RNA kissing-loop interactions supported by bifacial peptide nucleic acid

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Christopher Amedeo DeSantis

Graduate Program in Chemistry

The Ohio State University

2017

Dissertation Committee:

Professor Dennis Bong, Advisor

Professor Venkat Gopalan

Professor Jovica Badjic
Abstract

The dissertation to follow is an account of research efforts to utilize bifacial peptide nucleic acids (bPNA) triplexes to support kissing-loop complexes. A kissing-loop complex is composed of two nucleic acid stem-loops hybridized via Watson-Crick base pairing between their loop regions. For our studies, we employed two kissing-loop complexes in which a bPNA binding site, composed of ten sequential U-U mismatches, was inserted into the duplex stem of the interacting hairpins. The first system was a tRNA in which the anticodon stem was mutated with a 10-mer bPNA binding site and the loop region extended to nine nucleotides to facilitate a six-nucleotide loop-loop duplex during heterodimerization with a complementary tRNA mutant. The second system, based on the ColE1 plasmid replication control kissing-loop complex, involves two small hairpins which, after kissing complex formation, bind the protein Rop. The hairpins (RNAI-1 and RNAII-1) were mutated with a 10-mer bPNA binding site and have complementary seven-nucleotide loops.

In the first part if this work, kissing-loop and Rop complex formation was analyzed in titration binding assays and native gel electrophoresis. tRNA duplex-triplex kissing complexes show a 5-fold decrease in affinity while omission of bPNA decreases affinity by a further 3-fold. Triplex-triplex tRNA mutant binding is 9-fold weaker compared to duplex kissing-loops and also show greater instability when bPNA is omitted. Triplex-duplex mutants in the ColE1 system show similar affinity to the wild type system with Rop
present. Triplex-triplex kissing-loops bind 4-fold tighter than the wildtype system without the aid of Rop. Furthermore, Rop is able to bind to duplex-triplex and triplex-triplex kissing-loops; the first time a bPNA triplex has been installed at a protein-nucleic acid binding site without abolishing binding.

In the latter part of this work, fluorescently-labeled bPNA were used to monitor kissing-loop formation in vitro. Triplex-triplex kissing-loop formation was monitored by FRET using Cy3- and Cy5-labeled bPNA triplexes. Competitive displacement of the U10 mutant tRNA by duplex tRNA in the triplex-triplex kissing-loop was observed via a decreased FRET signal. In the ColE1 system, Rop-catalyzed competitive displacement of a wild type hairpin by a U10 mutant in a WT-U10 kissing complex was monitored via increased FRET signal. In a separate study, fluorescence turn-on probes to detect the presence of a hairpin were developed. The fluorogen thiazole orange was conjugated to bPNA by several designed linkers and a series of synthetic RNA probes with a nine-nucleotide loop were tested for maximum fluorescence turn-on. A peptide with a 2-(2-(2-aminoethoxy)ethoxy)acetic acid linker combined with a probe in which the 1, 2 and 9 positions of the loop was mutated with propyl linkers provided 2.3-fold turn-on effect.

In summary, we have shown that hairpins mutated with a bPNA triplex in the stem region are able to form kissing-loop complexes. Furthermore, proteins which interact with kissing-loops are still able to bind complexes with the triplex mutation. Using fluorescently labeled bPNA, we have proposed new technology to site selectively label and monitor RNA-RNA interactions.
To my family.
Acknowledgments

First and foremost, I would like to thank my family. They were a constant source of support and love when I needed it most throughout this experience. Their sacrifices made moving all the way to Ohio for school possible and endurable. Their visits provided a welcomed reprieve and delicious Italian treats while reminding me what was most important in life. I love you Laura, Claudia, Mom, and Dad.

Secondly, I would like to thank my friends Kendra Dewese, Laura Dzugan, Amanda Natter, Mike Davala, Johnny Wysocki, Dan Dilts, Emma Wei, and Nao Oda for reminding me to have fun. Their encouragement meant the world to me and lightened my heart on the toughest of days. Here is to many more home cooked meals together.

I owe a debt of gratitude to Dr. Dennis Bong for the years of support as my advisor. Your constant encouragement and problem-solving discussions helped me when the projects stalled and progress looked bleak. By working in your group, I have become a better scientist and for this I thank you.

I would also like to thank Jie Mao. My constant companion throughout grad school who was always ready with a friendly smile and encouraging words. My teachers, Xijun Piao, Xin Xia, and Zhun Zhou for putting up with my incessant questions and teaching me a myriad of lab skills. I would also like to thank the other lab group members for their helpful discussions. Finally, I would like to thank Dr. Gopalan and Dr. Badjic.
Vita

2007 ................................................................. Hamilton High School East

2011 ................................................................. B.A. Chemistry, Rutgers, the State University of New Jersey

2011 to present .................................................. Graduate Research Associate, The Ohio State University

Publications


Fields of Study

Major Field: Chemistry
Table of Contents

Abstract ................................................................................................................................. ii
Acknowledgments ................................................................................................................. v
Vita......................................................................................................................................... vi
Publications .......................................................................................................................... vi
Fields of Study ..................................................................................................................... vi
Table of Contents ................................................................................................................ vii
List of Tables ......................................................................................................................... xii
List of Figures ...................................................................................................................... xiii
Chapter 1: Structure of nucleic acids and their mimics ......................................................... 1
  1.1.1 SECONDARY STRUCTURES OF NUCLEIC ACIDS: bulges and loops. ........... 6
  1.1.2 SECONDARY STRUCTURES OF NUCLEIC ACIDS: junctions...................... 8
  1.1.3 SECONDARY STRUCTURES OF NUCLEIC ACIDS: G-quadruplex ............... 9
  1.2 COMPONENTS OF TERTIARY STRUCTURE OF NUCLEIC ACIDS ............. 9
     1.2.1 TERTIARY STRUCTURE OF NUCLEIC ACIDS: coaxial stacking......... 10
1.2.2 COMPONENTS OF TERTIARY STRUCTURE OF NUCLEIC ACIDS:
kissing-loops and pseudoknots ................................................................. 11

1.2.3 COMPONENTS OF TERTIARY STRUCTURE OF NUCLEIC ACIDS:  base
triples. ............................................................................................................. 12

1.2.4 COMPONENTS OF TERTIARY STRUCTURE OF NUCLEIC ACIDS:
tetraloops ........................................................................................................ 13

1.2.5 COMPONENTS OF TERTIARY STRUCTURE OF NUCLEIC ACIDS:
triplexes. .......................................................................................................... 14

1.3 MODIFICATIONS TO NUCLEIC ACIDS.................................................. 16

1.3.1 MODIFICATIONS TO NUCLEIC ACIDS:  natural modifications......... 17

1.3.2 MODIFICATIONS TO NUCLEIC ACIDS:  metal-coordinating bps. .... 18

1.3.3 MODIFICATIONS TO NUCLEIC ACIDS:  hydrogen bonding attenuation. 19

1.3.4 MODIFICATIONS TO NUCLEIC ACIDS:  triplex promoting modifications.
.................................................................................................................. 20

1.3.5 MODIFICATIONS TO NUCLEIC ACIDS:  triplex formation via Janus
wedge.................................................................................................................. 21

1.4 MODIFICATIONS TO NUCLEIC ACID BACKBONES.............................. 23

1.4.1 MODIFICATIONS TO NUCLEIC ACID BACKBONES:  phosphodiester
modifications. .................................................................................................. 23
1.4.2 MODIFICATIONS TO NUCLEIC ACID BACKBONES: ribose modifications ............................................................................................................. 25

1.4.3 MODIFICATIONS TO NUCLEIC ACID BACKBONES: addressing backbone charge repulsion. ............................................................................................................. 26

1.4.4 MODIFICATIONS TO NUCLEIC ACID BACKBONES: ribose substitutes. .................................................................................................................. 27

1.5 PEPTIDE NUCLEIC ACIDS .................................................................................................................................................................................. 29

1.5.1 PEPTIDE NUCLEIC ACIDS: incorporation of Janus wedge motifs. ........ 32

1.5.2 PEPTIDE NUCLEIC ACIDS: native peptide backbone.................................................. 33

1.5.3 BIFACIAL PEPTIDE NUCLEIC ACIDS (bPNA) .................................................. 35

REFERENCES FOR CHAPTER 1 .................................................................................................................................................................................. 39

Chapter 2: bPNA promoted kissing-loops .......................................................................................................................... 71

2.1 INTRODUCTION TO KISSING-LOOPS: .............................................................................................................................. 71

2.1.1 HIV-1 DIMERIZATION INITIATION SITE: ........................................................................................................................ 78

2.1.2 ColE1 KISSING COMPLEX: ........................................................................................................................................................ 82

2.2 HIV-DIS: EXPERIMENTAL DESIGNS AND RESULTS ....................................................................................................................... 87

2.3 ColE1 KISSING COMPLEX: EXPERIMENTAL DESIGNS AND RESULTS .. 96

2.4 DISCUSSION AND CONCLUSIONS: ............................................................................................................................................. 105

2.5 SUPPORTING INFORMATION CHAPTER 2 ............................................................................................................................. 110
List of Tables

Table 1: Heterodimers in initial HIV-DIS system. .............................................................. 91
Table 2: RNA and DNA templates for kissing-loop forming nucleic acids. ............ 111
Table 3: DNA templates for HIV-DIS mutants ................................................................. 112
Table 4: DNA templates for tri-stem-loop ColE1 mutants ............................................. 113
Table 5: Turn-on factor of hybridized TO-bPNA conjugates and 9-membered loop probes .......................................................... 179
Table 6: Turn-on factor of hybridized TO-bPNA conjugates and 7-membered loop probes ........................................................................................... 180
List of Figures

Figure 1: Canonical nucleosides and bps. .......................................................... 2
Figure 2: Duplex structure of nucleic acids. ...................................................... 3
Figure 3: Alternative base pairings. ................................................................. 5
Figure 4: Hairpin and bulge structure. ............................................................. 7
Figure 5: Structure three and four-way junctions. ........................................... 8
Figure 6: G-quadruplex structure. ................................................................. 9
Figure 7: General structure of a kissing-loop and pseudoknot .......................... 12
Figure 8: Base triples. ................................................................................. 13
Figure 9: Structures of tetraloops and receptors. ......................................... 14
Figure 10: Triplex-forming nucleic acids. ..................................................... 16
Figure 11: Examples of metal-mediated bps and triples.................................. 18
Figure 12: Examples of triplex promoting nucleic acids. ............................. 21
Figure 13: Janus wedge triplex forming motifs. ............................................ 22
Figure 14: Examples of phosphoramidates and phosphorothioate. ............. 24
Figure 15: Examples of locked and unlocked nucleic acids. ......................... 26
Figure 16: Example of nucleobase modifications which introduce charge. ... 27
Figure 17: Examples of ribose substitutes. ..................................................... 28
Figure 18: Structure of PNA based on protein backbone. ............................ 30
Figure 19: Mechanism of PNA strand invasion of duplex nucleic acids................. 31
Figure 20: Examples of Janus wedge motifs used in PNA................................. 33
Figure 21: Native peptide backbone PNA with triazine ring recognition moiety. .... 34
Figure 22: Structure of bPNA........................................................................ 36
Figure 23: bPNA as an allosteric switch.......................................................... 37
Figure 24: Structure of hairpins and kissing-loops............................................. 72
Figure 25: Intramolecular loop-loops interactions of Neurospora Varkud satellite
ribozyme and tRNA.................................................................................. 73
Figure 26: Cis-encoded antisense RNAs and their targets................................. 76
Figure 27: DIS in the 5' leader of the HIV genome............................................. 78
Figure 28: Crystal and NMR structure of DIS dimer.......................................... 81
Figure 29: Plasmid copy number control mechanism of the ColE1 plasmid........... 82
Figure 30: Structure of Rop as dimer in solution................................................. 84
Figure 31: Solution structure ColE1 kissing-loop.............................................. 85
Figure 32: Proposed experiment with mutated HIV-DIS hairpins........................ 87
Figure 33: Initial design of HIV-DIS mutants.................................................... 89
Figure 34: Dimerization of initial HIV-DIS system.............................................. 90
Figure 35: Structures kissing-loop competent tRNA mutants............................. 92
Figure 36: Heterodimerization of tRNA mutants................................................. 94
Figure 37: Dissociation constants of tRNA mutants............................................ 95
Figure 38: Initial designs of ColE1 kissing complex mutants............................. 96
Figure 39: Heterodimerization of ColE1 mutants.............................................. 98
Figure 40: Design of ColE1 stem loop 1 mutants.......................................................... 99
Figure 41: Dimerization of stem-loop 1 mutants of ColE1................................. 100
Figure 42: Complexation of RNAI-1U10 and RNAII-1U10 with Rop....................... 100
Figure 43: Dissociation constant of RNAII-1U10 and RNAI-1WT........................ 102
Figure 44: Dissociation constants of ColE1 mutants............................................. 103
Figure 45: Proton NMR of compound 1................................................................. 119
Figure 46: Proton NMR of Cy3.............................................................................. 120
Figure 47: Proton NMR of Cy5.............................................................................. 122
Figure 48: HPLC trace and MALDI-TOF of Cy5-(EM)10................................. 124
Figure 49: HPLC trace and MALDI-TOF of Cy3-(EM)10................................. 125
Figure 50: Negative control for non-specific dimerization of HD1U10 and HD1Nat. 126
Figure 51: Negative control for non-specific dimerization of HD2U10 and HD2Nat. 126
Figure 52: Mg$^{2+}$ dependence of HD1U10-bPNA triplex and HD2Nat complexation... 127
Figure 53: Mg$^{2+}$ dependence of HD1U10-bPNA triplex and HD2U10-bPNA triplex complexation........................................................................................................... 127
Figure 54: K_d determination of HD1Nat and HD2U10-bPNA triplex.............. 128
Figure 55: K_d determination of HD1U10-EM10 and HD2U10-EM10................. 129
Figure 56: K_d determination of HD1Nat and HD2Nat........................................ 130
Figure 57: K_d determination of HD1Nat and unstructured HD2U10............. 131
Figure 58: K_d determination unstructured HD1U10 and unstructured HD2U10. 132
Figure 59: K_d determination unstructured HD1U10 probe and unstructured HD2U10 probe .................................................................................................................. 133
Figure 60: Control for non-specific dimerization of RNAI ................................................................. 134
Figure 61: Control for non-specific dimerization of RNAII ................................................................. 134
Figure 62: Control for Rop binding to bPNA triplex ................................................................. 134
Figure 63: Rop complex of RNAII-1U10 and RNAI-1WT ................................................................. 135
Figure 64: Kissing complex of RNAI-1U10 and RNAII-1U10 without Rop ............................................ 136
Figure 65: Kissing complex of RNAI-1U10 and RNAII-1U10 with Rop ............................................. 137
Figure 66: Complex formation of unstructured RNAII-1U10 and RNAI-WT with Rop ... ................................. 138
Figure 67: Complex formation of unstructured RNAII-1U10 and RNAI-WT without Rop ................................................................. 139
Figure 68: Rop complex formation with preformed U10-WT kissing-loop .................................. 140
Figure 69: Rop complex formation with preformed U10-U10 kissing-loop .................................. 141
Figure 70: Fluorescent kissing-loop probes .................................................................................. 167
Figure 71: Experimental design of U10-U10 kissing-loop FRET experiment ...................................... 168
Figure 72: FRET experiments with U10 mutants ......................................................................... 169
Figure 73: tRNA competition assay ................................................................................................. 171
Figure 74: ColE1 competition assay ................................................................................................. 172
Figure 75: Rop dependence on displacement in ColE1 FRET competition assay ......................... 174
Figure 76: Kissing-loop turn on probe design .................................................................................. 176
Figure 77: TO-bPNA kissing-loop detection ..................................................................................... 178
Figure 78: Fluorescent controls for FRET with Cy3-bPNA and unlabeled bPNA in tRNA kissing-loop ................................................................. 186
Figure 79: Effect of kissing loop complex formation on Cy3 fluorescence in ColE1 system. .......................................................... 186
Figure 80: EMSA gel of kissing complex formation with probe 4....................... 187
Figure 81: EMSA gel of kissing complex formation with probe 3....................... 187
Figure 82: EMSA gel of kissing complex formation with probe 2....................... 187
Figure 83: EMSA gel of kissing complex formation with probe 1....................... 188
Figure 84: EMSA gel of kissing complex formation with probe 6 and 7.............. 188
Figure 85: Proton NMR of thiazole orange acetate. ......................................... 189
Figure 86: Attempted TO-peptides that could not be purified............................. 191
Figure 87: HPLC trace and MALDI-TOF of TO-(PEG)2(ME)10G-CNH2.............. 192
Figure 88: HPLC trace and MALDI-TOF of TO-(PEG)(ME)10G-CNH2............... 193
Figure 89: HPLC trace and MALDI-TOF of TO-(bAla)2(EM)10G-CNH2............. 194
Figure 90: HPLC trace and MALDI-TOF of TO-(bAla)(EM)10G-CNH2............... 195
Figure 91: HPLC trace and MALDI-TOF of TO-dP-G-(EM)10G-CNH2............... 196
Figure 92: HPLC trace and MALDI-TOF of TO-G-dP-G-(EM)10G-CNH2............ 197
Figure 93: HPLC trace and MALDI-TOF of TO-G-(ME)10G-CNH2................... 198
Figure 94: HPLC trace and MALDI-TOF of TO-(ME)10G-CNH2...................... 199
Figure 95: HPLC trace and MALDI-TOF of TO-(bAla)3-(EM)10G-CNH2............ 200
Chapter 1: Structure of nucleic acids and their mimics

Through evolution, nature has developed a code describing all the components of a cell necessary for the propagation of life. The code is encompassed in a polymer, named nucleic acid or polynucleotide, composed of monomers termed nucleotides which in turn are made up of a heterocycle and a sugar. The heterocycles, called nucleobases, typically found in the polymer are adenine (A), guanine (G), cytosine (C), Thymine (T), and uracil (U), though these aromatic compounds do not represent the only ring structures present in the code and are often heavily modified.\(^1\) The C1 position of the ribose ring is covalently bound to purines via N9 and pyrimidines via N1. The sugars are covalently bound to each other via phosphodiester bonds through the 5’-hydroxy and 3’-hydroxy of ribose. The backbone of the polymer is therefore alternating phosphate and ribose moieties. Nucleotide polymers with ribose are called ribonucleic acids (RNA) and nucleotide polymers with 2’-deoxyribose are called deoxyribonucleic acids (DNA) (Figure 1).

The four nucleobases create the code which is read by other nucleobases in complementary polynucleotides through hydrogen bonding which anneal, or hybridize, together into duplexes. Polynucleotides hybridize in solution through an enthalpically favorable process driven by non-covalent interactions between the nucleobases.\(^2-4\) Typical stacking interactions between two nucleobases contributes about 1 kcal/mol to the free energy of hybridization.\(^5\) Through complementary hydrogen bonds and favorable base
stacking, complementary strands of polynucleotides can form helical duplexes.\(^6\)

Hybridization relies on the specific hydrogen bonding of the nucleobases with one another and are thought to contribute a. \(-0.25\) to \(-1.6\) kcal/mol to the free energy of the system.\(^7\)

Forming base paris (bp) like rungs in a ladder, A recognizes U/T and G recognizes C via Watson-Crick (W-C)-type hydrogen bonding and stack on one another, though, these are

![Canonical nucleosides and bps](image)

**Figure 1: Canonical nucleosides and bps.**
A) Nucleosides which are the main monomers in nucleic acid polymers. B) Examples of Watson-Crick hydrogen bonding. Hydrogen bonding shown as dotted lines.

not the only bps possible (Figure 1). Base pairing forces the ribose-phosphate backbone into a helical structure (Figure 2). The type of helix formed by hybridized polynucleotides is dependent upon the type (RNA or DNA) and the components of the solution in which they are dissolved. DNA-DNA duplexes are most commonly in the B-form helix with 10-bps per turn and 3.32 Å rise per bp. Another possible structure is the A-form helix, common for RNA-RNA and DNA-RNA duplexes, with 11-bps per turn and 2.3 Å rise per
Lastly, duplexes can also adopt a rarer conformation designated Z form with 3.8 Å rise per bp and 12-bases per turn. Z-form DNA is thought to arise from a dehydrated state usually associated with high concentrations of positively charged counterions and certain types of sequences (Figure 2). The type of helix is largely predicated on the conformation of the sugar. The ribose ring in the B-form structure has a C2′ endo pucker whereas, in A form, the ribose has a C3′ endo pucker. According to crystallographic data, RNA duplexes prefer A form to accommodate the hydroxyl group at C2′ and there is a larger energy barrier to convert double stranded RNA to a B form helix due to steric interactions brought about by the C2′ hydroxyl group.

**Figure 2: Duplex structure of nucleic acids.**
This figure shows the various forms of DNA and RNA duplexes. Image from reference.
Watson-Crick type hydrogen bonding was the first type of base pairing observed in early DNA and RNA studies. However, nucleobases are not limited to these types of interactions. The nucleobases display more than one face capable of forming hydrogen bonds (Figure 3). One such non-W-C interaction is Hoogsteen hydrogen bonding which has been identified in nucleic acids bound to proteins, small molecules, RNA, DNA, and damaged DNA. Recently, Hoogsteen hydrogen bonding has been found to transiently occur in canonical DNA duplexes and even utilized by DNA polymerase-tau during DNA polymerization. Reverse Hoogsteen hydrogen bonding, which involves a 180° flip of one base, has also been observed. Other examples of non-W-C base pairing include wobble bps which involve G-U, inosine-uracil (I-U), inosine-adenine (I-A), and inosine-cytosine (I-C) as alternative nucleobase recognition events (Figure 3). Wobble type hydrogen bonding is widely used throughout nucleic acid interactions and have proven to be critical for cellular processes such as translation, where the third bp of the codon-anticodon interaction utilizes wobble bps to faithfully translate mRNA into the correct protein sequence. Besides the above examples of non-W-C interactions, there are thousands of non-canonical hydrogen bonding motifs found in the literature and efforts are underway to predict their prevalence in structurally ambiguous oligonucleotides.

Upon hybridization, two grooves, a major and minor groove, are formed (Figure 2). In B-form helices, the major groove is wide and deep, while the minor groove is narrow and deep. In contrast, the A-form helix has a narrow and deep major groove and a broad,
Figure 3: Alternative base pairings.
The above figure shows a few examples of non-Watson-Crick type bps. Image from reference.\textsuperscript{32}
shallow minor groove. The unusual structure of Z-form helices reveals a deep minor groove and a flattened out major groove. The major and minor grooves are solvent accessible and are binding sites for large macromolecules and small molecules. Typically, larger molecules such as proteins bind to larger, shallower grooves and small molecules, such as polyamides developed by the Dervan group, bind to the narrower, deeper groove where more intimate contacts can be made between small molecules and sides/floor of the groove. The walls of the grooves are lined by the backbone phosphate and sugar moieties. The floor of the grooves is made up of the edge of the bps not engaged in Watson-Crick type hydrogen bonding and can provide sequence specific information. In B-form helices, the floor of the major groove is lined by the Hoogsteen face of the nucleobases which can engage in alternative base pairing interactions. A discussion of major/minor groove binding molecules and Hoogsteen-type interactions will follow later in this chapter.

1.1.1 SECONDARY STRUCTURES OF NUCLEIC ACIDS: bulges and loops.

While the helix is the most prominent structure for RNA and DNA, these biopolymers display an array of diverse and complicated secondary and tertiary structures. Furthermore, duplexes need not be perfect bp matches. Small numbers of mismatches within duplexes create bulges and, internal loops form via a large number of unpaired nucleobases (Figure 4). The mismatches may be symmetric on either side of the duplex or asymmetric with one strand of the duplex having more unpaired bases. Up to this point, only intermolecular helical duplexes made up two discreet polynucleotides have
been discussed, but duplexes may also be formed intramolecularly. For an intramolecular duplex to form, a polynucleotide must be self-complementary. The single strand will bend back on itself, creating a loop with unpaired nucleobases, and then hybridize with itself to create a duplex stem (Figure 4). This type of structure, called a hairpin or stem-loop, is the major secondary structure studied in this work and will be discussed in great detail in later chapters. The nucleobases in the loop region of a hairpin are unpaired and are therefore able to bp, and they commonly engage in non-Watson-Crick like interactions. Hairpin motifs are common recognition sites for numerous polynucleotides, proteins, and small molecules and are often important for the stabilization of RNA and DNA tertiary structure.

**Figure 4: Hairpin and bulge structure.**
Hairpins structures are composed of an intramolecular duplex region bridged by single stranded RNA. Bulges are composed of unpaired nucleobases within a double helix.
1.1.2 SECONDARY STRUCTURES OF NUCLEIC ACIDS: junctions

Other examples of RNA secondary structure are junctions. As polynucleotides begin to fold, it is not uncommon for them to contain various helical regions that intersect. The intersection of helices separate by a region of zero or more single stranded sequences are called junctions. Common examples of junctions include three and four-way junctions found in hammerhead ribozyme and tRNA. An interesting feature of junctions is the propensity of helices to coaxially stack upon one another which has long range effects on the overall structure and stability of the nucleic acid.

