h^{\circ}(T) + R\,\mathrm{d}\,\alpha/\mathrm{d}(1/T) \tag{2.90}$$

$$\Delta s(T, M_{\mathrm{A}}^{+}) = \Delta s^{\circ}(T) + R\,\mathrm{d}\,\alpha/\mathrm{d}(T) - 1.15\,\eta R\,\log M_{\mathrm{A}}^{+} \tag{2.91}$$

where:

$$\Delta h^{\circ} = \mathrm{d}(\Delta\mu^{o}/T)/\mathrm{d}(1/T) \tag{2.92}$$

and,

$$\Delta s^{\circ} = -\mathrm{d}\Delta\mu^{\circ}/\mathrm{d}T \tag{2.93}$$

The point pertinent to this discussion is that the equation for free energy (2.89) indicates how $\Delta g$ is a function of $\log M_{\mathrm{A}}^{+}$ at any fixed temperature $T$, but the $\Delta s$ equation (2.91) predicts that the dependence of $\Delta g$, and thus $T_{\mathrm{m}}$, on salt concentration is entirely entropic in origin, and that $\Delta h$ (2.92) is independent of salt concentration. These equations compared well with the early calorimetric observations at the time from Privalov, Ptitsyn, and Birshtein[225], and also Krakauer and Sturtevant[226].

Expanding on one example, Manning's theory of the entropic nature of ion stabilization of nucleic acid duplexes is demonstrated in the calorimetric study of Privalov and Filimonov[227] where they investigated the denaturation of poly(rA):poly(rU) RNA duplexes as a function of [NaCl] (0.01-0.1 M). Shown in Table 2.6 below is data from their manuscript:

**Table 2.6) Salt Effects on Oligonucleotide Melting Temperature is Entropic in Origin**

| [Na+] (M) | $T_m$ (K) | $\Delta H^{(A:U)}$ (kJ mol$^{-1}$) | $\Delta S^{(A:U)}$ (J K$^{-1}$ mol$^{-1}$) | $\Delta G^{(A:U)}$ (kJ mol$^{-1}$) |
|---|---|---|---|---|
| 0.01 | 310.5 | 34.0 | 110 | 1.3 |
| 0.05 | 324.0 | 34.0 | 105 | 2.8 |
| 0.10 | 331.0 | 34.0 | 102 | 3.5 |

Data from Privalov & Filimonov, 1978.

In this table presented by Privalov and Filimonov[227] the enthalpic, entropic and free energy terms have been corrected using $\Delta C_P$ for the transition. The details of this correction are omitted here for brevity, but the equations used are covered in detail within their manuscript. From this corrected data it was suggested that enthalpy of base pairing is independent of ionic strength. As mentioned earlier, previous calorimetric studies on the enthalpy of denaturation of T2 phage DNA[225] and calf thymus DNA[228] as a function of salt concentration show a similar dependence of $\Delta H^{o}_{Tm}$ on $T_m$ to the observations of Privalov[227], and together this evidence supports the generalization that the effect of electrolyte concentration on the stability of the double helix is primarily entropic in origin.

## THE DEPENDENCE OF MELTING TEMPERATURE ON [SALT] INFLUENCES THE USE OF 1.0 M [SALT] AS IDEAL FOR DNA NEAREST-NEIGHBOR PARAMETER DETERMINATION

DNA melting transitions are highly dependent on ionic strength and this influence has been investigated extensively. In early studies it was shown that the melting temperature of DNA increases linearly with the logarithm of sodium cation concentration up to ~1M[195,199,229-231], and more recently such an effect of ion concentration on melting temperature has been noted in DNA microarray applications[232].

A relation between the melting temperature, the observed transition enthalpy, bulk solution salt concentration, and condensed counterions on the duplex was put forward by Record[220]:

$$\frac{dT_m}{d\ln M^{\pm}} = \alpha \frac{RT_m^2}{\Delta H^{o}_{obs}} \frac{\Delta\theta_{M^+}}{2} \tag{2.94}$$

where terms repeat their definitions from previous sections, the quantity $\alpha \equiv [1 - d\ln\gamma_{M^{\pm}}/d\ln M^{\pm}]$, $\gamma_{M^{\pm}}$ is the activity coefficient of the salt, and $M^{\pm}$ is the concentration of the monovalent salt such that the assumption $M^+ \cong M^{\pm}$ is valid.

Calorimetric studies on nucleic acid denaturation reveal that $RT_m^2/\Delta H_{obs}^o$ is a constant independent of temperature, counterion concentration, and chain length[225,226,228,233], and in simplification $RT_m^2/\Delta H_{obs}^o$ is substituted by the symbol $\beta$. Reintroducing Record's thermodynamic monovalent counterion association parameters ($\psi_h$ and $\psi_c$)[223] discussed earlier, the equation then simplifies to:

$$\frac{dT_m}{d\ln M^{\pm}} = \alpha\beta\Delta\psi \tag{2.95}$$

where, over the salt concentration range of $10^{-3}$ to $10^{-1}$ M, $\alpha \cong 0.95 \pm 0.03$[220]. However at higher salt concentrations such as those in the molar range, the equation becomes inadequate to describe salt effects because it neglects anion and solvent interactions with the nucleic acid, since its polyelectrolyte theory basis is limited to cation-nucleic acid interactions. Melting experiments of Hamaguchi & Geiduschek[234], Schildkraut & Lifson[195] , and Gruenwedel & Hsu[230] observed that $dT_m/d\ln M^{\pm}$ increases over the $10^{-3}$ to $10^{-1}$ M salt range, decreases above 0.1 M salt, becomes zero near 1 M salt, and takes on negative values above 1 M salt. It is generally suggested that at high salt concentrations, the $Cl^-$ anion acts as a hydrophobic bond-breaking agent. For other salts this action is anion-specific and the effect on melting temperature follows the lyotropic Hoffmeister series[235]. The exact mechanism of anion-specificity is not entirely clear, however it is likely to involve charge density differences and to hydration shell structure as the size and polarizability of anions changes[236], and the differential interaction of these moieties with the polar groups on unstacked nucleobases[237].

Because of the minimal impact of salt on the melting temperature at concentrations near 1M ($dT_m/d\ln M^{\pm} \approx 0$ at 1.0M NaCl), it is reasonable to suggest 1M NaCl is the optimal salt concentration for which to calculate nearest neighbor parameters for DNA thermodynamics.

### NEAREST-NEIGHBOR INTERACTIONS: ASSUMPTIONS ABOUT THE DUPLEX ENDS

Using the nearest-neighbor (NN) method for describing thermodynamic stability of DNA, the duplex is decomposed into a linear combination of NN doublets. As described earlier there are 10 possible unique internal NN doublets (INN) combinations, but this leaves the remaining ends of the duplex unaccounted for. For

longer DNA polymers the ends are nearly negligible due to being greatly outnumbered by the INNs but for short oligomer duplexes the effect of end interactions becomes substantial. Using the notation of "E" for end of the molecule, there are four possible types of end interactions:

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

Adding these interactions, the total number of possible n-n sequence dependent interactions then becomes 14, but various models have adopted different approaches for reducing this number because it is not possible to derive all 14 theoretically feasible parameters without making assumptions. This is because the NN model, by definition, requires a neighbor on both sides of a base pair. A more detailed description of this mathematical limitation and explanation for the maximum of 12 unique parameters can be investigated further in the Ph.D. dissertation of Richard Owczarzy[238]. A comparison of various assumptions and the resulting number of unique parameters is tabulated below:

**Table 2.7) The Number of Possible NN Doublets, Unique Parameters, and Assumptions for Variations of the NN Model**

| Model Type | General N-N | Allawi & Santa Lucia | INN Model + General Initiation Constant | INN Only Model |
|---|---|---|---|---|
| Assumption About End Interactions | none | AE/ET = TE/EA<br>CE/EG = GE/EC | AE/ET = TE/EA = CE/EG = GE/EC | AE/ET = 0<br>TE/EA = 0<br>CE/EG = 0<br>GE/EC = 0 |
| # of NN | 14 | 12 | 11 | 10 |
| # of Unique Parameters | 12 | 12 | 11 | 10 |

Various assumptions are made to reduce the number of unique parameters to 12, which is the most that can be determined, or fewer. This is an abbreviated version of a larger table provided in Owczarzy, 1999.

## NEAREST-NEIGHBOR INTERACTION PARAMETER SETS: SINGLET VS. DOUBLET FORMATS

Using the NN approach the sequence dependence of duplex energetics is considered to arise from the cumulative contributions of the Watson-Crick hydrogen bonding between base pairs and the stacking interactions of neighboring

nucleobases. There are two approaches to calculating the total doublet interaction energies, referred to as the singlet and doublet formats. In the singlet format the H-bonding and stacking interactions are considered separately, whereas in the doublet format the entire NN interaction is considered in a single parameter[239]. The choice format is left to the investigator since the calculated total thermodynamic values from both formats are numerically equivalent within roundoff error[213], but in practice the doublet format provides a simpler calculation. This dissertation focuses solely on parameters developed using the doublet format because of two assumptions made in the singlet approach that are inappropriate for the aims of this work. The first is the initial equivalency of all A:T and all G:C base pair melting temperatures (NN differences are accounted for afterwards), which seems to contradict the philosophy behind NN benefits over the single base pair approach. In the doublet format sequence context is inherently present from the start of the calculation, whereas in the singlet format sequence context is accounted for afterwards in the free energy portion of parameters, which could perhaps be interpreted as a pseudo correction. The second and most important reason for choosing the doublet format is due to the singlet format's assumption that the entropy change of base pair melting $\Delta S_{\mathrm{bp}}$ is independent of sequence and [$Na^+$] over the range from 0.02 to 1.0 M $Na^+$, with an average value of $\Delta S_{\mathrm{bp}} = 24.85 \pm 1.74$ cal $mol^{-1}K^{-1}$. This constant was determined from analysis[190] of nearly 30 $\Delta S = \Delta H/T_{\mathrm{m}}$ values from calorimetric studies such as those mentioned previously[190,225,226,228,233]. Using the singlet approach, the total transition entropy of the entire duplex comprising $N_{\mathrm{A:T}}$ A:T base pairs and $N_{\mathrm{G:C}}$ G⁝C base pairs would thus be calculated as:

$$\Delta S_{\mathrm{duplex}} = \Delta S_{\mathrm{bp}}(N_{\mathrm{A:T}} + N_{\mathrm{G:C}}) \tag{2.96}$$

Given that this chapter focuses on the energetic contributions of dithioate substitution in every INN pair, where distinct differences are observed, consideration of singlet parameter sets were avoided due to this generalized transition enthalpy value. The "dithioate correction" parameters developed in this

dissertation are easier for non-experts to apply when using the doublet format, and are presented as such.

Below in Table 2.8 is a summary of popular NN parameter sets developed experimentally, but from here forward only the doublet format of NN parameters is discussed. Further information on calculating thermodynamic values from the singlet format can be found in the manuscript out of Albert Benight's lab[239].

**Table 2.8) List of Experimentally Developed NN Parameter Sets by Year**

| *NN Parameter Set* | *Year* | *[Na+]* | *Format* |
|---|---|---|---|
| Gotoh and Tagashira | 1981 | 19.5 mM | Singlet |
| Ornstein and Fresco | 1983 | 50 mM | Singlet |
| Wartell and Benight | 1985 | 100 mM | Singlet |
| Breslaeur et al. | 1986 | 1.0 M | Doublet |
| McCampbell et al. | 1989 | 102 mM | Singlet |
| Delcourt and Blake | 1991 | 75 mM | Singlet |
| Doktycz et al. | 1992 | 115 mM | Singlet |
| SantaLucia et al. | 1996 | 1.0 M | Doublet |
| Sugimoto et al. | 1996 | 1.0 M | Doublet |
| Allawi and SantaLucia | 1997 | 1.0 M | Doublet |

Adapted from a more comprehensive table in Owczarzy, 1999.

### COMPARISON OF DOUBLET FORMAT NN PARAMETER SETS IN 1.0 M SALT

When published NN parameter sets are compared they are remarkably similar despite being developed independently with vastly different sequence compositions[213]. This evidence supports the validity of developing NN parameter sets for the prediction of nucleic acid thermodynamic values. As shown in Table 2.8 above, four groups of NN parameters can be easily compared because they have been calculated using the doublet format in a standardized salt concentration of 1.0 M NaCl. The importance of this specific salt concentration was discussed in detail previously where it was explained that $dT_m/d\ln M^{\pm} \approx 0$ at 1.0M NaCl. The entropic effect of counterions on duplex stability is more significant at other salt concentrations, therefore 1.0 M NaCl can be considered the safest concentration from which parameters from different labs can be compared and used for prediction, since buffer composition and preparation might vary across different

hands. For example, a common mistake in buffer preparation by young investigators in is the neglect of the contribution of the buffering agent on the total concentration of dissolved $Na^+$, particularly when sodium phosphate or sodium citrate is used. This problem is exacerbated by the impurity of salt reagents (e.g 95% vs 99%), imperfect stock preparations, or varying water quality; all of which would require additional titration of acid/base to achieve specific pH values in the final buffer. Because of these variables and the point that $dT_m/d\ln M^{\pm} \approx 0$ at 1.0 M NaCl, this is generally accepted as the optimal salt concentration for the development of nucleic acid thermodynamic parameters. In practice, experimental salt concentrations are much lower and duplex thermodynamic values calculated from 1.0 M NaCl NN parameters are modified *ex post facto* by empirically determined salt concentration correction terms so that the calculated values are appropriate for whichever condition the investigator is using.

A comparison of these four parameter sets shows remarkable consistency in the determined thermodynamic values (Table 2.9). Even where discrepancies exist, it must be noted that the calculated free energy values ($\Delta G = \Delta H - T\Delta S$) end up being very similar due to enthalpy-entropy compensation. At the bottom of the table are initiation energies, which will be discussed next, and symmetry correction terms, which account for unfavorable energetics of complementary single strand interactions that must be overcome. The nucleation initiation term is a general value that currently encompasses many energetic factors, but was historically introduced to account for the thermodynamic difference between formation of the first base pair between separate strands and the remaining internal base pairs. This "intiation" term evolved over time to include, in addition to differences between terminal and internal NNs, the interactions of the terminal ends of the linear duplex with surrounding solvent and counterion condensation[213,240,241].

**Table 2.9) Comparison of NN Thermodynamic Parameters (Doublet, 1.0 M NaCl)**

| Sequence (i,j) | Breslauer et al. | | SantaLucia et al. | | Sugimoto et al. | | Allawi and SantaLucia | |
|---|---|---|---|---|---|---|---|---|
| | $\Delta H_{i,j}$ | $\Delta S_{i,j}$ | $\Delta H_{i,j}$ | $\Delta S_{i,j}$ | $\Delta H_{i,j}$ | $\Delta S_{i,j}$ | $\Delta H_{i,j}$ | $\Delta S_{i,j}$ |
| **AA/TT** | -9.1 | -24.0 | -8.4 | -23.6 | -8.0 | -21.9 | -7.9 | -22.0 |
| **AC/GT** | -6.5 | -17.3 | -8.6 | -23.0 | -9.4 | -25.5 | -8.4 | -22.4 |
| **AG/CT** | -7.8 | -20.8 | -6.1 | -16.1 | -6.6 | -16.4 | -7.8 | -21.0 |
| **AT/AT** | -8.6 | -23.9 | -6.5 | -18.8 | -5.6 | -15.2 | -7.2 | -20.4 |
| **CA/TG** | -5.8 | -12.9 | -7.4 | -19.3 | -8.2 | -21.0 | -8.5 | -22.7 |
| **CG/CG** | -11.9 | -27.8 | -10.1 | -25.5 | -11.8 | -29.0 | -10.6 | -27.2 |
| **CC/GG** | -11.0 | -26.6 | -6.7 | -15.6 | -10.9 | -28.4 | -8.0 | -19.9 |
| **GA/TC** | -5.6 | -13.5 | -7.7 | -20.3 | -8.8 | -23.5 | -8.2 | -22.2 |
| **GC/GC** | -11.1 | -26.7 | -11.1 | -28.4 | -10.5 | -26.4 | -9.8 | -24.4 |
| **TA/TA** | -6.0 | -16.9 | -6.3 | -18.5 | -6.6 | -18.4 | -7.2 | -21.3 |
| Initiation if at least one terminal G:C | 0.0 | -16.8 | 0.0 | -5.9 | 0.6 | -9.0 | NA | NA |
| Initiation with both terminal A:T | 0.0 | -20.1 | 0.0 | -9.0 | 0.6 | -9.0 | NA | NA |
| Symmetry correction | 0.0 | -1.3 | 0.0 | -1.4 | 0.0 | -1.4 | 0.0 | -1.4 |
| 5' T:A correction | - | - | 0.4 | 0.0 | - | - | - | - |
| Initiation for each terminal A:T | NA | NA | NA | NA | NA | NA | 2.3 | 4.1 |
| Initiation for each terminal G:C | NA | NA | NA | NA | NA | NA | 0.1 | -2.8 |

Enthalpies ($\Delta H_{i,j}$) are in kcal·mol$^{-1}$ and entropies ($\Delta S_{i,j}$)are in cal·mol$^{-1}$K$^{-1}$ (yellow). Initiation energies compose the bottom rows (red), and this concept evolved over time (i.e. the NA, not applicable designation). There is also a small symmetry correction term for self-complementary sequences.

A thorough examination of many parameter sets (both singlet and doublet) was performed by John SantaLucia and published as a "unified parameter set" in 1998[213]. However, the values in this unified table are actually identical to a parameter set published by Allawi and SantaLucia in 1997[210], so the "unification" should be interpreted more as a validation using predicted/experimentally observed regression unity rather than true unification by averaging methods. Regardless, the current *de facto* thermodynamic NN doublet parameter set is that of Allawi and SantaLucia[210] and is shown in full form in Table 2.10 below.

**Table 2.10) Unified Nearest Neighbor Parameters (1.0 M NaCl)**

| NN Sequence (i,j) | $\Delta H^o$ (kcal·mol$^{-1}$) | $\Delta S^o$ (cal·mol$^{-1}$·K$^{-1}$) | $\Delta G^o_{37}$ (kcal·mol$^{-1}$) |
|---|---|---|---|
| **AA/TT** | -7.9 ± 0.2 | -22.0 ± 0.8 | -1.00 ± 0.01 |
| **AC/GT** | -8.4 ± 0.5 | -22.4 ± 2.0 | -1.44 ± 0.04 |
| **AG/CT** | -7.8 ± 0.6 | -21.0 ± 2.0 | -1.28 ± 0.03 |
| **AT/AT** | -7.2 ± 0.7 | -20.4 ± 2.4 | -0.88 ± 0.04 |
| **CA/TG** | -8.5 ± 0.6 | -22.7 ± 2.0 | -1.45 ± 0.06 |
| **CG/CG** | -10.6 ± 0.6 | -27.2 ± 2.6 | -2.17 ± 0.05 |
| **CC/GG** | -8.0 ± 0.9 | -19.9 ± 1.8 | -1.84 ± 0.04 |
| **GA/TC** | -8.2 ± 0.6 | -22.2 ± 1.7 | -1.30 ± 0.03 |
| **GC/GC** | -9.8 ± 0.4 | -24.4 ± 2.0 | -2.24 ± 0.03 |
| **TA/TA** | -7.2 ± 0.9 | -21.3 ± 2.4 | -0.58 ± 0.06 |
| **Init. w/ Term. A:T** | 2.3 ± 1.3 | 4.1 ± 0.2 | 1.03 ± 0.05 |
| **Init. w/ Term. G:C** | 0.1 ± 1.1 | -2.8 ± 0.2 | 0.98 ± 0.05 |
| **Symmetry Corr.** | 0 | -1.4 | 0.4 |

## CALCULATION OF PREDICTED DUPLEX THERMODYNAMICS USING NN PARAMETERS

Thermodynamic parameters allow the prediction of transition enthalpy, transition entropy, and the free energy of the DNA duplex. This is accomplished by calculating the sum of the individual thermodynamic parameters of the constituent NN doublets in the duplex, including initiation energies and the symmetry correction, if necessary. That is:

$$\Delta H_{\text{duplex}} = \sum N_{\text{i,j}} \Delta H_{\text{i,j}} + \Delta H_{\text{init}} \tag{2.97}$$

where $N_{\text{i,j}}$ is number of times the specific (i,j) NN doublet appears, and $\Delta H_{\text{i,j}}$ is the parameter value hashed from the database. Since the symmetry correction term is entropic in origin (Table 2.10), it is neglected from the transition enthalpy calculation. Using an analogous approach the duplex transition entropy can also be calculated:

$$\Delta S_{\text{duplex}} = \sum N_{\text{i,j}} \Delta S_{\text{i,j}} + \Delta S_{\text{init}} + \Delta S_{\text{sym}} \tag{2.98}$$

The free energy of the duplex can also be calculated by hashing values from a parameter database (Table 2.10) using a similar equation:

$$\Delta G^{37}_{\text{duplex}} = \sum N_{\text{i,j}} \Delta G^{37}_{\text{i,j}} + \Delta G^{37}_{\text{i,j}} + \Delta G_{\text{sym}} \tag{2.99}$$

Unlike transition enthalpy and transition entropy which are applicable for the melting transition temperature, the free energy parameters are reported at a standard temperature, since the transition free energy, by definition, would be zero at the melting temperature. In older parameter tables the traditional standard temperature of 25°C was reported[204], however, recently this has been supplanted by 37°C[210], which corresponds to the human body temperature. The free energy values are oftentimes not even reported, such as in SantaLucia's unified parameter set[213], because the free energy can be calculated from enthalpy and entropy values using the relation $\Delta G = \Delta H - T\Delta S$, at temperature $T$. This calculation is useful for investigators interested in duplex stability at temperatures far from standard temperatures, such as the stability of DNA primers at the annealing temperature during PCR reactions. As a demonstration on comparing predicted versus observed thermodynamic values we can analyze published melting temperature data for a duplex from an earlier SantaLucia manuscript[209]:

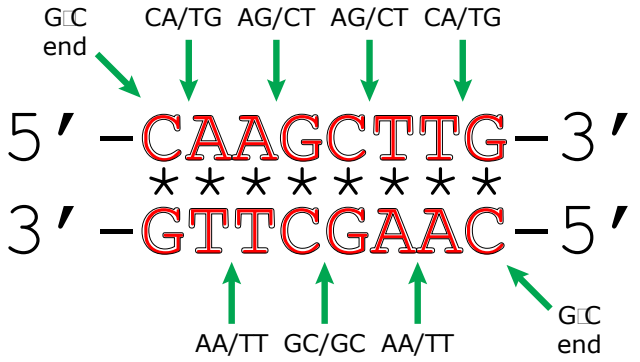



**Fig. 2.5) Calculating Duplex Thermodynamic Values from NN Parameters**

The number of appearances of each (i,j) NN doublet is determined:

$N_{\text{AA/TT}} = 0$ $N_{\text{AC/GT}} = 0$ $N_{\text{CA/TG}} = \mathbf{2}$ $N_{\text{CC/GG}} = 0$ $N_{\text{GC/GC}} = \mathbf{1}$

$N_{\text{AG/CT}} = 0$ $N_{\text{AT/AT}} = \mathbf{2}$ $N_{\text{CG/CG}} = 0$ $N_{\text{GA/TC}} = \mathbf{2}$ $N_{\text{TA/TA}} = 0$

As well as the necessary initiation and symmetry correction terms:

$N_{\text{A:T}} = 0$ $N_{\text{G:C}} = \mathbf{2}$ Symmetry? = **yes**

Which allows one to calculate the predicted transition enthalpy:

$$\Delta H_{\text{duplex}} = N_{\text{AT/AT}}\Delta H_{\text{AT/AT}} + N_{\text{CA/TG}}\Delta H_{\text{CA/TG}} + N_{\text{GA/TC}}\Delta H_{\text{GA/TC}} + N_{\text{GC/GC}}\Delta H_{\text{GC/GC}} + N_{\text{G : C end}}\Delta H_{\text{G : C init}}$$

$$\Delta H_{\text{duplex}} = 2(-7.2) + 2(-8.5) + 2(-8.2) + 1(-9.8) + 2(0.1)$$

$$\Delta \boldsymbol{H}_{\text{duplex}} = \mathbf{-57.4}\ \text{kcal·mol}^{-1} \qquad (\Delta \boldsymbol{H}_{\text{observed}} = \mathbf{-56.1}\ \text{kcal·mol}^{-1})$$

The predicted enthalpy value agrees well with an experimentally determined quantitiy for this duplex system[209]. Following a similar calculations, the transition entropy can be determined:

$$\Delta S_{\text{duplex}} = N_{\text{AT/AT}}\Delta S_{\text{AT/AT}} + N_{\text{CA/TG}}\Delta S_{\text{CA/TG}} + N_{\text{GA/TC}}\Delta S_{\text{GA/TC}} + N_{\text{GC/GC}}\Delta S_{\text{GC/GC}} + N_{\text{G : C init}}\Delta S_{\text{G : C init}} + \Delta S_{\text{sym}}$$

$$\Delta S_{\text{duplex}} = 2(-20.4) + 2(-22.7) + 2(-22.2) + 1(-24.4) + 2(-2.8) + (-1.4)$$

$$\Delta \boldsymbol{S}_{\text{duplex}} = \mathbf{-162.0}\ \text{cal·mol}^{-1}\text{·K}^{-1} \qquad (\Delta \boldsymbol{S}_{\text{observed}} = \mathbf{-157.3}\ \text{cal·mol}^{-1}\text{·K}^{-1})$$

And also the free energy, at 37°C:

$$\Delta G^{37}_{\text{duplex}} = N_{\text{AT/AT}}\Delta G^{37}_{\text{AT/AT}} + N_{\text{CA/TG}}\Delta G^{37}_{\text{CA/TG}} + N_{\text{GA/TC}}\Delta G^{37}_{\text{GA/TC}} + N_{\text{GC/GC}}\Delta G^{37}_{\text{CA/TG}} + N_{\text{G : C init}}\Delta G^{37}_{\text{G : C init}} + \Delta G^{37}_{\text{sym}}$$

$$\Delta G^{37}_{\text{duplex}} = 2(-0.88) + 2(-1.45) + 2(-1.30) + 1(-2.24) + 2(0.98) + (0.4)$$

$$\Delta \boldsymbol{G}^{37}_{\text{duplex}} = \mathbf{-7.14}\ \text{kcal·mol}^{-1} \qquad (\Delta \boldsymbol{G}^{37}_{\text{observed}} = \mathbf{-7.3}\ \text{kcal·mol}^{-1})$$

## 2.3) Extracting Thermodynamic Values from Non-Calorimetric Equilibrium Melting Curves

### INTRODUCTION

Equilibrium measurements of DNA's helix-to-coil dissociative conformational change can be used to extract thermodynamic information about the transition state. An ideal starting point in the literature for any biophysicist embarking on such studies would be manuscripts by Luis Marky and Kenneth Breslaeur[242,243], which cover in detail the mathematical models for such thermodynamic parameter

extraction. Any observable phenomenon that measures the relative population of both states is suitable for providing insight into the association and dissociation processes of complementary oligonucleotide strands. Historically, the absorption of ultra-violet light at the wavelength appropriate for DNA (~260 nm) has been used to monitor the strand dissociation as a function of temperature. As the temperature increases the stabilizing forces of the duplex are reduced and electrostatic repulsion from the anionic phosphate backbones drive forward the dissociation process. When on single strands the aromatic nucleobases have higher absorbance compared to the stacked pi-pi orientation in the duplex state, and this phenomenon is termed a hyperchromic shift. Although temperature-induced hyperchromicity is most commonly studied, other methods such as pH and chemical denaturants can also be used to drive the helix to coil transition. This dissertation, however, focuses solely on temperature-dependent equilibrium melting curves, and specifically those determined through a high-throughput fluorescence assay rather than classic UV absorbance. This method was chosen because of the ease at which it can be scaled up using modern real-time PCR machines, which have a higher sample throughput compared to UV spectrophotometers. Fluorescence melting curves are processed in an analogous manner to UV melting curves and to explain the analysis techniques an example of fluorescence data collected on the thermal transition of the H820-H822 duplex at 2 μM concentration is shown over the next few pages.

### FRET DNA HYBRIDIZATION ASSAY

Nucleic acid hybridization was first monitored by fluorescence resonance energy transfer (FRET) in the late 1980s[244]. This technique involves a quencher, that when in close proximity to a fluorophore, results in a reduction in the fluorescence intensity measured by a spectrofluorometer. This method is useful for monitoring DNA conformational changes when the system is designed with appropriately placed moieties. Upon dissociation a rapid increase in fluorescence is observed as the fluorophore and quencher become more distant. In the studies presented in this dissertation a fluorescence probe is attached the 5'-end of one

strand of the DNA duplex and a quencher on the adjacent 3'-end of the complementary strand, according to the schematic below.

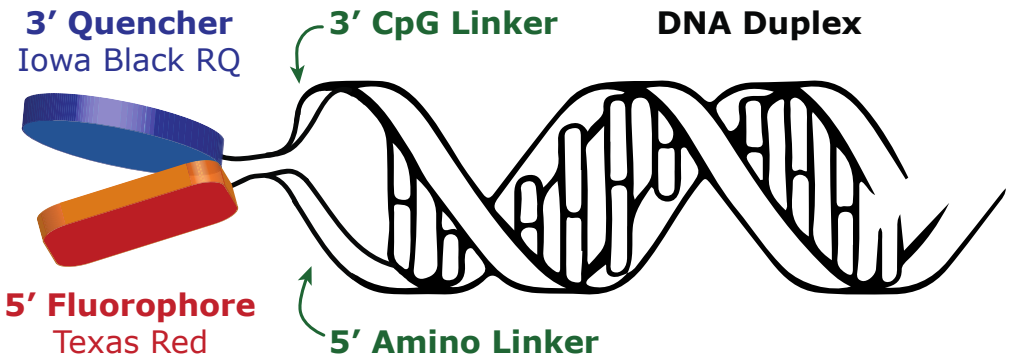



**Fig. 2.6) Schematic of FRET DNA Duplex Hybridization Assay**

Complementary strands contain one fluorophore and quencher, on the 5'- and 3'-end, respectively. In the duplex form, nearly all fluorescence is quenched. Upon denaturation during the helix-to-coil transition a significant increase in fluorescence is observed. The intensity of fluorescence is the observable for monitoring the equilibrium transition and is comparable to traditional UV absorbance measurements. The fluorophore used in this dissertation is Texas Red and it is coupled to the 5'-end via an amino linker. The 3'-end of the complementary strand contains an Iowa Black RQ quencher, which comes from a specially modified CpG support that is precoupled to the quencher. Methods for synthesis and coupling are covered in Chapter 1.

The melting transition of the DNA duplex is coupled to the separation of fluorophore and quencher and a concomitant increase in fluorescence intensity. The FRET-based experimental approach benefits from the high sensitivity of fluorescence spectroscopy as labeled oligonucleotides can reach the nanomolar concentration and still remain observable. This is amenable to thermodynamic measurements because a wide range of oligo concentrations, across several orders of magnitude, can be investigated. The major advantage of FRET-based methods is ability to scale up to large numbers of samples, as the development of modern real-time PCR equipment has enabled the measurement of fluorescence for hundreds of small volume samples in parallel on plastic plates. FRET-based melting temperature determination using real-time PCR has previously been performed on DNA quadruplexes[245-247], molecular beacons[248,249], duplexes, triplexes[250], and nanostructures[251]. However, complete thermodynamic analysis involves determination of parameters beyond just melting temperatures and performance of FRET-based methods in this area has struggled because of difficulties with non-

linear baselines that complicate analysis. Richard Owczarzy, a collaborator on this project, put forward a set of experiments in 2011 that addresses this concern by optimizing the fluorophore/quencher combination[252]. Twenty-two commercially available fluorophores and quenchers were assayed for their ability to reliably report thermodynamics of melting transitions in comparison with traditional UV spectroscopic methods. Of these, optimal performance was observed using Texas Red or ROX dyes with Iowa Black RQ or Black Hole quenchers. These labels provide accurate melting temperatures, and transition free energies, enthalpies, and entropies while retaining the two-state nature of duplex melting[252]. The fluorescence of common dyes such as rhodamine-green, carboxytetramethyl-rhodamine (TAMRA), 6-carboxyfluorescein (FAM), Cy3, Cy5, and various Alexa Fluors suffer from significant temperature dependence that makes them unsuitable for thermal melting studies an should be avoided. The studies in this dissertation use the Texas Red fluorophore and Iowa Black RQ quencher.

### CONVERTING RAW FLUORESCENCE DATA TO EQUILIBRIUM Θ TERM

The helix-coil transition is an equilibrium process and it is assumed that the fractional change in fluorescence monitors the extent of the dissociation reaction. Raw fluorescence information is converted graphically to an equilibrium term, Θ, which is defined in this dissertation as the fraction of oligos in the single strand coil state. This is performed using baseline subtraction from "helix" and "coil" lines according to the formula:

$$\Theta = \frac{F(T) - F_L(T)}{F_U(T) - F_L(T)} \tag{2.100}$$

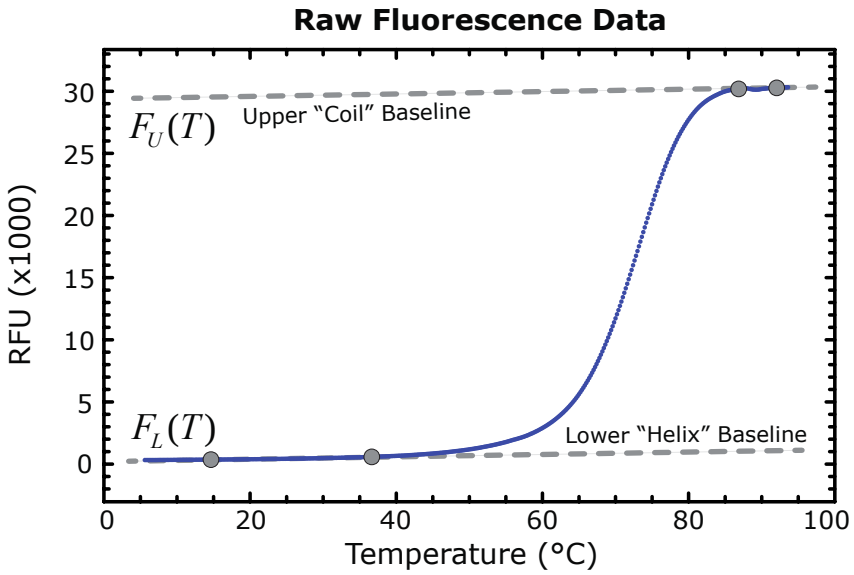



**Fig. 2.7) Raw Fluorescence DNA Thermal Melting Data**

The observed fluorescence intensity increases as function of temperature as the DNA double helix dissociates into the coil state. This is raw data corresponding to the H820-H822 duplex (table shown later) at 2 μM concentration, which is the second highest concentration studied. The apparent lack of fluorescence in the helix state is indicative of the efficient quenching of the Texas Red fluorophore by the Iowa Black RQ quencher.

where $F(T)$, $F_L(T)$, and $F_U(T)$ are the fluorescence values of the experimental curve, the lower baseline (helix), and the upper baseline (coil), respectively, at a given temperature $T$. The lower and upper baselines are determined by linear least-squares fitting of the measured fluoresence from regions before and after the melting transition, respectively, according to temperature points chosen by the experimenter[196]. There is admittedly some arbitrariness in baseline selection, and each investigator will have a slightly different opinion on which points to use. In general, baseline should be selected so that they are approximately parallel to each other by having slopes within 15%. To remove instrumental error from the observation, the raw fluorescence data can be smoothed using a Savitsky-Golay filter[253], although with a well-performing real-time PCR machine this smoothing effect is nearly negligible. The corresponding Θ curve for the given H820-H822 fluorescence melting data is:

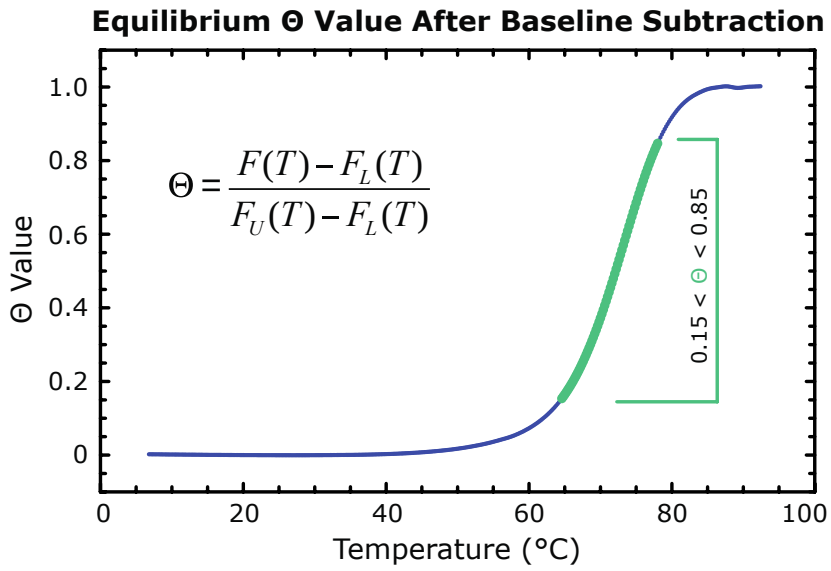



**Fig. 2.8) Converting the Raw Fluorescence Observable to an Equilibrium Term**

In this example the raw fluorescence melting data of H820-H822 DNA duplex is converted to an equilibrium term Θ using baseline subtraction for a two-state model, assuming that the fitted $F_L$ "lower" line corresponds to the intact duplex state, and that the fitted $F_U$ "upper" line corresponds to the dissociated coil state.

## DETERMINING THE MELTING TEMPERATURE: 1ST DERIVATIVE VS. Θ VALUE METHODS

The portion of the Θ curve colored green indicates the "transition" region in the helix-to-coil equilibrium, where Θ falls between 0.15 and 0.85. This region is important because it is the transition region from which thermodynamic information is extracted. The melting temperature can be determined using two different methods, and the resulting $T_m$'s normally differ by 1-2°C. Deviations greater than this indicate non-two-state melting behavior. The first method for determining melting temperature is to use the numerical first derivative of the Θ curve, with respect to temperature[242,243]:

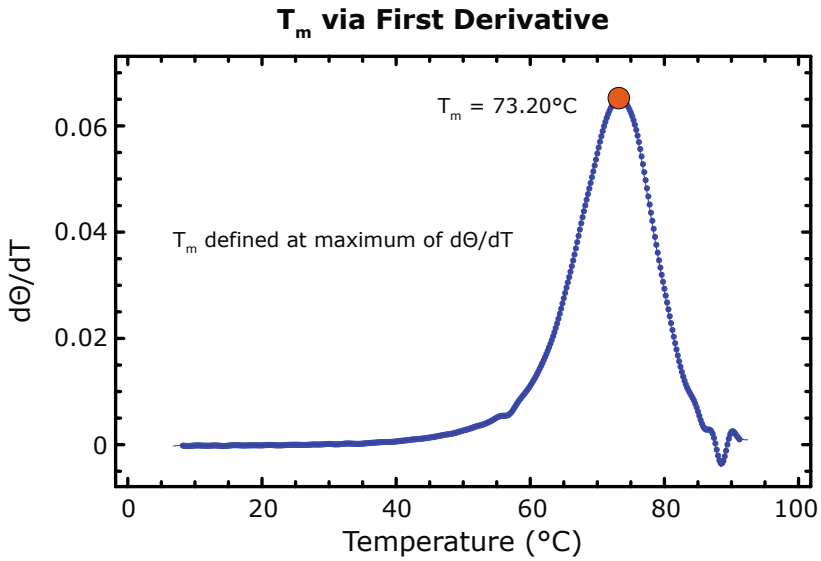



**Fig. 2.9) Determining Melting Temperature via the Derivative Method**

The first derivative of the Θ value with respect to temperature is the first way to determine the melting temperature. In this case of the 2 uM H820-H822 melting profile the melting temperature via 1st derivative is 73.20°C. This technique can also be used on the differential of the raw observable (*e.g.* absorbance vs. temp, or fluorescence vs. temp), which Is a useful calculation for software built in to thermal melting instruments.

Using this method the melting temperature of the duplex is taken as the transition, or inflection point, of the Θ curve. This corresponds to the temperature at which the first derivative of the Θ curve is at its maximum, resulting in a melting temperature of 73.20°C. The advantage of this approach is that it can be calculated *in silico* without the need for baseline selection, if need be. Most built-in thermal melting software supplied with instruments is equipped with some form of this analysis. The second method is to define the melting point as the temperature at which $\Theta = 0.5$. At this point, the fraction of single strands is equally distributed between the helix and coil states[242,243]:

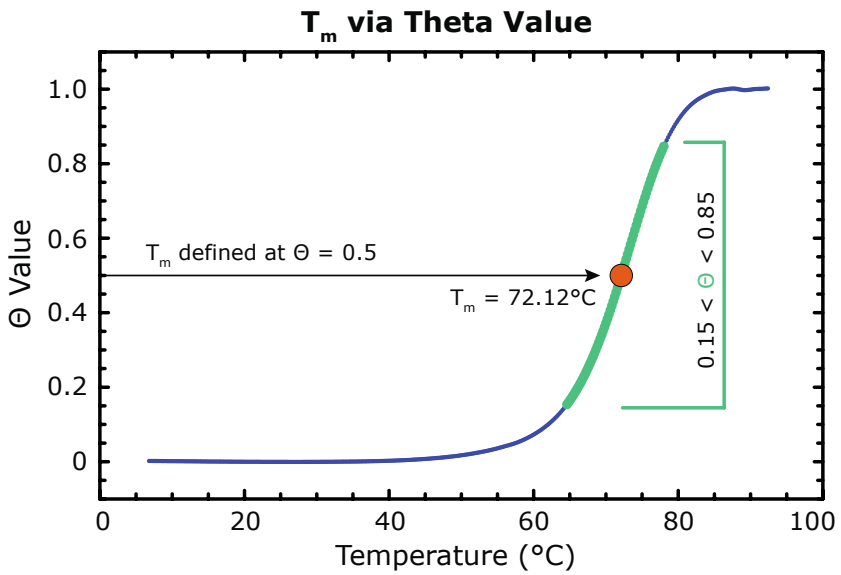



**Fig. 2.10) Determining Melting Temperature via the Theta Value**

The second method for acquiring the melting temperature is to determine where the Θ value is equal to 0.5, which is exactly half way in between duplex and coil states. In this case of the 2 uM H820-H822 melting profile the melting temperature via Θ value is 72.12°C.

In this example of H820-H822, the $\theta = 0.5$ melting temperature is 72.12°C. It is normal for the two $T_m$ determination methods to have results that differ by 1-2°C. The decision on which method is ideal depends on the specific application in regards to transition behavior and cooperativity, and typically during in-depth analysis the two values are compared. In high-resolution melting analysis, the $\theta = 0.5$ baseline method has an advantage of being less susceptible to deviations from non-ideal melting curve shapes. Later it is shown that each approach has its own method for extracting thermodynamic parameters.

## DNA Duplex Hybridization: Concentration and the Equilibrium Constant

Provided in the next few sections is an abbreviated version of the mathematical explanation for extracting thermodynamic information from melting transitions of non-self-complementary DNA duplexes. These techniques are covered in greater depth inside the manuscripts of Luis Marky and Kenneth Breslauer[242,243], which elaborate in detail situations where molecularity differs from that of two-strand DNA duplexes, and their texts are recommended for further insight. Another useful manuscript is out of Albert Benight's laboratory, the place

from which a collaborator on this project, Richard Owczarzy, completed his dissertation[238,239].

The equilibrium $\theta$ term used for the thermal melts in this dissertation describes dissociated DNA strands, meaning that the $\theta$ value is zero when the two single strands are associated as a duplex and $\theta$ increases to a value of one when the sample heats and dissociates into single strands.

$\theta = 0,$ none of the duplexes are melted, the fraction is zero

$\theta = 1,$ all of the duplexes are melted, the fraction is one

This $\theta$ definition is explicitly described here because the $\theta$ nomenclature differs from the classic texts of Marky and Breslaeur[242,243], which instead use an alpha $\alpha$ term that describes associated DNA strands. If one is comparing the equations derived in this dissertation with the references mentioned, recognize that $\theta = 1 - \alpha$, and $\alpha = 1 - \theta$. To add more confusion, the manuscripts of the Benight lab use $\theta_{\text{ext}}$, but the term is applied in a manner analogous to Marky and Breslauer's $\alpha$. To reemphasize, in the context of this dissertation, the explicit definition of $\theta$ as a measure of melted duplexes is utilized. Later manuscripts by Owczarzry[254] tend to use $\theta$ in a similar manner as to this dissertation.

In the case of a bi-molecular DNA duplex, the molecularity $n = 2$, and it is assumed that the concentration of both strands is equimolar, $[S_1] = [S_2]$. The concentration of each strand mixed during the experiment setup is, by definition, $C_{\text{T}}/n$. The total concentration of the DNA strands $C_{\text{T}}$ in the sample doesn't change as the experiment progresses, but its constituents do. The $C_{\text{T}}$ at all times is the sum of the duplex concentration $[D]$ and single strands, that is:

$$C_{\text{T}} = [S_1] + [S_2] + 2[D] \tag{2.101}$$

This establishes the assumption that the process is a two-state all-or-none transition with no population of intermediate states. The two strands are independent, $S_1 \neq S_2$, and non-self-complementary. The observed $\theta$ at any temperature within the hybridization experiment is simply a representation of the relative concentration of each species:

$$\theta_{\text{obs}} = \frac{[S_1] + [S_2]}{2[D] + [S_1] + [S_2]} = \frac{[S_1] + [S_2]}{C_{\text{T}}} \tag{2.102}$$

Hybridization of complementary DNA strands is a temperature dependent equilibrium process:

$$S_1 + S_2 \leftrightarrow D \tag{2.103}$$

The equilibrium constant for the DNA hybridization association reaction:

$$S_1 + S_2 \rightarrow D \tag{2.104}$$

would be:

$$K_{\mathrm{a}} = \frac{[D]}{[S_1][S_2]} \tag{2.105}$$

and obviously the inverse equilibrium constant:

$$K_{\mathrm{d}} = \frac{[S_1][S_2]}{[D]} \tag{2.106}$$

would describe the dissociation reaction:

$$D \rightarrow S_1 + S_2. \tag{2.107}$$

The derivations presented here utilize $K_{\mathrm{a}}$ rather than $K_{\mathrm{d}}$ because traditionally the association constant for the annealing reaction is used in nucleic acid hybridization literature. A keen reader may find it counterintuitive to use a dissociated DNA term $\theta$ with an association constant (hence, why Marky and Breslauer used the association $\alpha$ term), but it's best to just accept that this combination works well when analyzing thermal melting curves. This is because when selecting regions for baseline subtraction it is easier to visually judge the fits if equilibrium melting is processed as a $\theta$ term, since the coil state has a smaller temperature window for fitting due to the high melting temperature from the 1M salt conditions. Additionally, the upper fitted line for $\theta$ is not occluded by the graph's x-axis, which occurs if one uses the $\alpha$ term. It should be mentioned that in actual hybridization experiments, data from both the forward melting curve as well as the reverse annealing curve are used, and in practice it is often easier to just re-sort the data points so that temperature increases from low to high values in all temperature ramps.

One can obtain an equilibrium constant in terms of the experimentally available terms $\theta$ and $C_{\mathrm{T}}$ by substituting the equilibrium constant $K_{\mathrm{a}}$ equation (2.105) with equations (2.101) and (2.102) under the assumption that $[S_1] = [S_2]$:

$$K_{\mathrm{a}} = \frac{2(1-\theta)}{\theta^2 C_{\mathrm{T}}} \tag{2.108}$$

At the melting temperature $\theta = 0.5$, the association constant at the melting temperature simplifies to

$$K_{\mathrm{a}}(\text{at } T_{\mathrm{m}}) = \frac{2(1-0.5)}{(0.5)^2 C_{\mathrm{T}}} = \frac{4}{C_{\mathrm{T}}} \tag{2.109}$$

### Extracting $\Delta H_{\mathrm{VH}}$ and $\Delta S_{\mathrm{VH}}$ from the Shape of the Equilibrium Melting Curve

A general expression for calculating the transition enthalpy is derived by substituting the equilibrium constant into the van't Hoff equation:

$$\Delta H^{\circ}_{VH} = RT^2 \left[\frac{\mathrm{d} \ln K_{\mathrm{a}}}{\mathrm{d}T}\right] \tag{2.110}$$

and differentiating and then solving for $\Delta H_{VH}$ at the Tm, where $\theta = 0.5$. In the case of a bimolecular DNA duplex with a molecularity $n$ of two, this simplifies to an equation for the van't Hoff transition enthalpy, where $\partial\theta/\partial T$ is at its maximum:

$$\Delta H^{\circ}_{VH} = -6RT_{\mathrm{m}}^2 \left(\frac{\partial \theta}{\partial T}\right)_{\max} \tag{2.111}$$

This technique for calculating van't Hoff enthalpy is useful because of the speed at which these calculations can be performed, since derivative methods are amenable to automation. Software that ships with thermal melting instrumentation uses this method. For more rigorous analyses, the van't Hoff enthalpy $\Delta H^{\circ}_{VH}$ can also be determined graphically through the fundamental thermodynamic relations $\Delta G° = \Delta H° - T\Delta S°$ and $\Delta G° = -RT \ln K_{\mathrm{a}}$, which are equivalent through $\Delta G°$:

$$-RT \ln K_{\mathrm{a}} = \Delta H^{\circ}_{VH} - T\Delta S^{\circ}_{VH} \tag{2.112}$$

Solving for $\ln K_{\mathrm{a}}$, a linear equation becomes apparent and a second method for calculating van't Hoff enthalpy follows as:

$$\ln K_{\mathrm{a}} = -\frac{\Delta H^{\circ}_{VH}}{RT} + \frac{\Delta S^{\circ}_{VH}}{R} \tag{2.113}$$

In this linear equation of the form $y = mx + b$, the dependent variable $y$ is $\ln K_{\mathrm{a}}$, the slope $m$ is $-\Delta H^{\circ}_{VH}/R$, the dependent variable $x$ is $(1/T)$, and the intercept $b$ is $\Delta S^{\circ}_{VH}/R$. It becomes apparent that a least-squares linear fit of a plot of $\ln K_{\mathrm{a}}$ versus $(1/T)$ reveals the van't Hoff enthalpy and entropy for the duplex transition.

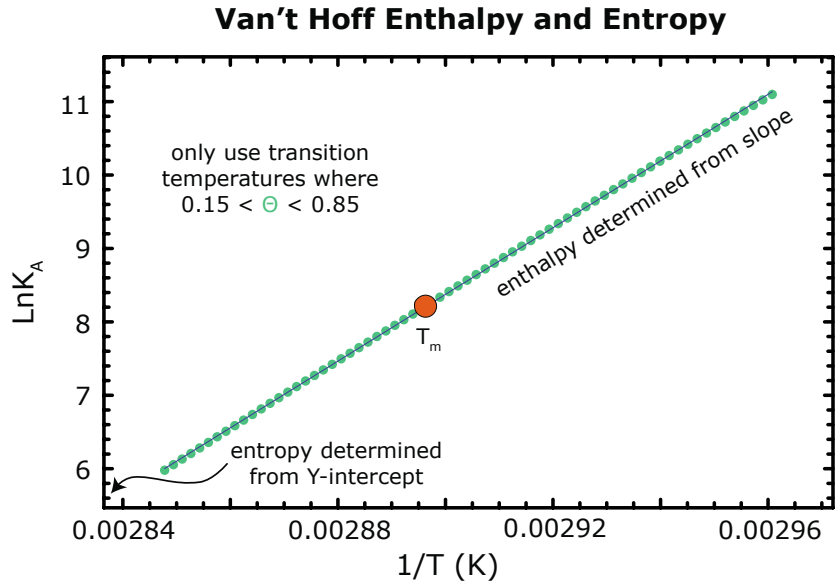



**Fig. 2.11) Extracting Thermodynamic Values from the van't Hoff Plot**

Thermodynamic values can be extracted from a plot of the equilibrium constant vs. temperature (van't Hoff plot), over the range of temperatures that corresponds to the region where $0.15 < \theta < 0.85$. From the linear fit, the slope yields $-\Delta H^{\circ}_{VH}/R$ and the y-intercept yields $\Delta S^{\circ}_{VH}/R$.

The two methods for calculating van't Hoff enthalpy and entropy will obviously yield different values, and each lab has their own preference. If an experimenter determines melting temperature of a bimolecular duplex using the maximum first derivative approach, then the $\Delta H^{\circ}_{VH} = -6RT_{\mathrm{m}}^2 \left(\frac{\partial\theta}{\partial T}\right)_{\max}$ method is an obvious choice since the maximum $\partial\theta/\partial T$ value has already been determined. An advantage of this approach when working with melting curve data in spreadsheet format is that it can avoid the need for graphical interpretation as calculations can be completed *in silico* quickly, without review. Nevertheless, there is an advantage in using the graphical approach with $T_{\mathrm{m}}$ determination by $\theta = 0.5$ since deviation from a linear $\ln K_{\mathrm{a}}$ versus $(1/T)$ plot may indicate problems with the shape of the melting curve around the transition temperature. Especially in cases of non-ideal melting curves, the maximum $\partial\theta/\partial T$ value is a less reliable melting temperature determination method since it is greatly affected by the curve shape. Additionally, the maximum $\partial\theta/\partial T$ $T_{\mathrm{m}}$ method suffers from resolution issues, since the maximum $\partial\theta/\partial T$ temperature is only as accurate as the interval between measurements (*i.e.* temperature point spacing). As stated earlier, it is beneficial to compare both

melting temperature determination techniques. The $\theta = 0.5$ approach by design implements interpolation that can solve the melting temperature at fractional degrees between data point measurements. In high-resolution melting experiments it is the preferred choice and was implemented as the approach of this dissertation (although both methods were calculated and compared). There has been an effort recently to harmonize previously published melting data with current and future approaches through corrective terms[254].

Just as there are two methods to determine van't Hoff enthalpy, there is also a second method to determine van't Hoff entropy using the maximum $\partial\theta/\partial T$ $T_{\mathrm{m}}$ method through van't Hoff enthalpy:

$$\Delta S^{\circ}_{VH} = \frac{\Delta H^{\circ}_{VH}}{T_{\mathrm{m}}} - R \ln\left(\frac{C_{\mathrm{t}}}{4}\right) \tag{2.114}$$

**Table 2.11) Two Approaches to Determine van't Hoff Enthalpy and Entropy from the Shape of Melting Curves**

| | **Value from Max of $\partial\theta/\partial T$** | **$\ln K_{\mathrm{a}}$ vs. $1/T$ fit [$y = mx + b$]** |
|---|---|---|
| **Melting Temp.** | $T_{\mathrm{m}}$ @ Max $\partial\theta/\partial T$ | $T_{\mathrm{m}}$ @ $\theta = 0.5$ |
| **Enthalpy** | $\Delta H^{\circ}_{VH} = -(2+2n)RT_{\mathrm{m}}^2\left(\frac{\partial\theta}{\partial T}\right)_{\max}$ | From slope ($m$) of $\ln K_{\mathrm{a}} = \left(\frac{-\Delta H^{\circ}_{VH}}{R}\right)\left(\frac{1}{T}\right) + \frac{\Delta S^{\circ}_{VH}}{R}$ |
| **Entropy** | $\Delta S^{\circ}_{VH} = \frac{\Delta H^{\circ}_{VH}}{T_{\mathrm{m}}} - R\ln\left(\frac{C_{\mathrm{T}}}{4}\right)$ | From intercept ($b$) of $\ln K_{\mathrm{a}} = \left(\frac{-\Delta H^{\circ}_{VH}}{R}\right)\left(\frac{1}{T}\right) + \frac{\Delta S^{\circ}_{VH}}{R}$ |

## EXTRACTING ENTHALPY ($\Delta H^{\circ}$) AND ENTROPY ($\Delta S^{\circ}$) FROM THE CONCENTRATION DEPENDENCE OF MELTING TEMPERATURE

The formation of DNA duplexes and any other complex of molecularity $n$ greater than one will result in a concentration dependent equilibrium melting temperature[242]. To understand the fundamentals behind this concept the $K_{\mathrm{a}}$ equation at the melting temperature (Eq. 2.109) can be plugged directly into the relation $-RT\ln K_{\mathrm{a}} = \Delta H^{\circ} - T\Delta S^{\circ}$:

$$RT_{\mathrm{m}}\ln\left(\frac{C_{\mathrm{T}}}{4}\right) = \Delta H^{\circ} - T_{\mathrm{m}}\Delta S^{\circ} \tag{2.115}$$

The expression can be rearranged into the familiar form:

$$\frac{1}{T_{\mathrm{m}}} = \frac{R}{\Delta H^{\circ}} \ln\left(\frac{C_{\mathrm{T}}}{4}\right) + \frac{\Delta S^{\circ}}{\Delta H^{\circ}} \tag{2.116}$$

A linear equation of the form $y = mx + b$ becomes apparent. In this equation, the dependent variable $y$ is $1/T_{\mathrm{m}}$, the slope $m$ is $R/\Delta H^{\circ}$, the independent variable $x$ is $\ln(C_{\mathrm{T}}/4)$, and the intercept $b$ is $\Delta S^{\circ}/\Delta H^{\circ}$. A plot of $1/T_{\mathrm{m}}$ versus $\ln(C_{\mathrm{T}}/4)$ should yield a straight line that can be linearly fit through least-squares methods. The slope of such a plot is the transition enthalpy the intercept yields the entropy.

