Invader Probe Targeting of Chromosomal DNA

A Thesis

Presented in Partial Fulfillment of the Requirements for the

Degree of Masters of Science

with a

Major in Chemistry

in the

College of Graduate Studies

University of Idaho

by

Caroline P. Shepard

Major Professor: Patrick J. Hrdlicka, Ph.D.

Committee Members: Peter B. Allen, Ph.D.; Eric B. Brauns, Ph.D.

Department Administrator: Ray von Wandruszka, Ph.D.

August 2020

Authorization to Submit Thesis

This thesis of Caroline P. Shepard, submitted for the degree of Masters of Science with a Major in Chemistry and titled "Invader Targeting of Chromosomal DNA," has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: _____ Date: _____
Patrick J. Hrdlicka, Ph.D.

Committee Members: _____ Date: _____
Eric B. Brauns, Ph.D.

_____ Date: _____
Peter B. Allen, Ph.D.

Department Administrator: _____ Date: _____
Ray von Wandruszka, Ph.D

Abstract

Interest in developing probes capable of targeting chromosomal DNA in cells has grown to meet diagnostic and therapeutic needs. DNA has a stable predicable double-stranded structure that has been the subject of study to identify agents that can specifically bind to the duplex (CHAPTER 1). Success stories from probe technologies such as triplex-forming oligonucleotides, peptide nucleic acids (PNAs), and minor-groove-binding polyamides have been well characterized. However, they suffer limits of detection with experimental conditions requirements (homopurine targets; denaturing steps; low ionic strengths; short target sequences). Newly discovered CRISPR-Cas9 has garnered much attention; however, the approach requires transfection of plasmids encoding CRISPR-Cas9 components. To address these shortcomings, our laboratory has developed Invader probes. Placement of 2'-*O*-(pyren-1-yl)methyl RNA monomers in +1 interstrand zipper arrangements destabilizes the probe duplex as the intercalating pyrene moieties *vie* for the same space between two Watson-Crick base pairs. These 'energetic hotspots' activate the double-stranded probe and, in concert with the high affinity for complementary DNA (cDNA) displayed by individual probe strands, provide the driving force for recognition of mixed-sequence target sites. The capabilities and features of Invader probes specific detection of chromosomal DNA fluorescent *in situ* hybridization (FISH) assays at near physiological conditions were statistically analyzed. Based on these results, Optimized Invader probes were synthesized and showed improved efficiency in chromosomal DNA detection (CHAPTER 2). Moreover, the combination of Invader FISH probes with the powerful detection properties of flow cytometry was used to quantify thousands of specifically label isolated nuclei, offering a potential advantage over the laborious microscope evaluation of detection (APPENDIX A).

Acknowledgements

I would like to thank Dr. Patrick Hrdlicka for his continued guidance, and challenging me to always critically think to find the ‘whys’ behind every event, occurrence, and idea. This is a skill that has brought me opportunity and success in my endeavors. I am also grateful to my lab-mates who took the time to teach me and readily extended a helping hand.

Dedication

To my friends and family that did not let me waiver

.

Table of Contents

Authorization to Submit Thesis ii
Abstract iii
Acknowledgements iv
Dedication v
Table of Contents vi
List of Tables vii
List of Figures ix
Chapter 1: DNA Targeting Agents in the Context of Imaging and Detection of Chromosomal DNA 1
1.1 DNA as a Target 1
1.2 History of DNA Targeting Desings 4
1.3 Introduction to Invaders 13
1.4 Development of Efficient Invaders 14
1.5 Invader Probe Applications 16
1.6 Conclusion 17
1.7 References 18
Chapter 2: Recogition of Chromosomal DNA Targets Using Invader Probes 22
2.1 Introduction 24
2.2 Results and Discussion 26
2.3 Conclusions 55
2.4 Supplementary Material 57
2.5 References 97
Appendix A: Analysis of Select Invader Probes by Flow Cytometry ... **Error! Bookmark not defined.**
A.1 Introduction 100
A.2 Results and Discussion 101
A.3 Conclusions 106
A.4 Acknowledgements 107
A.5 Supplementary Material 107
A.6 References 109

List of Tables

Table 2.2-1. Thermal denaturation temperatures (T_m), Thermal Advantage Values (TA_{isq} and TA_{DH}), Percent Modification (Mod%), and GC-content (GC%) of *DYZ1*-targeting Invader probes and duplexes between individual probe strands and cDNA.[a]……..30

Table 2.2-2. C_{50} values at 15 h for Invader probes studied herein.[a]……...38

Table 2.2-3. Percentages of nuclei presenting clear signal (coverage) in d-FISH and nd-FISH assays using *DYZ1*-targeting Invader probes……44

Table 2.2-4. Spearman Rank Correlation with nd-FISH Coverage Results……..47

Table 2.2-5. Thermal denaturation temperatures (T_m), Thermal Advantages (TA_{isq} and TA_{DH}), Percent Modification (Mod%), and Percent GC-content (GC%) of Optimized *DYZ1*-targeting Invader probes and duplexes between individual probe strands and cDNA.[a]…….49

Table 2.2-6. C_{50} values at 15 h for Invader probes studied herein.[a]……52

Table 2.2-7 Percentages of nuclei presenting clear signal (coverage) in d-FISH and nd-FISH assays using *DYZ1*-targeting Optimized Invader probes……55

Table 2.4-1. MALDI-MS of individual Invader probe strands denoted up (u) or down (d)...66

Table 2.4-2. MALDI-MS of individual Optimized Invader probe strands denoted up (u) or down (d)……...67

Table 2.4-3. Change in Gibbs free energy (ΔG^{310}) at upon formation of duplexes and change in reaction free energy upon Invader-mediated recognition of dsDNA targets (ΔG^{310}_{rec}).[a]……73

Table 2.4-4. Change in enthalpy (ΔH) upon formation of duplexes and change in reaction enthalpy upon Invader-mediated recognition of dsDNA targets ΔH_{rec}.[a]……..74

Table 2.4-5. Change in entropy ($-T\Delta S^{310}$) upon formation of duplexes and change in reaction entropy upon Invader-mediated recognition of dsDNA targets ($-T\Delta S^{310}_{rec}$).[a]…….75

Table 2.4-6. Change in Gibbs free energy (ΔG^{310}) for Optimized Invader probes at upon formation of duplexes and change in reaction free energy upon Invader-mediated recognition of dsDNA targets (ΔG^{310}_{rec}).[a]……76

Table 2.4-7. Change in enthalpy (ΔH) for Optimized Invader probes upon formation of duplexes and change in reaction enthalpy upon Invader-mediated recognition of dsDNA targets ΔH_{rec}.[a]……...76

Table 2.4-8. Change in entropy ($-T\Delta S^{310}$) for Optimized Invader probes upon formation of duplexes and change in reaction entropy upon Invader-mediated recognition of dsDNA targets ($-T\Delta S^{310}_{rec}$).[a]……77

Table 2.4-9. Sequence and intramolecular T_m of the DNA hairpins used herein.[a]……78

Table 2.4-10. C_{50} values at 2.5 h for Invader probes studied herein.[a]……..84

Table 2.4-11. C_{50} values at 2.5 h for Optimized Invader probes studied herein.[a]……..…85

Table 2.4-12. Quality of signal intensity (I) and background (B) in representative images from d-FISH experiments in which Invader probes were used at 1x concentration (30 ng of probe per 200 μL of PCR buffer), and nd-FISH experiments in which Invader probes were used at 1x-4x concentration……..94

Table A.2-1. Sequences of Invader probes used for Flow-FISH Analysis……...102

List of Figures

Figure 1.1-1. Structure of DNA and Watson-Crick base pairing (dashed lines denotes hydrogen bonds)……3

Figure 1.2-1. Schematic of current dsDNA targeting approaches……6

Figure 1.2-2. Structures of PNA and γ-PNAs. (B denotes nucleobases A, C, G, and T)…....9

Figure 1.2-3. Pseudocomplementary (pc) PNA base pair……...12

Figure 1.3-1. Double duplex invasion mechanism of Invader probes……14

Figure 2.1-1. (a) Illustration of Invader-mediated recognition of dsDNA *via* a double-duplex invasion process. (b) Structure of Invader monomers used herein…….……27

Figure 2.2-1. Principle Hairpin Assay……33

Figure 2.2-2. a) Representative gel electrophoretograms from recognition experiments at 15 h between a 100-fold molar excess of Invader probes **INV1**-**INV10** and their corresponding DNA hairpin targets **DH1**-**DH10**. b) Histograms depicting averaged results from at least three recognition experiments with error bars representing standard deviation. TC = ternary complex. DIG-labeled **DH1**-**DH10** (34.4 nM, sequences shown in Table 2.4-9) were incubated with the corresponding Invader probe in HEPES buffer (50 mM HEPES, 100 mM NaCl, 5 mM $MgCl_2$, pH 7.2, 10% sucrose, 1.44 mM spermine tetrahyrdochloride) for 15 h at 37 °C. Incubation mixtures were resolved on 12% non-denaturing TBE-PAGE slabs (~70 V, ~4 °C, ~1.5 h)……..36

Figure 2.2-3. Dose-response curves for **INV1**, **INV2, INV3** (upper left panel), **INV5**, **INV6**, and, **INV7** (upper right panel), and **INV8**, **INV9**, and **INV10** (lower panel) at 37 °C for 15 h. Experimental conditions are as described in Figure 2.2-1, except for variable probe concentrations……..38

Figure 2.2-4. Binding specificity. Representative electrophoretograms for experiments in which the specified Invader probe (3.44 μM) was incubated at 37 °C for 2.5 h with a mixture of nine non-target DNA hairpins (each hairpin present at 34.4 nM). For example, **INV1** was incubated with **DH2**-**DH10**, while **INV10** was incubated with **DH1**-**DH9**. Conditions are otherwise as described in Figure 2.2-2……..39

Figure 2.2-5 Images from FISH experiments using *DYZ1*-targeting Invader probes **INV2** and **INV10** under denaturing (80 °C, 5 min) (left), or non-denaturing (3 h, 38.5 °C) conditions (right). Images are representative of the signal intensity and background, and the size and morphology of all analyzed nuclei (~200). Fixed isolated nuclei from male bovine kidney cells were incubated with probes in a Tris buffs (20 mM Tris-Cl, 100 mM KCl, pH 8.0) and counterstained with DAPI. Images are obtained by overlaying Cy3 (red) and DAPI (blue) filter settings and adjusting the exposure. Nuclei were viewed at 60X magnification using a Nikon Eclipse T*i*-S inverted microscope. The scale bar represents 16 μm. For corresponding images for other Invader probes, see Figures 2.4-23, 2.4-24, and 2.4-25……42

Figure 2.2-6. Images from FISH experiments using *DYZ1*-targeting Invader probes **INV4** under denaturing (80 °C, 5 min) (left), or non-denaturing (3 h, 37.5 °C) conditions (right). Incubation and imaging specifications are described in Figure 2.2-5. Scale bar represents 16 µm……44

Figure 2.2-7. INV2 (left panel) and **INV10** (right panel) incubated in female bovine fibroblast cells under denaturing conditions. Note the lack of red Cy3 signal. Incubation and imaging specifications are described in Figure 2.2-5. Scale bar represents 16 µm……46

Figure 2.2-8. a) Representative gel electrophoretograms from recognition experiments at 15 h between a 100-fold molar excess of Optimized Invader probes **OPT6**, **OPT8**, and **OPT9** and their corresponding DNA hairpin targets **DH6**, **DH8**, and **DH9**. b) Histograms depict averaged results from at least three recognition experiments with error bars representing standard deviation. TC = ternary complex. DIG-labeled **DH6**, **DH8**, and **DH9** (34.4 nM, sequences shown in Table SX) were incubated with the corresponding Invader probe in HEPES buffer (50 mM HEPES, 100 mM NaCl, 5 mM MgCl2, pH 7.2, 10% sucrose, 1.44 mM spermine tetrahyrdochloride) for 15 h at 37 °C. Incubation mixtures were resolved on 12% non-denaturing TBE-PAGE slabs (~70 V, ~4 °C, ~1.5 h)……51

Figure 2.2-9. Dose-response curves for **OPT6, OPT8, and OPT9** at 37 °C for 15 h. Experimental conditions are as described in Figure 2.2-8, except for variable probe concentrations……..…52

Figure 2.2-10. Images from FISH experiments using DYZ1-targeting **OPT6**, **OPT8**, and **OPT9** under denaturing (80 °C, 5 min) (left), or non-denaturing (3 h, 38.5 °C) conditions (right). Incubation and imaging specifications are described in Figure 2.2-5 Scale bar represents 16 µm……..…54

Figure 2.4-1. MALDI-MS Spectra of Individual Strands (up and down) of Invader Probes **INV1**, **INV2**, **INV3**, and **INV4**……63

Figure 2.4-2. MALDI-MS Spectra of Individual Strands (up and down) of Invader Probes **INV5**, **INV6**, and **INV7**……..64

Figure 2.4-3. MALDI-MS Spectra of Individual Strands (up and down) of Invader Probes **INV8**, **INV9**, and **INV10**……65

Figure 2.4-4. MALDI-MS Spectra of Individual Strands (up and down) of Invader Probes **OPT6**, **OPT8**, and **OPT9**……..67

Figure 2.4-5. HPLC Spectra of Optimized Probes……..68

Figure 2.4-6. Position of sequences within the *DYZ-1* satellite gene on the bovine (*Bos taurus*) Y chromosome targeted by the different Invader probes.[26] The target sequence for **INV4** (shown as a green line) is presented six times within the tandem repeat (~6 x 10^4 tandem repeats); all other target sequences are present only once……69

Figure 2.4-7. Representative thermal denaturation curves for Invader probes **INV1**, **INV2**, **INV3**, and **INV5** and the corresponding duplexes between individual probe strands and cDNA, and unmodified reference DNA duplexes (**DNA1**, **DNA2**, **DNA3**, **DNA5**)……70

Figure 2.4-8. Representative thermal denaturation curves for Invader probes **INV6**-**INV10** and the corresponding duplexes between individual probe strands and cDNA, and unmodified reference DNA duplexes (**DNA6**-**DNA10**)……..71

Figure 2.4-9. Representative thermal denaturation curves for Optimized Invader probes **OPT6**, **OPT8**, and **OPT9** and the corresponding duplexes between individual probe strands and cDNA, and unmodified reference DNA duplexes (**DNA6**, **DNA8**, and **DNA9**)……72

Figure 2.4-10. Invader and corresponding hairpins with the Invaders at a 5-fold molar excess to the hairpin. Shows that the hairpin in indeed the lower band. All other conditions are the same as Figure 2.4-12……..79

Figure 2.4-11. Screen of 100-fold of corresponding Invader probes against all 10 hairpin targets (one complementary nine non-complementary) in each lane. Incubation for 2.5 hours at 37 °C, all other conditions are as reported in Figure 2.4-12. Shows that recognition of complementary hairpin is not hindered by presence of all other hairpins……...……79

Figure 2.4-12. a) Representative gel electrophoretograms from recognition experiments at 2.5 h between a 100-fold molar excess of Invader probes **INV1**-**INV10** and their corresponding DNA hairpin targets **DH1**-**DH10**. b) Histograms depict averaged results from at least three recognition experiments with error bars representing standard deviation. TC = ternary complex. DIG-labeled **DH1**-**DH10** (34.4 nM, sequences shown in Table 2.4-9) were incubated with the corresponding Invader probe in HEPES buffer (50 mM HEPES, 100 mM NaCl, 5 mM $MgCl_2$, pH 7.2, 10% sucrose, 1.44 mM spermine tetrahyrdochloride) for 2.5 h at 37 °C. Incubation mixtures were resolved on 12% non-denaturing TBE-PAGE slabs (~70 V, ~4 °C, ~1.5 h)……..80

Figure 2.4-13. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) and different concentrations of Invader probes a) **INV1**, b) **INV2**, c) **INV6**, and d) **INV7** at 37 °C for 2.5 h. Experimental conditions are as specified in Figure 2.4-12. For dose-response curves, see Figure 2.4-15……...81

Figure 2.4-14. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) and different concentrations of Invader probes a) **INV8**, b) **INV9**, and c) **INV10** at 37 °C for 2.5 h.. Experimental conditions are as specified in Figure 2.4-12. For dose-response curves, see Figure 2.4-15……..82