![3-way junction and 4-way junction](image)

**Figure 5:** Structure three and four-way junctions.
1.1.3 SECONDARY STRUCTURES OF NUCLEIC ACIDS: G-quadruplex

Another motif found in nucleic acids whose biological relevance is only just beginning to be understood is the G-quadruplex.\textsuperscript{40} Formed by one or more nucleic acid strands, G-quadruplexes form when four guanine bases form Hoogsteen hydrogen bonds (Figure 6). Typically stabilized by monovalent cations such as potassium\textsuperscript{41}, G-quadruplexes form stacks on one another creating flat, large, and hydrophobic surfaces which have been used advantageously for drug design\textsuperscript{42-43} and fluorescence imaging.\textsuperscript{44-46} Studies into the biological relevance are underway.\textsuperscript{47}

![G-quadruplex structure.](image)

**Figure 6: G-quadruplex structure.**
Squares are guanine and large spheres are potassium ions. Image modified from reference.\textsuperscript{43}

1.2 COMPONENTS OF TERTIARY STRUCTURE OF NUCLEIC ACIDS

To this point, the discussion has focused on primary and secondary structure of nucleic acids. However, nucleic acids in cells are often large and their tertiary structure is
complicated and intricate. Polynucleotides fold into compact and fascinating structures that all start with basic secondary structures discussed above. The three-dimensional structure of RNA and DNA in the aqueous milieu of the cell often involves non-canonical and unexpected contacts made within the molecules that help stabilize the overall folding of the molecule. Interactions that stabilize tertiary interactions include coaxial stacking, the adenosine platform, 2’-hydroxy-mediated helical interactions, base triples and triplexes, the tetraloop motif, metal-core motif, ribose zipper, kissing-loop, pseudoknot, kink turns, hook turns, and non-canonical hydrogen bonding and are covered extensively in reviews.48-49 A discussion of a few of these interactions follows below.

1.2.1 TERTIARY STRUCTURE OF NUCLEIC ACIDS: coaxial stacking.

Just as the major driving force for duplex formation is aromatic stacking, the major driving force for folding of nucleic acids into their most stable tertiary structure is base stacking, which is maximized through coaxial stacking of helices after duplex formation.50 As helices terminate near other helices or at junctions, the terminal bps are unstacked leaving the base of the helix exposed to solvent. When other helices terminate near each other, helices will stack the terminal bps and become coaxially stacked in order to maximize stacking and stabilize their structure.51-53 The stacking interactions can afford stability gains of 0.5 to 3.0 kcal/mole.51 These collinear helices are best exemplified in the tRNA crystal structures54, where the acceptor stem and T-arm stack to form a coaxial stack while the D-arm and anticodon arm form a separate coaxial stack. The coaxial stacking of
helices at junctions is determined not only by the sequence of the helix, but also by the topology of the junction, the most common being two, three, and four-way junctions.\textsuperscript{36, 50}

Generally, the ionic conditions\textsuperscript{55-56} of the solution and type/number of intervening single nucleotides control\textsuperscript{57-58} how the helices interact and if they will coaxially stack. Coaxial stacking is further stabilized by long-range interactions such as kissing-loops and pseudoknots.

### 1.2.2 COMPONENTS OF TERTIARY STRUCTURE OF NUCLEIC ACIDS:

**kissing-loops and pseudoknots**

A very common long-range interaction is the kissing-loop. A kissing-loop structure occurs when the two loops of two hairpins hybridize via Watson-Crick base pairing to from a duplex (Figure 7). Kissing-loops will be discussed in greater detail in Chapter 2 of this thesis. Like the kissing-loop, pseudoknots also involve the hybridization of a hairpin, except, in the case of pseudoknots, the complementary sequence is on a linear, single-stranded nucleic acid of the same RNA. Again, coaxial stacking can be invoked to describe the stability of the structure\textsuperscript{59} however, there is a dependence on Mg\textsuperscript{2+} or high concentrations of sodium.\textsuperscript{60} Pseudoknots provide structuring elements for catalytic RNAs.\textsuperscript{61} In an exquisite example of a pseudoknot structuring a small piece of RNA for self-splicing, two pseudoknots involved in the structuring of the hepatitis delta virus ribozyme fold on top of one another to form the catalytic pocket of the ribozyme.\textsuperscript{62}
1.2.3 COMPONENTS OF TERTIARY STRUCTURE OF NUCLEIC ACIDS: base triples.

As discussed above, nucleobases have more than one face of the molecule which can hydrogen bond and these faces are exposed in the major and minor groove of the helices. Nucleic acids take advantage of these information-rich faces via triplex formation in which an unpaired base will hydrogen bond to a bp (Figure 8). Base triples can occur in both the major and minor groove and are prevalent throughout many structures.\textsuperscript{49, 63-64} For example, tRNA\textsuperscript{Phe} displays base triples in both the major and minor groove.\textsuperscript{65} The D-arm in tRNA\textsuperscript{Phe} forms two base triples via Hoogsteen face binding with bases A9, A23, and U12 and also 46-22-13.\textsuperscript{65-66} Through the minor groove, a base triple in tRNA\textsuperscript{Phe} is observed involving nucleobases 10-25-45.\textsuperscript{66}
Figure 8: Base triples.
A) A-U-A  B) C-G-C\textsuperscript{+}  C) C-G-\textsuperscript{m7}G  D) A-U-A (reverse Hoogsteen hydrogen bonding).
Image from reference.\textsuperscript{49}

1.2.4 COMPONENTS OF TERTIARY STRUCTURE OF NUCLEIC ACIDS:
   tetraloops

Tetraloop motifs are a widely-used interaction in nucleic acids used to stabilize tertiary structure and are composed of a hairpin motif with a loop of four nucleotides, typically UNCG or GNRA where N is any nucleotide and R is A or G. Tetraloops have been shown to have unusual intramolecular interactions and stabilities.\textsuperscript{67-69} While these hairpins themselves are unusually stable, they still interact with other motifs called tetraloop receptors.\textsuperscript{70-71} They involve Watson-Crick type interactions and non-canonical interactions including stacking of the nucleobases in the loop (Figure 9).\textsuperscript{72} In another
example, the tetraloop binds to the minor groove of duplexes through intramolecular interactions.\textsuperscript{73}

**Figure 9:** Structures of tetraloops and receptors.
A) Structure of a tetraloop. B-D) Tetraloop interacting with various helixes. Image taken from reference.\textsuperscript{74}

1.2.5 COMPONENTS OF TERTIARY STRUCTURE OF NUCLEIC ACIDS: triplexes.

Base triples play an important role in stabilizing RNA and DNA tertiary structures through hydrogen-bonding and stacking interactions. Despite the possibility of base
triples, the biological occurrence of triple-stranded nucleic acids is thought to be much less prevalent and transient. Triple-stranded nucleic acids occur when two Watson-Crick hybridized nucleic acid interacts with third strand in either the minor or major groove of the duplex (Figure 10). Rich first discovered nucleic acid triplexes in which a poly-U and poly-A duplex interacted with another poly-U to form the triplex in the presence of divalent cations.\textsuperscript{75} Since then, much research has been performed to discover other triplexes which occur under physiological and non-physiological conditions\textsuperscript{76-81} such as the acidic conditions required for the C\textsuperscript{+}-G*C triplex.\textsuperscript{82} The triplex hydrogen bonding interaction occurs via Hoogsteen and reverse-Hoogsteen interactions and can either be parallel or anti-parallel depending on the orientation of third strand to the strand to which it is hydrogen bonded.\textsuperscript{83} Work by Breslauer and Dervan has been completed to study the energetics of intermolecular triplex formation from three individual DNA strands, finding that enthalpy of formation to be about 2 kcal/mole per bp or about one third that of a base pairing in a helix.\textsuperscript{84} Marky and coworkers studying intramolecular triplex formation found that formation of TAT/TAT, TAT/CGC\textsuperscript{+}, and CGC\textsuperscript{+}/CGC\textsuperscript{+} stacks had enthalpies of -24 kcal/mol, -23 kcal/mol, and -22 kcal/mol, respectively.\textsuperscript{85} Furthermore, intramolecular triplexes were found to be more robust due to the decrease in the entropic cost of triplex formation and stacking contributions from the loop region.\textsuperscript{86} The biological role of triplexes is just starting to be investigated but are thought to be important in long, non-coding RNAs, ribozymes, telomerases and other cellular functions.\textsuperscript{87-89} Work has also been performed to use triplexes to detect double stranded nucleic acids via directed redox chemistry and fluorescence.\textsuperscript{90-93}
Figure 10: Triplex-forming nucleic acids.
The green colored third strand binds to the duplex DNA or RNA in either a parallel or anti-parallel orientation. Inset shows the NMR structure for a nucleic acid triplex (PDB-id:1BWG). Image taken from reference.89

1.3 MODIFICATIONS TO NUCLEIC ACIDS

Through the course of evolution, nature has utilized different selection pressures and sampling of chemical space to design the nucleic acid polymer that utilizes alternating phosphate and ribose moieties as the backbone. The nucleic acid’s polymeric backbone is a versatile scaffold used to display the nucleobases and the backbone has been tailored to its purpose. The deoxyribose-phosphate backbone for DNA, which catalogs all genetic information, is so stable that intact DNA is extracted from long deceased organisms. Conversely, RNA’s backbone, composed of ribose and phosphate, is unstable and easily
degraded which suits its role perfectly as RNA levels are regulated by stress and environment, so chemical stability and persistence would greatly inhibit cell viability.

The uniformity of DNA and RNA has led chemists to investigate alternative backbones and bases in order to study how variants in nucleic acid structure would affect function within a living organism and to probe nucleic acid design. Modifications to the nucleobases and backbone have been performed to both strengthen and weaken hybridization, promote triplex formation, expand the genetic library, and provide alternate modes of interaction. Various modifications such as methylation of the canonical nucleobases, proposing a completely redesigned backbone, to introduction of new nucleobases have been attempted with varying degrees of success.

1.3.1 MODIFICATIONS TO NUCLEIC ACIDS: natural modifications

Nature has proven very adept at modifying nucleobases with over 100 reported modifications. These modifications are performed on the canonical nucleobases, meaning each new base has the same heterocycles at its core. Modifications can be as simple as methylation or as complicated as the addition of amino acids to form amino acid-nucleobase conjugates such as lysidine. Modifications to native nucleobases increases specificity, improves stability to enzymatic digestion, promotes signal transport to different cell compartments, and regulates splicing.
1.3.2 MODIFICATIONS TO NUCLEIC ACIDS: metal-coordinating bps.

Evolutionary pressure has developed specialized machinery to modify and recognize modified nucleobases in order to expand the genetic code to fit many new purposes. Similarly, chemists have begun synthesizing new nucleobases to improve upon nucleic acid binding, expand the genetic code, create an alternate genetic code, study the principles behind nature’s design, and create new, interesting interactions. Nature settled on two purines and two pyrimidines to construct the recognition elements of its genetic code, presumably through a selection of different candidates.\textsuperscript{101} Chemists have used their creativity and knowledge to explore other chemical space to create new bps. Metal complexes have been explored\textsuperscript{102-103} as nucleobase surrogates (Figure 11). Mercury has

![Figure 11: Examples of metal-mediated bps and triples.](image)

Taken from references.\textsuperscript{104-106}

been used as a way of bridging weakly interacting T-T mismatches in DNA.\textsuperscript{104} Platinum coordinated to base triples has been used as a receptor for cytosine.\textsuperscript{105} Tanaka and coworkers developed a silver-mediated bp to increase duplex stability,\textsuperscript{107} and a similar
system was later used as a way to trigger conformational change in DNA from a hairpin to duplex initiated by the addition of silver ions.\textsuperscript{106} pH triggered binding of metal ion was also studied on halogenated uracil mismatches in DNA.\textsuperscript{108}

1.3.3 MODIFICATIONS TO NUCLEIC ACIDS: hydrogen bonding attenuation.

Hydrogen bonding is an important interaction in supramolecular chemistry and provides a huge driving force in molecular recognition.\textsuperscript{109-112} Many chemists have created variants of the canonical nucleobases or have tried new heterocycles with various amounts of success. Freier and Altmann performed extensive experiment on over 200 nucleobase variants and found modifications which strengthen hydrogen bonding, neutralize charge repulsion, and pre-organize the backbone for hybridization to strengthen duplex stability.\textsuperscript{113} Some researchers have been able to completely do away with hydrogen bonding interactions while still being able to maintain specificity and canonical nucleic acid base stacking upon hybridization.\textsuperscript{114-117} The labs of Schultz and Kool have even shown that hydrogen bonding is not necessary for nucleic acid replication. Using isosteres incapable of hydrogen bonding, enzymes were able to faithfully insert the proper isostere according to the template strand even though no hydrogen bonds are present.\textsuperscript{7, 118-121} Knowing that hydrogen bonding is not an absolute necessity for duplex formation and stability, many groups have designed fluorescent analogs of nucleobases as a way to study nucleic acids. Typically, the substitute is a fluorescent reporter acting through a quenching
mechanism or turn-on mechanism. These types of modified bases will be discussed in Chapter 3.

1.3.4 MODIFICATIONS TO NUCLEIC ACIDS: triplex promoting modifications.

Nucleic acid triplex formation is typically not accessible under physiological conditions with the canonical bases. Using unmodified nucleotides, a triplex can form under acidic conditions and is highly dependent on divalent cations. Efforts have been made to promote triple-stranded nucleic acid formation under more physiologically relevant conditions via modifications to nucleobases (Figure 12). Early efforts focused on tailoring the nucleobases so that triplex formation could occur at neutral pH. Lee and coworkers produced a nucleic acid incorporating 5-methylcytosine that was able to form a stable triplex with duplex DNA at pH 8.122 Xodo and coworkers showed methylation produced an entropically favorable interaction.123 This methylation strategy is not general across all nucleic acids as Kool and coworkers showed a destabilization of the triplex upon the incorporation of 5-methylcytosine in RNA triplexes.124 However, they did report a stabilization of RNA triplexes when using 5-methyluracil.

Most modifications have focused on improving binding to purines in the major groove by modulating the artificial nucleobases ability to form Hoogsteen and reverse Hoogsteen hydrogen bonds and is covered extensively in a review.125 A few examples of these modifications include introducing new aromatic heterocycles to reduce pH dependence126, extending the π-system to induce intercalation127, increasing hydrogen
bonding to increase stability\textsuperscript{128}, introducing positive charge to shield electrostatic repulsion\textsuperscript{129}, incorporation of reactive moieties for covalent linking\textsuperscript{130}, and other more intricate designs.\textsuperscript{125, 131-132}

![Chemical structures](image)

**Figure 12:** Examples of triplex promoting nucleic acids. Bottom image from reference.\textsuperscript{129}

### 1.3.5 MODIFICATIONS TO NUCLEIC ACIDS: triplex formation via Janus wedge.

The final modification to be discussed involves the Janus wedge motif. Named after the Roman god Janus with two faces looking forward and backwards, Janus wedge motifs present two faces on the molecule capable of interaction with two separate nucleobases via hydrogen bonding and base stacking (Figure 13). Janus wedge motifs have been designed to insert into base mismatches within RNA/DNA and to create triplexes with
two non-interacting strands of nucleic acids. Efforts have been made in materials chemistry to create complex assemblies using Janus wedges and nucleobases \(^{105, 133-135}\). Several groups have sought Janus wedge motifs in an effort to drug various forms of malignant DNA and RNA such as G or C nucleotides in viral genomes\(^{136}\), G-U mismatches in Tau pre-mRNA\(^{137}\), and repeated U-U mismatches in the DMPK gene causing myotonic dystrophy type 1.\(^{138}\) Baranger and Zimmerman used a melamine heterocycle attached to an intercalator to target U-U mismatches. The melamine moiety provided a discriminatory motif that would hydrogen bond with two uracil Watson-Crick faces while the intercalator acridine provided the binding driving force.\(^{138}\) Shin and Tor were able to create a bicyclic nucleotide surrogate capable of stably forming DNA duplexes with A or T.\(^{139}\) Similarly, Perrin and coworkers developed a nucleotide which was used to target A-T bps,
preferentially binding to the A-analogue.\textsuperscript{140} In a series of elegant experiments, McLaughlin and coworkers were able to use PNA to insert a series of Janus wedge motifs into 8 continuous T-C mismatches, A-T bps, and G-C bps by forming PNA triplexes with DNA.\textsuperscript{141-142} Finally, melamine and cyanuric acid were used to create hexameric rosettes of DNA which further assembled into fibers.\textsuperscript{143}

1.4 MODIFICATIONS TO NUCLEIC ACID BACKBONES.

Before Bruce Merrifield invented solid phase synthesis\textsuperscript{144} in the 1960s, nucleic acids, especially designed nucleic acids, were difficult to obtain. Thanks to solid phase synthesis and phosphoramidite chemistry, chemists are able to chemically synthesize any nucleic acid sequence imaginable in high yields and purity. Chemists have begun to experiment with the structure of nucleic acids by introducing new and unusual nucleic acid monomers by modifying the ribose and phosphate moieties. Due to the possible application of nucleic acids as therapeutic reagents\textsuperscript{145-146}, efforts have been made to strengthen their binding to target sequences and protect them against degradation within biological systems.

1.4.1 MODIFICATIONS TO NUCLEIC ACID BACKBONES: phosphodiester modifications.

Examples of modifications to the phosphodiester bond are less prevalent than modifications to the sugar moiety of nucleic acids (Figure 14). Eckstein and coworkers
introduced phosphorothioates in 1967\textsuperscript{147} which showed resistance to enzymatic degradation.\textsuperscript{148} Sulfur atom-substituted oligonucleotides have been used to form stable triplexes with targeted duplex nucleic acids with minimal change in triplex stability.\textsuperscript{149} Phosphorothioate nucleotides have been effective in inhibition of transcription through triplex formation in an antisense strategy and effective antisense reagents in gene knockdown experiments.\textsuperscript{150-153} Improvements in duplex stability was achieved by Gryaznov and coworkers by introducing phosphoramidates into the backbone of nucleic acids by replacing the 3′-oxygen atom in the 2′-deoxyfuranose system with an NH group. Increased stability is thought to arise from a more rigid backbone.\textsuperscript{154-155} These zwitterionic nucleotides are also able to form triplexes in low salt, possibly due to reduced electrostatic repulsion.\textsuperscript{156-157} Other modifications have aimed at conformationally restricting the backbone of the nucleic acids via cyclization. Leumann and coworkers have created a bicyclodeoxynucleoside (bcdX) with a furanose ring as the core sugar moiety. While the bcdX nucleosides showed slightly weaker duplex stability, the triplex stability incorporating the bicyclic moiety showed greater stability.\textsuperscript{158}

![Figure 14: Examples of phosphoramidates and phosphorothioate.](image-url)
1.4.2 MODIFICATIONS TO NUCLEIC ACID BACKBONES: ribose modifications.

The ribose moiety has been targeted for modification not only to increase stability, but also as a target for the study of evolution of the basic building blocks of life (Figure 15). A rather simple modification was the methylation of the 2’-hydroxy moiety which has shown to increase stability of duplexes and triplexes by preventing 2’-hydroxyl-catalyzed hydrolysis.\textsuperscript{159-160} Removal of the C-C bond between C2’ and C3’ has created a new class of nucleotides called unlocked nucleic acids (UNA) which are resistant to enzymatic degradation.\textsuperscript{161-162} UNAs destabilize a duplex with respect to its native sequence due to greater mobility in the backbone. The greater entropy of the UNA is utilized to create synthetic oligonucleotides with varied specificity. Preceding UNA was the locked nucleic acid (LNA) which greatly stabilized duplex formation when hybridized with native nucleic acids due to greater preorganization of the backbone via introduction of a 2’-C, 4’-C-oxymethylene linkers. LNAs show enhanced stability to enzymatic degradation.\textsuperscript{163-164} Imanishi’s lab has shown the ability of LNAs to form triplexes with DNA in more basic conditions, pH 7.0, possibly due to greater rigidity.\textsuperscript{165-167} LNAs modified with a 5’-acridine derivative have formed stable triplexes with DNA at physiologically relevant temperature and pH with sub micromolar efficacy in gene silencing.\textsuperscript{166, 168} A fluorinated LNA at the C2’ and C4’ has shown increased activity in gene silencing studies.\textsuperscript{169} Usage of LNA has become a popular modification seeing use in various applications in bioengineering,\textsuperscript{170-172} and LNAs have been further modified with the 2’-amino
functionality which allows for the appendage of fluorescent reporters.\textsuperscript{173-175} The diastereomer of LNA, 2′-amino-\(\alpha\)-L-LNA, has also been developed further increasing the toolbox of modified base technology as a stabilizing moiety that allows for fluorescent labeling of target RNA with abasic sites.\textsuperscript{171,176-177} \(\alpha\)-L-LNA have been shown to form stable triplexes with DNA at pH 6.8 with greater thermal stability when compared to its DNA counterpart.\textsuperscript{161,178}

![Figure 15: Examples of locked and unlocked nucleic acids.](image)

Images taken from references.\textsuperscript{161,176}

**1.4.3 MODIFICATIONS TO NUCLEIC ACID BACKBONES: addressing backbone charge repulsion.**

One of the major factors affecting hybridization of nucleic acids is charge. Nucleic acids fold in a way which minimizes electrostatic repulsion between negatively charged phosphate groups while maximizing base stacking and hydrogen bonding.\textsuperscript{179-180} One way to stabilize duplexes and triplexes is the introduction of positively charged moieties which help alleviate charge repulsion from the backbone (Figure 16). Atsumi and coworkers found that introduction of an amino group at the C4′ position via ethyl and propyl linkers stabilized triplex formation whereas methyl and 2-[N-(2-aminoethyl)carbamoyl]oxy]ethyl-
linker destabilized triplexes.\textsuperscript{181} Introduction of 2′-aminoethoxy oligonucleotides also stabilizes triplex formation\textsuperscript{182-183} via hydrogen bonding to the nearest phosphate.\textsuperscript{183-184} Fox and Brown added additional positively charged groups to DNA systems via addition of a nucleobase modified with 5-propargylamino group.\textsuperscript{185-186} These systems were even more stable than 2′-amionethoxy modified systems due to the additive effects of two positive charges.\textsuperscript{187}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure16.png}
\caption{Example of nucleobase modifications which introduce charge. Images taken from references.\textsuperscript{181, 183}}
\end{figure}

1.4.4 MODIFICATIONS TO NUCLEIC ACID BACKBONES: ribose substitutes.

Chemists have explored surrogates for the ribose moiety (Figure 17). Substitution with a morpholine ring for the ribose results in a phosphoramidite bond and a new class of antisense nucleotides. Phosphorodiamidate morpholino oligomers also form triplexes but are still limited to low pHS for complexation to occur.\textsuperscript{188} Another ribose substitution which destabilized both duplex (\(-4\) to \(-2\)^\circ C T_m\) change) and triplex (\(-13\) to \(-1\)^\circ C T_m\) change) formation was presented by Leumann and coworkers in which they introduced pyrrolidino-C-nucleosides.\textsuperscript{189-190} Herdewijn and coworker substituted the ribose for 1′,5′-
anhydroxhexitol into nucleic acids (HNA) and found, under specific conditions (50 times excess) that triplex formation was favored. Utilizing a rather interesting W-shaped nucleic acid (WNA), which incorporates a bicyclic ringed system and a benzene ligand to provide stacking interactions, Sasaki and coworkers formed stable triplexes with great selectivity for oligo TA duplexes. Finally, Zhang and coworkers found that glycol nucleic acids (GNA), in which a simple 1,2-ethandiol links the phosphate groups, have greater duplex thermal stability than canonical DNA and RNA by 22.5 °C and 20.5 °C, respectively, for the sequences they studied. The unusual stability arises from preorganization of the backbone into a helix, as determined by CD experiments. Pedersen and coworkers interspersed GNA-intercalating conjugate monomers into native ssDNA to

---

**Figure 17: Examples of ribose substitutes.**

Images taken from references. 188, 190-191, 193-195
increase stability of duplex but could only modestly stabilize triplexes at physiologically relevant pH. A peptoid nucleic acid was attempted by Zarra and coworkers, but no CD or thermal melting signature could be found.

1.5 PEPTIDE NUCLEIC ACIDS

After the discovery of nucleic acids, chemists have explored not only why this polymer has achieved such a privileged position, but if any other polymer could suitably replace the natural backbone of nucleic acids. Jones and coworkers was able to weakly bind DNA with various polyacrylate and polyacrylamide polymers functionalized with nucleobases. Other groups have worked with polyvinyl polymers and acryloylnucleosides as nucleobase displaying polymers with some success. Work in Albert Eschenmoser’s lab has shown that the ribose sugar is not privileged structure and that other sugars are as effective as ribose in displaying genetic information. Lo and Sleiman used DNA-mimetic π-conjugated poly(p-phenylenebutadiynylene) as a DNA targeting molecule which was able to bind and d[A$_{20}$] and d[T$_{20}$] nucleic acids and enhance fluorescence. In similar studies, the Bong group has been able to bind both RNA and DNA with ten sequential U-U/T-T mismatches using a polyacrylate polymer displaying a Janus-wedge motif.