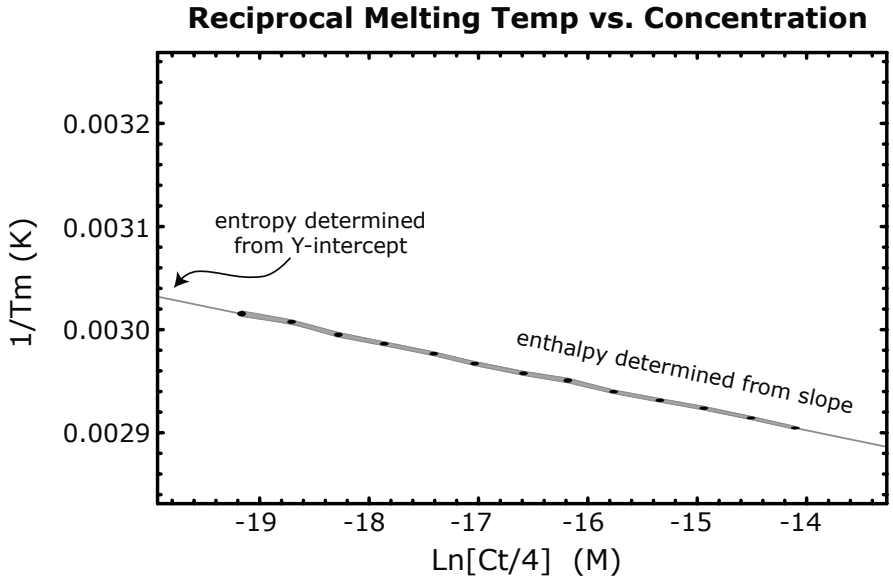



**Fig. 2.12) Extracting Thermodynamic Values from the Concentration Dependence of Melting Temperature**

Thermodynamic values can be extracted from a plot of the reciprocal melting temperature vs. total strand concentration. In this example the melting temperatures from 13 concentrations over a range of 19 nM – 3 μM are used to generate the graph. From the linear fit using Equation 2.116, the slope yields $R/\Delta H^{\circ}$ and the y-intercept yields $\Delta S^{\circ}/\Delta H^{\circ}$.

## 2.4) Methods: Experimental Design for Phosphorodithioated DNA Thermodynamic Studies

### INTRODUCTION

About a dozen studies reported in the literature have probed the thermal stability of oligonucleotides containing non-bridging (mostly mono-) thioated phosphates in comparison with unmodified duplexes[171,255-263]. Some other manuscripts worked towards comparing $R_p$ and $S_p$-stereospecific enantiomers of monothioates without direct comparison to unmodified duplexes[145,264], a couple more that focused on chemistry strategies for backbone thioation also reported thermal melting profiles[144,265], and one paper studied thioation of an RNA hairpin[266]. There was great interest in the late 1980s into 1990s surrounding sulfur-modified-DNA/RNA hybrids because of the potential for using thioated DNA as a therapeutic drug, because thioated DNA features enhanced nuclease resistance while retaining similar charge characteristics and hybrid oligo structure[267]. The majority of these mentioned reports studied monothioation of the phosphate backbone which is, as discussed in Chapter 1, a simpler mechanism than incorporating a dithioate substitution into the phosphate. Heteroatom substitution of the phosphate creates a chiral center about the phosphorus (Fig. 1.14), therefore monothioation results in a diastereomeric mixture of $R_p$ and $S_p$ forms with a diversity that scales exponentially as $2^n$, where $n$ is the number of phosphates. Since enantiomeric mixtures are quickly rejected by the US Federal Drug Administration, any therapeutic containing thioated phosphates must either A) be created using stereo-specific monothioation chemistry, or B) utilize achiral dithioated phosphates instead of monothioates.

With the objective of creating phosphorthioated oligonucleotide therapeutics, information on the thermodynamic behavior of modified hybrids becomes very important, especially for applications involving aptamer folding or hybridization with cDNA or mRNA. An abbreviated table summarizing literature on thioated melting studies is presented below (Table 2.12). The table focuses solely on melting temperature differences since the majority of these studies did not

explicitly elucidate enthalpic and entropic contributions. In analyzing these studies a few trends become apparent:

- Thioation (both PS and PS2) decreases the stability of an oligonucleotide duplex compared to its unmodified counterpart.
- PS2 substitutions are more destabilizing than PS substitutions.
- Melting temperature changes (both PS and PS2) appear sequence-dependent (the sequences are not listed in the table, but most of the references comment on this effect and should be sought for further details).
- $S_p$-PS-DNA·PO-DNA hybrids are more stable than $R_p$-PS-DNA·PO-DNA hybrids.
- $R_p$-PS-DNA·PO-RNA hybrids are typically more stable than $S_p$-PS-DNA·PO-RNA hybrids[262,264], although exceptions exist [259].

Despite the demonstrated energetic differences between normal duplexes and phosphorothioated duplexes, there has yet to be a comprehensive thermodynamic study that analyzes in depth the sequence effects of either monothiophosphates or dithiophosphates. The studies so far all have varied sequence length, inconsistence thioation patterns, frequently deviate from true two-state behavior, and have not efficiently represented all 10 nearest-neighbor pair phosphorothioates. The work presented in this dissertation aims to address this lack of thermodynamic information, and focuses exclusively on short phosphorodithioates DNA duplexes that melt in a two-state manner. Dithioates were chosen over monothioates because they allow a simpler, geometrically symmetrical interpretation of results, whereas monothioates have the sulfur oriented towards the major groove ($R_p$) or minor groove ($S_p$). By minimizing structural perturbations, the uniform dithio substitution potentially enables the generalization of observed trends to other sulfur-for-oxygen substitutions.

**Table 2.12) Literature Reports of Phosphorothioates Hybridization Studies**

| Reference | Duplex Hybrid Type | Notes | ΔTm, per thioation site |
|---|---|---|---|
| **LaPlanche et al., 1986** | PO-DNA·PO-DNA<br>PS-DNA·PS-DNA | • 1 duplex sequence<br>• 8-mer self-complementary<br>• 1 thioation site per strand<br>• 2 thioation sites per duplex<br>• Studied $R_p$ vs $S_p$ stability | PS($S_p$): ~0.0°C<br>PS($R_p$): ~1.0°C |
| **Stein et al., 1988** | PO-DNA·PO-DNA<br>PS-DNA·PO-DNA | • Numerous sequences<br>• Various lengths and complementarity<br>• 100% thioation | PS: ~0.8°C |
| **Bjergarde and Dahl, 1991** | PO-DNA·PO-DNA<br>PS-DNA·PO-DNA<br>**PS2-DNA**·PO-DNA<br>**PS2-DNA·PS2-DNA** | • 2 duplex sequences<br>• 17-mers non self-complementary<br>• 100% thioation | PS: ~0.7°C<br>**PS2: ~1.1°C** |
| **Ghosh et al., 1993** | PO-DNA·PO-DNA<br>PS-DNA·PO-DNA<br>**PS2-DNA**·PO-DNA | • 1 duplex sequence<br>• 17-mer non self-complementary<br>• 100% thioation | PS: ~0.7°C<br>**PS2: ~1.0°C** |
| **Cummins et al., 1996** | PO-DNA·PO-DNA<br>PS-DNA·PO-DNA<br>**PS2-DNA**·PO-DNA | • 2 duplex sequences<br>• 20-mer, 15-mer non self-complementary<br>• various % dithioation | PS: ~0.4°C<br>**PS2: ~0.9°C** |
| **Kanehara et al., 1996** | PO-DNA·PO-DNA<br>PO-DNA·RNA<br>PS-DNA·PO-DNA<br>PS-DNA·RNA | • 4 duplex sequences (8 total, DNA and RNA versions)<br>• 6-mers non self-complementary<br>• 1 or 2 (adjacent) thioation sites<br>• Stereospecific study ($R_p$ vs. $S_p$) | PS: ~2.6°C<br>• $S_p$-PS-DNA·DNA more stable than $R_p$-PS-DNA·DNA by 2.1°C<br>• $R_p$-PS-DNA·RNA and $S_p$-PS-DNA·RNA are about equivalent |
| **Clark et al., 1997** | RNA·RNA<br>PO-DNA·RNA<br>PS-DNA·RNA<br>PO-DNA·PO-DNA<br>PS-DNA·PO-DNA<br>PS-DNA·PS-DNA | • 1 duplex sequence<br>• 12-mer non self-complementary<br>• 100% thioation | PS: ~0.7°C |
| **Jaroszewski et al., 1999** | PO-DNA·PO-DNA<br>PS-DNA·PS-DNA<br>**PS2-DNA·PS2-DNA** | • 1 duplex sequence<br>• 17-mer self complementary<br>• 100% thioation | PS: ~0.9°C<br>**PS2: ~2.1°C** |
| **Kanaori et al., 1999** | RNA·RNA<br>PO-DNA·PO-DNA<br>PS-DNA·RNA<br>PS-DNA·PO-DNA | • 1 duplex sequence<br>• 6-mer self-complementary<br>• 2 adjacent thioation sites, $R_pR_p$- and $S_pS_p$-stereospecific<br>• $^1$H NMR melting | • $R_p$-PS-DNA·RNA (0.5°C) more stable than $S_p$-PS-DNA·RNA (2.5°C)<br>• $S_p$-PS-DNA·DNA (1.0°C) more stable than $R_p$-PS-DNA·DNA (2.0°C) |
| **Yu et al., 2000** | PS-DNA·RNA | • Various sequences and lengths<br>• Exclusively a stereospecific study ($R_p$ vs. $S_p$) rather than comparison with PO-DNA hybrids | $R_p$-PS·RNA more stable than $S_p$-PS·RNA |
| **Boczkowska et al., 2002** | PO-DNA·PO-DNA<br>PS-DNA·PO-DNA | • Various sequences (but only 5 summarized here)<br>• 100% thioation | PS: ~0.7°C<br>• $S_p$-PS-DNA·DNA (0.6°C) more stable than $R_p$-PS-DNA·DNA (0.8°C) |

## EXPERIMENTAL DESIGN RATIONALE

Currently dithioates are costly to produce, so there is great need to be able to predict the behavior of such molecules prior to synthesis so that their efficiency and performance is ensured for a study of interest. Numerous thermodynamic parameter sets have been developed empirically for the prediction of thermodynamic behavior for unmodified DNA duplexes (Table 2.8), with the most prevalent contemporary set being SantaLucia's[213] "unified parameters". The utility of these accurate predictions has spurred commercialization of oligo production houses that satiate the now immense demand for synthetic oligonucleotides. Areas of research such as PCR, molecular cloning, recombinant protein expression, siRNA knockdown, and targeted oligonucleotide aptamer design all critically rely upon well-designed nucleic acid constructs. Thermodynamic parameter sets are an invaluable component of such design as a lack of predictive knowledge would limit the serviceability of these constructs because investigators would waste significant manpower and financial resources if properties such as melting temperature had to be determined manually prior to each study. When parameter sets for PO-DNA duplexes were determined they benefited from large amounts of previously published melting data as well as a cooperative comparison with alternative parameter sets from other labs. Unfortunately this situation does not apply for phosphorodithioate nearest-neighbors since less than 20 dithioate sequences have published melting data, and this dissertation is the first comprehensive analysis of a dithioate nearest-neighbors.

The simplest way to achieve the ability to predict thermodynamic behavior of DNA dithioated phosphates was to develop a set of difference parameters, which is a correction factor for each NN doublet determined by comparing modified and unmodified duplexes. Values obtained from such analysis are $\Delta T_{\mathrm{m}}$, $\Delta\Delta G°$, $\Delta\Delta H°$, and $\Delta\Delta S°$. Using such an approach the thermodynamic properties can be calculated first using standard prediction parameters (e.g. SantaLucia) and then corrected using dithioate difference parameters.

Because of the enormity of such an experimental design, a collaboration was formed with Dr. Richard Owczarzy of the biotech firm Integrated DNA Technologies,

who demonstrated in 2011 the development of difference parameters for locked nucleic acid-modified (LNA) oligonucleotides[268]. A similar experimental system was developed that could determine thermodynamic difference parameters for all 10 dithioated NN doublets (5'→3'/5'→3': AA/TT, AC/GT, AG/CT, AT/AT, CA/TG, CC/GG, CG/CG, GA/TC, GC/GC, TA/TA). This type of analysis requires thousands of experimental melting curves, so the high-throughput FRET duplex hybridization method, which was also developed by Owczarzy[252,268], is the only realistic approach to accomplish such a task. In the dithioate studies presented here, complementary oligonucleotides sequences are 11-bases long and contain either a 5'-Texas Red fluorophore or a 3'-Iowa Black RQ quencher, which are located on the same side of duplex for maximum intensity quenching. Two copies of a specific NN doublet are contained within each strand. In the "top" strand, which is arbitrarily assigned to the fluorophore-containing oligonucleotide, the NN doublet is located four internucleotide linkages away from the 5'-end and another three linkages away from the 3'-end. The "bottom" sequence, assigned to the quenching strand, contains the NN doublet three linkages away from the 5' end and another four linkages away from the 3'-end. The spacing between the NN doublets is four internucleotide linkages apart. As a result of the phosphoramidite synthetic method the top strand does not contain a 3'-phosphate, but it gains an additional phosphate from the Texas Red amide coupling, for a total of 11 phosphates and -11 overall charge. The bottom strand gains a phosphate from the structure of the 3'-Iowa Black RQ CpG moiety and when combined with remaining phosphates a total of 11 phosphates and a -11 overall charge are present.

Normal phosphodiester NN doublets can be expressed with just one term (*e.g.* AG/TT), but dithioation of the top or bottom strand creates a unique NN pair that requires independent analysis and separate duplex constructs(*e.g.* A-ps2-G/TT and AG/T-ps2-T are different). An unmodified phosphodiester NN doublet-containing duplex was also prepared for use as a control, from which the ΔΔ parameters were calculated. Finally, an additional duplex containing both top and bottom strand dithioate was also studied to assess whether the effects of the modification were additive. Thus, four different duplexes were prepared for each

NN doublet resulting in a total of 40 duplexes employed in the analysis (10 NN doublets × 4 duplexes).

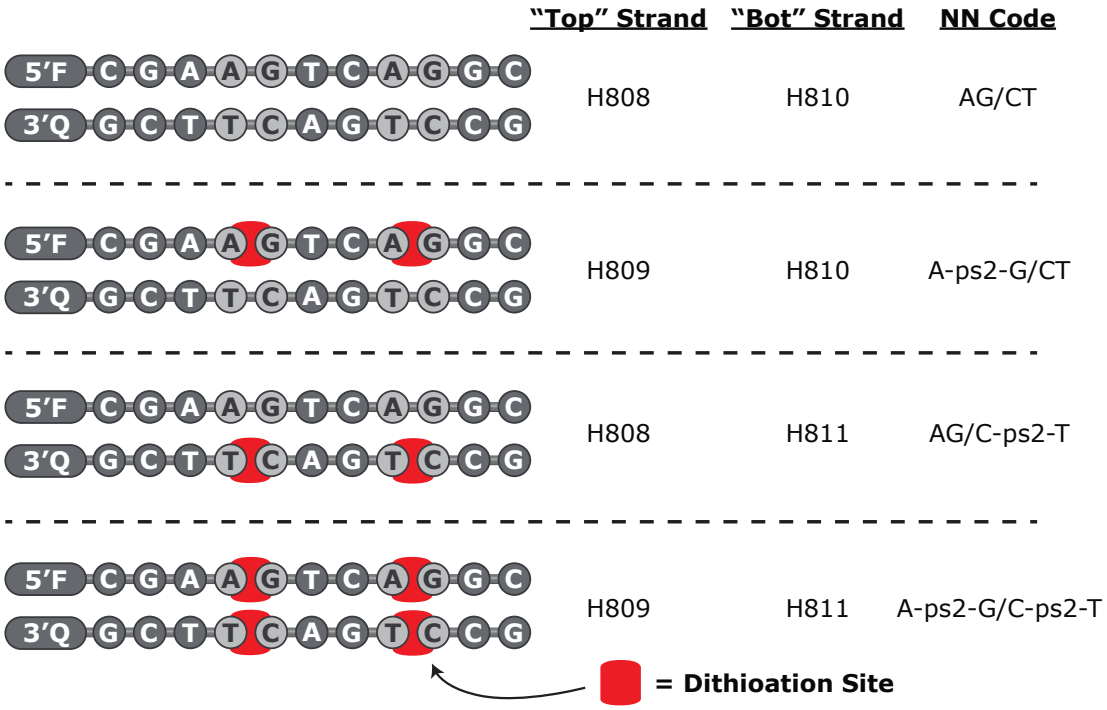



**Fig. 2.13) Four Duplexes Are Used to Study Each NN Doublet (*e.g.* AG/CT)**

Four DNA duplexes are used to probe each NN doublet. The first is an unmodified phosphodiester control (*e.g.* H808·H810). Because dithioation on the "top" or "bottom" strand is not symmetrical (*e.g.* A-ps2-G/TT ≠ AG/T-ps2-T), modification of the top NN doublet (*e.g.* H809·H810) and bottom NN doublet (*e.g.* H808·H811) require separate consideration. To test whether the effect of dithioate is additive, modifications in both top and bottom strands were also measured (*e.g.* H809·H811).

**Table 2.13) Experimental Design for the 40 Duplexes Used in the Analysis**

| Duplex Design | Top | Bot | NN Code |
|---|---|---|---|
|  | H800 | H802 | AA/TT |
| | H801 | H802 | A-ps2-A/TT |
| | H800 | H803 | AA/T-ps2-T |
| | H801 | H803 | A-ps2-A/T-ps2-T |
|  | H804 | H806 | AC/GT |
| | H805 | H806 | A-ps2-C/GT |
| | H804 | H807 | AC/G-ps2-T |
| | H805 | H807 | A-ps2-C/G-ps2-T |
|  | H808 | H810 | AG/CT |
| | H809 | H810 | A-ps2-G/CT |
| | H808 | H811 | AG/C-ps2-T |
| | H809 | H811 | A-ps2-G/C-ps2-T |
|  | H812 | H814 | AT/AT |
| | H813 | H814 | A-ps2-T/AT |
| | H812 | H815 | AT/A-ps2-T |
| | H813 | H815 | A-ps2-T/A-ps2-T |
|  | H816 | H818 | CA/TG |
| | H817 | H818 | C-ps2-A/TG |
| | H816 | H819 | CA/T-ps2-G |
| | H817 | H819 | C-ps2-A/T-ps2-G |
|  | H820 | H822 | CG/CG |
| | H821 | H822 | C-ps2-G/CG |
| | H820 | H823 | CG/C-ps2-G |
| | H821 | H823 | C-ps2-G/C-ps2-G |
|  | H824 | H826 | CC/GG |
| | H825 | H826 | C-ps2-C/GG |
| | H824 | H827 | CC/G-ps2-G |
| | H825 | H827 | C-ps2-C/G-ps2-G |
|  | H828 | H830 | GA/TC |
| | H829 | H830 | G-ps2-A/TC |
| | H828 | H831 | GA/T-ps2-C |
| | H829 | H831 | G-ps2-A/T-ps2-C |
|  | H832 | H834 | GC/GC |
| | H833 | H834 | G-ps2-C/GC |
| | H832 | H835 | GC/G-ps2-C |
| | H833 | H835 | G-ps2-C/G-ps2-C |
|  | H836 | H838 | TA/TA |
| | H837 | H838 | T-ps2-A/TA |
| | H836 | H839 | TA/T-ps2-A |
| | H837 | H839 | T-ps2-A/T-ps2-A |

## METHOD: REAL-TIME PCR DNA THERMAL MELTS

All 40 probe-modified oligonucleotides used in the analysis were synthesized using phosphoramidite chemistry and purified by RP-HPLC according to methods covered in Chapter 1. The production process, including quality control and resynthesis if necessary, required approximately eight months of work. Nucleic acid samples were at least 95% pure when assessed by capillary electrophoresis and the correct MW was verified using ESI-MS. Table 3 in Chapter 1 contains the predicted molecular weights and observed molecular weights of the purified oligos during quality control. Purified single strands were quantified using extinction coefficients at 260 nm[269], incorporating additional extinction coefficients for the Texas Red and Iowa Black RQ moieties of 14400 and 44510 $L \cdot mol^{-1} \cdot cm^{-1}$, respectively. Complementary strands were annealed by heating to 95°C and slowly cooled for ~30 minutes at room temperature, within the buffer for experimental melts so that dilution was simplified. This buffer consists of 1M NaCl, 3.87 mM $NaH_2PO_4$, 6.13 mM $Na_2HPO_4$, 1 mM $Na_2EDTA$ adjusted to pH 7.0 with NaOH. Macromolecule absorption onto the surface of microcentrifuge plastic tubes is well-known, and at concentrations below 300 nM this appears exacerbated by the addition of dithioate modifications and hydrophobic dyes and quenchers. To mitigate this effect, concentrated oligonucleotide stocks were prepared at the highest experimental concentration of 3 μM and the remaining samples were made by dilution from this stock. Low-binding microcentrifuge tubes such as Costar (Corning, Wilkes Barre, PA) are the preferred choice for minimizing potential sample loss[252]. For long-term storage, oligonucleotide duplex solutions are best kept in sealed O-ring tubes, which inhibits spontaneous water evaporation that occurs in snap-cap microcentrifuge tubes.

After QC of the oligos the process of final sample preparation and data collection was performed by technicians at Integrated DNA Technologies (Coralville, IA), which was a collaborator on the project and operates the real-time PCR instrument needed to complete the FRET analysis. Richard Owczarzy, a senior scientist at IDT, is the one who authored the manuscript detailing measuring DNA hybridization by high-throughput FRET analysis and he was consulted for

experiment design and analysis[252]. Aliquots of the 3 μM solution were diluted to produce 13 concentrations ($C_t$) of 19, 30, 46, 70, 110, 160, 250, 375, 570, 870 nM, and 1.3, 2.0, 3.0 μM. These values give uniformly separated data in the x-dimension on the $\ln C_t$ scale. Samples were loaded into 96-well Extreme Uniform thin walled plates (Cat#B70501, BIOplastics, Landgraaf, Netherlands) using volumes of 25 μL per well, and spun at 660 rcf for 2 min and then equilibrated at starting temperature (5°C) for 5 minutes. Each concentration was measured on its own individual plate, which is recommended to avoid issues with automatic gain detection that skews fluorescence readings if multiple concentrations are on the same plate. Two independent wells for each sample were on a plate, creating duplicate measurements. Fluorescence intensity was measured on an iQ5 real-time PCR system (Bio-Rad Laboratories, CA) that had a tungsten-halogen lamp source and filters specific for the Texas Red fluorophore. The machine was operating Bio-Rad iQ5 Optical System Software (version 2.0). Temperature was ramped from 5°C to 90°C at a rate of 25°C $h^{-1}$. Fluorescence intensity was recorded every 0.2°C and the values from the two duplicate wells were averaged. The software subtracted background fluorescence of plate wells automatically. A total of four temperature ramps were used for each sample: the first forward melt and reverse cooling profile, and then a second forward melt and reverse cooling profile. The melting profiles were completely reversible and thermodynamic information was extracted from each curve and then averaged.

### METHOD: THERMAL MELT ANALYSIS

The fluorescence measurements took four months to complete and produced a massive quantity of data that required over six months to process, and even longer to fully analyze. Since forty separate duplexes were studied and there were two replicates and four melting curves, a total of 4160 melting profiles were analyzed. With the temperature spacing for fluorescent reads, the total experiment for the forty duplexes contained 1,842,880 fluorescence intensity vs. temperature data points. Data was analyzed using a complex Visual Basic for Applications software operating in Microsoft Excel, which was authored iteratively over several years by

Richard Owczarzy of IDT. The software quickened the pace of the calculations covered in great detail inside the Extracting Thermodynamic Parameters section of this dissertation. Upper and lower baselines were selected and baseline subtraction yielded the equilibrium $\theta$ term. If duplex melting proceeds in a true two-state (all-or-none) manner, then partially melted duplexes are negligible throughout the transition and $\theta$ reflects the fraction of melted duplexes. The melting temperature was determined as the temperature at which $\theta = 0.5$ and this value was compared to the derivative melting temperature as initial evidence for two-state behavior ($T_{\mathrm{m}}$'s within 2°C). For statistical error measurements, a fixed standard deviation of the $T_{\mathrm{m}}$ equaling 0.4°C was used, which was an assumption supported by the data.

Thermodynamic values of $\Delta H°$, $\Delta S°$, and $\Delta G°$ were determined using two methods: van't Hoff melting curve analysis, and reciprocal melting temperature vs. concentration plots. In the van't Hoff method, equilibrium constants were least-squares fitted to the van't Hoff relationship,

$$-\ln K_{\mathrm{a}} = \frac{\Delta H°}{RT} - \frac{\Delta S°}{R} \tag{2.117}$$

where $R$ is the ideal gas constant. The $\Delta H°$ and $\Delta S°$ values were estimated from the slopes and intercepts of a fitted linear regression of $\ln K_{\mathrm{a}}$ versus $1/T$ plots, using the range of $\theta$ values between 0.15 and 0.85. The free energy $\Delta G°$ was calculated from the thermodynamic relation $\Delta G° = \Delta H° - T\Delta S°$. The $\Delta H°$, $\Delta S°$, and $\Delta G°$ at each concentration were averaged over all thirteen concentrations to arrive at the final thermodynamic values for each duplex. In the second method, the values were determined from the dependence of melting temperature on duplex concentration. The reciprocal melting temperatures were plotted against $\ln C_t/4$ and fitted to linear relationship:

$$\frac{1}{T_{\mathrm{m}}} = \frac{R}{\Delta H^{\circ}} \ln\left(\frac{C_{\mathrm{T}}}{4}\right) + \frac{\Delta S^{\circ}}{\Delta H^{\circ}} \tag{2.116}$$

The values from each method were compared and averaged to produce final thermodynamic values. If the results vary significantly, then the assumption of the two-state (all-or-none) behavior of duplex melting is likely invalid[270]. In these studies, duplexes were flagged as non-two-state if the values deviated by more than 15%. Difference thermodynamic parameters for modifying nearest-neighbor pairs

by dithioation were determined by subtracting the unmodified duplex control from the dithioated values.

## 2.5) Results: Dithioated Phosphate Melting Data

### RESULT: PERFORMANCE OF THE ASSAYS

Overall, the massive real-time PCR DNA hybridization experiment incorporating 4160 melting profiles had excellent performance, and produced high quality transition profiles that could be used for the extraction of thermodynamic values for the duplexes. The capability of the high-throughput FRET approach was verified, enabling the analysis such a large dataset that would have been intractable using traditional UV absorbance methods. Fluorescence profiles for the melting (forward) and annealing (reverse) reactions demonstrated excellent reversibility, having reproducible melting temperatures with an error of ~0.3°C for the $\theta$ method and ~0.4°C for the $\partial\theta/\partial T$ method. A total of 8 melting profiles (replicate A: 1F, 1R, 2F, 2R; replicate B: 1F, 1R, 2F, 2R) were used for each duplex to mitigate these effects, for each of the 13 concentrations. The thermodynamic values ($T_{\mathrm{m}}$, $\Delta G°$, $\Delta H°$, $\Delta S°$) of the 8 profiles were compared side-by-side, which enabled easy identification of outlier ramps (>15% deviation from the mean values). The melting curves of those significantly deviating from the mean were scrutinized for an explanation and the formation of air bubbles was by far the most common explanation. The technique uses small sample volumes of 25 μL and as a consequence is susceptible to sample evaporation if the plate seal has imperfections, as well as fluorescence disruption from the formation of air bubbles at higher temperatures. These bubbles result in a temporary jump up or jump down in fluorescence intensity that corrects itself after ~10°C. On the other hand, in the case of evaporation, a permanent increase in fluorescence is seen, which is easy to identify. Regardless, these problems were rare and <2% of data was excluded from further calculations.

Two methods were used to extract thermodynamic values: the melting curve van't Hoff analysis and the $T_m^{-1}$ vs. Ln[Ct/4] method. With the melting curve approach, thermodynamic values were calculated for each profile and then averaged over all 8 replicates, for each concentration. The averages for each concentration were then averaged together to produce a final value for the duplex. The thermodynamic calculations using this method are affected by the shape of the $\ln K_a$ vs. $1/T$ curve over the transition temperature window where $0.15 < \theta < 0.85$. Concentrations at the low- and high-end of the concentration window, specifically 19 nM, 30 nM, and 3 μM, tend to have non-ideal $\ln K_a$ vs. $1/T$ curve shapes and were omitted from the melting curve averaging so as not to skew data, which is the common practice[252,268]. The second method utilizes just the melting temperatures and concentration, and is less affected by non-ideal melting curve shapes. Therefore, data from all concentrations were used in the $T_m^{-1}$ vs. Ln[Ct/4] approach, with each linear fit incorporating $13 \times 8 = 104$ experiments.

**Table 2.14) Two Methods for Extracting Thermodynamic Values**

| Method | Melting Curve Fit | $T_m^{-1}$ vs. Ln[Ct/4] |
|---|---|---|
| **Ramps** | 8 total ramps are used for each concentration<br>Replicate A: 1F, 1R, 2F, 2R<br>Replicate B: 1F, 1R, 2F, 2R | |
| **Approach** | Each ramp is fit individually to extract $\Delta G°$, $\Delta H°$, and $\Delta S°$ using Eq. 2.117.<br><br>The $\Delta G°$, $\Delta H°$, and $\Delta S°$ values for the 8 ramps at one concentration are averaged. | $T_m$ is determined for each ramp using $\theta = 0.5$.<br><br>The $T_m$'s for the 8 ramps at each concentration are averaged. |
| **Final Values** | Average $\Delta G°$, $\Delta H°$, and $\Delta S°$ values from 10 of the 13 concentrations are averaged (46, 70, 110, 160, 250, 370, 570, 870 nM, 1.1, 2.0 μM). | Average $T_m$'s are plotted for all 13 concentrations (19, 30, 46, 70, 110, 160, 250, 370, 570, 870 nM, 1.1, 2.0, 3.0 μM) and then fit to Eq. 2.116. |
| **Outliers Present?** | Outliers in $\Delta G°$, $\Delta H°$, and $\Delta S°$ can occur at the 8-ramp averaging stage. Susceptible to curve-shape issues, which is why highest and lowest concentrations are omitted. | Outliers are rare since the s.d of $T_m$'s is very low. All data points for the duplex are fit in a single calculation. |
| **Reported Values** | Final values from both methods are compared, and averaged. If the $\Delta G°$ value deviates by more than 15% between the two methods, then the duplex has a non-two-state transition.<br><br>Reported values can be the average of the two methods, or the $T_m^{-1}$ vs. Ln[Ct/4] values, which are more reproducible between labs. In this dissertation, the average-value approach was implemented. | |

The two approaches for extracting thermodynamic values from DNA melting curves are summarized. Historically, melting curve shape-analysis dominates the literature because fewer data points are required. But this technique is more susceptible to non-ideal melting curve shapes from non-two-state transitions. The reciprocal melting temperature approach is more reproducible between labs because $T_m$ deviation is low, whereas the shape of the curve can fluctuate but much more data is required. High throughput real-time PCR fluorescence-based assays were utilized in this dissertation to develop a large enough dataset to compare both approaches. Significant deviation between the two methods was used an indicator of non-two-state melting behavior for the duplex.

## RESULT: CONCENTRATION DEPENDENCE OF MELTING TEMPERATURE AND RECIPROCAL MELTING TEMP VS. LN[CT/4] PLOTS FOR H800-H839 PS2 DUPLEXES

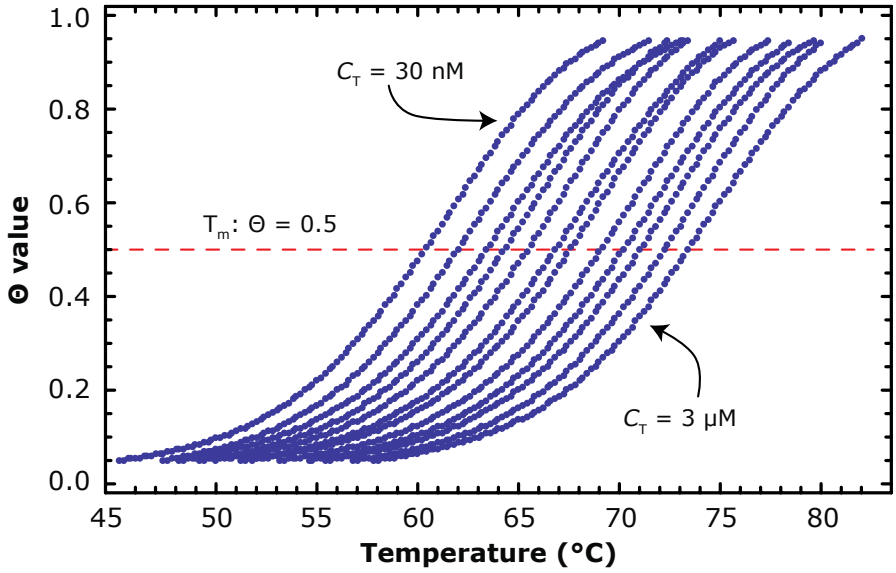



**Fig. 2.14) Concentration Dependence of Theta Curves**

The hybridization of DNA strands having molecularity $n > 1$ exhibits strong concentration dependence. Shown here are representative Θ curves for a 100-fold total strand concentration $C_T$ distribution observed for H820-H822 duplex from fluorescence measurements. The plotted range of $C_T$ is 30, 70, 110, 160, 250, 375, 570, 870 nM, and 1.1, 2.0, 3.0 μM, which are concentrations evenly spaced on the $\mathrm{Ln}[C_T/4]$ axis.

Oligonucleotide duplexes of $n = 2$ melted in a manner that was concentration dependent. This effect is illustrated in Fig. 2.14, which is a plot of representative $\theta$ curves vs. temperature for the H820-H822 duplex. All 40 duplexes in this study performed in a similar manner. Graphs of $T_m^{-1}$ vs. Ln[Ct/4] resulted in a straight line that could be linearly fit using equation 2.116. The standard deviation of $T_m$ measurements was approximately 0.3°C for all duplexes. As a test of outliers[268], values were removed if the $T_m^{-1}$ data point deviated from the fitted straight line by a value more than twice the value of the propagated error. Using this criterion only a single $T_m^{-1}$ data point required removal, which was from the H813-H814 duplex at 570 nM, and was attributed to sample evaporation. The $T_m^{-1}$ vs. Ln[Ct/4] plots (ovals) and fitted lines are shown in Figures 2.16-2.20. The plots are arranged by each NN grouping (*i.e.* AA/TT), with four duplexes plotted per graph (none, top, bottom, both). The height of each oval equals the thickness of the plotted colored band, which represents the standard deviation of the $T_m^{-1}$, each line encompasses 104 melting curves.

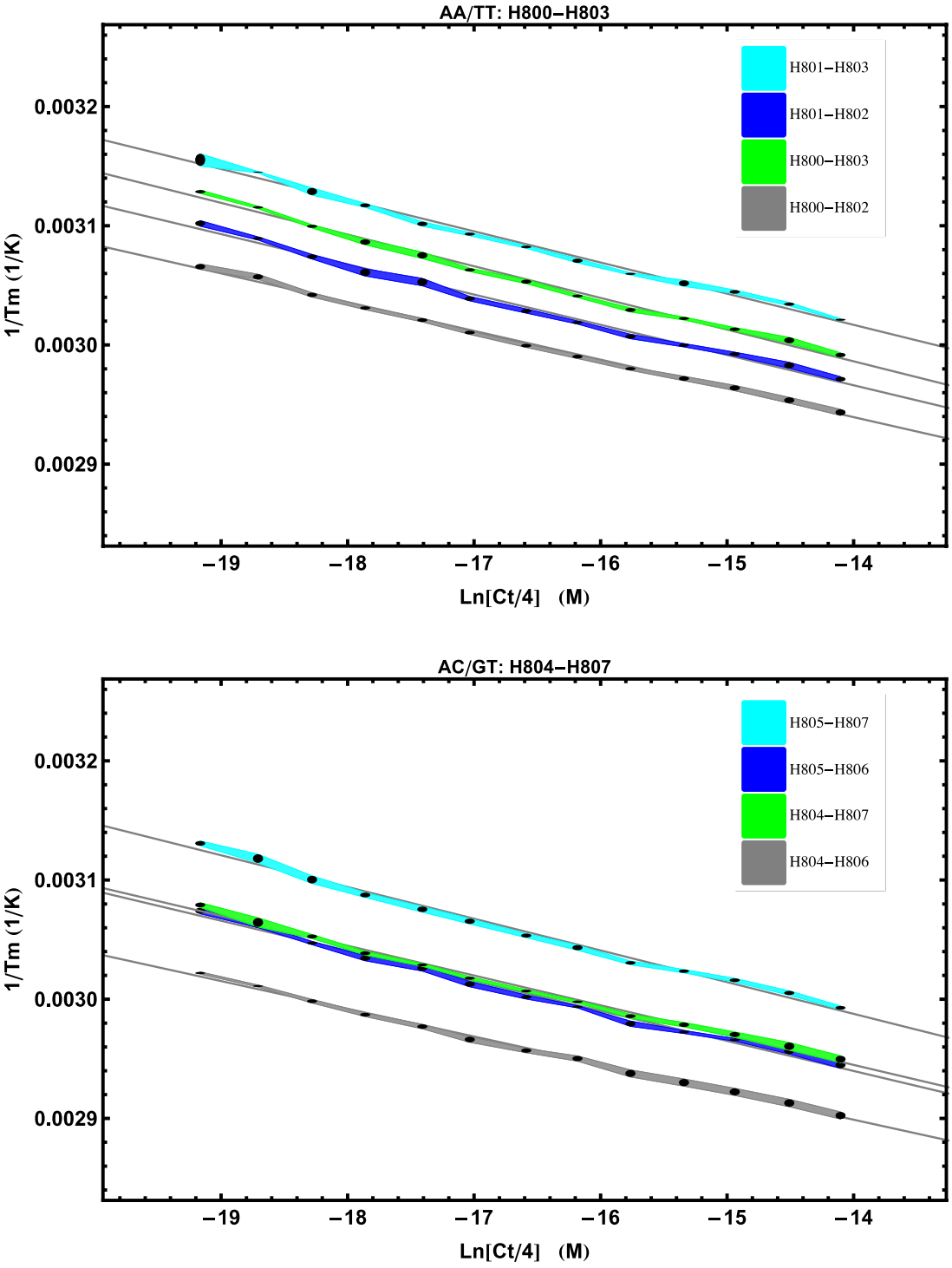



**Fig. 2.15) $T_m^{-1}$ vs. Ln[Ct/4] Plots for PS2 AA/TT and AC/GT Nearest Neighbor Doublets**

Melting temperatures are plotted as a function of duplex concentration (19 nM – 3 uM) using fixed-width black ovals. The height of each oval equals the thickness of the colored band, which represents the S.D. of eight replicates at that concentration. Each linear fit thus encompasses 104 separate Tm measurements (8 rep. x 13 conc.). Four duplexes for each NN were studied: the unmodified control (gray), the "top" dithioated strand (blue), the "bottom" dithioated strand (green), and the "both" dithioated version (cyan).

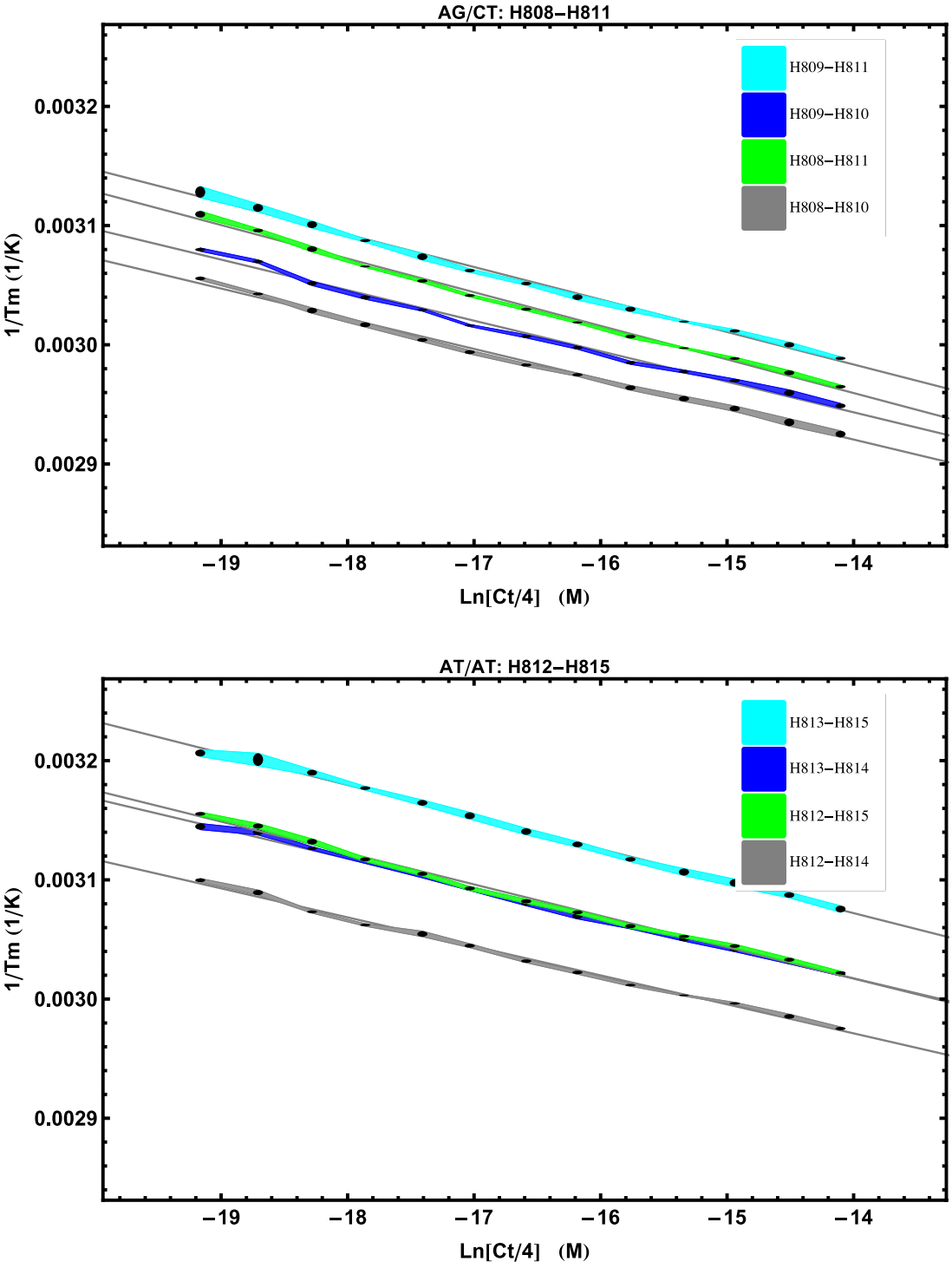



**Fig. 2.16) $T_m^{-1}$ vs. Ln[Ct/4] Plots for PS2 AG/CT and AT/AT Nearest Neighbor Doublets**

Melting temperatures are plotted as a function of duplex concentration (19 nM – 3 uM) using fixed-width black ovals. The height of each oval, as well as the color bar, represents the S.D. of eight replicates at that concentration. Each linear fit thus encompasses 104 separate Tm measurements (8 rep. x 13 conc.). Four duplexes for each NN were studied: the unmodified control (gray), the "top" dithioated strand (blue), the "bottom" dithioated strand (green), and the "both" dithioated version (cyan).

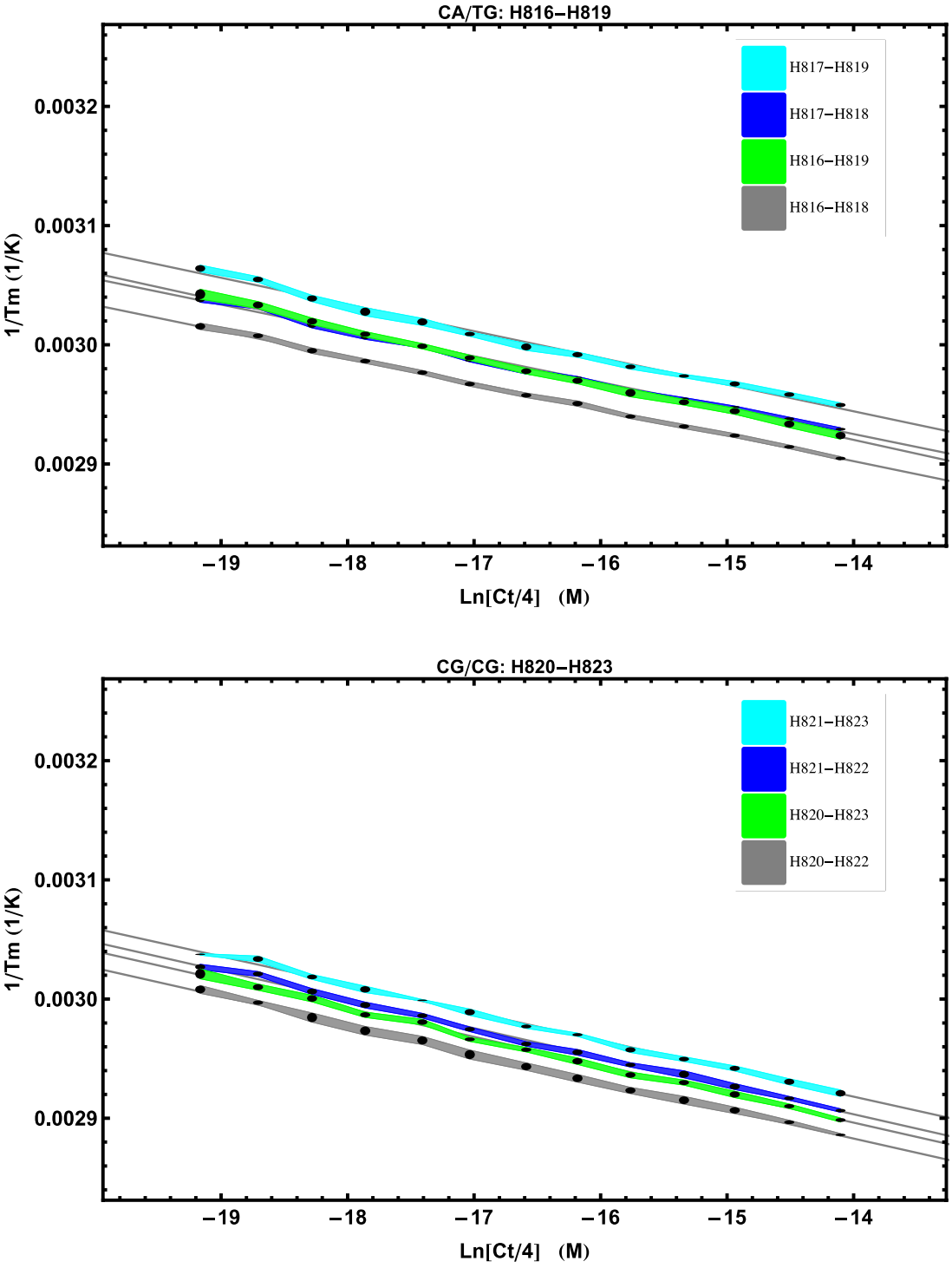



**Fig. 2.17) $T_m^{-1}$ vs. Ln[Ct/4] Plots for PS2 CA/TG and CG/CG Nearest Neighbor Doublets**

Melting temperatures are plotted as a function of duplex concentration (19 nM – 3 uM) using fixed-width black ovals. The height of each oval, as well as the color bar, represents the S.D. of eight replicates at that concentration. Each linear fit thus encompasses 104 separate Tm measurements (8 rep. x 13 conc.). Four duplexes for each NN were studied: the unmodified control (gray), the "top" dithioated strand (blue), the "bottom" dithioated strand (green), and the "both" dithioated version (cyan).

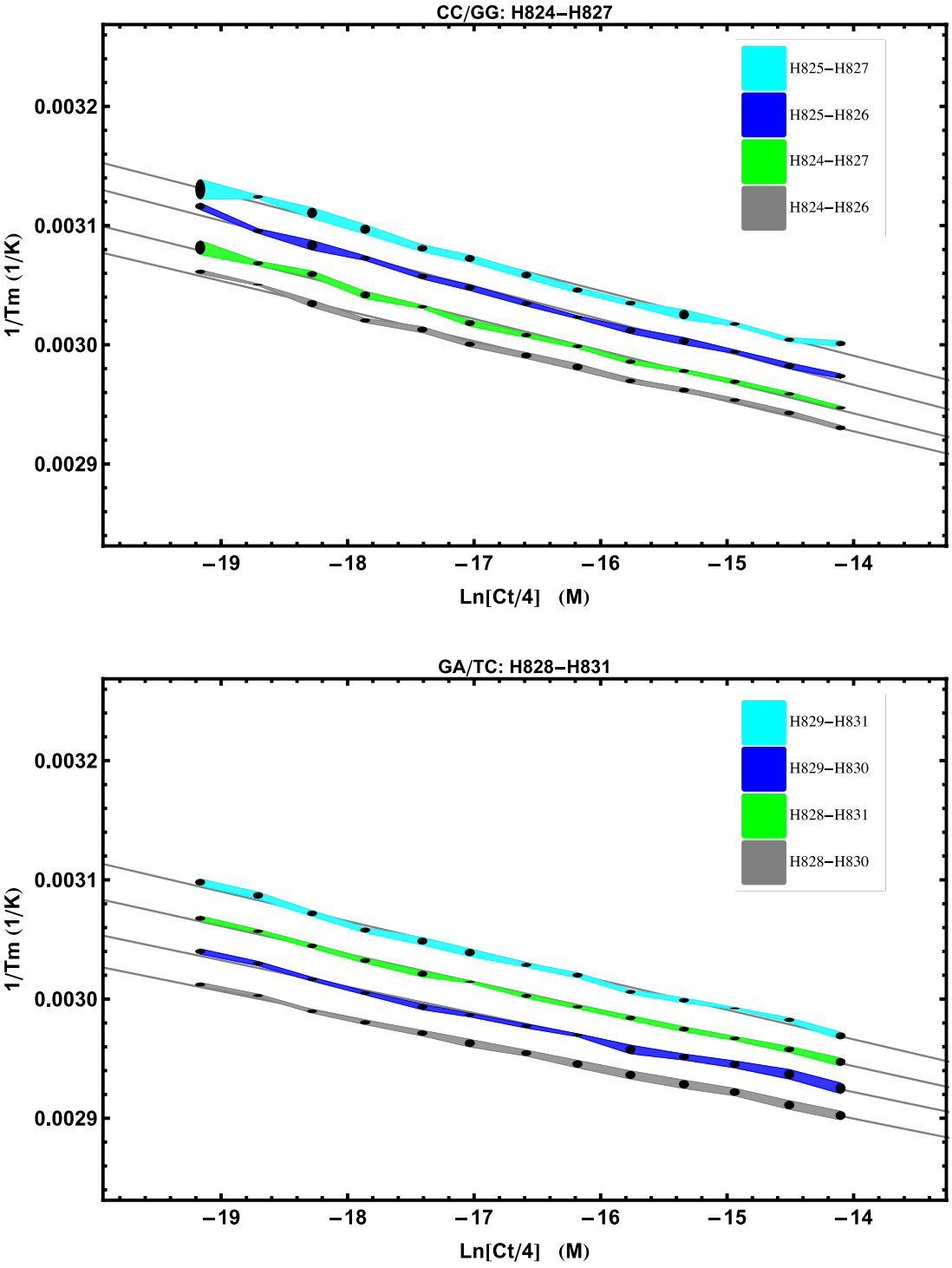



**Fig. 2.18) $T_m^{-1}$ vs. Ln[Ct/4] Plots for PS2 CC/GG and GA/TC Nearest Neighbor Doublets**

Melting temperatures are plotted as a function of duplex concentration (19 nM – 3 uM) using fixed-width black ovals. The height of each oval, as well as the color bar, represents the S.D. of eight replicates at that concentration. Each linear fit thus encompasses 104 separate Tm measurements (8 rep. x 13 conc.). Four duplexes for each NN were studied: the unmodified control (gray), the "top" dithioated strand (blue), the "bottom" dithioated strand (green), and the "both" dithioated version (cyan).

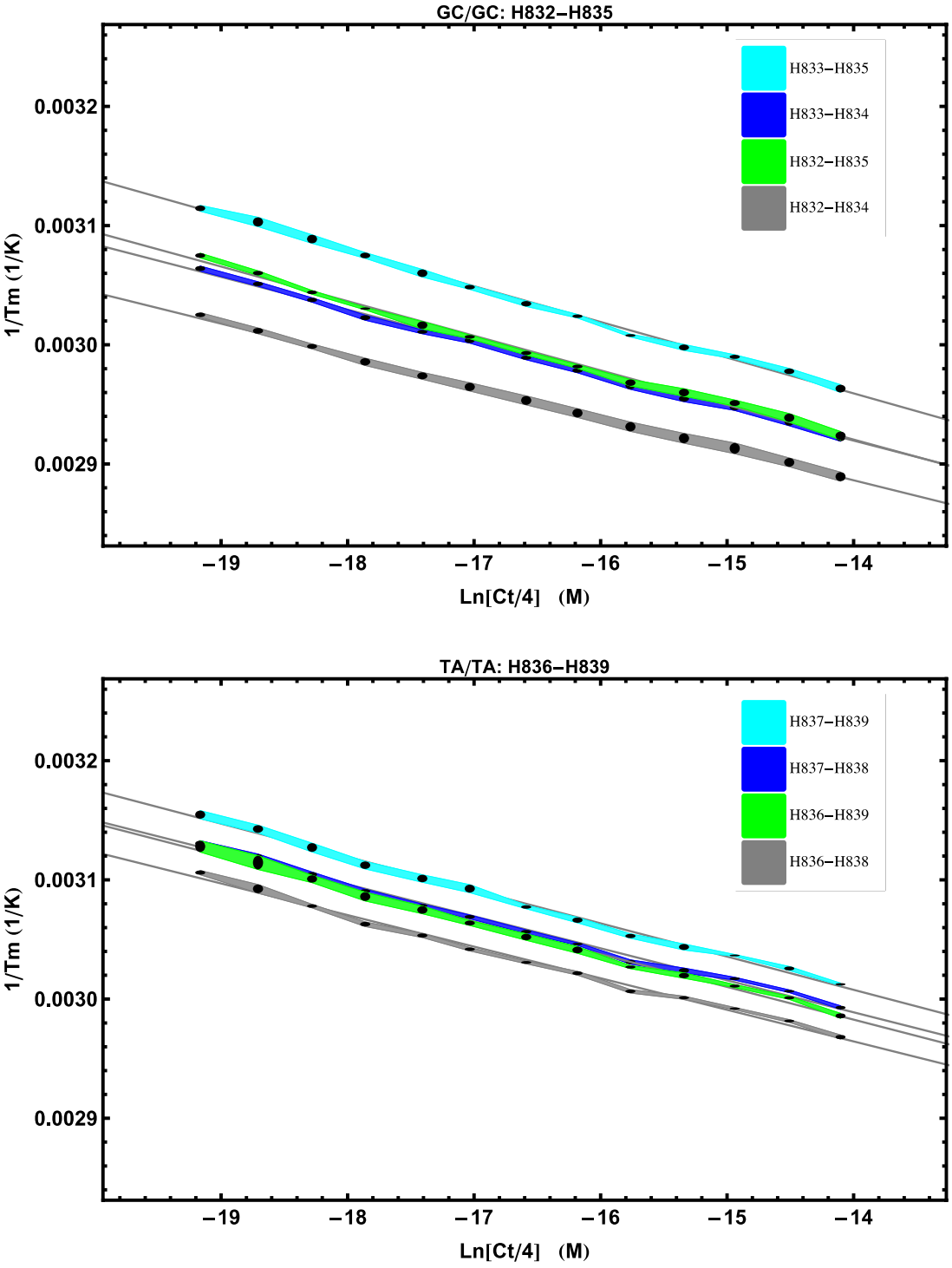



**Fig. 2.19) $T_m^{-1}$ vs. Ln[Ct/4] Plots for PS2 GC/GC and TA/TA Nearest Neighbor Doublets**

Melting temperatures are plotted as a function of duplex concentration (19 nM – 3 uM) using fixed-width black ovals. The height of each oval, as well as the color bar, represents the S.D. of eight replicates at that concentration. Each linear fit thus encompasses 104 separate Tm measurements (8 rep. x 13 conc.). Four duplexes for each NN were studied: the unmodified control (gray), the "top" dithioated strand (blue), the "bottom" dithioated strand (green), and the "both" dithioated version (cyan).