Figure 2.4-15. Dose-response curves for **INV1**, **INV2**, **INV6**, and **INV7** (upper panel), and **INV8**, **INV9**, and **INV10** (lower panel) at 37 °C for 2.5 h. Experimental conditions are as described in Figure 2.4-12, except for variable probe concentrations……..83

Figure 2.4-16. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) and different concentrations of Optimized Invader probes a) **OPT6**, b)

OPT8, c) **OPT9** at 37 °C for 2.5 h. Experimental conditions are as specified in Figure 2.2-8. For dose-response curves, see Figure 2.4-17……84

Figure 2.4-17. Dose-response curves for **OPT6, OPT8, and OPT9** at 2.5 h at 37 C. Experimental conditions are as described in Figure 2.2-8, except for variable probe concentrations……...……...……85

Figure 2.4-18. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) and different concentrations of Invader probes a) **INV1**, b) **INV2**, c) **INV3**, and d) **INV5** at 37 °C for 15 h. Experimental conditions are as specified in Figure 2.2-2. For dose-response curves, see Figure 2.2-3……86

Figure 2.4-19. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) and different concentrations of Invader probes a) **INV6**, b) **INV7**, and c) **INV8** at 37 °C for 15 h. Experimental conditions are as specified in Figure 2.2-2. For dose-response curves, see Figure 2.2-3……87

Figure 2.4-20. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) and different concentrations of Invader probes a) **INV9** and b) **INV10** at 37 °C for 15 h. Experimental conditions are as specified in Figure 2.2-2. For dose-response curves, see Figure 2.2-3……88

Figure 2.4-21. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) and different concentrations of Optimized Invader probes a) **OPT6**, b) **OPT8**, c) **OPT9** at 37 °C for 15 h. Experimental conditions are as specified in Figure 2.2-8. For dose-response curves, see Figure 2.2-9……89

Figure 2.4-22. Representative images of **INV1** and **INV2** from FISH experiments using *DYZ1*-targeting Invader probes under denaturing (80 °C, 5 min) (left), or non-denaturing (3 h, 38.5 °C) conditions (right). Incubation and imaging specifications are described in Figure 2.2-5. Scale bar represents 16 μm……..90

Figure 2.4-23. Representative images of **INV3**, **INV4** and **INV5** from FISH experiments using *DYZ1*-targeting Invader probes. Incubation and imaging specifications are described in Figure 2.2-5. Scale bar represents 16 μm……91

Figure 2.4-24. Representative images of **INV6**, **INV7** and **INV8** from FISH experiments using *DYZ1*-targeting Invader probes. Incubation and imaging specifications are described in Figure 2.2-5. Scale bar represents 16 μm……92

Figure 2.4-25. Representative images of **INV9** and **INV10** from FISH experiments using *DYZ1*-targeting Invader probes. Incubation and imaging specifications are described in Figure 2.2-5. Scale bar represents 16 μm……93

Figure 2.4-26. Representative Images for FISH experiments of nuclei pre-treated with DNase I prior to incubation with **INV2** and **INV10**. Note the absence of signal in the DNase I treatment indicating the Invader probes' targeting of the chromosomal DNA. Incubation and imaging specifications are described in Figure 2.2-5. Scale bar represents 16 μm……94

Figure 2.4-27. Representative Images for FISH experiments of nuclei pre-treated with RNase A and Proteinase K prior to incubation with **INV2** and **INV10**. Note the continued presence of signal in the RNase A and Proteinase K pre-treated nuclei (albeit there is a reduction in signal coverage which could be attributed to the loss of genetic material/number of nuclei during these enzyme treatments, not seen in DNase treatment). Incubation and imaging specifications are described in Figure 2.2-5 Scale bar represents 16 µm……95

Figure 2.4-28. Images from FISH experiments using DYZ1-targeting **OPT6**, **OPT8**, and **OPT9** under denaturing (80 °C, 5 min) (left), or non-denaturing (3 h, 37.5 °C) conditions (right). Incubation and imaging specifications are described in Figure 2.2-5. Scale bar represents 16 µm……..96

Figure A.2-1. Gating strategy utilizing DAPI (PB450-A) and Cy3 (PE-H) channels. Anaylzed nuclei (colored peaks) were incubated with **INV10** in 30-fold excess of target sequence……103

Figure A.2-2. Probe Cy3 intensity (PE-H) values for nuclei populations (count) incubated with a) **INV4**, b) **INV4D**, and c) **MM4**. Grey peak represents debris in sample. Colored peaks are isolated nuclei……105

Figure A.2-3. Comparison of Mean Intensities between Unmodified **INV4D**, Mismatch **MM4**, and **INV**4. Mean intensity values for each probe were taken from three separate populations (~3000 nuclei) incubated in 30-fold excess of **INV4** target sequence……...106

Chapter 1: DNA Targeting Agents in the Context of Imaging and Detection of Chromosomal DNA

Caroline Shepard

1.1 DNA as a Target

The critical role of DNA

DNA is the central molecule of life, providing the code for the structure and development of living organisms. Housed in the nucleus of cells, chromosomal DNA is the source of protein synthesis, cell regulation, and function. Based on the order of adenine (A), thymine (T), cytosine (C), and guanine (G) nucleobases that constitute genes, or rather the DNA's 'code,' RNA polymerases can 'read' the bases to synthesize pre-mRNA. Subsequently, this is processed into mature mRNA to be translated into usable structure and function proteins.[1] The genetic flow from DNA to protein synthesis is entitled the Central Dogma of Molecular Biology as it represents the core foundation for life. As the core, abnormalities occurring at any point along this path can be detrimental for living things. For instance, unwanted or unusual expression of genes can result in negative phenotypic traits and disease. This is the case for those with suffering from Huntington's disease as an expansion of a trinucleotide (CAG) repeat region on human Chromosome 4 results in a mutant, toxic huntingtin protein that causes severe motor and cognitive deficits.[2] Moreover, Duchenne Muscular Dystrophy (DMD) is caused by exon skipping, resulting in misprocessed pre-mRNA into mature mRNA. Once ribosomes translate this incorrect mRNA, it forms dysfunctional protein that leads to muscle breakdown.[2] For such diseases, the development of detection tools and treatments are a constant focal point for scientific and biomedical research, with much effort focused on the targeting of proteins. Despite much success, this work is limited as many

diseases are not the result of mis-expressed proteins, and the majority of transcribed DNA is not translated into protein. This incentivizes moving focus 'upstream' to the nucleic acid level to increase the number of targets within DNA and RNA. Additionally, the predictability of nucleic acid structures (i.e. a sugar-phosphate backbone attached to the four nucleobases) (Figure 1.1-1) provides easier determination of targeting molecules in comparison to the complexity of protein structures taking on any number of conformations. Currently, there is extensive work in the development of RNA targeting;[4] however, focusing further upstream at the DNA level could prove more advantageous. For RNA targeting agents, higher concentrations of probe must be utilized to match the thousands of RNA existing in a cell. If DNA was the target, theoretically, there would only be a need for two targeting molecules per cell as an individual gene is present twice within chromosomal DNA. This brings much interest to DNA for potential diagnostic and therapeutic applications.

Structure of DNA

In order to target DNA, its unique double-stranded structure must be understood (Figure 1.1-1). Each strand is constructed of repeating nucleotide units supported by a negatively charged phosphodiester backbone. Nucleotide monomer units consist of a phosphate group, a 2'-deoxy-B-D-*erythro*-pentosefuranose, and a nucleobase (A, C, G, and T) connected to the C1 of the furanose sugar ring. C and T are termed pyrimidines for their 6-membered ring structure. The bicyclic nucleobases, G and A, are classified as purines.[5] These units determine the DNA's 'code,' which RNA polymerases transcribe into RNA that will eventually be translated into proteins or be used as functional units themselves. Hydrogen bonds connect complementary nucleobases on the opposite strand to form the double-stranded structure. Determined by the location of hydrogen bond donor and acceptor atoms,

A will bind to T (A:T) and C will bind to G (C:G).[5] These are known as Watson-Crick (WC) base pairs (bps). Strands hybridize in an antiparallel fashion, where the 5' end of one strand will align with 3'end of the other strand. This predictability in structure makes DNA an excellent target. If the sequence (i.e. order of nucleobases) is known, probes can be designed with the target's complementary base pair.

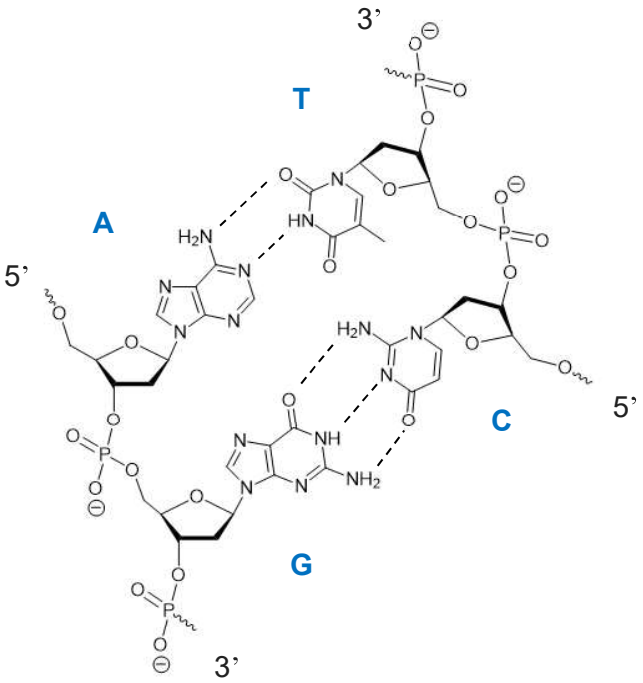


Figure 1.1-1. Structure of DNA and Watson-Crick base pairing (dashed lines denotes hydrogen bonds)

The formed duplex can take on different 3D helical structures, termed: A-type, B-type, and Z-type.[5,6] All three helical geometries form double-stranded duplexes determined by base complementarity, stabilization by intra- and inter-strand π-πstacking of the planar nucleobases, hydration, and positive counterions preventing the negatively charged strands from repelling each other. They differ in the handedness of the helix (right or left), tightness of twist, number of base pairs per turn of helix, and size and depth of minor and major grooves. Chromosomal DNA takes on the B-type structure, which is characteristically: right-

handed with narrow, deep minor grooves and wide, deep major grooves to give 10 bps per turn. Furthermore, the sugar ring plays a role in determining the 3D structure. In B-type, the furanose ring adopts a South Type conformation, or 'sugar pucker,' where the C2 moves out of the ring's plane in an *endo* position[6-8]. Consideration of sequence and overall helical geometry is used to design DNA targeting probes capable of forming Watson-Crick base pairs and binding within the major and minor grooves.

Considerations for the advancement to target chromosomal DNA

There has been varying levels of success in targeting chromosomal DNA with consideration to biological challenges and diagnostic relevance.[9-12] This is no easy feat as i) chromosomal DNA is housed in the nucleus of mammalian cells which implicates accessibility issues, requiring the targeting tool to cross the cell and nuclear membranes, ii) DNA is tightly packed into chromosomes. The duplex is supercoiled, wrapped around histones, and further compacted into chromatin that constitute chromosomes, raising further questions of target accessibility, and iii) double-stranded DNA (dsDNA) forms a stable duplex, raising a myriad of questions on how and where on the DNA structure to target.[5,13] Discussion of established DNA targeting designs follows in the subsequent section, outlining how these questions have addressed, with a focus on DNA detection and imaging. Before designing DNA targeting tools for *in vivo* use, their ability to recognize dsDNA with high specificity must be determined. Detection and imaging results highlight their potential as diagnostic tools and successful recognition gives promise to future *in vivo* and therapeutic applications.

1.2 History of DNA Targeting Designs

The field of Nucleic Acid Chemistry has sought to design dsDNA targeting probes (Figure 1.2-1), beginning with the development of Triplex Forming Oligonucleotides (TFOs). It was

noted in 1957 that naturally occurring polyadenylic acid (a single-stranded oligonucleotide consisting of only A nucleobases) will hybridize to polyuridylic acid (a single-stranded oligonucleotide made up of only U nucleobases). Another polyuridylic acid strand was shown to bind to this duplex, forming a 1:2 triplex.[14] Based on this observation, synthetic oligonucleotides termed TFOs were designed to mimic this structure, binding as a third strand to dsDNA.[15,16] Single-stranded in structure, TFOs form hydrogen bonds to short polypurine segments of dsDNA in the major groove. These contacts are known as Hoogsteen base pairs, and can be formed as the TFO aligns in a parallel (Hoogsteen) or anti-parallel (reverse Hoogsteen) orientation relative to its polypurine target.

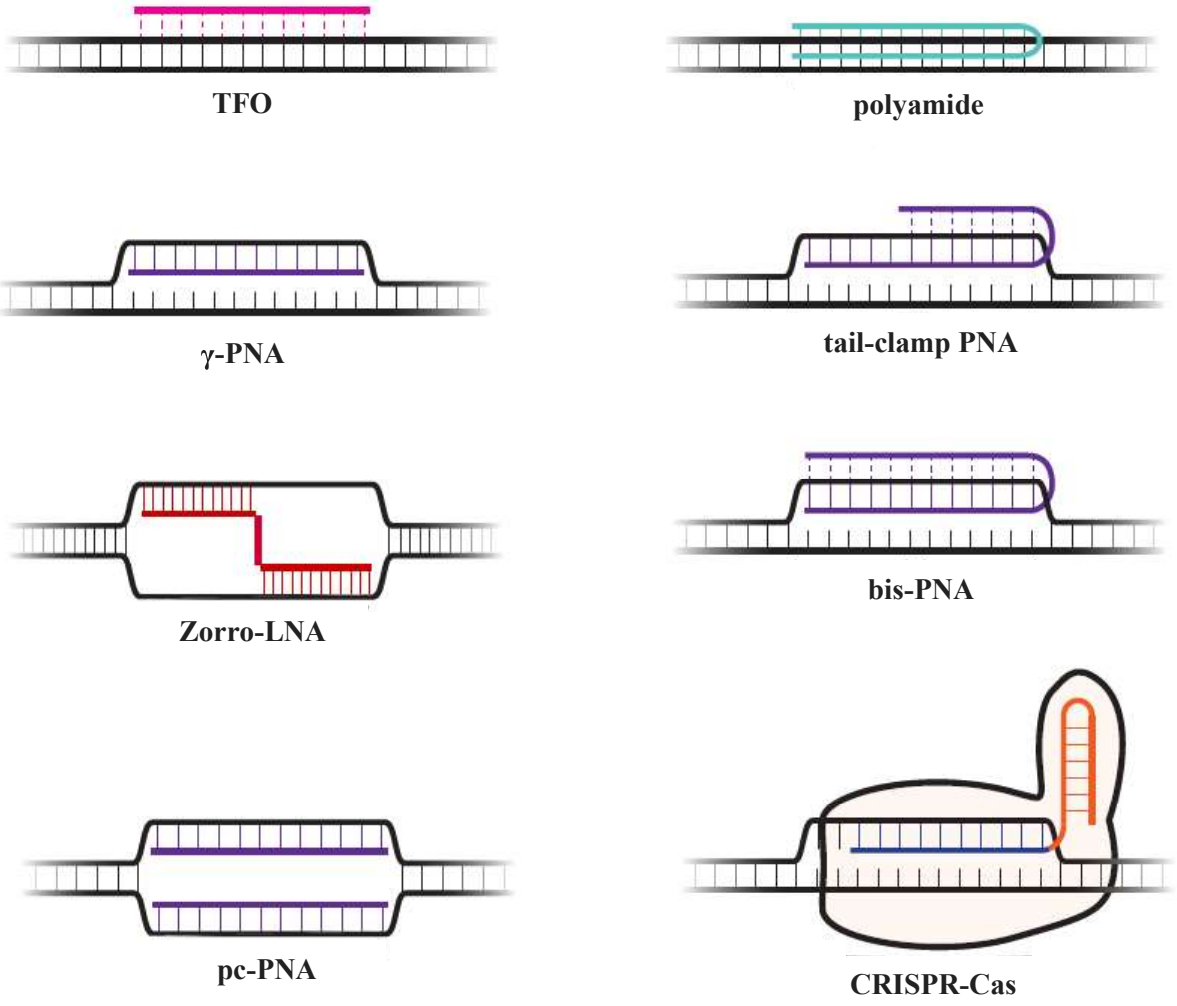


Figure 1.2-1. Schematic of current dsDNA targeting approaches

There have been many successful applications of TFOs in the targeting of chromosomal dsDNA, including, and not limited to: mutagenesis and modulation of gene expression.[17,18] Moreover, these constructs have shown use in detection and imaging studies. For instance, fluorescent TFOs were synthesized for *in situ* hybridization assays to target the chromosomal DNA of isolated, denatured and non-denatured metaphase and interphase nuclei. Termed TISH (Third-strand *in situ* hybridization), this assay was used to labeled specific sequences within human and mouse chromosomes.[19]

Despite these successes, TFOs suffer many drawbacks. In order to have unlimited targeting of chromosomal DNA, probes must be able efficiently recognize any sequence context at physiological relevant conditions (i.e. cellular environments). The polypurine sequence requirement severely limits the number of potential targets, which may preclude TFOs from detecting genes and sequences of interest. Additionally, the environmental pH must be acidic (<6.5) in order to protonate the cytosine TFO nucleobases to form the C+:CG triplets. This requirement makes TFOs suboptimal at a more neutral physiological pH. Much work has been done to overcome these challenges, such as modifying strands with crossover moieties to recognize sequences with an interrupting pyrimidine in an otherwise polypurine region, development of 5-methylcytosine (mC) to waive the pH requirement, reduction of charge repulsion between the negatively charged TFO and negatively charged dsDNA with positive charge addition to nucleobases, and increasing the overall affinity to target through affinity-enhancing modifications such as Locked Nucleic Acids (LNA).[20-23] These advancements have had varying levels of success, leaving much room for the development of new probe designs.