Surprisingly, a backbone similar to the poly-α-amino acid bearing the canonical nucleobases was discovered which was capable of binding single stranded nucleic acids.
Using achiral 2-aminoethylglycine units connected via amide bonds, Nielsen and coworkers introduced peptide nucleic acids (PNA) which were able to bind to oligo-A

![Diagram of PNA structure](image)

**Figure 18: Structure of PNA based on protein backbone.** Image from reference.\textsuperscript{208}

DNA and outcompete native, complementary Oligo-T to form triplexes (Figure 18).\textsuperscript{209}

Later work expanded the recognition of PNA to all nucleic acid sequences and showed its ability to form duplexes with DNA, RNA and other PNA molecules through Watson-Crick type hydrogen bonding with base discrimination.\textsuperscript{210-213} The superior stability of PNA complexes, as shown by higher $T_m$ values, appears to stem from the neutral backbone which eliminates the charge repulsion found in the natural nucleic acid duplex. Several structures have been solved via different techniques for the RNA-PNA duplex, RNA-PNA duplex, and DNA-PNA\textsubscript{2} triplex. The RNA-PNA duplex structure seems to be biased towards RNA as it appears to adopt an A-form structure similar to RNA-RNA duplexes.\textsuperscript{214} DNA-PNA complexes adopt a conformation completely different from the typical A-form and B-form of natural nucleic acid duplexes. The duplex and triplex of DNA-PNA adopts
a new formation called P-form, which has characteristics of 3.4 Å rise per nucleobase similar to B-form DNA but with 16 bps per turn. PNA can form duplexes and triplexes with single stranded RNA and DNA, and it is able to displace a hybridized nucleic acid in a duplex through a mechanism called strand invasion. Nielsen and coworkers investigated the mechanism using an elegant CD experiment and proposed that the PNA invasion takes place after non-specific association of the PNA, breathing of the Watson-Crick bps, and subsequent binding of the PNA (Figure 19). Later work by Nielsen and coworkers improved invasion efficiency via covalent linking of intercalators to the PNA and showed the ability to perform double duplex invasion with diaminopurine-thiouracil bps. A major obstacle for using PNAs as an antisense type drug was its propensity to form invasive triplexes with a PNA₂/DNA ratio. Researchers were eventually able to out-compete the invasive mechanism by using pseudoisocytosine, oligo-lysine, and 9-aminoacridine conjugated to the PNA oligomers to achieve PNA/DNA₂ ratio.

Figure 19: Mechanism of PNA strand invasion of duplex nucleic acids. Image from reference.
1.5.1 PEPTIDE NUCLEIC ACIDS: incorporation of Janus wedge motifs.

PNA recognizes single-stranded and double-stranded nucleic acids through hydrogen bonding via Watson-Crick and Hoogsteen faces of the nucleobases. This approach works well for triplexes T-A-T, C\textsuperscript{+}-G-C, A-A-T, and G-G-C but does not cover the full scope of possibilities, such as mismatched nucleobases, due to difficulty in triplex formation with pyrimidine bases. Lehn initially proposed the use of a Janus wedge motif which would present Watson-Crick like hydrogen bonds on both faces of a heterocycle to form a base triple as a way of binding mismatched nucleic acids (Figure 20).\textsuperscript{133} McLaughlin and coworkers successfully used an oligo-PNA with a Janus wedge motifs to target T-C mismatches in DNA.\textsuperscript{141} Using a heterocycle called W which presents hydrogen bonding motifs that would complement the Watson-Crick faces of thymine and cytosine, they were able to form stable triplexes with a duplex DNA with 8 consecutive T-C mismatches flanked by eleven canonical bps. In later work, heterocycles targeting A-T and G-C bps were able to successfully form triplexes when canonical bps were flanked by multiple mismatched bases and discriminate against non-complementary nucleobases.\textsuperscript{142} Ganesh and coworkers utilized a cyanuric acid moiety displayed on a PNA backbone to strengthen triplex formation with oligo-A DNA.\textsuperscript{222} They found that a single cyanuric acid moiety in the middle of an oligo-T displaying PNA stabilized the triplex by a 12 °C in T\textsubscript{m} measurements by providing a Watson-Crick competent binding face no matter the rotamer of the cyanuric acid presented to the DNA nucleobase.
1.5.2 PEPTIDE NUCLEIC ACIDS: native peptide backbone.

After Nielsen discovered PNA as a competent nucleic acid mimic, other groups have looked at native peptide backbones as a way of displaying genetic information. Lenzi first attempted a native peptide backbone nucleic acid with a self-complementary sequence incorporating thymine and adenine nucleobases, but the hairpin had a modest Tm of 19 °C, possibly due to a lack of flexibility in the backbone. Diederichsen synthesized alanyl peptide nucleic acids with alternating stereo centers to create nucleobase displaying beta sheets, again with modest success compared to PNA. Impressively, Huang and coworkers introduced αPNA in which the nucleobases are displayed on an alpha-helical peptide. Using a positively-charged helical structure, tight binding to the DNA sequence GGAGG with a Kd of 200 nM was achieved as well as high melting temperatures in a possible triplex structure.

Eschenmoser and Krishnamurthy, interested in the primordial evolution of genetic material, studied α-peptides and peptoids displaying triazine heterocycles as potential precursors to modern nucleic acids (Figure 21). In their systems 2,4-diamino-1,3,5-triazine and 2,4-dioxo-1,3,5-triazine rings were displayed on a glutamyl, aspartyl,
iminodiacetic acid, and ethylenediamine monomers with alternating aspartic acid or dialkyl-ammonium groups for solubility. Their findings showed that the binding of these small oligos, displaying as few as six nucleobase mimics, were robust in binding to RNA and DNA when containing the 2,4-diamino-1,3,5-triazine heterocycle displayed on the longest linker. They found that the 2,4-dioxo-1,3,5-triazine nucleobase mimic did not bind to nucleic acid targets due to deprotonation in the assay buffer. While they ruled out that the triazine heterocycles as possible nucleobase precursors, they did show that α-peptides

![Figure 21: Native peptide backbone PNA with triazine ring recognition moiety.](image)

Figure 21: Native peptide backbone PNA with triazine ring recognition moiety.

Image from reference.226

are competent as nucleic acid analogs even with negatively charged backbones. Furthermore, while not a major finding of the paper, they found that the 2,4-diamino-1,3,5-triazine is effective as a nucleobase mimic capable of recognizing uracil and thymine. In a follow up communication using α-peptides, the researchers showed that pKa of the nucleobase is important.227 Their findings yielded a relationship between pKa of the base and pH of the solution, in that if the base becomes deprotonated or tautomerizes in the
buffer (has a pKa near or equal to the pH of the solution), then molecular recognition is weak.

1.5.3 BIFACIAL PEPTIDE NUCLEIC ACIDS (bPNA)

In 2012, the Bong laboratory at The Ohio State University reported the binding of a 21mer α-protein bPNA, composed of alternating lysine and glutamic acid amino acids displaying 2,4,6-diamino-1,3,5-triazine on the lysine side chain, to oligo-thymine DNA (Figure 22). The system was designed to bind oligo thymine and oligo uracil using melamine as a Janus wedge motif capable of forming a triplex with the nucleobases.\footnote{228-229} Fluorescence anisotropy and molecular beacon experiments indicate discreet triplex formation with two dT_{10} tracks with high affinity (dT_{10} K_d \approx 4000 \text{nM}^2). Triplex formation can also occur via hybridization of bPNA with two dT_{10} tracks attached via a single stranded region to form a hairpin triplex (K_d= 2.7 nM) with high thermal stability (54 °C).\footnote{228} The sequence of bPNA follows the spacing presented by Eschenmoser’s work with the negative charge providing no destabilizing electrostatic effects. Aside from the backbone structure and unnatural nucleobase, bPNA differs from triplex-forming nucleic acids and PNA via its mechanism of formation. Triplex-forming nucleic acids interact with a preformed duplex through interaction with the major or minor groove and standard PNA involves the invasion of a preformed duplex. bPNA acts through an associative mechanism bringing two non-interacting nucleic acids together during triplex formation. In contrast to PNA and TNA which require a preformed duplex, bPNA binds to two non-interacting strands of nucleic acids to form a triplex.
The Bong lab has employed this bPNA a number of systems as a regulatory switch (Figure 23). They have shown that bPNA stably binds to oligo-uracil RNA in a comparable fashion to the DNA system.\textsuperscript{229} bPNA triplex formation has been able to compete with native enzymes to block transcription, exonuclease digestion, and reverse transcription due to significant structural reorganization.\textsuperscript{229} Enzymes are able to function until their encounter with the bPNA triplex, at which point enzymatic activity is aborted. While bPNA has been used to turn off enzymatic processing, it has also been used to restore function to aptamers and ribozymes. As an allosteric switch, bPNA has been shown to refold crippled ribozymes and restore self-cleaving activity.\textsuperscript{230} Similarly, bPNA is able to rescue crippled RNA and DNA aptamers which bind to a small fluorogen and protein, respectively.\textsuperscript{230} Furthermore, using oligo-thymine tracks as templates, the Bong lab have

**Figure 22: Structure of bPNA.**
Recognition unit melamine (Left), structure of bPNA (Center) and proposed triplex structure (Right). Image from reference.\textsuperscript{228}
been able to catalyze ligation of bPNA modules to restore function to a crippled ribozyme.\textsuperscript{230}

**Figure 23: bPNA as an allosteric switch.**
A) hammerhead ribozyme B) spinach aptamer C) protein aptamer. Images from reference.\textsuperscript{230}

Up to now, the triplex-incorporating bPNA has been used as a surrogate for a duplex region of a nucleic acid. When the triplex has been formed within the context of a hairpin, the loop sequence is usually an innocuous sequence with no function other than to bridge the binding tracks for bPNA. However, the loop regions of stem-loop structures are often utilized by nature as docking sites and readily bind to various targets. Hairpins may dock to proteins, bind small molecules, or bind other RNAs intermolecularly and intramolecularly. In the following work, we study the extent to which bPNA triplexes can promote loop-loop hybridization of hairpins, the so called “kissing-loop” complex. In Chapter 2, two kissing-loop systems were mutated with a bPNA binding site in their stem regions in order to examine the behavior of bPNA promoted kissing-loop complexes. Data will be presented showing the effect of bPNA triplex incorporation on hairpin binding.
hybridization and show compelling evidence that bPNA structuring of the stem aids in complex formation. In Chapter 3, we will investigate the use of bPNA triplexes as a selective and specific modification that can be used to chemically tag and modify a target hairpin nucleic acid. Efforts have been made to turn the triplex-modified hairpins into fluorescence reporters to detect not only the presence of a hairpin, but also kissing-complex formation. We hope to use this directed chemistry approach to fluorescently label and modify specific nucleic acid structures through functionalization of bPNA with reactive moieties.
REFERENCES FOR CHAPTER 1


39


201. Gait, M. J.; Jones, A. S.; Jones, M. D.; Shepherd, M. J.; Walker, R. T., Synthetic analogues of polynucleotides. Part 15. The synthesis and properties of poly[5′-prime or minute]-amino-3[prime or minute]-O-carboxymethyl-2[prime or minute],5[prime or
minute]-dideoxy-erythro-pentonucleosides) containing 3[prime or minute](O)[rightward arrow] 5[prime or minute](C) acetamidate linkages. *Journal of the Chemical Society, Perkin Transactions 1* **1979**, (0), 1389-1394.


Chapter 2: bPNA promoted kissing-loops

2.1 INTRODUCTION TO KISSING-LOOPS:

The hairpin motif, also called a stem-loop motif is one of the most common secondary structures in nucleic acids. Hairpins are formed when two complementary nucleic acid sequences within the same oligo nucleic acid hybridize to form a duplex with a single stranded sequence connecting the two strands at one end of the duplex (Figure 24A). The intervening unhybridized loop displays information-rich nucleobases which engage in a wide variety of recognition events with proteins, small molecules, and other nucleic acids. The loop region can hybridize with single stranded nucleic acids through Watson-Crick type interactions such as the binding between the anticodon loop of tRNA and mRNA inside the ribosome. Alternatively, two hairpins could hybridize with one another by Watson-Crick base pairing to form a duplex with the two loop regions and this has been termed a “kissing” loop complex (Figure 24B). Kissing complex formation can occur with as few as two bps and has a very strong Mg\textsuperscript{2+} dependence. The Mg\textsuperscript{2+} ion sits in a pocket at the kissing-loop interface, stabilizing the loop-loop duplex by minimizing the charge repulsion of the nucleic acid backbones.
Loop-loop interactions can form intramolecularly and help to stabilize the tertiary structure of nucleic acids. This is best exemplified in the very first nucleic acid crystal structure, yeast phenylalanine tRNA, in which the canonical L-shape tertiary structure is stabilized via loop-loop interactions through the D-loop and T-loop hairpins (Figure 24).

**Figure 24: Structure of hairpins and kissing-loops.**
A) Intramolecular formation of a hairpin with duplex stem (red and blue squares) and single stranded loop (green squares). B) Kissing complex formation via Mg$^{2+}$ dependent hybridization of single stranded loop regions between hairpins.

In an example of kissing complexes in regulation of cellular processes, viral genomes often utilize intramolecular kissing-loops to stabilize their genome in order to obtain the necessary conformation for binding to target proteins. Disruption of these kissing-loops negatively disrupts viral mRNA levels and even inhibits infectivity. Catalytic RNA and riboswitches also depend on kissing-loop formation in order to achieve their function. Group II introns stabilize their tertiary structure with a kissing-loop between loops $\alpha$ and $\alpha'$ in order to achieve splicing activity during regulation of gene expression. Similarly, Neurospora Varkud satellite ribozyme requires kissing-loop formation to induce
structural changes in the RNA and loss of the kissing-loop lowers ribozyme activity and favors cleavage over ligation (Figure 25A). Cells control gene expression through riboswitches which bind a small molecule in a process that monitors cellular concentrations of the ligand. As the concentration of small molecules change in a cell, a reversible binding event causes structural changes to the RNA and modulates gene expression. Often, riboswitches rely upon kissing-loops to fold and stabilize the pocket for ligand binding, and when these kissing-loops are abolished, riboswitch function is similarly abolished.

From the examples above, it is shown that long range interactions through kissing-loop formation play important roles in cellular activities.

**Figure 25: Intramolecular loop-loops interactions of Neurospora Varkud satellite ribozyme and tRNA.**

A) Intramolecular kissing-loop of Neurospora Varkud satellite ribozyme as indicated by dotted line. B) Intramolecular loop-loop interaction of tRNA as indicated by red box. Images take from reference.16, 21

Stem loops need not be part of the same nucleic acid to form a kissing-loop complex. Intermolecular loop-loop interactions have also been evolved to mediate important molecular recognition events in cellular and viral life cycles.27-28 There are many examples of intermolecular kissing-loops and a few interesting examples involve...
Retroviruses carry two near identical copies of their RNA genome and often form homodimers of their genomic RNAs to allow for more efficient operations such as recombination during reverse transcription, translation, and packaging. The dimerization initiation site (DIS) within the dimer linkage structure (DLS) of the HIV retroviral genome typically employs a stem-loop structure with a self-complementary loop sequence which allows homodimerization. Retroviruses that utilize kissing-loops in genomic dimerization include HIV-1, HIV-2, HTLV-I, human foamy virus, bovine leukaemia virus, avian retroviruses, HaSV, MuSV and MoMuLV. Kissing-loop formation has proven to be vital for HIV-1 and avian retroviruses dimerization, however, it does not sufficient enough to promote dimerization in other retroviruses which utilize further tertiary contacts. A discussion of the DIS in HIV-1 will follow below.

Cells utilize kissing-loop competent RNAs to control the copy number of genomic material. One such example of regulation in the production of genomic material is cis-encoded antisense RNAs which are produced from the same DNA locus as their target sequence and prime their target sequence for destruction. Being from the same DNA locus, the antisense sequence is perfectly complementary to their target’s sequence because the DNAs they are transcribed from are complementary. These antisense RNAs work by affecting transcription, primer formation, and translation by causing conformational changes or creating persistent hybrids with sense RNAs. There are many examples throughout nature, but we will only highlight a few here involving kissing-loop complexes. The first example involves translation of initiator protein RepA which is required for
replication of plasmid R1 (Figure 26C). RepA is translated from repA mRNA. CopA is an antisense RNA which binds to the stem-loop region CopT in the repA mRNA. CopA-CopT hybridization is initiated by kissing-loop formation followed by reorganization into an extended duplex. The extended duplex is primed for digestion by nucleases which prevent overproduction of the RepA protein. The next example involves the expression of RepR protein in Streptococcus agalactiae. RepR controls plasmid copy number through a role in the initiation complex during replication and is controlled by antisense RNAIII binding to the plasmid pIP501 (Figure 26A). Transcription of the repR gene is prevented via a multistep process which first involves kissing-loop complex formation and then duplex formation causing transcription abortion. Repression of RepA and RepR proteins involve priming of RNA target for destruction via RNA-RNA interactions. The next two examples require a protein to mediate the RNA-RNA binding event. An example of protein expression control by antisense RNA is the TraJ protein and antisense RNA FinP (Figure 26B). Antisense FinP is a small RNA with a stem-loop structure which binds to traJ mRNA, preventing translation. This RNA kissing complex is stabilized by another protein, FinO which increase the half-life of the complex. It is thought to do so by a chaperone mechanism in which FinP is protected against degradation and duplex formation is promoted between sense and antisense RNA. The FinP-traJ interaction is an example in which two negative feedback controls are required for regulation, both a nucleic acid and a protein. Another example of protein-mediated antisense control of plasmid copy number can be found in E. coli (Figure 29). The ColE1 plasmid replication is primed by RNAII binding to the plasmid DNA followed an enzymatically catalyzed cleavage event to form
the mature primer. RNAII binding is controlled by antisense RNAI which induces a conformational change in RNAII preventing primer formation. The interaction between sense and antisense RNAs is mediated by kissing-loop formation between three complementary stem loop structures in each RNA. The protein Rop (or Rom) encoded in the plasmid then binds the kinetically unstable kissing complex to create a stable protein-RNA complex. Again, two feed-back systems are employed to regulate plasmid production, both a nucleic acid and a protein. A more detailed discussion of the ColE1 kissing complex will follow below.

Figure 26: Cis-encoded antisense RNAs and their targets. This image shows intermolecular kissing-loop interactions used to control transcription, translation, and replication. A) Suppression of RepR protein production via duplex formation first initiated by RNAII and mRNA kissing-loop formation. B) Antisense FinP forms kissing-loop complex traJ mRNA which is stabilized by FinO. Duplex formation prevents translation. C) Cop A forms kissing complex with CopT in mRNA encoding repA. Duplex formation prevents expression of repA. Images from references.²⁸, ⁴⁰
Nature has adopted the kissing-loop interaction to form specific binding events with high affinity and it only makes sense that chemists would take advantage of this reliable interaction to create synthetic molecules. One such example by Scarabino and coworkers utilizes *in vitro* selection to create an aptamer for tRNA.\textsuperscript{50} Using an 80-nucleotide random library, they were able to pull out aptamers which were able to bind to the anticodon loop of the phenylalanine tRNA of *Saccharomyces cerevisiae*. These aptamers formed perfect kissing-loop complexes with Watson-Crick type interactions. A second set of aptamers complexed to the D-loop and T-loop showed improved affinity when the kissing complex between the D-loop and T-loop was abolished. Furthermore, a simple single-stranded complement did not bind to the anti-codon loop of the tRNA. Luc Jaeger and coworkers have utilized kissing-loops to engineer RNA-based materials.\textsuperscript{51-53} Using designed kissing-loop interactions based on HIV-1 DIS kissing-loop, they have been able to create 3 dimensional, cubic RNA scaffolds and two-dimensional programmable materials that can be used to coat surfaces. The complexity of their systems is possible due to the use of orthogonal loop sequences to ensure proper base pairing and reliable structuring. Finally, the Peixuan Guo group has created pRNA nanoparticles to deliver RNA based drugs using kissing-loop interactions.\textsuperscript{54-57} To create cyclic trimers and heptamers of small RNAs, orthogonal kissing-loop interactions were mutated into bacteriophage phi29 DNA packaging motor. These motors are then loaded with drug protein and associate with viral proteins to form the nano-delivery vehicle. The above examples are just a few attempts to use kissing-loops to design nucleic acid materials using naturally occurring motifs. Thanks
to databases which store structures and sequences, chemists can design nucleic acid structures and kissing-loops can serve as key motifs for non-covalent assembly.

2.1.1 HIV-1 DIMERIZATION INITIATION SITE:

**Figure 27: DIS in the 5′ leader of the HIV genome.**
A) Dimerization initiation site of the HIV genome. B) Kissing-loop formed by DIS and loop sequences of various subtypes. Image taken from reference.29

Retroviruses, such as HIV-1, commonly have dimerized copies of genomic RNA which is thought to be tied to key events in the viral life cycle. The HIV-1 DIS, the structure responsible for dimerization, is in the DLS of the 5′-leader (Figure 27). The DIS is thought to be a stem-loop with a palindromic loop sequence which allows for homodimerization
via kissing complex formation.\textsuperscript{13, 31, 58} The DIS stem-loop hybridization to dimeric RNA is controlled by an equilibrium in which the DIS is exposed for dimerization or unexposed for replication.\textsuperscript{59} When unexposed, the DIS is bound to an internal sequence known as the unique-5' (U5). The conformation is believed to be controlled by the number of guanosines at the 5' end of the leader (1G, 2G, or 3G) which causes a shift in base pairing.\textsuperscript{60} When exposed, the DIS displays a nine-nucleotide loop with two major sequences subtypes; subtype A (AGGUGCACA) and subtype B (AAGCGCGCA).\textsuperscript{61} Perturbations to the kissing-loop sequence have been found to be detrimental to dimerization. For example, alkylation of the phosphate groups and N7-methylation of the guanine residues completely abolishes dimerization.\textsuperscript{61} Furthermore, single mismatches in the loop sequences and unstructuring of the stem loop results in abolished dimerization.\textsuperscript{31, 62-63} However, compensatory mutations and strategically placed wobble bps allowed for the formation of homodimers while introduction of RNAs with complementary sequences resulted in heterodimers.\textsuperscript{31, 63} The unpaired purines flanking the self-complementary dimerization sequences have proven to be very important for dimerization.\textsuperscript{64} While DIS mutants with deletion of one of the flanking adenines on either the 5' or 3' side of the loop remain dimerization competent, deletion of two or more results in unstable dimerization.\textsuperscript{31, 63, 65} Mutations of the purines to uracil results in a dimerization competent mutant but alters the kinetics.\textsuperscript{65} Aptamers selected for dimerization competence also have provided evidence of the importance of the purine heterocycles.\textsuperscript{66-67} In these studies, adenines were selected preferentially in position 1, 2 and 9 due to an optimal sheared bp geometry. Substitution of an unpaired flanking base with another bp did not improve complex stability. The exact
role of these purine bases is still under debate though evidence supports that they interact in non-canonical interactions and base triples.68-69

The structure of the HIV DIS kissing complex has been studied by NMR, crystallography, and FRET. The structures presented in NMR and crystallography differ greatly in structural assignment (Figure 28). In the NMR presented by Parslow and coworkers, the duplex stems are not coaxially stacked onto the kissing-loop duplex.70 The planes of the kissing-loop’s bps are perpendicular to the bps in the stem duplex. The unpaired adenines are projected into the kissing-loop region forming non-canonical interactions with the base-paired palindromic sequence and partially stacked in an intermolecular fashion with one another. In the crystal structure published by Ehresmann and coworkers, the kissing-loop duplex is coaxially stacked onto the duplex stem with the bps in the kissing complex parallel to the bps in the stem.8 The unpaired adenines are bulged out to form crystal contacts in the form of a purine-zipper.71 Later crystal structures showed there was in fact a Mg\(^{2+}\) ion bound at the kissing-loop interface which could be displaced by spermine.71 The cation dependence was studied in depth in work by Weixlbaumer and coworkers using UV-melting experiments.72 Their findings show a 4 kcal/mol more stable kissing duplex compared to a standard RNA helix and the phenomena is sequence independent. Furthermore, the T\(_m\) of the loop-loop complex is more ion dependent than a standard helix by a factor of 2 or more per bp. Conformational analysis using single-molecule FRET analysis of the kissing complex, with dyes covalently bound the termini of the stem loop duplexes, reveals a bent geometry.73 This bent geometry seems to agree with the solution structure presented by Parslow. The structure of the kissing
complex formed by the DIS of HIV-1 is understood thanks to the work described above, however, the actual structure within the context of the full RNA genome is not yet known.

Figure 28: Crystal and NMR structure of DIS dimer. Left) Crystal structure. Right) NMR structure. Images taken from references. 8, 70

Due to the importance of this structure and the HIV-AIDS epidemic, the HIV-DIS has been studied as a possible drug target. Researchers have found success using various glycosides as drugs after noticing the similarities between the HIV-DIS kissing-loop and the site targeted by glycosides in bacterial ribosomal A site. 74-78 Ennifar and coworkers were able to crystalize the DIS kissing complex with neamine, ribostamycin, neomycin, and lividomycin. These aminoglycosides were able to stabilize the kissing complex by
acting as a bridge, binding to both RNAs cooperatively. They also found that aminoglycoside binding inhibited extended duplex formation of the DIS hairpins.

2.1.2 ColE1 KISSING COMPLEX:

Figure 29: Plasmid copy number control mechanism of the ColE1 plasmid. Primer maturation is inhibited by RNAI via kissing complex formation with RNAII which is stabilized by Rop protein. Image from reference.