**Table 2.15) Comparison of Results from the Two Thermodynamic Parameter Extraction Methods: The Two-State Test (H800-H819)**

| *Duplex ID* | *$Tm^{-1}$ vs. Ln(Ct/4) Plots (Eq. 2.116)* | | | *Melting Curve Fits (Eq. 2.117)* | | | *Two-State Test* | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\Delta G^o_{37}$ kcal·mol$^{-1}$ | $\Delta H^o$ kcal·mol$^{-1}$ | $\Delta S^o$ cal·mol$^{-1}$·K$^{-1}$ | $\Delta G^o_{37}$ kcal·mol$^{-1}$ | $\Delta H^o$ kcal·mol$^{-1}$ | $\Delta S^o$ cal·mol$^{-1}$·K$^{-1}$ | $\Delta G^o_{37}$ % Diff. | $\Delta H^o$ % Diff | $\Delta S^o$ % Diff. |
| H800-H802 | -15.91 ± 0.15 | -82.5 ± 2.2 | -214.8 ± 6.6 | -15.92 ± 0.17 | -82.6 ± 2.4 | -215.0 ± 7.4 | 0.0 | 0.1 | 0.1 |
| H801-H802 | -14.91 ± 0.12 | -78.6 ± 2.0 | -205.3 ± 6.1 | -14.21 ± 0.10 | -67.5 ± 1.5 | -171.9 ± 4.5 | -4.8 | -15.2 | -17.7 |
| H800-H803 | -14.15 ± 0.10 | -75.0 ± 1.9 | -196.0 ± 5.7 | -14.24 ± 0.08 | -76.2 ± 1.8 | -199.7 ± 5.5 | 0.6 | 1.6 | 1.8 |
| H801-H803 | -13.52 ± 0.09 | -76.2 ± 2.0 | -202.0 ± 6.1 | -13.25 ± 0.09 | -70.3 ± 1.6 | -183.8 ± 5.0 | -2.0 | -8.1 | -9.4 |
| H804-H806 | -17.25 ± 0.19 | -85.6 ± 2.3 | -220.2 ± 6.7 | -17.54 ± 0.25 | -88.8 ± 3.1 | -229.8 ± 9.3 | 1.6 | 3.7 | 4.3 |
| H805-H806 | -15.59 ± 0.14 | -79.1 ± 2.0 | -204.6 ± 6.0 | -14.95 ± 0.22 | -69.9 ± 3.0 | -177.3 ± 9.0 | -4.2 | -12.2 | -14.3 |
| H804-H807 | -15.52 ± 0.14 | -79.7 ± 2.0 | -206.9 ± 6.2 | -15.25 ± 0.13 | -75.7 ± 1.9 | -194.9 ± 5.7 | -1.8 | -5.1 | -6.0 |
| H805-H807 | -14.11 ± 0.10 | -74.9 ± 1.9 | -196.0 ± 5.7 | -13.41 ± 0.13 | -62.2 ± 2.0 | -157.3 ± 6.2 | -5.1 | -18.6 | -21.9 |
| H808-H810 | -16.03 ± 0.15 | -78.6 ± 2.0 | -201.7 ± 5.9 | -16.65 ± 0.18 | -86.4 ± 2.6 | -225.0 ± 7.8 | 3.8 | 9.5 | 10.9 |
| H809-H810 | -15.40 ± 0.13 | -77.8 ± 2.0 | -201.2 ± 5.9 | -15.11 ± 0.13 | -73.6 ± 2.0 | -188.7 ± 6.2 | -1.9 | -5.5 | -6.4 |
| H808-H811 | -14.42 ± 0.10 | -70.7 ± 1.6 | -181.4 ± 5.0 | -14.89 ± 0.09 | -77.8 ± 1.5 | -202.7 ± 4.7 | 3.2 | 9.5 | 11.1 |
| H809-H811 | -14.08 ± 0.10 | -73.1 ± 1.8 | -190.2 ± 5.4 | -13.82 ± 0.13 | -68.6 ± 2.1 | -176.7 ± 6.3 | -1.8 | -6.3 | -7.4 |
| H812-H814 | -15.04 ± 0.13 | -81.8 ± 2.2 | -215.2 ± 6.6 | -14.72 ± 0.13 | -76.6 ± 1.8 | -199.5 ± 5.5 | -2.2 | -6.6 | -7.6 |
| H813-H814 | -13.70 ± 0.10 | -79.1 ± 2.2 | -210.9 ± 6.7 | -13.58 ±0.10 | -76.5 ± 3.0 | -202.9 ± 9.4 | -0.9 | -3.3 | -3.9 |
| H812-H815 | -13.48 ± 0.09 | -75.6 ± 1.9 | -200.3 ± 6.0 | -13.32 ± 0.11 | -72.0 ± 2.5 | -189.2 ± 7.7 | -1.2 | -4.9 | -5.7 |
| H813-H815 | -12.12 ± 0.06 | -74.0 ± 1.9 | -199.4 ± 6.0 | -12.03 ± 0.09 | -70.8 ± 3.4 | -189.4 ± 10.6 | -0.7 | -4.4 | -5.1 |
| H816-H818 | -17.71 ± 0.21 | -91.0 ± 2.6 | -236.2 ± 7.6 | -17.59 ± 0.14 | -89.6 ± 1.5 | -232.2 ± 4.4 | -0.7 | -1.5 | -1.7 |
| H817-H818 | -17.13 ± 0.20 | -91.7 ± 2.7 | -240.3 ± 7.9 | -14.68 ± 0.26 | -60.7 ± 3.2 | -148.4 ± 9.4 | -15.4 | -40.6 | -47.3 |
| H816-H819 | -16.67 ±0.18 | -85.4 ± 2.3 | -221.5 ± 6.9 | -16.84 ± 0.13 | -87.4 ± 2.1 | -227.6 ± 6.3 | 1.0 | 2.4 | 2.7 |
| H817-H819 | -16.35 ± 0.18 | -88.9 ± 2.5 | -234.1 ± 7.6 | -14.27 ± 0.25 | -60.3 ± 3.3 | -148.3 ± 9.8 | -13.6 | -38.4 | -44.8 |

Thermodynamic values from the two methods are compared for each duplex. If the values deviate by more than 15%, they are considered to be non-two state transitions (red). Approximately 88% of the duplexes were two-state. The sign of the values indicates that the values represent the annealing reaction, which is the standard method of reporting.

**Table 2.16) Comparison of Results from the Two Thermodynamic Parameter Extraction Methods: The Two-State Test (H820-H839)**

| *Duplex ID* | *$Tm^{-1}$ vs. Ln(Ct/4) Plots (Eq. 2.116)* | | | *Melting Curve Fits (Eq. 2.117)* | | | *Two-State Test* | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\Delta G^o_{37}$ kcal·mol$^{-1}$ | $\Delta H^o$ kcal·mol$^{-1}$ | $\Delta S^o$ cal·mol$^{-1}$·K$^{-1}$ | $\Delta G^o_{37}$ kcal·mol$^{-1}$ | $\Delta H^o$ kcal·mol$^{-1}$ | $\Delta S^o$ cal·mol$^{-1}$·K$^{-1}$ | $\Delta G^o_{37}$ % Diff. | $\Delta H^o$ % Diff | $\Delta S^o$ % Diff. |
| H820-H822 | -17.45 ± 0.19 | -83.3 ± 2.1 | -212.5 ± 6.3 | -17.93 ± 0.19 | -88.7 ± 2.1 | -228.1 ± 6.2 | 2.7 | 6.2 | 7.1 |
| H821-H822 | -16.84 ± 0.17 | -82.6 ± 2.1 | -212.0 ± 6.3 | -16.29 ± 0.21 | -76.1 ± 2.6 | -192.8 ± 7.7 | -3.4 | -8.2 | -9.5 |
| H820-H823 | -17.05 ± 0.18 | -82.8 ± 2.1 | -212.1 ± 6.3 | -17.49 ± 0.22 | -88.1 ± 3.0 | -227.7 ± 8.8 | 2.6 | 6.2 | 7.1 |
| H821-H823 | -16.64 ± 0.17 | -84.5 ± 2.3 | -218.7 ± 6.7 | -15.93 ± 0.17 | -75.7 ± 2.3 | -192.6 ± 6.9 | -4.4 | -11.0 | -12.7 |
| H824-H826 | -15.89 ± 0.15 | -78.9 ± 2.0 | -203.2 ± 5.9 | -16.74 ± 0.17 | -90.0 ± 2.6 | -236.3 ± 7.8 | 5.2 | 13.2 | 15.1 |
| H825-H826 | -14.41 ± 0.10 | -72.4 ± 1.7 | -187.0 ± 5.2 | -14.44 ± 0.14 | -72.6 ± 2.2 | -187.5 ± 6.7 | 0.2 | 0.3 | 0.3 |
| H824-H827 | -15.21 ± 0.12 | -75.4 ± 1.8 | -193.9 ± 5.5 | -16.02 ± 0.13 | -86.7 ± 2.4 | -227.8 ± 7.3 | 5.2 | 14.0 | 16.1 |
| H825-H827 | -13.91 ± 0.09 | -73.1 ± 1.8 | -190.8 ± 5.5 | -13.81 ± 0.15 | -70.7 ± 2.5 | -183.4 ± 7.5 | -0.7 | -3.4 | -4.0 |
| H828-H830 | -17.99 ± 0.23 | -93.1 ± 2.7 | -242.1 ± 8.0 | -18.01 ± 0.16 | -93.1 ± 2.0 | -242.3 ± 6.2 | 0.1 | 0.1 | 0.0 |
| H829-H830 | -17.07 ± 0.20 | -90.1 ± 2.6 | -235.4 ± 7.6 | -15.74 ± 0.30 | -73.4 ± 3.6 | -185.9 ± 10.6 | -8.1 | -20.4 | -23.5 |
| H828-H831 | -15.99 ± 0.16 | -84.6 ± 2.3 | -221.4 ± 6.9 | -15.85 ± 0.17 | -82.5 ± 2.2 | -215.0 ± 6.5 | -0.9 | -2.5 | -2.9 |
| H829-H831 | -15.05 ± 0.13 | -80.3 ± 2.1 | -210.4 ± 6.4 | -14.16 ± 0.23 | -66.4 ± 3.2 | -168.3 ± 9.7 | -6.1 | -19.0 | -22.2 |
| H832-H834 | -16.56 ± 0.15 | -75.7 ± 1.8 | -190.7 ± 5.3 | -17.03 ± 0.18 | -81.0 ± 2.2 | -206.4 ± 6.7 | 2.8 | 6.8 | 7.9 |
| H833-H834 | -15.47 ± 0.12 | -72.4 ± 1.7 | -183.7 ± 5.0 | -15.94 ± 0.19 | -78.6 ± 2.7 | -202.0 ± 8.2 | 3.0 | 8.1 | 9.5 |
| H832-H835 | -15.09 ± 0.11 | -68.6 ± 1.5 | -172.6 ± 4.5 | -15.65 ± 0.17 | -76.0 ± 2.7 | -194.4 ± 8.1 | 3.7 | 10.2 | 11.9 |
| H833-H835 | -14.08 ± 0.09 | -66.4 ± 1.4 | -168.7 ± 4.4 | -14.59 ± 0.14 | -74.4 ± 2.1 | -192.9 ± 6.5 | 3.5 | 11.4 | 13.4 |
| H836-H838 | -14.68 ± 0.11 | -75.1 ± 1.8 | -195.0 ± 5.6 | -14.96 ± 0.12 | -79.5 ± 2.1 | -208.2 ± 6.2 | 1.9 | 5.7 | 6.6 |
| H837-H838 | -14.04 ± 0.10 | -74.2 ± 1.8 | -193.9 ± 5.6 | -14.35 ± 0.07 | -79.6 ± 1.3 | -210.5 ± 3.8 | 2.2 | 7.1 | 8.2 |
| H836-H839 | -14.06 ± 0.10 | -72.5 ± 1.7 | -188.6 ± 5.3 | -14.41 ± 0.13 | -78.6 ± 1.9 | -207.0 ± 5.7 | 2.5 | 8.0 | 9.3 |
| H837-H839 | -13.42 ± 0.08 | -71.4 ± 1.7 | -187.0 ± 5.3 | -13.75 ± 0.07 | -77.8 ± 1.5 | -206.5 ± 4.8 | 2.5 | 8.6 | 9.9 |

Thermodynamic values from the two methods are compared for each duplex. If the values deviate by more than 15%, they are considered to be non-two state transitions (red). Approximately 88% of the duplexes were two-state. The sign of the values indicates that the values represent the annealing reaction, which is the standard method of reporting.

## RESULT: TWO-STATE MELTING BEHAVIOR OF THE 40 DUPLEXES

The experiment was designed with short 11-basepair duplexes that could be expected to melt in a two-state (all-or-none) manner. Application of thermodynamic parameter extraction methods depends on the absence of intermediate states, or half-melted duplexes, as the equilibrium θ term accounts for only a fully folded bimolecular duplex state and an unfolded single strand coil state. In this study it was found after analysis of percent differences between the van't Hoff method and concentration dependence of melting temperature that the majority (88%) of the duplexes indeed melted in a two-state manner (Tables 2.14 and 2.15). The duplexes that did not melt in a two-state were the single strand dithioated NN doublets C-ps2-A/TG (H817·H818) and G-ps2-A/TC (H29-H830) and the double strand dithioated NN doublets C-ps2-A/T-ps2-G (H817·H819), G-ps2-A/T-ps2-C (H829·H831), and A-ps2-C/G-ps2-T (H805·H807). Of these five questionable melting behaviors most were only marginally non-two-state, having differences around 20% in the enthalpic and entropic terms, which only slightly exceeds the arbitrary 15% criteria that was established. It was only the C-ps2-A motif (H817) that had notable non-two-state behavior in the enthalpic and entropic terms, around 40%. There are two possible explanations for this observation, the first being that the C-A internucleotide phosphate linkage is more sensitive to dithioation in respect to DNA two-state DNA melting than the others. The second explanation is due to an artifact of the experimental design. Duplexes have a terminal dye—quencher pair and it is possible these moieties interact with neighboring base pairs. Perhaps an interaction with the dithioate C-A changes duplex melting behavior and local base pair cooperativity. Additionally, because the fluorescence depends on dye—quencher proximity and orientation, the observed fluorescence is more sensitive to non-two-state behavior than the UV absorbance signal. The non-two-state nature of the transition may be amplified from its true behavior if dissociation of the dye from the quencher does not coincide precisely with duplex melting. This would cause significant discrepancies in thermodynamic analysis and an inaccuracy of thermodynamic values. Nevertheless, in all of the

questionable-behavior samples the free energy term shows better agreement than the $\Delta H°$ and $\Delta S°$ terms, resulting in only ~5-15% $\Delta G°$ differences.

### RESULT: GENERAL CONSEQUENCES OF DITHIOATION ON DNA DUPLEX THERMODYNAMICS

The average value difference thermodynamic value Table 2.18 was created to show the effect of dithioation compared to unmodified phosphodiester backbones in DNA hybridization. The values presented here are halved values of the differences observed between duplexes, termed "per appearance" because each strand contains two copies of the dithioation site. This format of reporting allows for a simpler comparison with previously published thermodynamic data on dithioate melting. From these data presented here it is apparent that dithioation always decreases the melting temperature of the duplex, with a $\Delta T_m$ range for single dithioation of a single strand of -0.8 to -2.7°C, with an average $\Delta T_m$ of -1.7 $\pm$ 0.7°C (Table 2.18, Fig. 2.20.A from Table 2.17). These values compare favorably with previously published $\Delta T_m$ values per appearance for dithioates (Table 2.12), which are ~1.1°C[171], ~1.0°C[257], ~0.9°C[258], and ~2.1°C[261]. This melting temperature difference for dithioates comes directly from free energy term. Dithioation of the phosphate always increases $\Delta G°$ in the unfavorable (positive) direction, with $\Delta\Delta G°$ values ranging from +0.2 to +1.06 kcal·mol$^{-1}$, and an average increase +0.67 $\pm$ 0.26 kcal·mol$^{-1}$ (Table 2.18). The enthalpy $\Delta H°$ of the duplex also sees an unfavorable (positive) effect from dithioation, and there is a large difference between those having the greatest and least effect (Fig. 2.20.C). The $\Delta\Delta H°$ values range from +0.2 to +7.0 kcal·mol$^{-1}$, with an average increase of +3.3 $\pm$ 2.1 kcal·mol$^{-1}$. On the other hand, the entropy $\Delta S°$ of the duplex almost always sees a favorable (positive) effect of dithioation for the majority of the NN substitutions (Fig. 2.20.D). Only T-ps2-A, C-ps2-G, and A-ps2-T have nearly negligible effects. The $\Delta\Delta S°$ values range from -0.3 to 19.9 cal·mol$^{-1}$·K$^{-1}$ with an average of +8.6 $\pm$ 6.0 cal·mol$^{-1}$·K$^{-1}$. The generalized consequence of dithioation of the phosphate is an unfavorable enthalpic and favorable entropic term that combines into a slightly unfavorable free energy, which results in a decreased duplex melting temperature. From the ordering of the average difference values, there is an obvious NN sequence dependency.

## Table 2.17) Average Thermodynamic Values of the Duplexes Used In This Study

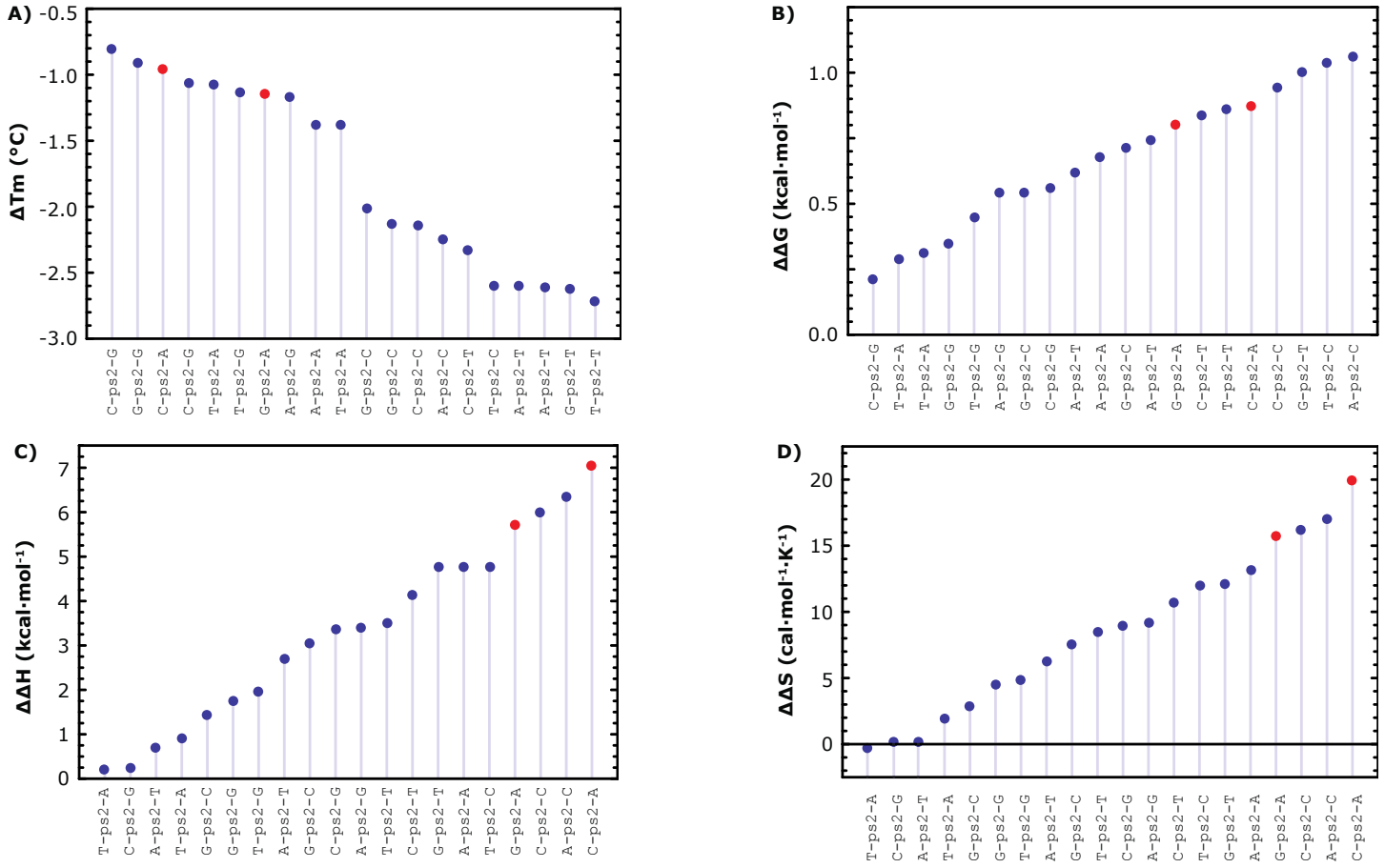
| Duplex ID | NN Doublet | Tm 2 µM (°C) | Average Values (from Tables 2.14 & 2.15) $\Delta G^{o}_{37}$ kcal·mol$^{-1}$ | $\Delta H^{o}$ kcal·mol$^{-1}$ | $\Delta S^{o}$ cal·mol$^{-1}$·K$^{-1}$ | Two-State? |
|---|---|---|---|---|---|---|
| H800-H802 | AA/TT | 65.62 | -15.91 ± 0.11 | -82.6 ± 1.6 | -214.9 ± 4.9 | Yes |
| H801-H802 | A-ps2-A/TT | 62.85 | -14.56 ± 0.08 | -73.1 ± 1.3 | -188.6 ± 3.8 | Yes |
| H800-H803 | AA/T-ps2-T | 60.19 | -14.20 ± 0.07 | -75.6 ± 1.3 | -197.9 ± 3.9 | Yes |
| H801-H803 | A-ps2-A/T-ps2-T | 57.05 | -13.39 ± 0.06 | -73.2 ± 1.3 | -192.9 ± 3.9 | Yes |
| H804-H806 | AC/GT | 70.32 | -17.40 ± 0.16 | -87.2 ± 1.9 | -225.0 ± 5.7 | Yes |
| H805-H806 | A-ps2-C/GT | 65.82 | -15.27 ± 0.13 | -74.5 ± 1.8 | -191.0 ± 5.4 | Yes |
| H804-H807 | AC/G-ps2-T | 65.07 | -15.39 ± 0.10 | -77.7 ± 1.4 | -200.9 ± 4.2 | Yes |
| H805-H807 | A-ps2-C/G-ps2-T | 60.47 | -13.76 ± 0.08 | -68.6 ± 1.4 | -176.7 ± 4.2 | No |
| H808-H810 | AG/CT | 67.55 | -16.34 ± 0.12 | -82.5 ± 1.6 | -213.3 ± 4.9 | Yes |
| H809-H810 | A-ps2-G/CT | 65.22 | -15.25 ± 0.09 | -75.7 ± 1.4 | -194.9 ± 4.3 | Yes |
| H808-H811 | AG/C-ps2-T | 62.90 | -14.66 ± 0.07 | -74.2 ± 1.1 | -192.0 ± 3.4 | Yes |
| H809-H811 | A-ps2-G/C-ps2-T | 60.61 | -13.95 ± 0.08 | -70.8 ± 1.4 | -183.4 ± 4.1 | Yes |
| H812-H814 | AT/AT | 62.15 | -14.88 ± 0.09 | -79.2 ± 1.4 | -207.4 ± 4.3 | Yes |
| H813-H814 | A-ps2-T/AT | 56.94 | -13.64 ± 0.07 | -77.8 ± 1.9 | -206.9 ± 5.8 | Yes |
| H812-H815 | AT/A-ps2-T | 56.95 | -13.40 ± 0.07 | -73.8 ± 1.6 | -194.8 ± 4.9 | Yes |
| H813-H815 | A-ps2-T/A-ps2-T | 51.04 | -12.07 ± 0.05 | -72.4 ± 1.9 | -194.4 ± 6.1 | Yes |
| H816-H818 | CA/TG | 70.11 | -17.65 ± 0.13 | -90.3 ± 1.5 | -234.2 ± 4.4 | Yes |
| H817-H818 | C-ps2-A/TG | 68.20 | -15.90 ± 0.17 | -76.2 ± 2.1 | -194.4 ± 6.2 | No |
| H816-H819 | CA/T-ps2-G | 67.85 | -16.76 ± 0.11 | -86.4 ± 1.6 | -224.6 ± 4.7 | Yes |
| H817-H819 | C-ps2-A/T-ps2-G | 65.94 | -15.31 ± 0.15 | -74.6 ± 2.1 | -191.2 ± 6.2 | No |
| H820-H822 | CG/CG | 72.12 | -17.69 ± 0.14 | -86.0 ± 1.5 | -220.3 ± 4.4 | Yes |
| H821-H822 | C-ps2-G/CG | 69.99 | -16.57 ± 0.14 | -79.3 ± 1.7 | -202.4 ± 5.0 | Yes |
| H820-H823 | CG/C-ps2-G | 70.50 | -17.27 ± 0.14 | -85.5 ± 1.8 | -219.9 ± 5.4 | Yes |
| H821-H823 | C-ps2-G/C-ps2-G | 68.35 | -16.29 ± 0.12 | -80.1 ± 1.6 | -205.7 ± 4.8 | Yes |
| H824-H826 | CC/GG | 66.67 | -16.31 ± 0.11 | -84.5 ± 1.6 | -219.8 ± 4.9 | Yes |
| H825-H826 | C-ps2-C/GG | 62.38 | -14.42 ± 0.09 | -72.5 ± 1.4 | -187.3 ± 4.3 | Yes |
| H824-H827 | CC/G-ps2-G | 64.86 | -15.62 ± 0.09 | -81.0 ± 1.5 | -210.8 ± 4.6 | Yes |
| H825-H827 | C-ps2-C/G-ps2-G | 59.81 | -13.86 ± 0.09 | -71.9 ± 1.5 | -187.1 ± 4.6 | Yes |
| H828-H830 | GA/TC | 70.44 | -18.00 ± 0.14 | -93.1 ± 1.7 | -242.2 ± 5.0 | Yes |
| H829-H830 | G-ps2-A/TC | 68.16 | -16.40 ± 0.18 | -81.7 ± 2.2 | -210.7 ± 6.5 | No |
| H828-H831 | GA/T-ps2-C | 65.25 | -15.92 ± 0.12 | -83.6 ± 1.6 | -218.2 ± 4.8 | Yes |
| H829-H831 | G-ps2-A/T-ps2-C | 62.97 | -14.60 ± 0.13 | -73.3 ± 1.9 | -189.4 ± 5.8 | No |
| H832-H834 | GC/GC | 71.55 | -16.79 ± 0.12 | -78.4 ± 1.4 | -198.6 ± 4.3 | Yes |
| H833-H834 | G-ps2-C/GC | 67.52 | -15.71 ± 0.11 | -75.5 ± 1.6 | -192.9 ± 4.8 | Yes |
| H832-H835 | GC/G-ps2-C | 67.29 | -15.37 ± 0.10 | -72.3 ± 1.5 | -183.5 ± 4.6 | Yes |
| H833-H835 | G-ps2-C/G-ps2-C | 62.74 | -14.33 ± 0.08 | -70.4 ± 1.3 | -180.8 ± 3.9 | Yes |
| H836-H838 | TA/TA | 62.52 | -14.82 ± 0.08 | -77.3 ± 1.4 | -201.6 ± 4.2 | Yes |
| H837-H838 | T-ps2-A/TA | 59.75 | -14.19 ± 0.06 | -76.9 ± 1.1 | -202.2 ± 3.4 | Yes |
| H836-H839 | TA/T-ps2-A | 60.37 | -14.24 ± 0.08 | -75.6 ± 1.3 | -197.8 ± 3.9 | Yes |
| H837-H839 | T-ps2-A/T-ps2-A | 57.60 | -13.59 ± 0.06 | -74.6 ± 1.2 | -196.8 ± 3.6 | Yes |

**Table 2.18) Average Value Differences of PS2 vs. PO NNs, Per Appearance**

| NN | Duplex ID | Dithioation | $\Delta T_m$ (2 µM) °C | $\Delta\Delta G^o_{37}$ kcal·mol$^{-1}$ | $\Delta\Delta H^o$ kcal·mol$^{-1}$ | $\Delta\Delta S^o$ cal·mol$^{-1}$·K$^{-1}$ |
|---|---|---|---|---|---|---|
| | H801-H802 | A-ps2-A/TT | -1.38 | +0.68 | +4.8 | +13.2 |
| **AA/TT** | H800-H803 | AA/T-ps2-T | -2.71 | +0.86 | +3.5 | +8.5 |
| | H801-H803 | A-ps2-A/T-ps2-T | -4.28 | +1.26 | +4.7 | +11.0 |
| | H805-H806 | A-ps2-C/GT | -2.25 | +1.06 | +6.3 | +17.0 |
| **AC/GT** | H804-H807 | AC/G-ps2-T | -2.63 | +1.01 | +4.7 | +12.1 |
| | H805-H807 | A-ps2-C/G-ps2-T | -4.92 | +1.82 | +9.3 | +24.2 |
| | H809-H810 | A-ps2-G/CT | -1.17 | +0.54 | +3.4 | +9.2 |
| **AG/CT** | H808-H811 | AG/C-ps2-T | -2.33 | +0.84 | +4.1 | +10.7 |
| | H809-H811 | A-ps2-G/C-ps2-T | -3.47 | +1.19 | +5.8 | +15.0 |
| | H813-H814 | A-ps2-T/AT | -2.61 | +0.62 | +0.7 | +0.2 |
| **AT/AT** | H812-H815 | AT/A-ps2-T | -2.60 | +0.74 | +2.7 | +6.3 |
| | H813-H815 | A-ps2-T/A-ps2-T | -5.56 | +1.40 | +3.4 | +6.5 |
| | H817-H818 | C-ps2-A/TG | -0.96 | +0.87 | +7.0 | +19.9 |
| **CA/TG** | H816-H819 | CA/T-ps2-G | -1.13 | +0.45 | +1.9 | +4.8 |
| | H817-H819 | C-ps2-A/T-ps2-G | -2.08 | +1.17 | +7.8 | +21.5 |
| | H821-H822 | C-ps2-G/CG | -1.06 | +0.56 | +3.3 | +8.9 |
| **CG/CG** | H820-H823 | CG/C-ps2-G | -0.81 | +0.21 | +0.3 | +0.2 |
| | H821-H823 | C-ps2-G/C-ps2-G | -1.89 | +0.70 | +3.0 | +7.3 |
| | H825-H826 | C-ps2-C/GG | -2.15 | +0.95 | +6.0 | +16.3 |
| **CC/GG** | H824-H827 | CC/G-ps2-G | -0.91 | +0.35 | +1.7 | +4.5 |
| | H825-H827 | C-ps2-C/G-ps2-G | -3.43 | +1.23 | +6.3 | +16.3 |
| | H829-H830 | G-ps2-A/TC | -1.14 | +0.80 | +5.7 | +15.8 |
| **GA/TC** | H828-H831 | GA/T-ps2-C | -2.60 | +1.04 | +4.8 | +12.0 |
| | H829-H831 | G-ps2-A/T-ps2-C | -3.74 | +1.70 | +9.9 | +26.4 |
| | H833-H834 | G-ps2-C/GC | -2.01 | +0.54 | +1.4 | +2.9 |
| **GC/GC** | H832-H835 | GC/G-ps2-C | -2.13 | +0.71 | +3.0 | +7.5 |
| | H833-H835 | G-ps2-C/G-ps2-C | -4.40 | +1.23 | +4.0 | +8.9 |
| | H837-H838 | T-ps2-A/TA | -1.39 | +0.31 | +0.2 | -0.3 |
| **TA/TA** | H836-H839 | TA/T-ps2-A | -1.07 | +0.29 | +0.9 | +1.9 |
| | H837-H839 | T-ps2-A/T-ps2-A | -2.46 | +0.62 | +1.4 | +2.4 |

"Per Appearance" implies that the ΔΔ values of thermodynamic differences in Table 2.15 and Table 2.16 have been halved, since each strand contains two dithioation sites. Although it's a simplifying assumption, this reporting method allows easier comparison with previously published PS2 thermodynamic melting data, which have different dithioated oligo designs.

**Fig. 2.20) Graphical Representation of Ordered Average Value Differences (per Appearance) for $\Delta T_m$, $\Delta\Delta G$, $\Delta\Delta H$, and $\Delta\Delta S$**




Average value $\Delta T_m$, $\Delta\Delta G°$, $\Delta\Delta H°$, and $\Delta\Delta S°$ differences for dithioate NNs (per appearance) from Table 2.17 are ordered by magnitude. Blue dots indicate date is from two-state verified, whereas red are from non-two-state transitions (Tables 2.14-2.15)

## RESULT: ENTHALPIC/ENTROPIC COMPENSATION IN DITHIOATED NN PHOSPHATES

The differences in $\Delta\Delta G°$ values for dithioated NNs are small compared to the differences of the enthalpic and entropic component. The $\Delta\Delta H°$ and $\Delta\Delta S°$ contributions were directly correlated ($R^2 = 0.994$), as demonstrated by the linear plot in Fig. 2.21. This establishes a strong enthalpy-entropy compensation for dithioated phosphates, having the relation:

$$\Delta\Delta S° = \Delta\Delta H°/347\ (\mathrm{K}) - 1.03\ (\mathrm{cal{\cdot}mol^{-1}{\cdot}K^{-1}}) \qquad (2.118)$$

This means that if the enthalpic contribution in the unfavorable direction is large, then there is a nearly equivalent contribution from the temperature-driven entropic component in the favorable direction. Such an observation for dithioates has not yet been reported, but these results are comparable to the enthalpic/entropic compensation equation of $\Delta\Delta S° = \Delta\Delta H°/360\ (\mathrm{K}) - 0.67$ $\mathrm{cal{\cdot}mol^{-1}{\cdot}K^{-1}}$ developed in Ignacio Tinoco's laboratory for base pair mismatches[271].

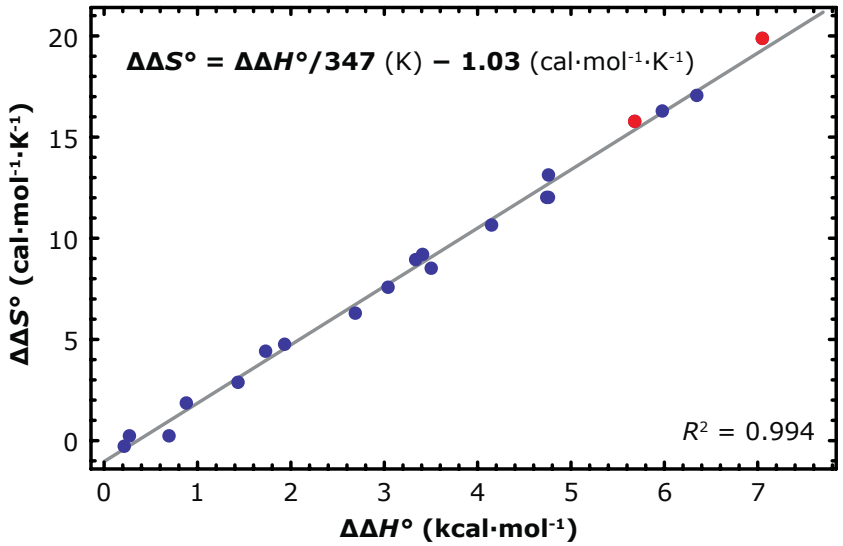



**Fig. 2.21) Enthalpic/Entropic Compensation in Dithioated NN Phosphates**

A direct correlation between $\Delta\Delta H°$ and $\Delta\Delta S°$ average value differences (Table 2.17) is observed for dithioated NN phosphates. Experimental values for two-state (blue) and non-two-state (red) single strand dithioates are plotted in relation to the line of least-squares fit: Eq. 2.118 $\Delta\Delta S° = \Delta\Delta H°/347\ (\mathrm{K}) - 1.03\ (\mathrm{cal{\cdot}mol^{-1}{\cdot}K^{-1}})$. This equation can be used to predict $\Delta\Delta S°$ from $\Delta\Delta H°$ values from earlier dithioate publications.

The vast majority of published dithioate melting data uses consecutive and fully dithioated sequences, which extends beyond the isolated NN experimental design of this project. Of all the dithioate melting studies previously published, only

the manuscript out of the Caruthers lab[258] contains sequences with an isolated dithioate modification that can be used to validate the average difference values of Table 2.17 and predictive ability of Equation 2.118. In this manuscript, $T_m$ measurements and $\Delta\Delta H°$ values were reported using maximum $\delta\alpha/\delta T$ calculations. The duplex name RBC-S2(7) is composed of a 5'-GATTCAG-ps2-CTAGTCCA-3' sequence hybridized with an unmodified complementary strand. This can be represented by the G-ps2-C/CG dithioate NN, and RBC-S2(7) has a reported $\Delta T_m$ of -1.6°C and $\Delta\Delta H°$ of +0.5 kcal·mol$^{-1}$. These values agree well with the results from this study, listed in Table 2.17, which are -2.01°C and +0.54 kcal·mol$^{-1}$ for the G-ps2-C/CG moiety. These are remarkable agreements, considering that the $\Delta T_m$ range for single dithioation of a single strand is -0.8 to -2.7°C, and the $\Delta\Delta H°$ values range from +0.2 to +7.0 kcal·mol$^{-1}$. Using equation 2.118 the predicted $\Delta\Delta S°$ for RBC-S2(7) would be +0.4 cal·mol$^{-1}$·K$^{-1}$. There is one more duplex from the Caruthers study that is useful for validation, RBC-S2(5,10) with the sequence 5'-GATTC-ps2-AGCTA-ps2-GTCCA-3'. This contains an isolated C-ps2-A/GT and isolated A-ps2-G/TC moiety, with a reported $\Delta T_m$ for the duplex of -1.8°C and $\Delta\Delta H°$ of +11 kcal·mol$^{-1}$. Again, these experimental values compare well with the results from Table 2.17, which predicts a combined $\Delta T_m$ of -2.13°C and combined $\Delta\Delta H°$ of +10.3 kcal·mol$^{-1}$. The predicted $\Delta\Delta S°$ for RBC-S2(5,10) using their reported $\Delta\Delta H°$ would be +2.1 cal·mol$^{-1}$·K$^{-1}$. Beyond these two sequences, using additional sequences from the Caruthers study[258] for validation of the values reported here becomes difficult because the remaining sequences are composed of consecutive and terminal dithioates.

## RESULT: ADDITIVITY OF PS2 THERMODYNAMIC EFFECTS

Four duplexes were used to study each dithioate NN: an unmodified control, the "top" single strand, the "bot" single strand", and the "top+bot" dual strand dithioate. The additivity of dithioate effects was investigated, that is, if the top and bottom strand differences sum to the dual strand dithioated duplex difference value:

$$\Delta T_{\mathrm{m}}^{\mathrm{top+bot}} = \Delta T_{\mathrm{m}}^{\mathrm{top}} + \Delta T_{\mathrm{m}}^{\mathrm{bot}} \tag{2.119}$$

These calculations and values reported in Table 2.19 are for the full duplex difference values (Table 2.17), not the halved "per appearance" data (Table 2.18). The data shows an apparent additivity in the $T_{\mathrm{m}}$ and $\Delta\Delta G°$ terms, within 0.8°C and 0.5 kcal·mol$^{-1}$, respectively (Table 2.19). Additivity is best observed for melting temperatures and free energies because of low relative errors ($T_{\mathrm{m}}$ error is <1% and $\Delta\Delta G°$is about 4%), $\Delta\Delta H°$ and $\Delta\Delta S°$ values typically have errors about 8% and are more noisy. The results for the $T_{\mathrm{m}}$ and $\Delta\Delta G°$ terms suggests that dithioate thermodynamic effects are very local and extend only to short distances.

## RESULT: MODELING OF PS2 NN

Because two of the duplexes C-ps2-A/TG (H817·H818) and G-ps2-A/TC (H29-H830) melted in non-two-state fashion, there is not enough data to obtain complete 16 NN parameters, and the design matrix is singular. To reduce the number of parameters needed to describe the NNs, alternative models were proposed utilizing a purine/pyrimidine term instead of base identity at positions before or after the dithioated phosphate; *e.g.* A-ps2-C and A-ps2-T are both described by A-ps2-Pyr, according to the Pur/Pyr identity of their 3' base. Ten different models were created using a range of 1-8 parameters; the exact parameter list is covered in detail in Table 2.20. The dataset started with melting data from 18 different duplexes after the non-two-state data H817-H818, and H829-H830 were removed from the original group of twenty single-strand dithioate duplexes. An additional A-ps2-T dataset from an earlier pilot study (not shown) was also included in the analysis, bringing the duplex melting data count up to 19.

## Table 2.19) Additive Properties of Single and Dual Stranded Dithioation

| ID | NN Doublet | Sequence | $\Delta T_m$ (°C) | $\Delta T_m$ Top + $\Delta T_m$ Bot | $\Delta T_m$ Diff. (%) | $\Delta\Delta G°$ (kcal·mol$^{-1}$) | $\Delta\Delta G°$ Top + $\Delta\Delta G°$ Bot | $\Delta\Delta G°$ Diff. (%) | $\Delta\Delta H°$ Diff. (%) | $\Delta\Delta S°$ Diff. (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| **↓Two-State Dual Strand Dithioated Duplexes (4 PS2 Total: 2 per Strand)** | | | | | | | | | | |
| H801-H803 | A-ps2-A/T-ps2-T | TXRD-CGAA-ps2-ACCA-ps2-AGC<br>IBRQ-GCTT-ps2-TGGT-ps2-TCG | -8.57 | -8.20 | 4.3% | 2.53 | 3.07 | 21.6% | 76.9% | 97.4% |
| H809-H811 | A-ps2-G/C-ps2-T | TXRD-CGAA-ps2-GTCA-ps2-GGC<br>IBRQ-GCTT-ps2-CAGT-ps2-CCG | -6.94 | -6.98 | 0.6% | 2.39 | 2.76 | 15.8% | 29.4% | 32.9% |
| H813-H815 | A-ps2-T/A-ps2-T | TXRD-CGAA-ps2-TCTA-ps2-TGG<br>IBRQ-GCTT-ps2-AGAT-ps2-ACC | -11.12 | -10.42 | 6.2% | 2.81 | 2.72 | 3.1% | 1.0% | 0.5% |
| H821-H823 | C-ps2-G/C-ps2-G | TXRD-CACC-ps2-GAGC-ps2-GTC<br>IBRQ-GTGG-ps2-CTCG-ps2-CAG | -3.77 | -3.74 | 0.6% | 1.40 | 1.50 | 9.8% | 21.3% | 24.8% |
| H825-H827 | C-ps2-C/G-ps2-G | TXRD-AGAC-ps2-CTTC-ps2-CAG<br>IBRQ-TCTG-ps2-GAAG-ps2-GTC | -6.86 | -6.11 | 11.1% | 2.45 | 2.59 | 5.7% | 22.7% | 26.8% |
| H833-H835 | G-ps2-C/G-ps2-C | TXRD-TCTG-ps2-CGAG-ps2-CCG<br>IBRQ-AGAC-ps2-GCTC-ps2-GGC | -8.81 | -8.29 | 5.9% | 2.46 | 2.51 | 2.2% | 12.4% | 17.0% |
| H837-H839 | T-ps2-A/T-ps2-A | TXRD-CGCT-ps2-AGAT-ps2-ACG<br>IBRQ-GCGA-ps2-TCTA-ps2-TGC | -4.92 | -4.92 | 0.1% | 1.23 | 1.21 | 2.1% | 19.8% | 34.4% |
| **↓Non-Two-State Dual Strand Dithioated Duplexes (4 PS2 Total: 2 per Strand)** | | | | | | | | | | |
| H805-H807 | A-ps2-C/G-ps2-T | IBRQ-GCTT-ps2-GAGT-ps2-GCG<br>TXRD-CGAC-ps2-ATCC-ps2-AGC | -9.85 | -9.75 | 0.9% | 3.63 | 4.13 | 13.7% | 19.0% | 20.3% |
| H817-H819 | C-ps2-A/T-ps2-G | TXRD-CGAC-ps2-ATCC-ps2-AGC<br>IBRQ-GCTG-ps2-TAGG-ps2-TCG | -4.17 | -4.17 | 0.0% | 2.34 | 2.64 | 12.6% | 14.7% | 15.0% |
| H829-H831 | G-ps2-A/T-ps2-C | TXRD-ACGG-ps2-ACTG-ps2-AGC<br>IBRQ-TGCC-ps2-TGAC-ps2-TCG | -7.47 | -7.47 | 0.0% | 3.40 | 3.68 | 8.5% | 5.7% | 5.1% |

**Table 2.20) Description of Models Used for Dithioated NNs**

| *Model* | *# of Parameters* | *Model Parameter and Appearance Frequency (#)* |
|---|---|---|
| Single | 1 | **Single ps2**<br>ps2 (38) |
| PurPyr3 | 2 | **Purine/Pyrimidine character at the 3'-end of the ps2**<br>ps2-Pur-3' (16) ps2-Pyr-3' (22) |
| BasePairs3 | 2 | **AT/GC base pair at 3'-end of ps2**<br>ps2-(AT) at 3' (18) ps2-(GC) at 3' (20) |
| BaseGroup3 | 2 | **A+C or G+T group at 3'-end of ps2 (test of disagreement, for a "bad" model)**<br>ps2-(A or C) at -3' (16) ps2-(G or T) at 3' (22) |
| Base3 | 4 | **Base identity at 3'-end of ps2**<br>ps2-A-3' (6) ps2-C-3' (10) ps2-G-3' (10) ps2-T-3' (12) |
| Base5 | 4 | **Base identity at 5'-end of ps2**<br>5'-A-ps2 (12) 5'-C-ps2 (8) 5'-G-ps2 (8) 5'-T-ps2 (10) |
| PurPyr | 4 | **Purine and pyrimidine character of NN bases (both 5'- and 3'-ends of ps2)**<br>5'-Pur-ps2-Pur-3' (6) 5'-Pyr-ps2-Pyr-3' (8) 5'-Pur-ps2-Pyr-3' (14) 5'-Pyr-ps2-Pur-3' (10) |
| Base5PurPyr | 8 | **Base identity at 5'-end of ps2 and purine/pyrimidine character at 3'-end of ps2**<br>5'-A-ps2-Pur (4) 5'-A-ps2-Pyr (8) 5'-C-ps2-Pur (4) 5'-C-ps2-Pyr (4)<br>5'-G-ps2-Pur (2) 5'-G-ps2-Pyr (6) 5'-T-ps2-Pur (6) 5'-T-ps2-Pyr (4) |
| PurPyrBase3 | 8 | **Base identity at 3'-end of ps2 and purine/pyrimidine character at 5'-end of ps2**<br>Pur-ps2-A-3' (2) Pyr-ps2-A-3' (4) Pur-ps2-C-3' (6) Pyr-ps2-C-3' (4)<br>Pur-ps2-G-3' (4) Pyr-ps2-G-3' (6) Pur-ps2-T-3' (8) Pyr-ps2-T-3' (4) |
| Base3PurPyr | 8 | **Base identity at 3'-end of ps2 and purine/pyrimidine character of the next 3'-end base (long distance effect)**<br>ps2-A-Pur-3' (3) ps2-A-Pyr-3' (3) ps2-C-Pur-3' (5) ps2-C-Pyr-3' (5)<br>ps2-G-Pur-3' (5) ps2-G-Pyr-3' (5) ps2-T-Pur-3' (7) ps2-T-Pyr-3' (5) |
| NN | 16 | **Complete Nearest-Neighbor model**<br>A-ps2-A (2) A-ps2-C (2) A-ps2-G (2) A-ps2-T (6) C-ps2-A (0) C-ps2-C (2) C-ps2-G (4) C-ps2-T (2)<br>G-ps2-A (0) G-ps2-C (4) G-ps2-G (2) G-ps2-T (2) T-ps2-A (4) T-ps2-C (2) T-ps2-G (2) T-ps2-T (2) |

With a dataset containing 19 duplexes it is not possible to derive through regression parameters for every type of dithioated nearest neighbor possibility (16 types, see NN), because the degrees of freedom is too low. A set of models was therefore developed to reduce the number of parameters describing the dithioated NN. Context descriptions of each parameter on the right are followed by the number of appearances of the parameter in the dataset (in parenthesis). The 19 duplexes had 2 dithioates each, for a total of 38 dithioate appearances.

To test the models, difference data from the thermodynamic averages (Table 2.17) were subjected to multivariate linear regression. Equations were a linear combination of parameters and their appearance frequency in the dataset, and the observed thermodynamic difference values were taken as constants. The multivariate fits were calculated using the Excel LINEST function and errors were estimated from residuals of the fit. The results were loosely ranked (Table 2.21) according to the mean absolute errors of the predicted $T_m$ and ΔΔG° values from the fitted parameters to observation, the numbers of degrees of freedom in the model ($v$), and the reduced chi-squared value ($\chi^2/v$). The best models for the NN were: PurPyr, which had four parameters using the purine/pyrimidine character of bases on both the 5'- and 3'-end of the dithioated phosphate (*i.e.* 5'-Pur/Pyr-ps2-Pur/Pyr-3'); PurPyr3, which had two parameters using the Pur/Pyr character of the base on the 3'-end of the dithioate (*i.e.* ps2-Pur/Pyr-3'); and Base3, which had four parameters using the base identity on the 3'-end of the dithioated phosphate (*i.e.* ps2-A/C/G/T-3'). The fitted parameters for the top four models are shown in Table (2.21).

Analyzing the rank order of the models (Table 2.21) a few points should be discussed: First, dithioated phosphate effects are sequence specific, which is supported by comparing values from the "single" PS2 model ($\Delta T_m$ MAE = 1.32°C and $\Delta T_m$ $\chi^2/v = 12.36$), with those of other, better models ($\Delta T_m$ MAE $< \sim$0.5°C and $\Delta T_m$ $\chi^2/v < \sim$2.5). Secondly, the 3'-end of the dithioated phosphate confers greater sequence specificity than 5'-end, which is supported by a direct comparison of the "Base3" model ($\Delta T_m$ $\chi^2/v = 2.13$, $\Delta\Delta G°$ $\chi^2/v = 6.22$), against the "Base5" model ($\Delta T_m$ $\chi^2/v = 13.61$, $\Delta\Delta G°$ $\chi^2/v = 13.04$). The top six fitted models produced parameters that could accurately predict the melting temperature of the duplex within a mean average error of 0.5°C. Of these, one model uses 2 parameters, two models use 4 parameters, and three models use 8 parameters. With these data, the inclusion of the additional parameters beyond four does not appreciably increase the accuracy of prediction and their use is not justified. Nevertheless, the choice of model preference is up to the individual investigator and for these purposes a list of the most useful parameters is presented in Table (2.21).

## Table 2.21) Ranked Order of Models for Dithioates in Single Strand

| | | | | $\Delta T_m$ prediction (°C) | | $\Delta\Delta G°$ prediction (kcal·mol$^{-1}$) | | $\Delta\Delta H°$ prediction (kcal·mol$^{-1}$) | | $\Delta\Delta S°$ prediction (cal·mol$^{-1}$·K$^{-1}$) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| # | Model | P | $v$ | MAE | $\chi^2/v$ | MAE | $\chi^2/v$ | MAE | $\chi^2/v$ | MAE | $\chi^2/v$ |
| 1 | PurPyr | 4 | 15 | 0.49 | 2.48 | 0.26 | 4.54 | 2.4 | 1.83 | 6.7 | 1.70 |
| 2 | PurPyr3 | 2 | 17 | 0.49 | 2.12 | 0.28 | 5.56 | 2.7 | 2.31 | 7.9 | 2.16 |
| 3 | Base3 | 4 | 15 | 0.46 | 2.13 | 0.28 | 6.22 | 2.7 | 2.45 | 7.8 | 2.28 |
| 4 | PurPyrBase3 | 8 | 11 | 0.42 | 2.51 | 0.21 | 5.44 | 2.0 | 1.90 | 5.7 | 1.77 |
| 5 | Base5PurPyr | 8 | 11 | 0.39 | 2.30 | 0.23 | 6.26 | 2.1 | 2.15 | 6.1 | 1.98 |
| 6 | Base3PurPyr | 8 | 11 | 0.39 | 2.09 | 0.27 | 7.21 | 2.6 | 3.23 | 7.5 | 3.03 |
| 7 | Single | 1 | 18 | 1.32 | 12.36 | 0.46 | 11.85 | 3.1 | 2.72 | 8.7 | 2.40 |
| 8 | BasePairs3 | 2 | 17 | 1.25 | 12.47 | 0.45 | 12.86 | 3.1 | 2.83 | 8.8 | 2.48 |
| 9 | BaseGroup3 | 2 | 17 | 1.34 | 13.45 | 0.46 | 13.10 | 3.1 | 2.89 | 8.8 | 2.54 |
| 10 | Base5 | 4 | 15 | 1.22 | 13.61 | 0.46 | 13.04 | 2.9 | 2.96 | 7.9 | 2.60 |
| 11 | NN | 16 | 3 | 0.19 | 4.09 | 0.11 | 11.40 | 1.0 | 3.68 | 2.8 | 3.35 |

Difference data from the thermodynamic averages Table (2.16) were subjected to multivariate linear regression using models described in Table (2.19). Modeling results are ranked (#) loosely according to mean absolute errors (MAE) of predicted values for $T_m$ and ΔΔG° values, the number of parameters (P) and degrees of freedom ($v$) and then according to the reduced chi-squared $\chi^2/v$.

## Table 2.22) Fitted Thermodynamic Difference Parameters for Various NN Models

| *Model* | *Parameters* | $\Delta\Delta G°$ (kcal·mol$^{-1}$) | $\Delta\Delta H°$ (kcal·mol$^{-1}$) | $\Delta\Delta S°$ (cal·mol$^{-1}$·K$^{-1}$) |
|---|---|---|---|---|
| **Single** | -ps2- | 0.67 | 3.06 | 7.71 |
| **PurPyr3** | -ps2-Pur-3' | 0.42 | 2.07 | 5.29 |
| | -ps2-Pyr-3' | 0.85 | 3.79 | 9.46 |
| **PurPyr** | 5'-Pur-ps2-Pur-3' | 0.52 | 3.30 | 8.94 |
| | 5'-Pyr-ps2-Pyr-3' | 0.92 | 4.60 | 11.86 |
| | 5'-Pyr-ps2-Pur-3' | 0.36 | 1.33 | 3.10 |
| | 5'-Pur-ps2-Pyr-3' | 0.82 | 3.32 | 8.09 |
| **Base3** | -ps2-A-3' | 0.43 | 1.95 | 4.91 |
| | -ps2-C-3' | 0.86 | 4.31 | 11.13 |
| | -ps2-G-3' | 0.42 | 2.13 | 5.52 |
| | -ps2-T-3' | 0.85 | 3.35 | 8.07 |
| **PurPyrBase3** | Pur-ps2-A-3' | 0.68 | 4.76 | 13.16 |
| | Pyr-ps2-A-3' | 0.30 | 0.55 | 0.79 |
| | Pur-ps2-C-3' | 0.77 | 3.61 | 9.14 |
| | Pyr-ps2-C-3' | 0.99 | 5.38 | 14.13 |
| | Pur-ps2-G-3' | 0.45 | 2.57 | 6.84 |
| | Pyr-ps2-G-3' | 0.40 | 1.85 | 4.65 |
| | Pur-ps2-T-3' | 0.85 | 3.11 | 7.30 |
| | Pyr-ps2-T-3' | 0.85 | 3.82 | 9.59 |

## RESULT: STRUCTURAL EXPLANATION FOR DITHIOATE EFFECTS (BI VS. BII PROPENSITY)

An attempt was made to correlate the observed thermodynamic difference parameters for dithioated NNs with the BI vs. BII propensity of the NN. Oligonucleotide backbones are classified as BI or BII according to the dihedral angles of $\varepsilon$ (C4'-C3'-03'-P) and $\zeta$ (C3'-03'-P-05'). Both orientations have been observed crystallographically[272,273], but the BI is more common. The BI state places the phosphate in a roughly symmetric position with respect to both grooves, where as the BII state swings the phosphate around towards the minor groove[274]. This transition involves coupled changes of the $\varepsilon$ and $\zeta$ dihedral angles which pass from (t,g-) in BI to (g-,t) in BII, and the conformation of the phosphate can be characterized by the difference $\varepsilon - \zeta$ which passes from roughly -90° in BI to roughly +90° in BII state. Two studies have analyzed dozens of NMR[272] and crystallography data[272,273] to determine propensities of dinucleotide junctions.

Dithioate difference values for $\Delta T_{\mathrm{m}}$ positively correlate with %BII propensities ($R = 0.78$) (Fig. 2.23). This indicates that there is a relation between inherent phosphate backbone flexibility in DNA NNs and the magnitude of phosphorodithioation effect on melting temperature. The more flexible the NN inherently is, the weaker the phosphorodithioation effect. No correlation was found between %BII propensities and $\Delta\Delta G°$ ($R = -0.34$), $\Delta\Delta H°$ ($R = 0.08$), or interestingly, the entropic $\Delta\Delta S°$ term ($R = 0.13$).