To address sequence limitations, minor-groove binding polyamides were developed. Non-nucleotidic monomer units *N*-methyl pyrrole (Py), hydroxy *N*-methyl pyrrole (Hp), *N*-methyl imidazole (Im) are joined by amide bonds to constitute this short, minor groove binding probe (Figure 1.2-1). Each subunit binds to a different DNA base pair (Py → A or C, Hp → T, Im → G) based on shape complementarity.[24] Due to their unique binding mode, polyamides can recognize mixed sequences, providing an advantage of TFOs. Particularly, Py/Im polyamides have shown potential in detecting chromosomal DNA targets. For example, pericentromeric heterochromatin (important for gene expression and cellular

differentiation) in murine chromosomal DNA was detected in a fluorescent *in situ* hybridization (FISH) assay with FITC-labeled pyrrole-imidazole polyamide probes as well as telomere labeling.[25,26] However, this design leaves much to be desired as they cannot distinguish between AT and TA base pairs, which raises questions about target fidelity and specificity. Moreover, they are most useful against short sequence targets (<8 bp). As the length of the target increases, it loses its shape complementarity and binding capabilities.[12, 24,27-29]

There has been extensive work in designing Peptide Nucleic Acid (PNA) probes (Figure 1.2-1) that have been used to target chromosomal DNA for purposes such as: manipulation of gene expression, site-specific mutagenesis, and dsDNA labeling.[10,29,30] In one study, this design's DNA detection potential was highlighted by fluorescently labeling Chromosome 9 under denaturing conditions in a PNA-COMBO-FISH assay.[31] Similar to TFOs, these probes can form triplexes in a parallel or anti-parallel fashion with double-stranded polypurine DNA sequences. Nucleobases (A, C, G, T) are attached to an achiral, uncharged N-(2-aminoethyl)glycine backbone (Figure 1.2-2). The lack of charge facilitates fast association and immensely high binding affinity with negatively charged DNA, offering an advantage over other negatively charged probes. Triplex formation is the preferred binding mode at physiological conditions and low PNA concentrations, which confers the same targeting restrictions to polypurine regions.[32]

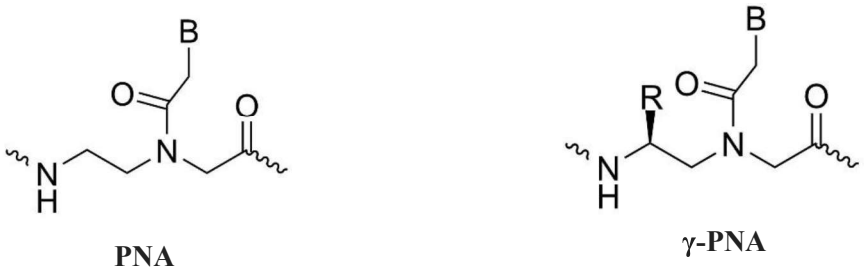


Figure 1.2-2. Structures of PNA and γ-PNAs. (B denotes nucleobases A, C, G, and T)

However, PNAs are a diverse class of probes that display binding modes.[32] Triplex formation can facilitate access the Watson-Crick faces of dsDNA, which is termed triplex invasion. During invasion, the hydrogen bonds of the target duplex are disrupted and the nucleobases of the PNA probe form classical Watson-Crick hydrogen bonds with the target strand. In the case of triplex invasion, one PNA strand will form a triplex with a polypurine target, disrupting the dsDNA duplex to allow the hybridization of a second PNA strand with Watson-Crick base pairing. As a result, the non-complementary, adjacent target strand, is displaced, forming a single-stranded loop. Triplex invasion has been further enhanced by tethering two PNA strands with a linker to form bis-PNA (Figure 1.2-1).[30] Linking the two strands increases the local concentration of the PNA and reduces entropic costs to facilitate efficient binding. This has been used to indirectly detect chromosomal and mitochondria DNA with a combined incubation of bis-PNA and an additional padlock probe.[33] Tail-clamp PNA bind similarly to bis-PNA; however, the PNA strand meant to form Watson-Crick base pairs is extended to improve binding affinity and confer some level of mixed-sequence recognition in this extended region limitations (Figure 1.2-1).[34] Unfortunately, these designs suffer similar polypurine sequence limitations as TFOs as triplex invasion necessitates triplex formation.

The targeting of mixed sequences is the 'gold standard' for nucleic acid probes. To address past sequence limitations, next generation single-stranded γ-PNAs were developed. Introduction of chirality through substitution at the γ carbon position of the glycine backbone, enables the recognition of mixed sequences (Figure 1.2-2).[35,36] This additional chiral group preorganizes the PNA strand to match the chiral right-handed conformation of dsDNA, enabling the probe to overcome entropic costs to invade B-type DNA. Moreover, its improved target affinity is facilitated by reduction in electrostatic repulsion due to its uncharged backbone. This probe binds through duplex invasion, displacing one of the target DNA strands to form Watson-Crick base pairing with one of the target strands (Figure 1.2-1). This differs from triplex invasion, as the formation of an initial triplex is not needed. Several designs and optimizations have been deployed such as G-clamps, diethylene glycol substitution, addition of acridine units, and increasing probe length.[37-39] Improvements have resulted in γ-PNAs capable of mixed-sequence DNA detection (e.g. chromosomal labeling of telomeres).[40] However, current efforts are still solving aggregation, solubility issues, and poor cellular uptake issues in order to move to future *in vivo* studies.

There is another design to note that has resulted in mixed-sequence recognition of DNA. Zorro-LNA is also two-stranded probe where the strand ends hybridize to form a linker, leaving their 3'-ends single-stranded, forming a "Z" shaped probe (Figure 1.2-1). For the classic Zorro-LNA structure, each strand must have an inversion of polarity (i.e. align anti-parallel) in order to hybridize in the linker region. Additionally, affinity enhancing units, such as LNA, must be incorporated within the linker and single-stranded overhang in order to bind to target. There have been developments of other architectures such as single-stranded Zorro-LNA, single-stranded Zorro-LNA with a stiffener strand, and utilization of non-

nucleotide linkers.[41] There is evidence of its capabilities in recognizing mixed-sequence dsDNA targets, even so far as to inhibit transcription at the plasmid level. However, this design requires a pre-annealing step, presents issues of self-hybridization due to the affinity enhancing LNA monomers, and there is evidence that the binding mechanism might require initial triplex formation of one probe molecule to facilitate invasion.

The final binding mode that has been utilized and will serve as the basis for the research discussed later herein, is double-duplex invasion. Differing from duplex invasion, each strand of a double-stranded probe will form Watson-Crick base pairs with its complementary target strand. Both target strands hybridize to the probe, leaving no displaced strand as is in the cases of triplex invasion and duplex invasion. Pseudocomplementary (pc) PNA probes take advantage of this approach mode (Figure 1.2-1).[42,43] This novel design is constructed as two hybridized PNA strands forming a duplex. To reduce affinity towards its complementary PNA strand and to increase each PNA strand's affinity for its DNA target strand, A and T nucleobases are replaced with 2,6-diaminopurine (D) and 2-thiouracil (S) (Figure 1.2-3). When these monomers are placed across from each other on opposing strands of the double-stranded probe, the NH_2 of the D and the sulfur of the S clash due to sterics, resulting in destabilization. Thus, the more labile probe is 'activated' to follow a thermodynamic gradient (i.e. unstable to stable) driving recognition. As it encounters its target, the labile probe strands easily displace from each other. The D and S monomers of the now single-stranded pc-PNA serve to enhance each strands affinity to its complementary DNA (cDNA) strand, driving recognition. There has been reported success in targeting chromosomal DNA under molecular crowding conditions and for the purpose of gene modification in mouse fibroblast cells.[44,45] However, this design requires a high level of modification per strand (i.e. many D

and S units) and low ionic strength to increase lability of target site, especially region rich in GC-content, to facilitate recognition.

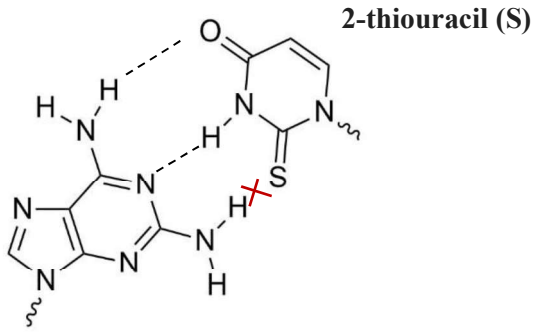


Figure 1.2-3. Pseudocomplementary (pc) PNA base pair

Worth mentioning are engineered protein designed to target dsDNA. CRISPR Cas has made a recent debut as a potential tool for gene detection and manipulation (Figure 1.2-1). It is a cluster of regulatory interspaced short palindromic repeats within CRISPR associated proteins which have an RNA-guided nuclease function. Originally utilized by bacteria in virus immune responses, this construct can be 'hijacked' and redesigned to target chromosomal DNA for the purposes of gene manipulation and detection.[46,47] For example, a CASFISH assay was developed by fluorescently labeling nuclease-deactivated Cas9 protein. With this design, pericentromere, centromere, G-rich telomere, and different coding genes in genomic DNA were fluorescently labeled and detected.[48] While there is success in *in vitro* assays, engineered proteins face many challenges such as inadequate specificity, limited targets, and *in vivo* hurdles with stimulation of immunogenic responses and problematic delivery.

1.3 Introduction to Invaders

With the targeting and environmental challenges posed in other DNA targeting probes, there is a need to design tools that accurately recognize mixed sequence dsDNA at physiological relevant conditions for diagnostic and therapeutic purposes. To meet this need, our lab has developed double-stranded probes, termed Invaders. They follow a similar thermodynamic driving force for recognition as pc-PNA (Figure 1.3-1). The strategic placement of 2'-*O*-(pyren-1-yl)methyl RNA monomers creates a destabilized oligonucleotide probe, activated for target recognition. By arranging these monomers in a +1 interstrand architecture (termed energetic hotspots), the pyrene moieties are forced to occupy the same space, violating the nearest neighbor exclusion principle (NNEP). The ingenuity of this design in part due to the use of pyrenes as DNA intercalators. They slide in between base pairs and remain stably bound in this region by π-π stacking interactions. According to the NNEP, intercalators are limited to binding every second base pair as there is a limit to the extent that the DNA helix can expand to accommodate intercalators. Violating this rule and forcing the pyrenes to occupy the space between the same base pairs, leading to a less stable duplex. As the Invader encounters its target, the probe will easily denature into its individual strands and bind with high affinity to their respective complementary strands. The stacking interactions by the pyrene moieties of individual probe strands help create stable probe-target duplexes (i.e. recognition complex), driving double duplex invasion. stability.[49-51]

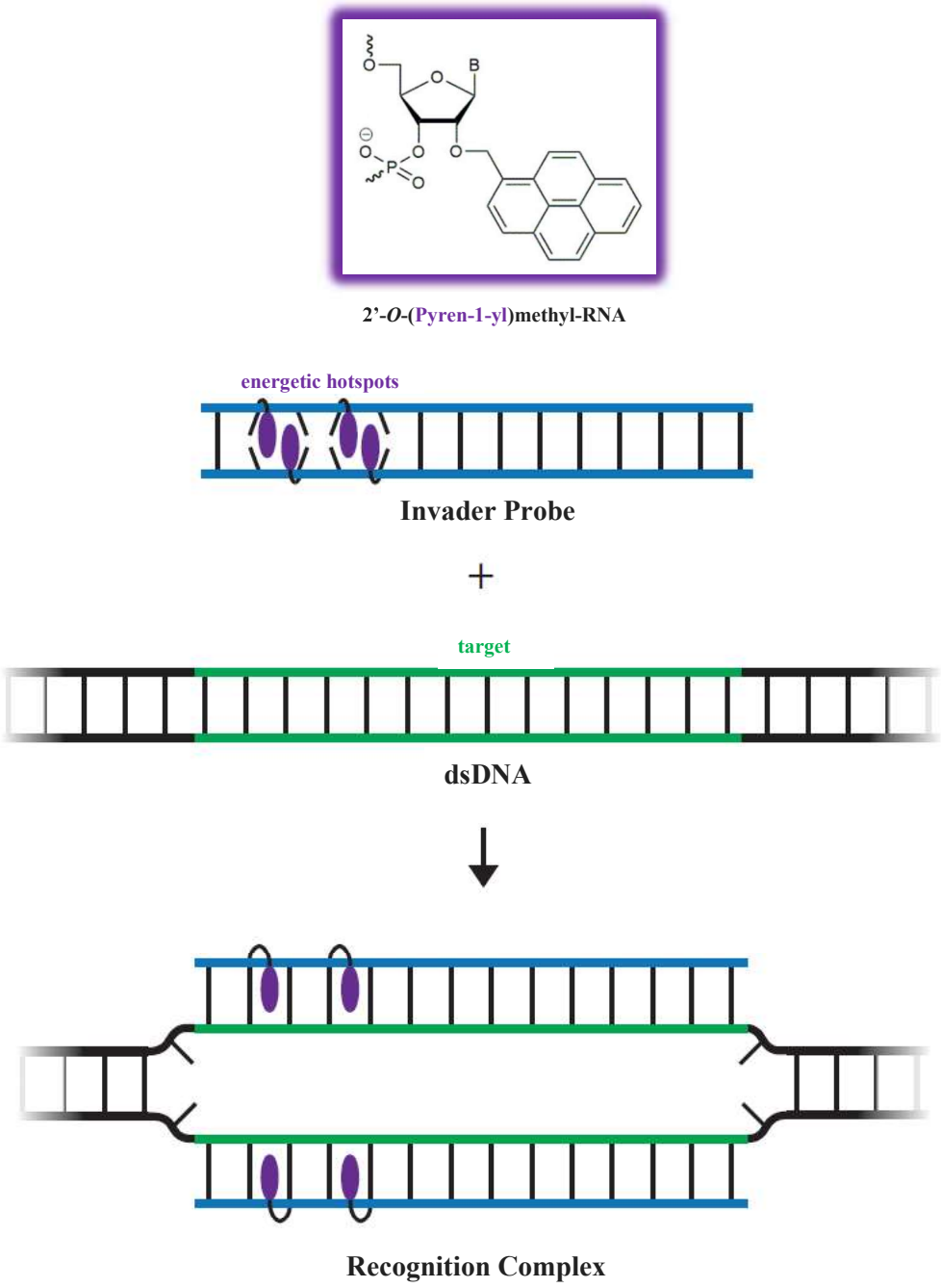


Figure 1.3-1. Double duplex invasion mechanism of Invader probes.