The kissing complex used in the process to control ColE1 plasmid replication is an intricate study on design dictating function. In the HIV DIS system, the kissing-loop complex is quick to form and very stable, as might be expected given its role as an anchor to bring two identical RNAs together to form a stable dimer for packaging. The kissing complex in the ColE1 system has been designed to control the copy number of the plasmid.
and therefore must be not be persistent but able to adapt to the various conditions of the cell. Reversibility and carefully designed kinetics allow the cells to modulate plasmid production by forming and degrading the RNA kissing complex.\(^79\) Plasmid replication begins by the hybridization of RNAII, which is produced via transcription, to the plasmid at the origin site of replication (Figure 29).\(^80\) The DNA-RNA hybrid is then cleaved by RNAase H to form the primer for DNA synthesis.\(^80\) As transcription proceeds, a small RNA, RNAI, is produced on the opposite strand of DNA in the same location as RNAII. RNAI controls the priming of the DNA template by binding to RNAII via three successive kissing-loop interactions.\(^81\) The kinetics of kissing-loop complexation is vastly different from the HIV DIS kissing-loop in that the kissing-loop complex is slow to form.\(^83\) A protein, called Rop, which is encoded in the plasmid, stabilizes the initial unstable kissing complex.\(^45, 47\) Rop is a 63-amino acid protein which dimerizes to form a four-helical bundle\(^84\) and has low affinity for the individual RNAs but high affinity for the kissing complex (Figure 30).\(^85\) Rop binds to the kissing complex using two phenylalanine side chains and polar interactions which appear to interact with a general shape of kissing complexes rather than a specific sequence.\(^48, 86\) Rop increases binding affinity by decreasing the dissociation rate of the kissing complex.\(^48\) This kissing complex forces RNAII into a conformation which prevents hybridization to the plasmid and shuts down replication. Through this mechanism, plasmid copy number is controlled by the changing concentrations of two components. The kissing-loop complex is unstable and unable to completely shut down the replication process until enough Rop protein synthesis has
occurred in the cell. The levels of Rop would only be sufficient when plasmid copy number is sufficient.\textsuperscript{87}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure30.png}
\caption{Structure of Rop as dimer in solution. Image from reference.\textsuperscript{86}}
\end{figure}

The interesting behavior of the larger RNA transcripts lead to studies of individual ColE1 hairpins. Tomizawa and Eguchi first noticed that when the three kissing-loops in the ColE1 complex were individually studied they did not have the same stability in spite of having the same G-C bp content and loop size.\textsuperscript{46} In later work, they showed that inverting the loop sequence of hairpin 1 increased complex stability 350 fold and simply swapping positions of 1 and 7 in the loop increased stability 133 fold. Furthermore, mutants with 6-membered and 8-membered loops were able to bind more efficiently than the wild type kissing-loop and were recognized by Rop.\textsuperscript{48} Crothers and Gregorian carried out further investigations into the loop sequence of ColE1 and confirmed the importance
of the identity of the nucleobase at positions 1 and 7 when they observed a 19°C increase in $T_m$ for the inverted loop sequence and 15°C increase in $T_m$ when position 1 and 7 were swapped.\cite{87} This effect seemed to be specific to G1-U7 sequence with a complementary A1-C7 mutation in the binding partner. This sequence specificity may have something to do with preorganization within the stem loops of the wild type which destabilized the kissing complex and inter-strand interactions, which are present in the inverted loop, are not present in the wild type.\cite{87} Marino and Crothers have analyzed the structure of the individual hairpins by NMR and found a highly-structured loop in RNAI wild type, which supports this idea of a preorganized wild type stem-loop.\cite{88} Initial NMR solution structure

![Figure 31: Solution structure ColE1 kissing-loop.](image)

of the kissing complex of isolated stem-loops pointed to a fully base stacked, bent structure at the interface of hybridization with all seven nucleotides involved in Watson-Crick type base recognition (Figure 31).\cite{87} The bend is towards the major groove and allows for the continuous stacking of the bps. Critically, the inverse sequence has a purine-purine
stacking interaction at the stem-loop interface and this may be the key interaction for increased stability of the inverse loop sequence.\textsuperscript{89} The kissing loop for the ColE1 plasmid replication was selected to be meta-stable in order for copy number to be concentration dependent.

Clearly hairpins and their kissing-loop complexes play a key role in numerous cellular activities across all cellular domains. Therefore, it would be of great use to develop biologically compatible systems capable of reporting, both \textit{in vitro} and \textit{in vivo}, on stem-loop life time, complexation, and localization within cells. We believe the bPNA triplex is a promising biotechnology that could fit this need based on the ease of synthesis of the bPNA peptides and tolerance of bPNA triplexes in functional nucleic acid. bPNA presents an easily modifiable molecule that can adorned with any functional molecule that is compatible with peptide synthesis. We have shown that as a duplex surrogate, bPNA triplexes are well tolerated in various functional nucleic acids, bPNA binds specifically to poly-U and poly-T domains, and is resistant to nucleases.\textsuperscript{90-94} It is therefore reasonable to expect that mutations of the duplex stem in hairpins to include a bPNA binding site would still be able to function as wild type hairpins in kissing complexes. To test this hypothesis, we chose to study two systems of kissing-loops modeled after the HIV-1 DIS and ColE1kissing complexes. We chose these two systems to explore how triplex mutations affect kissing complexes with different loop sizes, and, in the case of the ColE1 system, how proteins such as Rop which interact with kissing-loop interfaces, bind to kissing complexes incorporating a bPNA triplex. This chapter will discuss various designs to achieve bPNA-supported kissing complexes and the characterization of these complexes.
Figure 32: Proposed experiment with mutated HIV-DIS hairpins. Hybridization of bPNA with U10 mutant would fold RNA into a hairpin structure which would induce kissing complex formation when Mg²⁺ is present.
The bPNA binding mutant HIV-DIS experimental design is shown in Figure 32. We would mutate a stem-loop structure in the 5’-leader sequence of the HIV genome called the HIV-DIS (nucleotides 240-280) by inserting the bPNA binding site, which is ten consecutive U-U mismatches, causing the hairpin to become unstructured. Upon addition of bPNA with the sequence G(EM)_{10}, triplex formation would refold the RNA into the required stem-loop structure allowing for kissing-loop complex formation. Our initial design of the HIV-DIS system aimed at maintaining as much of the structure of the wild type stem-loop as possible (Figure 33) and we included the bulged region which separates the upper stem and lower stem. The loop region would have the sequence 5’-AAGCGCGCA-3’ with the 6-nucleobase palindromic sequence in the middle which would be able to form homodimers. We designed four variants of the mutated HIV-DIS in which the duplex stems of three of the hairpins would be mutated to have ten consecutive uracil bases while maintaining a G-C clamp at the triplex-loop interface to preserve the native structure of the loop. The designs included a lower stem modification (DIS-LS), upper stem modification (DIS-US), a positive control (DIS-PC), and negative control (DIS-NC) (Figure 33). The most ambitious of the designs, DIS-US, would extend the upper stem to ten bases and place the triplex near the interface of the loop just below the G-C clamp. A more conservative design, DIS-LS, would extend the lower stem to ten bases with a G-C clamp before the bulge. The positive control, DIS-PC, would have the bPNA binding site appended onto the 3’-end of the wild type hairpin. Finally, the negative control, DIS-NC, would have the same mutation to the upper stem as DIS-US but a non-complementary loop sequence would prevent dimerization.
Figure 33: Initial design of HIV-DIS mutants.
This figure shows the structures of the initial designs of the mutate HIV-DIS systems. The upper stem, lower stem, bulge, and loop region are labeled on the wild type structure. The dots in the structures represent predicted bps.

The RNAs were produced via T7 transcription runoff from templates hybridized to a T7 primer, purified by denaturing acrylamide gel electrophoresis, and RNA stocks stored in water at -20°C. bPNA was prepared by solid phase synthesis, labeled with cyanine dyes and stored in water with 5 mM of sodium bicarbonate to increase solubility. Samples were prepared by diluting stocks of RNA and bPNA into 1x dimerization buffer to appropriate concentrations, annealing at 95°C for 5 min, and snap cooling on ice for 5 min. 1x dimerization buffer was composed of 50 mM Tris-Cl pH=7.6, 10% glycerol, 250 mM NaCl, and 5 mM MgCl₂. Dimerization of the mutant hairpins were determined using native PAGE assays in 1x TBM1 with 0.1 mM Mg²⁺. The result for dimerization experiments are shown in Figure 34.

As expected, the negative control showed no dimerization even at 10 μM and the positive control showed near complete dimerization even at 10 nM, the lowest concentration tested and visible by SYBR gold staining. DIS-LS also showed near
Figure 34: Dimerization of initial HIV-DIS system.
This figure shows the dimerization efficiency, by native PAGE in TBM, of the various HIV-DIS mutants. Sample buffers contain 10 mM MgCl$_2$ while running buffer contained 0.1 mM MgCl$_2$. Concentration of unlabeled (EM)$_{10}$ was 1:1 with U10 RNA. RNA was stained with SYBR gold.

complete dimerization, however, DIS-US would not completely dimerize even at high concentrations such as 20 μM. We found this a troubling result considering the literature $K_d$ of this system is 0.3 ± 0.14 nM.$^{67}$ This unstable dimerization could be due to the triplex either inducing a conformational change upon folding or disruption of important contacts within the kissing complex which destabilizes the system. Finding it difficult to drive the complexation to completion using higher Mg$^{2+}$ concentrations (Using 1x TBM2 with 10 mM Mg$^{2+}$), we turned our attention to forming heterodimers using loop sequences derived from an in vitro evolution by Lodmell and coworkers (Table ).$^{67}$ We created five pairs of heterodimers, called DIS-HD#, with complementary sequences in positions 3 through 9 of the loop, leaving the flanking adenines intact, in the DIS-US scaffold. Heterodimers were
chosen as it would give us better control as to when dimerization would occur. Unfortunately, all variants tested were unable to form stable heterodimers in our hands due to either homodimerization or incomplete dimerization (Table 1).

Table 1: Heterodimers in initial HIV-DIS system.
This table shows the results of the native PAGE dimerization assay of heterodimeric DIS-HD#. The number in the name is simply a means to name the sequences.

<table>
<thead>
<tr>
<th>DIS</th>
<th>Loop Sequence</th>
<th>Heterodimerization</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD1</td>
<td>5’-AAGCGCACA-3’</td>
<td>Failed</td>
<td>Homodimerization</td>
</tr>
<tr>
<td>HD2</td>
<td>3’-ACGCGTGAA-5’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HD3</td>
<td>5’-AAGUGGACA-3’</td>
<td>Failed</td>
<td>Homodimerization</td>
</tr>
<tr>
<td>HD4</td>
<td>3’-ACACCUGAA-5’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HD5</td>
<td>5’-AAGCAGGCA-3’</td>
<td>Failed</td>
<td>Homodimerization</td>
</tr>
<tr>
<td>HD6</td>
<td>3’-ACGUCCGAA-5’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HD7</td>
<td>5’-AACCUCGGA-3’</td>
<td>Failed</td>
<td>Incomplete</td>
</tr>
<tr>
<td>HD8</td>
<td>3’-AGGAGGCAA-5’</td>
<td></td>
<td>Dimerization</td>
</tr>
<tr>
<td>HD9</td>
<td>5’-AACGCUUCA-3’</td>
<td>Failed</td>
<td>Incomplete</td>
</tr>
<tr>
<td>HD10</td>
<td>3’-AGCGAAGAA-5’</td>
<td></td>
<td>Dimerization</td>
</tr>
</tbody>
</table>

Following these disappointing results, we decided to completely redesign the system. One issue we wished to address, from a practical standpoint, was the low transcription efficiency of the RNA transcripts which caused major delays as material for the experiments required multiple transcriptions. It has been shown that using a tRNA
scaffold during in vivo RNA synthesis of small constructs greatly increased the yield of RNA and we expected that in vitro T7 runoff transcription would also benefit from such a reliably transcribed scaffold. As small RNAs are very difficult to transcribe in vitro, we decided to mutate the anticodon stem of a tRNA scaffold by adding the bPNA binding site and loop sequence. Of course, the biggest issue to address was the inefficient dimerization, which could be attributed to disruption of a key structure at the loop and stem interface. We decided to remove the G-C clamp at the stem-loop junction and simply bring the triplex into the loop sequence. The triplex mutant would have the body of the tRNA, a bPNA binding site in the anticodon stem, and a designed loop. We also performed a mutation to the tRNA scaffold incorporating the wild type duplex, to create a wild type mimic. Finally, we chose to create a heterodimer system with a more heterogeneous complementary loop sequence, 5ˈ-AAGCCCCGA-3ˈ and the complementary sequence 5ˈ-AACGGGGCA-3ˈ with the expectation of suppressing homodimerization (Figure 35).

**Figure 35: Structures kissing-loop competent tRNA mutants.**
This figure shows the redesign of the mutant HIV-DIS. The sequence in red designates the tRNA scaffold while the sequence in black denotes the mutant DIS sequence.
We used the same conditions reported above for the synthesis, purification, and binding assays of the new RNA constructs. The redesigned mutants with tRNA scaffolds did in fact have higher yields in transcription efficiency. Perhaps more importantly, both duplex and triplex containing mutants showed efficient heterodimerization in kissing-loop assays ran on agarose gels in 1x TBM1 (Figure 36). We were able to visualize kissing complexes with the wild type mimics, mixed duplex-triplex heterodimers, and triplex-triplex heterodimers via a gel-shift assay. There is clear shift from a monomeric form to a discreet, slower moving heterodimeric complex. A negative control using a triplex containing tRNA mutant and duplex mutant with the same loop sequence showed no dimerization indicating dimerization occurs due to complementary loop sequences forming a kissing complex and not non-specific interactions between tRNA mutants (Supporting information, Figure 50 and 51). As kissing complexes are Mg$^{2+}$ dependent, we were interested to examine if our mutant tRNA kissing complex could form without divalent cations. As expected, in buffers lacking Mg$^{2+}$ in the reaction or gel complexation was suppressed (Supporting information, Figure 52 and 53). This result indicates the mutant kissing complexes also require Mg$^{2+}$ to form stable heterodimeric complexes much like naturally occurring kissing-loops. To further characterize our mutant complexes and the role bPNA plays, we proceeded to radio label the mutant variants in order to determine their dissociation constants by native PAGE gel.

The results for $K_d$ determination are shown in Figure 37. Literature value for the wild type loop was reported as $0.8 \pm 0.3$ nM$^{67}$ The mutants designed to mimic the wild type complex have a dissociation constant in near agreement with the literature value.
Figure 36: Heterodimerization of tRNA mutants.
This figure shows the heterodimerization of tRNA mutants. The gels were electrophoresed using 1x TBM1 in agarose matrix. One tRNA was held at a constant concentration of 100 nM and the complementary tRNA titrated in. Complexes with U10 species are visualized using a Cy5 labeled bPNA in a 1:1 complex. The wild type mimic was imaged using a $^{32}$P-labeled RNA and phosphorimager.

with a $K_d$ of $2.4 \pm 0.6$ nM. Confident in our design and assay, we then determined the dissociation constants of the U10 mutants with and without the presence of bPNA. Our findings indicate that the triplex is tolerated in the kissing complex. A duplex-triplex mutant heterodimer has a $K_d$ of $9.3 \pm 2.0$ nM which about 5-fold weaker than the wild type heterodimer. A triplex-triplex mutant heterodimer is 2-fold weaker than the duplex triplex kissing complex but is still a relatively strong binding interaction with a $K_d$ of $18.5 \pm 3.2$.
Figure 37: Dissociation constants of tRNA mutants.
This figure shows the dissociation constants for the various mutant tRNAs. The last two dissociation constants were determined without bPNA.

nM. To our surprise, the heterodimer between U10 tRNA and duplex tRNA without bPNA only showed a modest decrease in affinity with a $K_d$ of $30.3 \pm 10.1$ nM which is about 3-fold weaker than when bPNA is bound in the triplex. This surprisingly tight binding could be due to the tRNA body whose structuring may rescue what one would expect to be a crippled dimerization due to the increase in entropy of the unstructured U10 tRNA mutant. A 2-fold decrease in binding affinity was observed in the U10-U10 tRNA dimerization without bPNA which we attribute to the somewhat less structured, larger loops. It should be noted that when bPNA is not incorporated, complexation of the tRNA is incomplete most likely due to an alternative structure incapable of forming a kissing complex. We also determined the apparent $K_d$ for the loop sequences without the tRNA body and bPNA to be $14.4 \pm 3.7$ nM, however two complex bands were formed. The small changes in $K_d$ when bPNA is left out is consistent with Bouchard and Legault’s observation that kissing loops are no more stable than the theoretical stability of a corresponding duplex.\textsuperscript{96}
Encouraged by these results, we began using the U10 tRNA mutants with fluorescently-labeled bPNA as reporting molecules to monitor the wild type tRNA mutants during kissing-loop formation, the results of which will be discussed in Chapter 3. With these results in hand, we turned our attention to a kissing-loop interaction with a more homogeneous loop sequence, the ColE1 kissing-loop.

2.3 ColE1 KISSING COMPLEX: EXPERIMENTAL DESIGNS AND RESULTS

The ColE1 kissing complex was chosen for a few key reasons: 1) the loop size was smaller, 2) the loop sequence was more homogeneous, 3) with three consecutive stem-loops capable of forming kissing complex, we assumed complex formation would be facile even in U10 mutants were weaker binding, and 4) interest in exploring if triplex insertion

Figure 38: Initial designs of ColE1 kissing complex mutants.
This figure shows the mutations made to the proposed ColE1 kissing complex. A) The proposed wild type structure of the ColE1 kissing-loop from reference 46. Structure of Native sequence. B) Closed loop mutation of stem 1. C) Open loop mutation of stem 1.
would inhibit Rop binding to a kissing complex. Our two initial designs of the ColE1 system is shown in Figure 38. Two designs were considered with mutations taking place in the first stem loop of each construct RI-I and RII-I. The primary difference between the constructs is the loop-stem interface. RNAI and RNAII maintained the native structure in stem 1. RNAI CL an RNAII CL had the ten consecutive U-U mismatches in the stem with two G-C clamps at the stem and loop interface. RNAI OL and RNAII OL also had ten consecutive U-U mismatches but with an A-A mismatch at the loop and stem interface to provide more flexibility in the loop by increasing its size by two nucleotides. The constructs were again synthesized using T7 runoff transcription from PCR amplified templates and purified via gel electrophoresis. Dimerization of these constructs were tested via native PAGE binding assay in 1x TBM2 after incubation in a buffer containing 20 mM Tris-Cl pH = 7.6, 100 mM NaCl, and 10 mM Mg$^{2+}$. Dimerization of these constructs was again inefficient under our assay conditions (Figure 39). We had expected that three potential kissing-loops would be enough for effective complexation, however our designed constructs could not form the desired complexes efficiently. Poor complexation could arise through either improper base pairing or perhaps, without the larger body of the RNA constructs to provide proper spatial orientation, the mutants were not properly oriented to interact. Even with the addition of Rop to the reaction solution, we were not able to observe complete complexation in the gel assay (data not shown). As we were working on this system at the same time as the HIV-DIS system mentioned in the above section, we took the lessons from that successful attempt and redesigned the ColE1 mutants. Simplicity was again adopted and rather than three consecutive stem-loops in our construct, the new
designs would take the form of a single stem-loop similar to those studied in previous reports. Single hairpins from the ColE1 system, specifically stem-loop one, have shown to be efficient at kissing-loop formation and Rop binding. The redesign would
also have the triplex binding site brought up to the loop sequence without a clamp or extra nucleobase added to the loop to influence flexibility. As experiments with the new constructs would involve Rop, we avoided the tRNA scaffold to prevent steric blocking of Rop binding or Rop binding to the intramolecular tRNA loop-loop interface. Finally, RNAI-1’s construct would be clamped with two G-C pairs at the termini to prevent slippage in the structure as the ten U-U mismatches would run into the first two uracil nucleobases in the loop sequence. The newly designed ColE1 mutants are shown in Figure 40 and were chemically synthesized.

![Figure 40: Design of ColE1 stem loop 1 mutants.](image)

This figure shows the single stem-loops designed for the ColE1 mutants. The hairpin design is derived from stem-loop 1 of RNAI and RNAII. RNAI-1U10 has two G-C clamps to ensure proper alignment.

With the new constructs in hand, we began testing them by studying their dimerization by native PAGE assay in 1x TBM2, which contained 10 mM Mg$^{2+}$. Control experiments for non-specific dimerization showed no complexation when RNAI-1U10 was titrated into RNAI-1WT (Supporting information, Figure 60 and 61). Again, simplicity in design seemed to be key as we were able to observe dimerization of the wild type RNA
and U10 triplex hairpin, and the binding of the Rop protein to the mutant kissing complex (Figure 41). Furthermore, we were able to observe dimerization between RNAI-1U10 and RNAII-1U10 along with complexation of the Rop protein to the U10-U10 kissing complex (Figure 42). Control experiments showed that Rop did not bind to U10 mutants alone (Supporting information, Figure 62).

Figure 41: Dimerization of stem-loop 1 mutants of ColE1.
Dimerization and binding of Rop to the ColE1 mutants. RNAI-1WT was held at 1 μM while RNAII-1WT (lanes 2-7) and RNAII-1U10 (lanes 8-13) were titrated in. For Rop complexation experiments, Rop was held at 10 μM.

Figure 42: Complexation of RNAI-1U10 and RNAII-1U10 with Rop.
A) Kissing complex formation between RNAI-1U10 and RNAII-1U10 with RNAI-1U10 held constant at 500 nM and visualized using a Cy5 fluorescence from Cy5-bPNA. B) Rop complex formation between Rop and U10-U10 kissing complex with kissing complex held constant at 4 μM and visualized with Cy5 fluorescence.
Characterization of the various mutant complexes were performed using a gel shift assay (Figure 43 and 44). Literature $K_d$ values for wild type system kissing complex is 200 nM without Rop and 1.5 nM in the presence of 10 μM Rop. We have determined the dissociation constant for the U10-WT kissing complex without Rop to be $3.1 \pm 3.9$ nM at a concentration of 500 nM with a poor fit to our 1:1 binding model using a cyanine labeled bPNA to visualize the gel (Figure 43B and C). This value must be viewed with caution. The gel shows two clear species in dimer region with a shift to one species at high concentrations of WT RNA. We believe these two bands to be caused by the dye projecting towards and away from the loop region (Figure 43A) based on radiolabeled gel-shift assays which shows only one band with unlabeled bPNA (Figure 43 E and F). Attempts were made to determine the apparent $K_d$, however the binding constant determination is also problematic at lower concentrations. At 200 nM and 100 nM, the fitting to the gel is still poor (data not shown). When the concentration of U10 RNA is dropped to 50 nM and the titration experiment is repeated with the cyanine labeled bPNA, the complex does not show a clear shift in the gel image but instead shows a concentration dependent decrease in band mobility even though our previous fitting provides a $K_d$ of 3 nM (Figure 43D). It is possible that at the two different concentrations of our assays, the kinetic parameters of complexation differ due to a change in binding mechanism. For example, at high concentrations, the on and off rate of kissing loop complex formation may be compatible with the timescale of our gel shift assay, but the experiment may be too far above the $K_d$ and provide a poor fitting. Conversely, at low concentrations the mechanism of association might change and the time scale of complex formation may be incompatible with the time
scale of the gel shift assay. Further kinetic and thermodynamic studies must be completed to decipher these results.

However, characterization of the other complexes was successful with our gel shift assay. The WT-U10 (0.33 ± 0.07 nM) and WT-WT (1.5 nM) kissing complex have similar

![Image](image.png)

**Figure 43:** Dissociation constant of RNAII-1U10 and RNAI-1WT.
A) Hypothesized result of binding assay with dye-bPNA producing two complex bands for gel in panel C. B) Fitting of titration experiment at 500 nM in panel C to Equation 1. C) Gel shift of RNAII-1U10 and RNAI-WT visualized with Cy3 fluorescence at 500 nM. D) Gel shift of RNAII-11U10 and RNAI-WT visualized with Cy3 fluorescence at 50 nM. E) Predicted complexation result of binding assay with unlabeled bPNA in panel F. F) Radiolabeled RNAI-1WT with RNAIIU10-(EM)_{10} showing only one complex band.