**Table 2.23) A Review of BII Propensities in DNA NNs**

| Pur/Pyr | NN | % BII |
|---|---|---|
| RpR | AA | 12 |
| | AG | 15 |
| | GA | 37 |
| | GG | 48 |
| YpY | CC | 32 |
| | CT | 0 |
| | TC | 9 |
| | TT | 0 |
| RpY | AC | 6 |
| | GT | 0 |
| | GC | 22 |
| | AT | 0 |
| YpR | CA | 52 |
| | CG | 36 |
| | TA | 15 |
| | TG | 31 |

Values are calculated from Hartmann et al. 2006, which is a literature review of NMR data.

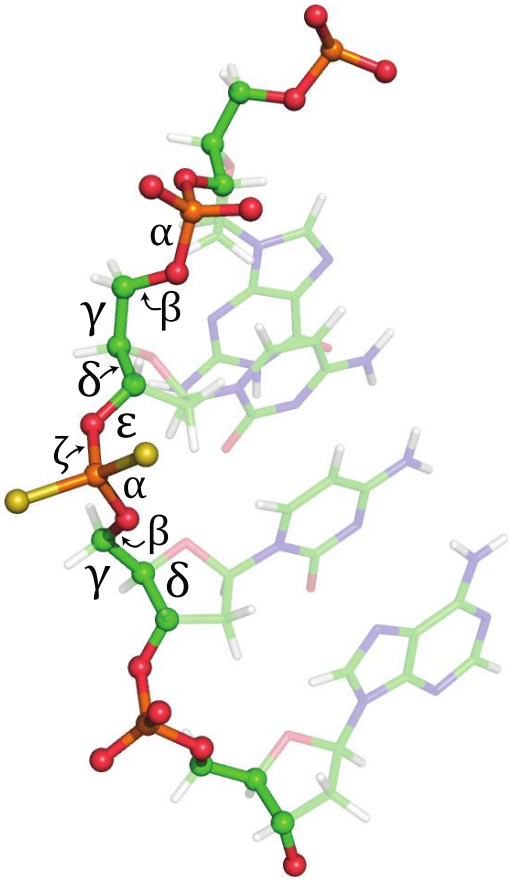



**Fig. 2.22) Dihedral Angles in a Phosphate Backbone**

fadfsdfasfas

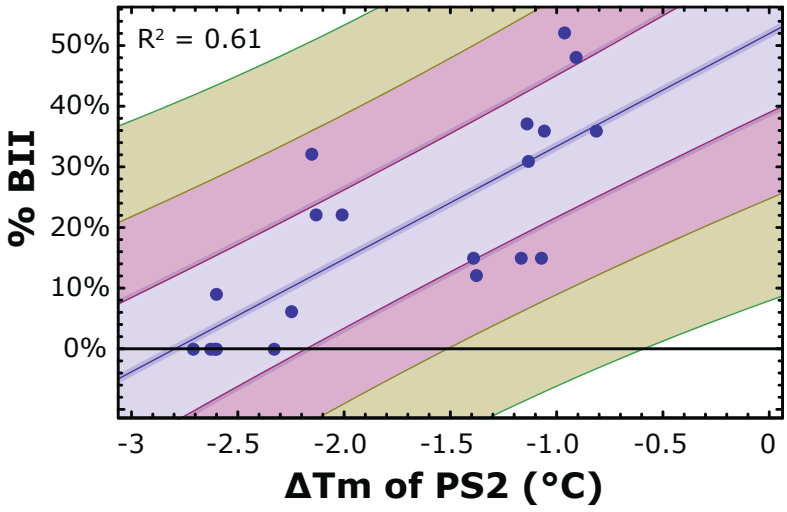



**Fig. 2.23) Correlation of %BII with $\Delta T_m$ for PS2**

The %BII propensity from Hartmann et al., 2006 (NMR data only) plotted vs $\Delta T_m$ values for dithioates from this study. A moderate correlation is observed, indicating that rigid NNs are more susceptible to phosphorodithioation $\Delta\Delta T_m$ effects. The data fall within one-sigma (light blue) or two-sigma (pink) range.

## 2.6) Original Scientific Contributions

### First Comprehensive PS2 DNA Melting Studies

- The first detailed thermodynamic analysis of dithioated DNA melting was performed, all 10 NNs and 16 PS2 NNs (Table 2.13) were represented in the experiment. Using 13 different concentrations and 40 duplexes, a total of 4160 melting profiles were analyzed and thermodynamic information was extracted. The effects of dithioation on the energetics of the duplex could be generalized (Table 2.18). Upon dithioation, the melting temperature is lowered, the free energy of the duplex becomes less favorable (positive), the enthalpic term becomes less favorable (positive), and the entropic term becomes more favorable (positive).
- Since a few duplexes melted in a non-two-state manner (Tables 2.14-2.15), dithioate NN parameters for all 16 unique motifs could not be determined. Regardless, the melting data presented in this dissertation can be combined with results from future studies to fully develop the 16 unique parameters. This is standard procedure for nucleic acid thermodynamic studies; the amount of data required to develop NN parameters is immense and multiple datasets from several labs are usually combined to improve the statistical power of NN modeling.

### Nearest-Neighbor Parameter Modeling for PS2 Thermodynamic Difference Effects

- Numerous NN models with a lower number of parameters (Table 2.20) were successfully fit to the melting data (Table 2.22). These models reduced the parameters from 16 PS2 NNs to 1, 2, 4, and 8 parameters; which were defined using descriptors such as base identity (A/C/G/T) and purine/pyrimidine character of the bases 5'- and 3'- to dithioation site (Table 2.20). All of the modeled parameters (Table 2.22) can be used as correction factors for dithioated DNA sequences in the prediction of nucleic acid thermodynamics from nearest-neighbor basis sets (Table 2.9 and Table

2.10). Modeling for a single parameter representing any dithioate regardless of sequence context the difference values are: $\Delta\Delta G° = +0.67$ kcal·mol$^{-1}$, $\Delta\Delta H° = +3.1$ kcal·mol$^{-1}$, $\Delta\Delta S° = +7.7$ cal·mol$^{-1}$·K$^{-1}$.

- $\Delta T_m$ effects across the NN doublet pair are additive** (Table 2.19). *E.g.*

  $\Delta T_m$(A-ps2-C/G-ps2-T) = $\Delta T_m$(A-ps2-C/GT) + $\Delta T_m$(A-C/G-ps2-T)

** The upper limit of this additivity was not tested. (i.e. Would 50 PS2 modifications produce an oligo having a negative °C melting temperature?)

### DITHIOATE ENTHALPY-ENTROPY COMPENSATION

- Strong enthalpy-entropy compensation in dithioated NNs is observed (Fig. 2.21), so the PS2 NN enthalpic difference value and entropic difference values are proportional (i.e. low enthalpy values also have low entropy values). In previous phosphorodithioate melting studies typically only $T_m$ is reported, and occasionally the van't hoff enthalpy. Using evidence from this dissertation, differential entropy (in units of cal·mol$^{-1}$) can be calculated from enthalpy differential values (in units of kcal·mol$^{-1}$), according to the empirically determined relation (Eq. 2.118):

$$\Delta\Delta S° \text{ (cal·mol}^{-1}\text{·K}^{-1}) = \Delta\Delta H°/347 \text{ (K)} - 1.03$$

### CORRELATION OF $\Delta T_M$ FOR PS2 NNS AND THE BI/BII PROPENSITY OF THE CORRESPONDING PHOSPHATE

- $\Delta T_m$ effects for single dithioated NN doublets (*i.e.* top or bottom, but not both) are correlated (R = 0.78, Fig. 2.23) with BI-BII propensity values for the unmodified phosphate NN pairs determined from analysis of dozens of NMR structures[272] (Table 2.23). This is interesting because it shows that inherently rigid DNA NNs are more susceptible to $T_m$ perturbation upon dithioation than more flexible NNs.

# CHAPTER 3: DITHIOATED PHOSPHATES IN PROTEIN-DNA INTERACTIONS

## 3.1) Introduction to Protein-DNA Interactions

From Kossel's early work[1,3] on nuclein purification it was obvious that DNA interacts with proteins to form macromolecular complexes. The substance Kossel purified from DNA was later clarified into histones, a class of proteins responsible for spooling and condensing genomic material by electrostatically neutralizing the polyanionic phosphate backbone. This is accomplished by the protein's region of high alkaline amino acid composition, which provides electrostatic avidity towards the anionic phosphate backbone. The other side of the histone molecules is for associating with analogous histones to form a multicomponent strand-packing nucleosome. In nearly every DNA-binding protein there is a DNA-association region and then a protein-association region. The role fulfilled by this associated protein is what ascribes biological function to the protein-DNA complex.

Another prevalent population of DNA-binding proteins is the transcription factors, which have a protein-binding region that initiates assembly of RNA transcription complexes that progress gene expression. Protein-DNA interactions are critical to the central dogma of molecular biology[275], and without them genetic material could not be stored nor converted to biological function. Analysis of functional genome databases[276] supports this importance, showing that ~6-7% of eukaryotic genes encode proteins with DNA-binding domains[277].

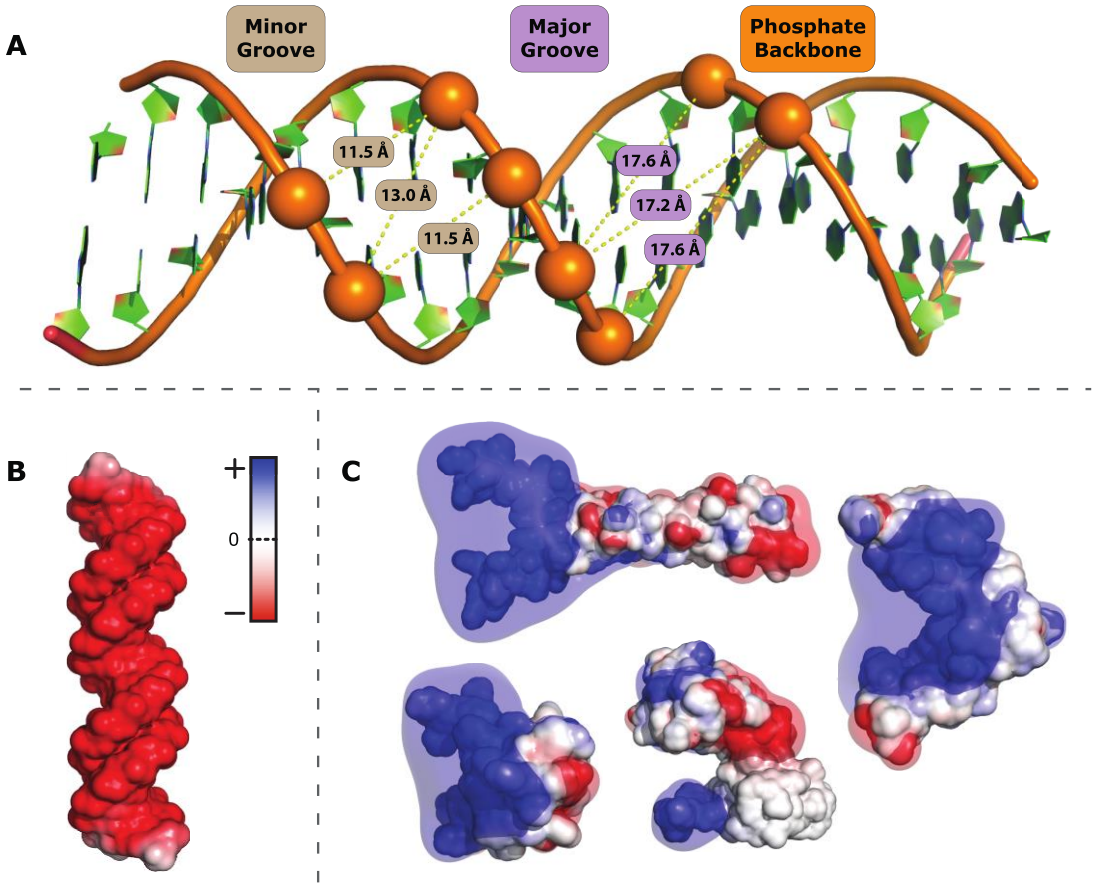



**Fig. 3.1) Electrostatic Attraction Between DNA and Proteins**

**A)** The DNA duplex has a high concentration of negative charge due to its polyanionic phosphate backbone. The interstrand phosphate distances across B-form duplex grooves are shown, major groove (~17 Å) and minor groove (~12 Å). **B)** DNA is negatively charged **C)** DNA-binding proteins contain regions rich in alkaline amino acids for attraction to polyanionic DNA. From left to right: λ-repressor, PDB: 1lmb[278] (repressor family); HoxD9, homology model (helix-turn-helix), Zif268 zinc finger, PDB: 1aay[279] (ββα-zinc finger); c-Fos/c-Jun, PDB: 1fos[280] (bZIP heterodimer).

## DOES PROTEIN READ DNA SEQUENCE THROUGH NUCLEOBASE INTERACTIONS?

Functionalization of stored genomic data is a sequence-specific phenomenon that involves an elegant search process in which non-specific interactions are fluidly formed and broken until a particular DNA sequence is recognized. This area has been of great interest to investigators, and discussions on target-site recognition using sliding, hopping, and trans-segmental transfer movements has been discussed[281]. When considering the sequence dependence of recognition it is natural to first focus on nucleobase-mediated contacts as a mechanism for target-site recognition. Such interactions occur deep within the major and minor groove of duplexes where the edge of planar nucleobase rings presents nitrogen and oxygen

moieties that provide attractive points of interaction for amino acid side chains[282]. DNA sequence, therefore, is responsible for presenting these moieties in the appropriate positions creating a direct sequence recognition mechanism. Although such contacts with base planar edges are present in almost every protein-DNA structure, the search for patterns and a direct readout recognition code of protein sidechains with DNA sequene has been unsuccessful. A clear relationship between the two is not apparent and direct readout cannot, by itself, sufficiently explain protein-DNA recognition and specificity. Instead there is a second process that contributes, which is the indirect readout of DNA shape prior to and during the binding process[283].

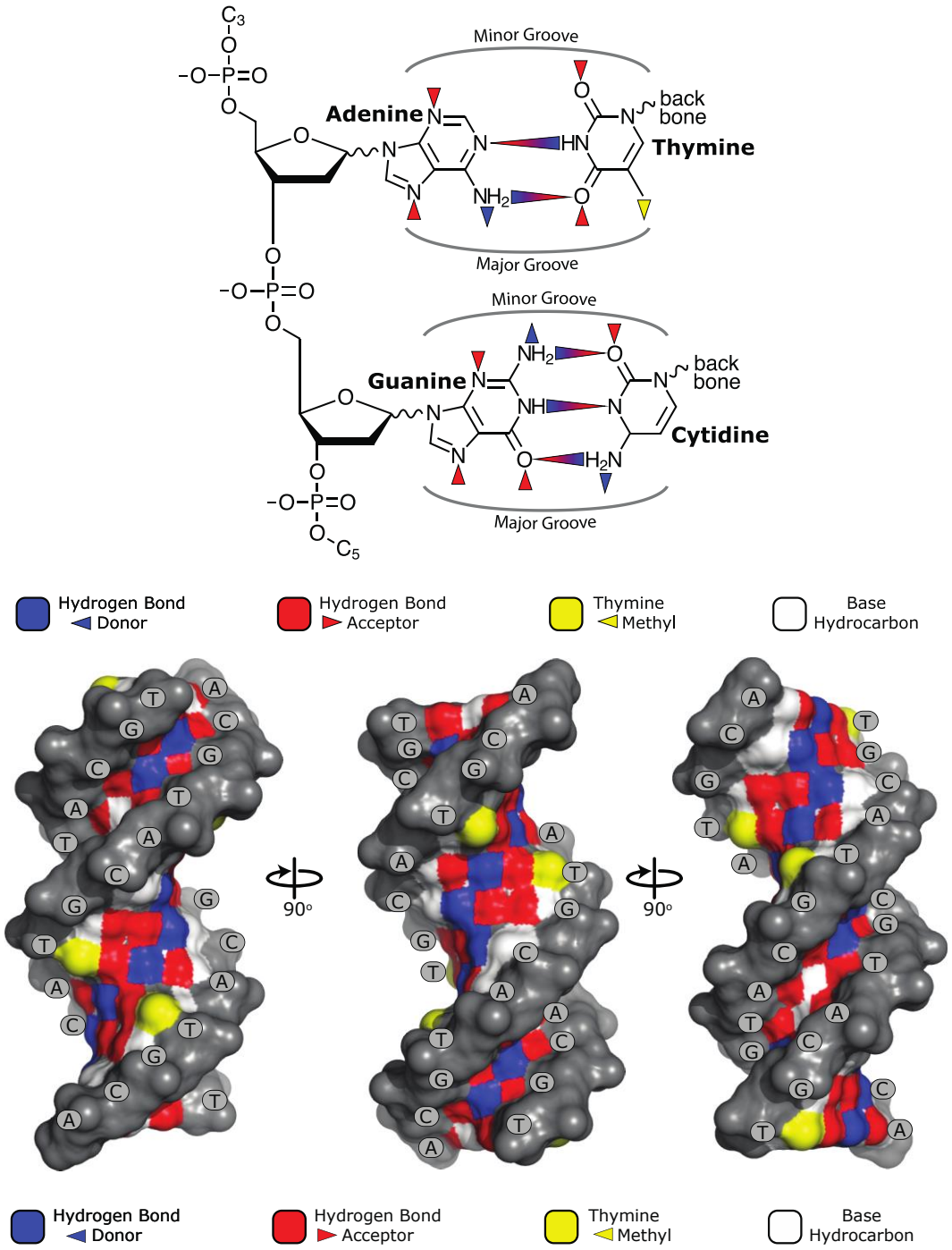



**Fig. 3.2) Nucleobase Edge Contact Points in the Major and Minor Grooves**

**A)** In a DNA duplex the nucleobases stack in a pi-pi orbital-stabilized structure, orienting the planar edges of nucleobases towards the major and minor grooves. Each basepair planar edge faces a specific groove in B-form DNA, and are labeled. Hydrogen bond donors, acceptors, and thymine methyl locations are represented in the basepair chemical structures using colored triangles. **B)** Surface representation of B-form DNA with groove contact points colored as hydrogen bond donor, acceptor, thymine methyl, or base hydrocarbons. These combinations, unique to each base pair, provide the basis for direct DNA sequence readout by nucleobase contacts. Figure 3B adapted from Rohs et. al[283].

## PROTEIN INDIRECTLY READS DNA SEQUENCE BY THE PHOSPHATE BACKBONE SHAPE

Structural information helps elucidate recognition mechanisms, but deriving atomic resolution information for protein-DNA complexes is difficult due to polymorphism in crystal packing that occurs during co-crystallization. The first insights into structural explanations for recognition came by solving the apo-structure of proteins and modeling in DNA duplexes. The first DNA-associating proteins for which this was accomplished were the catabolite gene activator protein[284], the Cro repressor[285], and the λ repressor[286], during the early 1980s. Crystallographic methods slowly improved, it was shown that truncating the DNA duplex binding site into half-sites[287] aided co-crystallization. The first full length DNA site was co-crystallized with its protein[288], but formed fragile crystals that produced diffraction patterns 3-4 Å in resolution, below atomic scale. To form better crystals it was shown that instead of using blunt-ended DNA duplexes, having a single base overhang would aid crystal formation by allowing a crystal contact point for stabilization[289]. After a year of growth, crystals of the *trp* repressor/operator were able to solve first atomic resolution structure of a protein-DNA complex[290]. The most striking finding for the authors was the absence of direct hydrogen-bonded or non-polar contacts between protein side-chains and the nucleobases, instead these contacts appeared water-mediated. All of the direct hydrogen-bonded contacts instead occurred with the operator's phosphate backbone. Although now it appears this system is rather unique in that zero direct nucleobase contacts occur, it provided the first evidence that the phosphate backbone plays a larger role in the energetic stabilization of protein-DNA complexes than expected. In the two decades since, thousands of high-resolution protein-DNA complexes have been deposited by researchers across the globe into the protein databank[291,292], and investigators can now mine this data in search of general themes for association[293-296]. An in-depth analysis encompassing examples of known DNA-binding structural motifs was completed about 10 years ago, and its statistics show that phosphate backbone contacts are indeed the most prevalent points of interaction in protein-DNA complexes[277,297].

Just as the sequence of an oligonucleotide arranges the planar side interaction sites of nucleobases for amino acids, the sequence of the oligo also plays a role in forming the shape of the phosphate backbone. It has been proposed that specificity in protein-DNA complexes arises not from direct readout of groove nucleobase information but rather through an indirect readout of shape variations in the phosphate backbone and oligo topology which are themselves sequence dependent[298,299]. Comparison of apo- and holo-form crystal structures has shown that many oligos involved in protein-DNA complexes adopt a shape similar to their bound form even prior to association, and that this shape aids in sequence recognition[300-302].

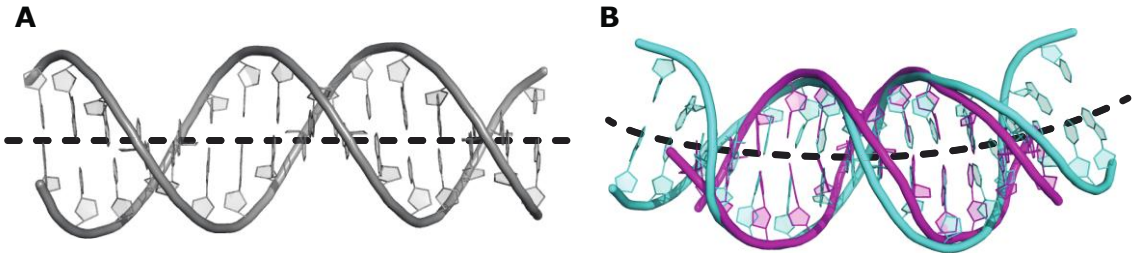



**Fig. 3.3) Indirect DNA Readout Relies on Phosphate Backbone Shape**

The phosphate backbone adopts sequence-dependent conformations, which are the basis of indirect (shape readout) DNA recognition. A good example is the transcriptional regulator (E2) for human papilloma virus. **A)** Here the specific DNA recognition site for HPV-18 E2 is modeled as an ideal linear B-form duplex. **B)** However, two X-ray crystal structures for the same sequence reveals a bent DNA duplex. The free duplex (unbound, PDB: 1ilc[302]) is colored magenta, and the structure extracted from the protein-DNA complex (bound, PDB: 1jj4[303]) is colored cyan. The duplex clearly adopts the A-tract curvature prior to binding to protein, implicating the importance of phosphate backbone shape in DNA binding site recognition by proteins.

From a processivity perspective, it is faster to initially discriminate between specific and non-specific oligo regions using an indirect phosphate shape readout and complement this method with direct nucleobase contacts deep in the grooves afterwards, to confirm sequence specificity. DNA shape at both the local and global levels has been found to be important for regulatory recognition and gene expression, and these structural characteristics are evolutionarily constrained. Stretches of topography-informed genomic regions correlate with functional noncoding elements, including enhancers, better than regions identified solely on the basis of nucleotide sequence[304]. New bioinformatics tools have emerged for detecting previously undiscovered transcription factor binding sites by emphasizing structural characteristics of DNA in addition to sequence-only searches[305,306].

## PROTEIN SIDE CHAINS THAT INTERACT WITH DNA

Interactions with DNA occurs via protein side chains, and although any polar amino acid side chain has the potential to interact with DNA the basic amino acids lysine and arginine are most involved in protein-DNA complexes where they encompass roughly 50% of all protein-DNA interactions[297]. These two amino acids contain high pKa side chain functional groups that are positively charged at neutral pH, which provides electrostatic attraction to DNA with its highly anionic charge properties. Structurally, the arginine side chain is planar, large, and rigid, and it contains three nitrogens (Nη1, Nη2, and Nε) available for hydrogen bonding. These nitrogens form single, bidentate, and complex interactions with either nucleobases or the phosphate backbone. A recent survey analyzing 129 diverse protein-DNA crystal structures shows that approximately 40% of arginine's interactions involve nucleobases contacts deep in the DNA grooves, with a base preference in the order of G>T≒A>C. The remaining 58% of arginine interactions are to the phosphate backbone, and just 2% involve DNA backbone sugar interactions[297].

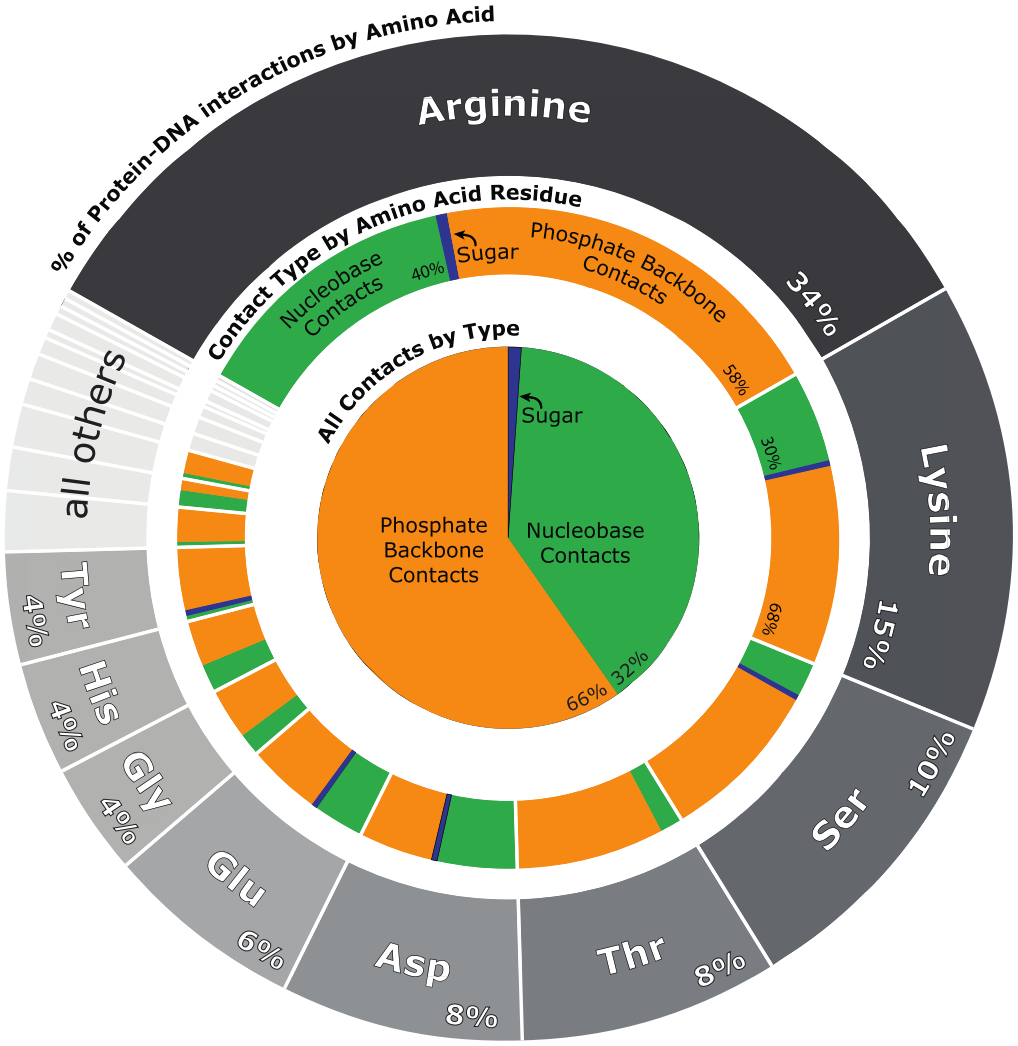



**Fig. 3.4) Distribution of Phosphate vs. Nucleobase Contacts in Protein-DNA Complexes**

Stacked pie chart created from data published in Luscombe et. al[297] detailing the distribution of contacts within 129 diverse protein-DNA crystal structures. **Outer ring)** Breakdown by amino acid residue type. Arginine and lysine encompass ~50% of all protein-DNA interactions. The remainder of high-frequency side chains charged or are polarizable. **Middle Ring)** Each amino acid is broken down into either nucleobase, phosphate backbone, or sugar contacts. **Inner Pie)** The distribution for all contacts is calculated and shown. Phosphate backbone-mediated contacts dominate in protein-DNA interactions.

Lysine on the other hand, is composed of a flexible alkyl chain with a terminal $NH_3^+$ group (protonated amine). Its high degree of conformational freedom makes this versatile functional group optimized for interactions with the phosphate backbone, which itself is very dynamic, alternating between BI and BII conformation states[307,308]. As protein binds, the anionic phosphates are neutralized and the dynamics of the DNA will increase[309], which emphasizes the importance of having a flexible binding partner. Lysine side chains are well suited for this, as

almost 70% of lysine interactions in protein-DNA complexes are to the phosphate backbone. Because lysine side chains are long in length, they can also reach deep within DNA grooves and the remaining 30% of interactions are towards moieties on the planar edge of nucleobases. This dissertation focuses on the former, probing interaction of lysine side chains with both a normal phosphate and a dithiophosphate.

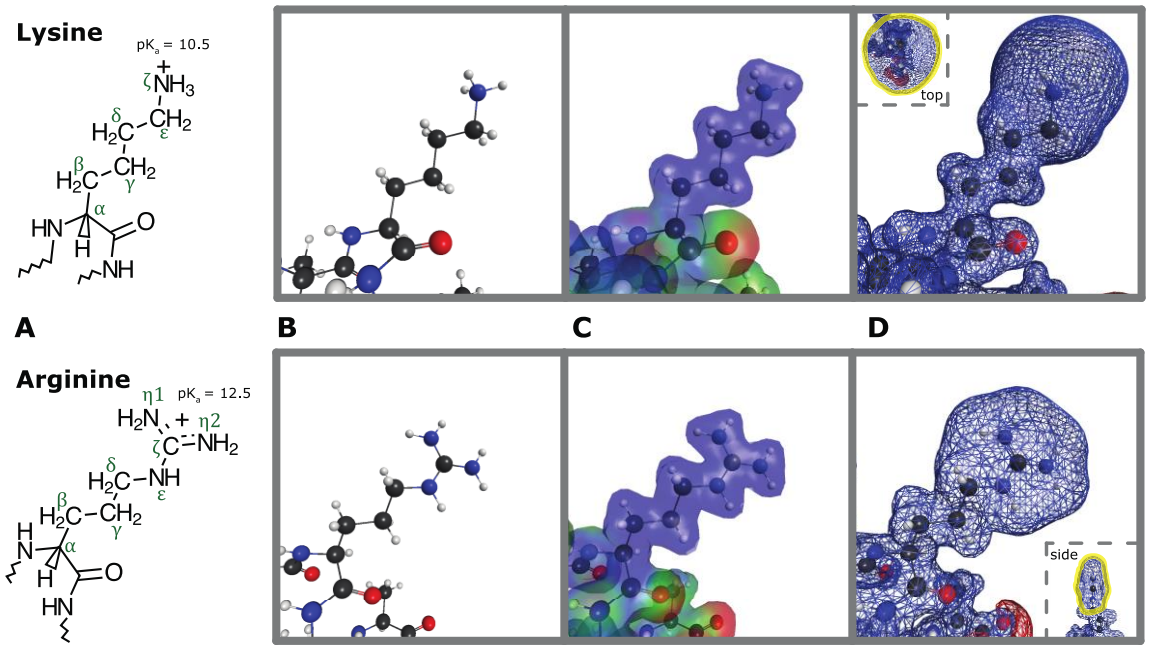



**Fig. 3.5) Properties of Lysine and Arginine Side Chains**

**A)** 2D chemical structure drawings of amino acid peptide bond and side chains. **B)** 3D ball-and-stick model. **C)** Electrostatic density map, placed over figure B. **D)** Surface electrostatic potential map, placed over figure B. **Top)** Lysine sidechain. The terminal $NH_3^+$ group is a flexible bulbous sphere. **Bot)** Arginine side chain. The terminal $NH_2$ groups are planar and rigid

## LYSINE SIDE CHAIN NMR METHODOLOGY DEVELOPMENT

The $NH_3$ group at the terminal end of lysine side chains has two properties that make it suitable for observation by nuclear magnetic resonance (NMR) spectroscopy. These are its hydrogens atoms that allow the application of highly sensitive proton NMR ($^1$H NMR), and its nitrogen atom that allows isotopic labeling for $^{15}$N NMR. Labeling is important because it replaces the $^{14}$N atom, which has an integer nuclear spin, with the $^{15}$N atom, which has a fractional nuclear spin of ½. This is change from a quadrupolar nucleus to a dipolar nucleus results in

significantly sharper NMR peaks compared to the naturally abundant (>99.6%) $^{14}N$ nuclei peaks, which are usually too broad to detect. However this increase in resolution using $^{15}N$ comes at a cost of reduced sensitivity, so in practice a 2D $^{1}H$-$^{15}N$ heteronuclear correlation experiment is used, invoking proton detection that greatly improves sensitivity. This technique was first described in 1980 by Geoffrey Bodenhausen and David Ruben of the Massachusetts Institute of Technology when they developed the Heteronuclear Single Quantum Coherence (HSQC) experiment for amide protons in the protein backbone[310]. The underlying principle of this technique is that after the amide proton is magnetized, an INEPT (Insensitive Nuclei Enhanced by Polarization Transfer) step is used to create proton anti-phase magnetization, which then transfers to the adjacent $^{15}N$ nuclei via scalar J-coupling. This $^{15}N$ magnetization is left to evolve with its chemical shift during t1 evolution time, and then a retro-INEPT step transfers this magnetization back to the proton as anti-phased magnetization. After refocusing, there is a decoupling procedure and then the $^{1}H$ signal is then captured during an acquisition stage, t2. The result of the $^{1}H$-$^{15}N$ HSQC experiment is a direct observation of the $^{1}H$ chemical shift, and an indirect observation of the $^{15}N$ chemical shift, making the technique a 2D NMR method.

Although the 2D $^{1}H$-$^{15}N$ HSQC experiment was developed for amide protons (N-H group), all hydrogen-nitrogen bonds can potentially produce peaks in the spectra, appearing in different chemical shift regions. For example in a normal HSQC spectra asparagine and glutamine side chain $NH_2$ groups appear in the top right corner as doublets with the same nitrogen but different hydrogen ppm values[311]. Despite this, applying the $^{1}H$-$^{15}N$ HSQC experiment to observe lysine side chains ($NH_3$) is not as straightforward as it might seem. This is primarily due to lysine $NH_3$ rapid exchange rate ($k_{\mathrm{ex}}^{\mathrm{water}}$) with the water solvent[312], which can be as fast as 4000 $s^{-1}$ under biological conditions[313]. This is too fast for observation, since $NH_3$ groups with water exchange rates greater than 100 $s^{-1}$ are undetectable by NMR[314].

Great progress in solving this problem for detecting $NH_3$ groups was made by Junji Iwahara during his post-doc tenure under Marius Clore at the National Institutes of Health[314-319], and later during his career at the University of Texas

Medical Branch[320-324]. He reasoned that if the $NH_3$ proton exchange rate were slowed down then $NH_3$ cross-peaks could be observed. There are several techniques for reducing the hydrogen exchange rate with the solvent: reducing temperature to lower kinetic motion, decreasing pH to increase free protons that compete with $NH_3$ proton exchange, and avoiding solvent interactions by adding a binding partner to block the $NH_3$ proton from solution. Using these three approaches Iwahara showed that lysine $NH_3$ groups from a protein could be observed during its interaction with phosphate backbone of DNA using $^{1}H$-$^{15}N$ HSQC[314]. The chemical shift appears at ~33 ppm, far upfield of the region normally viewed ($^{15}N$: ~100-130 ppm), and shows a distinct quartet splitting pattern indicative of an $NH_3$ group. Their appearance existed only when protein was present, implying their involvement in protein-DNA interactions.

Although this was a significant first step, the $NH_3$ peaks observed using HSQC and the similar heteronuclear experiment HMQC (Heteronuclear Multiple Quantum Coherence)[325] were much broader than amide peaks, and these unique $^{15}N$ line shapes were attributed to scalar relaxation occurring via transient water exchange[314]. This result was initially surprising, as previously HMQC was shown to be the best experiment for optimal $^{1}H$-$^{13}C$ correlation of $CH_3$ groups[326], which like $NH_3$ groups, are also classified as $AX_3$ first-order four spin systems[327]. To solve this issue, Iwahara developed a modified 2D $^{1}H$-$^{15}N$ correlation experiment that exclusively observed in-phase $^{15}N$ transverse coherence, the HISQC (Heteronuclear In-phase Single Quantum Coherence)[314]. This work was derived from the "decoupled HSQC" experiment[328] and utilizes continuous WALTZ-16[329,330] $^{1}H$-decoupling rather than a single 180° pulse. This forces the $^{15}N$ transverse magnetization during t1 evolution to always be in-phase ($N_x$ or $N_y$) with respect to $^{1}H$, resulting in strikingly sharper $^{15}N$ line shapes with higher intensities. With a few small modifications to the pulse sequence, Iwahara created the HISQC heteronuclear $^{1}H$-$^{15}N$ experiment that allows selective observation of the $NH_3$ groups in proteins, providing an invaluable tool for probing lysine side-chain interactions with DNA.

## 3.2) Developing an Experimental Model: HoxD9-DNA Complex

The purpose of this dissertation is to elucidate an understanding of sulfur-for-oxygen substitution at the basic science level. Therefore when developing a system for studying protein-DNA interactions it is important to keep things simplified so that analysis is straightforward and interpretation involves small logical conclusions rather than invoking complex hypothetical solutions. The homeodomain-DNA system was therefore chosen because it is one of the most well characterized protein-DNA systems, and every major technique used in this dissertation (CD, NMR, FP, ITC) has data from a homologous homeodomain system that provides reference information for experimental design as well as interpretation. HoxD9 was used specifically because my collaborator and co-mentor Junji Iwahara has previously optimized production, NMR, and binding experiments for the system[314,316,317]. My contribution to the field is the investigation of phosphate dithioation on interactions with a lysine side-chains in a protein-DNA complex. I make a single dithioate modification at a specific location, and then compare measurements with the unmodified phosphate.

### HOXD9 HOMEODOMAIN

HOXD9 is a sequence-specific DNA binding transcription factor encoded by the *hoxd9* gene (previously known as *hox4C* gene) located on chromosome 2 in the 2q31-2q37 region[331]. It is one member of the larger homeobox family, which are developmental regulatory proteins located in the cell nucleus. There are four similar homeobox gene clusters, HOXA, HOXB, HOXC, and HOXD, which reside on different chromosomes. Each gene encodes a specific homeoprotein, and the function of these was first characterized using *Drosophila melanogaster* homologues where they were shown to regulate embryonic development along the anterior-posterior axis[332,333]. A striking example of their function was demonstrated through the mutation a Hox gene subsequently named Antennapedia, which produced a fruit fly that developed its second pair of legs as ectopic antenna on top of its head[334].

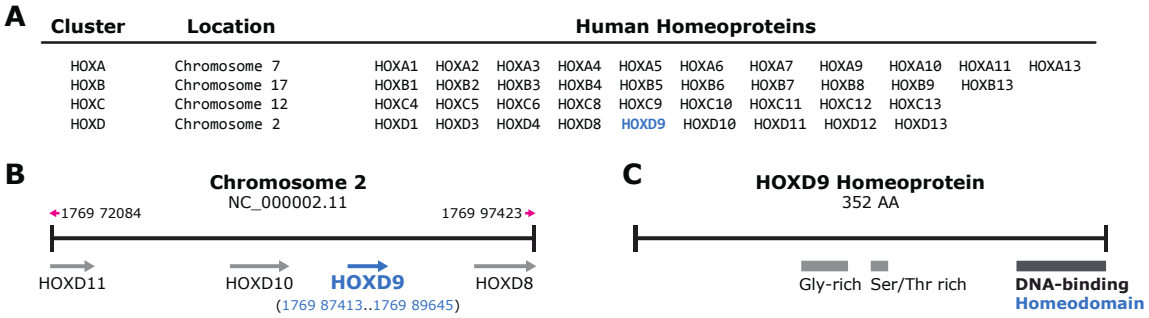



**Fig. 3.6) HoxD9 Gene Information and Conservation**

**A)** HoxD9 is one of many human homeoproteins spread across 4 gene clusters, each located on separate chromosomes. **B)** The HOXD gene cluster is located on chromosome 2 of humans; full length *hoxd9* is positioned between 8 and 9 and consists of 2232 bases. **C)** The functional HoxD9 protein 352 amino acids long, with its DNA-binding domain located at the C-terminus, consisting of roughly 60 amino acids.

The full HOXD9 protein is 352 amino acids long and contains a C-terminal DNA-binding domain of 60 amino acids, which is called the homeodomain. Across all multicellular organisms homeodomains are highly conserved due to the specific nature protein-DNA interactions and the conservation of regulatory elements in genes pertaining to embryonic morphogenesis[335,336]. This domain is structurally categorized as a helix-turn-helix motif, which is a class that invariably binds in the major groove of DNA. HoxD9 consists of three helices separated by flexible loops. Its third helix, the recognition helix, is long in length and inserts with a parallel orientation into the major groove. The remaining portion of HoxD9 beyond the homeodomain portion shown here consists of a flexible linker and a protein-association interface for transcription initiating complexes.

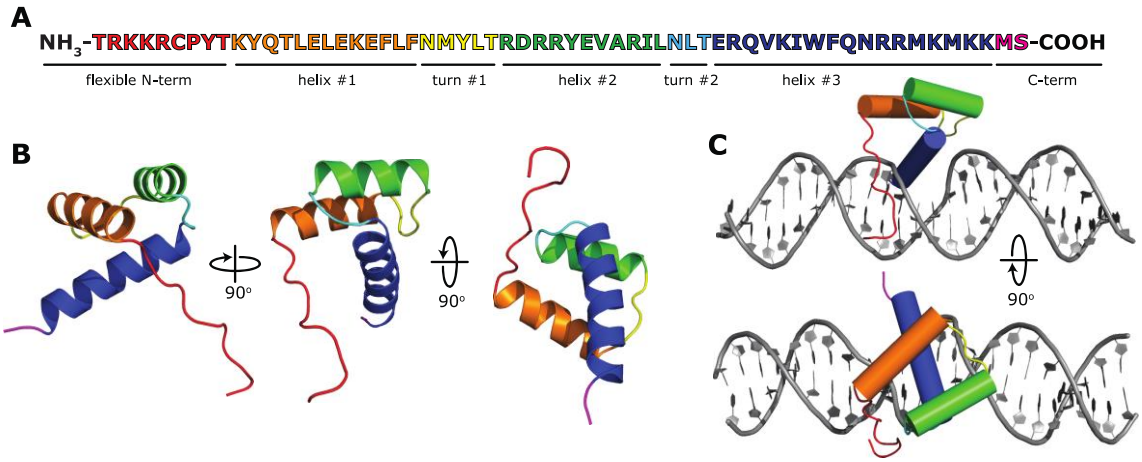



**Fig. 3.7) HoxD9 Homeodomain Helix-Turn-Helix Structure**

**A)** HoxD9 homeodomain protein sequence used in the studies, color-coded by region (helices and turns). **B)** Various perspectives of the 3D model of HoxD9 showing the helix-turn-helix motif. HoxD9 homeodomain is composed of three helices, two turns, and one flexible N-terminal region. **C)** Homology model developed for HoxD9 in association with the 24-bp duplex used in these studies (details of model covered in next section). The third helix (blue) is the "recognition helix", and inserts into the major groove. The flexible N-terminal region (red) binds in the adjacent minor groove.

## METHOD: BUILDING THE EXPERIMENTAL MODEL OF THE HOXD9-DNA DUPLEX

A 24-bp duplex containing the consensus HoxD9 binding site was used as the DNA construct in these studies. This exact complex has been investigated previously by Iwahara[315-317] and the dithioate work is an extension of these studies. At the center is a 12-bp recognition sequence, which is flanked on both ends by 6-bp. Because the crystal structure for this, or any other HoxD9 complex, has yet to be solved, structures of homeodomain homologs were relied on heavily for modeling purposes. These computational simulations described here were performed with the expertise of Dr. Miguel Elizondo-Riojas, a post-doc in the Gorenstein lab and resident theoretical chemist. Briefly, the homology model for HoxD9/24-bp DNA was constructed using the mouse HoxA9 homeodomain crystal structure (PDB: 1puf[337]) using the technique of amino acid side chain replacement followed by minimization, and then molecular dynamics (MD) simulation.

Mouse HoxA9 and human HoxD9 homeodomains are very similar, only six amino acids required substitution to generate the crude side-chain replaced HoxD9 structure. Next, the DNA duplex from our study was modeled in place of the shorter (20bp + 1 base overhang) duplex used in the HoxA9 structure, using a minimized

phosphate-backbone RMSD approach focused on the DNA-binding center portion. The additional base pairs were modeled onto the ends to complete the blunt-ended HoxD9 duplex. Side chain orientations, peptide backbone, and DNA torsional angles were then geometry optimized using basic energy minimization involving local quadratic approximation of the bond and angle potential energy surfaces[338]. This initial "crude" model of the HoxD9/24-bp DNA complex was used as a starting structure for a 100 ns MD simulation (details covered later). After the MD run an "average" structure was calculated, which was extracted and then energy minimized one last time to produce the final homology model for the HoxD9/24-bp system. This process was completed independently for both the normal PO duplex as well as the PS2 duplex. MD simulation for the PS2 required generation of force-field parameters for the dithiophosphate because it is a modified phosphate that is not included in the standard AMBER parameter set.

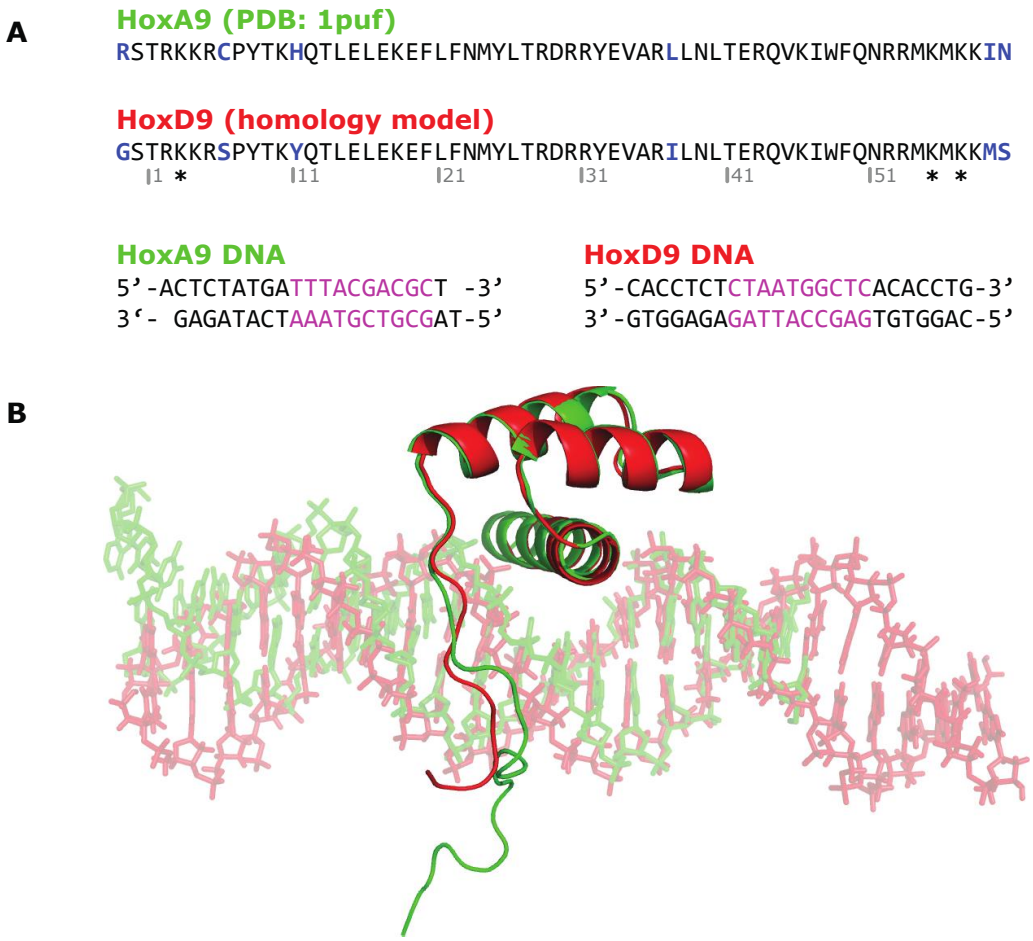



**Fig. 3.8) Homology Model of HoxD9 vs. Parent HoxA9**

**A)** Top: The amino acid composition of the HoxA9 protein from the crystal structure (PDB 1puf[337]) compared the HoxD9 sequence used in these studies. Only 6 amino acids needed to be changed (blue). Specific lysines in HoxD9 probed by NMR are highlighted with asterisks. Bot: The corresponding DNA sequences from HoxA9 and HoxD9. DNA binding sites are colored magenta. B) 3D carton representation of HoxA9 (green) and HoxD9 (red). DNA duplexes are drawn as transparent sticks, showing good alignment of the phosphates within the DNA-binding region.

## METHOD: MOLECULAR DYNAMICS SIMULATIONS AND DITHIOATE PARAMETERS

Although molecular simulations were not the focus of this dissertation, they were used in a supplementary manner to aid experiment planning and interpretation, so a general overview of the methods is needed. Importantly, the MD simulations considered explicit water based on the TIP3P model[339-342], which greatly increases the calculation time compared to an implicit continuum water solvent approximation. Therefore the simulations were performed at the Texas Advanced Computing Center (TACC) at the University of Texas at Austin running the latest Amber 12 software[343], rather than locally on laboratory machines. To prepare

the crude HoxD9 homology model (substituted side-chain, minimized) for MD the system's charge was first neutralized with 35 $Na^+$ atoms. An unrestrained 100 ns simulation was set up at 300 K using a fixed number of particles *N*, fixed volume *V*, and fixed temperature T, which is the canonical ensemble appropriate for simulating biological systems. The force-field used was parmbsc0, a refinement of the AMBER parm99 force field, that is optimized for nucleic acids[344]. The dithiophosphate is a modification not included in this release and requires its own set of custom parameters that must be calculated from scratch or borrowed from similar moieties, which will be covered in detail during the next section. Long-range electrostatics was accounted using the Particle-Mesh Ewald summation method, which helps stabilize trajectories for nucleic acids[345]. Inter-atom distances beyond the real space cutoff distance of 8 Å were ignored. The SHAKE algorithm[346] was enabled, which is a holonomic constraint for bond stretching that helps stabilize the trajectory by eliminating high-frequency bond length calculations. The time step over the 100 ns simulations was 2 fs, and each MD simulation took approximately 2 weeks to complete at the TACC.

The AMBER force field equation (Fig. 3.9) is composed of four terms describing the potential energy of the system: two-atom bond length energy, three-atom bond angle energy, four-atom torsional dihedral geometry energy, and two-atom non-bonding energy (van der Waals and electrostatics). Parameters describing these forces for the simulation of protein, nucleic acids, and organic molecules have been calculated and published in the highly cited "Cornell et al." force field, ff94[347]. Further enhancements were made, such as the ff99 force field[348], the optimization for nucleic acids[344] mentioned earlier, and various others[349]. Unfortunately, since the dithioated phosphate is a unique modification, its parameters are not included in these standard AMBER force fields and need to be determined using quantum mechanics calculations prior to MD simulation. 3D models obtained from a geometrical and vibrational DFT study of phosphorodithioate linkages[350] were considered as starting structures for calculations, after ions in their study were removed. Values for the P-S distance, and S-P-S and S-P-O angles were obtained by using a single-point calculation (Gaussian 03[351]) using HF/6-31G* level of theory (the same level of theory used for AMBER

force field parameterization). Charges were then fit to this electrostatic potential using RESP (AMBER program), assigning equal partial charges to equivalent atoms. This process was repeated for both *g-g-* and *g-t* sugar pucker and the charges, bond angles, and distances were averaged to produce the dithioate MD parameters.

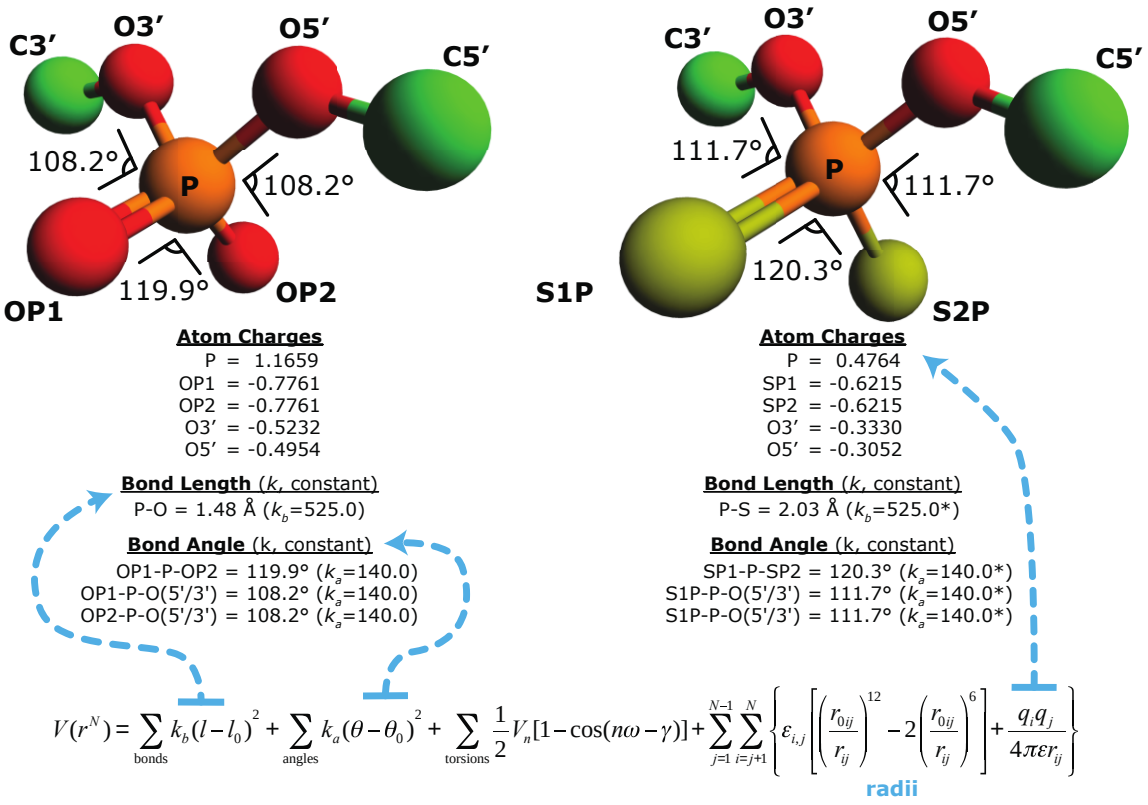



**Fig. 3.9) AMBER Force-Field Parameters for Dithioated Phosphate**

**(Top)** Calculated charges, bond lengths, and bond angles for normal (PO2) and dithioated (PS2) dimethyl phosphate. This is a suitable model for developing phosphate parameters for dithioate DNA AMBER MD simulations. QM calculations show that dithioation alters charge distribution over the phosphate, with a reduced negative charge on the nonbridging phosphoryl atoms (S1P/S2P vs. O1P/O2P). Sulfur in the phosphorodithioate PS2 is more polarizable compared to oxygen in the phosphate PO. In the PS2, electron charge density is more localized over the phosphorus atom. The bond length P-S is considerably larger than P-O, and bond angles are slightly larger. Force constants for PS2 indicated with an asterisk * were borrowed from the phosphate counterparts. **(Bot)** The AMBER potential energy function: The four terms used in the equation represent bond energies, bond angle energies, dihedral torsional energies, and non-bonded van der Waals and electrostatic energy. Blue arrows indicate the connection between parameters and the AMBER equation. Torsional informational (middle portion of Eq.) on a PS2 phosphate is not known, and values were borrowed from the PO.

Because of a lack of experimental or theoretical information, force field constants for PS2 dithioates were borrowed from PO phosphates, which are indicated with an asterisk in Fig. 3.9. This assumption should be minor as there is likely little appreciable difference between PO2 and PS2 bond angle force constants

that would affect MD simulations. However, a significant and perhaps poorer assumption was made in the case of the backbone torsion dihedral angles. Unfortunately there is no structural information on backbone dynamics of dithioates that provides information on BI-BII transition, which is rotation about the ε-ζ dihedral angle. Given the altered bond angles and larger P-S bond length, there is likely to be a subtle difference between PO2 and PS2 phosphates at this specific torsion that could potentially be revealed in MD. Determining these values experimentally would require an in-depth NMR analysis[352] suitable for its own manuscript, so it was decided to move forward by adopting the PO2 torsional angle constraints for PS2. This is a reasonable decision since torsional angles govern the BI-BII transition, the BI conformation is dominant in B-DNA duplexes[272], and dithioates also form B-DNA duplexes[353]. Nevertheless, since the most critical parameters for MD simulation are the charge, bond lengths, and bond angles, the parameters developed and presented here are suitable starting point for force field parameters that simulate dithioated phosphates.