1.4 Development of Efficient Invaders

The recognition of mixed-sequenced DNA by Invader probes is driven by the thermodynamic gradient determined by a destabilized probe moving to a highly stable

recognition complex. The origins of Invader probes began with development and incorporation of 2'-amino-L-α-LNA monomers.[52] A 2′-C,4′-C-oxymethylene linker 'locks' the ribose ring into the *C2'-endo* sugar conformation. The use of a conformationally restricted oligonucleotides enhances the affinity of the probes as the oligonucleotide is restricted to a conformation that matches its target, leading to a reduction of entropic cost. Additionally, intercalators attached to these monomers are in a favorable position and conformation for intercalation once bound to its target. While these locked nucleic acids offer efficient targeting, generation of these probes are synthetically difficult which led to the development of the Invader chemistry. Though more flexible than their LNA predecessor, the 2'-*O*-(pyren-1-yl)methyl-RNA monomers (Figure 1.3-1), they retain adequate hybridization properties.[49]

Upon development of Invaders, it was crucial to establish design rules and determination of optimal probe architectures. First, the effect of sequence context of energetic hotspots was evaluated to determine if the nucleobase (B) of the Invader monomer and/or its unmodified flanking base pairs within its sequence affected recognition. To this end, 2'-*O*-(pyren-1-yl)methyl-RNA monomers **A**, **T**, **C**, and **G** were incorporated into Invader probe strands.[53] It was discovered that the **G** was less stabilizing than the other three monomers, which may be attributed to weakened **G**:C base pair within the recognition, rather than a lack of intercalation by the pyrene. Furthermore, it was shown that these monomers provided more stabilization in the recognition complex when flanked by a 3'-purine. Lastly, probes incorporated with **U** and **C** monomers showed greater potential for dsDNA recognition. These studies laid the groundwork for all future Invader probes, necessitating use of **U** and **C**

with limited use of **A**, and an avoidance of **G** incorporation. Additionally, incorporated monomers should by flanked by a 3'-purine.

Second, the optimal level of modification was evaluated to determine the required number of energetic hotspots to facilitate efficient dsDNA recognition. Recent studies have shown the need to balance high binding affinity with binding specificity. A small number of hotspots gives low binding affinity, while increasing the number of incorporations reduces specificy.[54,55] Probes with 15-30% of their sequence modified with Invader monomers display adequate affinity and specificity towards their dsDNA targets.

1.5. Invader Probe Applications

DNA targeting tools have great diagnostic application. Current studies have shown the potential of Invaders for these purposes. They can detect different pathogenic bacteria that are common food contaminates through use of sandwich assay.[50] In this study, Invader probes were immobilized on a plate. Bacterial DNA was introduced and hybridized to the immobilized duplexes, leaving a single-stranded region of bacterial DNA. Subsequently, an additional biotin-labeled Invader probe was added to bind to the available base pairs of the single-stranded region. After a wash step, streptavidin-horseradish peroxidase conjugate and substrate were added to produce a fluorescent signal that can be used for quantification. This detection could be advantageous in food preparation as Invader probes have mixed sequence recognition capabilities (i.e. target sequence is not a limitation), avoid laborious sample preparation, occurs isothermally, and is highly sensitive (capable of recognition at 20 pM). Results and design of this assay have the potential to expand to other fields for medical diagnostic and biotechnological applications.

Extensive work has been done to evaluate the use of these Invaders in non-denaturing fluorescent *in situ* hybridization (nd-FISH) assays in the detection of chromosomal DNA. Towards this end, Invader probes have been Cy3-labeled in order to be visualized by fluorescent microscopy. These fluorescently labeled probes are incubated with fixed nuclei on microscope slides under physiological relevant conditions. Following incubation, nuclei are DAPI counterstained in order to overlay the Cy3-labeled probe signal to determine the probe's binding to chromosomal DNA. This technique has been shown to effectively target the telomeres and Y chromosome of bovine cells.[54,55] Control experiments have verified these findings through relevant mismatch studies, incubation with female bovine fibroblast cells (displayed no signal), pre-treatment with DNase (disappearance of signal), and pre-treatment of RNase and proteinase (signal is maintained). This represents an advancement in DNA targeting probes as Invaders can: i) recognize mixed-sequence chromosomal DNA, ii) target in non-denaturing conditions suggesting that target sequences do not need to be denatured in order to bind, and iii) recognize chromosomal DNA near physiological conditions.

1.6 Conclusion

Mixed sequence double-stranded DNA has remained an elusive target for nucleic acid targeting probes for diagnostic, detection, and biological applications. However, the advantages in effective targeting of DNA fuel the development of usable probes. Our Invader probe design utilizes the prior knowledge of previous DNA targeting approaches to take the next step in that direction. The ability to target mixed-sequence chromosomal DNA under physiological conditions gives immense promise for future diagnostic and detection applications as well as biological applications.

1.7 References

1. T. Schneider-Poetsch and M. Yoshida, *Annu. Rev. Biochem.*, 2018, **87**, 391-420.
2. M. Jackson, L. Marks, G. May and J. B. Wilson, *Essays Biochem.*, 2018, **62**, 643-723.
3. M. Pertea, *Genes*, 2012, **3**, 344-360.
4. S. T. Crooke, J. L. Witztum, C. F. Bennet and B. F. Baker, *Cell Metabolism*, 2018, **27**, 714-739.
5. A. Travers and G. Muskhelishvili, *FEBS Journal*, 2015, **282**, 2279-2295.
6. R. Dickerson, H. Drew, B. Connor, R. Wing, A. Fratini, and M. Kopka, *Science*, 1982, **216**, 475-485.
7. A. Rich, *Nat. Struct. Mol. Biol.*, 2003, **10**, 247-249.
8. S. Arnott. *Trends Biochem. Sci.*, 2006, **31**, 349-354.
9. I. Ghosh, C. I. Stains, A. T. Ooi and D. J. Segal, *Mol. Biosyst*., 2006, **2**, 551-560.
10. P. E. Nielson, *Biodiv*., 2010, **7**, 786.
11. Y. Aiba, J. Sumaoka and M. Komiyama, *Chem. Soc. Rev*., 2011, **40**, 5657-5668.
12. T. Vaijayanthi, T. Bando, G. N. Pandian and H. Sugiyama, *ChemBioChem*., 2012, **13**, 2170
13. M. Radman-Livaja and O. J. Rando, *Developmental Biology*, 2010, **339**, 258-266.
14. G. Felsen, D. R. Davies and A. Rich. *J. Am. Chem. Soc*., 1957, **79**, 2023-2024.
15. M. Duca, P. Vekhoff, K. Oussedik, L. Halby and P. B. Arimondo*, Nucleic Acids Res*., 2008, **36**, 5123-5138.
16. S. Buchini, *Curr. Opin. Chem. Biol.*, 2003, **7**, 717-736.
17. K. M. Vasquez, L. Narayanon and P. M. Glazer, *Science*, 2000, **290**, 530-533.

18. K. R. Fox, D. A. Rusling, V. J. Broughton-Head and T. Brown, *Curr. Chem. Biol.*, 2008, **2**, 1-10.

19. M. D. Johnson III and J. R. Fresco, *Chromosoma*, 1999, **108**, 181-189.

20. Y. Hari, M. Akabane and S. Obika, *Chem. Commun*., 2013, **49**, 7421-7423.

21. S. F. Singleton and P. B. Dervan, *Biochemistry*, 1992, **31**, 10995-11003.

22. A. S. Cardew, T. Brown and K. R. Fox, *Nucleic Acids Res.,* 2011, **40**, 3753-3762.

23. S. P. Sau, P. Kumar, B. A. Anderson, M. E. Østergaard, L. Deobald, A. Paszczynski, P. K. Sharma and P. J. Hrdlicka, *Chem. Comm*., 2009, 6756-6758.

24. S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird and P. B. Dervan, *Nature*, 1998, **391**, 468-471.

25. A. Sasaki, S. Ide, Y. Kawamoto, T. Bando, Y. Murata, M. Shimura, K. Yamada, A. Hirata, K. Nokihara, T. Hirata, H. Sugiyama and K. Maeshima, *Sci. Rep*., 2016, **6**, 29261.

26. Y. Kawamoto, T. Bando and H. Sugiyama, *Bioorg. Med. Chem*, 2018, **26**, 1393-1411,

27. P. B. Dervan and B. S. Edelson, *Curr. Opin. Struct. Biol*., 2003, **13**, 284-299.

28. M. S. Blackledge and C. Melander, *Bioorg. Med. Chem*., 2013, **21**, 6101-6114.

29. P. E. Nielsen, M. Egholm and O. Buchardt, *Science*, 1991, **254**, 1497-1500.

30. K, Kaihatsu, B. A. Janowski and D. R. Corey, *Chem. Biol*., 2004, **11**, 748-758.

31. P. Müller, J. Rößler, J. Schwarz-Finsterle, E. Schmitt and M. Hausmann, *Experimental Cell Res.*, 2016, **345**, 51-59.

32. M. E. Hansen, T. Bentin and P. E. Nielson, *Nucleic Acid Res*., 2009, **37**, 4498-4507.

33. A. I. Yaroslavsky, I. V. Smolina, *Chem. Biol,* 2013, **20**, 445-453.

34. T Bentin, H. J. Larsen and P. E. Nielson, *Biochemistry*, 2003, **42**, 13987-13995.

35. R. Bahal, B. Sahu, S. Rapireddy, C. M. Lee and D. H. Ly, *ChemBioChem*, 2012, **13**, 56-60.

36. S. Rapireddy, R. Bahal and D. H. Ly, *Biochemistry*, 2011, **50**, 3913-3918.

37. S. Rapireddy, G. He, S. Roy, B. A. Armitage and D. H. Ly, *J. Am. Chem. Soc*., 2007, **129**, 15596-15699.

38. V. Chenna, S. Rapireddy, B. Sahu, C. Ausin, E. Pedroso and D. H. Ly, *ChemBioChem* 2008, **9**, 2388-2391.

39. S. He, S. Rapireddy, R. Bahal, B. Sahu and D. H. Ly, *J. Am. Chem. Soc*., 2009, **131**, 12088-12090.

40. H. H. Pham, C. T. Murphy, G. Sureshkumar, D. H. Ly, P. L. Opresko and B. A. Armitage, *Org. Biol. Chem*., 2014, **12**, 7345-7354.

41. E. M. Zaghloul, A. S. Madsen, P. M. Moreno, I. I. Oprea, S. El-Andaloussi, B. Bestas, P. Gupta, E. B. Pedersen, K. E. Lundin, J Wengel and C. I. Smith, *Nucleic Acids Res*., 2011, **39**, 1142-1154.

42. V. V. Demidov, E. Protozanova, K. I. Izvolsky, C. Price, P. E.Nielsen and M. D. Frank-Kamenetskii, *Proc. Natl. Acad. Sci.*, 2002, **99**, 5953-5958.

43. J. Lohse, O. Dahl and P. E. Nielsen, *Proc. Natl. Acad. Sci*., 1999, **96**, 11804-11808.

44. J. Sumaoka and M. Komiyama, *Chem. Lett*., 2014, **43**, 1581-1583.

45. P. Lonkar, K.H. Kim, J.Y. Kuan, J.Y. Chin, F.A. Rogers, M.P. Khauert, R. Kole, P.E. Nielsen and P.M. Glazer, *Nucleic Acids Res.*, 2009, **37**, 3635-3644.

46. A. C. Komor, A. H. Badran and D. R. Liu, *Cell*, 2017, **168**, 20–36.

47. P. D. Hsu, E. S. Lander and F. Zhang, *Cell*, 2014, **157**, 1262– 1278.

48. W. Deng, X. Shi, R. Tjian, T. Lionnet and R. H. Singer, *Proc. Natl. Acad. Sci.*, 2015, **112**, 11870-11875.

49. P. A. Sau, A. S. Madsen, P. Podbevsek, N. K. Anderson, T. S. Kumar, S. Anderson, R. L. Rathju, B. A. Anderson, D. C. Guenther, S. Karmakar, P. Kumar, J. Plavec, J. Wendel and P. J. Hrdlicka, *J. Org. Chem*., 2013, **78**, 9560-9570.

50. B. Denn, S. Karmakar, D. C. Guenther and P. J. Hrdlicka, *Chem. Commun*., 2013, **49**, 9851-9852.

51. B. A. Didion, S. Karmakar, D. C. Guenther, S. P. Sau, J. P. Verstegen and P. J Hrdlicka, *ChemBioChem*, 2103, **14**, 1534-1538.

52. P. J. Hrdlicka, T. S. Kumar and J Wengel, *Chem. Commun*., 2005, 1534-1538.

53. S. Karmakar, D. C. Guenther and P. J. Hrdlicka, *J. J. Org. Chem*., 2013, **78**, 12040-12048.

54. D. C. Guenther, G. H. Anderson, S. Karmakar, B. A. Anderson, B. A. Didion, W. Guo, J. P. Verstegen and P. J. Hrdlicka, *Chem. Sci*., 2015, **6**, 5006-5015.

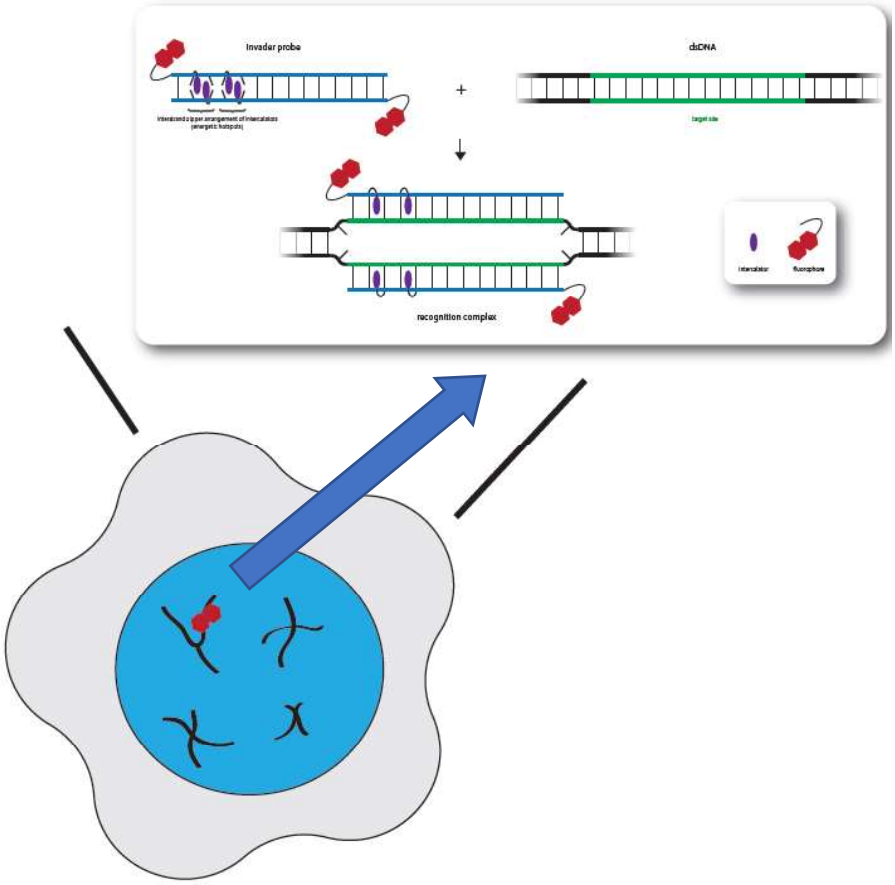
55. R. G. Emehiser, E. Hall, D. C. Guenther, S. Karmakar and P. J. Hrdlicka, *Org. Biol. Chem.*, 2020, **18**, 56-65.