K_d values in the presence of Rop. Interestingly, the U10-U10 kissing complex formed a tighter kissing complex, with a K_d 4-fold smaller than the wild-type system (40.3 ± 3.1 nM), however Rop did not show any stabilizing affect towards U10-U10 kissing complex formation. The 4-fold lower binding affinity was pleasantly surprising given the significant
structural differences in the triplex RNA compared to the wild type. Rop appears to recognize the shape of the WT-U10 kissing complex and bind but does not change the $K_d$ value of the kissing complex formation. However, our results agree with Eguchi and Tomizawa’s findings that stem sequence does not play a large role in the overall stability of the ColE1 kissing complex.\textsuperscript{46, 48} Our assays yielded dissociation constants for the Rop

**Dissociation constants of ColE1 system**

<table>
<thead>
<tr>
<th>Literature $K_d$ values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_d = 200 \text{nM}$ (1.5 nM with Rop)</td>
</tr>
<tr>
<td>$K_d = 200 \text{nM}$</td>
</tr>
</tbody>
</table>

Kissing complex formation RNA1-IWT and RNAII-IU10

$K_d = 0.33 \pm 0.07 \text{nM}$

Binding of unstructured RNA1-IU10 and RNAII-IWT

$K_d = 227.1 \pm 24.3 \text{nM}$

Rop binding to U10-WT kissing complex

$K_d = 1.2 \pm 0.4 \text{\mu M}$

Rop binding to U10-U10 kissing complex

$K_d = 15.1 \pm 3.1 \text{\mu M}$

**Figure 44: Dissociation constants of ColE1 mutants.**
The figure shows the various dissociation constants for the mutant ColE1 hairpin. Literature values for wild type system were taken from reference 48. The brown bean represents Rop protein. Dissociation constants were obtained via gel shift assay and fitted according to the supporting material.
complex with the mutant kissing-loops U10-WT and U10-U10 to be 1.15 nM and 15.1 μM, respectively. We were also gratified to find that Rop was still able to recognize the single mutant kissing complexes with only a modest decrease (5-fold) in affinity for the WT-U10 complex in spite of the non-native triplex structure’s proximity to the protein binding location. The 80-fold decrease in affinity of Rop for the U10-U10 kissing complex could indicates a distortion in the kissing-loop structure in Rop’s binding footprint. As the affinity of the kissing complex did not change with Rop’s presence, the nature of the interaction could be different than mechanism of interaction observed in the wild type system. Alternatively, the mutant kissing complexes may be in a significantly different conformation than the wild type system, and upon the addition of Rop, the complex may reorganize to accommodate protein binding. In either case, the gel shift assay results coupled with Rop’s inability to alter the stability of U10-U10 kissing complex indicate Rop binds poorly when two mutations are introduced. Further structural analysis of the mutant kissing complexes must be performed to determine how Rop binds to the U10 mutant kissing loops.

We again studied the role of bPNA in the kissing complex formation. Without bPNA to refold the U10 mutant, the affinity of the wild type hairpin for the linear sequence was 227.1 nM which is nearly similar to the wild type kissing complex without Rop present. The results from the above experiments with our mutant ColE1 system gave us confidence that we could use the hairpin for fluorescent assays to detect wild type dimerization and kissing complexes between U10 mutants. A discussion of our fluorescent studies will follow in Chapter 3.
2.4 DISCUSSION AND CONCLUSIONS:

A quick lesson from these series of experiments is that simple designs yield the best answers. Major redesigns were necessary in order to achieve efficient kissing complex formation. Attempts to keep the loop and stem interface more “native” by closing the loop with a G-C pair to separate bPNA binding site from the loop was an inferior design compared to bringing the bPNA binding site directly into the loop.

From our experiments with the redesigned HIV-DIS we can speculate on the perturbations made to the kissing complex structure. The redesigned HIV-DIS system with tRNA scaffold maintained its reliance on divalent cations in spite of its heavily engineered design. This sensitivity to divalent cations indicates that we have maintained, to some extent, the binding pocket for the cation which is characteristic of kissing-loops. We were not surprised to see a 5-fold decrease in affinity upon insertion of the bPNA triplex in one stem of the kissing complex as the mutation could disrupt intermolecular interactions between RNAs. For example, in the NMR structure of HIV-DIS, the G-C pair at the loop-stem junction are not bp’ed but instead engage in non-canonical interactions with the unpaired adenine bases flanking the binding sequence in the loop and we may be deleting these interactions. Similarly, the crystal structure, while having the G-C bp’ed at the terminus of the stem, also has the unpaired adenine bases engaged in non-canonical interactions and insertion of the bPNA triplex may abolish these stabilizing interactions. The bPNA triplex could also cause a misalignment of hydrogen bonds within in the
complex due to the larger size of the triplex. Another possibility is that the triplex may destabilize an already structured hairpin, which would lead to a greater entropic cost to form the kissing complex.

The ColE1 system provides a more complex picture to decipher due to its intricate kinetics and interactions with of Rop. The wild type system uses kinetically unstable kissing-loops leading to a 100-fold difference in dissociation constants when Rop is present. While we could determine $K_d$ for insertion of one bPNA triplex in the presence of Rop, we were unable to determine a binding constant for the same kissing complex without Rop with confidence. In our experiments, a binding shift could be observed at 500 nM but not at 50 nM. This leads us to believe that the $K_d$ would be between 500 nM and 50 nM, and that the kinetics of complex formation may not be as simple as duplex formation causing incompatibility with our assay. Interestingly, when a second triplex is substituted into the system, the complex binding appears to be dramatically affected, as stable complexation is observed without the presence of Rop. The U10-U10 triplex forms stable kissing complexes at low concentrations (40 nM) with higher affinity when compared to the wild type system studied by Tomizawa. The increase in affinity could be caused by the bPNA triplex re-structuring the loop, priming it for complexation. Alternatively, the triplex may be affecting the kinetic parameters of the binding such as in the inverted loop mutant which presented superior affinity when compared the wild type loop. Nature evolved the ColE1 system to favor faster kinetics rather than tighter binding and future studies will have to be completed to see how the triplex mutants alter the kinetics of the system.
Rop has been shown to bind to kissing-loop complexes of various sizes in a sequence independent manner relying more on shape of the kissing complex rather than nucleobase composition.\textsuperscript{49, 86, 97-100} Rop appears able to recognize the kissing complex of both one and two triplex insertions, though binding is much weaker as the number of mutations increases. The fact that Rop recognizes the kissing complexes could indicate the triplex modified kissing complexes can adopt a conformation at the kissing interface similar to the wild type structure. Exactly how Rop affects kissing complex formation is not known for U10-WT kissing complex. Rop does not affect the formation of the U10-U10 kissing complex formation and appears to only weakly bind to this mutant system. However, given the fact that Rop has been shown to interact with the first three bps of the stem immediately after the loop\textsuperscript{46}, binding of the protein with the mutant systems is remarkable. The bPNA triplex structure must be able to achieve a pseudo native-like structure for Rop to bind.

In the experiments presented above, bPNA was shown to promote kissing complex formation in the mutant tRNA system and ColE1 plasmid replication suppression system. The presence of bPNA in the tRNA system improved the stability of the kissing complex by four-fold. In the case of the U10-U10 tRNA kissing-loop, with two less structured 29mer loop, the $K_d$ is only 2-fold higher. On explanation could be the tRNA bodies provide some preorganization to the system, negating the entropic penalty that the less-structured system would pay during complexation without bPNA. However, kissing loops have been shown to be no more stable than a typical duplex, so a dramatic decrease in complex stability should not be expected when bPNA is omitted from the binding assay with mutant
kissing loops. The role of bPNA in the ColE1 system is less clear due to unusual binding observed in our gel assays. While the binding of the wild type hairpin to the unstructured (no bPNA) U10 compliment approaches the literature value of the kissing-loop complex, bPNA’s effect on U10-WT kissing complex formation cannot be discovered until a $K_d$ is determined. What is clear is that the triplex mutants form a tighter complex than the wild type.

With the work presented above, we have been able to show that bPNA-triplexes are capable of promoting kissing-loop complexes. Incorporation of the bPNA-triplexes influences kissing-loop stability in a case dependent manner as the HIV-DIS system shows a weaker complex but the ColE1 system shows a more stable complex. We have also shown that proteins which recognize kissing-loop complexes, such as Rop, are still able to bind to kissing complexes incorporating one or two bPNA triplexes. Additionally, this is the first example in which bPNA triplexes were incorporated into a protein binding sites without abolishing protein recognition. Based on the dependence of $\text{Mg}^{2+}$ for kissing complex formation and Rop recognition, we have shown that the triplex kissing complexes retain at least some wild-type-like structure. Clearly thermodynamic and kinetic experiments must be performed on both kissing-loop systems to fully ascertain the impact of bPNA-triplexes in the context of kissing-loop complexes. However, with this work in hand, we began to ask the question: Can the bPNA triplex be used in directed chemistry within the context of a kissing-loop? Specifically, we wished to know if bPNA, modified with a fluorophore or fluorogen could be used to fluorescently monitor intermolecular
RNA interactions *in vitro*. We began to explore these concepts and our efforts are discussed in Chapter 3.
2.5 SUPPORTING INFORMATION CHAPTER 2

2.5.1 Materials and instrumentation

Chemicals for Fmoc-protected amino acids, peptide, acrylamide solutions and dye synthesis were purchased from Chem-Impex, TCI America, AAPPTec, and VWR and used without further purification. Rink Resin LS (100-200 mesh, 0.28 mmol/g) was purchased from CreoSalus. 40% acrylamide and bis-acrylamide solution (19:1), agarose, and TEMED were purchased from Amerersco and Bio-Rad. Oligonucleotides were purchased from Integrated DNA technologies. Radioactive material was purchased from Perkin Elmer. P30 spin columns were purchased from BioRad. SYBR gold was purchased from Life Technologies. rNTPs were purchased from Sigma Aldrich. T4 PNK kinase was purchased from New England Biolabs. T7 polymerase and Pfu DNA polymerase enzymes were produced in house. Rop was generously donated from the lab of Professor Thomas Magliery and was produced by Anusha Kumar and Kimberly Stephany from The Ohio State University.

The concentration of all DNA/RNA oligomers and synthetic peptides was determined by UV absorbance using a Thermo Scientific Nanodrop 200c spectrophotometer. MALDI-TOF Mass spectra were acquired on Bruker Microflex MALDI-TOF instrument. Electrospray ionization mass spectrometry (ESI) data were acquired on Bruker MicroTOF equipped with an ESI source. NMR spectra were acquired on Bruker Advance DPX 400 instrument. Solid phase peptide synthesis was performed on AAPPTec Apex 396 peptide synthesizer. All peptides were purified on a Hitachi L-7150 pump, Hitachi D-7000 interface and Hitachi L-7400 UV detector on an AAPPTec RPC18
preparative HPLC column and the purity was confirmed on an AAPPTec RP-C18 analytical column. Data and curve fittings were performed on KaleidaGraph 4.0. Gel electrophoresis experiments were at least duplicated, analyzed and results averaged.

**RNA and DNA sequences:**

**Table 2: RNA and DNA templates for kissing-loop forming nucleic acids.**

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD1Nat</td>
<td>5ˈ-tggcgccgaacaggacttgacccctgacccgaccggctttgcgttcgagcaagggcgcgtcttacgactga gctcgggcttccagcaaggccgcgtctctaccgactgagcatccgggcataatcgggtactaatgtagctatatta ttc -3ˈ</td>
</tr>
<tr>
<td>HD2Nat</td>
<td>5ˈ-tggcgccgaacaggacttgacccctgacccgaccggctttgcgttcgagcaagggcgcgtctctaccg aactgagcatccgggcataatcgggtactaatgtagctatatttc -3ˈ</td>
</tr>
<tr>
<td>HD1U10</td>
<td>5ˈ-tggcgccgaacaggacttgacccctgacccgaccggctttgcgttcgagcaagggcgcgtctctaccg aactgagcatccgggcataatcgggtactaatgtagctatatttc -3ˈ</td>
</tr>
<tr>
<td>HD2U10</td>
<td>5ˈ-tggcgccgaacaggacttgacccctgacccgaccggctttgcgttcgagcaagggcgcgtctctaccg aactgagcatccgggcataatcgggtactaatgtagctatatttc -3ˈ</td>
</tr>
<tr>
<td>FP</td>
<td>5ˈ- tggcgccgaacaggac -3ˈ</td>
</tr>
<tr>
<td>RP</td>
<td>5ˈ- gaaattaatacgactcactatagccg -3ˈ</td>
</tr>
<tr>
<td>RNAI-1WT</td>
<td>5ˈ - GCACCGUUGGUAGCGGUGC -3ˈ</td>
</tr>
<tr>
<td>RNAI-1U10</td>
<td>5ˈ - GGUUUUUUUUUUUUUGGUAGUUUUUUUUUUCC -3ˈ</td>
</tr>
<tr>
<td>RNAII-1WT</td>
<td>5ˈ - GCACCGCUACCAACCGGUGC -3ˈ</td>
</tr>
<tr>
<td>RNAII-1U10</td>
<td>5ˈ - UUUUUUUUUUUUCUACCAAUUUUUUUUU -3ˈ</td>
</tr>
</tbody>
</table>

RNA is in capital letters. DNA is in lowercase letters. Underlined bases are T7 primer and not present in final RNA. Loop sequences in DNA are highlighted in bold, red letters.
### Table 3: DNA templates for HIV-DIS mutants

<table>
<thead>
<tr>
<th>HIV DIS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RP</strong></td>
<td>5’-CCCCTCGCCTAAAAAAAAAAAAAAAAAG-3’</td>
</tr>
<tr>
<td><strong>FP</strong></td>
<td>5’-GGAAATTAATACGACTCAGG-3’</td>
</tr>
<tr>
<td><strong>Wild type</strong></td>
<td>5’-CCCCTCGCCTTTGCGGCTTCAGCAAGCCGAGTCCC TATAGTGAGTCGATTAAATTTC-3’</td>
</tr>
<tr>
<td><strong>US</strong></td>
<td>5’-CCCCTCGCCTTAAGGTTGTCGTCGAGTCG-3’</td>
</tr>
<tr>
<td><strong>LS</strong></td>
<td>5’-CCCCAAAAGCGCCTTTGCGGCTTCAGCAAGC CAAATATCCCTATAGTGAGTCGATTAAATTTC-3’</td>
</tr>
<tr>
<td><strong>NC</strong></td>
<td>5’-CCCCTCGCCTTACCAAGGTTGTCGTCGAGTCG-3’</td>
</tr>
<tr>
<td><strong>PC</strong></td>
<td>5’-AACCGCTTACCAAGGTTGTCGTCGAGTCG-3’</td>
</tr>
<tr>
<td><strong>HIV-DIS</strong></td>
<td>Heterodimers</td>
</tr>
<tr>
<td><strong>HD1</strong></td>
<td>5’-CCCCTCGCCTTTGCGGCTTCAGCAAGCCGAGTCCC TATAGTGAGTCGATTAAATTTC-3’</td>
</tr>
<tr>
<td><strong>HD2</strong></td>
<td>5’-CCCCTCGCCTTTGCGGCTTCAGCAAGCCGAGTCCC TATAGTGAGTCGATTAAATTTC-3’</td>
</tr>
<tr>
<td><strong>HD3</strong></td>
<td>5’-CCCCTCGCCTTTGCGGCTTCAGCAAGCCGAGTCCC TATAGTGAGTCGATTAAATTTC-3’</td>
</tr>
<tr>
<td><strong>HD4</strong></td>
<td>5’-CCCCTCGCCTTTGCGGCTTCAGCAAGCCGAGTCCC TATAGTGAGTCGATTAAATTTC-3’</td>
</tr>
<tr>
<td><strong>HD5</strong></td>
<td>5’-CCCCTCGCCTTTGCGGCTTCAGCAAGCCGAGTCCC TATAGTGAGTCGATTAAATTTC-3’</td>
</tr>
<tr>
<td><strong>HD6</strong></td>
<td>5’-CCCCTCGCCTTTGCGGCTTCAGCAAGCCGAGTCCC TATAGTGAGTCGATTAAATTTC-3’</td>
</tr>
<tr>
<td><strong>HD7</strong></td>
<td>5’-CCCCTCGCCTTTGCGGCTTCAGCAAGCCGAGTCCC TATAGTGAGTCGATTAAATTTC-3’</td>
</tr>
<tr>
<td><strong>HD8</strong></td>
<td>5’-CCCCTCGCCTTTGCGGCTTCAGCAAGCCGAGTCCC TATAGTGAGTCGATTAAATTTC-3’</td>
</tr>
<tr>
<td><strong>HD9</strong></td>
<td>5’-CCCCTCGCCTTTGCGGCTTCAGCAAGCCGAGTCCC TATAGTGAGTCGATTAAATTTC-3’</td>
</tr>
<tr>
<td><strong>HD10</strong></td>
<td>5’-CCCCTCGCCTTTGCGGCTTCAGCAAGCCGAGTCCC TATAGTGAGTCGATTAAATTTC-3’</td>
</tr>
<tr>
<td>Table 4: DNA templates for tri-stem-loop ColE1 mutants</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>ColE1</strong></td>
<td></td>
</tr>
<tr>
<td><strong>ColE1 RP</strong> 5’-GAAATTAATACGACTCACTATAGGG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>RNAI Native FP</strong> 5’-AACAAAAAAAACCACCCTAC-3’</td>
<td></td>
</tr>
<tr>
<td><strong>RNAII Native FP</strong> 5’-ACAGTATTTTGTATCTGCGC-3’</td>
<td></td>
</tr>
<tr>
<td><strong>RNAI Native</strong> 5’-AACAAAACCCACCCGCTACCAACGGGTGTTTTTGTGCCGGATCAAGAGCTACCAACTCCTTTTCCGGAAGGTAACTGCGCTTCAGCAAAAGCGCATACCAAAATACTGTCCCTATAGTGAGTCGTATTAAATTCC-3’</td>
<td></td>
</tr>
<tr>
<td><strong>RNAII Native</strong> 5’-ACAGTATTTGTATATCTGCGCTCTGCTGAAGCCAGTTAACTTCGGAAAAAGAGGTGTTAGCTCTTGAATCCGGCGAAACAAAAACCACCATTGGTATCAGGTTTTTTTGGTCTACTGTCCCTATAGTGAGTCGTATTAAATTCC-3’</td>
<td></td>
</tr>
<tr>
<td><strong>RNAI CL FP</strong> 5’-AACAGTATTTTGTTATCTGCGC-3’</td>
<td></td>
</tr>
<tr>
<td><strong>RNAII CL FP</strong> 5’-ACAGTATTTTGTTATCTGCGC-3’</td>
<td></td>
</tr>
<tr>
<td><strong>RNAI CL</strong> 5’-AACAAAAACCCACCCGCTACCAACGAAGGTTTTTGTGCCGGATCAAGAGCTACCAACTCCTTTTCCGGAAGGTAACTGCGCTTCAGCAAAAGCGCATACCAAAATACTGTCCCTATAGTGAGTCGTATTAAATTCC-3’</td>
<td></td>
</tr>
<tr>
<td><strong>RNAII CL</strong> 5’-ACAGTATTTTGTTATCTGCGCTCTGCTGAAGCCAGTTAACTTCGGAAAAAGAGGTGTTAGCTCTTGAATCCGGCGAAACAAAAACCACCATTGGTATCAGGTTTTTTTGGTCTACTGTCCCTATAGTGAGTCGTATTAAATTCC-3’</td>
<td></td>
</tr>
<tr>
<td><strong>RNAI OL FP</strong> 5’-AACGAAAAAATCTACCAATAAACAAAAACGC-3’</td>
<td></td>
</tr>
<tr>
<td><strong>RNAI OL RP</strong> 5’-GAAATTAATACGACTCACTATAGGGACAGTAT-3’</td>
<td></td>
</tr>
<tr>
<td><strong>RNAI OL</strong> 5’-AACGAAAAAAATCTACCAATAAACAAAAACGCAGCGATACCAACTCCTTTTCCGGAAGGTAACTGCTTGTCAGCAAAAGCGCATACCAAAATACTGTCCCTATAGTGAGTCGTATTAAATTCC-3’</td>
<td></td>
</tr>
<tr>
<td><strong>RNAII OL</strong> 5’-ACAGTATTTTGTTATCTGCGCTCTGCTGAAGCCAGTTAACTTCGGAAAAAGAGGTGTTAGCTCTTGAATCCGGCGAAACAAAAACGTTTTTGTCTACTGTCCCTATAGTGAGTCGTATTAAATTCC-3’</td>
<td></td>
</tr>
</tbody>
</table>

FP and RP are forward primer and reverse primer, respectively. The underlined portion of each template is the T7 promoter region and does not appear in the transcript RNA.
2.5.2 Methods and Procedures.

DNA amplification.

DNA was used as received and PCR amplified using a Personal Eppendorf Mastercycler. Pfu PCR 10x buffer contained 200 mM Tris-Cl pH 8.8, 100 mM KCl, 100 mM (NH₄)₂SO₄, 20 mM MgSO₄, 1 mg/ml nuclease free BSA, 1% v/v triton X-100. PCR was performed using the following components reported as final concentrations: 1x of Pfu PCR buffer, 500 nM forward and reverse primer, 6 nM DNA template, 200 μM dNTP, 1.35 μl of stock Pfu DNA polymerase per 100 μl reaction solution. Thermocycling was performed as follows: Denature (95°C, 2 min), 25x [Denature (95°C, 30 s), Anneal (55°C, 30 s), Elongation (72°C, 60 s)], Polish (72°C, 4 min), and hold at 4°C. DNA was purified using a Qiagen MinElute PCR purification kit and eluted from purification column with water.

RNA transcription.

tRNAs were made using T7 runoff transcription by T7 RNA polymerase. RNA transcription buffer 10x: 1M HEPES-KOH pH 7.5, 100 mM MgCl₂, 20 mM Spermidine-HCl, 400 mM DTT. Transcription was performed in the following concentrations: 1x Buffer with the addition of 10 mM DTT, 20 mM MgCl₂, 20 mM rNTP, 7.5% v/v glycerol, 350 nM DNA template from PCR, and 1.9 μl of T7 polymerase stock (100000 units/mL) (per 100 μl of transcription reaction solution). Transcription took place at 37°C for 3-4 h at which point 1.5 equivalents of EDTA was added to quench the reaction and 2x TBE/urea loading buffer was added. Samples were heated at 95°C for 10 min, cooled to room
temperature, centrifuged at 13200 pm, and loaded onto 15% w/v acrylamide (19:1 acrylamide:bisacrylamide) denaturing gel (8.3 cm X 7.3 cm X 1.5 mm) with 8M urea as the denaturant. The denaturing gel was electrophoresed for 3 h. at 250 volts. The RNA was visualized using UV shadowing, excised from the gel, the gel piece pushed through the end of a 1 ml syringe, and soaked in water overnight at room temperature with agitation. The sample was then centrifuged and the supernatant removed, discarding the gel pieces. The RNA was precipitated from the supernatant by the addition of 5 M NH₄OAc pH 5.2 to a concentration of 450 mM and 2.5 volume equivalents of 200-proof ethanol followed by incubation at -20° C overnight and centrifugation at 13200 rpm for 30 min at 4° C. The RNA pellet was washed with 150 μl of ice-cold 30% v/v water in ethanol and centrifuged again at high speed for 10 min at 4° C. RNA pellets were then dried in a speed-vac at room temperature for 20 min and suspended in filtered water and stored in a freezer at -20° C. Stock concentrations were determined by UV-vis spectroscopy, measuring absorbance at 260 nm. RNA purity was checked on a 15% w/v denaturing gel as described above except gels were electrophoresed at 250 V for 1 h. and visualized using SYBR gold staining. To obtain radiolabeled RNAs, 5μl of 3000 uCi/mmol Alpha-[P³²]-GTP was incorporated during transcription and normal procedure was followed.

**5’-end labeling of ColE1 RNA.**

End labeling was performed as described in the protocol provided by New England Biolabs. 400 pmoles of RNA, 40 pmoles of [γ⁻³²P] ATP (250 μCi, 3000 Ci/mmol 10 mCi/ml), 760 pmole ATP, and 5 μl of T4 PNK kinase (10,000 units/ml) were mixed in 40
μl of 1X T4 Polynucleotide Kinase Reaction Buffer (70 mM Tris-HCl pH 7.6, 10 mM MgCl2, 5 mM DTT). Reaction was incubated for 1.5 h at 37°C and then heated at 72°C for 20 min. Excess radiolabel was removed using a P30 spin column. RNA’s were purified on a denaturing 20% acrylamide gel (19:1) with urea in 1X TBE. RNAs were excised from the gel, the gel slab pushed through a syringe to crush it, and soaked overnight in water at room temperature. The soaked gel was then centrifuged and the supernatant removed, discarding the gel pieces. The RNA was precipitated by the addition of 5 M NH4OAc pH 5.2 to a concentration of 450 mM and 2.5 volume equivalents of 200-proof ethanol followed by incubation at -20°C overnight and centrifugation at 13200 rpm for 30 min at 4°C. The RNA pellet was washed with 150 μl of ice-cold 30% v/v water in ethanol and centrifuged again at high speed for 10 min at 4°C. RNA pellets were then dried at room temperature for 20 min and suspended in water and stored at -20°C. Stock concentrations were determined by UV-vis spectroscopy, measuring absorbance at 260 nm.

tRNA EMSA gel assay.

A parent RNA stock solution was diluted into 1x kissing buffer (20 mM Tris-Cl pH 7.6, 10 mM MgCl2, and 100 mM NaCl), glycerol (final concentration 10% v/v), and, if necessary, bPNA to a final RNA concentration 3-fold greater than desired for the gel assay. A separate stock was created diluting the complementary RNA in 1x buffer, with or without bPNA, to concentrations appropriate for titration. Stock solutions were annealed at 90°C for 5 min and then snap cooled for 5 min on ice. 10 μl of parent RNA stock was added to each tube and the complementary RNA was added in increasing
concentrations using 1x kissing buffer to bring the final reaction solution to 30 μl. RNAs were then mixed and incubated for 30 min on ice prior to loading. The kissing complex was visualized on a 2.5% w/v agarose gel in 1x TBM1 (89 mM Tris, 89 mM boric acid, and 0.1 mM MgCl₂). Gels were electrophoresed at 120 volts for 2 h at 4 °C. Gels were imaged on a GE Healthcare Typhoon FLA 9500 using the phosphorimager or cyanine dye fluorescence. Alternatively, a 4-15% gradient acrylamide gel (70:1) was used electrophoresing at 120 V for 3 h. Gels with intrinsic P³²-labeling were dried using a temperature gradient for 2.5 h. holding at 65°C. Gels were quantified using ImageQuant 5.0 and binding curves fitted using KaleidaGraph 4.0 using a 1:1 binding model using Equation 1. For Rop binding, [RNA₂] was replaced by [Rop] in Equation 1. Experiments in most cases were performed in triplicate and results after work up were averaged.