## RATIONALE FOR DITHIOATION LOCATION IN THE SYSTEM

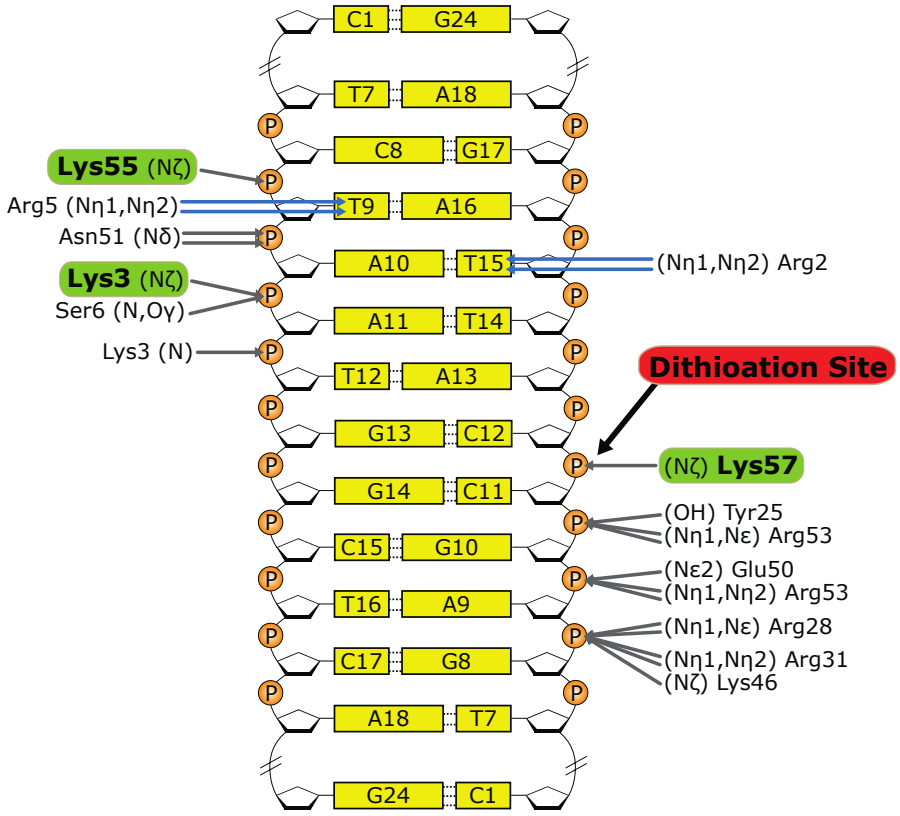



**Fig. 3.10) Protein-DNA Interaction Sites in the HoxD9/DNA Homology Model**

Interaction sites between HoxD9 and the 24bp duplex (Yellow) are determined according to measurements taken from the HoxD9/DNA homology model. Phosphate interactions are marked with grey arrows, and interior nucleobase contacts marked with blue arrows. An interaction is defined as having a hydrogen bond donor-acceptor distance < 3.2 Å. With this criteria, four lysines are shown as having contacts at 300 K (26.85°C), the temperature of MD simulations. The three lysines observed at the NMR temperature of 35°C are highlighted in bold. The specific phosphate interacting with Lys57, which was the site for dithioation in these studies, is identified.

The NUCPLOT[354] contact-mapping program inspired this image

The rationale behind choosing the specific phosphate for dithioation was determined according to Dr. Iwahara's previous NMR studies of the HoxD9/DNA complex which showed three high quality lysine $NH_3$ cross peak signals to work with (Lys3, Lys55, and Lys57) from the HISQC experiment at 35°C[314]. In total there are eight lysines in HoxD9, however at the elevated temperature of 35°C the remaining five lysine $NH_3$ cross peaks are broadened beyond detection, presumably because they are not involved in tight protein-DNA interactions and freely engage in hydrogen exchange with the solvent. At lower temperatures, the exchange rate is

reduced and these peaks begin to appear. Therefore it was reasoned that Lys3, Lys55, and Lys57 must be in direct contact with the DNA duplex, since their cross peaks are visible even at elevated temperatures. This presumption was confirmed by structural insight from the homology model for HoxD9/DNA, which allows measurement of expected $NH_3$ Nζ-P distances in the complex. Lysine 57 was chosen as the target for site-specific dithioation, which according to the homology model puts two sulfur atoms in place of non-bridging phosphoryl oxygens between the two cytosines in the HoxD9 binding site (5'-TAATGG-3'/3'-ATTA**CC**-5') near the middle of the 24-bp duplex. Hereafter, the terms "dithio", "dithioate", and "PS2" are used interchangeably to represent this single-dithioate containing duplex modified at the phosphate interacting with Lys57. Likewise, the terms "normal", "PO", and "PO2" are used as shorthand notation for the unmodified 24-bp HoxD9-specific duplex.

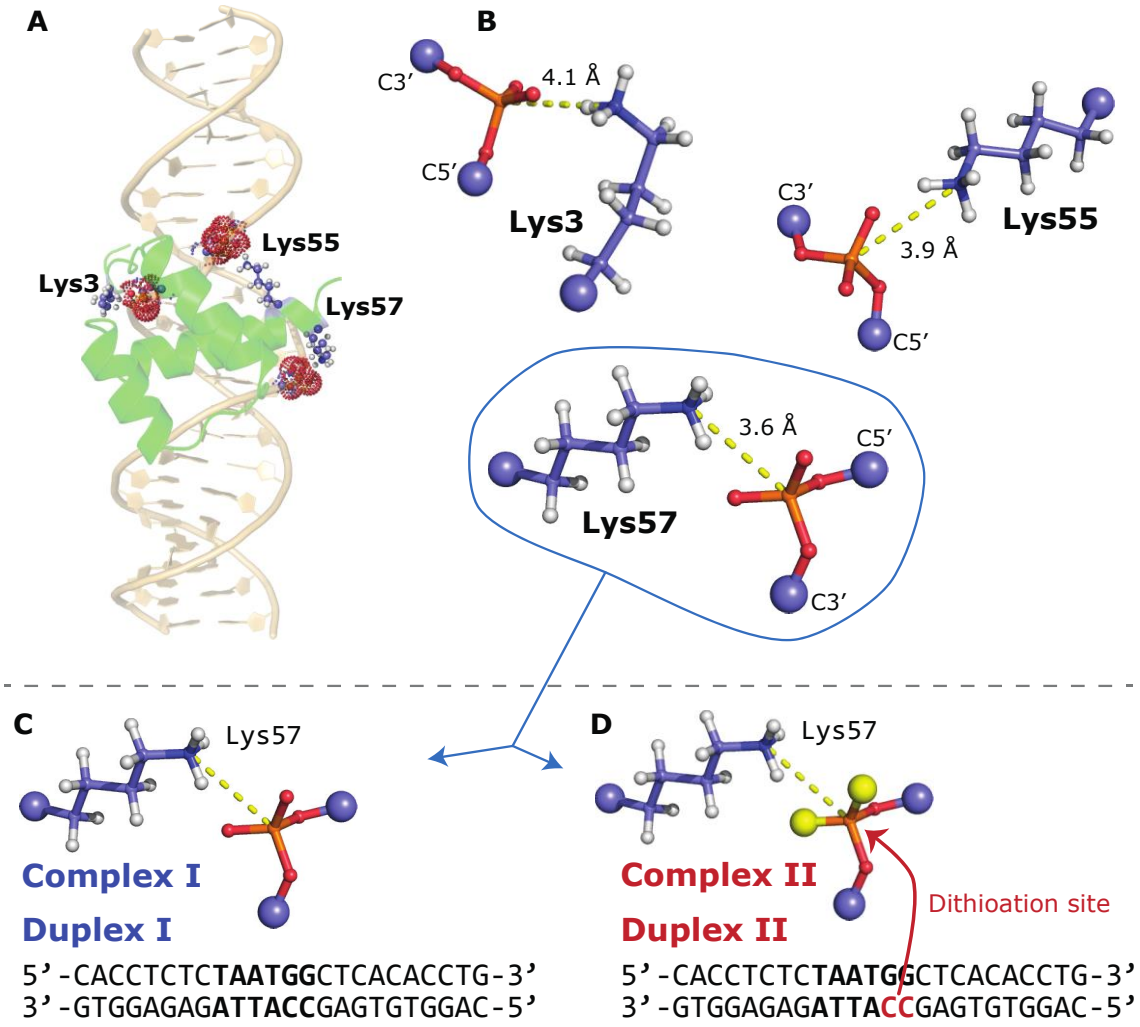



**Fig. 3.11) Focus on Three Lysines Residues in the HoxD9-DNA complex**

**A)** Cartoon representation of the modeled complex, with the three lysines studied shown as colored sticks and the interacting DNA phosphates as dotted spheres. The phosphate interacting Lys57 was dithioated in these studies. **B)** Close up view of the three lysines studied and their respective phosphates. The measured nitrogen-phosphorus distances are shown. **C)** Complex I contains a normal phosphate at the location interacting with Lys57. **D)** Complex II contains a dithioated phosphate at the location interacting with Lys57.

## METHOD: HOXD9-DNA COMPLEX SAMPLE PRODUCTION

HoxD9 was obtained in purified form from the lab of Dr. Junji Iwahara following standard production procedures for recombinant proteins, using methods published previously[315,316,319]. Unmodified DNA strands were purchased from Integrated DNA Technologies and came with QC diagnostics. The DNA strand containing the phosphorodithioate moiety at the site for Lys57 interaction was synthesized according to methods discussed in Chapter 1, and purified using

reverse-phase HPLC purification via a 5′-dimethoxytrityl group. ESI-MS analysis confirmed oligo identity and successful incorporation of both sulfur atoms. Equimolar concentrations of complementary strands were mixed in moderate salt annealing buffer (10 mM TE, 100 mM NaCl, 1 mM EDTA) and heated at 90°C for 5 minutes. Annealing occurred over an hour long cooling at room temperature. Due to small uncertainties in oligo concentration determination using UV spectroscopy, minor single-stranded DNA excesses were eliminated using cation-exchange chromatography to produce pure oligo duplexes. Annealed oligos were injected onto a Mono Q 10/100 column (GE Healthcare) pre-equilibrated with running buffer (10 mM TE, 100 mM NaCl, 1 mM EDTA). A gradient using a similar high salt buffer (1M NaCl) was used to elute the purified duplexes. As discussed in the first chapter, dithioated oligos have a tendency to produce curious looking purification runs, and the same held true here with cation exchange. Even though >98% purity single strands were annealed, the dithioate chromatograph appears remarkably complex. Therefore, another quality test of the annealed duplexes was performed by native PAGE, in which 1 μg of oligo material was loaded. The resulting bands for both the PS2 and PO duplex were devoid of single strand contamination and both prepared duplexes were assumed pure.

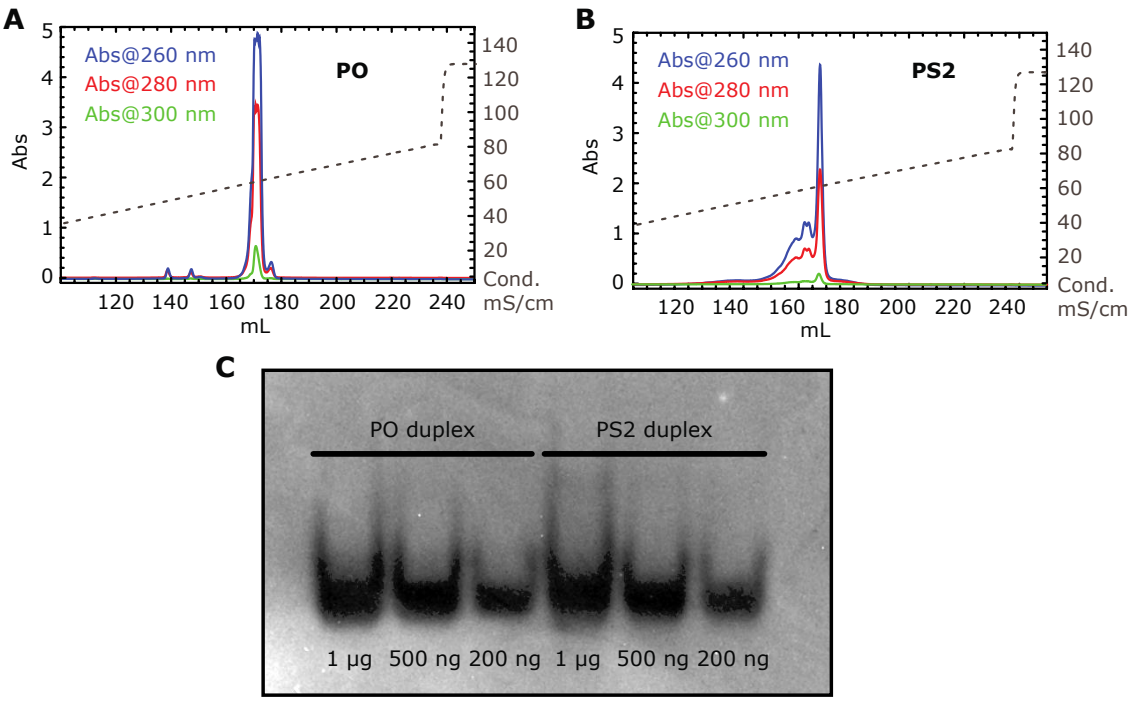



**Fig. 3.12) Purification of Annealed Duplexes and QC by Native PAGE**

After annealing, the PO and PS2 duplexes were purified by cation-exchange chromatography. **A)** PO duplex showed an ideal purification run, the largest peak was collected as the final sample. **B)** The dithioate produced a complex purification chromatogram, which is very surprising given that both strands were QC'd exceeding >98% purity prior to annealing. Peculiar looking column runs appear to be the normal course for dithioated molecules. **C)** Regardless, a native PAGE run with up to 1 µg oligo loaded shows very pure duplexes devoid of single strands for both PO and PS2 samples.

## 3.3) Biophysical Studies of HoxD9 and HoxD9-DNA Interactions

### INTRODUCTION: ASSAYING COMPLEX FORMATION USING CIRCULAR DICHROISM

A frequent occurrence with helix-turn-helix proteins is cooperative folding of the recognition helix upon DNA association. With homeodomains some show this tendency, but others do not, it really is a mixed bag. In the case of HoxD9 there has not been published information on this matter, but this question must be analyzed since cooperative folding of the recognition helix complicates the analysis for other sections.

Circular Dichroism (CD) was used to monitor structural changes between free protein and complexes with DNA to determine whether α-helix formation within HoxD9 occurs upon cooperatively upon binding, or whether the motif is fully folded prior to binding. When light enters matter its properties of intensity, polarization, wavelength and velocity will change. Biological molecules contain numerous chiral centers that absorb right- and left-circularly polarized light to varying degrees due to differences in absorption extinction coefficients for each polarized ray. This orientation-specific difference is called circular dichroism (CD), and is a powerful technique for monitoring conformational changes in macromolecules. The far-UV wavelength region of 230-178 nm is particularly useful for observing the effects of conformational changes of the protein backbone, while the near-UV region (>250 nm) is used for monitoring tertiary changes. CD spectra should be interpreted as a linear combination of spectral contributions from each type of secondary structure (e.g. α-helix, β-sheet, turns, random coils, etc…) weighted by their abundance in the protein. Of these, CD is most accurate for helical secondary structures because they are homogenous by classification, have well-defined angles, and produce similar spectra regardless of sequence composition. So being a helix-turn-helix motif, HoxD9 is well suited for structural observation by CD.

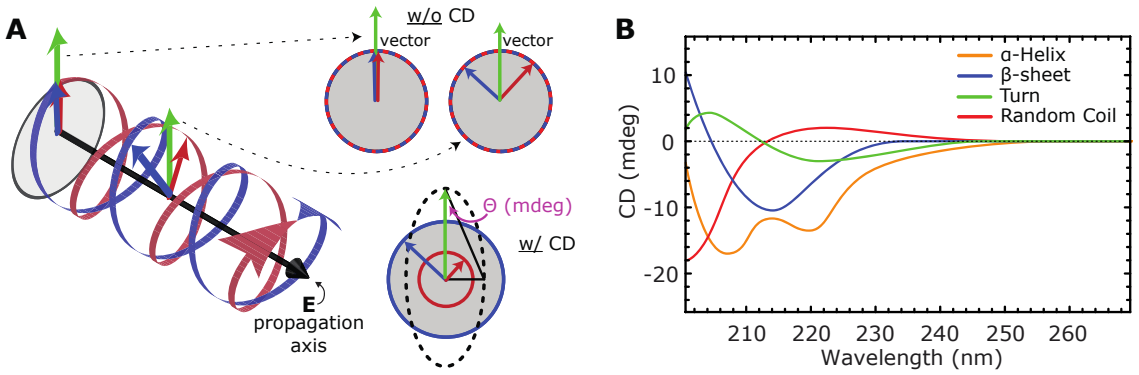



**Fig. 3.13) Principles of Circular Dichroism**

Electromagnetic radiation (light) is a complex waveform that can be interpreted as having two perpendicular components: the electric (E) and magnetic (M). Because they are perpendicular, it is sufficient to consider only the E-component in describing the wave. In circularly polarized light the magnitude of the electronic component is constant but the direction of oscillation changes as the wave progresses. **A)** CD spectrometers produce light that is polarized in both the right (blue curve) and left (red curve) directions. The direction of oscillation (red arrows, blue arrows) can be in any direction that is perpendicular to the direction of propagation (black cone-tipped arrow). The left- and right-polarized radiation sums to a vector (green arrow). In the absence of absorption (**top circles)** the magnitudes of counter-rotating E-components are equal. In biological samples the molar extinction coefficients for right- and left-polarized light are unequal (**bottom circles**), resulting in a decrease in the magnitude of one oscillation vs. the other (red inner circle vs. blue outer). This is the definition of circular dichroism (CD). The measured ellipticity is defined as the tangent of the ratio of the minor to major elliptical axis formed (pink, mdeg). **B)** Canonical CD spectra of secondary structure components in the wavelength region used in these studies. This information should be referred to when analyzing CD results in the next section.

## METHOD: CIRCULAR DICHROISM

Purified oligonucleotide duplexes (PO and PS2) and HoxD9 were dialyzed into experimental buffer using rigorous buffer exchange, and samples were prepared at 10 μM concentration. Experiments were performed on a JASCO J-715 Circular Dichroism (CD) Spectropolarimeter using a 0.1 cm round eye CD cuvette. Data was collected at 0.5 nm resolution using a band width of 0.5 nm, a sensitivity of 20 mdeg, a response time of 0.5 seconds, and scanning speed of 100 nm/min. The first channel was set to CD mode (mdeg) and the second channel was set to high-tension voltage (HT) to monitor abnormal wavelength-dependent electrical gain throughout the measurement. The wavelength window chosen was 200-270 nm, which was as low as the machine could reliably produce a CD signal from the HT voltage. Below 200 nm the signal became too noisy for the concentrations of solutes and macromolecules in the buffer. This situation was not ideal since there is

significant structural information in the 180-200 nm range, but I was unfortunately limited by the quality of the spectrometer. Modern machines have more advanced nitrogen purging and are capable of reaching these wavelengths, but a machine with this capability was not available. The region beyond 270 nm was not collected because this range is dominated by oligo signal, which masks any information on HoxD9 secondary structure. Data was processed using ellipticity units (mdeg) rather than converting to molar ellipticity or mean residue molar ellipticity. because the purpose of these studies was to ascertain whether HoxD9 undergoes significant α-helical formation upon binding, which could be accomplished using a linear summation of duplex, complex, and free protein measurements in mdeg units, if the concentrations are kept equal.

## RESULT: CD REVEALS HOXD9 BINDING TO DNA IS NOT ACCOMPANIED BY HELIX FORMATION

CD experiments showed that a single dithioation of the 24-bp duplex did not alter the overall structure of the duplex, as both spectra were near-perfect overlays. A similar result was found for the PO vs. PS2 complex, with identical overlays. This was encouraging, as any difference between the "complex" spectra would indicate a gross deviation of the association model between the two systems. Next, an "in-complex" HoxD9 spectrum was calculated by subtracting "free" HoxD9 protein data from the "complex" spectra. The in-complex spectrum revealed no significant difference in HoxD9 structure occurring upon complex formation. This is an important finding, as formation of the recognition α-helix upon binding the major groove is common event that occurs with homeodomains. Such a finding using CD analysis was reported for homeodomain homologues MATα2[355], as well as NK-2[356] where significant differences were shown between their in-complex and free spectra in the 205-225 nm region. These manuscripts are recommended for those interested in what α-helix formation upon DNA binding looks like when observed by CD. However, my data suggests that unlike these homeodomains, HoxD9 is a fully formed helix-turn-helix motif prior to association with DNA.

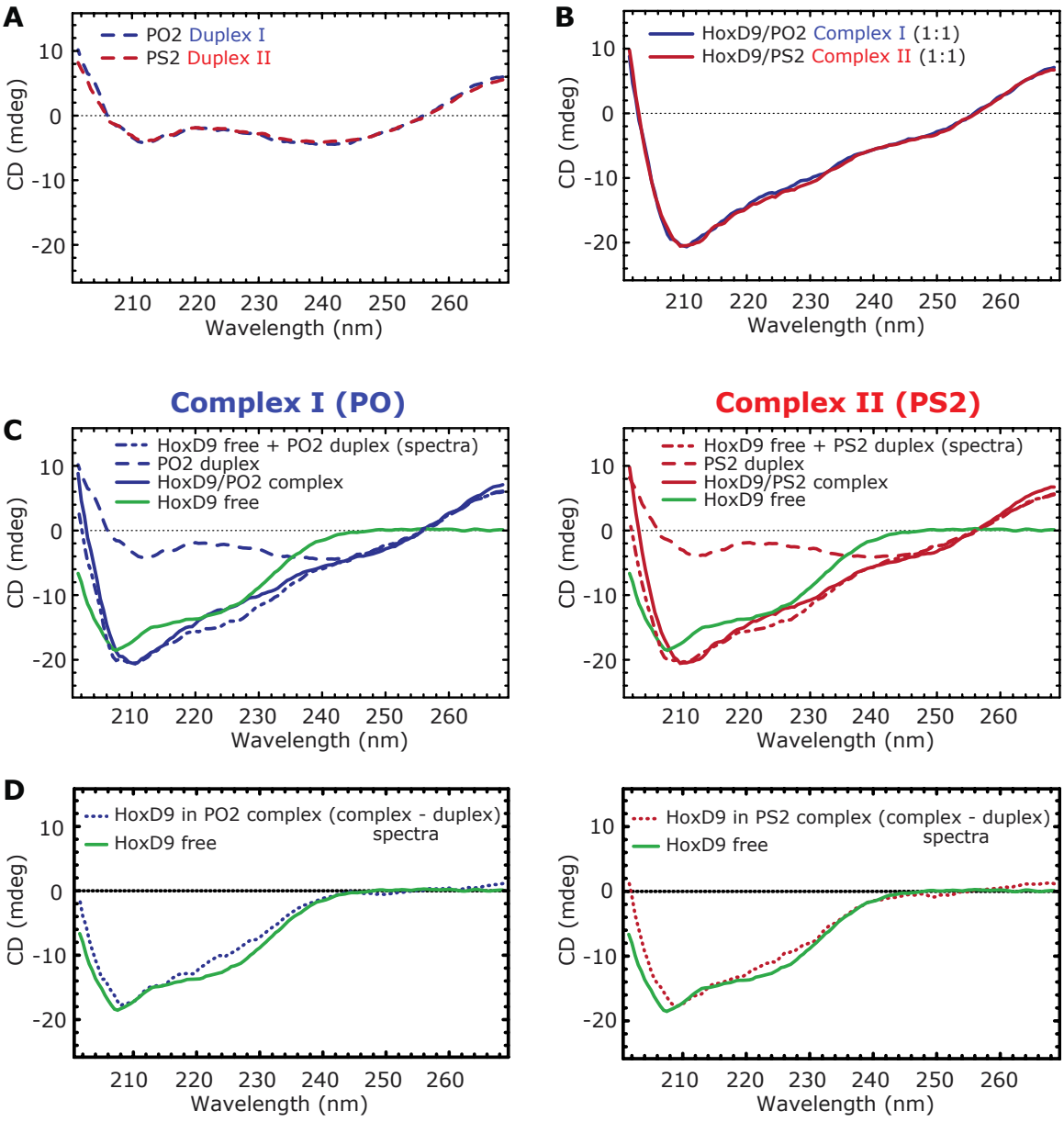



**Fig. 3.14) CD Spectra of Free HoxD9, HoxD9 Complexes I & II, and PO & PS2 Duplexes**

**A)** Both the PO and PS2 duplexes show identical CD spectra, indicating stability of the duplex despite dithioation. **B)** The PO and PS2 complexes (1:1, protein:DNA) both produce identical CD spectra, indicating similarity of the two associated complexes. **C)** Comparison of the complex (solid blue/red) with free HoxD9 (green) can be made after addition of contributions from the duplex (dashed blue/red), into a linear combination of spectra (hashed blue/red). **D)** Determination that HoxD9 does not undergo α-helical formation upon protein binding: By subtracting the duplex contributions from the complex spectra, a calculated "HoxD9 in complex" spectra is obtained, which appears very similar to the "free" HoxD9 spectra. If formation of α-helices had occurred, the "free" HoxD9 data would be well above the calculated "in complex" spectra at the wavelengths indicated in Fig. 3.14.B (~205-225 nm). Although a small difference is seen here between the two, it is in the wrong direction and of such a small magnitude that it is likely due to poor accuracy of the machine at the utilized scanning rate, which is exacerbated by the assumption of true linear summation of spectra components. Therefore, the two spectra ("in complex" and "free") should be interpreted as so similar that no major structural changes are occurring upon binding.

## INTRODUCTION: ASSAYING HOXD9 THERMAL STABILITY USING INTRINSIC FLUORESCENCE

Many homeodomains at human body temperature are partially denatured in apo form and fold upon DNA binding. Previously it was shown by CD (Fig. 3.14) that HoxD9 does not show significant structural changes between apo and bound forms, but these experiments were performed only at ambient room temperature. Some of the NMR experiments performed later in this dissertation take place at 35°C, which is near the human body temperature and is potentially thermally active enough to destabilize a helix-turn-helix motif. A stability assay for HoxD9 over a wide temperature range was pursued that utilized intrinsic fluorescence within the protein.

Tryptophan, tyrosine, and phenylalanine are aromatic amino acids that fluoresce when excited with UV light. The shape and magnitude of this intrinsic fluorescence is dependent on the local environment surrounding these amino acids, which makes intrinsic fluorescence an ideal non-labeled technique for monitoring the folding state of a protein. As the protein unfolds (denatures), previously buried aromatic acids become exposed to the solvent and fluorescence increases due to lack of quenching from nearby interactions. Denaturation in protein samples is typically induced using the following methods: thermally, by heating the sample to break hydrogen bonds stabilizing tertiary and secondary structure, chaotropically, using chemicals such as urea, guanidinium chloride, or lithium perchlorate, and finally by addition of strong acids or organic solvents. Since the stability of HoxD9 at human body temperatures was in question, an experiment was performed using thermally-induced protein denaturation coupled with UV spectroscopy to determine the melting temperature and stability of HoxD9 homoeodomain. With eight intrinsically fluorescence amino acids, HoxD9's unusually high percentage of fluorophores make the UV technique a great approach for assaying structural characteristics.

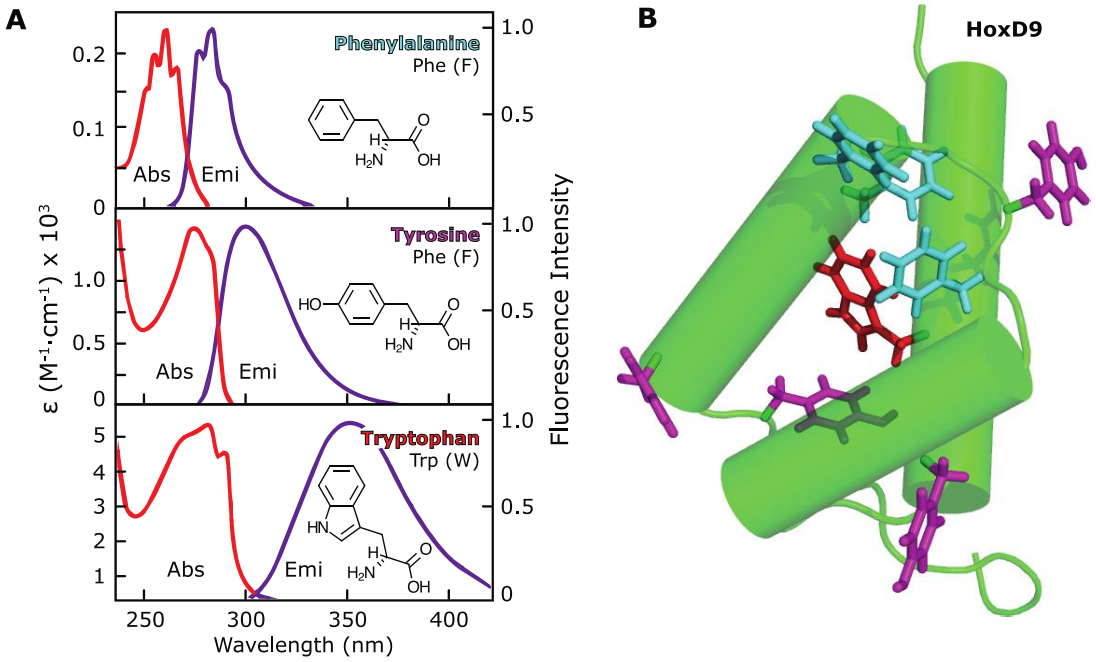



**Fig. 3.15) HoxD9 Intrinsic Fluorophores**

**A)** The absorbance and emission spectra for the three intrinsically fluorescent amino acids are shown. Tryptophan fluoresces the strongest, followed by tyrosine and then phenylalanine (extinction coefficients values on left axis). This figure is reproduced from an image in Lakowicz's book[357]. **B)** HoxD9 contains many fluorescent amino acids: three phenylalanines, four tyrosines, and one tryptophan. Eight of 60 amino acids in HoxD9 are fluorescent, which is an unusually high percentage for a protein, making intrinsic fluorescence an ideal technique.

## METHOD: HOXD9 THERMAL MELTING

Fluorescence measurements were performed using an ISS PC-1 spectrofluorometer in a fluorescence-grade (four polished sides) 1 cm x 1 cm quartz cuvette. With a lamp electrical current of 19 amps, the fluorescence emission spectrum was collected at 20°C using an excitation wavelength of 280 nm. The emission spectrum was recorded over a range of 270-400 nm at an interval of 1 nm and scanning rate of 1 nm/sec. Under these parameters a 1 μM concentration of HoxD9 in Experimental Buffer (100 mM NaCl, 20 mM sodium phosphate, pH 5.0 (120 mM total [Na+]) gave strong fluorescence intensity and signal-to-noise ratio. Values were scanned 10 times and the mean presented as the final data point. A maximum fluorescence emission wavelength of 303 nm was determined and utilized for observation in fluorescence melting experiments. Measurements were made inside a sealed quartz cuvette that was thermostated by a Peltier assembly using a temperature range of 3°C-80°C at a rate of 0.7°C/min and interval of 0.5°C.

To prevent atmospheric water condensation on the outside of the cuvette below room temperature, 0.22 μm-filtered air was continuously blown at moderate velocity over all faces. This did not affect the temperature of the solution, as determined by comparing the Peltier block temperature with a digital thermometer probe placed inside the cuvette in a separate test. Each data point represents the mean of 10 fluorescence intensity readings. Raw data was processed according to the baseline method, where regions representing folded and unfolded states are linearly fit and an "α" term defining the fraction of folded protein is calculated by subtracting the baselines. This is identical to the equilibrium melting analysis presented by Marky and Breslauer[242], which was discussed in the previous chapter. This α plot was smoothed via a Savitzky-Golay filter using a cubic polynomial and a moving window of seven data points. A function was then interpolated between the smoothed data points and HoxD9's melting temperature of 48.8°C was calculated as the point of mid transition by solving where the interpolated α function equals 0.5, the temperature at which the population is 50% folded and 50% unfolded.

### RESULT: HOXD9 THERMAL STABILITY

Fluorescence intensity diminished as the sample was heated until the protein started to unfold, at which point there was a sigmoidal increase in fluorescence indicative of protein denaturation. This melting process consisted of a single transition, and was found to be reversible. Using the baseline method an "α"-fraction folded plot was generated and the melting temperature of free HoxD9 was determined to be 48.8°C. The data suggests that at 35°C, the temperature used for NMR experiments, HoxD9 is at least 95% folded. It's likely that the small degree of unfolding (remaining 5%) is due to the intrinsically disordered N-terminal region of the protein, which binds the protein. Nevertheless, binding to DNA duplexes greatly enhances the transition temperature for protein denaturation so it can be assumed that the protein is fully folded when associated with DNA at all experimental temperatures used in this dissertation.

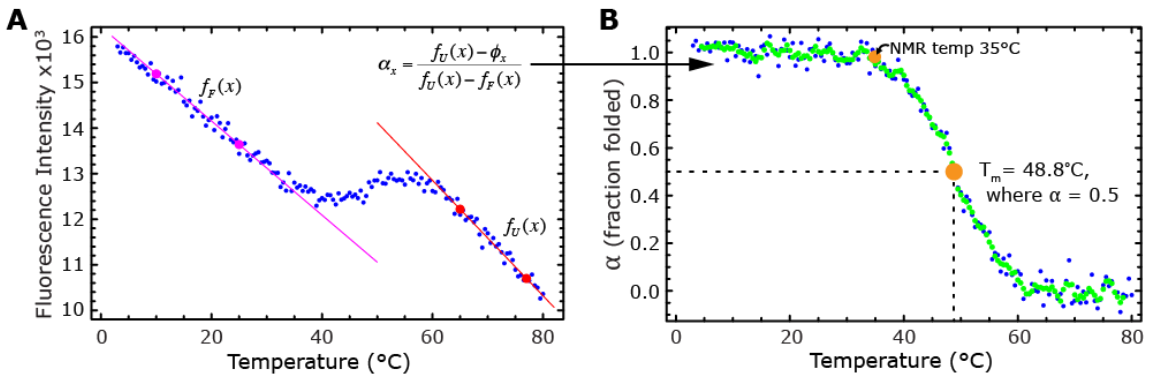



**Fig. 3.16) Thermal Denaturation of HoxD9**

**A)** Raw intrinsic fluorescence data (blue points) obtained for the thermal melt of HoxD9. Regions associated with the folded and unfolded state were fit to linear functions and values were converted to an "α" plot (fraction folded) plot by baseline subtraction. **B)** After data smoothing (green points), the melting point was calculated as the temperature at which α = 0.5 (50% folded), which in this case with HoxD9 was 48.8°C.

## INTRODUCTION: BINDING AFFINITY OF HOXD9 WITH PS2- AND PO-DNA

Quantifying the affinity of HoxD9 with both duplexes is an important aspect of characterizing the effects of site-specific dithioation on protein association. A solution-based approach was pursued utilizing fluorescence anisotropy measurements. In this technique, anisotropy of a rhodamine fluorophore attached to the 5'-end of one strand in the duplex is measured as function of protein concentration. This value increases in a sigmoidal fashion as protein is added, indicating that protein is binding to the fluorophore-duplex. Anisotropy increases because of the tumbling rate for the macromolecular complex is lower than that of the free fluorophore, because of the difference in hydrodynamic volume.

Anisotropy measurements begin with excitation of the sample by plane-polarized light at the wavelength of the fluorophore's maximum absorbance. A second set of polarizers resolve the fluorescence emitted into its horizontal and vertical components and the intensity is quantified. Anisotropy $\langle r \rangle$ is calculated using the following relation:

$$\langle r \rangle = \frac{I_{VV} - G * I_{VH}}{I_{VV} + 2 * G * I_{VH}}, \text{ where } G = \frac{I_{VH}}{I_{HH}}$$

In this equation, the notation $I_{VH}$ refers to the intensity of light that has passed through the initial vertical polarizer, the sample, and then the final horizontal polarizer before hitting the detector. $I_{VV}$, $I_{HH}$, etc… can be deduced, and G

stands for the g-factor that is determined at the start of each experiment. Anisotropy values vary in amplitude between 0.4 and –0.2 for samples in which the molecules are oriented randomly, and increases upon complex formation due to an enlargement of the hydrodynamic radius that reduces the random motion of the fluorophore-coupled macromolecule.

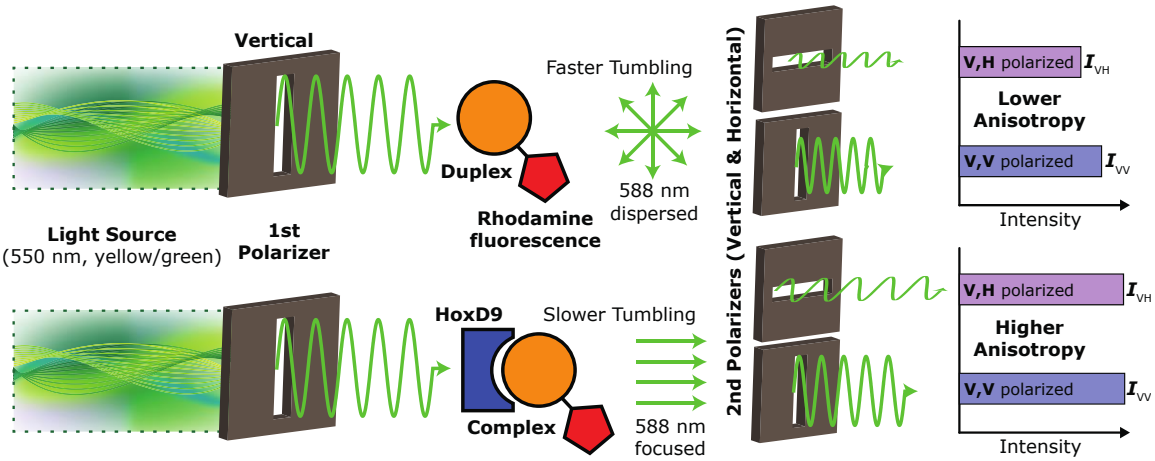



**Fig. 3.17) Principles of Fluorescence Anisotropy**

Light of a fluorophore-specific excitation wavelength is linearly plane-polarized using mechanical filters before reaching the sample. The fluorescence emitted, when quantitated by the vertical and horizontal intensities, is used to calculate the anisotropy value of the sample. Unbound fluorophores tumble faster in solution than when associated in complex, due to hydrodynamic forces, and have lower anisotropy vales. Anisotropy increases sigmoidally until binding is saturation.

## METHOD: FLUORESCENCE ANISOTROPY MEASUREMENTS

The same ISS PC-1 spectrofluorometer was used for anisotropy measurements with similar power settings as previously described. A 5'-amine-linked rhodamine-modified 24-base "top" strand was purchased from Integrated DNA Technologies and arrived HPLC purified, with accompanying QC reports. The PO and PS2 duplexes were annealed as described previously and then purified using native PAGE in mini-cassette form. The fluorescent band representing the duplex was excised and eluted from the crushed gel fragments in annealing buffer using overnight incubation. The experimental design involved a 2 mL solution of a fixed concentration of duplex in the cuvette that was titrated with concentrated HoxD9 protein to achieve final concentrations ranging from 0.1 nM to 1 μM. Initial experiments showed the association to be high affinity (<50 nM), so the duplex concentration was then optimized at 4 nM to keep the probe concentration below

the $K_d$. of the highest affinity measured. The prerequisite for reasonable anisotropy signals is a good signal-to-noise ratio and I found 4 nM to be at the lowest concentration of probe that could be used effectively. Even so, the deviation (σ) of measurements was quite large at this low concentration, reaching as high as 20% (~0.01 units) of the difference between bound and unbound anisotropy values (~0.05 units). Therefore, 20 acquisitions were made to improve the standard error according to the relation $SE_{\text{mean}} = \sigma/\sqrt{N}$, where $N$ is the number of measurements. Mean anisotropy values were plotted as a function of HoxD9 concentration and the data was fitted to the model:

$$f(x) = \langle r_{\text{free}} \rangle + \left( \frac{\langle \text{r}_{\text{bound}} \rangle - \langle r_{\text{free}} \rangle}{2c} \right) \cdot \sqrt{(K_{\text{D}} + x + c) - (x + c + K_{\text{D}})^2 - 4c \cdot x}$$

where $x$ is the HoxD9 concentration values, $f(x)$ is the observed anisotropy values, $c$ is the concentration of duplex used (4 nM), and the three parameters fitted are $\langle \text{r}_{\text{bound}} \rangle$, $\langle r_{\text{free}} \rangle$ , and $K_D$. In the analysis a weighted fitting procedure was used so that the measurement error could be considered, as opposed to just fitting the mean values. Each data point was weighted using $1/\Delta y_i{}^2$, where $\Delta y_i$ is the standard error for mean $y_i$ value. Parameter fitting standard errors were also computed from these weights, rather than an estimated variance scale using the sum of squares of the residuals. The binding experiments were performed at 20°C using a thermostated Peltier assembly. At higher temperatures, specifically 35°C, the fluorescence intensity of the probe dropped significantly and the standard deviation of measurements became too large. Attempts to increase the probe concentration higher than 4 nM, but below 10 nM, were unsuccessful in improving this problem so near-room temperature measurements were all that could be captured. Each experiment was completed in triplicate and a mean $K_D$ was calculated for HoxD9 affinity to both the PO and PS2 duplexes.

### RESULT: PS2 – LYS57 COMPLEX HAS IMPROVED BINDING COMPARED TO PO – LYS57

The mean dissociation constant for the PO duplex to HoxD9 was determined to be 33 ± 8 nM, and the PS2 was 13 ± 4 nM, a 2½-fold difference. From this $K_D$

data the free energy difference of the dithioate complex can be calculated according to:

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ, \qquad \Delta G^\circ = -RT\ln K_A, \qquad K_A = 1/K_D$$

$$K_{D,\,PO} = 39.9\text{ nM} \;\; \therefore \Delta G_{PO} = -10.0\text{ kcal·mol}^{-1}$$

$$K_{D,\,PO} = 12.6\text{ nM} \;\; \therefore \Delta G_{PO} = -10.6\text{ kcal·mol}^{-1}$$

$$\Delta\Delta G^\circ_{PS2} = \Delta G^\circ_{PO} - \Delta G^\circ_{PO} = -0.6\text{ kcal·mol}^{-1}$$

The difference between free energies of binding is calculated to be $\Delta\Delta G_{PS2} = -0.6\,\text{kcal}\cdot\text{mol}^{-1}$. This difference will manifest itself in either the enthalpic term, $\Delta\Delta H_{PS2}$, and the entropic term, $-T\Delta\Delta S_{PS2}$. Insight into hydrogen bond enthalpy for sulfur and oxygen has been recently been published using a dimethylsulfide-methanol (DMS-MeOH) and dimethyl ether-methanol (DME-MeOH) system[358]. Experimentally using IR spectroscopy the authors determined H···S, and H···O hydrogen bond enthalpy to be -3.5 kcal·mol$^{-1}$, and -4.5 kcal·mol$^{-1}$, respectively, yielding a $\Delta\Delta H_{DMS-MeOH}$ of +1.0 kcal·mol$^{-1}$. The same system was analyzed theoretically using statistical thermodynamics and quantum mechanics and they calculated a $\Delta H_{DMS-MeOH}$of -4.6 kcal·mol$^{-1}$ and $\Delta H_{DME-MeOH}$ of -4.9 kcal·mol$^{-1}$, for a $\Delta\Delta H_{DMS-MeOH}$ of +0.3 kcal·mol$^{-1}$. The authors' data together implies that a sulfur-for-oxygen substitution should be thought of as nearly equivalent in the magnitude of hydrogen-bond enthalpy, with the sulfur being approximately +0.5 kcal·mol$^{-1}$ (less favorable). Applying this interpretation to my data for a single dithioate-lysine interaction, the $\Delta\Delta G_{PS2} = -0.6\,\text{kcal}\cdot\text{mol}^{-1}$observed in binding studies can be suggested to arise from the entropic component $-T\Delta\Delta S_{PS2}$.

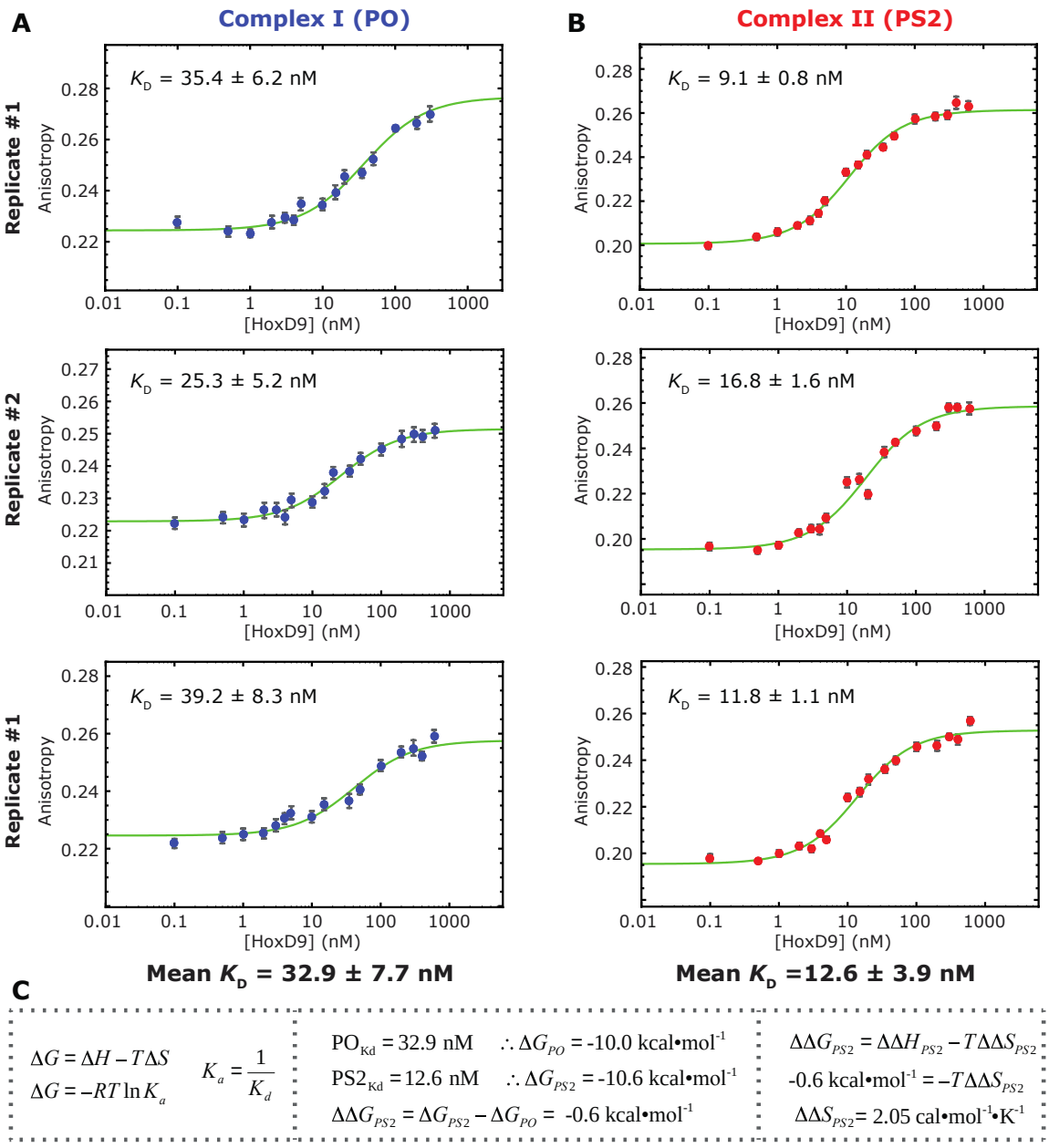



**Fig. 3.18) Binding Affinity Determination Using Fluorescence Anisotropy**

Fluorescence anisotropy was used to determine affinity constants for the two duplexes with HoxD9. A rhodamine fluorophore was attached to the duplex, which remained at a constant concentration while HoxD9 was titrated into the cell. Twenty acquisitions on anisotropy values were used for each datapoint and the standard error of each measurement, as described in the methods, is indicated by the error bars. The data were fit to a 3-parameter model (Kd, free value, bound value) that considered the oligo probe concentration. Binding data suggests a 2.5-fold increase in binding affinity upon dithioation of the phosphate interacting with Lys57, resulting in a $\Delta\Delta G_{PS2}$ of -0.6 kcal·mol$^{-1}$.

## INTRODUCTION: ISOTHERMAL TITRATION CALORIMETRY

An attempt was made to decompose the observed free energy difference from anisotropy binding assays ($\Delta\Delta G_{PS2} = -0.6\ \text{kcal·mol}^{-1}$) into its enthalpic and entropic component using Isothermal Titration Calorimetry (ITC), which is a direct measurement of binding enthalpy[359]. In the ITC apparatus there are two identical cells, one containing the sample (receptor) and the other containing buffer alone. In the sample cell there is an injection syringe containing a ligand. During an experiment known concentrations of ligand are titrated from the syringe into the sample cell, where a binding reaction occurs which produces some quantity of thermodynamic heat. Sensitively calibrated thermocoupled circuits attached to both cells detect temperature differences between the two cells. Binding processes are either exothermic or endothermic, and the electrical power (measured in $\text{J·s}^{-1}$) needed to maintain temperature equivalency between the two cells is quantified. There is a direct relation of the power applied to the reference cell during injection $i$ and heat of interaction of the $i^{\text{th}}$ injection, $q_{i,app}$, which is integrated with respect to time to yield the apparent heat $\Delta q_{i,app}$. The observed $\Delta q_{i,app}$ encompasses the heat of both specific (binding) and non-specific (dilution) processes, $\Delta q_{i,app} = \Delta q_i + \Delta q_{i,ns}$. In ITC experiments the $\Delta q_{i,ns}$ heat of dilution is determined using titrations of the ligand into buffer alone, or the heat of excess titrations after the ligand has fully saturated the available binding sites. A plot of $\Delta q_i$ vs. the mole fraction of ligand:receptor is plotted and can be fit using non-linear regression analysis with various binding models to determine numerous parameters from a single titration experiment: at temperature $T$ the $n$, stoichiometry; $K_a$, the association constant, and $H_{app}$, the apparent enthalpy of the interaction.

The HoxD9 system used here has an affinity that lies on the outer edge of interactions that be reliably fit all three parameters using this approach, so instead ITC was used to determine just $H_{app}$. ITC provides a direct measurement of binding enthalpy, as opposed to other methods that rely on indirect calculation of enthalpy using the van't Hoff relation (Ch. 2). Sometimes significant differences between van't Hoff enthalpies and enthalpies determined using calorimetric methods are

seen[360]. Since ITC is capable of operating over a wide temperature range it was able to provide binding information at temperature of 35°C, which is the temperature of NMR experiments. The anisotropy measurements were not capable of operating at this temperature due to fluorescence degradation. Using ITC the energetic signature for HoxD9-24bp DNA system was probed over a wide range of temperatures, producing a thermodynamic landscape of this protein-DNA interaction.

### METHOD: ITC PILOT BINDING STUDIES

It is critical in ITC that buffer composition of the receptor and ligand be identical to minimize non-specific heats arising from protein dilution. The purified HoxD9 protein as well as the PO and PS2 duplexes were simultaneously dialyzed into experimental buffer, which was identical to the buffer used for anisotropy binding studies (20 mM sodium acetate pH 5.0, 100 mM NaCl). After dialysis, samples were thoroughly degassed using the ThermoVac system operated with a spinning mini magnetic stir bar in the vial. The syringe injection protocol had HoxD9 protein titrated as the ligand into fixed concentrations of the PO- or PS2-DNA duplexes in the sample cell. Experiments were performed using the Microcal $ITC_{200}$, which is a modern ITC instrument featuring enhanced sensitivity and miniature cell size volume of 204.8 μL. This model requires much less sample than the popular MicroCal VP-ITC instrument from the previous generation, which has lower sensitivity and a cell volume of 2.5 mL. In the sequential flow of HoxD9 studies these ITC experiments were performed last, and material was running low. The high-affinity association constants revealed by anisotropy studies mean that ITC fitting cannot capably determining $n$, $K_a$, and $H_{\text{app}}$ because the inflection of the binding curve is too sharp, so experiments instead focused on capturing the apparent enthalpy of the HoxD9-DNA interaction. As a pilot study the decision was made to perform many lower resolution experiments using the $ITC_{200}$ with its reduced sample concentration requirement rather than fewer high-resolution large-volume experiments on the VP-ITC. A snapshot of the energetic signature of HoxD9-DNA interactions over a broad temperature range could be obtained, rather than focusing on a narrow temperature window. Many biological recognition processes,

especially those of DNA-binding proteins, have thermodynamic enthalpy values that cross between positive and negative values within the temperature range of 10-40°C. The HoxD9-DNA system has not yet been characterized and it would have been a gamble to select an arbitrary temperature to perform all experiments without prior knowledge on the enthalpy the system. If a temperature near the enthalpic transition point was accidentally chosen, the observed heat in the experiment would be near zero and the sample would be wasted.

Since the number of experiments was maximized, sample concentration was kept low and there was battle with signal-to-noise limitations of the instrument. Increasing the concentration of sample 10-fold would have improved S/N ratio drastically, but the sample stock would be consumed 10 times quicker. To compensate, experiments were designed with a fewer, larger titrations volumes to increase the magnitude of the heat signal produced in the cell, which limited the number of injections to 16 for most studies, and required faster stirring. This type of experimental design is suitable for apparent enthalpy measurements but is inappropriate for binding affinity measurements because of poor resolution around the binding curve sigmoidal inflection point from fewer data points. The heat produced from each titration was near the lower limit recommended by the manufacturer for the $ITC_{200}$. It is possible that the older MicroCal VP-ITC and similar generation machines will not perform well with these concentration values.

Since the cell volume in ITC is fixed, there is an upper limit to the available number of ligand injections before the liquid spills out of the fixed volume. Therefore to capture all binding processes, three different categories of experiment designs were considered based on molar ratio range for protein:DNA. The initial set of experiments were archetypical ITC studies at protein:DNA concentrations near the stoichiometric equivalency. These initial studies showed the apparent association of HoxD9 to DNA to be more complex that previously expected, having both a high-affinity and low-affinity site, each producing a thermodynamic signature. Presumably these are the HoxD9 recognition sequence, and non-specific DNA association to the remainder of the duplex, respectively. Concentrations were optimized to capture the apparent enthalpy of both interactions across three experimental designs: high-affinity saturation $0 < n < \sim 0.5$, (16 2.49 µL injections

of 36 μM HoxD9 into 15 μM duplex); normal saturation $0 < n < \sim 2.1$ (16 2.49 μL injections of 100 μM HoxD9 into 10 μM duplex), and low-affinity saturation $0 < n < \sim 5.3$, (24 1.66 μL injections of 250 μM HoxD9 into 10 μM duplex). Experiments were performed over the temperature range of 5-35°C at 5°C intervals for as many of the experiment types that sample quantity allowed. Calorimeter settings for all experiments were as follows: pre-titration delay, 60s; stirring speed, 1000 rpm, reference power, 1-4 μcal/sec; feedback mode, 2 (high). Because these partial saturations cannot be accurately fitted using models requiring full saturation, enthalpies of association at fractional saturation were instead calculated from heat averages:

$$\sum_{m}^{1\,m} \Delta q_{i,app} = \sum_{i=1}^{m} \left(\Delta q_{i,app} + \Delta q_{i,ns}\right) \times m^{-1}$$

where m is the number of injections. This approach is valid because the heats of association are identical for each injection until saturation of the targeted binding site begins to occur (the early stages of sigmoidal curve inflection). This works as long as the affinities for two sites are at least two orders of magnitude apart, which was assumed in this case.