Chapter 2: Recognition of Chromosomal DNA Targets Using Invader Probes

Caroline P. Shepard, Raymond G. Emehiser, Saswata Karmakar, and Patrick J. Hrdlicka*

Department of Chemistry, University of Idaho, Moscow, Idaho 83844-2343, USA

E-mail: hrdlicka@uidaho.edu



Abstract

Future advancements in chemical biology will necessitate probes capable of targeting specific sites on chromosomal DNA to regulate biological processes and aid in diagnostic applications. Established probe technologies such as triplex-forming oligonucleotides, peptide nucleic acids (PNAs), and minor-groove-binding polyamides only detect DNA under limited experimental conditions (homopurine targets; denaturing steps; low ionic strengths; short target sequences), whereas the CRISPR-Cas9 approach requires transfection of plasmids encoding CRISPR-Cas9 components. To address these shortcomings, our laboratory has developed Invader probes, which are energetically activated for sequence-unrestricted DNA recognition. Placement of 2'-*O*-(pyren-1-yl)methyl RNA monomers in +1 interstrand zipper arrangements destabilizes the probe duplex as the intercalating pyrene moieties *vie* for the same space between two Watson-Crick base pairs. These energetic hotspots, in concert with the high affinity for complementary DNA (cDNA) displayed by individual probe strands, provides the driving force for recognition of the target site (Figure 2.1-1). Herein, we evaluated ten Invader probes designed against different targets within the DYZ-1 satellite region (~6 x 10^4 tandem repeats) of the bovine (Bos Taurus) Y chromosome. Fluorescence *in situ* hybridization (FISH) assays conducted under non-denaturing conditions revealed specific binding of several of the Invader probes, resulting in high signal intensity. A subsequent Spearman Rank Correlation analysis revealed properties and design features of Invader probes that result in efficient recognition of chromosomal DNA. Optimized Invader probes based on these factors confirmed our statistical analysis. This represents a step in the continual improvement of the Invader probe design that will help develop efficient probes for applications in molecular biology, nucleic acid diagnostics, and biotechnology.

2.1 Introduction

Oligonucleotide-based technologies for targeting specific sequences of double-stranded DNA have shown great success in identification, regulation, and manipulation of genes; however, the design of established approaches remain suboptimal.[1-7] For example, Triplex-forming oligonucleotides (TFOs) bind in the major groove of double-stranded DNA (dsDNA) by forming Hoogsteen base pairs.[8,9] There is a limitation in number of target sequences as major groove binding of TFOs necessitates long purine tracts as only adenine and guanine have the optimal hydrogen bond acceptor and donor group geometry to interact with the TFO. Additionally, Pyrrole-Imidazole (Py-Im) Polyamides have been designed to target specific sites by binding to the minor groove; however, it is challenging to design Py-Im probes that target sequences longer than 8 nucleotides because its structure loses flexibility with increasing length and no longer adequately fits in the minor groove.[10,11] Additionally, they cannot distinguish between AT and TA base pairs, reducing their specificity.[12,13] Utilizing a different approach, Peptide nucleic acids (PNAs) have base pairs attached to an uncharged *N*-[2-aminoethyl)glycine backbone which reduces charge repulsion associated with duplex formation.[14,15] This results in very stable duplexes with cDNA *via* Watson-Crick and Hoogsteen base pairing, allowing for strand invasion of dsDNA. Due to the need for Hoogsteen base pairing, PNAs are limited to targeting sequences with a short stretch of polypurines. Different designs have been employed to overcome these limitations, such as tail-clamp PNA and bis-PNA that utilize Hoogsteen base pairing to perturb the target duplex so their long Watson-Crick arm can stably hybridize.[15,16] Additionally, the γ-PNA design has been used to reduce sequence limitations by its significant increase in binding affinity.[17,18] Introduction of a chiral center induces preorganization of the probe to reduce entropic costs, thus increasing

affinity to cDNA. Although γ-PNA can recognize mixed-sequence dsDNA target regions (150-300 bp), they require low ionic conditions for duplex invasion, making them less than ideal for physiologically relevant ionic strengths. The recent discovery CRISPR-Cas9 and its programmability interduce single specific cuts in a genome has been of enormous value for gene-editing. Moreover, by deactivating its nuclease activity, CRISPR-Cas9 can be used for recognition of specific sequences of DNA in diagnostic applications.[19,20] Unfortunately, these abilities are limited in *in vivo* studies due to the need for plasmid transfection.[20,21]

To overcome limitations in target sequences and experimental conditions, we have developed dsDNA probes, termed Invaders, which have targeting capabilities of mixed-sequence chromosomal DNA through double-duplex invasion (Figure 2.1-1a). The driving force for dsDNA recognition is the energy difference between the less stable reactants (i.e. the double-stranded probe and DNA target) and the highly stable products (i.e. the two probe-target duplexes formed as part of the recognition complex). In this approach, the Invader probe is destabilized or "activated" through steric clashing of intercalators at the energetic hotspots. The term "energetic hotspots" denotes the +1 interstrand zipper architecture in which intercalating moieties *vie* for the same space between two Watson-Crick base pairs, promoting unwinding and destabilization of the double-stranded probe. The Invader design employed herein utilizes 2'-*O*-(pyren-1-yl)methyl-RNA monomers (Figure 2.1-1b), which are the successor to the classic 2'-*N*-(pyren-1-yl)methyl-2'-amino-α-L-LNA monomers.[22-24] These monomers increase affinity to target DNA by sterically constricting the probe strand to reduce entropic costs and increase base stacking interactions *via* the intercalator and the nucleobases. The destabilization of Invader probe in concert with the high affinity for cDNA displayed by

individual probe strands provides the driving force for recognition of target site (Figure 2.1-1a).

Previously reported Invader 4 (**INV4**) (Table 2.2-1) displayed the ability to detect chromosomal DNA, specifically the Y chromosome of male bovine kidney cells under non-denaturing conditions (3 h, 37.5 °C, 20 mM Tris, 100 mM KCl, pH 8.0) to yield a single localize fluorescent signal in ~90% of observed nuclei.[25] Encouraged by these results, the ability to detect chromosomal DNA by nine other Invaders with different energetic hotspot arrangements and targeting different sequences in the 1175 base-pair repeat (Table 2.2-1 and Figure 2.4-6) was determined. Thermal denaturing analysis, thermodynamic properties, and dose-response experiments were used as measures of the physical properties and invasion potential for each of the ten probes. A subsequent DNA FISH assay was conducted at denaturing and non-denaturing conditions to determine the Invader probes' capability to detect chromosomal DNA. Based on the results of these experiments and structural features of each probe, a Spearman Rank Correlation analysis was conducted to determine factors contributing to efficient FISH signaling under physiologically relevant, non-denaturing conditions. Based on these factors, three of the ten screened Invaders were optimized and evaluated to determine the validity of our statistical analysis. Results revealed crucial factors in detection of chromosomal DNA and will contribute to future probe design.

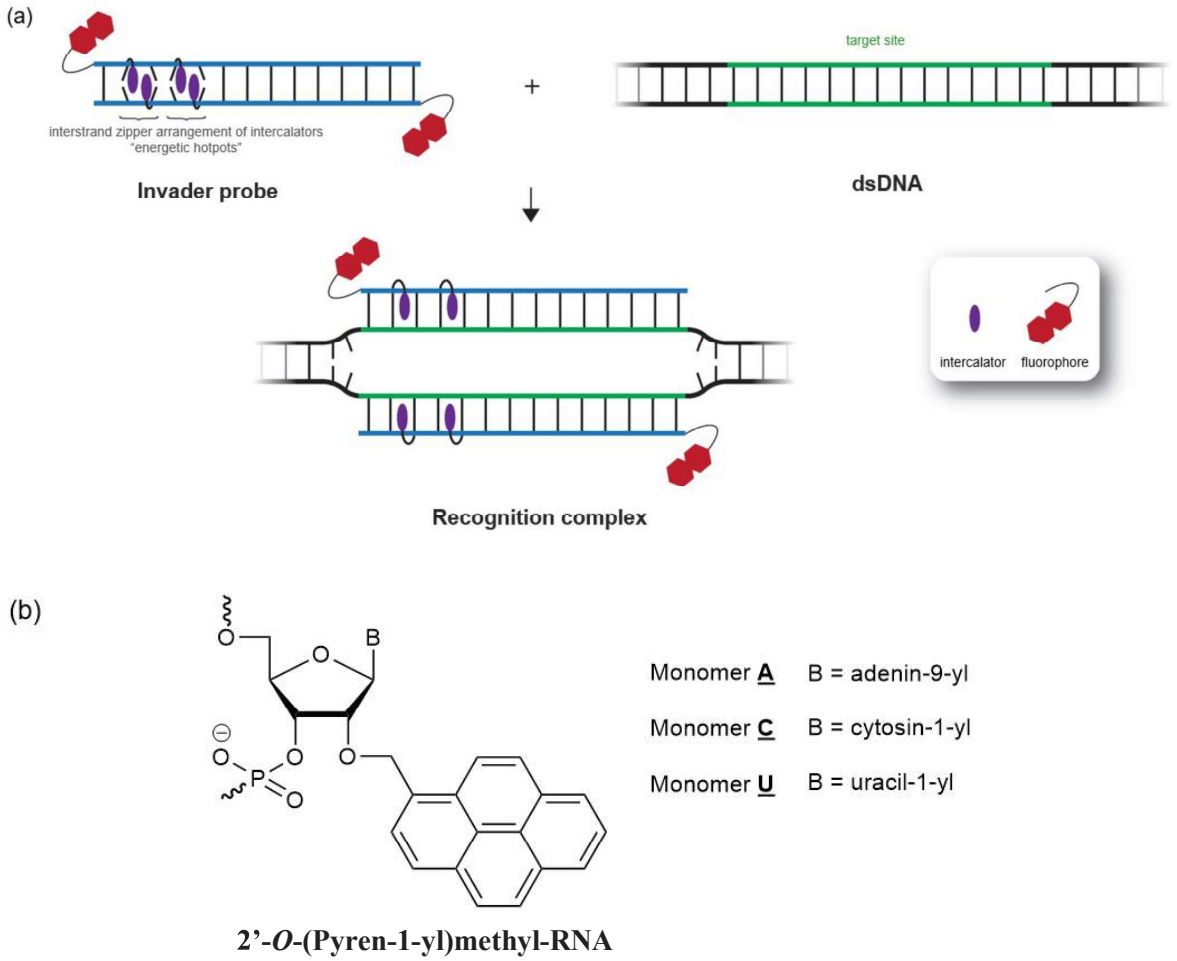


Figure 2.1-1. (a) Illustration of Invader-mediated recognition of dsDNA *via* a double-duplex invasion process. (b) Structure of Invader monomers used herein.

2.2 Results and Discussion

Experimental design of Invader probes

Ten oligodeoxyribonucleotide (ON)-based Invader probes - varying in length (14-16 base pairs (bps)), number of energetic hotspots (3-4), and GC-content (GC%) (31-71%) and designed to target corresponding complementary sequences in the *DYZ-1* satellite gene (~6 x 10^4 tandem repeats of the ~1175 bp region) on the bovine (*Bos taurus*) Y chromosome (NCBI code: M26067, Figure 2.4-6)[26] – were obtained using previously established machine-assisted solid-phase DNA synthesis protocols (Table 2.2-1).[27] Each Invader probe targeted one region within

each ~1175 bp repeat with the exception of **INV4** whose target sequence is present 6 times within each repeat. The identity and purity of the modified ONs was established by MALDI-TOF (Table 2.4-1 and Figures 2.4-1, 2.4-2, and 2.4-3) and ion-pair reverse phase HPLC respectively. The lengths of the probes were selected to minimize unintended binding to non-targets, whereas the relative modification density (mod%) (18.8-28.6%, Table 2.2-1) was chosen based on prior work[25,27] in which we showed that Invader probes with a hotspot content of 20-25% strike a favourable balance between high binding affinity and excellent binding specificity. Less densely modified probes (e.g., ~15%) result in lower binding affinity, whereas more densely modified probes (e.g., ~30%) display reduced specificity.

Thermal denaturation properties of Invader probes

Thermal denaturation temperatures (T_ms) were determined for the double-stranded Invader probes and the corresponding duplexes between individual probe strands and cDNA (Table 2.2-1). The Invader probe duplexes are slightly destabilized, displaying T_ms which on average are 1.6 °C lower than the T_ms of the corresponding unmodified DNA duplexes **DNA1**-**DNA10**, though ΔT_m of probe duplexes range widely (-18.5 °C to +9.5 °C, Table 2.2-1). Conversely, duplexes between individual Invader probe strands and cDNA display T_ms that on average are 12.8 °C higher than for the unmodified DNA duplexes (ΔT_ms range between +3.5 °C and +22.0 °C, Table 2.2-1). Greater relative increases in T_ms are observed for more densely modified probe-target duplexes (ΔT_m = ~17 °C and ~10 °C for duplexes with hotspot contents of >20% and ≤20%, respectively).[1] These observations follow our expectations that Invader probes are

[1] By design, Invader probes targeting dsDNA sequences with a low GC-content, will tend to be more densely modified (in the present study, probes with a GC-content below 50% have, on average, a hotspot contents of ~24%, whereas probes with a GC-content at or above 50% have, on average, a hotspot content of ~20%. This is because Invader probes with hotspots constructed of 2'-*O*-(pyren-1-yl)methyl RNA pyrimidine monomers, display higher dsDNA-affinity than probes constructed using the corresponding purine monomers.[28]

more labile[22,29] as the neighbor exclusion principle is violated due to high localized intercalator densities,[30,31] whereas individual probe strands form more stable duplexes with cDNA since the neighbor exclusion principle is not violated, which enables favorable stacking interactions between the pyrene moieties and neighboring base-pairs.[22,29] Spearman correlation coefficients (r_s) comparing modification density to ΔT_ms of probe-target duplexes confirms these observations. There is a correlation between increasing modification with increasing ΔT_ms of probe-target duplexes 5'-**INV**:cDNA (r_s = 0.774, p-value = 0.009) and 3'-**INV**:cDNA (r_s = 0.661, p-value = 0.037).

Consequently, probe-target duplexes with low GC-contents display greater relative increases (ΔT_m ~16 °C and ~9.5 °C for duplexes with GC contents of <50% and ≥50%, respectively)

Table 2.2-1. Thermal denaturation temperatures (T_m), Thermal Advantage Values (TA_{isq} and TA_{DH}), Percent Modification (Mod%), and GC-content (GC%) of *DYZ1*-targeting Invader probes and duplexes between individual probe strands and cDNA.[a]

Probe	Sequence	T_m [ΔT_m] (°C)			TA_{isq} (°C)	TA_{DH} (°C)	Mod%	GC%
		Probe duplex	5'-INV: cDNA	3'-INV: cDNA				
INV1	5'-Cy3-T**U**ATCAGCAC**U**G**U**GC-3' 3'- AA**U**AGTCGTGA**C**A**C**G-Cy3-5'	52.0 [-4.0]	65.5 [+9.5]	66.0 [+10.0]	+23.5	+3.5	20%	47%
INV2	5'-Cy3-A**U**AC**U**GGTTTG**U**G**U**TC-3' 3'- TA**U**GA**C**CAAACA**C**A**A**G-Cy3-5'	34.5 [-18.5]	66.0 [+13.0]	66.0 [+13.0]	+44.5	+25.5	25%	38%
INV3	5'-Cy3-T**U**G**U**GCCCTGGC**A**AC-3' 3'- AA**C**A**C**GGGACCGT**U**G-Cy3-5'	NT	64.0 [+5.5]	62.0 [+3.5]	ND	ND	20%	60%
INV4	5′-Cy3-**A**GCCC**U**GTGCC**C**TG-3' 3′- T**C**GGGA**C**ACGGG**A**C-Cy3-5'	61.5 [+1.0]	69.5 [+9.0]	75.5 [+15.0]	+23.0	+1.5	21.4%	71%
INV5	5'-Cy3-G**A**TTTCAGCCA**U**G**U**GC-3' 3'- CT**A**AAGTCGGTA**C**A**C**G-Cy3-5'	45.0 [-12.0]	63.0 [+6.0]	69.5 [+12.5]	+30.5	+11.5	18.8%	50%
INV6	5'-Cy3-C**U**G**U**GCAACTGGT**U**TG-3' 3'- GA**C**A**C**GTTGACCAA**A**C-Cy3-5'	63.0 [+5.0]	65.5 [+7.5]	69.0 [+11.0]	+13.5	-3.5	18.8%	50%
INV7	5'-Cy3-C**U**G**U**GCAA**U**ATTT**U**GT-3' 3'- GA**C**A**C**GTTA**T**AAAA**C**A-Cy3-5'	55.0 [+4.0]	73.0 [+22.0]	71.0 [+20.0]	+38.0	+21.0	25%	31%
INV8	5'-Cy3-TT**C**ACAGCCC**U**G**U**GC-3' 3'- AAG**U**GTCGGGA**C**A**C**G-Cy3-5'	58.5 [-1.5]	70.5 [+10.5]	74.5 [+14.5]	+26.5	+6.0	20%	60%
INV9	5'-Cy3-T**U**A**U**ATGCTG**U**TCTC-3' 3'- AA**U**A**U**ACGACA**A**GAG-Cy3-5'	55.0 [+9.5]	58.0 [+12.5]	64.0 [+18.5]	+21.5	0	20%	33%
INV10	5'-Cy3-G**U**G**U**AGTG**U**A**U**ATG-3' 3'- CA**C**A**U**CACA**U**A**U**AC-Cy3-5'	45.5 [+2.0]	65.0 [+21.5]	64.5 [+21.0]	+40.5	+22.0	28.6%	36%

[a]ΔT_m = change in T_m values relative to corresponding unmodified duplexes. T_ms for the corresponding unmodified DNA duplexes **DNA1** = 56.0 °C, **DNA2** = 53.0 °C, **DNA3** = 58.5 °C, **DNA4** = 60.5 °C (previously reported in reference 25), **DNA5** = 57.0 °C, **DNA6** = 58.0 °C, **DNA7** = 51.0 °C, **DNA8** = 60.0 °C, **DNA9** = 45.5 °C, and **DNA10** = 43.5 °C. Thermal denaturation curves were recorded in medium salt buffer ($[Na^+]$ = 110 mM, $[Cl^-]$ = 100 mM, pH 7.0 (NaH_2PO_4/Na_2HPO_4), [EDTA] = 0.2 mM) and each [ON] = 1.0 μM; see main text for definition of TA_{isq} and TA_{DH}. NT = no clear transition (very broad) at wavelengths 230-280 nm. ND = not determined. For structures of **A**, **C** and **U**, see Figure 2.1-1.