Equation 1:

\[
[\text{complex}] = \frac{K_d + [RNA_1] + [RNA_2]}{2} - \left(\sqrt{(K_d + [RNA_1] + [RNA_2])} - (4 * [RNA_1] * [RNA_2])\right)/2
\]

CoIE1 EMSA gel assay.

Samples were prepared in the same manner as described above in section tRNA EMSA gel assay. Assays ran in the presence of Rop used 10x Rop Buffer: 200 mM Tris-Cl pH 7.6, 3M NaCl, 100 mM MgCl₂, 50 mM DTT, which was diluted to 1x in assay. Rop was added as necessary after RNA was annealed to the Parent RNA stock solution. Rop concentrations are reported as monomer concentration unless otherwise stated. After incubation on ice for 30 min, samples were loaded onto a 15% w/v acrylamide gel (8.3 cm
X 7.3 cm X 1.5 mm) composed of 70:1 acrylamide to bis-acrylamide which had be pre-run for 2 h at 120 volts. Gels were electrophoresed for 4-6 h at 120 volts in 1x TBM2 (89 mM Tris, 89 mM boric acid, and 10 mM MgCl₂). Experiments to determine binding affinity were performed with a cyanine dye labeled bPNA or radiolabeled RNA and directly imaged as described in tRNA EMSA gel assay. If unlabeled bPNA was used, RNA was stained with SYBR gold and then imaged as described in the section tRNA EMSA gel assay. Gels were quantified using ImageQuant 5.0 and binding curves fitted in KaleidaGraph 4.0 using a 1:1 binding model using Equation 1. Experiments were performed in triplicate and results after work up were averaged.

2.5.3 Synthetic Procedures and compound characterization.

Dye Synthesis: Dye synthesis was completed by modifying a previously reported procedure.¹⁰²

Synthesis of 1-(5-carboxypentyl)-2,3,3-trimethyl-3H-indol-1-ium bromide (1): In a round bottom flask was added 18.8 mmol (3 g) of 1,3,3-Trimethyl-2-methyleneindoline in 10 ml of nitromethane. To this was added 21 mmol (4.13 g) of 6-bromohexanoic acid and
the solution was refluxed for 36 h. Upon completion, the reaction was cooled to room temperature and slowly added to 150 ml of ethyl acetate with vigorous stirring. The product was collected by vacuum filtration and washed with 3 x 50 ml of ethyl acetate to yield 15.3 mmol (5.39 grams, 81% yield) of 1 as a purple powder. Calculated [M+]: 274.1802 Found: 274.1803

Figure 45: Proton NMR of compound 1. 
$^1$H NMR (400 MHz, DMSO-D6) of 1-(5-carboxypentyl)-2,3,3-trimethyl-3H-indol-1-ium bromide (1)

Synthesis of Cy3 1-(5-carboxypentyl)-3,3-dimethyl-2-((E)-3-((Z)-1,3,3-trimethylindolin-2-ylidene)prop-1-en-1-yl)-3H-indol-1-ium (2): To a round bottom flask was added 0.5 grams of 1 (1.4 mmol), 340 mg of $N,N'$-diphenylformamidine, and 7.5 ml of acetic anhydride. This solution was heated at 120°C for 0.5 h. The solution was then
cooled to room temperature followed by the addition of 310 mg of 1,3,3-Trimethyl-2-methyleneindoline and 7.5 ml of pyridine. The solution was stirred overnight at room temperature. The following day, 200 ml of ethyl ether was added and the product precipitated overnight at -20°C. The precipitate was collected by centrifugation at 3500 rpm for 10 min, dried with a nitrogen stream, and purified by flash column chromatography using 100% EtOAc, then a gradient of methanol/dichloromethane gradient of 2%-5%-7% yielding 358 mg of 2 as a dark red foam. \( R_f = 0.3 \) (10% MeOH in DCM). Calculated [M+] 457.2850 Found: 457.2853

Figure 46: Proton NMR of Cy3.
\(^1\)H NMR (400 MHz, CDCl3) Cy3 1-(5-carboxypentyl)-3,3-dimethyl-2-((E)-3-((Z)-1,3,3- trimethylindolin-2-ylidene)prop-1-en-1-yl)-3H-indol-1-ium (2)
Synthesis of Cy5 1-(5-carboxypentyl)-3,3-dimethyl-2-((1E,3Z)-5-((E)-1,3,3-trimethylindolin-2-ylidene)penta-1,3-dien-1-yl)-3H-indol-1-ium (3): To a round bottom flask was added 0.5 grams of 1 (1.4 mmol), 450 mg of N-(3-(phenylamino)-2-propenylidene)aniline hydrochloride, and 7.5 ml of acetic anhydride. This solution was heated at 120°C for 0.5 h. The solution was then cooled to room temperature followed by the addition of 310 mg of 1,3,3-Trimethyl-2-methyleneindoline and 7.5 ml of pyridine. The solution was stirred overnight at room temperature. The following day, 200 ml of diethyl ether was added and the product precipitated overnight at -20°C. The precipitate was collected by centrifugation, dried with a nitrogen stream, and purified by flash column chromatography using 100% EtOAc, then a gradient of methanol/dichloromethane gradient of 2%-5%-7% yielding 413 mg of 3 as a dark blue foam. R_f = 0.3 (10% MeOH in DCM). Calculated [M+]: 483.3006 Found: 483.3002
Figure 47: Proton NMR of Cy5.
$^1$H NMR (400 MHz, CDCl3) Cy5 1-(5-carboxypentyl)-3,3-dimethyl-2-((1E,3Z)-5-((E)-1,3,3-trimethylindolin-2-ylidene)penta-1,3-dien-1-yl)-3H-indol-1-ium (3)

Peptide Synthesis: Fmoc-Lys(Mel)-OH and NH$_2$-(EM)$_{10}$G-CONH$_2$ was synthesized as previously reported.$^{90}$ Peptides were synthesized on an AAPPTec Apex 396 peptide synthesizer using standard solid phase Fmoc chemistry on Rink Resin LS with 0.28 mmol/g loading. 0.25 M of Amino acids were coupled with 0.25 M of HOBr and 0.25 M DIC in 2 ml DMF. The coupling reactions were mixed for 1.5 h at which point they were rinsed with 3 x 2 ml DMF. Fmoc protecting groups were cleaved by subjecting resin to 2 x 2 ml of 20% v/v piperidine solution in DMF for ten min followed by another rinse cycle. Cyanine dyes were coupled using three equivalents of dye, 3.3 equivalents of
HBTU, and 3.3 equivalents of DIEA in 2 ml DMF. Dye and coupling reagents were allowed to react for fifteen min before addition to resin. Peptides were cleaved from the resin using 95:5 v/v TFA:water solution for 4 h (beads will lose dye color due to solvatochromic effect, but dye does survive cleavage). Cleaved peptide solutions were filtered through cotton to remove the beads, peptide crashed out in cold diethyl ether, centrifuged to collect the solid peptide, washed 3 times with cold ether, and dried using a nitrogen gas stream to remove excess ether and TFA. Peptides were purified by HPLC with a preparatory C18 column and water/acetonitrile gradient containing 0.1% v/v TFA. The purification used solvent A (water, 0.1% v/v TFA) and solvent B (80% v/v acetonitrile, 0.1% v/v TFA) gradient as follows: 0-5 min 10% B, 5-20 min increase to 40% B, 40% B for 20 min, then 100% B for 5 min. Cyanine dye conjugated peptide purity was checked on an analytical C18 HPLC column using an isocratic gradient starting at 10% B for 0-5 min, 5-20 min reaching 35% B, holding at 35% solvent B for 20 min, and 100% wash.
Figure 48: HPLC trace and MALDI-TOF of Cy5-(EM)$_{10}$.
A) HPLC trace of Cy5(EM)$_{10}$G-CONH$_2$. Cy5(EM)$_{10}$G-CONH$_2$ on a RP-C18 analytical column monitored with UV-Vis detector at 220 nm. B) MALDI-TOF spectra of Cy5(EM)$_{10}$G-CONH$_2$: Calculated [M+]: 4201.102 Found: 4198.010.
Figure 49: HPLC trace and MALDI-TOF of Cy3-(EM)$_{10}$.  
A) HPLC trace of Cy3(EM)$_{10}$G-CONH$_2$. Cy3(EM)$_{10}$G-CONH$_2$ on a RP-C18 analytical column monitored with UV-Vis detector at 220 nm. B) MALDI-TOF spectra of Cy3(EM)$_{10}$G-CONH$_2$: Calculated [M+]: 4175.0864 Found: 4172.451
2.5.4 tRNA mutant EMSA gel results and dissociation constant determination:

Figure 50: Negative control for non-specific dimerization of HD1U10 and HD1Nat. The gel above is the negative control agarose gel for tRNA kissing. Imaged by Cy5 fluorescence. HD1U10-Cy5(EM)10 complex is held constant at 200 nM while HD1Nat is titrating in with a range 0-1000 nM.

Figure 51: Negative control for non-specific dimerization of HD2U10 and HD2Nat. The gel above is the negative control agarose gel for tRNA kissing. Imaged by Cy5 fluorescence. HD2U10-Cy5(EM)10 complex is held constant at 200 nM while HD2Nat is titrating in with a range 0-1000 nM.
Figure 52: Mg\textsuperscript{2+} dependence of HD1U10-bPNA triplex and HD2Nat complexation. The gel above displays kissing complex formation of HD1U10-Cy5(EM)\textsubscript{10} triplex and HD2Nat without Mg\textsuperscript{2+} in the reaction buffer or gel running buffer. HD2Nat was titrated from 0-1000 nM into a HD1U10-Cy5(EM)\textsubscript{10} held constant at 200 nM. Gel ran in TBE (89 mM Tris, 89 mM boric acid, and 2 mM EDTA). Imaged by Cy5 fluorescence. Lower band is free HD1U10. Upper band is kissing complex.

Figure 53: Mg\textsuperscript{2+} dependence of HD1U10-bPNA triplex and HD2U10-bPNA triplex complexation. The gel above looks at kissing complex formation of HD1U10-Cy5(EM)\textsubscript{10} triplex and HD2U10-(EM)\textsubscript{10} without Mg\textsuperscript{2+} in the reaction buffer or gel running buffer. HD2U10-(EM)\textsubscript{10} was titrated from 0-2000 nM into a HD1U10-Cy5(EM)\textsubscript{10} held constant at 500 nM. Gel ran in TBE (89 mM Tris, 89 mM boric acid, and 2 mM EDTA). Imaged by Cy5 fluorescence.
**A)**

Kissing complex
HD1Nat

Average $K_d = 9.3 \pm 2.0$ nM

**B)**

Dissociation constant of HD1Nat and HD2U10-Cy5EM10

Figure 54: $K_d$ determination of HD1Nat and HD2U10-bPNA triplex.

A) Gel image of HD1Nat and HD2U10-bPNA kissing complex formation.  B) The graph above represents the $K_d$ determination of HD1Nat with HD2U10 triplex (bound to (EM)10). This was determined by gel shift assay with radiolabeled HD1Nat held constant at 50 nM with HD2U10 bound to bPNA (structured anticodon loop) titrated in from 0-800 nM.
A) Kissing complex
HD1U10-bPNA

Average $K_d = 18.5 \pm 3.2 \text{ nM}$

B) Dissociation constant of HD1U10-EM10 and HD2U10-EM10

Figure 55: $K_d$ determination of HD1U10-EM10 and HD2U10-EM10.
A) Gel image of HD1U10-bPNA and HD2U10-bPNA kissing complex formation. B) The graph above represents the $K_d$ determination of HD1U10 triplex (bound to (EM)$_{10}$) with HD2U10 triplex (bound to (EM)$_{10}$). This was determined by gel shift assay with radiolabeled HD1Nat held constant at 50 nM with HD2U10 bound to bPNA (structured anticodon loop) titrated in from 0-800 nM.
A) Gel image of HD1Nat and HD2Nat kissing complex formation. Average $K_d = 2.4 \pm 0.6$ nM

B) The graph above shows the $K_d$ determination of HD1Nat and HD2Nat. Gel experiments using a radiolabeled HD1Nat held constant at 20 nM and titrating in HD2Nat from 0-320 nM.

**Figure 56:** $K_d$ determination of HD1Nat and HD2Nat.
A) Gel image of HD1Nat and HD2Nat kissing complex formation. B) The graph above shows the $K_d$ determination of HD1Nat and HD2Nat. Gel experiments using a radiolabeled HD1Nat held constant at 20 nM and titrating in HD2Nat from 0-320 nM.
A) Kissing complex

**HD1Nat**

Average $K_d = 30.3 \pm 10.1$ nM

B) Dissociation constant for HD1Nat and HD2U10 without bPNA

![Graphs showing $K_d$ determination](image)

- $K_d = 19.5$ nM, $R = 0.99583$
- $K_d = 27.5$ nM, $R = 0.99099$
- $K_d = 43.8$ nM, $R = 0.99425$

**Figure 57: $K_d$ determination of HD1Nat and unstructured HD2U10.**

A) Gel image of HD1Nat and HD2U10 kissing complex formation. B) The graph above represents the $K_d$ determination of HD1Nat with HD2U10 (no bPNA). This was determined by gel shift assay with radiolabeled HD1Nat held constant at 50 nM with HD2U10 without bPNA (unstructured anticodon loop) titrated in from 0-800 nM.
A) Gel image of HD1U10 and HD2U10 without bPNA kissing complex formation.  B) The graph above represents the $K_d$ determination of unstructured HD1U10 with unstructured HD2U10 (no bPNA). This was determined by gel shift assay with radiolabeled HD1U10 held constant at 1 μM with HD2U10 titrated in from 0-6 μM, both without bPNA (unstructured anticodon loop). Raw data does not show a complete gel shift.

Figure 58: $K_d$ determination unstructured HD1U10 and unstructured HD2U10.  A) Gel image of HD1U10 and HD2U10 without bPNA kissing complex formation.  B) The graph above represents the $K_d$ determination of unstructured HD1U10 with unstructured HD2U10 (no bPNA). This was determined by gel shift assay with radiolabeled HD1U10 held constant at 1 μM with HD2U10 titrated in from 0-6 μM, both without bPNA (unstructured anticodon loop). Raw data does not show a complete gel shift.
A) Gel image of HD1U10 probe and HD2U10 probe without bPNA kissing complex formation.  

B) The graph above represents the K_d determination of unstructured HD1U10 probe with HD2U10 probe (no bPNA). This was determined by gel shift assay with radiolabeled HD1U10 probe held constant at 50 nM with HD2U10 titrated in from 0-800 nM, both without bPNA (unstructured anticodon loop).

**Figure 59: Kd determination unstructured HD1U10 probe and unstructured HD2U10 probe**

A) Gel image of HD1U10 probe and HD2U10 probe without bPNA kissing complex formation.  

B) The graph above represents the K_d determination of unstructured HD1U10 probe with HD2U10 probe (no bPNA). This was determined by gel shift assay with radiolabeled HD1U10 probe held constant at 50 nM with HD2U10 titrated in from 0-800 nM, both without bPNA (unstructured anticodon loop).
2.5.5 ColE1 mutant EMSA gel results and dissociation constant determination:

Figure 60: Control for non-specific dimerization of RNAI.
RNAI-1U10-Cy5EM10 complex is held constant at 500 nM while RNAI-1WT is titrated in with a range 0-8000 nM. Imaged by Cy5 fluorescence.

Figure 61: Control for non-specific dimerization of RNAII.
RNAII-1U10-Cy5EM10 complex is held constant at 500 nM while RNAII-1WT is titrated in with a range 0-8000 nM. Imaged by Cy5 fluorescence.

Figure 62: Control for Rop binding to bPNA triplex.
RNAII-1U10-Cy5EM10 complex is held constant at 100 nM while Rop is titrated in with a range 0-16000 nM. Imaged by Cy5 fluorescence.
A) The gel above shows the gel shift of RNAII-1U10-Cy5EM10 upon binding to RNAI-1WT in the presence of Rop. RNAII-1U10-Cy5EM10 is held constant at 10 nM, Rop is held constant at 10 μM, and RNAI is titrating in with a range of 0-160 nM. B) The curve fitting for the gel shift fitted to Equation 1. Lower band is RNAII-1U10. Upper band is Rop complex.

Figure 63: Rop complex of RNAII-1U10 and RNAI-1WT.
A) **Kissing complex**

RNAII-1U10-bPNA

Average $K_d = 40.3 \pm 3.1$ nM

B) **Dissociation constant of RNAI-1U10 and RNAII-1U10 with Cy5EM10**

- **Figure 64:** Kissing complex of RNAI-1U10 and RNAII-1U10 without Rop.
  
  A) The gel above shows the gel shift of RNAII-1U10-Cy5EM10 upon binding to RNAI-1U10-Cy5EM10 without Rop. Radiolabeled RNAII-1U10-Cy5EM10 is held constant at 40 nM and RNAI is titrating in with a range of 0-640 nM. B) The curve fitting for the gel shift fitted to Equation 1. Bottom band is RNAII-1U10. Top band is the kissing complex.
A) The gel above shows the gel shift of RNAII-1U10-Cy5EM10 upon binding to RNAI-1U10-Cy5M10 with 20 μM Rop. RNAI-1U10-Cy5EM10 is held constant at 40 nM and RNAI is titrating in with a range of 0-640 nM. B) The curve fitting for the gel shift fitted to Equation 1. Bottom band is RNAII-1U10. Top band is the Rop complex.

Figure 65: Kissing complex of RNAI-1U10 and RNAII-1U10 with Rop.
A) Average $K_d = 196.7 + 54.5 \text{ nM}$

B) Dissociation constant of unstructured RNAII-1U10 and RNAI-1WT with 10 μM Rop

- $K_d = 254.9 \text{ nM}$  
  $R = 0.98234$
- $K_d = 123.9 \text{ nM}$  
  $R = 0.98888$
- $K_d = 211.2 \text{ nM}$  
  $R = 0.97512$

**Figure 66: Complex formation of unstructured RNAII-1U10 and RNAI-1WT with Rop.**  
A) The gel above shows the gel shift of RNAII-1U10 upon binding to RNAI-1WT in the presence of Rop. RNAII-1U10 is held constant at 500 nM, Rop is held constant at 10 μM, and RNAI is titrating in with a range of 0-2500 nM. B) The curve fitting for the gel shift fitted to Equation 1. Complex fluorescence was corrected for SYBR gold turn on by multiplying complex intensity by 0.58.
A) Average $K_d = 227.1 \pm 24.3 \text{ nM}$

B) Dissociation constant of unstructured RNAII-1U10 and RNA1-WT

![Graphs showing complex formation and dissociation constants](image)

**Figure 67: Complex formation of unstructured RNAII-1U10 and RNA1-WT without Rop.**

A) The gel above shows the gel shift of RNAII-1U10 upon binding to RNAI-WT. RNAII-1U10 is held constant at 500 nM and RNAI is titrating in with a range of 0-2500 nM. B) The curve fitting for the gel shift fitted to Equation 1. Complex fluorescence was corrected for SYBR gold turn on by multiplying complex intensity by 0.8.
A) The gel above shows the gel shift of kissing complex between RNAII-1U10-Cy5EM10 and RNAI-1WT binding to Rop. The kissing complex is held constant at 1 μM and Rop is titrated in with a range of 0-6 μM. B) The curve fitting for the gel shift fitted to Equation 1.

Average $K_d = 1.2 + 0.4 \mu M$

B)

**Dissociation constant of Rop and WT-U10 kissing complex**

![Graphs showing dissociation constants](image)

- $K_d = 1782.0 \text{ nM}$  
  $R = 0.98576$
- $K_d = 860.7 \text{ nM}$  
  $R = 0.9949$
- $K_d = 976.3 \text{ nM}$  
  $R = 0.99396$

**Figure 68: Rop complex formation with preformed U10-WT kissing-loop.**

A) The gel above shows the gel shift of kissing complex between RNAII-1U10-Cy5EM10 and RNAI-1WT binding to Rop. The kissing complex is held constant at 1 μM and Rop is titrated in with a range of 0-6 μM. B) The curve fitting for the gel shift fitted to Equation 1.
Average $K_d = 15.1 \pm 3.1 \mu$M

**B)**

Dissociation constant of Rop binding to preformed kissing complex triplex to triplex

![Graphs showing dissociation constant](image)

**Figure 69: Rop complex formation with preformed U10-U10 kissing-loop.**
A) The gel above shows the gel shift of kissing complex between RNAI-1U10-Cy3EM10 and RNAII-1U10-EM10 binding to Rop. The kissing complex is held constant at 4 μM and Rop is titrated in with a range of 0-20 μM. 
B) The curve fitting for the gel shift fitted to Equation 1. $K_d$ is average of two independent experiments and the above is a representative plot.
2.6 REFERENCES CHAPTER 2


58. Muriaux, D.; Girard, P.-M.; Bonnet-Mathonière, B.; Paoletti, J., Dimerization of HIV-1Lai RNA at Low Ionic Strength AN AUTOCOMPLEMENTARY SEQUENCE IN


73. Mundigala, H.; Michaux, J. B.; Feig, A. L.; Ennifar, E.; Rueda, D., HIV-1 DIS stem loop forms an obligatory bent kissing intermediate in the dimerization pathway. Nucleic Acids Research 2014, 42 (11), 7281-7289.


85. Helmer-Citterich, M.; Anceschi, M. M.; Banner, D. W.; Cesareni, G., Control of ColE1 replication: low affinity specific binding of Rop (Rom) to RNAI and RNAII. The EMBO Journal 1988, 7 (2), 557-566.


Chapter 3: Directed Chemistry with bPNA

3.1 INTRODUCTION TO FLUORESCENTLY LABELING RNA:

Key advances in the understanding of cellular processes on a molecular level have been made possible through the plethora of biological and chemical tools available to researchers. Through the ligation of chemical probes, researchers have been able study molecules of interest to determine their production, function, and life time. Interactions of molecules can be studied \textit{in vitro} and \textit{in vivo} through fluorescent techniques which allows the observation of specific interactions, probe binding sites, and track the transit of materials in a cell. Fluorescent experiments require various labeling techniques in order to specifically tag the molecule of interest. A vast amount of resources has been devoted to development of methodology in order to fluorescently label a target either covalently or non-covalently. Below is a discussion, with a few examples, of various techniques for labeling of RNA and DNA through both chemical and genetic tagging.