**Table 3.1) List of all ITC Experiments Performed**

| *Temp.* | *Low-Saturation* | | *Normal-Saturation* | | *High-Saturation* | |
|---|---|---|---|---|---|---|
| | **PO** | **PS2** | **PO** | **PS2** | **PO** | **PS2** |
| 5°C | ✔ | ✔ | ✔ | ✔ | ✔ | |
| 10°C | ✔ | ✔ | ✔ | ✔ | | |
| 15°C | | | ✔ | ✔ | | |
| 20°C | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ |
| 25°C | | | ✔ | ✔ | | |
| 30°C | ✔ | ✔ | ✔ | ✔ | | |
| 35°C | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ |

**Table 3.2) ITC Titration Design**

| | *Low-Saturation* | *Normal-Saturation* | *High-Saturation* |
|---|---|---|---|
| | "High-Affinity Site" | | "Low Affinity Site" |
| **[DNA]** | 15 μM | 10 μM | 10 μM |
| **Cell Vol.** | 205 μL | 205 μL | 205 μL |
| **[HoxD9]** | 36 μM | 100 μM | 250 μM |
| **Inj. Vol.** | 2.49 μL | 2.49 μL | 1.66 μL |
| | | | |
| **Injection #** | **Mole Fraction (protein:DNA)** | | |
| 1 | 0.03 | 0.12 | 0.20 |
| 2 | 0.06 | 0.25 | 0.41 |
| 3 | 0.09 | 0.37 | 0.62 |
| 4 | 0.12 | 0.50 | 0.82 |
| 5 | 0.15 | 0.63 | 1.03 |
| 6 | 0.18 | 0.76 | 1.25 |
| 7 | 0.21 | 0.89 | 1.46 |
| 8 | 0.24 | 1.02 | 1.67 |
| 9 | 0.28 | 1.15 | 1.89 |
| 10 | 0.31 | 1.29 | 2.11 |
| 11 | 0.34 | 1.43 | 2.33 |
| 12 | 0.38 | 1.57 | 2.55 |
| 13 | 0.41 | 1.71 | 2.77 |
| 14 | 0.44 | 1.85 | 3.00 |
| 15 | 0.48 | 1.99 | 3.22 |
| 16 | **0.51** | **2.13** | 3.45 |
| 17 | | | 3.68 |
| 18 | | | 3.91 |
| 19 | | | 4.15 |
| 20 | | | 4.38 |
| 21 | | | 4.62 |
| 22 | | | 4.86 |
| 23 | | | 5.10 |
| 24 | | | **5.34** |

## RESULT: ITC REVEALS HIGH- AND LOW-AFFINITY BINDING SITES ON THE 24-BP DUPLEX FOR HOXD9

The initial ITC binding experiments were performed at saturation levels typical for 1:1 interactions (i.e. titration saturation at ~2.5:1 protein:DNA). However at this saturation level, there remained a large enthalpic signature, indicating that binding reactions were still occurring (Fig. 3.19). The receptor, in this case the DNA duplex, is not fully bound unless the enthalpy of each ligand

injection (HoxD9) returns to the enthalpy of its dilution. Because the sample dialysis preparation was extensive, the enthalpy of diluting HoxD9 in the experiments was nearly negligible, within instrument noise near ~0 kcal/mol over all temperature ranges 5-35°C, even when using protein injection concentrations as large as 250 μM. The observed excess enthalpy was therefore interpreted as non-specific association of HoxD9 to the duplex at areas beyond the 10-bp wide high-affinity binding site. This mode of non-specific interaction has been reported previously with HoxD9, using diffusion and hopping studies[317]. Hand-constructed molecular models suggest that up to five HoxD9 monomers can fit onto the 24-bp duplex without significant steric clashes and the ITC data supports this suggestion, showing that 3-4 HoxD9 monomers easily associate with the 24-bp duplex.

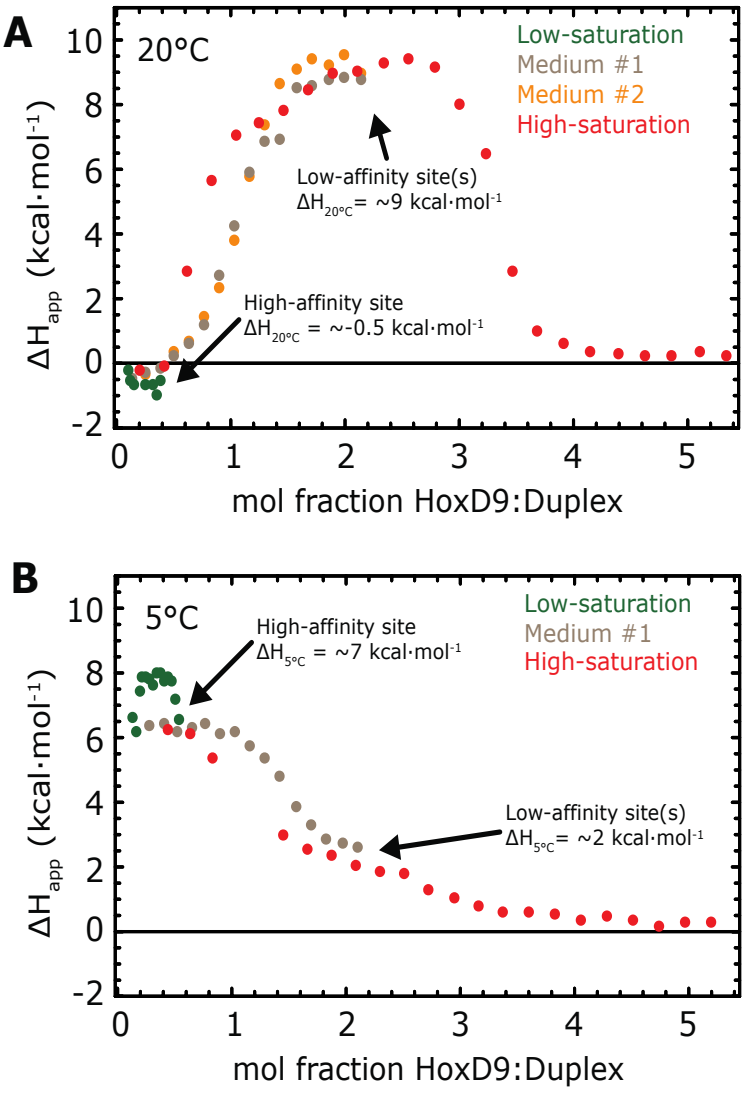



**Fig. 3.19) ITC Reveals High-Affinity and Low-Affinity Binding Sites**

**A)** All three experimental concentration data points at 20°C show nice uniformity across concentration setups at this temperature. **B)** Larger temperature differences between methods was observed at 5°C.

## RESULT: THERMODYNAMIC SIGNATURE OF HOXD9-DNA ASSOCIATION

The thermodynamic signature of HoxD9 association with DNA is complicated, having two phenomena distinctly contributing to the observed heat. The molar ratio of protein to DNA during the ITC experiments indicates the presence of one site high-affinity site and another low-affinity site. Their enthalpic signatures are distinctly different at temperatures beyond 10°C, with the high-affinity site becoming more exothermic as temperature increases while the non-

specific binding becomes exceptionally endothermic. A 3D surface plot of the data (Fig. 3.20) shows this effect nicely.

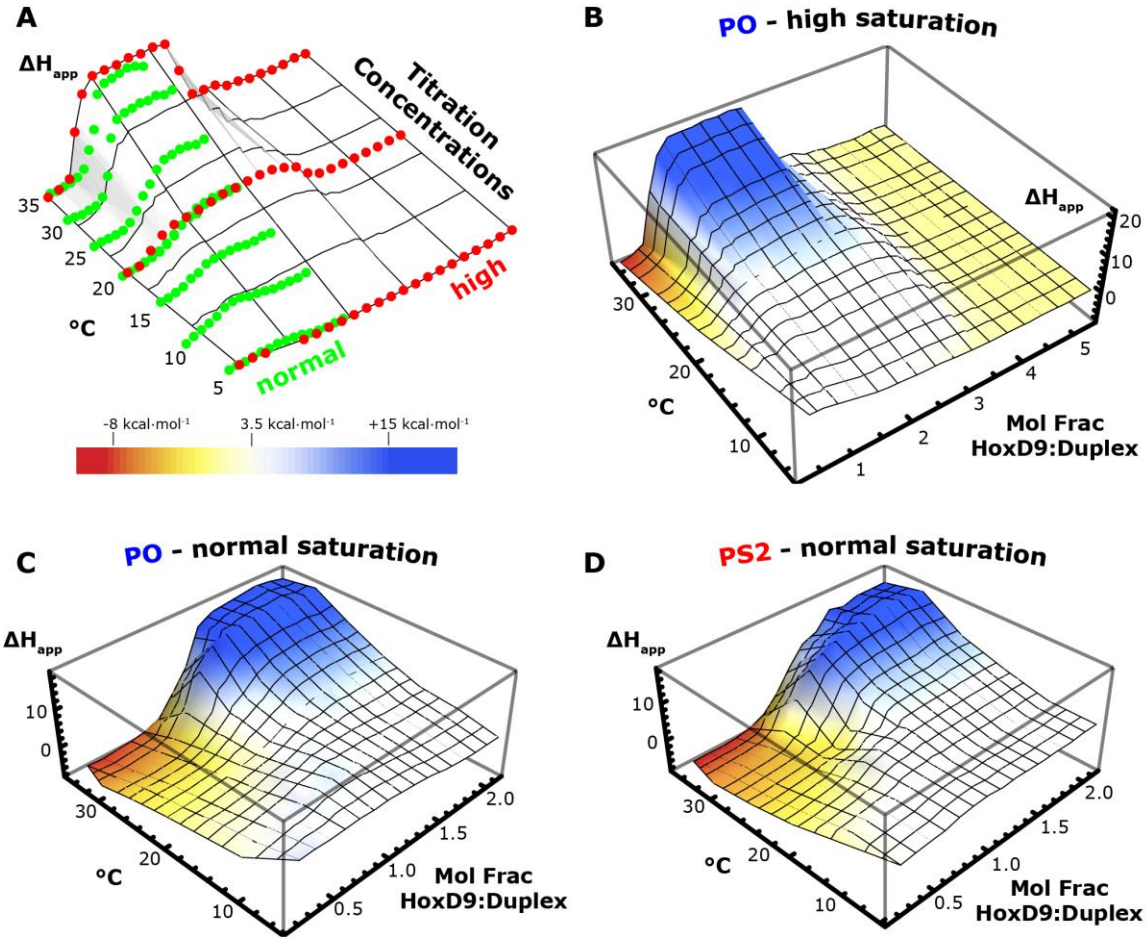



**Fig. 3.20) Enthalpic Signature of HoxD9-DNA Interaction**

**A)** The enthalpic signature over the temperature range of 5°C to 35°C was produced using the datapoints shown here. The normal saturation experiments are indicated with green dots, and the high saturation experiments are indicated with red dots. **B)** The enthalpic signature of HoxD9 associating with DNA is plotted in a 3D surface plot from the high-saturation experiment (green dots), colored by the magnitude of the enthalpy value. In this coloring scheme, red corresponds to exothermic, yellow is centered around 0 kcal·mol$^{-1}$, white is slightly positive, and blue corresponds to a strong endothermic reaction. Two binding modes are observed: the initial high-affinity binding site that is exothermic at 35°C and becomes less exothermic as the reaction temperature is decreased; and then a second low-affinity "site" that is strongly endothermic at high temperatures. It is assumed that the low-affinity "site" corresponds to non-specific binding that occurs as the protein saturates the DNA molecule. After an approximate molar ratio of 3:1 HoxD9:DNA, additional titration of HoxD9 does not result in observed enthalpy of association, seemingly because no more HoxD9 can fit onto the duplex. **C)** The normal saturation experiment (red dots) for the PO duplex is shown. The distinction between high-affinity and low-affinity sites is obvious. Since there is still an observed enthalpy at 2:1 molar ratio, complete binding is not complete and more HoxD9 is needed for saturation. **D)** The normal saturation experiment (red dots) for the PS2 duplex. The PO and PS2 enthalpic signatures are very similar and cannot be distinguished with the resolution of the ITC binding experiments performed.

In this representation X- and Y-values are the mole fraction of HoxD9:DNA duplex, and experimental temperature, respectively. The Z-values are the $\Delta H_{app}$ at these conditions, and $\Delta H_{app}$ values are used to generate the coloring scheme for the surface. Endothermic enthalpy values are colored blue and exothermic values are red. The center color (white) is skewed slightly off, centered at 3.5 kcal·mol$^{-1}$ instead of zero. Instead, yellow represents zero enthalpy. Figure 3.19.A shows the data points that were used to generate the frame for the 3D surface plot in Fig. 3.20.B, which summarizes seven different experimental temperatures and numerous HoxD9:DNA concentrations. The bottom half of the figure, Fig. 3.20.C and 3.20.D, show the results for normal saturation experiments for the PO- and PS2-HoxD9 association studies, respectively. The apparent enthalpy of HoxD9 association was very similar for the PO and PS2 complexes, producing nearly identical thermodynamic signatures that are just about indistinguishable at the resolution of the high-saturation and normal-saturation experiments.

The fluorescence binding data suggests that HoxD9 association with the PS2 duplex at the high affinity site is 0.6 kcal·mol$^{-1}$ more favorable than the PO duplex. ITC experiments focusing on this high affinity site (low saturation, $<$ 0.5 HoxD9:DNA) provided insight in the enthalpy of association for both Complex I and Complex II. It is assumed that each titration has the same enthalpy of association until the high-affinity binding site is half saturated. Both the low affinity saturation (15 values) and the normal saturation (3 values) had data points that fell at this concentration range (Table 3.2). These observed enthalpy values were averaged and taken to represent the enthalpy of the high-affinity site, as described in the methods. For the most part there was good agreement between low-saturation and normal-saturation measurements (Fig. 3.20), except for 5°C and 10°C. This was a good result, because only the first few injections from the normal saturation experiment were used, and initial titrations in ITC are often noisy due to unavoidable syringe leakage in ITC apparatuses. Visual analysis of the 3D high saturation plot (Fig. 3.20.B) shows that the difference between the high-affinity and low-affinity sites becomes less distinct at lower temperatures, which may contribute to the observed discrepancies between low- and normal-saturation observed enthalpies at 5°C and 10°C. Nevertheless, the goal here was to ascertain whether

dithioation of the phosphate backbone has any obvious enthalpic consequences. Analysis of the high affinity site enthalpic signature (Fig. 3.20) shows no significant difference between HoxD9 association with the PO and PS2 duplex. For both systems the temperature dependence of apparent binding enthalpy appears linear over the 5°C - 35°C temperature range and transitions from being endothermic to exothermic at approximately 20°C. The absence of curvature at higher temperatures confirms the CD result that no folding of the recognition helix occurs upon DNA binding, which would have a distinct enthalpic effect from hydrogen bonding that occurs in such a process. As for the differences between Complex I (normal) and Complex II (dithioate at Lys57), within instrument noise, the apparent enthalpies of association for the PO Complex I and PS2 Complex II are indistinguishable.

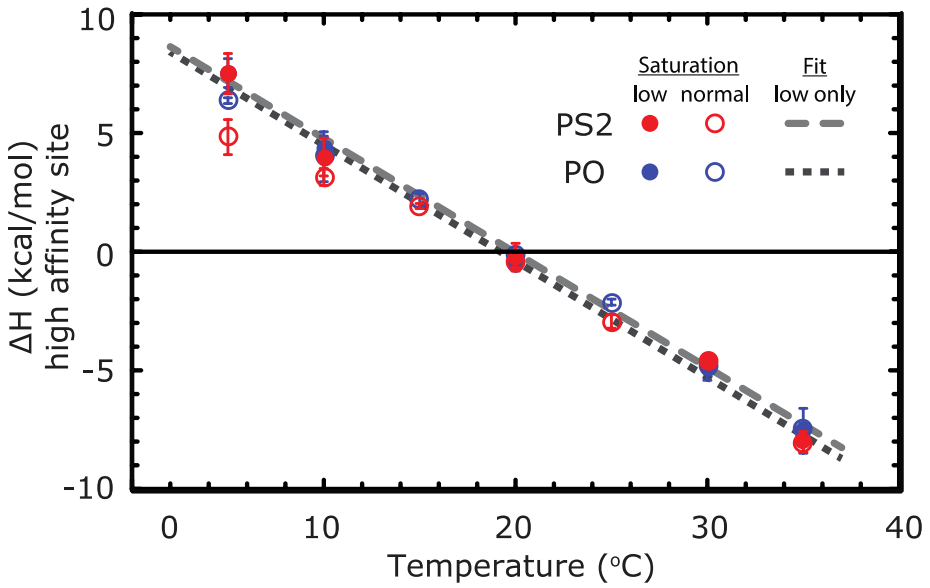



**Fig. 3.21) Apparent Enthalpy of High Affinity Site Similar for PO and PS2**

High-affinity binding site enthalpies from low-saturation experiments (closed circles, average of up to 15 injections), and normal-saturation (open circles, average of up to 3 injections) are plotted for HoxD9 association to the PO DNA (blue) and PS2 DNA (red) duplexes. Because of limited (only 3 injections) and poorer accuracy (slight leakage from syringe causes concentration uncertainties) of the normal-saturation experiments, the linear fitted line considers data only from the low-saturation experiments. The apparent enthalpies pass through zero, becoming exothermic as temperature increases, crossing over at ~20°C.

The resolution of ITC is near $\pm$ 1 kcal·mol$^{-1}$, so it would have been difficult to quantify the contribution of a single sulfur-for-oxygen substitution using ITC given the size of the heat of the system. However, these experiments do succeed in

providing the insight that no major structural differences in association is occurring, such as if the PS2 caused structural perturbations such as a kink or bend. Such a situation would have produced an obvious enthalpic effect, which was not observed.

## 3.4) Probing Ion Pairing with NMR Spectroscopy

The interaction of the positive charge of lysine $NH_3^+$ group (cation) with the anionic phosphate in the DNA backbone (anion) is an example of a biological ion pair. Ion pairing is a fundamental atomic interaction that is the foundation for many chemical and biological processes, yet they are difficult to detect. Because of this, there has been a lot of debate as to the existence of ion pairs in solution over the past hundred years.

### THE CONCEPT OF ION PAIRS DEVELOPS FROM EXPERIMENTS WITH ELECTROLYTE SALTS

Ions were discovered in the 1800s using conductivity and activity coefficient measurements of electrolyte salts, where the behavior of an electrolyte solution was shown to deviate from that an ideal solution. The model adopted for a strong 1:1 electrolyte was that of two fully dissociated ions in solution.

$$\text{strong electrolyte}(aq) = \text{cation}^+(aq) + \text{anion}^-(aq)$$

It was simple and could explain experimental observations of the time, which focused on dilute aqueous solutions of strong electrolytes, well enough to warrant adoption. Fundamental aspects of electrolyte solubility equilibria were put forward by J.H. van't Hoff and L.T. Reicher[361], Walther Nernst[362], and Arthur Noyes[363] in a way that described electrolyte experimental data while also satisfying both the van't Hoff's theory of dilute solutions[364,365] and Svante Arrhenius' theory of salt dissociation[365,366]. William Sutherland[367,368] and Niels Bjerrum[369,370] refined the model of a 1:1 electrolyte, describing binary charges as completely dissociated, yet still interacting through long-range Coulombic electrostatic forces as electric doublets. Further equations were advanced by Peter Debye and Erich Hückel[371,372] and later Von Lars Onsanger[373] that expanded on this, and worked for describing observed activity coefficients and conductivities of dilute electrolyte solutions, respectively. The term "dilute" is important, the equations proposed were valid

only for experiments in low electrolyte concentrations ($\lesssim$ 3 mM), because any higher concentrations saw significant deviations from prediction.

Looking closer at Debye-Huckel theory for activity coefficients from a modern perspective, there are several obvious assumptions the model makes with its simplicity that helps explain its failures at higher concentrations.

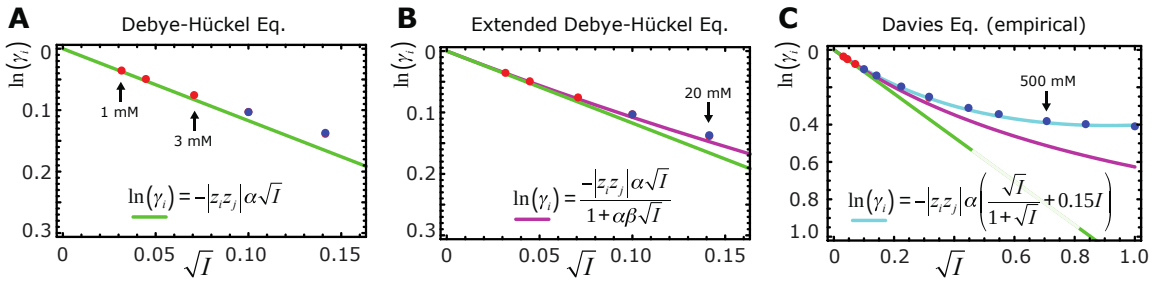



**Fig. 3.22) Concentration-Dependent Terms Improve Ion Theory**

Early ion theory was developed for dilute salts, but it struggles at higher concentrations due partly to deficiencies in modeling concentration-dependent ion association. The following equations and graphs predict the activity coefficient $\gamma$ for a 1:1 salt at various concentrations. The points in the graph are published experimental activity coefficient data for low[374] (red) and high[375] (blue) ion concentrations. **A)** The Debye-Hückel approximation of $\gamma$ (green) is valid only for very low ionic strengths (see 3 mM point). **B)** The Extended Debye-Hückel equation (purple) has an additional parameter to improve $\gamma$ prediction, but it is only effective for another order of magnitude (see 20 mM point). **C)** The Davies equation (cyan) contains an empirically determined parameter for ion concentration, and accurately predicts $\gamma$ up into the 1.0 M range.

First, it assumes that ions are non-polarizable hard spheres, inaccurate since many ions have large, polarizable electron densities. Second, Debye-Hückel describes the solvent as a structure-less medium with a uniform dielectric constant and the electrostatic interaction of solvent molecules with solute ions are ignored (i.e. implicit vs. explicit solvent). The third and most pertinent aspect of Debye-Hückel theory is the assumption that in strong electrolytes ions are completely dissociated into free cations and anions. At the time ion pairing was difficult to detect with available technology and was considered relatively unimportant compared to interactions with the bulk solvent[376]. However as can been seen in Figure 3.21B incorporation of parameters considering the effect of the ionic strength and interaction of ions (the β parameter, the hallmark of the extended Debye-Hückel equation) increases the accuracy of prediction up to concentrations an order of magnitude higher (~20 mM). This emphasis on an ion concentration-dependent term was considered in greater detail when C.W. Davies produced his empirical extension of the Debye-Hückel equation. He analyzed data of many

experiments and used this empirical evidence, instead of a theoretical formulation, to create an expanded concentration-dependent term. The result was a model capable of accurately predicting activity coefficients up to 0.5 M in ion concentration. Interestingly, the Davies (or the similar summed Pitzer version) is still used to this day in many computational programs because of its quick calculation speed and relative accuracy.

Others during this period were also recognizing the importance ion association, developing concepts of ion association constants derived from mass action law that could describe solubility equilibria at higher electrolyte concentrations. This idea was supported by Julius Stieglitz[377], Noyes[378,379], Gilbert Lewis[380] and others who were working on solubility studies involving the salting-in and salting-out of mixed ions. Their comments regarded the phenomenon that relatively insoluble salts could be made soluble after the addition of a second species of salt[378,379,381-384]. How else, they reasoned, could this be explained without initial salt dissociation in the solvent followed by formation of new ion pairs of different molecular composition and solubility properties? Expanding on this thought Johannes Brønsted presented his theory of specific ionic association, stating that ions were uniformly influenced by other ions of their own charge and that an electrolyte's activity coefficient was dependent on the combined action of these like charges as well as that of the solvent, and also the work required to separate two ions of opposite charges[385-387]. This is an important development in electrolyte solution theory because it helped reconcile the difference between canonically strong electrolytes, which should be fully dissociated, and weak electrolytes, which should be only partly dissociated. Instead of a dichotomous definition, electrolytes should be thought of as falling somewhere between these two states, with varying degrees of dissociation.

Based on this concept Bjerrum advanced a new mathematical model describing ion pairing of electrolytes, developing a term describing the electrostatic work required to separate ions of opposite charges and defined the probability of ion pair formation as a function the distance, $r$, between the ion centers. This function has a minimum that falls at a distance, $q$, such that ions where $r \leq q$ should be considered as paired and those where $r > q$ should be considered as free. This

formal definition for a specific cutoff distance for ion pairs, $R$, was a significant development in ion association and ion pairing theory and is referred to as a Bjerrum ion pair. It is generally accepted today that the electrostatic attraction between the ion pair partners is the main driving force for the pair formation, and that these forces can be dealt with more less on the lines proposed 90 years ago by Bjerrum.

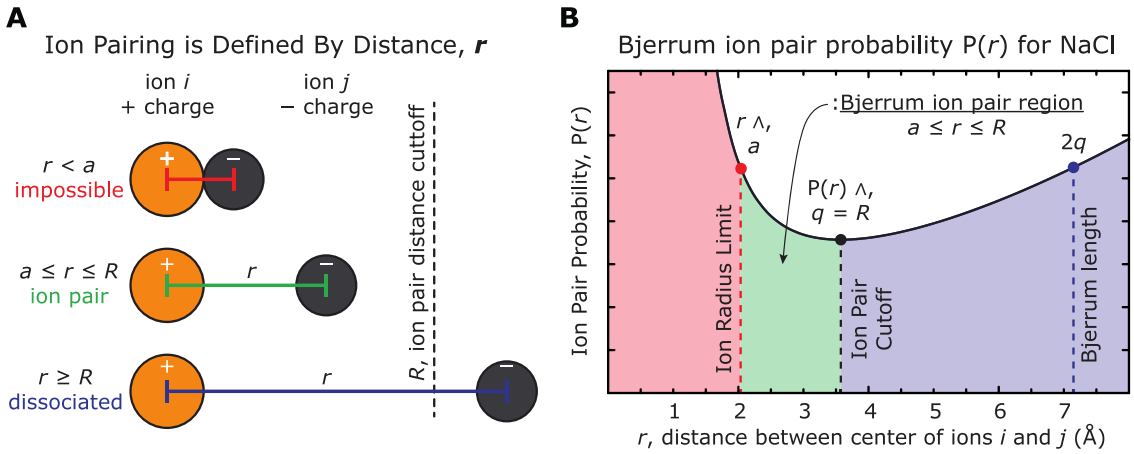



**Fig. 3.23) Bjerrum Ion Pairing in the Restricted Primitive Model (RPM)**

**A)** Ion pairing is determined by the distance ($r$) between cation (orange) and anion (black). Values of $r$ smaller than the ion radius limit are impossible, and values of $r$ greater than Bjerrum's calculated value of $R$ are considered dissociated. Anything in-between is considered an ion pair. **B)** As an example, Bjerrum's ion pair probability function is plotted for NaCl. The green region corresponds to the distance range for NaCl ion pairs in a dielectric constant of 78.5.

Over time the concept of an "ion pair" has grown from the Bjerrum ion pair to ever more complex models through the work of Raymond Fuoss[388], Ernest Grunwald[389], Eigen and Tamm[390,391], Krienke and Barthel[392] and others. Short range interactions are more ambiguous, involving repulsion of ions at very short distances and also considers interaction of the solvent molecules in the solvation shells. For an excellent review of this progress in advanced ion pair models one should look at the review by Marcus and Hefter[376]. The important point to take away is that in the modern view of electrolyte association, the solvent heavily influences the equilibria of ion pairs.

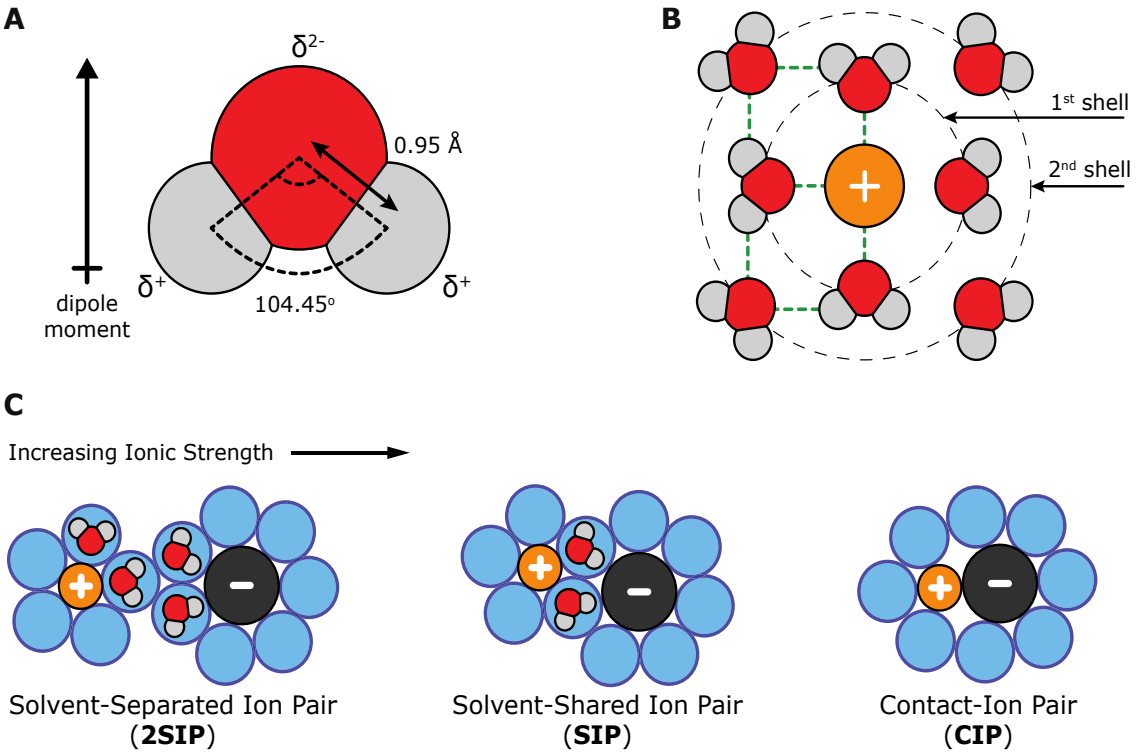



**Fig. 3.24) Solvent Interaction Determines the Type of Ion Pairing**

**A)** The water molecule's dimensions and dipole moment. $H_2O$: Oxygen, Red; Hydrogen, Gray. **B)** A solvated cation has several hydration shells. **C)** Solvated ion pairs can take on three forms depending on the degree of solvation shell sharing.

Ion pairs are classified into one of three categories (Fig. 3.24) based on their degree of solvent interaction: solvent-separated ion pairs (2SIP), solvent-shared ion pairs (SIP) and contact ion pairs (CIP). In the 2SIP configuration both the cation and anion electrostatically interact but are hydrated in their own independent solvation shells. In the SIP state, the ions are close enough to share a solvation shell. In the CIP state the ions are even closer, a significant portion of each ion is desolvated and they interact through direct contact, while extensively sharing solvation shells. Most present-day researchers agree that ion pairs can exist in all three of these states and each is at equilibrium with ions as free species, and can be accounted for using mass action law.

## Previous Experimental Studies on Contact Ion Pairing Limited to Small Molecules

Ion pair dynamics have been previously investigated in small organic compounds using time-resolved absorption spectroscopy, infrared spectroscopy, and Raman spectroscopy[393-397]. The timescale for transitioning between the CIP and SIP states in these systems was found to occur at the picosecond to nanosecond range, and free energy differences between the CIP and SIP states were found to be ∼1−2 kcal/mol. Despite the wealth of information available for small organic compounds, not much is known about the dynamics of ion pairing in biological macromolecules because their size and complexity means the techniques successful for small molecules cannot be applied successfully.

A new approach for understanding structural function of ion pairing in protein-DNA interactions is needed. In the upcoming results, such a mthodolgy is demonstrated. Using the HoxD9-DNA system the highly dynamic nature of the interfacial ion pairs in biological recognition is reported, and the biological significance of this behavior is discussed.

## $^{h3}J_{NP}$ Coupling as a Determinant of Contact Ion Pairing (CIP) in Protein-DNA Interactions

In NMR experiments there are two types of magnetic interactions between nuclei: direct coupling (dipole-dipole coupling), and indirect scalar coupling (J-coupling, or spin-spin splitting). Scalar couplings are through-bond magnetic interactions mediated by the interceding electrons, and are used to probe bonding patterns and angles between nuclei. During J-coupling processes, the spin of one nucleus polarizes the spins of intervening electrons over the bond, which in turn polarizes the spin of the next nuclei. The result is an increase or decrease in energy depending on the spin relation of the two nuclei, and a J-coupling value between them is determined (reported in Hz units). Because the effect is transmitted through the bonding electrons, the J value drops significantly as the number of bonds between nuclei increases. In these experiments three-bond inter-hydrogen-bond scalar couplings between the nitrogen nuclei and phosphorus nuclei ($^{h3}J_{NP}$)

were used as an indicator of ion pairing between lysine side chains and the phosphate backbone. Because of orbital overlaps in hydrogen bonds, only those nuclei involved in contact ion pairing (CIP) will exhibit this type of J-coupling (*i.e.* not 2SIP or SIP), and the J value for CIPs is predicted to be small, given the number of through-bond interactions necessary.

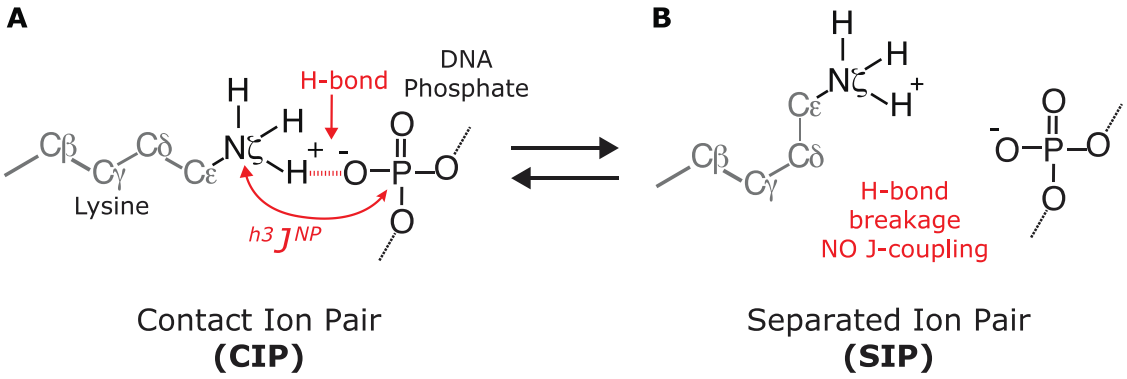



**Fig. 3.25) J-Coupling as a Determinant of Contact Ion Pairings (CIP)**

Three-bond inter-hydrogen-bond scalar couplings ($^{15}$Nζ-$^{1}$Hζ···O=$^{13}$C') between the nitrogen nuclei and phosphorus nuclei ($^{h3}J_{NP}$) was used as an indicator of ion pairing between lysine side chains and the phosphate backbone. Because *J*-coupling intensity diminishes as more through-bond interactions are considered, this type of *J*-coupling is observable only for the CIP state.

## METHOD: NMR $^{h3}J_{NP}$ COUPLING

NMR experiments were performed using a Bruker Avance III 600-MHz equipped with a QCI ($^{1}$H/$^{13}$C/$^{15}$N/$^{31}$P) cryogenic probe. The pulse sequences used for analyzing $^{h3}J_{NP}$ coupling between lysine $NH_3^+$ and the phosphate $^{31}$P are an extension of those developed by Iwahara[323] for long-range $J_{NC}$ coupling for lysine $NH_3^+$ groups to intraresidue $^{13}$Cγ and hydrogen bonding to the peptide backbone ($^{15}$Nζ-$^{1}$Hζ···O=$^{13}$C'). Separate $^{31}$P carrier positions were used for PS2 $^{31}$P and PO $^{31}$P because of their large chemical shift differences, which was confirmed independently using using 1D experiments with free duplex. Similar to the long-range H(N)CO experiment for $^{h3}J_{NC'}$ analysis[398], the $H_3$(N)P experiment was used to provide heteronuclear ($^{1}$H, $^{31}$P) correlations via $^{h3}J_{NP}$ coupling between the lysine $NH_3^+$ and the phosphate. Then the spin-echo $^{h3}J_{NP}$ modulation constant-time HISQC experiment was used to determine $|^{h3}J_{NP}|$ values, in a manner similar to spin-echo J-modulation experiments for protein χ1 angle analysis[399,400]. Two sub-spectra were recorded in an interleaved manner: one with $^{31}$P 180° pulses at positions *a* and the

other at position *b*. Only if $^{h3}J_{NP}$-modulation in the sub-experiment *a* is significantly larger than noise, then the difference spectra gives signals with an intensity corresponding to $I_b - I_a$. The J constant $|^{h3}J_{NP}|$ was calculated from this, according to: $I_a/I_b = cos\{2\pi J(Td + \delta)\}$. Further details on these methods can be found in the manuscript published for this work[324].

## RESULT: $^{H3}J_{NP}$ COUPLING PROVIDES DIRECT EVIDENCE FOR CIP STATE AT THE PROTEIN-DNA INTERFACE

All three lysines $NH_3^+$ groups in both complexes exhibited $^{h3}J_{NP}$ coupling, providing direct evidence that they form at least part-time contact ion pairs (CIP) with phosphate groups at the protein-DNA interface. Neither the 2SIP nor SIP state can exhibit $^{h3}J_{NP}$ coupling because of the through-bond requirement for indirect coupling (Fig. 3.25). The 2D heteronuclear correlation H3(N)P experiment was used to observe the coherence transfer via $^{h3}J$ coupling. The intermolecular correlation signals arising from $^{h3}J_{NP}$ evolution were small in magnitude (<1 Hz), but could be observed if long coherence transfer periods were used, due to the very slow $^{15}N$ transverse relaxation of $NH_3^+$ groups[323]. The absolute values of $^{h3}J_{NP}$ coupling constants for the intermolecular ion pairs in Complex I and II were determined using signal intensities from the a and b subspectra of spin−echo $^{h3}J_{NP}$-modulation constant-time HISQC experiments as described in methods. These absolute values ranged in magnitude from 0.06 to 0.63 Hz and are assigned in Figure 3.25B. In these measurements an additional loss in $^{h3}J_{NP}$ sensitivity occurred because of $^{31}P$ longitudinal relaxation, a partial self-decoupling effect, and required slight correction to the data. To accomplish this a 1-D $^{31}P$ inversion recovery experiment was performed using multiple recovery periods to determine a generalized $R_{1P}$ rate for the phosphates (~0.4-0.6 $s^{-1}$), and this was used to estimate the sensitivity loss due to $^{31}P$ longitudinal relation, which was up to ~10%. Further details on the sensitivity error estimation process is covered in the supplementary information of the published manuscript for this work[324]. The end result is that the measured $|^{h3}J_{NP}|$ values were corrected using the relation: $J_a = \sqrt{{J_{NP}}^2 - (R_{1P}/2\pi)^2}$, where $J_a$ is the apparent coupling constant and $R_{1P}$ was taken to be 0.5 $s^{-1}$. In Figure

3.25B the original $|^{h3}J_{NP}|$ values are present and the corresponding corrected values are indicated in parenthesis, in all cases the differences between the values were smaller than 0.05 Hz.

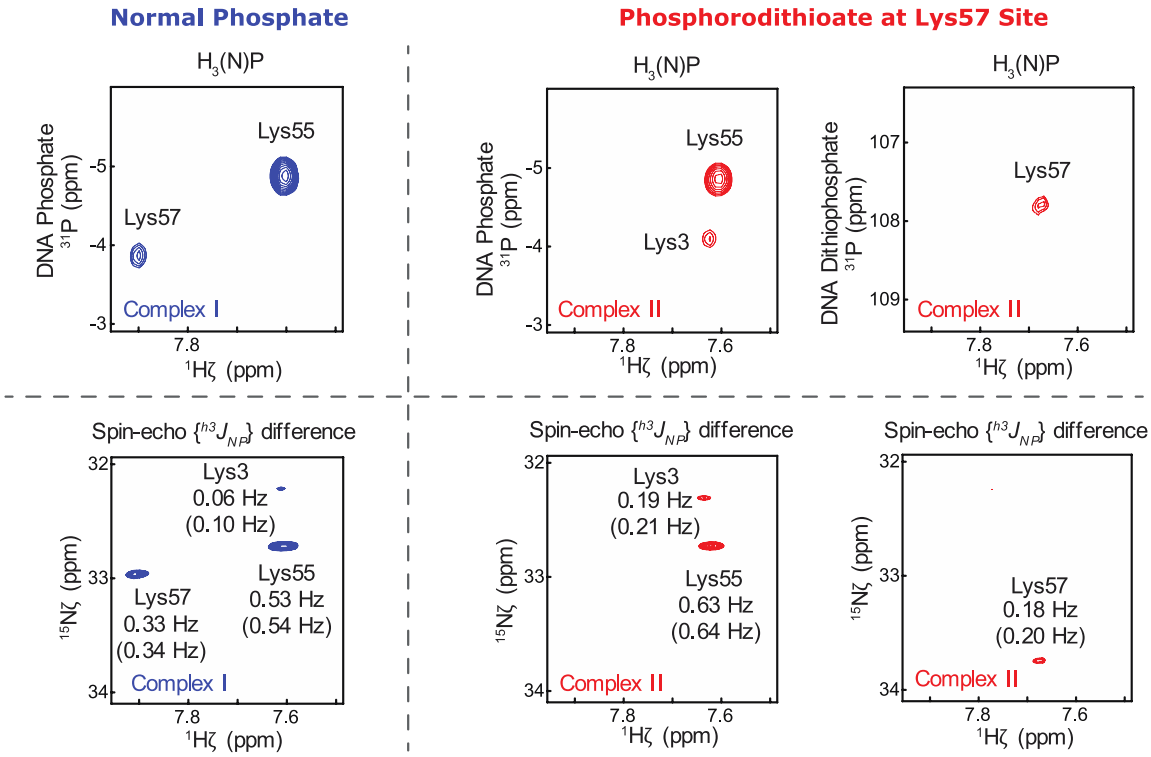



**Fig. 3.26) J-Coupling Provides Evidence for the CIP State in Ion Pairing Between Protein-DNA**

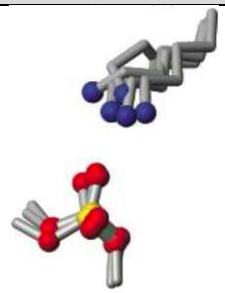
Hydrogen-bond scalar $^{15}N–^{31}P$ coupling $^{h3}J_{NP}$ between $^{15}N$ and $^{31}P$ nuclei provides evidence for the presence of CIP in lysine side-chain ion pairs with the DNA phosphate backbone. **(A, C, D)** 2D H3(N)P spectra that give heteronuclear $^{1}H–^{31}P$ correlation cross peaks via h3JNP coupling. 31P chemical shifts are relative to trimethylphosphate (TMP). **(B, E, F)** Difference spectra recorded by the spin–echo $^{h3}J_{NP}$–modulation constant-time HISQC experiment. Measured absolute values of $^{h3}J_{NP}$ are indicated; values in parentheses are $^{h3}J_{NP}$ values taking account of the correction due to partial self-decoupling arising from $^{31}P$ longitudinal relaxation (see text). Data for the normal phosphate is on the left **(A, B: Complex I)** and data for the phosphorodithioate at the position interacting with Lys57 is on the right **(C, D, E, F: Complex II)**. Carrier positions for $^{31}P$ were set to −3 ppm for panels for the normal DNA phosphate and 107 ppm for the DNA phosphorodithioate.

## Result: Comparing Observed $^{H3}J_{NP}$ Coupling Constants with Theoretical $^{H3}J_{NP}$ Coupling Constants Determined Through Density Functional Theory

A comparison was made between these experimentally observed values and $^{h3}J_{NP}$ coupling constants predicted by density functional theory (DFT). These calculations were performed by collaborators on the project, Mariappan Manoharan and Rafael Brüschweiler of the National High Magnetic Field Laboratory at Florida

State University, according to their previous methods[401]. The manuscript for this work[324], particularly the supplementary information, should be referenced for more details on the DFT preparation than what is provided here. Briefly, crystal structures used for calculations were of highly homologous homeodomain-DNA complexes in which Lys55 and Lys57 $NH_3^+$ groups exhibited a CIP state with DNA phosphates. Geometries from these ion pairs were used to construct a dimethylphosphate–lysine model, for which $^{h3}J_{NP}$ coupling constants were predicted using DFT calculations on Gaussian 09 software. Table 3.3 below summarizes the predicted $^{h3}J_{NP}$ coupling constants and bond geometries:

**Table 3.3) DFT Predicted $^{h3}J_{NP}$ Constants for Lys55 and Lys57**

| PDB code | N…O (Å) | N…P (Å) | N…O=P (°) | P=O…H (°) | Predicted $^{h3}J_{NP}$ (Hz) | Model Overlays |
|---|---|---|---|---|---|---|
| Lys55 – Phosphate | | | | | | |
| 2HDD [a] | 3.31 | 4.64 | 149 | 148 | -1.09 |  |
| 3HDD [b] | 3.16 | 4.58 | 159 | 141 | -2.27 | |
| 9ANT(i) [c] | 3.96 | 4.61 | 107 | 102 | 0.34 | |
| 9ANT(ii) [c] | 3.28 | 4.07 | 112 | 108 | 0.26 | |
| 2H1K [d] | 2.88 | 3.96 | 128 | 121 | -0.8 | |
| | | | Average $^{h3}J_{NP}$ : -0.72 (s.d. 1.08) | | | |
| Lys57 – Phosphate | | | | | | |
| 2HDD [a] | 2.74 | 3.86 | 129 | 128 | -1.47 |  |
| 3HDD [b] | 2.86 | 3.38 | 97 | 100 | 1.4 | |
| 9ANT(i) [c] | 3.68 | 4.35 | 107 | 109 | 0.49 | |
| 9ANT(ii) [c] | 3.45 | 4.35 | 118 | 122 | -2.79 | |
| 2H1K [d] | 2.42 | 3.39 | 119 | 121 | -0.48 | |
| | | | Average $^{h3}J_{NP}$ : -0.57 (s.d. 1.64) | | | |

DFT calculated hydrogen-bond scalar coupling is highly sensitive to subtle differences in hydrogen-bonding geometry and large variations were found in $^{h3}J_{NP}$ constants predicted for similar $NH_3^+$ groups from different crystal structures. However, when averaged together these predicted values (Lys55 $^{h3}J_{NP}$ = -0.72;

Lys57 $^{h3}J_{NP}$ = -0.57) agree in part with the experimental $^{h3}J_{NP}$ constants, determined as absolute values: (PO) Lys55 $|^{h3}J_{NP}| = 0.55$; Lys57 $|^{h3}J_{NP}| = 0.33$. This agreement suggests that the CIP state has a major presence in solution for the ion pairs of Lys55 and Lys57 $NH_3^+$ groups with a DNA phosphate group.

Dithioation of the phosphate had a marked effect on the experimental $^{h3}J_{NP}$ coupling, as the phosphorodithioate-Lys57 $NH_3^+$ ion pair had a significantly smaller $|^{h3}J_{NP}|$ constant (0.18 Hz) than that of the normal phosphate—Lys57 $NH_3^+$ ion pair (0.33 Hz). However, in contrast the DFT calculation suggested the opposite, with the average $^{h3}J_{NP}$ constant being larger (1.40 Hz) for the dithioate hydrogen-bonding geometry. This discrepancy and the smaller $|^{h3}J_{NP}|$ observed experimentally for the phosphorodithioate—Lys57 $NH_3^+$ ion pair could partly be attributed to a decreased solution presence of the CIP state in the PS2 ion pair compared with the normal PO ion pair, which is something not captured in DFT $^{h3}J_{NP}$ constant prediction from crystal structures but may have been captured experimentally. The culminating view, when the rest of the experiments are considered is that the dithioate shows enhanced ion pair dynamics (covered next), which possibly produces the lower $^{h3}J_{NP}$ coupling constant observed for dithioates.

## MEASURING ION PAIR DYNAMICS AT THE PROTEIN-DNA INTERFACE

The ion pair dynamics at the molecular interface in the HoxD9 homeodomain—DNA complex were probed using HISQC NMR spectroscopy. From these experiments $^{15}N$ relaxation data provide motional information on reorientation and bond rotations of $NH_3^+$ groups[322] (Fig. 3.27). This data was interpreted as a measure of ion pair dynamics between the $NH_3^+$ groups and phosphate/phosphorodithioate.

## METHOD: ORDER PARAMETERS FROM HISQC RELAXATION MEASUREMENTS

NMR experiments were performed using Bruker Avance III 600-MHz and 800-MHz spectrometers equipped with cryogenic probes. All relaxation experiments are HISQC-based, and were carried out at 35°C and pH 5.8, which is the condition at which Lys3, Lys55, and Lys57 are the prevalent peaks. The dynamics

of intermolecular ion pairing was described according to several amino-group symmetry axis order parameters described in Figure 3.25: $S^2$-axis, C—N bond rotational correlation times $\tau_f$, and reorientational correlation times $\tau_i$. Relaxation data from $^{15}N$ $R_1$, $^{15}N$ $R_{2,ini}$, $R_{(4N_zH_zH_z)}$, and heteronuclear NOE experiments were fit to determine these order parameters. Details on the pulse-sequences used, delay lengths, and non-linear regression analysis is covered in detail previously[322]. However in the fitting for this dataset, Eq. 17 rather than Eq. 16 in the reference[322] was used to calculated heteronuclear NOE because relaxation rates of $4N_zH_zH_z$ terms were greater than 60 $s^{-1}$, which makes making the two equations virtually identical). Based on the almost ideal tetrahedral geometry found for $NH_3^+$ groups[402], the order parameter used for bond rotation was 0.111, according to previous methods[322].

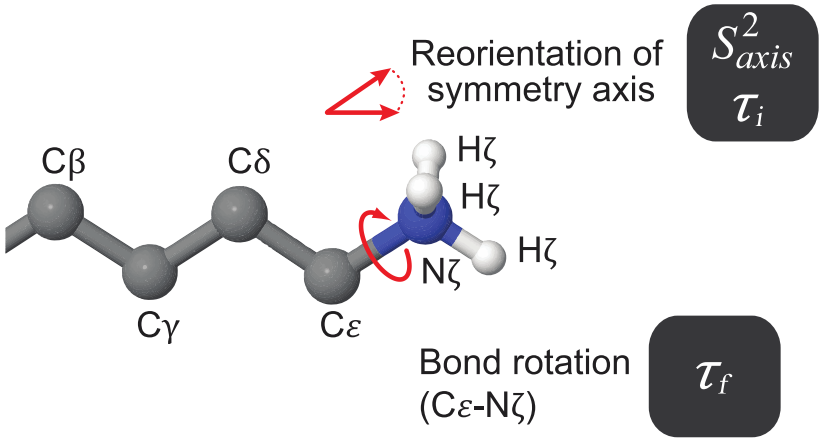



**Fig. 3.27) Motional Model Used to Describe Lysine $NH_3^+$ Dynamics**

Reorientation of symmetry axis and C–N bond rotation studied by $^{15}N$ relaxation for Lys side-chain $NH_3^+$ group. Parameters determined for $S^2$, $\tau_i$, and $\tau_f$ are shown in Table 3.4.

## RESULT: CHEMICAL SHIFT PERTURBATIONS UPON DITHIOATION OF THE PHOSPHATE

The three $^1H$-$^{15}N$ NMR cross-peaks of side-chain $NH_3^+$ groups of Lys3, Lys55, and Lys57 were studied using HISQC. The resonances for Lys57 $NH_3^+$ showed large perturbations upon the dithioation of its associated phosphate ($|\Delta\delta_H|$, 0.21 ppm; and $|\Delta\delta_N|$, 0.78 ppm), whereas resonances for unmodified lysine sites remained unchanged (Fig. 3.28). This pattern could be repeated with Lys55 as well as duel Lys55/Lys57 dithioated duplexes (not shown, but manuscript in preparation). This large chemical shift makes dithioation a handy technique for peak assignment, and

can supplement lysine and backbone triple resonance spectra for unambiguous assignment of HISQC peaks. This is a game changer for permitting the observation inter-macromolecular ion pairs from systems that are too large and complex for current NMR approaches.

As mentioned earlier, the $^{31}P$ carrier positions for J-coupling experiments were also different for the PS2 and PO, because of a large chemical shift difference between the two (Fig. 3.26 top). This was confirmed with a 1D $^{31}P$ NMR experiment that showed the $^{31}P$ shift to be 107.8 ppm for the PS2 Complex II and 108.7 ppm for the free PS2 Duplex II (Fig. 3.28.B). There was excellent agreement between these and the chemical shifts observed during J-coupling experiments (Fig. 3.26 top).

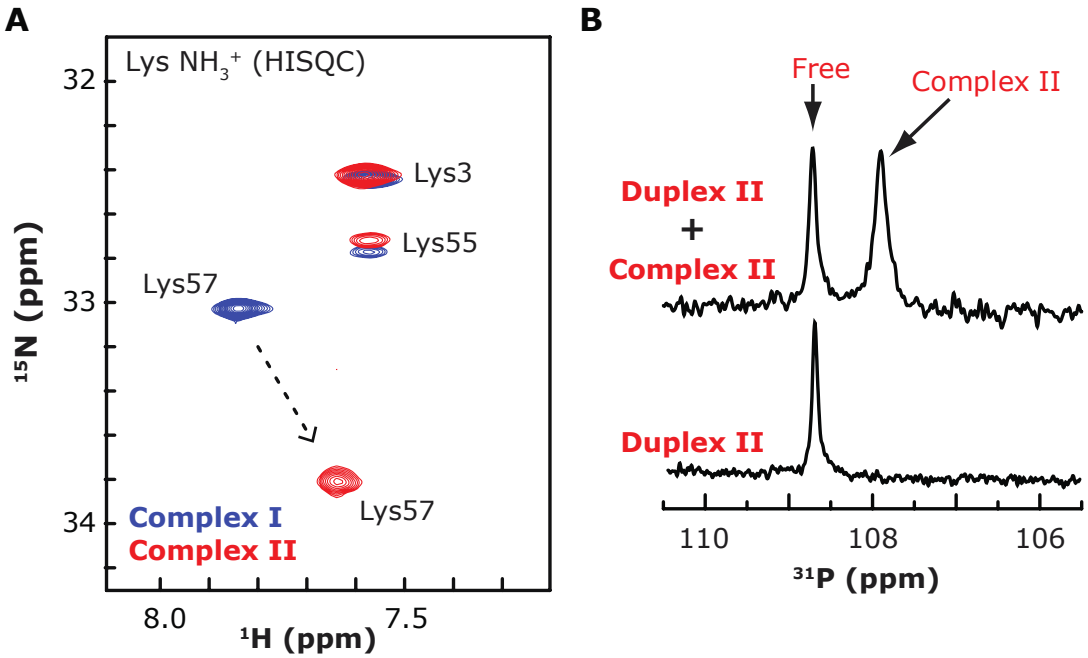



**Fig. 3.28) Overlaid HISQC Spectra of the Lys $NH_3^+$ Groups Recorded for HoxD9 Homeodomain-DNA Complexes with Duplex I and with Duplex II at 35 °C.**

**A)** Lysine-selective two-dimensional (2D) heteronuclear in-phase single quantum coherence (HISQC) spectra reveal $^{1}H–^{15}N$ cross-peaks for side-chain $NH_3^+$ groups for Lys3, Lys55, and Lys57. Oxygen-to-sulfur dithioation at the ion pair with Lys57 caused large perturbations in $^{1}H$ and $^{15}N$ resonances of Lys57 $NH_3^+$ group ($|\Delta\delta|$, 0.21 ppm; and $|\Delta\delta|$, 0.78 ppm), whereas signals from Lys3 and Lys55 $NH_3^+$ groups (which are unmodified) remained almost unchanged. **B)** 1-D $^{31}P$ spectra recorded for the phosphorodithioated Duplex II and Complex II + Duplex II mixture. The $^{31}P$ chemical shift of the cross peak agrees with $^{31}P$ chemical shift of the J-coupling signal in Fig. 3.26.

## RESULT: MOBILITY OF LYS $NH_3^+$ GROUPS IN ION PAIRS WITH DNA PHOSPHATES AND PHOSPHORODITHIOATES

**Table 3.4) $^{15}N$ Relaxation and Dynamics Parameters for Lys Side-Chain $NH_3^+$ Groups in the Intermolecular Ion Pairs in the HoxD9 Homeodomain-DNA Complexes**

| | Complex I (PO) | | | Complex II (PS2) | | |
|---|---|---|---|---|---|---|
| | Lys3 $NH_3^+$ | Lys55 $NH_3^+$ | Lys57 $NH_3^+$ | Lys3 $NH_3^+$ | Lys55 $NH_3^+$ | Lys57 $NH_3^+$ |
| **800 Mhz** | | | | | | |
| $^{15}N$ $R_1$ ($s^{-1}$) | 0.25 ± 0.01 | 0.43 ± 0.02 | 0.74 ± 0.01 | 0.29 ± 0.02 | 0.43 ± 0.03 | 0.33 ± 0.02 |
| $^{15}N$ $R_{2,ini}$ ($s^{-1}$)[A] | 0.98 ± 0.05 | 1.98 ± 0.21 | 2.27 ± 0.04 | 1.16 ± 0.07 | 2.18 ± 0.26 | 1.76 ± 0.05 |
| $^1H$–$^{15}N$ NOE | -2.40 ± 0.03 | -2.72 ± 0.08 | -3.00 ± 0.03 | -2.63 ± 0.06 | -2.72 ± 0.09 | -3.02 ± 0.13 |
| **600 Mhz** | | | | | | |
| $^{15}N$ $R_1$ ($s^{-1}$) | 0.27 ± 0.01 | 0.59 ± 0.04 | 0.80 ± 0.02 | 0.29 ± 0.02 | 0.54 ± 0.08 | 0.36 ± 0.02 |
| $^1H$–$^{15}N$ NOE | -3.07 ± 0.06 | -2.94 ± 0.17 | -3.17 ± 0.04 | -3.24 ± 0.10 | -3.08 ± 0.20 | -3.08 ± 0.13 |
| **Dynamics[B]** | | | | | | |
| $S^2_{axis}$ | 0.22 ± 0.02 | 0.46 ± 0.06 | 0.48 ± 0.01 | 0.26 ± 0.02 | 0.49 ± 0.06 | 0.39 ± 0.01 |
| $\tau_f$ (ps) | 3 ± 1 | 23 ± 2 | 113 ± 42 | 8 ± 4 | 26 ± 12 | 27 ± 7 |
| $\tau_i$ (ps) | 222 ± 16 | 249 ± 42 | 82 ± 132 | 163 ± 43 | 184 ± 87 | 36 ± 19 |

Relaxation rates and dynamics parameters determined for Complex I and Complex II, the latter of which contains a phosphorodithioate group at the ion pair with Lys57. **A)** The initial rate for intrinsically biexponential $^{15}N$ transverse relaxation of $NH_3^+$. **B)** Symbols are defined in Figure 3.25. The molecular rotational correlation time and anisotropy were determined to be 10.6 ns and 2.1, respectively, from backbone $^{15}N$ relaxation data.