Thermodynamic driving force for recognition of dsDNA targets.

The driving force for Invader-mediated recognition of isosequential dsDNA targets can be assessed by the term *thermal advantage*, which we define as $TA_{isq} = T_m$ (5'-**INV**:cDNA) + T_m (3'-**INV**:cDNA) - T_m (probe duplex) - T_m (dsDNA) (Table 2.2-1). Large positive TA_{isq} values indicate that a probe is activated for dsDNA recognition, meaning the probe duplex is sufficiently destabilized and its individual strands form stable duplexes with cDNA. Eight of the ten Invader probes display TA_{isq} values above 20 °C, indicating these probes have dsDNA recognition potential. Noteworthily, the three most densely modified Invader probes, display the most prominent TA_{isq} values (TA_{isq}s between 38.0-44.5 °C for **INV2**, **INV7** and **INV10**). Only **INV6**, which is less densely modified (18.8%), displays a *TA* value below 20 °C. Calculated Spearman correlation coefficients confirms a positive correlation between TA_{isq} and percent modification (r_s = 0.662, p-value = 0.052), indicating the activation for dsDNA-recognition by Invader monomer incorporation.

In order to fully understand the driving force for dsDNA recognition, thermodynamic parameters of hybridization were determined by using baseline fitting of the thermal denaturation curves of each of the ten Invader probes *via* the van't Hoff method.[32] The available free energy for dsDNA recognition can be described as $\Delta G^{310}_{rec} = \Delta G^{310}$ (5'-**INV**:cDNA) + ΔG^{310} (3'-**INV**:cDNA) – ΔG^{310} (probe duplex) – ΔG^{310} (dsDNA) (Table 2.4-3). Large negative values ($\Delta G^{310}_{rec} << 0$ kJ/mol) indicate a more prominent driving force and probes activated for dsDNA recognition. This driving force is achieved by a relative change in stability between the labile Invader probes to a stable recognition complex. In agreement with T_m-based trends, our probes are largely activated for invasion (ΔG^{310}_{rec} ranging between -19 to -56 kJ/mol) which can be attributed to the energy differential between the low stability of the ten Invader

Probes ($\Delta\Delta G^{310}$ averaging at +11.6 kJ/mol) and high stability of probe-target duplexes (5'/3'-**INV**:cDNA $\Delta\Delta G^{310}$ averaging at -6.6 kJ/mol) (Table 2.4-3). Additionally, recognition of dsDNA by probes (except for **INV9**) are highly enthalpically favored ($\Delta H^{310}_{rec} < 0$ kJ/mol), further highlighting the intercalation driving force of dsDNA recognition that results in stable recognition complex (Table 2.4-4). A combination of positive TA_{isq} values and negative ΔG^{310}_{rec} values these Invader probes show they have potential for recognition of dsDNA, which is confirmed in the subsequent experiments below.

Recognition of mixed-sequence model DNA hairpin targets – design and initial screen.

The dsDNA-recognition potential of the ten Invader probes were initially evaluated in an electrophoretic mobility shift assay[25] using 3'-digoxigenin (DIG)-labelled DNA hairpins (**DH**) as model targets (Figure 2.2-1). DNA hairpins **DH1**-**DH10** are comprised of a double-stranded stem which is complementary to the corresponding Invader probe and is linked at one end by a decameric thymidine (T_{10}) loop. The unimolecular nature of the DNA hairpins ensures that both target strands are present in equimolar amounts and that the stem regions are denaturing at high temperatures, thus rendering them as challenging dsDNA targets (compare T_{m}s for **DH1**-**DH10** and T_{m}s for **DNA1**-**DNA10**, 2.4-9 and footnote a of Table 2.2-1, respectively. T_{m}s for **DH1**-**DH10** are higher by 17-23 °C). The challenging nature of this double-stranded hairpin was designed as a representation of chromosomal DNA where both ends of the duplex are clamped. Invader-mediated recognition of the double-stranded stem regions is expected to result in the formation of ternary complexes (TC), similar to the formation of recognition complex in Figure 2.1-1, which are visualized as lower mobility bands on non-denaturing (nd) PAGE gels.

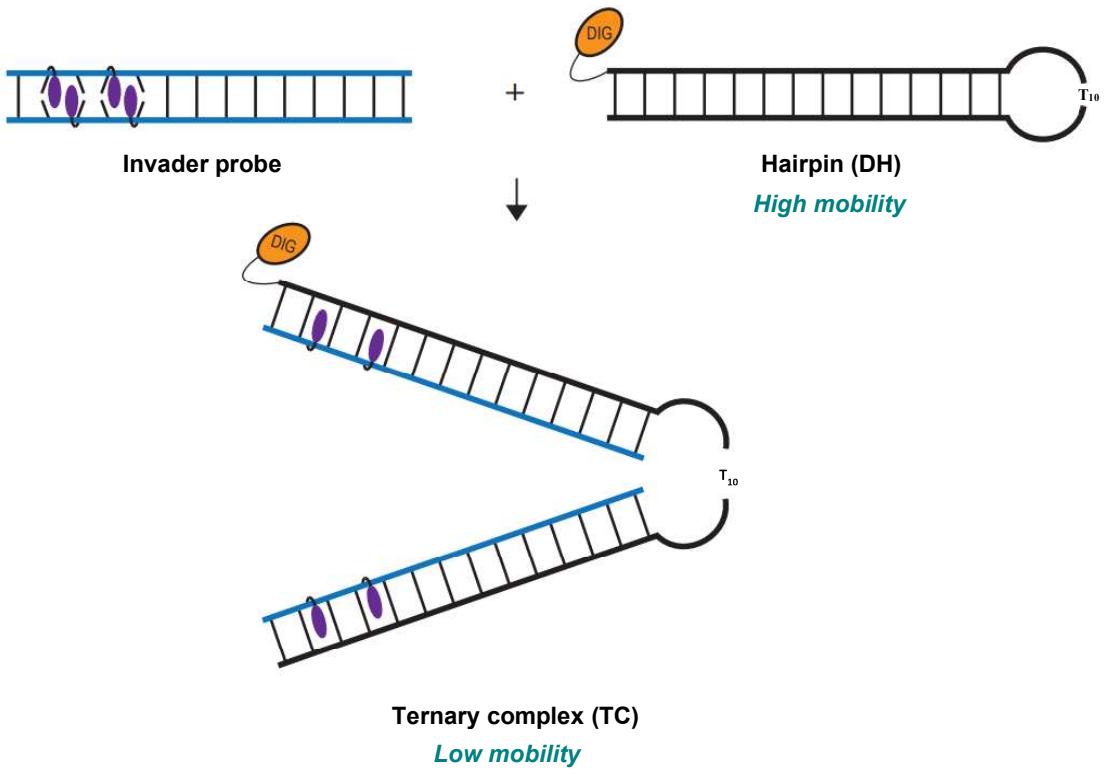


Figure 2.2-1. Principle Hairpin Assay

A 100-fold molar excess of each Invader probe was incubated with its corresponding DNA hairpin for 15 hours at 37 °C in a HEPES buffer containing 100 mM NaCl and 5 mM $MgCl_2$ (Figure 2.2-2) and percent recognition of hairpin was determined (i.e. the amount of TC formed relative to DH). This experiment was conducted at 15 hours to ensure that full recognition potential was reached for each Invader probe. **INV2**, **INV7**, **INV8**, and **INV10** resulted in almost full recognition (>96%) (Figure 2.2-2 and Table 2.2-2). **INV1**, **INV3**, and **INV9** exhibited moderately high recognition (60-66%) whereas **INV6** gave lower levels of recognition (41%). **INV4** gave no recognition in this screen.

For this assay, TA_{DH} values - defined as $TA_{DH} = T_m$ (5'-**INV**:cDNA) + T_m (3'-**INV**:cDNA) - T_m (probe duplex) - T_m (**DH**) – are evaluated (Table 2.2-1). TA_{DH} is being used for this assay as it better describes the Invader probes activated for recognition of DNA hairpin. Both TA_{isq} and

TA_{DH} follow a similar relative trend (i.e. higher values for more highly activated probes); however, hairpins (**DH1**-**DH10**) are a more challenging target (i.e. higher in T_m) in comparison to the isosequential DNA duplexes (**DNA1**-**DNA10**). A large positive TA_{isq} may indicate activation for dsDNA recognition; however, it might be inadequate to predict invasion of hairpin. TA_{DH} more clearly shows probes exhibiting a sufficient driving force for hairpin recognition. This is exemplified by **INV6**, which gives a positive TA_{isq} (+13.5 °C), but a negative TA_{DH} (-3.5 °C) (Table 2.2-1). This indicates **INV6** is not highly activated for hairpin recognition, which is reflected in its poor recognition in this screen (Table 2.2-2).

The high recognition efficiency of the quadruply modified **INV2**, **INV7**, and **INV10** probes was expected given their prominent ΔG_{rec}^{310} and TA_{DH} and their relatively low Invader duplex T_m (Table 2.4-3 and Table 2.2-1), which is expected to facilitate probe separation as recognition ensues. Although **INV8** is less densely modified (20%), it exhibited near full binding at 15 h, which can be attributed to its significant ΔG_{rec}^{310} (-52 kJ/mol), and favorable TA_{DH} (+6.0) (Table SX and Table 2.2-1). **INV1**, **INV3**, and **INV9** each have a relatively lower modification density (20%), which could rationalize the lower levels of recognition in comparison to highly modified **INV2**, **INV7**, and **INV10**, in addition to small TA_{DH}s (Table 2.2-1). ΔG_{rec}^{310} values could not be determined for **INV1** and **INV3**; however, low recognition by **INV5, INV6, and INV9** could also be explained by a less prominent ΔG_{rec}^{310} (between -29 kJ/mol and -19 kJ/mol) and low modification density (≥20%) (Table SX and Table 2.2-1). The absence of recognition by **INV4** may be attributed to less significant ΔG_{rec}^{310} (-29 kJ/mol) and TA_{DH} (+1.5), in addition to a highly GC rich target sequence (71%), contributing to the highest hairpin T_m of 82 °C (Table 2.4-9). Lending support to these rationales are Spearman correlation coefficients, as recognition at 15 h is highly correlated to TA_{DH} (r_s = 0.707, p-value = 0.033), GC% (r_s = -

0.640, p-value = 0.046), %mod (r_s = 0.699, p-value = 0.025), ΔG_{rec}^{310} of **DH** (r_s = -0.872, p-value = 0.005). This initial screen suggests: i) need for highly modified probes that are sufficiently activated to invade their target sequence, and ii) that most of our Invaders meet this requirement.

An additional recognition screen was performed using 100-fold molar excess of each Invader probe incubated with its corresponding DNA hairpin for 2.5 hours at 37 °C in a HEPES buffer containing 100 mM NaCl and 5 mM $MgCl_2$ (Figure 2.4-12). **INV2** and **INV10** result in high levels of recognition (>80%), whereas **INV7-INV9** result in intermediate levels of recognition (30-60%) and **INV1** and **INV3-INV6** result in low levels of recognition (<25%) (Table 2.4-10). Although the 2.5 h assay shows the potential for fast recognition, allowing 15 h for recognition is needed to reach the full binding potential of the Invader probes and, as discussed in subsequent sections, recognition after 15 h is better measure of potential binding in chromosomal DNA.

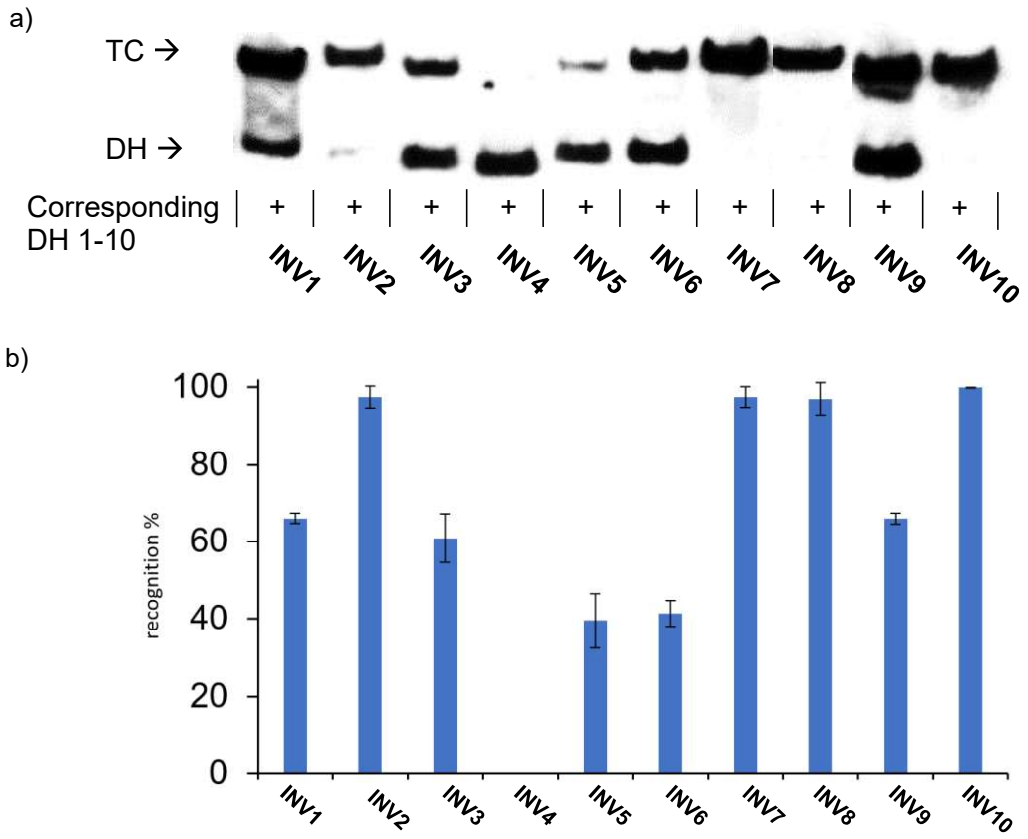


Figure 2.2-2. a) Representative gel electrophoretograms from recognition experiments at 15 h between a 100-fold molar excess of Invader probes **INV1**-**INV10** and their corresponding DNA hairpin targets **DH1**-**DH10**. b) Histograms depicting averaged results from at least three recognition experiments with error bars representing standard deviation. TC = ternary complex. DIG-labeled **DH1**-**DH10** (34.4 nM, sequences shown in Table 2.4-9) were incubated with the corresponding Invader probe in HEPES buffer (50 mM HEPES, 100 mM NaCl, 5 mM $MgCl_2$, pH 7.2, 10% sucrose, 1.44 mM spermine tetrahyrdochloride) for 15 h at 37 °C. Incubation mixtures were resolved on 12% non-denaturing TBE-PAGE slabs (~70 V, ~4 °C, ~1.5 h).