The simplest technique involves covalent attachment of a fluorescent molecule to a nucleic acid during solid phase synthesis and enzymatically catalyzed post synthetic modifications. The advantages of these strategies include: 1) site-specific labeling, 2)
enzymes can target a specific molecule in complex mixtures, 3) only the nucleic acid introduced into the system will be tagged, 4) immediate fluorescence without need for binding event/turn-on event, 5) tag is covalently bound and will not dissociate, and 6) can image low-abundance nucleic acids. The disadvantages associated with covalent labeling include: 1) size limitations achievable via solid-phase synthesis limits targets choices, 2) enzymes and substrates must be engineered and not every position on a nucleic acid has an enzyme that can modify it, 3) nucleic acid degradation during handling, 4) synthetic nucleic acids may lack nuclear factors important for localization, 5) nucleic acids may need to be injected into cells which perturbs the natural physiological state, and 6) probes are at higher concentrations than physiologically relevant. Fluorescent tags can be introduced via solid phase synthesis by incorporation of chemically modified nucleobases or fluorescent nucleobase mimics. Constructs can be labeled internally with a modified base\textsuperscript{1-4} or on the termini\textsuperscript{5-6} through standard phosphoramidite chemistry or click chemistry. Various groups have also been able to introduce fluorescent tags into the polymeric backbone through modifications to the phosphodiester\textsuperscript{7} and ribose moieties\textsuperscript{8-10} or other unique scaffolds.\textsuperscript{10-11} Ranjit and coworkers went so far as to directly incorporate a cyanine dye directly into the backbone.\textsuperscript{12} The Kool lab used a templated reaction to catalyze covalent attachment of two nucleic acid probes which upon reaction released a quencher molecule to turn-on the probe’s fluorescence.\textsuperscript{13}

For constructs that cannot be synthesized chemically, some researchers have been able to create nucleotides which are site selectively incorporated by naturally occurring enzymes or mutant enzymes which can site selectively alkylate nucleobases. Native
enzymes which selectively alkylate specific nucleobases\textsuperscript{14}, such as trimethylguanosinesynthases\textsuperscript{15} and tRNA\textsuperscript{\textit{Ile}}\textsuperscript{2}-agmatidine synthetase\textsuperscript{16}, or mutated enzymes, such as methyltransferase\textsuperscript{17}, have been used to install reactive groups for further modifications with fluorophores via biorthogonal reactions. Other groups have designed unnatural nucleotide triphosphates, modified typically with click reactive moieties, which are permittable for incorporation by polymerases. T7 RNA polymerase has been found to incorporate alkyne-modified dinucleotides\textsuperscript{18}, norbornene guanosine\textsuperscript{19}, 2′ and 3′ azide substituted ribose\textsuperscript{20}, and furan-modified uracil.\textsuperscript{21} Incorporation of unnatural bps designed to expand the genetic alphabet can also be modified for later click chemistry reactions to attach reporting molecules.\textsuperscript{22} Riedl and coworkers were able to polymerize a DFHBI-modified cytidine (DFHBI is a small molecule mimic of the green fluorescent protein chromophore\textsuperscript{23}) using KOD XL DNA polymerase to detect protein binding.\textsuperscript{24}

Site-specific incorporation of a fluorescent reporter on a synthetic nucleic acid allows for the study of injected/transfected species but does not allow observation of RNA or DNA produced within a cell. To study the transient production of specific sequences of RNA and DNA, chemists have developed reporting molecules which bind to their target. These binding probes come in two forms: one being hybridization probes which bind specific sequences and the other being small molecules which recognize a specific sequence or structure. Hybridization probes, composed of fluorescently labeled DNA, RNA, PNA, or LNA, form duplexes or triplexes with their targets. The advantages of these probes include: 1) no cellular or enzymatic engineering, 2) small size of probes which are less sterically demanding, 3) high-multiplexing potential to reduce background noise, and
4) sorting of cells based on mRNA expression. As with any method, there are disadvantages to this strategy such as: 1) prior knowledge of target sequence and structure, 2) may need to microinject probes, 3) low sensitivity, 4) degradation of probe within cell, 5) accumulation of probes in nucleus, 6) hybridization delay, and 7) non-specific binding cannot be completely avoided. DNA probes, which are stable in the cellular environment, have been adorned with fluorogens such as thiazole orange which become brighter upon binding to target due to intercalation of the dye.\textsuperscript{2, 25} The use of fluorogens improves background noise as only hybridized probes are “turned on” or able to emit light due to an interaction, such as intercalation, with the target. DNA probes to study oligo(dA) have been used with caged fluorochrome reporters which emit only after a stimulant, such as light, uncages the fluorophore.\textsuperscript{26} G-quadruplex formation was studied using pyrene labeled oligo(dG) which forms an eximer between the G-quartet and the pyrene small molecule.\textsuperscript{27} Levels of mRNA have also been studied with multiply labeled or binding of multiple DNA probes which reduces background as the mRNA concentrates the probes in one location within the cell after hybridization.\textsuperscript{28-29} RNA, usually stabilized by modification of the 2' position, has been used as a hybridization probe to study cellular levels of RNAs and during \textit{in vitro} studies.\textsuperscript{30-32} Peptide nucleic acids, which have superior complex stabilities and nuclease resistance compared to RNA and DNA, have been used as hybridization probes. Efforts to reduce background signal with PNA probes include the use of fluorogens and dual reporter systems. Most designs used the fluorogen thiazole orange which is either appended to the end of the probe\textsuperscript{33-34} or used as a base surrogate\textsuperscript{35-39} that turns on upon hybridization. In dual reporter system, two different probes which bind to sequences
which are very close to one another are introduced and bind to their targets. Upon binding, two reactive end groups are brought near one another and they interact to produce a fluorescent signal. Two examples of dual reporters include incorporation of a FRET pair or the use of a Staudinger reaction to release a quenching molecule to restore fluorescence.

Locked nucleic acids have successfully been used to measure nucleic acid levels in both in vivo and in vitro studies. Steroid labeled LNA were used to study microRNA levels and localization in zebrafish during the specimen’s life time via northern blot analysis. Pyrene is a popular reporting molecule which remains dark when probe is free due to nucleobase quenching. Upon hybridization, the pyrene probes become fluorescent as pyrene is excluded from the duplex structure.

Triplex hybrids have also been used as a way of fluorescently tagging a RNA or DNA sequence of interest. The fluorogen 5-(3-methylbenzofuran-2-yl)uracil was incorporated into a triplex-forming oligonucleotide, which upon hybridization to a DNA target induced rotational restriction and increased fluorescence more than 10 times.

Fluorescent cytosine mimic 4-amino-1H-benzo[g]quinazoline-2-one was used to probe protonation sites in triplex forming oligonucleotides.

A popular hybridization probe is the molecular beacon. Molecular beacons are pseudo stable, single stranded hairpin motifs which are functionalized with a fluorophore and quencher on the 5′ and 3′ termini. Upon hybridization, the hairpin unfolds moving the quencher and fluorophore away from one another, restoring fluorescence. Molecular beacons have been successfully used to image mRNA in cells. There has been a large
effort to improve molecular beacon specificity, stability, and reduce background noise. Stability has been achieved using LNA molecular beacons\(^\text{48}\) and $\gamma$-substituted PNAs\(^\text{49}\) which are resistant to nucleases. Specificity has been enhanced through various avenues. FRET systems using two independently binding molecular beacons which recognize two spatially close sequences adorned with complementary fluorophores have been developed\(^\text{50-51}\). Similarly, chemical reactions between two individual molecular beacons in a template-catalyzed reaction results in unquenching of the fluorophore. Reactions such as reductions\(^\text{52-53}\) and autoligation\(^\text{54}\) have been used to turn on fluorophores through chemical modification of the fluorophore itself or by release of a quencher. Ratiometric bimolecular beacons use a reference dye which is always on and a quenched dye in order to determine the difference between undesired localization of the probe and hybridization to target\(^\text{55}\). Triplex molecular beacons have also been developed\(^\text{56}\). Triplex molecular beacons allow for the tuning of specificity and sensitivity by controlling stability of the triplex through sequence choice. Furthermore, background noise can be reduced by adding quenching moieties on the TFO.

Small molecules offer a different avenue for structure recognition as they do not rely on sequence and thus are not prone to binding to similar sequences. Examples include the work by Baranger and Zimmerman which employs melamine bound to an acridine system which becomes quenched upon binding to U-U and T-T mismatches\(^\text{57}\). Derivatives of DFHBI which intercalate have been used as sensors of double stranded DNA\(^\text{58}\). G-quadruplexes have become popular targets for small dyes by researchers exploring their biological relevance. NapthoTASQ is a small molecule with a four guanosine bases
attached to naphthalene core which binds to quadruplexes. UV excitation of the quadruplexes sensitizes naphthalene fluorescence. A large body of work on staining G-quadruplexes has been completed by Shankar Balasubramanian and coworkers using pyridostatin type small molecules.\textsuperscript{59-64} Fluorogens such as thiazole orange\textsuperscript{65} and malachite green\textsuperscript{66} have been used as stains selective for quadruplexes. Xie and coworkers used an emissive uracil surrogate and a dye labeled aminoglycoside in a FRET system to measure sugar binding and subsequent expulsion of the free fluorescent uracil mimic.\textsuperscript{67}

A major issue with hybridization probes is the need to get the probe into a living cell which may not be feasible depending on the design of the probe, the cost and amount of time needed to make synthetic probes, and need for prior knowledge of target strand. Another way of tagging a molecule to measure it levels of production and localization is to add a genetically encoded tag. The most popular genetically encoded tag so far has been green fluorescent protein (GFP).\textsuperscript{68} When labeling a nucleic acid with GFP, the most common way of bringing the two components togethers is through an aptamer system. Typically, the GFP molecule is modified with a small protein sequence for which is the target of a known aptamer. The aptamer is mutated into the desired nucleic acid with multiple repeating aptamers appended in order to reduce background noise. Thanks to the development of fluorescent proteins with different colors and orthogonal aptamers systems, it is now possible to label more than one nucleic acid so that imaging of multiple targets is possible.\textsuperscript{69-70} One of the major issues with GFP tags is the fact that the protein’s fluorescence creates large background signals. Efforts to address this problem include incorporation of nuclear localization signals onto the GFP so that unbound protein is
shuttled to a predetermined location such as the nucleus. Another interesting technique is the that of split GFP which requires two targeting domains to guide the halves of GFP together which then refold and restores fluorescence. The pros of using a GFP tag include: 1) engineered cells can be used repeatedly, 2) no need to inject or transfect probe, 3) high affinity and specificity, and 4) track mRNA from nucleus to cytoplasm. Some of the problems associated with GFP tags include: 1) high background noise, 2) GFP is a large molecule and the need of multiple tags on a target greatly increases the its size and may impact cellular processes, 3) knowledge of target is essential to ensure the large tag will not interfere, and 4) relies on two signal outputs due to the protein and aptamer system.

Based on the above problems associated with GFP tags, there has been considerable effort to improve genetically tagged systems. Aptamers which bind small fluorescent molecules, including fluorogens have been devised. The small dyes are typically membrane permeable and aptamer size is considerably smaller than GFP + aptamer systems. Aptamers for malachite green, Hoechst 33258, dimethylindole red, the fluorogen thiazole orange, and dyes which interact with quenchers via photoinduced electron transfer has been developed. Pei and coworkers found a general system for any small molecule aptamers by attaching the target small molecule to a fluorogen. The most successful aptamer small molecule system is the spinach and broccoli systems which shows and extremely high turn-on when bound its target fluorogen. Spinach is an aptamer developed through SELEX to bind a small molecule mimic of the GFP chromophore DFHBI. Spinach and its successors have relatively strong binding affinities for the dye and very high turn-on to reduce background. Spinach has a G-quadruplex on which DFHBI
binds, turning on the fluorogen. The system has been improved to include fluorogens of various wavelengths, brighter turn-on (spinach2), and smaller size (broccoli). Similar strategies to improve the signal-to-noise ratio include a split aptamer and a secondary binding site for metabolites which refold the aptamer, allowing DFHBI to bind. As with all tagging methods, there are both pros and cons to aptamer-dye system deployment in studies. The positive attributes of aptamer-dye systems include: 1) the use of fluorogens to reduce background noise, 2) relatively small size, 3) only relies on RNA expression, and 4) faster maturation of fluorescent complex than GFP. The drawbacks of dye-aptamer systems include: 1) low affinity of aptamers for dyes, 2) dyes must be membrane permeable, 3) requirement for repeated units of aptamer on target, 4) dyes may sensitize radical production, and 5) engineering of RNA construct.

The labeling of RNA with a probe with some form of readout has been critical in deciphering the roles of various nucleic acid species in cells. The methods discussed above have their various merits and pitfalls. All of the methods require altering the cells natural state; that cannot be avoided with the current technology. Solid-phase synthesis of nucleic acids is limited in size therefore it is difficult to study large sequences. Furthermore, synthetic nucleic acid probes often require injection, transfection, or the use of fixed cell samples. A major drawback of synthetic probes is the need for prior knowledge of sequence and structure which are critical for labeling success. Site specific enzymatic labeling of a recombinantly produced nucleic acids requires either mutation of a native enzyme or an enzyme that already modifies a particular site. Mutation of an enzyme could be very time consuming and there is no general enzyme for site selective labeling. The use
of genetically encoded tags allows for visualization of RNA without having to know the exact structure. However, the most common tag, GFP is large and requires multiple binding events to reduce background fluorescence. Encoded aptamer tags are smaller and require a small molecule that is able to diffuse into the cell. However, aptamers are still rather large and cannot be placed within a construct.

We believe that the bPNA triplex offers a unique genetically encoded labeling moiety capable of being positioned at a terminus, such as traditional fluorescent tags, or internally in place of known duplexes regions. In this way, bPNA could be used to study both localization of RNA in a cell or monitor interactions of the RNA both in vitro and in vivo. Our lab has shown in previous publications that bPNA triplexes can be inserted into RNA constructs as duplex substitutes\textsuperscript{89-91}, and we have also shown in Chapter 2 of this dissertation that bPNA triplexes can support kissing-loop interactions. Furthermore, we have shown that bPNA triplexes are tolerated at protein-nucleic acid interfaces by proteins such as Rop yet stable enough to inhibit certain enzymatic processes.\textsuperscript{92} bPNA also displays superior specificity for long stretches of U-U and T-T mismatches with nanomole affinity. With this knowledge, we decided to test the hypothesis that bPNA triplexes could be used to site selectively modify a nucleic acid as a way of probing the mutant’s interactions in fluorescent assays. The mutant’s interactions with another nucleic acid or other U10 mutant could be monitored via Forster resonance energy transfer (FRET) or fluorogen turn on (Figure 68). The systems we chose to use in this study were the ColE1 and tRNA kissing-loops developed in Chapter 2. In the U10-U10 complexation study, we proposed to use two bPNAs labeled with dyes Cy3 and Cy5 as reporting molecules which when
bound to the U10 mutants would interact via FRET. To study kissing-loop formation with a native hairpin, the fluorogen thiazole orange would be appended to the terminus of bPNA which, upon loop-loop hybridization with the U10-triplex mutant would intercalate and fluoresce.

![Diagram A)

Cy3 - Cy5

![Diagram B)

TO

Figure 70: Fluorescent kissing-loop probes.
A) Dye labeled bPNA FRET reporters of mutant U10 kissing-loops. B) Fluorogen thiazole orange turn on as a reporter of hairpin via kissing-loop formation.

3.2 bPNA AS A FRET BASED PROBE: EXPERIMENTAL DESIGN AND RESULTS:

The FRET experiment would be performed in both the ColE1 mutants and the U10 tRNA due to the different characteristics of their respective kissing complexes. The synthesis of the bPNA labeled with Cy3 (Cy3EM10) and Cy5 (Cy5EM10) is shown in the appendix accompanying this chapter. First, the individual RNA constructs would be
hybridized separately with Cy3 and Cy5 labeled (EM)$_{10}$ in separate stock solutions. After complexation of the bPNA, the triplex hairpins would be mixed together in the presence of Mg$^{2+}$ and incubated to allow for kissing-loop complex formation. At this point, solutions would be analyzed on a fluorimeter by exciting Cy3 at 550 nM and monitoring both Cy3 and Cy5 emission at 570 nM and 670 nM, respectively. The experimental design is shown in Figure 71.

**Figure 71: Experimental design of U10-U10 kissing-loop FRET experiment.** Individual U10 mutants are mixed with either cyanine dye labeled bPNA to form a hairpin. Triplex hybrids are mixed and fluorescence measured for FRET. X and Y are nucleobases which are complementary in non-identical loop sequences.

We first ran control experiments with Cy3-(EM)$_{10}$ in the assay in which the Cy5-bPNA was substituted with unlabeled (EM)$_{10}$ and found that Cy3 fluorescence did not
decrease with kissing complex formation (Supporting information, Figure 78). The results for the FRET experiments are shown in Figure 72. A FRET type signal is observed when

Figure 72: FRET experiments with U10 mutants.

A) FRET experiment results of mutant tRNA. Mutant bound to Cy5-bPNA is titrated into Cy3-bPNA bound mutant and solution incubated 30 minutes on ice. Cy3-mutant complex held constant at 2 μM. B) FRET experiment results in mutant ColE1 stem 1. Mutant bound to Cy5-bPNA is titrated into Cy3-bPNA bound mutant and solution incubated 30 minutes on ice. Cy3-mutant complex held constant at 500 nM. Blue line is emission intensity at 670 nm and the black line is emission intensity at 570 nm in both A and B. In all experiments, U10 mutants were annealed separately with either Cy3 or Cy5 labeled bPNA.

the Cy5EM10-HD2U10 is titrated into Cy3EM10-HD1U10 where the Cy3 signal at 570 nM decreases and Cy5 signal at 670 nM increases until a 1:1 solution is obtained at
which point both signals plateau (Figure 72A). The same FRET signal is detectable when bPNAs are switched between tRNA constructs. A similar significant FRET signal was observed in the ColE1 system no matter the combination of dye-bPNA and U10 triplex formed (Figure 72B).

Excited with these results, we then began to look at competition experiments with the constructs and their duplex counterparts as a means of detecting unlabeled hairpins. The competition experiments were performed taking into account dissociation constants determined in the previous chapter; meaning the more stable kissing complex was expected to form. We therefore expected the FRET pair in the tRNA system to be broken up by the addition of duplex tRNA. For the tRNA system, a preformed kissing complex between the complementary U10 mutants bound to Cy3 and Cy5 labeled bPNA would be formed followed by the addition of duplex tRNA and incubation at 37°C for 1 h. As expected, the FRET interaction decreases as duplex tRNA was titrated into the solution due to displacement of a U10 tRNA (Figure 73). The duplex displaced the U10 mutant with complete displacement appearing to occur at a 1:1 ratio.

We hypothesized the FRET pair to form in the ColE1 system by displacing a wild type stem in a WT-U10 kissing complex. Competition assays with the ColE1 system were first performed in the presence of Rop. The ColE1 competition assay was performed by incubating a wild type hairpin with the complementary U10-mutant bound to a Cy3-(EM)$_{10}$ to form the kissing complex in the presence of 10 μM Rop. The U10-mutant with the same loop sequence as the wild type was then added and the reaction was incubated 3 h on ice. Displacement was observed as fluorescence emission at 570 nm decreased and emission at
Figure 73: tRNA competition assay.
The tRNA duplex is expected to displace the U10-tRNA with the same loop sequence in the kissing complex (top). The graphs show the fluorescent read out of the assay (bottom panels). The duplex tRNA is able to displace the U10-tRNA with the sequence and this is observed by a decrease in emission at 670 nm and an increase of emission at 570 nm. Kissing complex is 500 nm and reaction incubated at 37 °C for 1 h. Blue line is fluorescence emission at 670 nm and black line is emission at 570 nm.

670 nm increased as the FRET capable kissing complex was formed (Figure 74). The ColE1 displacement appears to be complete at a slightly greater than 1:1 ratio. The fact that
displacement of the wild type by a triplex mutant occurred in the presence of Rop was surprising, as we assumed Rop would favor a wild type duplex hairpin and repress exchange. Like the tRNA system, the ColE1 displacement assay appears to favor the thermodynamically stronger kissing complex, in this case the U10-U10 complex.

Figure 74: ColE1 competition assay.
Top: A kissing complex between a wild type stem-loop would be pre-formed and the wild type construct displaced by a U10 mutant with the same loop sequence. Brown bean is Rop. Bottom: Results of competition assay were blue line is fluorescence emission at 670 nm and black line is emission at 570 nm. Wildtype-U10 kissing complex held at 500 nM and U10 mutant with same loop sequence as the wild type construct is titrated in the presence of 10 μM Rop. Incubate 2 h on ice.

The ColE1 system, due to the effect Rop has on complex kinetics was further studied in the competition assays. Specifically, we sought to determine if the presence Rop protein made a difference in the equilibrium of the exchange. Previous experiments
revealed that Rop did not have a preference for the WT-U10 kissing complex as U10-U10 kissing complex was readily formed. The assay was repeated as described above but this time Rop was omitted from the reaction solution. The results indicated that Rop did make a significant difference in the displacement assay: displacement assays ran without Rop see only 50-70% of the fluorescence decrease at 570 nm compared to assays with Rop present in the reaction (Figure 73). In the assays without Rop, displacement was inefficient and did not appear to take place until after the addition of more than one equivalent of the competitive U10. There also appeared to be a sequence dependent affect in the displacement when Rop was not present. In the presence of Rop, the displacement was efficient and nearly complete after the addition of one equivalence of the triplex hybrid. The reason for the Rop dependence exchange could be for one of the following reasons: 1) Rop prefers the U10-U10 kissing complex 2) Rop has no preference but accelerates the exchange 3) A combination of 1 and 2. Our Rop binding experiment in chapter 2 and reported literature values indicate that Rop binds more tightly to the wild type kissing complex. It appears Rop has a kinetic affect and catalyzes the exchange of the triplex and duplex hairpins to the more stable kissing complex. Further studies are required to determine the cause of this phenomena.

From this work, we have successfully used bPNA triplexes to monitor intermolecular interactions between two RNAs. Furthermore, we have used a competition assay to measure displacement of a dye labeled, bPNA triplex mutant. In general, the thermodynamically stable kissing-loop formed. However, in the case of the ColE1 system, exchange was very slow and appeared to be accelerated by Rop.
Comparing Rop vs no Rop for ColE1 FRET competition

Figure 75: Rop dependence on displacement in ColE1 FRET competition assay. Kissing complexes were preformed between wild type and U10-Cy3EM10 complex at 500 nM. A U10 mutant with the same loop sequence as the wild type construct hybridized with Cy5EM10 was titrated into the reaction. Rop when present was at 10 μM. When Rop is absent, displacement of the wild type hairpin is slow (Black Xs and Blue □). When Rop is present, displacement of the wildtype hairpin is readily observed (Black ♦ and Green ◊).

3.3 bPNA-THIAZOLE ORANGE CONJUGATES AS TURN ON PROBES:

3.3.1 EXPERIMENTAL DESIGN AND RESULTS:
Having shown that we can use dye-labeled bPNA triplexes as a way to monitor kissing-loop interactions through FRET assays, we next wanted to develop a probe capable of identifying the presence of a hairpin target (Figure 76). The probe itself would be a small hairpin with a bPNA-binding site. The bPNA design would be a base (EM)$_{10}$ or (ME)$_{10}$ with a linker to covalently bind the peptide to a fluorescent molecule (Figure 76A). The fluorescent molecule would be the fluorogen called thiazole orange which would help to reduce background noise as it would “turn-on” via intercalation when the kissing complex was formed. Thiazole orange when unbound, relaxes via bond rotation after excitation but when bound to nucleic acids by an intercalation mechanism, is forced to relax via fluorescence as rotation is restricted.\textsuperscript{93} The designs for peptides and U10 probe RNA is shown in Tables 5 and 6 and briefly described below.

The linker between the peptide and fluorogen would be varied including no linker, a floppy linker, and a rigid turn linker in an effort to maximize fluorescence turn-on while minimizing background. The floppy linker would be composed of one to three β-alanine amino acids and one or two of a more hydrophilic 2-(2-(2-aminoethoxy)ethoxy)acetic acid linker. The rigid linker would be designed as a turn sequence based on the D-Pro-Gly β-turn motif developed by Gellman and coworkers.\textsuperscript{94} The turn linker was designed using the solution phase structures developed for the kissing-loops of HIV-DIS\textsuperscript{95} and ColE1\textsuperscript{96} which shows a bend in the kissing-loop complex. We hypothesized that if the fluorogen was projected out of the hairpin structure that upon kissing complex formation, the newly formed duplex would fold around the fluorogen (Figure 74B and D) and turn on the
Figure 76: Kissing-loop turn on probe design.
A) Basic structure of \((EM)_{10}\) and \((ME)_{10}\) peptide-thiazole orange conjugates. B) The rigid linker design based on the bent structure observed in NMR solution structure of kissing-loops. C) Floppy linker design for TO probe. D) Rigid turn linker design for TO probe. Fluorogen would be projected out of the triplex construct and kissing-loop would fold around the dye and turn on the fluorescence.

fluorescence. The synthetic β-turn would be flanked on either side by glycine in order to create turns of different lengths. A small molecule turn was also designed using 3-
aminobenzoic acid, however synthesis and purification of the designed peptides proved difficult due to aggregation (supporting information Figure 84).

Probe design was based on a small hairpin as initial studies with the tRNA mutants showed max turn-on when TO-bPNA conjugate was hybridized, presumptively through intercalation into the tRNA body (data not shown). Loops in the hairpin probes were designed to be as unstructured as possible in the monomeric form in an effort to turn off the background fluorescence observed in design 5 (Table 4). Unstructuring of the loop was accomplished via substitution of abasic propyl linkers (Pr) and triethylene glycol linker [(EG)$_3$] in place of the flanking adenosine nucleobases in the HIV-DIS mutant loop sequence and the substitution of unlocked nucleic acids (G$_u$) in positions one and six of the hexanucleotide hybridization sequence. Probes 1 and 3 maintained the backbone length of a nine-nucleotide loop by replacing the flanking dA in the loop region with a propyl linker. Probe 2 was designed to make a symmetric loop and the backbone length resembled a ten-nucleotide loop by adding Pr at positions 1, 2, 9, and 10 of the loop. Probe 4 which incorporated the triethylene glycol linker was designed to provide a side loop design creating the greatest asymmetry in loop of all our probes. The seven-nucleotide loop U10 mutants of the ColE1 system did not show turn on in our previous experiments (data not shown) so shorter loop sequences were used in designs 6 and 7.

All probes were first tested by native gel electrophoresis to ensure they were capable of forming kissing complexes (supporting information Figure 80-84). The fluorescence turn-on experimental results are shown in Table 5 and 6. An example of the observed fluorescence turn-on is shown in Figure 77. The five-nucleotide loop of the
ColE1 probes showed minimal improvement in fluorogen turn-on with most peptides, the exception being the combination of probe 7 with TO-bAla-(EM)$_{10}$ which showed a turn-on of 1.8-fold. There appears to be a sequence-specific effect with turn-on of the thiazole

![Figure 77: TO-bPNA kissing-loop detection.](image)

This figure shows the difference in TO fluorescence between the unbound TO-bPNA-probe triplex hybrid and kissing-loop complex. The top red spectrum indicates TO fluorescence for the kissing complex and bottom black spectrum indicates free TO-bound probe.

orange as the complementary sequence in probe 6 did not show the same level of TO turn on. Probes to detect complexation with the tRNA mutants also showed somewhat weak turn-on factors. The best results of about 2.3-fold turn-on were from probe 1 in which positions 1, 2 and 9 of the loop sequence were abasic propyl linkers and peptides TO-(bAla)$_2$(ME)$_{10}$ and TO-PEG(ME)$_{10}$ whose linkers are roughly the same size. The rigid linker did not appear to make a significant difference in reducing background turn-on
fluorescence. In general, the less structured the loop, the less turn-on was observed upon kissing complex formation.