Order parameters for Lys3, Lys55, and Lys57 have low values ($S^2$<0.5), suggesting a very dynamic interaction between lysine side-chains and the DNA phosphate backbone (summarized in Table 3.4). This is surprising given the expectation that hydrogen-bonding and short-range electrostatics from the ion pairing should cause a higher degree of motional restriction for the $NH_3^+$ group, which would result in values much higher. The order parameters here were comparable to those for intra-molecular $NH_3^+$ hydrogen bonding in human ubiquitin[322]. It is suggested that by retaining a mobile motional character, ion pairing of the dynamic type minimizes the loss of conformational entropy that occurs upon complex formation and is favorable for protein-DNA association.

The $S^2$ parameters for the Lys3 group in Complex I and Complex II (I: 0.22; II: 0.26) were lower than those of Lys55 (I: 0.46, II: 0.49) and Lys57 (I: 0.48; II: 0.39). This agrees with the homology model (Fig. 3.8, Fig. 3.11) that places Lys3 towards the outside edge of the complex with greater exposure to solvent, whereas Lys55 and Lys57 are more shielded. The previously covered $^{h3}J_{NP}$ coupling data suggested a major presence of CIP for the intermolecular ion pairs involving Lys55

and Lys57 $NH_3^+$ groups, and weaker evidence of CIP for Lys3 (Fig. 3.26 bottom). Additional support is provided by C—N bond-rotation correlation times, $\tau_f$, which show of $NH_3^+$ group rotation for Lys55 (I: 23 $\pm$ 2 ps; II: 26 $\pm$ 12 ps) and Lys57 (I: 113 $\pm$ 42 ps; II: 27 $\pm$ 7 ps) is much slower than that of Lys3 $NH_3^+$ group (I: 23 $\pm$ 2 ps; II: 26 $\pm$ 12 ps) (Table 3.4). Hydrogen bonds in the CIP state would render slower bond rotations of Lys55 and Lys57 $NH_3^+$ groups and reduced presence of CIP state for Lys3 $NH_3^+$ may be increasing its rotation. The faster C—N bond rotation of Lys3 $NH_3^+$ group implicates a higher population of the SIP state with no hydrogen bonds with DNA. The small $|^{h3}J_{NP}|$, short $\tau_f$, and small $S^2_{\text{axis}}$ collectively suggest a low population of the CIP state for the ion pair of Lys3 $NH_3^+$ group. The equilibrium between the CIP and SIP states is shifted toward the SIP, whereas the corresponding equilibria for Lys55 and Lys57 appears shifted toward the CIP.

Transient breakage of the hydrogen bonds occurs either from rotational permutations of hydrogen atoms within the CIP state for the $NH_3^+$ (propeller adjustment) or the complete transition from CIP to SIP. Given the major presence of CIP, it is likely that the time scale of the transient hydrogen-bond breakage for Lys55 and Lys57 $NH_3^+$ groups is faster than or comparable to their bond-rotation correlation times $\tau_f$. A similar relation between $\tau_f$ and hydrogen-bonding lifetimes for Lys $NH_3^+$ groups was previously seen in ubiquitin molecular dynamics simulations[322]. The $\tau_f$ and $\tau_i$ data in the current study suggest that the transient breakage of hydrogen bonds between Lys $NH_3^+$ and DNA phosphate/phosphorodithioate groups occurs on a subnanosecond time scale (Table 3.4).

### RESULT: EFFECT OF DITHIOATION ON ION PAIRING DYNAMICS

Upon dithioation, the Lys57 $S^2_{\text{axis}}$ order parameter is reduced from 0.49 in Complex I to 0.39 in Complex II. This suggests that ion pairing between a Lys $NH_3^+$ group and sulfur atoms ($NH_3^+\cdots S$) is more dynamic than $NH_3^+\cdots O$ ion pairs. This observation is interesting considering that the potential energy surface for an H···S hydrogen bond is flatter than that for an H···O hydrogen bond[403]. The term “flatter” means that a slight deviation from ideal hydrogen-bond geometry causes only

marginal enthalpic penalty, so a H···S hydrogen is able to adopt a wider conformational space without significant enthalpic loss.

Order parameters for fixed-length bond vectors provide information only on orientational distributions, but do not provide the breadth of positional distributions or any information on translational motions[404]. However, by assuming a motional model for the bond vectors (Fig. 3.27), qualitative information on conformational entropy can be obtained from such order parameters[405-408]. Although the entropy for a whole side chain can be empirically estimated for relatively short amino acid side chains from order parameters measured only at the tip of the side chain, such relationships are less well-defined for longer amino acid side chains such as lysine and arginine[406,409]. One method for interpreting conformational entropy, which is covered by Yang and Kay[410], is the popular "diffusion-in-a-cone"[408] model. Using this approach the entropy difference can be expressed as:

$$\Delta S = k_{\mathrm{B}} \ln\left\{\left[3-\left(1+8S_{\mathrm{axis,a}}\right)^{1/2}\right]/\left[3-\left(1+8S_{\mathrm{axis,b}}\right)^{1/2}\right]\right\}$$

Based on the experimental order parameters $S^2_{\mathrm{axis}}$ observed for Lys57 $NH_3^+$ groups in Complexes I and II, the increase in entropy for the symmetry axis of the $NH_3^+$ group by the oxygen-to-sulfur substitution was estimated to be 0.4 cal·mol$^{-1}$·K$^{-1}$. Of course this model is crude, the diffusion-in-a-cone concept may be too simplistic for ion pairs undergoing the CIP-SIP transitions however this approach does yield a good first approximation for the reorientational entropic consequence of phosphate dithioation on ion pairing.

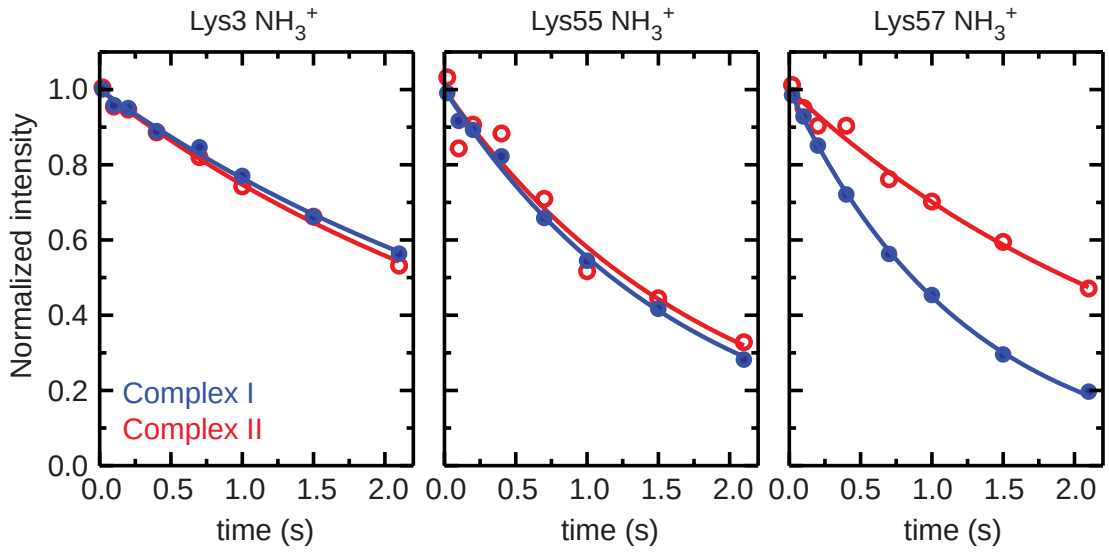



**Fig. 3.29) $^{15}$N Longitudinal Relaxation of Lys Sid-Chain $NH_3^+$ Groups at the Protein–DNA Interfaces of Complexes I (PO) and II (PS2)**

Upon oxygen-to-sulfur substitution in the DNA phosphate group, Lys57 $NH_3^+$ group exhibited substantially different relaxation behavior. $^{15}$N relaxation of the other $NH_3^+$ groups, where both complexes contained normal phosphates, was virtually unaffected.

In regards to rotational entropy it was also observed that the C—N bond rotation correlation time $\tau_f$ is significantly faster for Lys57 $NH_3^+$ ion pairing with a phosphorodithioate group (Table 3.4). This produces substantially different $^{15}$N $R_1$ rates of Lys57 $NH_3^+$ groups in Complexes I and II (Fig. 3.29, Table 3.4). As was shown previously with $CH_3$ groups[411], the rotational entropy ($S_{rot}$) of an $NH_3^+$ group is indirectly related to the bond rotation kinetics because the probability distribution function for rotation depends on the energy barrier for rotation. Applying the Eyring equation in transition-state theory, experimental $NH_3^+$ rotation $\tau_f$ data combined with the the analytical expression of $S_{rot}$[411] suggest that $NH_3^+$ rotational entropy increases by +0.4 cal·mol$^{-1}$·K$^{-1}$ upon dithioation of the phosphate. The cumulative entropic effect of dithioation in ion pairing (*i.e.*, reorientational + rotational) caused by mobilizing the $NH_3^+$ group in ion pairs of $NH_3^+$···S vs. $NH_3^+$···O is thus estimated to be ~0.8 cal·mol$^{-1}$·K$^{-1}$.

## 3.5) Original Scientific Contributions

### NEW BIOPHYSICAL INFORMATION ON THE HOXD9 HOMEODOMAIN AND HOXD9-DNA COMPLEX

- HoxD9/DNA interactions have been further characterized, adding to homeodomain biophysical literature. Here the thermal stability of HoxD9 homeodomain is reported, with a melting temperature of 48.8°C (Fig. 3.16) in the experimental buffer used, so HoxD9 will be folded at ~35°C.
- Room temperature circular dichroism spectroscopy (CD) shows that the HoxD9 homeodomain does not demonstrate the typical homeodomain tendency to fold into a recognition α-helix upon protein-DNA complex formation (Fig. 3.14). This is important for homeodomain calorimetry because α-helix folding involves hydrogen bond formation, which produces heat that obfuscates the apparent enthalpy of binding for homeodomains.
- The enthalpic signature of HoxD9-DNA interaction was determined by Isothermal Titration Calorimetry (ITC) over the temperature range of 5°C-35°C. Enthalpy measurements reveal multiple association processes occurring with the 24-base pair DNA duplex: an apparent high-affinity HoxD9 association and then low-affinity non-specific association. Saturation occurs at molar ratios just beyond ~3:1 HoxD9:DNA (Fig. 3.19). Specific and non-specific enthalpic signatures are distinctly different at temperatures beyond 10°C, with the high-affinity site becoming more exothermic as temperature increases while the non-specific binding becomes exceptionally endothermic (Fig. 3.20). The "cross-over" point between endothermic and exothermic binding reaction for the high-affinity site is approximately 20°C (Fig. 3.21). Future experiments should avoid this temperature to improve the observed enthalpy signal-to-noise ratio of the calorimeter.

### DITHIOATION AT LYS57 IMPROVES HOXD9 BINDING AFFINITY

- HoxD9-DNA binding was probed using fluorescence polarization assays. These experiments show that PS2 substitution at the phosphate interacting with Lys57 results in a 2½-fold improvement in binding affinity, from 32.9 $\pm$ 7.7 nM to 12.6 $\pm$ 3.9 nM (Fig. 3.18). This corresponds to a free energy change in the favorable direction of $\Delta\Delta G° = -0.6$ kcal·mol$^{-1}$.

### PHOSPHORODITHIOATION AS A MEANS FOR PEAK ASSIGNMENT IN LYSINE HISQC NMR

- Lysine-selective two-dimensional (2D) heteronuclear in-phase single quantum coherence (HISQC) NMR spectra revealed that $^{1}H-^{15}N$ cross-peaks for interacting side-chain $NH_3^+$ groups are severely perturbed upon oxygen-to-sulfur phosphorodithioation (Fig. 3.28.A). The $^{1}H$ and $^{15}N$ resonances of Lys57 $NH_3^+$ group were changed $\Delta\delta$ = -0.21 ppm and $\Delta\delta$ = 0.78 ppm, whereas signals from Lys3 and Lys55 $NH_3^+$ groups (which are unmodified) remained almost unchanged. This unique chemical shift upon dithioation can be exploited as a means for NMR peak identification in HISQC experiments for complexes typically too large for conventional resonance assignment.

### MOBILITY OF LYS $NH_3^+$ GROUPS IN ION PAIRS WITH DNA

- The mobility of three HoxD9 Lys $NH_3^+$ groups (Lys3, Lys55, Lys57) was probed by measuring nuclear magnetic resonance (NMR) using relaxation rates $R_1$, $R_2$, and $R(4N_zH_zH_z)$ and $^{1}H$-$^{15}N$ nuclear Overhauser enhancement (NOE). From this relaxation data order parameters were determined for the amino-group symmetry axis $S^2_{axis}$, C−N bond rotational correlation times $\tau_f$, and reorientational correlation times $\tau_i$, for Lys $NH_3^+$ groups (Fig. 3.27). The order parameters axis $S^2_{axis}$ were low ($S^2_{axis} < 0.5$) (Table 3.4), indicating that the ion pairs between Lys side-chain $NH_3^+$ and phosphate groups are highly mobile, despite short-range electrostatic interactions. This high degree of

flexibility of the intermolecular ion pairs should minimize the side-chain conformational entropic costs for protein−DNA association.

### DIRECT EVIDENCE FOR THE CIP STATE IN PROTEIN-DNA ION PAIR INTERACTIONS

- Hydrogen-bond scalar coupling $^{h3}J_{NP}$ between lysine side-chain $^{15}N$ and DNA $^{31}P$ nuclei provides direct evidence for the contact ion pair (CIP) state in protein-DNA interactions (Fig. 3.25). Spectra recorded from a 2-D heteronuclear correlation $H_3(N)P$ experiment show clear intermolecular $^{1}H-^{31}P$ correlation signals (Fig. 3.26 top), indicating at least part-time presence of the CIP states for the ion pairs between Lys $NH_3^+$ and DNA phosphate/phosphorodithioate groups.
- A spin−echo $^{h3}J_{NP}$-modulation constant-time HISQC experiment determined the absolute values of $^{h3}J_{NP}$ coupling constants (Fig. 3.26 bottom). These values were comparable to the average of density functional theory (DFT) predicted constants for highly homologous protein−DNA complexes (Table 3.3). Hydrogen-bond scalar coupling is highly sensitive to subtle differences in hydrogen-bonding geometry, and the $|^{h3}J_{NP}|$ constants suggest that the CIP state has a major presence in solution for the ion pairs of Lys55 and Lys57 $NH_3^+$ groups with a DNA phosphate group, and a low presence for Lys3, which is disordered in most homeodomain crystal structures.
- More evidence for the CIP state comes from the C−N bond rotational correlation times, $\tau_f$. Lys55 and Lys57 $NH_3^+$, which interact with DNA, have $\tau_f$ values that are an order of magnitude slower than Lys3 $NH_3^+$ (Table 3.4), which faces the solvent in our homology model and has a harder time curling back and interacting with DNA during MD simulations.

### HYDROGEN BONDING DYNAMICS OF INTERMOLECULAR ION PAIRS

- This present work provides the first experiment-based perspective of the ion pair dynamics in a biological macromolecular system. Hydrogen bonds with phosphate in the CIP state should render slower bond rotations of Lys55 and Lys57 NH3+ groups, observed here using C−N bond rotational correlation

times $\tau_f$, and reorientational correlation times $\tau_i$. The $\tau_f$ and $\tau_i$ data in this study suggest that the transient breakage of hydrogen bonds between Lys $NH_3^+$ and DNA phosphate/phosphorodithioate groups occurs on a subnanosecond time scale (Table 3.4).

### OXYGEN-TO-SULFUR SUBSTITUTION ENHANCES DYNAMICS OF ION PAIRING

- The $S^2_{axis}$ data indicate that Lys57 $NH_3^+$ is more mobile in an ion pair with the phosphorodithioate group than with the phosphate group (Table 3.4). The $^{h3}J_{NP}$ coupling for phosphorodithioate−Lys57 $NH_3^+$ ion pair (0.18 Hz) was found to be significantly smaller than that for phosphate−Lys57 NH3+ ion pair (0.33 Hz) (Fig. 3.26 bottom). The oxygen-to-sulfur substitution in DNA phosphate makes the intermolecular ion pair more dynamic, which seems to contribute to the affinity enhancement of this type of substitution.
- Using the order parameters $S^2_{axis}$ for Lys57 $NH_3^+$ groups in contact with normal and dithioated phosphates, the increase in entropy for the symmetry axis (reorientational entropy $S_{reorient}$) of the $NH_3^+$ group was estimated through the Yang and Kay model (see text) to be $+0.4$ cal·mol$^{-1}$·K$^{-1}$.
- The C−N bond rotation correlation time $\tau_f$ is significantly faster for Lys57 $NH_3^+$ group in the ion pair with a phosphorodithioate group (Table 3.4). This causes substantially different $^{15}N$ $R_1$ rates of Lys57 $NH_3^+$ groups in PO and PS2 complexes. The rotational entropy ($S_{rot}$) is indirectly related to bond rotational kinetics (see text), and the experimental $\tau_f$ values along with the analytical expression for $S_{rot}$ suggest that $NH_3^+$ rotational entropy increases by $+0.4$ cal·mol$^{-1}$·K$^{-1}$ upon dithioation of the phosphate.
- The overall increase in entropy (*i.e.* reorientional + rotational) upon dithioation of the phosphate in a protein-DNA ion pair is thus estimated to be $\sim +0.8$ cal·mol$^{-1}$·K$^{-1}$.

# CONCLUSION

## 4.1) Dissertation Overview

This dissertation studies biophysical effects of sulfur-for-oxygen substitution in the phosphate backbone of deoxyribonucleic acid. A dithioated phosphate is one in which both non-bridging phosphoryl oxygen atoms are replaced with sulfur atoms. Sulfur has a larger van der Waals radius therefore the phosphorodithioate has altered bond angles and lengths, which significantly alters electron density about the phosphorous, imparting unique biophysical properties to the moiety. Phosphate dithioation in DNA is useful in biomolecular experiments because of altered binding properties and potential for therapeutic applications due to enhanced nuclease resistance. Just about every biophysical measurement performed shows marked differences between normal and dithioated phosphates.

All modified nucleic acids were synthesized *de novo* in the laboratory, therefore Chapter 1 of this manuscript begins by highlighting major developments in synthetic nucleic acid chemistry that ultimately lead to the modern PIII phosphoramidite method for the production of short oligonucleotides. The chemistry of each reaction step in this type of synthesis is described in detail. Using this technique, inserting a single sulfur atom to make monophosphorothioated DNA is a simple alteration of the oxidation step in the phosphoramidite method. However, there are two oxygen locations for the sulfur and monothioation by the traditional sulfurization step produces a chiral center at the phosphorus atom, which limits purity due to the diastereomeric mixture of $S_P$ and $R_P$ monothioates. For therapeutic applications and potentially generalizing these dithiophosphate results to other sulfur-for-oxygen systems, both non-bridging oxygens must be replaced to make a dithioate, which is achiral. A historical record of various dithioation approaches is presented in detail, which ultimately leads to the development of the thiophosphoramidite that brings with it the additional sulfur atom. The modern chemistry of dithioation is covered, as well as the development

of various sulfurization reagents. The standard work-up and purification strategies for oligonucleotides and dithioated oligos post-synthesis are then discussed. Chapter 1 concludes by covering the quality control results for all oligonucleotides synthesized for the work in this dissertation, and includes discussion on various issues that arise when working with dithioates.

Phosphorodithioated DNA is a model system to probe sulfur-for-oxygen substitution because of its macromolecular size and outwards-facing sulfurs that simplify experimental design. There is a rich history of oligonucleotide biophysical studies with established protocols that enable clear comparison of normal and dithioated phosphates. The most well studied phenomenon is the melting experiment, or nucleic acid hybridization, in which two complementary single strand oligonucleotides anneal in a temperature-dependent manner into a single duplex. Such experiments provide insight into the thermodynamics of the transition between helix and coil states. Chapter 2 of this dissertation introduces the topic of thermodynamics at the introductory level. From first principles an attempt is made to explain various thermodynamic state functions and how they can be derived from the fundamental laws of thermodynamics. Discussion then focuses on nucleic acid thermodynamics and some biophysical properties of DNA that have been studied empirically and theoretically over the past several decades.

Since the melting behavior of DNA is dependent on GC content and ultimately sequence content, there has been a great effort in improving the ability to predict the melting behavior of DNA based solely on its sequence. There are obvious benefits of this knowledge in areas such as PCR and DNA micro-array applications. One successful approach is the Nearest-Neighbor (NN) doublet model, in which a DNA sequence is deconvoluted into a linear combination of base pair doublets. With four base choices and two basepairs there are 16 total possible combinations, of which redundancy reduces to 10 unique NN parameters that describe an oligonucleotide sequence. Several NN parameter sets have been derived from empirical melting of normal DNA duplexes, but only a handful of dithioate melting studies have been performed which is not enough to derive parameters for such modifications. The research presented in Chapter 2 is the first large-scale thermodynamic analysis of phosphorodithioated DNA. A set of 40 duplexes was

synthesized that contain all 10 NN motifs and dithioation of all three possibilities: top, bottom, and both internucleotide linkages of the NN doublet. Upon phosphate modification the 5'→3' symmetry of NNs is broken, and the full 16 possible NN motifs are needed to describe dithioated NNs instead of 10. A fluorescence resonance energy transfer melting experiment was developed in collaboration with Dr. Ricahrd Owczarzy of the biotechnology company Integrated DNA Technologies (Coralville, IA) to test dithioate differences with unmodified duplexes. This fluorescence approach enabled the assay to be completed in a high-throughput manner using Real Time PCR machines in 96-well format. Thousands of melting experiments were performed over 13 concentrations and thermodynamic information was extracted. Various approaches to extracting thermodynamic content from equilibrium melting curves are presented in detail in Chapter 2. Dithioation was shown to significantly reduce the stability and melting temperature of the DNA duplex. Upon dithioation the free energy of the duplex became more unfavorable (positive), the enthalpic term became more unfavorable (positive), and the entropic term became more favorable (positive). A table representing thermodynamic difference values for the NNs was produced. There was obvious sequence dependence to the magnitude of thermodynamic difference values. The original aim was to develop unique parameters describing all 10 dithioated NN doublets, but since a few of the duplexes melting in a non-two-state manner these NNs could not be fit. Instead, a rigorous set of NN models with a reduced number of parameters (less than 16 NNs) was analyzed. These models contained simplifying descriptors such as base identity (A/C/G/T) and purine/pyrimidine character of the bases 5'- and 3'- to dithioation site. Several models accurately described the data, and the general observation was that the identity and purine/pyrimidine character of the 3'-base was most important for determining the magnitude of dithioate thermodynamic difference values. Using the models presented, accurate predictions for the melting behavior and thermodynamic difference values for dithioated DNA molecules can be calculated. It isn't quite the full unique 16 parameter PS2 NNs, but nevertheless these reduced parameter models provide valuable insight for researchers working on dithioated DNA molecules. The melting information in this dissertation is high quality and can be combined with a future study to extract the

full 10 unique NN parameters. Interestingly, it was found that the effect of dithioation upon $T_m$ was also correlated to BI-BII propensity of the NN. More rigid NN pairs had larger $T_m$ perturbations upon dithioation compared to those of more flexible NN pairs.

The final chapter of this dissertation focuses on dithioated phosphates in protein-DNA interactions which came from a collaboration with the lab of Dr. Junji Iwahara of the University of Texas Medical Branch (Galveston, TX). Chapter 3 commences with general discussion on such complexes and covers finding that amino acid side-chain interactions with the phosphate backbone are the most prevalent type of contact at the protein-DNA interface. The shape of the phosphate backbone in DNA duplexes, which is sequence dependent, is what determines protein-DNA specificity (*i.e.* the arrangement of phosphates in 3-D space). The system used for experimentation, HoxD9 helix-turn-helix homeodomain binding to a 24bp duplex containing its recognition sequence, is introduced and notable literature on this protein and other homeodomains is covered. Specific reasons for choosing this system, such as previous work by Iwahara, are also discussed. This is followed by an introduction into the novel HISQC NMR experiment developed by Iwahara and how it is used to observe Lysine $NH_3^+$ protons in protein-DNA complexes.

The first experiments reported are general biophysical experiments for HoxD9-DNA complex formation. These include thermal stability studies of HoxD9, circular dichroism spectroscopy, and fluorescence polarization binding assays with both the normal and dithioated duplexes. After this a wide range of isothermal titration calorimetry studies are presented, which reveal the enthalpic signature of HoxD9 complex formation.

A rigorous set of NMR experiments is then presented that probes the ion pairing dynamics between Lys $NH_3^+$ groups and DNA phosphates/ phosphorodithioates. Various order parameters are developed from relaxation measurements and provide insight into the general nature of intermolecular ion pairs as being very dynamic with high mobility despite electrostatic restriction from ion pairing. This includes discussion on the time scale of hydrogen bonding dynamics of ion pairs and the conversion between contact-ion pair states and

separated-ion pair states in protein-DNA interactions. Finally, comparison is made between the ion pairing dynamics and relative mobility of Lys $NH_3^+$ groups in ion pairs with normal phosphates and phosphorodithioates. From these studies it was shown that phosphorodithioation results in enhanced dynamics of lysine side chain that is involved in site-specific ion pairing.

When combined with calorimetry data, which shows no obvious enthalpic difference between the two complexes, this NMR data suggests that the affinity enhancement observed in binding assays after site-specific phosphorodithioation is primarily entropic in origin.

## 4.2) Limitations of This Work and Future Directions

In retrospect it has been beneficial to scrutinize this work for limitations in experimental design that could affect certain conclusions. Some points became apparent in the preparation of this dissertation, and the following section is used to develop these criticisms. After this discussion, some future studies are mentioned, which would be the logical next steps in dithioate DNA research.

In the second chapter on analyzing dithioated phosphates in DNA duplex thermodynamics, one major limitation of the experimental design was that only one sequence for each NN motif was used. Despite the robustness of the melting study (>4000 experiments that covered all 10 normal PO NNs and 16 PS2 NNs), the construct design only contained one sequence for each NN moiety. In retrospect, multiple sequences for each NN and a larger experiment would have generated more data to fit. The lone motif approach became a problem after some duplexes melted in a non-two-state manner, which prevented the determination of all 16 unique PS2 NN parameters. Fitting these would have been possible with a larger dataset. In a similar vein, the dithioate duplex design was somewhat limited because only isolated phosphorodithioates were tested. Some questions are not answered with this sequence design, such as: consecutive dithioation, dithioation of duplex ends, and dithioates across the minor/major groove from each other, *etc*.

In regards to using the fluorescence-resonance transfer (FRET) assay for oligonucleotide thermodynamics, it should be mentioned that this was a non-

traditional approach for empirical determination of thermodynamic values. Historically, UV melting has been used to determine nucleic acid hybridization thermodynamics but this method cannot be scaled up in high-throughput format. The FRET method is amenable to "difference-value" studies, in which an unmodified and modified duplex are compared. In designing these experiments, this study relied upon other authors' work in validation of the FRET assay for thermodynamic difference value determination (*i.e.* FRET and UV values were not compared). This study could have used an independent confirmation by comparing that traditional UV difference values for a fluorophore-free duplex with those obtained by FRET analysis.

In preparing the third chapter on dithioated phosphates in protein-DNA interactions, a few issues became apparent with the study designs. First off, the DNA construct used for these studies was too large. After the ITC experiments showed potential non-specific binding, the 24-bp DNA construct should have been truncated to minimize the amount of non-specific HoxD9 association. The extra base pairs are unnecessary and potentially complicate analysis. In the future it would be suggested that the duplex be truncated to the minimum length required for specificity in the interaction to remain, yet minimize non-specific heats of association for non-specific interactions elsewhere on the duplex.

One aspect of the studies that was an unavoidable limitation was the temperature discrepancies between experiments. Unfortunately the methods utilized have optimal temperatures for best assay performance, but there is no agreement among the techniques. Fluorescence polarization binding assays were performed at 20°C, but the NMR relaxation measurements were performed at 35°C. This discrepancy could not be remedied as the NMR experiments for the observation of the three lysines involved in protein-DNA contact (Lys3, Lys55, Lys57) are optimized for 35°C, but fluorescence was unreliable at this temperature. This potentially complicates compilation of their respective results. Similarly, CD studies were performed at room temperature (22°C), which could not be thermostated.

One of the major findings in this dissertation was that site-specific dithioation at Lys57 enhanced affinity of the HoxD9-DNA association. However, the

binding affinities were determined using only one assay, which was fluorescence polarization. Ideally it would be better to employ a second method for validating the slight 2½-fold affinity improvement upon dithioation at the Lys57-interacting phosphate. Since the HoxD9-DNA association can apparently handle an oligo-attached fluorophore, perhaps a binding assay incorporating anchor attachment from the oligo would be achievable.

The major limitation of Chapter 3 of this dissertation is that it generalizes for all dithiophosphate-lysine interactions from just one site-specific substitution in a single system, using only one type of binding data. Therefore, the chance that this behavior is simply just an anomaly cannot be discredited. However, after having spent many years working with dithioates on a number of systems in the lab of Dr. Gorenstein, I confidently claim that the results presented here are an appropriate characterization of dithioate behavior in protein-DNA interactions. Dithioates seem to be stickier, have slower off-rates, and improve protein-DNA affinity constants. The focus of this dissertation, the HoxD9 homeodomain system, was the best approach for demonstrating these dithioation effects.

During the writing process, various starting points for future dithioate DNA studies became self-evident. The next step could be in any number of areas, but three research topics stood out as being appropriate for future students. The first type of "low-hanging fruit" project would involve continuing the PS2 FRET-based thermodynamic studies. The results presented in Chapter 2 clearly demonstrate the robustness of the FRET-based approach, but more sequence diversity and additional thermodynamic data are needed before singular PS2 NN parameters can be determined. These types of studies have a high probability of success, especially if the analysis methods and experimental design utilized in this dissertation are duplicated.

The next type of "middle-hanging fruit" project would be the structural characterization of a dithioate-lysine ion pair. Currently there is no high-resolution X-ray structure of a dithiophosphate, either in native duplex or protein complex. However, since the CD data presented in Chapter 3 shows no major structural perturbation of the protein-DNA complex upon single-site dithioation at a lysine, it is reasonable to expect PO and PS2 DNA to have similar crystallization conditions.

Many homeodomain structures have already been solved by X-Ray crystallography with their associated DNA duplexes, and it is likely that one of these systems is amenable to dithioation and crystallization. This type of project is fairly straightforward as it involves duplicating previous crystallization conditions, but its potential for payout is high.

A more ambitious "high-hanging fruit" project would be the determination of actual dithioate backbone torsional parameters using experimental NMR. Currently little is known about the effect of phosphate dithioation on DNA backbone torsional angles, which is necessary for accurate *in silico* predictions and molecular dynamics simulations. These dithioate torsional angle parameters can be determined directly from NMR structural constraints. Such experimental evidence would provide structural insight into the mechanism by which dithioates increase configurational entropy of ion pairs and enhance protein-DNA binding affinity.

## Bibliography


1. Kossel, A. Ueber die bildung von thymin aus fischsperma. *Zeitschrift für Physikalische Chemie* **22,** 188–190 (1897).
2. Miescher, F. Ueber die chemische zusammensetzung der eiterzellen. *Medicinisch-Chemische Untersuchungen* **4,** 441–460 (1871).
3. Kossel, A. Ueber das nuclein der hefe. *Zeitschrift für Physikalische Chemie* **3,** 284–291 (1879).
4. Levene, P. A. & Jacobs, W. A. Über die pentose in den nucleinsäuren. *Berichte der Deutschen Chemischen Gesellschaft* **42,** 3247–3251 (1909).
5. Altmann, R. Ueber Nucleinsäuren. *Archiv für Anatomie und Entwickelungsgeschichte* 524–536 (1889).
6. Steudel, H. & Kossel, A. Ueber das thymin. *Zeitschrift für Physikalische Chemie* **29,** 303–304 (1900).
7. Kossel, A. Ueber das nuclein der hefe. *Zeitschrift für Physikalische Chemie* **4,** 290–295 (1890).
8. Kossel, A. Ueber die basischen stoffe des zellkerns. *Zeitschrift für Physikalische Chemie* **22,** 176–187 (1887).
9. Kossel, A. Weitere beiträge zur chemie des zellkerns. *Zeitschrift für Physikalische Chemie* **12,** 248–264 (1886).
10. Kossel, A. Ueber das adenin. *Zeitschrift für Physikalische Chemie* **12,** 241–253 (1888).
11. Kossel, A. Üeber guanin. *Zeitschrift für Physikalische Chemie* **8,** 404–410 (1884).
12. Kossel, A. Ueber einen peptonartigen bestandtheil des zellkerns. *Zeitschrift für Physikalische Chemie* **8,** 511–515 (1884).
13. Kossel, A. Zur chemie des zellkerns. *Zeitschrift für Physikalische Chemie* **7,** 7–22 (1883).
14. Levene, P. A. & Medigreceanu, F. The action of gastro-intestinal juices on nucleic acids. *Journal of Biological Chemistry* **9,** 375–387 (1911).
15. Levene, P. A. & Medigreceanu, F. On nucleases. *Journal of Biological Chemistry* **9,** 389–402 (1911).
16. Levene, P. A. On the biochemistry of nucleic acids. *Journal of the American Chemical Society* **32,** 231–240 (1910).
17. Levene, P. A. & LaForge, F. B. Uber die hefenucleinsaure. V. Die struktur der pyrimidin-nucleoside. *Chem. Ber* **45,** 608 (1912).
18. Levene, P. A. & London, E. J. On the structure of thymonucleic acid. *Science* **68,** 572–573 (1928).
19. Levene, P. A. & London, E. S. The structure of thymonucleic acid. *Journal of Biological Chemistry* **83,** 793–802 (1929).
20. Levene, P. A. Hydrolysis of yeast nucleic acid with dilute alkali at room temperature. (Conditions of Steudel and Peiser.). *Journal of Biological Chemistry* **55,** 9–13 (1923).
21. Levene, P. A. & Mori, T. Ribodesose and xylodesose and their bearing on the structure of thyminose. *Journal of Biological Chemistry* **83,** 803–816 (1929).
22. Levene, P. A., Mikeska, L. A. & Mori, T. On the carbohydrate of thymonucleic acid. *Journal of Biological Chemistry* **85,** 785–787 (1930).
23. Brown, D. M. & Todd, A. R. 13. Nucleotides. Part X. Some observations on the structure and chemical behaviour of the nucleic acids. *Journal of the Chemical Society (Resumed)* 52–58 (1952).
24. Levene, P. A. & London, E. S. Guaninedesoxypentoside from thymus nucleic acid. *Journal of Biological Chemistry* **81,** 711–712 (1929).
25. Levene, P. A. & Tipson, R. S. The ring structure of thymidine. *Journal of Biological Chemistry* **109,** 623–630 (1935).
26. Michelson, A. M. & Todd, A. R. Nucleotides part XXXII. Synthesis of a dithymidine dinucleotide containing a 3′: 5′-internucleotidic linkage. *Journal of the Chemical Society (Resumed)* 2632–2638 (1955).
27. Andersen, W., Dekker, C. A. & Todd, A. R. 513. Nucleotides. Part XIV. A method suitable for the large-scale preparation of deoxyribonucleosides from deoxyribonucleic acids.

*Journal of the Chemical Society (Resumed)* 2721–2724 (1952).
28. Klein, W. Über Nucleophosphatase. *Zeitschrift für Physikalische Chemie* **207,** 125–140 (1932).
29. Cohn, W. E. The anion-exchange separation of ribonucleotides. *Journal of the American Chemical Society* **72,** 1471–1478 (1950).
30. Michelson, A. M. & Todd, A. R. 198. Nucleotides. Part XX. Mononucleotides derived from thymidine. Identity of thymidylic acid from natural sources with thymidine-5′ phosphate. *Journal of the Chemical Society (Resumed)* 951–956 (1953).
31. Michelson, A. M. & Todd, A. R. Deoxyribonucleosides and related compounds. Part V. cyclothymidines and other thymidine derivatives. The configuration at the glycosidic centre in thymidine. *Journal of the Chemical Society (Resumed)* 816–823 (1955).
32. Michelson, A. M. & Todd, A. R. Nucleotides. Part XXIII. Mononucleotides derived from deoxycytidine. Note on the structure of cytidylic acids a and b. *Journal of the Chemical Society (Resumed)* 34–40 (1954).
33. Hayes, D. H., Michelson, A. M. & Todd, A. R. Nucleotides. Part XXX. Mononucleotides derived from deoxyadenosine and deoxyguanosine. *Journal of the Chemical Society (Resumed)* 808–815 (1955).
34. Khorana, H. G., Tener, G. M., Moffatt, J. G. & Pol, E. H. A new approach to the synthesis of polynucleotides. *Chemistry and Industry* **1523,** (1956).
35. Khorana, H. G., Razzell, W. E., Gilham, P. T., Tener, G. M. & Pol, E. H. Syntheses of dideoxyribonucleotides. *Journal of the American Chemical Society* **79,** 1002–1003 (1957).
36. Jacob, T. M. & Khorana, H. G. Studies on polynucleotides. XXX. A comparative study of reagents for the synthesis of the C3‘-C5’ internucleotidic linkage. *Journal of the American Chemical Society* **86,** 1630–1635 (1964).
37. Marvel, C. S., Whitson, J. & Johnston, H. W. The dissociation of hexaarylethanes. XV. Methoxyl substituents. *Journal of the American Chemical Society* **66,** 415–417 (1944).
38. Hayatsu, H. & Khorana, H. G. Deoxyribooligonucleotide synthesis on a polymer support. *Journal of the American Chemical Society* **88,** 3182–3183 (1966).
39. Hogenkamp, H. & Oikawa, T. G. The Synthesis and Properties of 2', 5“-Dideoxyadenosylcobalamin and 5”-Deoxythymidylcobalamin. *Journal of Biological Chemistry* **239,** 1911–1916 (1964).
40. Gilham, P. T. & Khorana, H. G. Studies on polynucleotides. I. A new and general method for the chemical synthesis of the C5‘-C3’ internucleotidic linkage. Syntheses of deoxyribo-dinucleotides. *Journal of the American Chemical Society* **80,** 6212–6222 (1958).
41. Smith, M., Drummond, G. I. & Khorana, H. G. Cyclic phosphates. IV. Ribonucleoside-3‘, 5’ cyclic phosphates. A general method of synthesis and some properties. *Journal of the American Chemical Society* **83,** 698–706 (1961).
42. Smith, M., Rammler, D. H., Goldberg, I. H. & Khorana, H. G. Studies on polynucleotides. XIV. Specific synthesis of the C3'-C5' interribonucleotide linkage. Syntheses of uridylyl-(3'→ 5')-uridine and uridylyl-(3‘→ 5’)-adenosine. *Journal of the American Chemical Society* **84,** 430–440 (1962).
43. Khorana, H. G. Polynucleotide synthesis and the genetic code. *Federation Proceedings* **24,** 1473–1487 (1965).
44. Khorana, H. G. Polynucleotide synthesis and the genetic code. *Harvey Lectures* **62,** 79–105 (1967).
45. Nirenberg, M. W. & Matthaei, J. H. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences of the United States of America* **47,** 1588 (1961).
46. Holley, R. W. *et al.* Structure of a ribonucleic acid. *Science* **147,** 1462–1465 (1965).
47. Weiss, B. & Richardson, C. C. Enzymatic breakage and joining of deoxyribonucleic acid, I. Repair of single-strand breaks in DNA by an enzyme system from Escherichia coli infected with T4 bacteriophage. *Proceedings of the National Academy of Sciences of the United States of America* **57,** 1021 (1967).
48. Agarwal, K. L. *et al.* Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. *Nature* **227,** 27–34 (1970).
49. Khorana, H. G. *et al.* CIII. Total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast. *Journal of Molecular Biology* **72,** 209–217 (1972).
50. Weber, H. & Khorana, H. G. CIV. Total synthesis of the structural gene for an alanine

transfer ribonucleic acid from yeast. Chemical synthesis of an icosadeoxynucleotide corresponding to the nucleotide sequence 21 to 40. *Journal of Molecular Biology* **72,** 219–249 (1972).

51. Kumar, A., Ohtsuka, E. & Khorana, H. G. CVI. Total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast. Synthesis of two nonanucleotides and a heptanucleotide corresponding to nucleotide sequences 22 to 30, 41 to 49 and 28 to 34. *Journal of Molecular Biology* **72,** 289–307 (1972).
52. Ohtsuka, E., Kumar, A. & Khorana, H. G. CVII. Total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast. Synthesis of a dodecadeoxynucleotide and a hexadeoxynucleotide corresponding to the nucleotide sequences 1 to 12. *Journal of Molecular Biology* **72,** 309–327 (1972).
53. Kumar, A. & Khorana, H. G. CVIII. Total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast. Synthesis of an undecadeoxynucleotide, a decadeoxynucleotide and an octadeoxynucleotide corresponding to the nucleotide sequences 7 to 27. *Journal of Molecular Biology* **72,** 329–349 (1972).
54. Agarwal, K. L., Kumar, A. & Khorana, H. G. CIX. Total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast. Synthesis of a dodecadeoxynucleotide and a decadeoxynucleotide …. *Journal of Molecular Biology* (1972).
55. Caruthers, M. H., Van de Sande, J. H. & Khorana, H. G. CX. Total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast. Synthesis of three decadeoxynucleotides corresponding to the nucleotide sequence 51 to 70. *Journal of Molecular Biology* **72,** 375–405 (1972).
56. Caruthers, M. H. & Khorana, H. G. CXI. Total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast. Synthesis of a dodecadeoxynucleotide and a heptadeoxynucleotide corresponding to the nucleotide sequence 66 to 77. *Journal of Molecular Biology* **72,** 407–426 (1972).
57. Sgaramella, V. & Khorana, H. G. CXII. Total synthesis of the structural gene for an alanine transfer RNA from yeast. Enzymic joining of the chemically synthesized polydeoxynucleotides to form the DNA duplex representing nucleotide sequence 1 to 20. *Journal of Molecular Biology* **72,** 427–444 (1972).
58. Sgaramella, V., Kleppe, K., Terao, T., Gupta, N. K. & Khorana, H. G. CXIII. Total synthesis of the structural gene for an alanine transfer RNA from yeast. Enzymic joining of the chemically synthesized segments to form the DNA duplex corresponding to nucleotide sequence 17 to 50. *Journal of Molecular Biology* **72,** 445–456 (1972).
59. Van de Sande, J. H., Caruthers, M. H., Sgaramella, V., Yamada, T. & Khorana, H. G. CXIV. Total synthesis of the structural gene for an alanine transfer RNA from yeast. Enzymic joining of the chemically synthesized segments to form the DNA duplex corresponding to nucleotide sequence 46 to 77. *Journal of Molecular Biology* **72,** 457–474 (1972).
60. Büchi, H. & Khorana, H. G. CV. Total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast. Chemical synthesis of an icosadeoxyribonucleotide corresponding to the nucleotide sequence 31 to 50. *Journal of Molecular Biology* **72,** 251–288 (1972).
61. Turner, A. F. & Khorana, H. G. Studies on polynucleotides. VI. 1 Experiments on the chemical polymerization of mononucleotides. Oligonucleotides derived from thymidine-3' Phosphate. *Journal of the American Chemical Society* **81,** 4651–4656 (1959).
62. Tener, G. M. 2-Cyanoethyl phosphate and its use in the synthesis of phosphate esters. *Journal of the American Chemical Society* **83,** 159–168 (1961).
63. Letsinger, R. L. & Mahadevan, V. Stepwise synthesis of oligodeoxyribonucleotides on an insoluble polymer support. *Journal of the American Chemical Society* **88,** 5319–5324 (1966).
64. Letsinger, R. L., Caruthers, M. H. & Jerina, D. M. Reactions of nucleosides on polymer supports. Synthesis of Thymidylylthymidylylthymidine. *Biochemistry* **6,** 1379–1388 (1967).
65. Letsinger, R. L. & Ogilvie, K. K. Nucleotide chemistry. XIII. Synthesis of oligothymidylates via phosphotriester intermediates. *Journal of the American Chemical Society* **91,** 3350–3355 (1969).
66. Westheimer, F. H. Pseudo-rotation in the hydrolysis of phosphate esters. *Acc. Chem. Res.* **1,** 70–78 (1968).
67. Catlin, J. C. & Cramer, F. Deoxy oligonucleotide synthesis via the triester method. *The*

*Journal of organic chemistry* **38,** 245–250 (1973).
68. Katagiri, N., Itakura, K. & Narang, S. A. Use of arylsulfonyltriazoles for the synthesis of oligonucleotides by the triester approach. *Journal of the American Chemical Society* **97,** 7332–7337 (1975).
69. Itakura, K., Katagiri, N., Bahl, C. P., Wightman, R. H. & Narang, S. A. Improved triester approach for the synthesis of pentadecathymidylic acid. *Journal of the American Chemical Society* **97,** 7327–7332 (1975).
70. Letsinger, R. L. & Miller, P. S. Nucleotide chemistry. XIV. Protecting groups for nucleosides used in synthesizing oligonucleotides. *Journal of the American Chemical Society* **91,** 3356–3359 (1969).
71. Daub, G. W. & Van Tamelen, E. E. Synthesis of oligoribonucleotides based on the facile cleavage of methyl phosphotriester intermediates. *Journal of the American Chemical Society* **99,** 3526–3528 (1977).
72. Devine, K. G. & Reese, C. B. Highly reactive condensing agents for the synthesis of oligonucleotides by the phosphotriester approach. *Tetrahedron Letters* **27,** 5529–5532 (1986).
73. Letsinger, R. L. & Mahadevan, V. Oligonucleotide synthesis on a polymer support. *Journal of the American Chemical Society* **87,** 3526–3527 (1965).
74. Letsinger, R. L. & Kornet, M. J. Popcorn polymer as a support in multistep syntheses. *Journal of the American Chemical Society* **85,** 3045–3046 (1963).
75. Letsinger, R. L., Kornet, M. J., Mahadevan, V. & Jerina, D. M. Reactions on polymer supports. *Journal of the American Chemical Society* **86,** 5163–5165 (1964).
76. Letsinger, R. L., Finnan, J. L., Heavner, G. A. & Lunsford, W. B. Nucleotide chemistry. XX. Phosphite coupling procedure for generating internucleotide links. *Journal of the American Chemical Society* **97,** 3278–3279 (1975).
77. Letsinger, R. L. & Lunsford, W. B. Synthesis of thymidine oligonucleotides by phosphite triester intermediates. *Journal of the American Chemical Society* **98,** 3655–3661 (1976).
78. Sinha, N. D., Biernat, J. & Köster, H. β-Cyanoethyl N, N-dialkylamino/N-morpholinomonochloro phosphoamidites, new phosphitylating agents facilitating ease of deprotection and work-up of synthesized oligonucleotides. *Tetrahedron Letters* **24,** 5843–5846 (1983).
79. Beaucage, S. L. & Caruthers, M. H. Deoxynucleoside phosphoramidites-a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Letters* **22,** 1859–1862 (1981).
80. McBride, L. J. & Caruthers, M. H. An investigation of several deoxynucleoside phosphoramidites useful for synthesizing deoxyoligonucleotides. *Tetrahedron Letters* **24,** 245–248 (1983).
81. Dörper, T. & Winnacker, E.-L. Improvements in the phosphoramidite procedure for the synthesis of oligodeoxyribonucleotides. *Nucleic Acids Research* **11,** 2575–2584 (1983).
82. Adams, S. P., Kavka, K. S., Wykes, E. J., Holder, S. B. & Galluppi, G. R. Hindered dialkylamino nucleoside phosphite reagents in the synthesis of two DNA 51-mers. *Journal of the American Chemical Society* **105,** 661–663 (1983).
83. Amarnath, V. & Broom, A. D. Chemical synthesis of oligonucleotides. *Chemical Reviews* **77,** 183–217 (1977).
84. Caruthers, M. H. *et al.* New chemical methods for synthesizing polynucleotides. 215 (1980).
85. Matteucci, M. D. & Caruthers, M. H. The synthesis of oligodeoxyprimidines on a polymer support. *Tetrahedron Letters* **21,** 719–722 (1980).
86. Caruthers, M. H. *et al.* Deoxyoligonucleotide Synthesis via the Phosphoramidite Method. *Gene Amplification and Analysis* **3,** 1–26 (1983).
87. Gough, G. R., Brunden, M. J. & Gilham, P. T. Recovery and recycling of synthetic units in the construction of oligodeoxyribonucleotides on solid supports. *Tetrahedron Letters* **22,** 4177–4180 (1981).
88. Usman, N., Ogilvie, K. K., Jiang, M. Y. & Cedergren, R. J. The automated chemical synthesis of long oligoribuncleotides using 2'-O-silylated ribonucleoside 3“-O-phosphoramidites on a controlled-pore glass support: synthesis of a 43-nucleotide sequence similar to the 3”-half molecule of an Escherichia coli formylmethionine tRNA. *Journal of the American Chemical Society* **109,** 7845–7854 (1987).
89. Sinha, N. D., Biernat, J., McManus, J. & Köster, H. Polymer support oligonucleotide

synthesis XVIII1. 2): use of β-cyanoethyi-N, N-dialkylamino-/N-morpholino phosphoramidite of deoxynucleosides for the synthesis of DNA fragments simplifying deprotection and isolation of the final product. *Nucleic Acids Research* **12,** 4539–4557 (1984).

90. Nielsen, J., Taagaard, M., Marugg, J. E., Van Boom, J. H. & Dahl, O. Application of 2-cyanoethyl N, N, N′, N′-tetraisopropylphosphorodiamidite for in situ preparation of deoxyribonucleoside phosphoramidites and their use in polymer-supported synthesis of oligodeoxyribonucleotides. *Nucleic Acids Research* **14,** 7391–7403 (1986).
91. McBride, L. J., Kierzek, R., Beaucage, S. L. & Caruthers, M. H. Nucleotide chemistry. 16. Amidine protecting groups for oligonucleotide synthesis. *Journal of the American Chemical Society* **108,** 2040–2048 (1986).
92. McMinn, D. L. & Greenberg, M. M. Synthesis of oligonucleotides containing 3'-alkyl amines using N-isobutyryl protected deoxyadenosine phosphoramidite. *Tetrahedron Letters* **38,** 3123–3126 (1997).
93. Schulhof, J. C., Molko, D. & Teoule, R. The final deprotection step in oligonucleotide synthesis is reduced to a mild and rapid ammonia treatment by using labile base-protecting groups. *Nucleic Acids Research* **15,** 397–416 (1987).
94. Zhu, Q., Delaney, M. O. & Greenberg, M. M. Observation and elimination of N-acetylation of oligonucleotides prepared using fast-deprotecting phosphoramidites and ultra-mild deprotection. *Bioorganic & medicinal chemistry letters* **11,** 1105–1107 (2001).
95. Reddy, M. P., Hanna, N. B. & Farooqui, F. Ultrafast cleavage and deprotection of oligonucleotides, synthesis and use of CAc derivatives. *Nucleosides & Nucleotides* **16,** 1589–1598 (1997).
96. Berner, S., Mtihlegger, K. & Seliger, H. The reaction of tetrazole with phosphoramidites as a model for the nucleotide coupling STEP. *Nucleosides & Nucleotides* **7,** 763–767 (1988).
97. Nurminen, E. J., Mattinen, J. K. & Lönnberg, H. Kinetics and mechanism of tetrazole-catalyzed phosphoramidite alcoholysis. *Jounal of the Chemical Society - Perkins Transactions 2* 1621–1628 (1998).
98. Berner, S., Mūhlegger, K. & Seliger, H. Studies on the role of tetrazole in the activation of phosphoramidites. *Nucleic Acids Research* **17,** 853–864 (1989).
99. Nurminen, E. J., Mattinen, J. K. & Lönnberg, H. The effectivity of 1H- triazoles and tetrazoles as activators in acid- catalyzed phosphoramidite alcoholysis. *Helvetica chimica acta* **86,** 2005–2008 (2003).
100. Dahl, B. H., Nielsen, J. & Dahl, O. Mechanistic studies on the phosphoramidite coupling reaction in oligonucleotide synthesis. I. Evidence for nudeophilic catalysis by tetrazole and rate variations with the phosphorus substituents. *Nucleic Acids Research* **15,** 1729–1743 (1987).
101. Aronson, S., Epstein, P., Aronson, D. B. & Wieder, G. Ionic equilibriums in pyridine-iodine solutions. *J. Phys. Chem.* **86,** 1035–1037 (1982).
102. Tassaing, T. & Besnard, M. Ionization Reaction in Iodine/Pyridine Solutions: What Can We Learn from Conductivity Measurements, Far-Infrared Spectroscopy, and Raman Scattering? *J. Phys. Chem. A* **101,** 2803–2808 (1997).
103. Okitsu, T., Yumitate, S., Sato, K., In, Y. & Wada, A. Substituent effect of bis (pyridines) iodonium complexes as iodinating reagents: Control of the iodocyclization/oxidation process. *Chemistry - A European Journal* **19,** 4992–4996 (2013).
104. Skowrońska, A., Pakulski, M., Michalski, J., Cooper, D. & Trippett, S. The Arbuzov reaction of triethyl phosphite with elemental iodine. *Tetrahedron Letters* **21,** 321–322 (1980).
105. Nemer, M. J. & Ogilvie, K. K. Ribonucleotide analogues having novel internucleotide linkages. *Tetrahedron Letters* **21,** 4149–4152 (1980).
106. Nemer, M. J. & Ogilvie, K. K. Phosphoramidate analogues of diribonucleoside monophosphates. *Tetrahedron Letters* **21,** 4153–4154 (1980).
107. Ogilvie, K. K. & Nemer, M. J. The synthesis of phosphite analogues of ribonucleotides. *Tetrahedron Letters* **21,** 4145–4148 (1980).
108. Burgers, P. M. & Eckstein, F. Synthesis of dinucleoside monophosphorothioates via addition of sulphur to phosphite triesters. *Tetrahedron Letters* **19,** 3835–3838 (1978).
109. Eckstein, F. Nucleoside phosphorothioates. *Annual Review of Biochemistry* **54,** 367–402 (1985).