Recognition of model mixed-sequence dsDNA targets – dose-response and binding specificity.

Dose-response experiments conducted at 15 h were performed for all Invader probes displaying above 15% recognition in the preliminary screen (Figure 2.2-3) using the same electrophoretic mobility shift assay (EMSA) as described above to determine C_{50} values, i.e., the probe concentration that results in 50% recognition of **DH**. The C_{50} values follow a similar trend as percent recognition. **INV2** and **INV10** have the lowest C_{50} values (0.2 μM) followed

by **INV8** and **INV7** (0.6 μM and 0.7 μM) (Table 2.2-2), exhibiting excellent binding in our model system. In line with percent recognition trends, these are followed by **INV1**, **INV3**, **INV5**, and **INV9** with C_{50} values between 1.3 μM and 4.1 μM (Table 2.2-2). Once again, trailing the other probes are **INV6** and **INV4**, which have negligible levels of recognition. As they follow a similar trend to the initial screen, similar rationales related to TA_{DH}, ΔG^{310}_{rec}, modification density, and GC-content of each probe can be applied to explain their resulting C_{50} values. These dose-responses further support the rationale that densely modified probes sufficiently activated for invasion of dsDNA are needed for efficient dsDNA recognition (i.e. low C_{50} values). This is supported by Spearman correlation coefficients as C_{50} values are correlated to %mod (r_s = -0.850, p-value = 0.008) and ΔG^{310}_{rec} of **DH** (r_s = 0.852, p-value = 0.031).

Additional dose-response experiments were conducted at 2.5 h for all Invaders; however, four of the ten Invaders did not reach 50% recognition even at high levels of excess hairpin (Table 2.4-10) and, as will be discussed in subsequent sections, C_{50} values obtained at 15 h are more predictive of chromosomal binding.

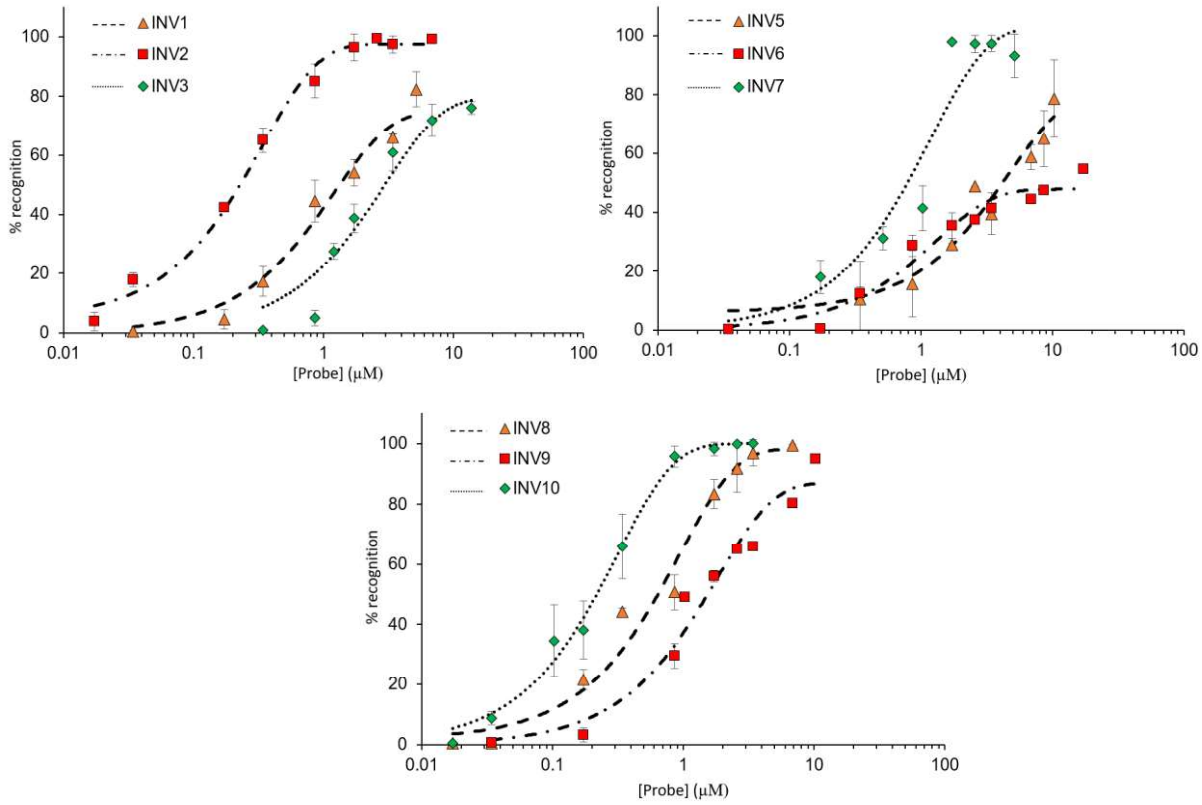


Figure 2.2-3. Dose-response curves for **INV1**, **INV2, INV3** (upper left panel), **INV5**, **INV6**, and, **INV7** (upper right panel), and **INV8**, **INV9**, and **INV10** (lower panel) at 37 °C for 15 h. Experimental conditions are as described in Figure 2.2-1, except for variable probe concentrations.

Table 2.2-2. C_{50} values at 15 h for Invader probes studied herein.[a]

Probes	C_{50} (μM)	Rec_{100X} (%)
INV1	1.3	66 ± 1.3
INV2	0.2	97 ± 2.8
INV3	2.9	60 ± 6.2
INV4	ND	ND
INV5	4.1	39 ± 7.0
INV6	>10	41 ± 3.4
INV7	0.7	97 ± 2.6
INV8	0.6	96 ± 4.2
INV9	1.5	66 ± 1.4
INV10	0.2	99 ± 0.0

[a] Rec_{100} = level of DNA hairpin recognition using 100-fold molar probe excess. C_{50} values were determined from dose-response curves shown in Figure 2.2-3.

Next, to evaluate their binding specificity, each Invader probe – present at 3.44 μM – was incubated with a mixture comprised of the nine non-target hairpins, each present at 34.4 nM

(e.g., **INV1** was incubated with a mixture of **DH2**-**DH10**). As expected, none of the Invader probes recognized the mismatch target hairpins and ternary complexes were not formed (Figure 2.2-4). To ensure the presence of the non-target hairpins did not interfere with potential non-specific binding to hairpin in our specificity assay, each Invader probe was individually incubated with a **DH1**-**DH10** mixture. The lack of interference was confirmed as similar levels of recognition of match target hairpin were seen for all Invader probes when each was incubated with the **DH1**-**DH10** mixture, though recognition could not be quantified due to presence of large quantities of non-targeting hairpin (Figure 2.4-11)

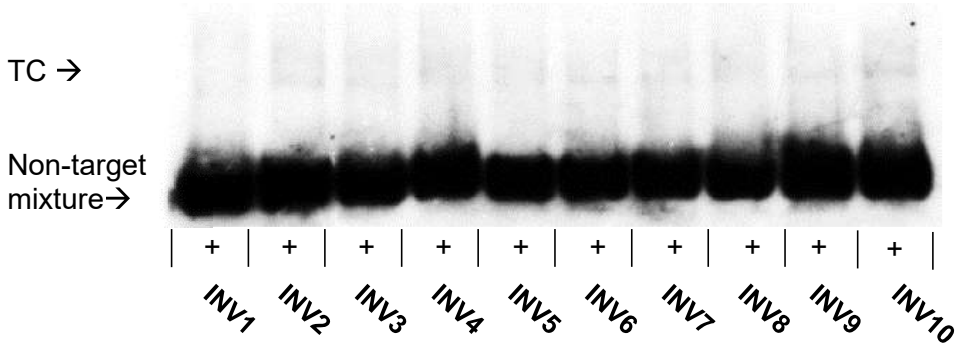


Figure 2.2-4. Binding specificity. Representative electrophoretograms for experiments in which the specified Invader probe (3.44 μM) was incubated at 37 °C for 2.5 h with a mixture of nine non-target DNA hairpins (each hairpin present at 34.4 nM). For example, **INV1** was incubated with **DH2**-**DH10**, while **INV10** was incubated with **DH1**-**DH9**. Conditions are otherwise as described in Figure 2.2-2.

Detection of chromosomal DNA: Fluorescence in situ hybridization (FISH) assays

The ten Cy3-labelled Invader probes were evaluated for their ability to recognize the corresponding *DYZ-1* target regions on the bovine Y chromosome using fluorescence in situ hybridization (FISH) assays. Fixed interphase nuclei and metaphase spreads from a male bovine kidney cell line (MDBK (NBL-1) (ATCC® CCL-22™)) were incubated with **INV1**-

INV10 under denaturing (d) or non-denaturing (nd) FISH conditions. Results from the d-FISH assay (80 °C, 5 min, 20 mM Tris, 100 mM KCl, pH 8.0) are expected to yield each Invader probes' full recognition potential for chromosomal DNA due to the increased lability of chromosomal DNA by high temperature conditions. Binding under these conditions would reflect targetable regions and probes capable of recognizing chromosomal DNA. More critically, nd-FISH results would reveal probes capable of binding at physiologically relevant conditions (3 h, 37.5 °C, 20 mM Tris, 100 mM KCl, pH 8.0)

When incubated with isolated nuclei, fluorescently labeled Invader probes display varying levels of signal strength, background, and coverage (i.e. number of nuclei displaying one strong localized signal) (Figure 2.4-23, 2.4-24 and 2.4-25 and Table 2.4-12). Thus, to determine acceptable conditions and Invader probe concentrations yielding many nuclei with a single strong signal against a weak background, probes were incubated under denaturing (80 °C, 5 min, 20 mM Tris, 100 mM KCl, pH 8.0) and with increasing amounts of Invader with non-denaturing conditions (3 h, 37.5 °C, 20 mM Tris, 100 mM KCl, pH 8.0) (Table 2.4-12). These preliminary studies revealed an excellent signal-to-background ratio with our d-FISH assay. Increasing the concentration of the Invader probes did not yield further improvements in quality of the signal in nd-FISH experiments and only increased the background, rendering it more difficult to discern signals (Table 2.4-12). Thus, 1x concentration (~30 ng of probe per 200 μL of incubation buffer) was used for subsequent experiments to yield quality signals, except for **INV4**, which required ¼ the concentration of the other Invader probes due to high levels of background.

The d-FISH assay results revealed Invader probes capable of recognizing their target in chromosomal DNA. **INV2**, **INV4**, and **INV10** display a single localized Cy3-signal (red dot)

with minimal background in ~90% of the analyzed nuclei, consistent with successful recognition of the corresponding target regions on the Y chromosome (Figure 2.2-5 and Figure 2.2-6 – left column; Table 2.2-3). These results demonstrate that Invader probes can bind to chromosomal DNA provided they exhibit a sufficiently high binding affinity and the target region is accessible. Consistent with this, **INV3** and **INV6-INV9** display clear signals in 30-70% of the analyzed nuclei at denaturing conditions (Figures 2.4-23, 2.4-24, 2.4-25 – left column; Table 2.2-3), a trend that reflects the more moderate recognition levels seen for these probes in the hairpin assay. Furthermore, **INV1** yields multiple patches of signals, indicative of significant non-specific binding (Figure 2.4-23, Row 1, Left), whereas neither signals nor background staining is observed with **INV5**, which agrees with the results from the DNA hairpin assay. This indicates that i) most of our probes can recognize their chromosomal DNA target, and ii) our hairpin model has predictive power in identifying which probes will have the ability to recognize chromosomal DNA. (3 h, 37.5 °C, 20 mM Tris, 100 mM KCl, pH 8.0)

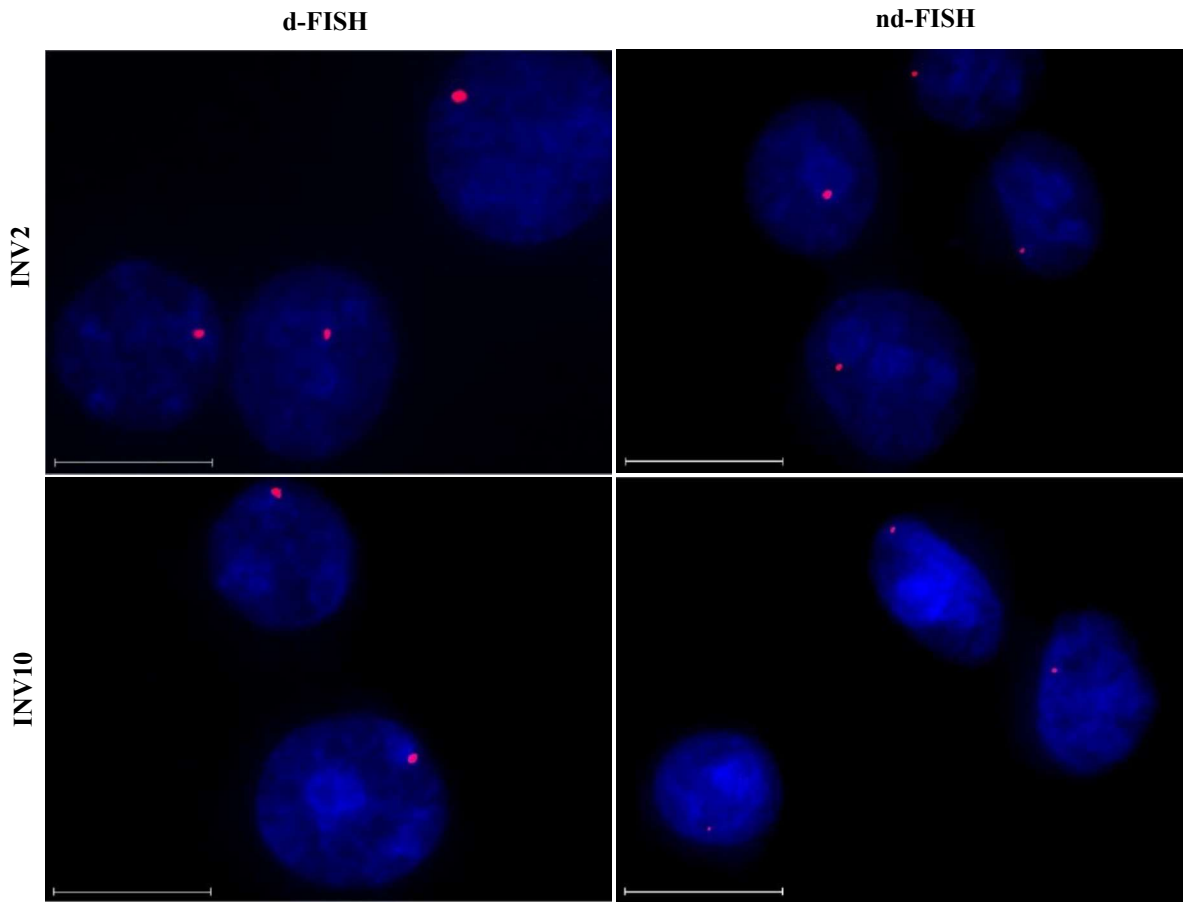


Figure 2.2-5 Images from FISH experiments using *DYZ1*-targeting Invader probes **INV2** and **INV10** under denaturing (80 °C, 5 min) (left), or non-denaturing (3 h, 38.5 °C) conditions (right). Images are representative of the signal intensity and background, and the size and morphology of all analyzed nuclei (~200). Fixed isolated nuclei from male bovine kidney cells were incubated with probes in a Tris buffs (20 mM Tris-Cl, 100 mM KCl, pH 8.0) and counterstained with DAPI. Images are obtained by overlaying Cy3 (red) and DAPI (blue) filter settings and adjusting the exposure. Nuclei were viewed at 60X magnification using a Nikon Eclipse T*i*-S inverted microscope. The scale bar represents 16 µm. For corresponding images for other Invader probes, see Figures 2.4-23, 2.4-24, and 2.4-25.