Table 5: Turn-on factor of hybridized TO-bPNA conjugates and 9-membered loop probes.

<table>
<thead>
<tr>
<th>Design</th>
<th>Probe Structure</th>
<th>Turn-on Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TO-(ME)₁₀</td>
<td>1.3</td>
</tr>
<tr>
<td>2</td>
<td>TO-G(ME)₁₀</td>
<td>1.5</td>
</tr>
<tr>
<td>3</td>
<td>TO-PEG(ME)₁₀</td>
<td>2.3</td>
</tr>
<tr>
<td>4</td>
<td>TO-P(GEM)₁₀</td>
<td>1.7</td>
</tr>
<tr>
<td>5</td>
<td>TO-bAla(EM)₁₀</td>
<td>1.6</td>
</tr>
<tr>
<td>6</td>
<td>TO-(bAla)₂(ME)₅₀</td>
<td>2.25</td>
</tr>
<tr>
<td>7</td>
<td>TO-(bAla)₃(ME)₅₀</td>
<td>1.9</td>
</tr>
<tr>
<td>8</td>
<td>TO-G₂PG(ME)₅₀</td>
<td>1.6</td>
</tr>
<tr>
<td>9</td>
<td>TO-G₃PG(ME)₅₀</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Table is reported as number of fold fluorescence turn-on of probe upon kissing loop complex formation.

Peptides: TO is thiazole orange represented as an orange circle. PEG in peptide design is 2-(2-(2-aminoethoxy)ethoxy)acetic acid. bAla is β-Alanine. DP is D-Proline. Probes: Designs 1-5 were tested against HD2Nat tRNA mutant. Pr in probe sequence indicates an abasic propyl linker. Gu is an unlocked nucleic acid. (EG)₃ is an abasic triethylene glycol linker.
Table 6: Turn-on factor of hybridized TO-bPNA conjugates and 7-membered loop probes.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>TO-G(ME)$_{10}$</th>
<th>TO-PEG(ME)$_{10}$</th>
<th>TO-(PEG)$<em>{2}$(ME)$</em>{10}$</th>
<th>TO-bAla(EM)$_{10}$</th>
<th>TO-(bAla)$<em>{3}$(ME)$</em>{10}$</th>
<th>TO-D$<em>{2}$PG-(ME)$</em>{10}$</th>
<th>TO-G$<em>{2}$D$</em>{2}$PG(ME)$_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.9</td>
<td>1.3</td>
<td>1.15</td>
<td>0.9</td>
<td>1.16</td>
<td>0.9</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Table is reported as number of fold fluorescence turn-on of probe upon kissing loop complex formation. Peptides: TO is thiazole orange represented as an orange circle. PEG in peptide design is 2-(2-aminoethoxy)ethoxy)acetic acid. bAla is β-Alanine. DP is D-Proline. Probes: 6 and 7 were tested against RNAI-1WT and RNAII-1WT, respectively.

In general, we were able to detect kissing complex formation through turn-on making this approach a viable option for in vitro detection. If a fluorogen-based bPNA triplex system is to be used in vivo, improvements must be made to reduce background turn on. The TO-bPNA conjugates already show a turn-on without hybridization to the U10 probe and further turn-on upon triplex formation is also a problem. Decreased background could be realized by changing the fluorogen, alternative probe redesigns, or the use of a small-molecule quencher whose affects are removed upon kissing-loop formation.
3.3.2 DISCUSSION AND CONCLUSIONS:

The specificity of bPNAs for sequential U-U mismatches offers a new approach to site specifically label a nucleic acid by mutating a duplex region to a bPNA binding site. We have successfully utilized bPNA as a fluorescent probe to monitor kissing-loop formation through both FRET competition assays and turn on probes. The tRNA mutant system in the competition assay successfully showed that hairpins with higher affinity can displace a weaker binding triplex mutant and the displacement could be monitored with FRET. This strategy could be used to monitor the expression of any sequence or as a structural probe with the target sequence displacing a FRET contributor. The strength of the complex can be tailored using common strategies such as UNAs and LNAs to adjust the stringency of the probe.

FRET competitions assays in the ColE1 system was less straightforward than the tRNA system. We expected the triplex hairpin to displace the wild type hairpin with the same loop sequence in a kissing complex with a complementary U10 mutant. Interestingly, at stoichiometric amounts, the displacement was not observed until an excess of competitor was added. However, when 10 μM of Rop was present, displacement of the wild type was efficient at sub-stoichiometric amounts. These results point to a catalytic role for Rop in the displacement assay. We were able to realize this unique interaction using two bPNAs functionalized with a pair of FRET dyes with which we could then site selectively label two different, interacting RNAs.
In our final study, we successfully utilized bPNA as a turn-on probe for complexation. Using a fluorogen probe, we could monitor kissing-loop formation with only one bPNA binding mutation in the system. While further work is required to achieve acceptable background levels for use of bPNA as an in vivo probe, fluorogen labeled bPNA such as the ones present above could be readily used for in vitro experiments. Fluorogen-bound bPNA could readily be used to measure nucleic acid interactions provided the bPNA triplex could be placed near the region of interest via duplex mutation. Strategic placement of the probe-functionalized bPNA triplexes near a site of interaction or in a region which undergoes a conformational change could potentially be observed as the probe environment changes, resulting in a change of the fluorescence signal of the probe. In summary, we have shown that bPNA triplexes could be used as site specific labels for nucleic acids.
3.4 SUPPORTING INFORMATION CHAPTER 3

3.4.1 Materials and instrumentation:

Chemicals for Fmoc-protected amino acids, peptide, acrylamide solutions and dye synthesis were purchased from Chem-Impex, TCI America, AAPPTec, and VWR and used without further purification. Oligonucleotides were purchased from Integrated DNA technologies. HD2Nat was synthesized as described in Chapter 2 of this document. rNTPs were purchased from Sigma Aldrich. T7 RNA polymerase and Pfu DNA polymerase enzymes were produced in house.

The concentration of all DNA oligomers and synthetic peptides was determined by UV absorbance using a Thermo Scientific Nanodrop 200c spectrophotometer. The concentration of thiazole orange peptides were determined by measuring the absorbance at 501 nm and using the molar extinction coefficient of 63,000 M\(^{-1}\) cm\(^{-1}\). Fluorescence experiment was performed on a Thermo Scientific NanoDrop 3300 fluorospectrometer. MALDI-TOF Mass spectra were acquired on Bruker Microflex MALDI-TOF instrument. Electrospray ionization mass spectrometry (ESI) was acquired on Bruker MicroTOF equipped with an electrospray ionization source. NMR spectra were acquired on Bruker Advance DPX 400 instrument. Solid phase peptide synthesis was performed on AAPPTec Apex 396 peptide synthesizer. All peptides were purified on a Hitachi L-7150 pump, Hitachi D-7000 interface and Hitachi L-7400 UV detector on an AAPPTec RPC18 preparative HPLC column and the purity was confirmed on an AAPPTec RP-C18 analytical column. Data was plotted on KaleidaGraph 4.0.
3.4.2 Fluorescence experiments procedure:

**FRET experiment:** U10-RNA and Cy3(EM)₁₀G-CONH₂ stock solutions at 3x concentrations were prepared in 1x buffer. Complementary U10-RNA and Cy5(EM)₁₀G-CONH₂ were diluted in a separate vial in 1x buffer. RNA/bPNA solutions were annealed at 90 °C for 5 min, and snap cooled on ice for 5 min. Rop was added if necessary to bring stock concentration to 30 μM of Rop (10μM final concentration after dilution with FRET partner and 1x buffer). Reaction solutions were made by aliquoting 10 μl of RNA/Cy3 stock into multiple tubes and adding an increasing amount of RNA/Cy5 to separate tubes, using 1x buffer to bring the total volume to 30 μl. The final concentration of U10-RNA/Cy3 was 2μM or 0.5 μM for the tRNA and ColE1 systems, respectively. U10-RNA/Cy5 ranged from 0 μM to 4 μM. For control experiments, an unlabeled bPNA was used in place of the Cy5-bPNA. RNAs were mixed and incubated at least 30 min on ice. The fluorescence experiment was performed on a Thermo Scientific NanoDrop 3300 fluorospectrometer by exciting Cy3 at 550 nm and monitoring at 570 nm and 670 nm.

**FRET competition experiments procedure.**

**tRNA:** HD1U10 was snap annealed with Cy3(EM)₁₀G-CONH₂ and HD2U10 was snap annealed with Cy5(EM)₁₀G-CONH₂ as described above in 1x kissing buffer. The triplex containing RNA’s were mixed with one another to give a complex concentration of 500 nM (500 nM of each RNA triplex) and incubated 20 min on ice. Varying amounts of HD1Nat or HD2Nat was titrated into the samples with a concentration ranging from 0-2000 nM and the samples brought up to 20 μl with 1x buffer. The samples were incubated
at 37°C for 1 h and then analyzed. Fluorescence was measured by exciting system at 550 nM and monitoring at 570 nM and 670 nM.

**ColE1:** RNA-1U10 and Cy3(EM)10G-CONH₂ were snap annealed in 1x kissing buffer. RNA-1WT was snap annealed in 1x buffer. RNA-1U10-Cy3EM10 and RNA-1WT with complementary loop sequences were mixed 1:1 to produce a stock solution 3x greater than assay concentrations and incubated 20 min on ice. 10 µl of stock U10-WT kissing complex is added to multiple tubes. At the same time, the U10 mutant with the same loop sequence as RNA-1WT was mixed with Cy5(EM)10G-CONH₂ and snap annealed in 1x buffer. To the tubes with kissing complex was added various amounts of RNA-Cy5EM10 triplex using 1x buffer to bring the volume to 30 µl. The final concentration of preformed kissing complex was 500 nM and competitor concentration ranged from 0-10 µM. The reaction was incubated for 2 h on ice and fluorescence measured by exciting at 550 nm and observing the emission at 570 nm and 670 nm. When Rop was present, the stock solution of kissing complex had Rop added at a concentration of 30 µM. Final Rop concentration was 10 µM.

**TO-probe turn on experiments:** U10-probe was mixed with TO-bPNA at 2 µM concentration in 1x kissing buffer and snap annealed. HD2Nat or ColE1 RNA-WT was diluted into 1x kissing buffer at a concentration of 3 µM and snap annealed. An equal volume of both solutions was mixed and incubated 30 min on ice. Fluorescence was then measured by exciting at 501 nm and measuring emission at 530 nm.
Figure 78: Fluorescent controls for FRET with Cy3-bPNA and unlabeled bPNA in tRNA kissing-loop.
The above graph shows the fluorescence change in triplex-tRNA HD1U10 bound to EM10 is titrated into a solution containing a constant 500 nM of triplex-tRNA HD2U10 bound to Cy3-EM10 in 1:1 ratio.

Figure 79: Effect of kissing loop complex formation on Cy3 fluorescence in ColE1 system.
This is a control experiment for Cy3 fluorescence when forming kissing complex in ColE1 system. The above graph shows the fluorescence change in triplex-RNAI-1U10 bound to EM10 is titrated into a solution containing a constant 500 nM of triplex-RNAII-1U10 bound to Cy3-EM10. RNAs and bPNA were mixed 1:1 and incubated separately on ice for 5 minutes. RNA-bPNA triplexes were titrated in and samples incubated 30 min on ice. Cy3 was excited at 550 nm and fluorescence monitored at 570 nm.
3.4.3 EMSA gels of probes with targets.

Gel experiments were performed as described in Chapter 2 of this document.

**Figure 80: EMSA gel of kissing complex formation with probe 4.**
The gel above shows the gel shift of probe 4-Cy5bPNA triplex upon binding to HD2Nat. Probe 4 is held constant at 1 μM and HD2Nat is titrating in with a range of 0-8 μM. Gel is imaged by Cy5 fluorescence. Top band is kissing complex. Bottom band is free probe 4.

**Figure 81: EMSA gel of kissing complex formation with probe 3.**
The gel above shows the gel shift of probe 3-Cy5bPNA triplex upon binding to HD2Nat. Probe 3 is held constant at 100 nM and HD2Nat is titrating in with a range of 0-1600 nM. Gel is imaged by Cy5 fluorescence. Top band is kissing complex. Bottom band is free probe 3.

**Figure 82: EMSA gel of kissing complex formation with probe 2.**
The gel above shows the gel shift of probe 2-Cy5bPNA triplex upon binding to HD2Nat. Probe 2 is held constant at 100 nM and HD2Nat is titrating in with a range of 0-200 nM.
Gel is imaged by Cy5 fluorescence. Top band is kissing complex. Bottom band is free probe 2.

Kissing complex

Figure 83: EMSA gel of kissing complex formation with probe 1. The gel above shows the gel shift of probe 1-Cy5bPNA triplex upon binding to HD2Nat. Probe 1 is held constant at 100 nM and HD2Nat is titrating in with a range of 0-200 nM. Gel is imaged by Cy5 fluorescence. Top band is kissing complex. Bottom band is free probe 1.

Figure 84: EMSA gel of kissing complex formation with probe 6 and 7. The gel above shows the gels shift of probe 6-Cy5(EM)_{10} and probe 7-(EM)_{10} kissing complex. Probe 7-Cy5(EM)_{10} is held constant at 500 nM and probe 6-(EM)_{10} is titrated in with a range of 0-2.5 μM. Gel is imaged by Cy5 fluorescence. Excess Probe 6 may be stained by excess Cy5(EM)_{10}.

3.4.4 Synthetic Procedures and compound characterization.

Thiazole orange acetate synthesis:

Synthesis of thiazole orange acetate was performed as reported.\(^97\) Yield was 164 mg (2.8%) of a bright orange powder.
Figure 85: Proton NMR of thiazole orange acetate. 
$^1$H NMR (400 MHz, MeOD-D4) of thiazole orange acetate.

**Peptide Synthesis:** Fmoc-Lys(Mel)-OH and NH$_2$-(EM)$_{10}$G-CONH$_2$ was synthesized as previously reported. Peptides were synthesized on an AAPPTec Apex 396 peptide synthesizer using standard solid phase Fmoc chemistry on Rink Resin LS with 0.28 mmol/g loading. 0.5 mmol of Amino acids were coupled with 0.5 mmol of HOBt and 0.5 mmol DIC in 2 ml DMF. The coupling reactions were mixed for 1.5 h at which point they were rinsed with 3 x 2 ml DMF. Fmoc protecting groups were cleaved by subjecting resin to 2 x 2 ml of 20% v/v piperidine solution in DMF for ten min followed by another rinse cycle. Peptides were cleaved from the resin using 95:5 v/v TFA:water solution for 4 h. Cleaved peptide solutions were filtered through cotton to remove the beads, peptide
crashed out in cold diethyl ether, centrifuged to collect the solid peptide, washed 3 times with cold ether, and dried using a nitrogen gas stream to remove excess ether and TFA. Peptides were purified using preparatory HPLC using a C18 column and water/acetonitrile gradient containing 0.1% v/v TFA. The purification used solvent A (water, 0.1% v/v TFA) and solvent B (80% v/v acetonitrile, 0.1% v/v TFA) gradient as follows: 0-5 min 10% B, 5-20 min increase to 38% B, 30% B for 15 min, then 100% B for 5 min.

**Thiazole Orange conjugation:** Thirty mg of thiazole orange and 23 mg of pyridinium p-toluenesulfonate (PPTS) were dissolved in 3 ml of DMF. To this mixture was added 172.5 μM N-hydroxysuccinimide and 172.5 μM N,N'-diisopropylcarbodiimide and the reaction is incubated 2 h at 38°C. The activated thiazole orange is then added to the peptide resin and reaction is incubated at 70°C for 2 h. Reaction was monitored using Kaiser test and looking at beads under microscope to examine color.
Figure 86: Attempted TO-peptides that could not be purified.
Figure 87: HPLC trace and MALDI-TOF of TO-(PEG)$_2$(ME)$_{10}$G-CNH$_2$
A) MALDI-TOF spectra of TO-(PEG)$_2$(ME)$_{10}$G-CNH$_2$: Calculated [M+]: 4357.050
Found: 4355.796. B) HPLC trace of TO-(PEG)$_2$(ME)$_{10}$G-CNH$_2$ on a RP-C18 analytical
column monitored with UV-Vis detector at 220 nm.
Figure 88: HPLC trace and MALDI-TOF of TO-(PEG)(ME)$_{10}$G-CONH$_2$
A) MALDI-TOF spectra of TO-(PEG)(ME)$_{10}$G-CONH$_2$: Calculated [M+]: 4211.976
Found: 4209.825. B) HPLC trace of TO-(PEG)(ME)$_{10}$G-CONH$_2$ on a RP-C18 analytical
column monitored with UV-Vis detector at 220 nm.
Figure 89: HPLC trace and MALDI-TOF of TO-(bAla)$_2$(EM)$_{10}$G-CONH$_2$
A) MALDI-TOF spectra of TO-(bAla)$_2$(EM)$_{10}$G-CONH$_2$: Calculated [M$^+ $]: 4208.976
Found: 4204.906. B) HPLC trace of TO-(bAla)$_2$(EM)$_{10}$G-CONH$_2$ on a RP-C18 analytical
column monitored with UV-Vis detector at 220 nm.
Figure 90: HPLC trace and MALDI-TOF of TO-(bAla)(EM)_{10}G-CONH2
A) MALDI-TOF spectra of TO-(bAla)(EM)_{10}G-CONH2: Calculated [M+]: 4137.939
Found: 4134.006. B) HPLC trace of TO-(bAla)(EM)_{10}G-CONH2 on a RP-C18 analytical
column monitored with UV-Vis detector at 220 nm.
Figure 91: HPLC trace and MALDI-TOF of TO-dP-G-(EM)$_{10}$G-CNH2
A) MALDI-TOF spectra of TO-dP-G-(EM)$_{10}$G-CNH2: Calculated [M+]: 4220.976
Found: 4219.422. B) HPLC trace of TO-dP-G-(EM)$_{10}$G-CNH2 on a RP-C18 analytical
column monitored with UV-Vis detector at 220 nm.
Figure 92: HPLC trace and MALDI-TOF of TO-G-dP-G-(EM)$_{10}$G-CONH$_2$
A) MALDI-TOF spectra of TO-G-dP-G-(EM)$_{10}$G-CONH$_2$: Calculated [M+]: 4277.998
Found: 4274.353.  B) HPLC trace of TO-G-dP-G-(EM)$_{10}$G-CONH$_2$ on a RP-C18 analytical column monitored with UV-Vis detector at 220 nm.
Figure 93: HPLC trace and MALDI-TOF of TO-G-(ME)$_{10}$G-CONH$_2$
A) MALDI-TOF spectra of TO-G-(ME)$_{10}$G-CONH$_2$: Calculated [M$^+$]: 4123.923 Found: 4121.694. B) HPLC trace of TO-G-(ME)$_{10}$G-CONH$_2$ on a RP-C18 analytical column monitored with UV-Vis detector at 220 nm.
Figure 94: HPLC trace and MALDI-TOF of TO-(ME)_{10}G-CONH2
A) MALDI-TOF spectra of TO-(ME)_{10}G-CONH2: Calculated [M+]: 4066.902 Found: 4061.909. B) HPLC trace of TO-(ME)_{10}G-CONH2 on a RP-C18 analytical column monitored with UV-Vis detector at 220 nm.
Figure 95: HPLC trace and MALDI-TOF of TO-(bAla)$_3$-(EM)$_{10}$G-CONH$_2$

HPLC trace of TO-(bAla)$_3$-(EM)$_{10}$G-CONH$_2$ on a RP-C18 analytical column monitored with UV-Vis detector at 220 nm.
3.5 REFERENCES CHAPTER 3


37. Kummer, S.; Knoll, A.; Socher, E.; Bethge, L.; Herrmann, A.; Seitz, O., Fluorescence Imaging of Influenza H1N1 mRNA in Living Infected Cells Using Single-


68. Tyagi, S., Imaging intracellular RNA distribution and dynamics in living cells. 


83. Strack, R. L.; Disney, M. D.; Jaffrey, S. R., A superfolding Spinach2 reveals the

84. Filonov, G. S.; Moon, J. D.; Svensen, N.; Jaffrey, S. R., Broccoli: Rapid Selection
of an RNA Mimic of Green Fluorescent Protein by Fluorescence-Based Selection and

85. Kikuchi, N.; Kolpashchikov, D. M., Split Spinach Aptamer for Highly Selective

Monitoring of RNA Assembly and Processing Using the Split-Spinach Aptamer. *ACS

87. Paige, J. S.; Nguyen-Duc, T.; Song, W.; Jaffrey, S. R., Fluorescence Imaging of

fluorescence imaging of intracellular metabolites and proteins in living bacteria. *Nat

89. Piao, X.; Xia, X.; Bong, D., Bifacial Peptide Nucleic Acid Directs Cooperative
Folding and Assembly of Binary, Ternary, and Quaternary DNA Complexes.
*Biochemistry* **2013**, *52* (37), 6313-6323.

90. Xia, X.; Piao, X.; Bong, D., Bifacial Peptide Nucleic Acid as an Allosteric Switch


CHAPTER 1 REFERENCES:


   


218


78. Howard, F. B.; Frazier, J.; Lipsett, M. N.; Miles, H. T., Infrared demonstration of two- and three-strand helix formation between poly C and guanosine mononucleotides


94. Muramatsu, T.; Nishikawa, K.; Nemoto, F.; Kuchino, Y.; Nishimura, S.; Miyazawa, T.; Yokoyama, S., Codon and amino-acid specificities of a transfer RNA are


Antiparallel polypurine phosphorothioate oligonucleotides form stable triplexes with the
rat α1(I) collagen gene promoter and inhibit transcription in cultured rat fibroblasts.
_Nucleic Acids Research_ **1997**, 25 (11), 2182-2188.

154. Gryaznov, S.; Chen, J.-K., Oligodeoxyribonucleotide N3’-P5’
(7), 3143-3144.

oligodeoxyribonucleotide N3’→P5’ phosphoramidates. _Nucleic Acids Research_ **1995**, **23**
(14), 2661-2668.

156. Schultz, R. G.; Gryaznov, S. M., Oligo-2’-Fluoro-2’-Deoxynucleotide N3’→P5’
2966-2973.

157. Mignet, N.; Gryaznov, S. M., Zwitterionic oligodeoxyribonucleotide N3’→P5’
phosphoramidates: Synthesis and properties. _Nucleic Acids Research_ **1998**, **26** (2), 431-
438.

158. Tarköy, M.; Bolli, M.; Leumann, C., Nucleic-Acid Analogues with Restricted
Conformational Flexibility in the Sugar-Phosphate Backbone (‘bicyclo-DNA’). Part 3.
Synthesis, pairing properties, and calorimetric determination of duplex and triplex
stability of decanucleotides from [(3’S,5’R)-2’-deoxy-3’,5’- ethano-β-D-


201. Gait, M. J.; Jones, A. S.; Jones, M. D.; Shepherd, M. J.; Walker, R. T., Synthetic analogues of polynucleotides. Part 15. The synthesis and properties of poly(5[prime or minute]-amino-3[prime or minute]-O-carboxymethyl-2[prime or minute],5[prime or minute]-dideoxy-erythro-pentonucleosides) containing 3[prime or minute](O)[rightward arrow] 5[prime or minute](C) acetamidate linkages. *Journal of the Chemical Society, Perkin Transactions 1* **1979**, (0), 1389-1394.


227. Mittapalli, G. K.; Osornio, Y. M.; Guerrero, M. A.; Reddy, K. R.; Krishnamurthy, R.; Eschenmoser, A., Mapping the Landscape of Potentially Primordial Informational Oligomers: Oligodipeptides Tagged with 2,4-Disubstituted 5-Aminopyrimidines as


**CHAPTER 2 REFERENCES:**


7. Li, P. T. X.; Bustamante, C.; Tinoco, I., Unusual mechanical stability of a minimal RNA kissing complex. *PNAS* 2006, 103 (43), 15847-15852.


246


17. Melchers, W. J.; Hoenderop, J. G.; Slot, H. J. B.; Pleij, C. W.; Pilipenko, E. V.; Agol, V. I.; Galama, J. M., Kissing of the two predominant hairpin loops in the coxsackie


258


**CHAPTER 3 REFERENCES:**


Seitz, O., LNA-enhanced DNA FIT-probes for multicolour RNA imaging. *Chemical

Fluorescent PNA Probes as Hybridization Labels for Biological RNA†. *Biochemistry
2006, 45* (19), 6066-6074.

41. Cai, J.; Li, X.; Yue, X.; Taylor, J. S., Nucleic Acid-Triggered Fluorescent Probe

42. Wienholds, E.; Kloosterman, W. P.; Miska, E.; Alvarez-Saavedra, E.; Berezikov,

43. Østergaard, M. E.; Cheguru, P.; Papasani, M. R.; Hill, R. A.; Hrdlicka, P. J.,
Glowing Locked Nucleic Acids: Brightly Fluorescent Probes for Detection of Nucleic

Ito, T.; Shirouzu, M.; Kuwasako, K.; Muto, Y.; Sekine, M.; Seio, K., Controlling the
Fluorescence of Benzofuran-Modified Uracil Residues in Oligonucleotides by Triple-

45. Godde, F.; Toulmé, J.-J.; Moreau, S., 4-amino-1H-benzo[g]quinazoline-2-one: a
fluorescent analog of cytosine to probe protonation sites in triplex forming