110. Goody, R. S. & Eckstein, F. Thiophosphate analogs of nucleoside di-and triphosphates. *Journal of the American Chemical Society* **93,** 6252–6257 (1971).
111. Schlimme, E., Lamprecht, W., Eckstein, F. & Goody, R. S. Thiophosphate- analogues and 1- N- oxides of ATP and ADP in mitochondrial translocation and phosphoryl- transfer reactions. *European Journal of Biochemistry* **40,** 485–491 (1973).
112. Eckstein, F. & Goody, R. S. Synthesis and properties of diastereoisomers of adenosine 5‘-(O-1-thiotriphosphate) and adenosine 5’-(O-2-thiotriphosphate). *Biochemistry* **15,** 1685–1691 (1976).
113. Eckstein, F. Phosphorothioate analogs of nucleotides. *Acc. Chem. Res.* **12,** 204–210 (1979).
114. Eckstein, F., Sternbach, H. & Haar, Von der, F. Stereochemistry of internucleotidic bond formation by tRNA nucleotidyltransferase from bakers' yeast. *Biochemistry* **16,** 3429–3432 (1977).
115. Burgers, P. M. & Eckstein, F. Structure of metal x nucleotide complex in the creatine kinase reaction. A study with diastereomeric phosphorothioate analogs of adenosine di- and triphosphate. *Journal of Biological Chemistry* **255,** 8229–8233 (1980).
116. Connolly, B. A., Haar, Von der, F. & Eckstein, F. Structure of the metal. nucleotide complex in the yeast phenylalanyl transfer ribonucleic acid synthetase reaction as determined with diastereomeric phosphorothioate analogs of ATP. *Journal of Biological Chemistry* **255,** 11301–11307 (1980).
117. Connolly, B. A. & Eckstein, F. Structures of the mono-and divalent metal nucleotide complexes in the myosin ATPase. *Journal of Biological Chemistry* **256,** 9450–9456 (1981).
118. Eckstein, F. & Gindl, H. Polyribonucleotides containing a phosphorothioate backbone. *European Journal of Biochemistry* **13,** 558–564 (1970).
119. Usher, D. A., Richardson, D. I. & Eckstein, F. Absolute stereochemistry of the second step of ribonuclease action. (1970).
120. Usher, D. A., Erenrich, E. S. & Eckstein, F. Geometry of the first step in the action of ribonuclease-A. *Proceedings of the National Academy of Sciences of the United States of America* **69,** 115–118 (1972).
121. Saenger, W., Suck, D. & Eckstein, F. On the mechanism of ribonuclease A. *European Journal of Biochemistry* **46,** 559–567 (1974).
122. Eckstein, F. Nucleoside phosphorothioates. *Journal of the American Chemical Society* **92,** 4718–4723 (1970).
123. Eckstein, F., Romaniuk, P. J., Heideman, W. & Storm, D. R. Stereochemistry of the mammalian adenylate cyclase reaction. *Journal of Biological Chemistry* **256,** 9118–9120 (1981).
124. Saenger, W. & Eckstein, F. Stereochemistry of a substrate for pancreatic ribonuclease. Crystal and molecular structure of the triethylammonium salt of uridine 2‘, 3’-0, 0-cyclophosphorothioate. *Journal of the American Chemical Society* **92,** 4712–4718 (1970).
125. Eckstein, F., Schulz, H. H., Rueterjans, H., Haar, W. & Maurer, W. Stereochemistry of the transesterification step of ribonuclease T1. *Biochemistry* **11,** 3507–3512 (1972).
126. Burgers, P. M. & Eckstein, F. Absolute configuration of the diastereomers of adenosine 5'-O-(1-thiotriphosphate): consequences for the stereochemistry of polymerization by DNA-dependent RNA polymerase from Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **75,** 4798–4800 (1978).
127. Yee, D., Armstrong, V. W. & Eckstein, F. Mechanistic studies on deoxyribonucleic acid dependent ribonucleic acid polymerase from Escherichia coli using phosphorothioate analog. 1. Initiation and pyrophosphate exchange reactions. *Biochemistry* **18,** 4116–4120 (1979).
128. Burgers, P. M. *et al.* Stereochemistry of hydrolysis of adenosine 3‘: 5’-cyclic phosphorothioate by the cyclic phosphodiesterase from beef heart. *Journal of Biological Chemistry* **254,** 9959–9961 (1979).
129. Sathyanarayana, B. K., Saenger, W. & Eckstein, F. Crystal and molecular structure of adenosine 5'-O-phosphorothioate Op-nitrophenyl ester (Sp diastereomer)[snake venom phosphodiesterase. *European Journal of Biochemistry* **100,** 585–591 (1979).
130. Bartlett, P. A. & Eckstein, F. Stereochemical course of polymerization catalyzed by avian myeloblastosis virus reverse transcriptase. *Journal of Biological Chemistry* **257,** 8879–8884 (1982).
131. Romaniuk, P. J. & Eckstein, F. A study of the mechanism of T4 DNA polymerase with

diastereomeric phosphorothioate analogues of deoxyadenosine triphosphate. *Journal of Biological Chemistry* **257,** 7684–7688 (1982).
132. De Clercq, E., Eckstein, F. & Merigan, T. C. Interferon induction increased through chemical modification of a synthetic polyribonucleotide. *Science* **165,** 1137–1139 (1969).
133. Eckstein, F. Diuridin-3′, 5′-thiophosphate. *Tetrahedron Letters* **8,** 3495–3499 (1967).
134. Vosberg, H. P. & Eckstein, F. Effect of deoxynucleoside phosphorothioates incorporated in DNA on cleavage by restriction enzymes. *Journal of Biological Chemistry* **257,** 6595–6599 (1982).
135. Potter, BV & Eckstein, F. Cleavage of phosphorothioate-substituted DNA by restriction endonucleases. *Journal of Biological Chemistry* **259,** 14243–14248 (1984).
136. Sayers, J. R., Schmidt, W., Wendler, A. & Eckstein, F. Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide. *Nucleic Acids Research* **16,** 803–814 (1988).
137. Spitzer, S. & Eckstein, F. Inhibition of deoxyribonucleases by phosphorothioate groups in oligodeoxyribonucleotides. *Nucleic Acids Research* **16,** 11691–11704 (1988).
138. Sayers, J. R., Olsen, D. B. & Eckstein, F. Inhibition of restriction endonuclease hydrolysis by phosphorothioate-containing DNA. *Nucleic Acids Research* **17,** 9495 (1989).
139. Olsen, D. B., Kotzorek, G., Sayers, J. R. & Eckstein, F. Inhibition of the restriction endonuclease BanII using modified DNA substrates. Determination of phosphate residues critical for the formation of an active enzyme-DNA complex. *Journal of Biological Chemistry* **265,** 14389–14394 (1990).
140. Eckstein, F. Nucleoside phosphorothioates. *Journal of the American Chemical Society* **88,** 4292–4294 (1966).
141. Hacia, J. G., Wold, B. J. & Dervan, P. B. Phosphorothioate oligonucleotide-directed triple helix formation. *Biochemistry* **33,** 5367–5369 (1994).
142. Lackey, D. B. & Patel, J. Biochemical synthesis of chirally pure Rp oligonucleotide phosphorothioates. *Biotechnology letters* **19,** 475–478 (1997).
143. Stec, W. J., Grajkowski, A., Koziolkiewicz, M. & Uznanski, B. Novel route to oligo (deoxyribonucleoside phosphorothioates). Stereocontrolled synthesis of P-chiral oligo (deoxyribonucleoside phosphorothioates). *Nucleic Acids Research* **19,** 5883–5888 (1991).
144. Stec, W. J. *et al.* Diastereomers of nucleoside 3'-O-(2-Thio-1, 3, 2-oxathia (selena) phospholanes): Building blocks for stereocontrolled synthesis of oligo(nucleoside phosphorothioate)s. *Journal of the American Chemical Society* **117,** 12019–12029 (1995).
145. Stec, W. J. *et al.* Deoxyribonucleoside 3'-O-(2-thio-and 2-oxo-"spiro"-4, 4-pentamethylene-1, 3, 2-oxathiaphospholane)s: Monomers for stereocontrolled synthesis of oligo (deoxyribonucleoside phosphorothioate) s and chimeric PS/PO oligonucleotides. *Journal of the American Chemical Society* **120,** 7156–7167 (1998).
146. Guga, P. & Stec, W. J. Synthesis of phosphorothioate oligonucleotides with stereodefined phosphorothioate linkages. *Current Protocols in Nucleic Acid Chemistry* 4.17. 1–4.17. 28 (2003).
147. Nawrot, B. *et al.* 1, 3, 2-Oxathiaphospholane approach to the synthesis of P-chiral stereodefined analogs of oligonucleotides and biologically relevant nucleoside polyphosphates. *Pure and Applied Chemistry* **80,** 1859–1871 (2008).
148. Connolly, B. A., Potter, B. V., Eckstein, F., Pingoud, A. & Grotjahn, L. Synthesis and characterization of an octanucleotide containing the EcoRI recognition sequence with a phosphorothioate group at the cleavage site. *Biochemistry* **23,** 3443–3453 (1984).
149. Stec, W. J., Zon, G. & Egan, W. Automated solid-phase synthesis, separation, and stereochemistry of phosphorothioate analogs of oligodeoxyribonucleotides. *Journal of the American Chemical Society* **106,** 6077–6079 (1984).
150. Iyer, R. P., Egan, W., Regan, J. B. & Beaucage, S. L. 3H-1, 2-Benzodithiole-3-one 1, 1-dioxide as an improved sulfurizing reagent in the solid-phase synthesis of oligodeoxyribonucleoside phosphorothioates. *Journal of the American Chemical Society* **112,** 1253–1254 (1990).
151. Iyer, R. P., Phillips, L. R., Egan, W., Regan, J. B. & Beaucage, S. L. The automated synthesis of sulfur-containing oligodeoxyribonucleotides using 3H-1, 2-benzodithiol-3-one 1, 1-dioxide as a sulfur-transfer reagent. *The Journal of organic chemistry* **55,** 4693–4699 (1990).
152. Vu, H. & Hirschbein, B. L. Internucleotide phosphite sulfurization with tetraethylthiuram

disulfide. Phosphorothioate oligonucleotide synthesis via phosphoramidite chemistry. *Tetrahedron Letters* **32,** 3005–3008 (1991).
153. Roelen, H., Kamer, P., Van den Elst, H., Van der Marel, G. A. & Van Boom, J. H. A study on the use of phenylacetyl disulfide in the solid- phase synthesis of oligodeoxynucleoside phosphorothioates. *Recueil des Travaux Chimiques des Pays-Bas* **110,** 325–331 (1991).
154. Rao, M. V., Reese, C. B. & Zhao, Z. Dibenzoyl tetrasulphide-A rapid sulphur transfer agent in the synthesis of phosphorothioate analogues of oligonucleotides. *Tetrahedron Letters* **33,** 4839–4842 (1992).
155. Zhang, Z., Nichols, A., Alsbeti, M., Tang, J. X. & Tang, J. Y. Solid phase synthesis of oligonucleotide phosphorothioate analogues using bis (ethoxythiocarbonyl) tetrasulfide as a new sulfur-transfer reagent. *Tetrahedron Letters* **39,** 2467–2470 (1998).
156. Stec, W. J. *et al.* Bis (O, O-diisopropoxy phosphinothioyl) disulfide-a highly efficient sulfurizing reagent for cost-effective synthesis of oligo (nucleoside phosphorothioate) s. *Tetrahedron Letters* **34,** 5317–5320 (1993).
157. Rao, M. V. & Macfarlane, K. Solid phase synthesis of phosphorothioate oligonucleotides using benzyltriethylammonium tetrathiomolybdate as a rapid sulfur transfer reagent. *Tetrahedron Letters* **35,** 6741–6744 (1994).
158. Efimov, V. A., Kalinkina, A. L., Chakhmakhcheva, O. G., Hill, T. S. & Jayaraman, K. New efficient sulfurizing reagents for the preparation of oligodeoxyribonucleotide phosphorothioate analogues. *Nucleic Acids Research* **23,** 4029–4033 (1995).
159. Xu, Q., Musier-Forsyth, K., Hammer, R. P. & Barany, G. Use of 1, 2, 4-dithiazolidine-3, 5-dione (DtsNH) and 3-ethoxy-1, 2, 4-dithiazoline-5-one (EDITH) for synthesis of phosphorothioate-containing oligodeoxyribonucleotides. *Nucleic Acids Research* **24,** 1602–1607 (1996).
160. Zhang, Z., Nichols, A., Tang, J. X., Han, Y. & Tang, J. Y. Solid phase synthesis of oligonucleotide phosphorothioate analogues using 3-methyl-1, 2, 4-dithiazolin-5-one (MEDITH) as a new sulfur-transfer reagent. *Tetrahedron Letters* **40,** 2095–2098 (1999).
161. Tang, J. Y., Han, Y., Tang, J. X. & Zhang, Z. Large-scale synthesis of oligonucleotide phosphorothioates using 3-amino-1, 2, 4-dithiazole-5-thione as an efficient sulfur-transfer reagent. *Organic Process Research & Development* **4,** 194–198 (2000).
162. Guzaev, A. P. Reactivity of 3 H-1, 2, 4-dithiazole-3-thiones and 3 H-1, 2-dithiole-3-thiones as sulfurizing agents for oligonucleotide synthesis. *Tetrahedron Letters* **52,** 434–437 (2011).
163. Lemaitre, M. M., Murphy, A. S. & Somers, R. L. Sulfurizing Reagent II and it's use in synthesizing oligonucleotide phosphorothiates. *Glen Research Supplementary Material* **18,** 1–6 (2006).
164. Nielsen, J., Brill, W. K.-D. & Caruthers, M. H. Synthesis and characterization of dinucleoside phosphorodithioates. *Tetrahedron Letters* **29,** 2911–2914 (1988).
165. Grandas, A., Marshall, W. S., Nielsen, J. & Caruthers, M. H. Synthesis of deoxycytidine oligomers containing phosphorodithioate linkages. *Tetrahedron Letters* **30,** 543–546 (1989).
166. Brill, W. K.-D., Nielsen, J. & Caruthers, M. H. Synthesis of dinucleoside phosphorodithioates via thioamidites. *Tetrahedron Letters* **29,** 5517–5520 (1988).
167. Farschtschi, N. & Gorenstein, D. G. Preparation of a deoxynucleoside thiophosphoramidite intermediate in the synthesis of nucleoside phosphorodithioates. *Tetrahedron Letters* **29,** 6843–6846 (1988).
168. Gorenstein, D. G. & Farschtschi, N. Process for preparing dithiophosphate oligonucleotide analogs via ... US Patent (1993).
169. Caruthers, M. H., Ma, Y. X., Yau, E., Nielsen, J. & Brill, W. K. Nucleoside thiophosphoramidites. US Patent (1993).
170. Brill, W. K., Tang, J. Y., Ma, Y. X. & Caruthers, M. H. Synthesis of oligodeoxynucleoside phosphorodithioates via thioamidites. *Journal of the American Chemical Society* **111,** 2321–2322 (1989).
171. Bjergårde, K. & Dahl, O. Solid phase synthesis of oligodeoxyribonucleoside phosphorodithioates from thiophosphoramidites. *Nucleic Acids Research* **19,** 5843–5850 (1991).
172. Wiesler, W. T., Marshall, W. S. & Caruthers, M. H. Synthesis and purifications of phosphorodithioate DNA. *Methods in Molecular Biology* **20,** 191–206 (1993).
173. Yang, X., Hodge, R. P., Luxon, B. A., Shope, R. & Gorenstein, D. G. Separation of synthetic oligonucleotide dithioates from monothiophosphate impurities by anion-exchange

chromatography on a Mono-Q column. *Analytical Biochemistry* **306,** 92–99 (2002).
174. Goodwin, H. M., Faraday, M., Hittorf, J. W. & Kohlrausch, F. W. G. The fundamental laws of electrolytic conduction: memoirs by Faraday, Hittorf and F. Kohlrausch. **7,** (1899).
175. Levine. *Physical Chemistry (Sie)*. (Tata McGraw-Hill Education, 2007).
176. DeVoe, H. *Thermodynamics and Chemistry*. (Pearson Education, Inc, 2014).
177. Neumann, C. Vorlesungen ü ber die Mechanische Theorie der Wärme. (1875).
178. Joule, J. P., Scoresby, W., Playfair, L. P. B. & Kelvin, W. T. B. *The Scientific Papers of James Prescott Joule*. **1,** (The Physical Society of London, 1884).
179. Eddington, A. S. *The Nature of the Physical World*. (Cambridge University Press, 1929).
180. Perlmutter, S. *et al.* Discovery of a supernova explosion at half the age of the Universe. *Nature* **391,** 51–54 (1998).
181. Riess, A. G. *et al.* Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal* **116,** 1009 (1998).
182. Perlmutter, S. *et al.* Measurements of Ω and Λ from 42 high-redshift supernovae. *The Astrophysical Journal* **517,** 565 (1999).
183. Nielsen, H. B. & Ninomiya, M. Unification of cosmology and the second law of thermodynamics—proposal for solving the cosmological constant and inflation problems. *Progress of Theoretical Physics* **116,** 851–871 (2006).
184. Nernst, W. *Theoretische Chemie, vom Standpunkte der Avogadroschen Regel und der Thermodynamik*. (Zeitschrift für anorganische Chemie, 1893).
185. Horn, J. R., Russell, D., Lewis, E. A. & Murphy, K. P. van't Hoff and calorimetric enthalpies from isothermal titration calorimetry: Are there significant discrepancies? *Biochemistry* **40,** 1774–1778 (2001).
186. Avogadro, A. Essai d'une manière de déterminer les masses relatives des molécules élémentaires des corps, et les proportions selon lesquelles elles entrent dans les combinaisons. *Journal de Physique* **72,** 58–76 (1811).
187. Gibbs, J. W. *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics*. (Charles Scribner's Sons, 1902).
188. Gibbs, J. W. *The scientific papers of J. Willard Gibbs*. **1,** (Longmans, Green and Company, 1906).
189. Dick, D. A. The distribution of sodium, potassium and chloride in the nucleus and cytoplasm of Bufo bufo oocytes measured by electron microprobe analysis. *The Journal of physiology* **284,** 37–53 (1978).
190. Delcourt, S. G. & Blake, R. D. Stacking energies in DNA. *Journal of Biological Chemistry* **266,** 15160–15169 (1991).
191. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H. & Zehfus, M. H. Hydrophobicity of amino acid residues in globular proteins. *Science* **229,** 834–838 (1985).
192. Rice, S. A. & Doty, P. The thermal denaturation of desoxyribose nucleic acid. *Journal of the American Chemical Society* **79,** 3937–3947 (1957).
193. Thomas, R. Recherches sur la d'enaturation des acides desoxyribonucléiques. *Biochimica et biophysica acta* **14,** 231–240 (1954).
194. Tinoco, I., Jr. Hypochromism in Polynucleotides. *Journal of the American Chemical Society* **82,** 4785–4790 (1960).
195. Schildkraut, C. & Lifson, S. Dependence of the melting temperature of DNA on salt concentration. *Biopolymers* **3,** 195–208 (1965).
196. Wartell, R. M. & Benight, A. S. Thermal denaturation of DNA molecules: a comparison of theory with experiment. *Physics Reports* **126,** 67–107 (1985).
197. Marmur, J. & Doty, P. Heterogeneity in deoxyribonucleic acids: I. Dependence on composition of the configurational stability of deoxyribonucleic acids. *Nature* **183,** 1427–1429 (1959).
198. Marmur, J. & Doty, P. Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *Journal of Molecular Biology* **5,** 109–118 (1962).
199. Owen, R. J., Hill, L. R. & Lapage, S. P. Determination of DNA base compositions from melting profiles in dilute buffers. *Biopolymers* **7,** 503–516 (1969).
200. Blake, R. D. & Hydorn, T. G. Spectral analysis for base composition of DNA undergoing melting. *Journal of Biochemical and Biophysical Methods* **11,** 307–316 (1985).
201. Blake, R. D. & Lefoley, S. G. Spectral analysis of high resolution direct-derivative melting curves of DNA for instantaneous and total base composition. *Biochimica et Biophysica Acta*

*(BBA)-Nucleic Acids and Protein Synthesis* **518,** 233–246 (1978).
202. Hirschman, S. Z. & Felsenfeld, G. Determination of DNA composition and concentration by spectral analysis. *Journal of Molecular Biology* **16,** 347–358 (1966).
203. Wada, A., Yabuki, S., Husimi, Y. & Brahms, J. G. Fine structure in the thermal denaturation of DNA: High temperature-resolution spectrophotometric studies. *Critical Reviews in Biochemistry and Molecular Biology* **9,** 87–144 (1980).
204. Breslauer, K. J., Frank, R., Blöcker, H. & Marky, L. A. Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences of the United States of America* **83,** 3746–3750 (1986).
205. Gotoh, O. & Tagashira, Y. Stabilities of nearest- neighbor doublets in double- helical DNA determined by fitting calculated melting profiles to observed profiles. *Biopolymers* **20,** 1033–1042 (1981).
206. Ornstein, R. L. & Fresco, J. R. Correlation of tm and sequence of DNA duplexes with δh computed by an improved empirical potential method. *Biopolymers* **22,** 1979–2000 (1983).
207. Doktycz, M. J., Goldstein, R. F., Paner, T. M., Gallo, F. J. & Benight, A. S. Studies of DNA dumbbells. I. Melting curves of 17 DNA dumbbells with different duplex stem sequences linked by T4 endloops: Evaluation of the nearest- neighbor stacking interactions in DNA. *Biopolymers* **32,** 849–864 (1992).
208. Sugimoto, N., Nakano, S.-I., Yoneyama, M. & Honda, K.-I. Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic Acids Research* **24,** 4501–4505 (1996).
209. SantaLucia, J., Allawi, H. T. & Seneviratne, P. A. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* **35,** 3555–3562 (1996).
210. Allawi, H. T. & SantaLucia, J. Thermodynamics and NMR of internal G⊙ T mismatches in DNA. *Biochemistry* **36,** 10581–10594 (1997).
211. Aida, M. Characteristics of the Watson- Crick type hydrogen- bonded DNA base pairs: An ab initio molecular orbital study. *Journal of Computational Chemistry* **9,** 362–368 (1988).
212. McCampbell, C. R., Wartell, R. M. & Plaskon, R. R. Inverted repeat sequences can influence the melting transitions of linear DNAs. *Biopolymers* **28,** 1745–1758 (1989).
213. SantaLucia, J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences of the United States of America* **95,** 1460–1465 (1998).
214. Sugimoto, N. *et al.* Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry* **34,** 11211–11216 (1995).
215. Manning, G. S. Limiting laws and counterion condensation in polyelectrolyte solutions I. Colligative properties. *Journal of Chemical Physics* **51,** 924–933 (1969).
216. Manning, G. S. On the application of polyelectrolyte 'limiting laws' to the helix- coil transition of DNA. I. Excess univalent cations. *Biopolymers* **11,** 937–949 (1972).
217. Imai, N. & Onishi, T. Analytical Solution of Poisson-Boltzmann Equation for Two-Dimensional Many-Center Problem. *J. Chem. Phys.* **30,** 1115–1116 (1959).
218. Record, M. T., Woodbury, C. P. & Lohman, T. M. Na+ effects on transitions of DNA and polynucleotides of variable linear charge density. *Biopolymers* **15,** 893–915 (1976).
219. Korolev, N., Lyubartsev, A. P. & Nordenskiöld, L. Application of polyelectrolyte theories for analysis of DNA melting in the presence of Na+ and Mg2+ ions. *Biophysical Journal* **75,** 3041–3056 (1998).
220. Record, M. T., Anderson, C. F. & Lohman, T. M. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Quarterly Reviews of Biophysics* **11,** 103–178 (1978).
221. Anderson, C. F., Record, M. T., Jr & Hart, P. A. Sodium-23 NMR studies of cation-DNA interactions. *Biophysical Chemistry* **7,** 301–316 (1978).
222. Bleam, M. L., Anderson, C. F. & Record, M. T., Jr. Sodium-23 nuclear magnetic resonance studies of cation-deoxyribonucleic acid interactions. *Biochemistry* **22,** 5418–5425 (1983).
223. Record, M. T., Jr, Lohman, T. M. & Haseth, P. de. Ion effects on ligand-nucleic acid interactions. *Journal of Molecular Biology* **107,** 145–158 (1976).
224. Record, M. T., Jr *et al.* Double helical DNA: conformations, physical properties, and interactions with ligands. *Annual Review of Biochemistry* **50,** 997–1024 (1981).

225. Privalov, P. L., Ptitsyn, O. B. & Birshtein, T. M. Determination of stability of the DNA double helix in an aqueous medium. *Biopolymers* **8,** 559–571 (1969).
226. Krakauer, H. & Sturtevant, J. M. Heats of the helix–coil transitions of the poly A–poly U complexes. *Biopolymers* **6,** 491–512 (1968).
227. Filimonov, V. V. & Privalov, P. L. Thermodynamics of base interaction in (A) n and (A· U) n. *Journal of Molecular Biology* **122,** 465–470 (1978).
228. Shiao, D. & Sturtevant, J. M. Heats of thermally induced helix–coil transitions of DNA in aqueous solution. *Biopolymers* **12,** 1829–1836 (1973).
229. Dove, W. F. & Davidson, N. Cation effects on the denaturation of DNA. *Journal of Molecular Biology* **5,** 467–478 (1962).
230. Gruenwedel, D. W. & Hsu, C. H. Salt effects on the denaturation of DNA. *Biopolymers* **7,** 557–570 (1969).
231. Frank Kamenetskii, M. D. Simplification of the empirical relationship between melting temperature of DNA, its GC content and concentration of sodium ions in solution. *Biopolymers* **10,** 2623–2624 (1971).
232. Fuchs, J., Fiche, J. B., Buhot, A., Calemczuk, R. & Livache, T. Salt concentration effects on equilibrium melting curves from DNA microarrays. *Biophysical Journal* **99,** 1886–1895 (2010).
233. Klump, H. & Ackermann, T. Experimental thermodynamics of the helix–random coil transition. IV. Influence of the base composition of DNA on the transition enthalpy. *Biopolymers* **10,** 513–522 (1971).
234. Hamaguchi, K. & Geiduschek, E. P. The effect of electrolytes on the stability of the deoxyribonucleate helix. *Journal of the American Chemical Society* **84,** 1329–1338 (1962).
235. Schleich, T. & Hippel, Von, P. H. Ion-induced water-proton chemical shifts and the conformational stability of macromolecules. *Biochemistry* **9,** 1059–1066 (1970).
236. Collins, K. D. Charge density-dependent strength of hydration and biological structure. *Biophysical Journal* **72,** 65–76 (1997).
237. Robinson, D. R. & Grant, M. E. The effects of aqueous salt solutions on the activity coefficients of purine and pyrimidine bases and their relation to the denaturation of deoxyribonucleic acid by salts. *Journal of Biological Chemistry* **241,** 4030–4042 (1966).
238. Owczarzy, R. Predictions of Short DNA Duplex Thermodynamics and Evaluation of Nearest-Neighbor Interactions. *Doctoral Dissertation. University of Illinois at Chicago* (1999).
239. Owczarzy, R. *et al.* Predicting sequence- dependent melting stability of short duplex DNA oligomers. *Biopolymers* **44,** 217–239 (1997).
240. Gray, D. M. 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* **42,** 783–793 (1997).
241. Gray, D. M. Derivation of nearest- neighbor properties from data on nucleic acid oligomers. II. Thermodynamic parameters of DNA· RNA hybrids and DNA duplexes. *Biopolymers* **42,** 795–810 (1997).
242. Marky, L. A. & Breslauer, K. J. Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers* **26,** 1601–1620 (1987).
243. Breslauer, K. J. Extracting thermodynamic data from equilibrium melting curves for oligonucleotide order-disorder transitions. *Methods in Molecular Biology* **26,** 347 (1994).
244. Cardullo, R. A., Agrawal, S., Flores, C., Zamecnik, P. C. & Wolf, D. E. Detection of nucleic acid hybridization by nonradiative fluorescence resonance energy transfer. *Proceedings of the National Academy of Sciences of the United States of America* **85,** 8790–8794 (1988).
245. Rachwal, P. A. & Fox, K. R. Quadruplex melting. *Methods* **43,** 291–301 (2007).
246. De Cian, A. *et al.* Fluorescence-based melting assays for studying quadruplex ligands. *Methods* **42,** 183–195 (2007).
247. Guédin, A., Lacroix, L. & Mergny, J.-L. Thermal melting studies of ligand DNA interactions. *Drug-DNA Interaction Protocols* 25–35 (2010).
248. Bonnet, G., Tyagi, S., Libchaber, A. & Kramer, F. R. Thermodynamic basis of the enhanced specificity of structured DNA probes. *Proceedings of the National Academy of Sciences of the United States of America* **96,** 6171–6176 (1999).
249. Tsourkas, A., Behlke, M. A., Rose, S. D. & Bao, G. Hybridization kinetics and thermodynamics of molecular beacons. *Nucleic Acids Research* **31,** 1319–1330 (2003).

250. Darby, R. A. *et al.* High throughput measurement of duplex, triplex and quadruplex melting curves using molecular beacons and a LightCycler. *Nucleic Acids Research* **30,** e39–e39 (2002).
251. Nangreave, J., Yan, H. & Liu, Y. Measuring thermodynamic details of DNA hybridization using fluorescence. *Biophysical Journal* **97,** 563–571 (2009).
252. You, Y., Tataurov, A. V. & Owczarzy, R. Measuring thermodynamic details of DNA hybridization using fluorescence. *Biopolymers* **95,** 472–486 (2011).
253. Savitzky, A. & Golay, M. J. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* **36,** 1627–1639 (1964).
254. Owczarzy, R. Melting temperatures of nucleic acids: Discrepancies in analysis. *Biophysical Chemistry* **117,** 207–215 (2005).
255. LaPlanche, L. A. *et al.* Phosphorothioate-modified oligodeoxyribonucleotides. III. NMR and UV spectroscoptc studies of the Rp-Rp, Sp-Sp, and Rp-Sp duplexes, [d (GGsAATTCC)], derived from diastereomeric O-ethyl phosphorothioates. *Nucleic Acids Research* **14,** 9081–9093 (1986).
256. Stein, C. A., Subasinghe, C., Shinozuka, K. & Cohen, J. S. Physicochemical properties of phosphorothioate oligodeoxynucleotides. *Nucleic Acids Research* **16,** 3209–3221 (1988).
257. Ghosh, M. K., Ghosh, K., Dahl, O. & Cohen, J. S. Evaluation of some properties of a phosphorodithioate oligodeoxyribonucleotide for antisense application. *Nucleic Acids Research* **21,** 5761–5766 (1993).
258. Cummins, L., Graff, D., Beaton, G., Marshall, W. S. & Caruthers, M. H. Biochemical and physicochemical properties of phosphorodithioate DNA. *Biochemistry* **35,** 8734–8741 (1996).
259. Kanehara, H., Wada, T., Mizuguchi, M. & Makino, K. Influence of a thiophosphate linkage on the duplex stability--does Sp configuration always lead to higher stability than Rp? *Nucleosides & Nucleotides* **15,** 1169–1178 (1996).
260. Clark, C. L., Cecil, P. K., Singh, D. & Gray, D. M. CD, absorption and thermodynamic analysis of repeating dinucleotide DNA, RNA and hybrid duplexes [d/r (AC)] 12·[d/r (GT/U)] 12 and the influence of phosphorothioate substitution. *Nucleic Acids Research* **25,** 4098–4105 (1997).
261. Jaroszewski, J. W., Clausen, V., Cohen, J. S. & Dahl, O. NMR investigations of duplex stability of phosphorothioate and phosphorodithioate DNA analogues modified in both strands. *Nucleic Acids Research* **24,** 829–834 (1996).
262. Kanaori, K. *et al.* Structure and stability of the consecutive stereoregulated chiral phosphorothioate DNA duplex. *Biochemistry* **38,** 16058–16066 (1999).
263. Boczkowska, M., Guga, P. & Stec, W. J. Stereodefined phosphorothioate analogues of DNA: relative thermodynamic stability of the model PS-DNA/DNA and PS-DNA/RNA complexes. *Biochemistry* **41,** 12483–12487 (2002).
264. Yu, D. *et al.* Stereo-enriched phosphorothioate oligodeoxynucleotides: synthesis, biophysical and biological properties. *Bioorganic & Medicinal Chemistry* **8,** 275–284 (2000).
265. Caruthers, M. H. *et al.* Chemical and biochemical studies with dithioate DNA. *Nucleosides & Nucleotides* **10,** 47–59 (1991).
266. Smith, J. S. & Nikonowicz, E. P. Phosphorothioate substitution can substantially alter RNA conformation. *Biochemistry* **39,** 5642–5652 (2000).
267. Marshall, W. S. & Caruthers, M. H. Phosphorodithioate DNA as a potential therapeutic drug. *Science* **259,** 1564–1570 (1993).
268. Owczarzy, R., You, Y., Groth, C. L. & Tataurov, A. V. Stability and mismatch discrimination of locked nucleic acid–DNA duplexes. *Biochemistry* **50,** 9352–9367 (2011).
269. Tataurov, A. V., You, Y. & Owczarzy, R. Predicting ultraviolet spectrum of single stranded and double stranded deoxyribonucleic acids. *Biophysical Chemistry* **133,** 66–70 (2008).
270. Petersheim, M. & Turner, D. H. Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp. *Biochemistry* **22,** 256–263 (1983).
271. Petruska, J. *et al.* Comparison between DNA melting thermodynamics and DNA polymerase fidelity. *Proceedings of the National Academy of Sciences of the United States of America* **85,** 6252–6256 (1988).
272. Heddi, B., Foloppe, N., Bouchemal, N., Hantz, E. & Hartmann, B. Quantification of DNA

BI/BII backbone states in solution. Implications for DNA overall structure and recognition. *Journal of the American Chemical Society* **128,** 9170–9177 (2006).
273. Svozil, D., Kalina, J., Omelka, M. & Schneider, B. DNA conformations and their sequence preferences. *Nucleic Acids Research* **36,** 3690–3706 (2008).
274. Hartmann, B., Piazzola, D. & Lavery, R. BI-BII transitions in B-DNA. *Nucleic Acids Research* **21,** 561–568 (1993).
275. Crick, F. Central dogma of molecular biology. *Nature* **227,** 561–563 (1970).
276. Riley, M. L. *et al.* PEDANT genome database: 10 years online. *Nucleic Acids Research* **35,** D354–D357 (2007).
277. Luscombe, N. M., Austin, S. E., Berman, H. M. & Thornton, J. M. An overview of the structures of protein-DNA complexes. *Genome biology* **1,** reviews001 (2000).
278. Beamer, L. J. & Pabo, C. O. Refined 1.8 Å crystal structure of the λ repressor-operator complex. *Journal of Molecular Biology* **227,** 177–196 (1992).
279. Elrod-Erickson, M., Rould, M. A., Nekludova, L. & Pabo, C. O. Zif268 protein–DNA complex refined at 1.6 Å: a model system for understanding zinc finger–DNA interactions. *Structure* **4,** 1171–1180 (1996).
280. Glover, J. N. M. & Harrison, S. C. Crystal structure of the heterodimeric bZIP transcription factor c-Fos–c-Jun bound to DNA. (1995).
281. Halford, S. E. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Research* **32,** 3040–3052 (2004).
282. Seeman, N. C., Rosenberg, J. M. & Rich, A. Sequence-specific recognition of double helical nucleic acids by proteins. *Proceedings of the National Academy of Sciences of the United States of America* **73,** 804–808 (1976).
283. Rohs, R. *et al.* Origins of specificity in protein-DNA recognition. *Annual Review of Biochemistry* **79,** 233–269 (2010).
284. McKay, D. B. & Steitz, T. A. Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. *Nature* **290,** 745 (1981).
285. Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. Structure of the cro repressor from bacteriophage λ and its interaction with DNA. *Nature* **290,** 754–758 (1981).
286. Pabo, C. O. & Lewis, M. The operator-binding domain of lambda repressor: structure and DNA recognition. *Nature* **298,** 443–447 (1982).
287. Anderson, W. F. *et al.* Crystallographic data for complexes of the Cro repressor with DNA. *Journal of Molecular Biology* **168,** 903–906 (1983).
288. Anderson, J., Ptashne, M. & Harrison, S. C. Cocrystals of the DNA-binding domain of phage 434 repressor and a synthetic phage 434 operator. *Proceedings of the National Academy of Sciences of the United States of America* **81,** 1307–1311 (1984).
289. Joachimiak, A. *et al.* Crystals of the trp repressor-operator complex suitable for X-ray diffraction analysis. *Journal of Biological Chemistry* **262,** 4917–4921 (1987).
290. Otwinowski, Z. *et al.* Crystal structure of trp represser/operator complex at atomic resolution. *Nature* **335,** 321–329 (1988).
291. Bernstein, F. C. *et al.* The Protein Data Bank: a computer-based archival file for macromolecular structures. *Journal of Molecular Biology* **112,** 535–542 (1977).
292. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Research* **28,** 235–242 (2000).
293. Pabo, C. O. & Sauer, R. T. Protein-DNA recognition. *Annual Review of Biochemistry* **53,** 293–321 (1984).
294. Pabo, C. O. & Sauer, R. T. Transcription factors: structural families and principles of DNA recognition. *Annual Review of Biochemistry* **61,** 1053–1095 (1992).
295. Suzuki, M. A framework for the DNA–protein recognition code of the probe helix in transcription factors: the chemical and stereochemical rules. *Structure* **2,** 317–326 (1994).
296. Mandel-Gutfreund, Y., Schueler, O. & Margalit, H. Comprehensive analysis of hydrogen bonds in regulatory protein DNA-complexes: in search of common principles. *Journal of Molecular Biology* **253,** 370–382 (1995).
297. Luscombe, N. M., Laskowski, R. A. & Thornton, J. M. Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level. *Nucleic Acids Research* **29,** 2860–2874 (2001).
298. Rohs, R., West, S. M., Liu, P. & Honig, B. Nuance in the double-helix and its role in protein–DNA recognition. *Current Opinion in Structural Biology* **19,** 171–177 (2009).
299. Rohs, R. *et al.* The role of DNA shape in protein–DNA recognition. *Nature* **461,** 1248–1253

(2009).
300. Hegde, R. S., Grossman, S. R., Laimins, L. A. & Sigler, P. B. Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DMA-binding domain bound to its DNA target. (1992).
301. Rozenberg, H., Rabinovich, D., Frolow, F., Hegde, R. S. & Shakked, Z. Structural code for DNA recognition revealed in crystal structures of papillomavirus E2-DNA targets. *Proceedings of the National Academy of Sciences of the United States of America* **95,** 15194–15199 (1998).
302. Hizver, J., Rozenberg, H., Frolow, F., Rabinovich, D. & Shakked, Z. DNA bending by an adenine–thymine tract and its role in gene regulation. *Proceedings of the National Academy of Sciences of the United States of America* **98,** 8490–8495 (2001).
303. Kim, S.-S., Tam, J. K., Wang, A.-F. & Hegde, R. S. The structural basis of DNA target discrimination by papillomavirus E2 proteins. *Journal of Biological Chemistry* **275,** 31245–31254 (2000).
304. Parker, S. C., Hansen, L., Abaan, H. O., Tullius, T. D. & Margulies, E. H. Local DNA topography correlates with functional noncoding regions of the human genome. *Science* **324,** 389–392 (2009).
305. Broos, S. *et al.* PhysBinder: improving the prediction of transcription factor binding sites by flexible inclusion of biophysical properties. *Nucleic Acids Research* (2013).
306. Zhou, T. *et al.* DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Research* (2013).
307. Trieb, M. *et al.* Dynamics of DNA: BI and BII phosphate backbone transitions. *Journal of Physical Chemistry* **108,** 2470–2476 (2004).
308. Tian, Y. *et al.* 31P NMR investigation of backbone dynamics in DNA binding sites. *Journal of Physical Chemistry* **113,** 2596–2603 (2008).
309. Okonogi, T. M., Alley, S. C., Harwood, E. A., Hopkins, P. B. & Robinson, B. H. Phosphate backbone neutralization increases duplex DNA flexibility: a model for protein binding. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 4156–4160 (2002).
310. Bodenhausen, G. & Ruben, D. J. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters* **69,** 185–189 (1980).
311. Volk, D. E. *et al.* NMR assignments of the sylvatic dengue 1 virus envelope protein domain III. *Biomolecular NMR assignments* **2,** 155–157 (2008).
312. Farmer, B. T. & Venters, R. A. Assignment of aliphatic side-chain 1 H N/15 N resonances in perdeuterated proteins. *Journal of Biomolecular NMR* **7,** 59–71 (1996).
313. Liepinsh, E. & Otting, G. Proton exchange rates from amino acid side chains—implications for image contrast. *Magnetic Resonance in Medicine* **35,** 30–42 (1996).
314. Iwahara, J., Jung, Y. S. & Clore, G. M. Heteronuclear NMR spectroscopy for lysine NH3 groups in proteins: Unique effect of water exchange on 15N transverse relaxation. *Journal of the American Chemical Society* **129,** 2971–2980 (2007).
315. Iwahara, J. & Clore, G. M. Direct observation of enhanced translocation of a homeodomain between DNA cognate sites by NMR exchange spectroscopy. *Journal of the American Chemical Society* **128,** 404–405 (2006).
316. Iwahara, J. & Clore, G. M. Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature* **440,** 1227–1230 (2006).
317. Iwahara, J., Zweckstetter, M. & Clore, G. M. NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proceedings of the National Academy of Sciences* **103,** 15062–15067 (2006).
318. Iwahara, J. & Clore, G. M. Sensitivity improvement for correlations involving arginine side-chain Nε/Hε resonances in multi-dimensional NMR experiments using broadband 15N 180° pulses. *Journal of Biomolecular NMR* **36,** 251–257 (2006).
319. Sahu, D., Clore, G. M. & Iwahara, J. TROSY-based Z-exchange spectroscopy: Application to the determination of the activation energy for intermolecular protein translocation between specific sites on different DNA molecules. *Journal of the American Chemical Society* **129,** 13232–13237 (2007).
320. Takayama, Y., Castañeda, C. A., Chimenti, M., García-Moreno, B. & Iwahara, J. Direct evidence for deprotonation of a lysine side chain buried in the hydrophobic core of a protein. *Journal of the American Chemical Society* **130,** 6714–6715 (2008).
321. Takayama, Y., Sahu, D. & Iwahara, J. Observing in-phase single-quantum 15N multiplets

for groups with two-dimensional heteronuclear correlation spectroscopy. *Journal of Magnetic Resonance* **194,** 313–316 (2008).
322. Esadze, A., Li, D.-W., Wang, T., Bruschweiler, R. & Iwahara, J. Dynamics of lysine side-chain amino groups in a protein studied by heteronuclear 1H− 15N NMR spectroscopy. *Journal of the American Chemical Society* **133,** 909–919 (2011).
323. Zandarashvili, L., Li, D.-W., Wang, T., Bruschweiler, R. & Iwahara, J. Signature of mobile hydrogen bonding of lysine side chains from long-range 15N–13C scalar J-couplings and computation. *Journal of the American Chemical Society* **133,** 9192–9195 (2011).
324. Anderson, K. M. *et al.* Direct observation of the ion-pair dynamics at a protein–DNA interface by NMR spectroscopy. *Journal of the American Chemical Society* **135,** 3613–3619 (2013).
325. Bax, A. D., Griffey, R. H. & Hawkins, B. L. Correlation of proton and nitrogen-15 chemical shifts by multiple quantum NMR. *Journal of Magnetic Resonance (1969)* **55,** 301–315 (1983).
326. Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E. & Kay, L. E. Cross-correlated relaxation enhanced 1H− 13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *Journal of the American Chemical Society* **125,** 10420–10428 (2003).
327. Kumar, A., Grace, R. & Madhu, P. K. Cross-correlations in NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy* **37,** 191–319 (2000).
328. Bax, A., Ikura, M., Kay, L. E., Torchia, D. A. & Tschudin, R. Comparison of different modes of two-dimensional reverse-correlation NMR for the study of proteins. *Journal of Magnetic Resonance (1969)* **86,** 304–318 (1990).
329. Shaka, A. J., Keeler, J., Frenkiel, T. & Freeman, R. An improved sequence for broadband decoupling: WALTZ-16. *Journal of Magnetic Resonance (1969)* **52,** 335–338 (1983).
330. Shaka, A. J., Keeler, J. & Freeman, R. Evaluation of a new broadband decoupling sequence: WALTZ-16. *Journal of Magnetic Resonance (1969)* **53,** 313–340 (1983).
331. Cannizzaro, L. A. *et al.* Human homeo box-containing genes located at chromosome regions 2q31----2q37 and 12q12----12q13. *American journal of human genetics* **41,** 1 (1987).
332. Gehring, W. J., Affolter, M. & Burglin, T. Homeodomain proteins. *Annual Review of Biochemistry* **63,** 487–526 (1994).
333. Kornberg, T. B. Understanding the homeodomain. *Journal of Biological Chemistry* **268,** 26813–26813 (1993).
334. Postlethwait, J. H. & Schneiderman, H. A. Pattern formation and determination in the antenna of the homoeotic mutant Antennapedia of Drosophila melanogaster. *Developmental Biology* **25,** 606–640 (1971).
335. Manak, J. R. & Scott, M. P. A class act: conservation of homeodomain protein functions. *Development* **1994,** 61–77 (1994).
336. Wilson, D. S., Sheng, G., Jun, S. & Desplan, C. Conservation and diversification in homeodomain-DNA interactions: a comparative genetic analysis. *Proceedings of the National Academy of Sciences of the United States of America* **93,** 6886–6891 (1996).
337. LaRonde-LeBlanc, N. A. & Wolberger, C. Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. *Genes & Development* **17,** 2060–2072 (2003).
338. Schlegel, H. B. Geometry optimization. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1,** 790–809 (2011).
339. Jorgenson, W. L. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *Journal of the American Chemical Society* **103,** 335 (1981).
340. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* **79,** 926 (1983).
341. Neria, E., Fischer, S. & Karplus, M. Simulation of activation free energies in molecular systems. *Journal of Chemical Physics* **105,** 1902 (1996).
342. Mark, P. & Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *Journal of Physical Chemistry* **105,** 9954–9960 (2001).
343. Case, D. A., Darden, T. A., Cheatham, T. E., III, SImmerling, C. & Kollman, P. A. AMBER 12. *University of California, San Francisco* (2012). doi:10.1021/ct200909j

344. Pérez, A. *et al.* Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophysical Journal* **92,** 3817–3829 (2007).
345. Cheatham, T. I., Miller, J. L., Fox, T., Darden, T. A. & Kollman, P. A. Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *Journal of the American Chemical Society* **117,** 4193–4194 (1995).
346. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23,** 327–341 (1977).
347. Cornell, W. D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **117,** 5179–5197 (1995).
348. Wang, J., Cieplak, P. & Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry* **21,** 1049–1074 (2000).
349. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65,** 712–725 (2006).
350. Dhaouadi, Z. *et al.* Geometrical and vibrational features of phosphate, phosphorothioate and phosphorodithioate linkages interacting with hydrated cations: A DFT study. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **73,** 805–814 (2009).
351. Frisch, M. J. *et al.* Gaussian 03, Revision C. 02. Wallingford CT: Gaussian. *Gaussian Inc.* (2004).
352. Gorenstein, D. G. Conformation and dynamics of DNA and protein-DNA complexes by 31P NMR. *Chemical Reviews* **94,** 1315–1338 (1994).
353. Volk, D. E. *et al.* Solution structure and design of dithiophosphate backbone aptamers targeting transcription factor NF-κB. *Bioorganic Chemistry* **30,** 396–419 (2002).
354. Luscombe, N. M., Laskowski, R. A. & Thornton, J. M. NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. *Nucleic Acids Research* **25,** 4940–4945 (1997).
355. Carra, J. H. & Privalov, P. L. Energetics of folding and DNA binding of the MATα2 homeodomain. *Biochemistry* **36,** 526–535 (1997).
356. Dragan, A. I. *et al.* Forces driving the binding of homeodomains to DNA. *Biochemistry* **45,** 141–151 (2006).
357. Lakowicz, J. R. *Protein Fluorescence*. 445–486 (Springer, 1999).
358. Howard, D. L. & Kjaergaard, H. G. Hydrogen bonding to divalent sulfur. *Physical Chemistry Chemical Physics* **10,** 4113–4118 (2008).
359. Read, C. M. & Jelesarov, I. Calorimetry of protein-DNA complexes and their components. *Methods in Molecular Biology* **148,** 511–534 (2001).
360. Naghibi, H., Tamura, A. & Sturtevant, J. M. Significant discrepancies between van't Hoff and calorimetric enthalpies. *Proceedings of the National Academy of Sciences of the United States of America* **92,** 5597–5599 (1995).
361. Van't Hoff, J. H. & Reicher, L. T. Die umwandlungstemperatur bei der doppelten zersetzung. *Zeitschrift für Physikalische Chemie* **3,** 482 (1889).
362. Nernst, W. Über gegenseitige Beeinflussung der Löslichkeit von Salzen. *Zeitschrift für Physikalische Chemie* **4,** 372–383 (1889).
363. Noyes, A. A. Über die gegenseitige beeinflussung der löslichkeit von dissociirten körpern. *Zeitschrift für Physikalische Chemie* **6,** 241–267 (1890).
364. van't Hoff, J. H. L'équilibre chimique dans les systèmes gazeux on dissous à l'état dilué. *Recueil des Travaux Chimiques des Pays-Bas* **4,** 424–427 (1885).
365. Pfeffer, W., Hoff, J. V., Raoult, F. M. & Arrhenius, S. The Modern Theory of Solution: Memoirs. (1899).
366. Arrhenius, S. Recherches sur la conductibilité galvanique des électrolytes. *Doctoral Dissertation, Uppsala University* (1884).
367. Sutherland, W. LXX. The electric origin of molecular attraction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **4,** 625–645 (1902).
368. Sutherland, W. XVIII. Ionization, ionic velocities, and atomic sizes. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **3,** 161–177 (1902).

369. Bjerrum, N. Der aktivitätskoeffizient der ionen. *Zeitschrift für Anorganische und Allgemeine Chemie* **109,** 275–292 (1920).
370. Bjerrum, N. Die dissoziation der starken elektrolyte. *Zeitschrift für Elektrochemie und Angewandte Physikalische Chemie* **24,** 321–328 (1918).
371. Debye, P. & Hückel, E. Zur theorie der elektrolyte. *Physikalische Zeitschrift* **24,** 185–206 (1923).
372. Debye, P. & Hückel, E. The theory of the electrolyte II: The border law for electrical conductivity. *Physikalische Zeitschrift* **24,** 305–325 (1923).
373. Onsanger, L. Zur Theorie der Electrolyte. II. *Physikalische Zeitschrift* **28,** 277–298 (1927).
374. Shedlovsky, T. The Activity Coefficients of LaCl3, CaCl2, KCl, NaCl and HCl in Dilute Aqueous Solutions. *Journal of the American Chemical Society* **72,** 3680–3682 (1950).
375. Truesdell, A. H. Activity coefficients of aqueous sodium chloride from 15 to 50 C measured with a glass electrode. *Science* **161,** 884–886 (1968).
376. Marcus, Y. & Hefter, G. Ion Pairing. *Chemical Reviews* **106,** 4585–4621 (2006).
377. Stieglitz, J. Note on the Solubility Product. *Journal of the American Chemical Society* **30,** 946–954 (1908).
378. Noyes, A. A., Boggs, C. R., Farrell, F. S. & Stewart, M. A. The effect of salts on the solubility of other salts. II. *Journal of the American Chemical Society* **33,** 1650–1663 (1911).
379. Noyes, A. A. & Bray, W. C. The effect of salts on the solubility of other salts. I. *Journal of the American Chemical Society* **33,** 1643–1649 (1911).
380. Lewis, G. N. The activity of the ions and the degree of dissociation of dtrong electrolytes. *Journal of the American Chemical Society* **34,** 1631–1644 (1912).
381. Harkins, W. D. & Winninghoff, W. J. The effect of salts upon the solubility of other salts VI. The solubility of difficultly soluble unibivalent salts. *Journal of the American Chemical Society* **33,** 1827–1836 (1911).
382. Harkins, W. D. The effect of salts upon the solubility of other salts. VII. Discussion of the solubility relations of unibivalent salts. *Journal of the American Chemical Society* **33,** 1836–1873 (1911).
383. Harkins, W. D. & Paine, H. M. The effect of salts upon the solubility of other salts. VIIIa. The solubility relations of a very soluble bi-univalent salt. *Journal of the American Chemical Society* **38,** 2709–2714 (1916).
384. Harkins, W. D. & Pearce, W. T. The effect of salts upon the solubility of other salts. VIIIb. The solubility relations of some extremely soluble salts. *Journal of the American Chemical Society* **38,** 2714–2717 (1916).
385. Bronsted, J. N. Studies on solubility. I. The solubility of salts in salt solutions. *Journal of the American Chemical Society* **42,** 761–786 (1920).
386. Bronsted, J. N. Studies on solubility. II. The solubility ratios of salts in strong homoionic solvents. *Journal of the American Chemical Society* **42,** 1448–1454 (1920).
387. Bronsted, J. N. Studies on solubility. IV. The principle of the specific interaction of ions. *Journal of the American Chemical Society* **44,** 877–898 (1922).
388. Fuoss, R. M. Ionic association. III. The equilibrium between ion pairs and free ions. *Journal of the American Chemical Society* **80,** 5059–5061 (1958).
389. Grunwald, E. Interpretation of data obtained in nonaqueous media. *Analytical Chemistry* **26,** 1696–1701 (1954).
390. Eigen, M. & Tamm, K. Schallabsorption in elektrolytlösungen als folge chemischer relaxation I. Relaxationstheorie der mehrstufigen dissoziation. *Zeitschrift für Elektrochemie und Angewandte Physikalische Chemie* **66,** 93–107 (1962).
391. Eigen, M. & Tamm, K. Schallabsorption in elektrolytlösungen als folge chemischer relaxation II. Meßergebnisse und relaxationsmechanismen für 2—2- wertige elektrolyte. *Zeitschrift für Elektrochemie und Angewandte Physikalische Chemie* **66,** 107–121 (1962).
392. Krienke, H. & Barthel, J. Association concepts in electrolyte solutions. *Journal of Molecular Liquids* **78,** 123–138 (1998).
393. Simon, J. D. & Peters, K. S. Picosecond dynamics of ion pairs: the effect of hydrogen bonding on ion-pair intermediates. *Journal of the American Chemical Society* **104,** 6542–6547 (1982).
394. Masnovi, J. M. & Kochi, J. K. Direct observation of ion-pair dynamics. *Journal of the American Chemical Society* **107,** 7880–7893 (1985).

395. Yabe, T. & Kochi, J. K. Contact ion pairs. Picosecond dynamics of solvent separation, internal return, and special salt effect. *Journal of the American Chemical Society* **114,** 4491–4500 (1992).
396. Peters, K. S. & Li, B. Picosecond dynamics of contact ion pairs and solvent-separated ion pairs in the photosolvolysis of diphenylmethyl chloride. *Journal of Physical Chemistry* **98,** 401–403 (1994).
397. Lü, J.-M., Rosokha, S. V., Lindeman, S. V., Neretin, I. S. & Kochi, J. K. "Separated" versus "contact" ion-pair structures in solution from their crystalline states: dynamic effects on dinitrobenzenide as a mixed-valence anion. *Journal of the American Chemical Society* **127,** 1797–1809 (2005).
398. Cordier, F. & Grzesiek, S. Direct Observation of Hydrogen Bonds in Proteins by Interresidue 3hJNC'Scalar Couplings. *Journal of the American Chemical Society* **121,** 1601–1602 (1999).
399. Bax, A. *et al.* [2] Measurement of homo-and heteronuclear J couplings from quantitative J correlation. *Methods in Enzymology* **239,** 79–105 (1994).
400. Hu, J. S., Grzesiek, S. & Bax, A. Two-dimensional NMR methods for determining (χ1) angles of aromatic residues in proteins from three-bond JC′ Cγ and JNCγ couplings. *Journal of the American Chemical Society* **119,** 1809 (1997).
401. Czernek, J. & Bruschweiler, R. Geometric dependence of 3h J (31P-15N) and 2h J (31P-1H) scalar couplings in protein-nucleotide complexes. *Journal of the American Chemical Society* **123,** 11079–11080 (2001).
402. Lehmann, M. S., Koetzle, T. F. & Hamilton, W. C. Precision neutron diffraction structure determination of protein and nucleic acid components. I. Crystal and molecular structure of the amino acid L-alanine. *Journal of the American Chemical Society* **94,** 2657–2660 (1972).
403. Wennmohs, F., Staemmler, V. & Schindler, M. Theoretical investigation of weak hydrogen bonds to sulfur. *Journal of Chemical Physics* **119,** 3208 (2003).
404. Iwahara, J. & Clore, G. M. Structure-independent analysis of the breadth of the positional distribution of disordered groups in macromolecules from order parameters for long, variable-length vectors using NMR paramagnetic relaxation enhancement. *Journal of the American Chemical Society* **132,** 13346–13356 (2010).
405. Akke, M., Brueschweiler, R. & Palmer, A. G., III. NMR order parameters and free energy: an analytical approach and its application to cooperative calcium (2+) binding by calbindin D9k. *Journal of the American Chemical Society* **115,** 9832–9833 (1993).
406. Li, D.-W. & Bruschweiler, R. A dictionary for protein side-chain entropies from NMR order parameters. *Journal of the American Chemical Society* **131,** 7226–7227 (2009).
407. Li, Z., Raychaudhuri, S. & Wand, A. J. Insights into the local residual entropy of proteins provided by NMR relaxation. *Protein Science* **5,** 2647–2650 (1996).
408. Brainard, J. R. & Szabo, A. Theory for nuclear magnetic relaxation of probes in anisotropic systems: application to cholesterol in phospholipid vesicles. *Biochemistry* **20,** 4618–4628 (1981).
409. Trbovic, N. *et al.* Protein side-chain dynamics and residual conformational entropy. *Journal of the American Chemical Society* **131,** 615–622 (2008).
410. Yang, D. & Kay, L. E. Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *Journal of Molecular Biology* **263,** 369–382 (1996).
411. Krishnan, M. & Smith, J. C. Response of small-scale, methyl rotors to protein−ligand association: A simulation analysis of calmodulin−peptide binding. *Journal of the American Chemical Society* **131,** 10083–10091 (2009).