Exhibiting similar trends shown in the d-FISH assay, the nd-FISH assay revealed probes capable of recognizing chromosomal DNA at near physiological conditions. Thus, no signals are observed for **INV1** and **INV5**, while clear signals are observed for **INV3** and **INV6-INV8** at a slightly lower coverage than in the d-FISH assay (20-40%, Table 2.2-3). Interestingly, **INV9**, which results in distinct signals at moderate coverage in the d-FISH assay, does not

produce signals (Table 2.2-3). As **INV9** can access its target under denaturing conditions, the lack of coverage may be due to insufficient energy that is needed invade its chromosomal DNA target (ΔG_{rec}^{310} = -19 kJ/mol, Table 2.4-3). The moderate level of coverage resulting from **INV7** was surprising as well considering it is a densely modified probe (25%) displaying favorable binding within the hairpin assay (Table 2.2-2 and Table 2.2-3). This, as well as the lack of signal by **INV1** and **INV5**, highlights the unpredictability and challenging targeting nature of chromosomal DNA. It may be that there is not sufficient energy to target the stable dsDNA duplex, or possibly, the target sequence tightly packed with supercoiling and/or wrapped around structural proteins, preventing full binding even under denaturing conditions.[33] Critically, **INV2**, **INV4**, and **INV10** retained clear signaling against a clean background with excellent coverage at nd-FISH conditions (80-90%, Table 2.2-3), corroborating the feasibility of these probe designs for recognition of chromosomal mixed-sequence dsDNA at physiologically relevant conditions. As discussed above, **INV2** and **INV10** are strongly activated for dsDNA-recognition (Table 2.2-1), thus providing a compelling rationale for their performance in the nd-FISH assay. The signaling efficiency of **INV4** (Table 2.2-3) is surprising given that the probe is weakly activated (TA_{isq} = +23.0 °C, ΔG_{rec}^{310} = -29 kJ/mol, TA_{DH} = +1.5 °C, Table 2.2-1) compared to the other Invader probes, has a high GC-content (~71%), and does not result in recognition of its corresponding DNA hairpin (Figure 2.2-2). A unique feature of **INV4** (and, hence, its corresponding target region) relative to the other probes is that it features two GGG-tracts, which may render the target region uniquely accessible due to G-quadruplex formation or other G-forming secondary structures.[34] An alternative - and perhaps more likely explanation for the unusual signaling characteristics of **INV4** - is that the corresponding target region is present six times within a single 1175 bp long *DYZ-1* unit, which

in turn is repeated ~60,000 times (Figure 2.4-6).[26] In contrast, the target regions of the other probes are only present once per repeating unit. The additional target sites may yield more strong signals that is reflected in the high level of coverage.

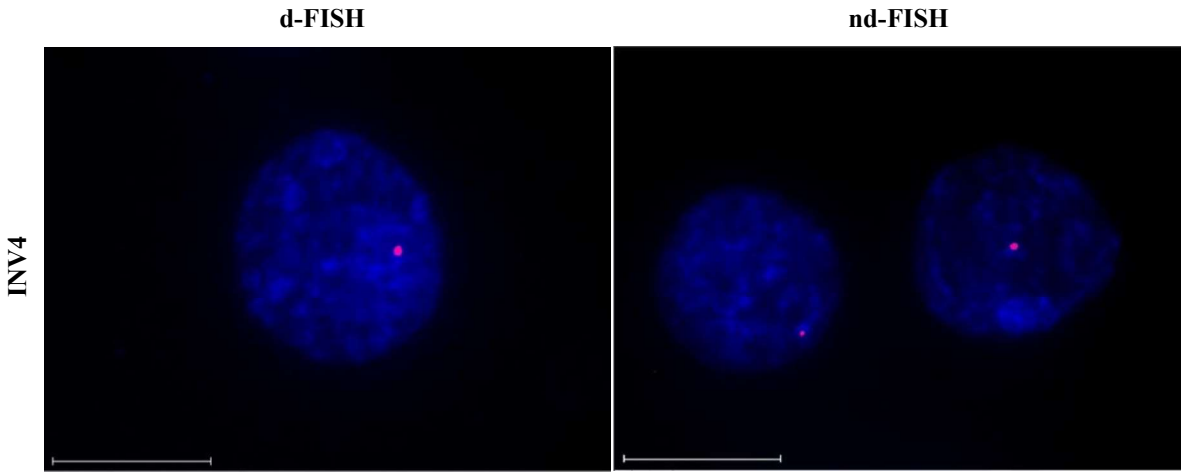


Figure 2.2-6. Images from FISH experiments using *DYZ1*-targeting Invader probes **INV4** under denaturing (80 °C, 5 min) (left), or non-denaturing (3 h, 37.5 °C) conditions (right). Incubation and imaging specifications are described in Figure 2.2-5. Scale bar represents 16 μm.

Table 2.2-3. Percentages of nuclei presenting clear signal (coverage) in d-FISH and nd-FISH assays using *DYZ1*-targeting Invader probes.

Probe	d-FISH	nd-FISH
INV1	0%	0%
INV2	90%	80-90%
INV3	30-50%	20-40%
INV4	90%	90%
INV5	0%	0%
INV6	50-70%	20%
INV7	50-70%	20-30%
INV8	50-70%	20-30%
INV9	50-70%	0%
INV10	90%	90%

Pre-treatment with enzymes to digest DNA (DNase I), RNA (RNase A), and protein (Proteinase K) prior to incubation with probes of interest (**INV2** and **INV10**) confirmed that Invader probes target chromosomal DNA, rather than binding to RNA or adhering to nuclear proteins. Nuclei pre-treated with DNase I (Figure 2.4-26) were completely devoid of signal, supporting that Invader probes are binding to chromosomal DNA. Treatment with RNase A and Proteinase K yielded signal for both **INV2** and **INV4**, albeit less and overall fainter signal (Figure 2.4-27). The signal reduction could be attributed to nuclei/chromosome loss during the RNase A and Proteinase K treatment prior to incubation with Invader probe that was not seen in the DNase I treatment. However, the presence of signal indicates that Invader probe signal is not caused by binding to RNA or protein.

Additionally, a negative control was run for both **INV2** and **INV10** (Figure 2.2-7). Both Y chromosome specific Invaders were incubated with a female bovine fibroblast cell line. The incubation was run under denaturing conditions to eliminate the barrier of strand invasion and ensure that a lack of signal is due to specificity of probe. All nuclei for the female controls for both probes were devoid of signal, supporting the specificity of these Y chromosome targeting Invader probes.

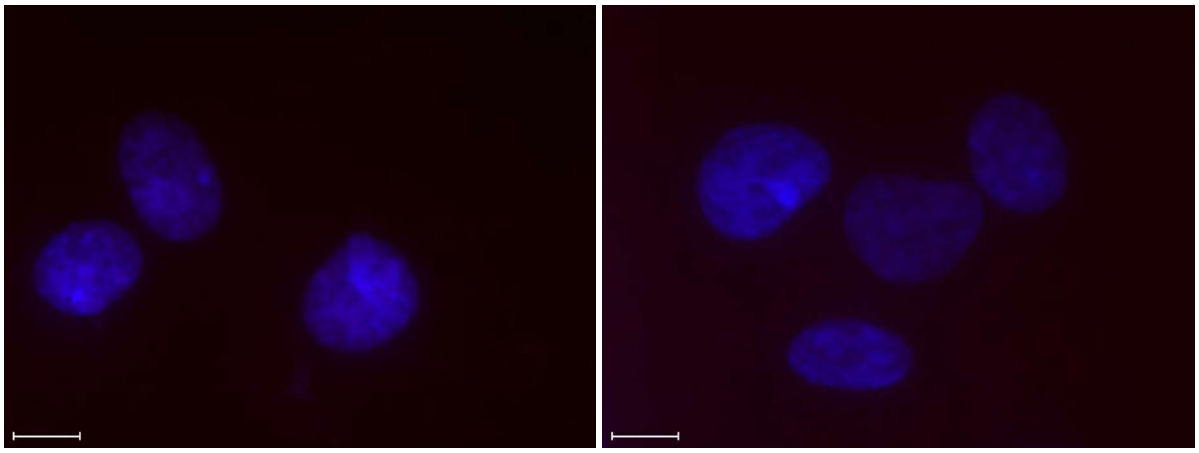


Figure 2.2-7. INV2 (left panel) and **INV10** (right panel) incubated in female bovine fibroblast cells under denaturing conditions. Note the lack of red Cy3 signal. Incubation and imaging specifications are described in Figure 2.2-5. Scale bar represents 16 µm.

Spearman rank correlation analysis: rules of probe design

The results from screening the ten Invader probes were used to determine the key parameters that result in efficient binding to chromosomal DNA at near physiological conditions. To this end, thermal denaturation results (Probe Duplex T_m, Probe Duplex ΔT_m, 5'-**INV**:cDNA T_m, 5'-**INV**:cDNA ΔT_m, 3'-**INV**:cDNA T_m, 3'-**INV**:cDNA ΔT_m, T_m cDNA, TA_{isq}, TA_{DH}), Probe design features (GC%, %mod, # of modifications [#mod], probe length, longest stretch of nucleotides between each hotspot [Unmodified Stretch]), thermodynamic parameters (Probe Duplex ΔG, Probe Duplex $\Delta\Delta G$, 5'-**INV**:cDNA ΔG, 5'-**INV**:cDNA $\Delta\Delta G$, 3'-**INV**:cDNA ΔG, 3'-**INV**:cDNA $\Delta\Delta G$, ΔG cDNA, ΔG_{rec}^{310} of **DH**, ΔG_{rec}^{310} of cDNA), EMSA hairpin assay results (C_{50} and Rec_{100X} at 2.5 and 15 h), and FISH assay results (denaturing and non-denaturing probe coverage) were evaluated *via* Spearman Rank Correlation analysis (Table 2.2-4). Results were ranked from 1 to 10 (i.e. low C_{50} values were ranked closer to 1 and higher C_{50} values ranked closer to 10) and based on each Invaders ranking, statistical relevance to the listed parameters was determined. The Spearman correlation coefficients between nd-FISH coverage (Table 2.2-

3) and all parameters listed above elucidated key factors in effective Invaders. Most strongly correlated to nd-FISH coverage are percent modification (%mod), C_{50} values at 15 h, and d-FISH coverage (Table 2.2-4). First, this suggests a positive denaturing control and C_{50} values once probes have reached their full recognition potential in the hairpin model are decent first approximations for determining which probes have the potential to bind to chromosomal DNA. Second, it gives one factor that should be considered in Invader probe design: percent modification. To this end, we sought verify these results by creating highly modified probes and observe their chromosomal binding capabilities.

Table 2.2-4. Spearman Rank Correlation with nd-FISH Coverage Results

Variables	Correlation (r_s)	p-value
Probe Duplex T_m	-0.063	0.871
Probe Duplex ΔT_m	-0.171	0.660
5'-**INV**:cDNA T_m	0.413	0.236
5'-**INV**:cDNA ΔT_m	0.268	0.454
3'-**INV**:cDNA T_m	0.081	0.036
3'-**INV**:cDNA ΔT_m	0.312	0.381
T_m cDNA	0.106	0.771
TA_{isq}	0.505	0.165
TA_{DH}	0.505	0.165
GC%	0.191	0.597
%mod	**0.738**	**0.015**
#mod	0.547	0.102
Probe Length	-0.360	0.306
Unmodified Stretch	-0.590	0.073
Probe Duplex ΔG	0.108	0.800
Probe Duplex $\Delta\Delta G$	-0.049	0.908
5'-**INV**:cDNA ΔG	0.094	0.797
5'-**INV**:cDNA $\Delta\Delta G$	-0.028	0.939
3'-**INV**:cDNA ΔG	-0.268	0.454
3'-**INV**:cDNA $\Delta\Delta G$	-0.563	0.090
ΔG cDNA	-0.157	0.665
ΔG_{rec} of DH	-0.418	0.303
ΔG_{rec} of cDNA	-0.583	0.129
C_{50} (2.5 h)	-0.211	0.559
Rec_{100X} (2.5 h)	0.224	0.533
C_{50} (15 h)	**-0.710**	**0.049**
Rec_{100X} (15 h)	0.324	0.361
d-FISH Coverage	**0.798**	**0.006**
nd-FISH Coverage	-	-

Design of Optimized Invader probes

We optimized three of the existing Invader probes using the same synthesis methods described above. In the nd-FISH assay, **INV2**, **INV4**, and **INV10** resulted in ~90% of nuclei displaying signal (Table 2.2-3), eliminating them as potential candidates for optimization as they already exhibit optimal levels of binding. Out of the remaining seven, **INV6**, **INV8**, and **INV9** were selected for optimization. These Invader probes were chosen due to their potential for modification and chromosomal DNA recognition. All three have a low percent modification (~20%; Table 2.2-1), many potential "good steps" (i.e. pyrimidine followed by a 3'-purine base pair) for effective Invader monomer incorporation,[28] and show some existing levels of recognition under denaturing and non-denaturing conditions, suggesting their sequences are accessible for binding. With an additional 2-3 hotspots, the percent modification of **INV6**, **INV8**, and **INV9** was increased by 12.5%, 6.6%, and 13.3% respectively to give the newly designed **OPT6**, **OPT8**, and **OPT9** (Table 2.2-5).

Table 2.2-5. Thermal denaturation temperatures (T_m), Thermal Advantages (TA_{isq} and TA_{DH}), Percent Modification (Mod%), and Percent GC-content (GC%) of Optimized *DYZ1*-targeting Invader probes and duplexes between individual probe strands and cDNA.[a]

Probe	Sequence	T_m [ΔT_m] (°C)			TA_{isq} (°C)	TA_{DH} (°C)	Mod%	GC%
		Probe duplex	5'-INV: cDNA	3'-INV: cDNA				
OPT6	5'-Cy3-C**U**G**U**G**C**AAC**U**GGT**U**TG-3' 3'- GA**C**A**C**G**U**TGA**C**CAA**A**C-Cy3-5'	49.0 [-9.0]	75.0 [+17.0]	75.0 [+17.0]	+43.0	+26.0	31.3%	50%
OPT8	5'-Cy3-TT**C**A**C**AGCCC**U**G**U**GC-3' 3'- AAG**U**G**U**CGGGA**C**A**C**G-Cy3-5'	38.0 [-22.0]	77.0 [+17.0]	76.5 [+16.5]	+55.5	+35.0	26.6%	60%
OPT9	5'-Cy3-T**U**A**U**A**U**GC**U**G**U**TCTC-3' 3'- AA**U**A**U**A**C**GA**C**A**A**GAG-Cy3-5'	29.0 [-16.5]	65.0 [+19.5]	64.0 [+18.5]	+54.5	+33.0	33.3%	33%

[a]ΔT_m = change in T_m values relative to corresponding unmodified duplexes. T_ms for the corresponding unmodified DNA duplexes **DNA6** = 58.0 °C, **DNA8** = 60.0 °C, and **DNA9** = 45.5 °C. Thermal denaturation curves were recorded in medium salt buffer ($[Na^+]$ = 110 mM, $[Cl^-]$ = 100 mM, pH 7.0 (NaH_2PO_4/Na_2HPO_4), [EDTA] = 0.2 mM) and each [ON] = 1.0 μM; see main text for definition of TA. NT = no clear transition (very broad) at wavelengths 230-280 nm. For structures of **A**, **C** and **U**, see Figure 2.1-1.

Thermal denaturation and thermodynamic properties of Optimized Invader Probes

To characterize the new probes, thermal denaturation temperatures (T_ms) were determined for the optimized probes and the corresponding duplexes between individual probe strands and cDNA (Table 2.2-5). First to note is the decrease in T_m of each optimized probe in comparison to the original Invader probes due to the additional instability instilled by additional hotspots (compare Table 2.2-1 and 2.2-5). Furthermore, there is an overall increase in stability each Optimized Invader probe strand *vs.* cDNA, facilitated by the additional pyrene moieties increasing stacking interactions within the recognition complex (compare Table 2.2-1 and 2.2-5). These differentials give higher *TA* values for both isosequential and hairpin target in comparison to the original ten Invaders, suggesting potentially superb target binding in the hairpin and FISH assays. This is supported by more prominent ΔG^{310}_{rec}c values for each Optimized Invader (-84, -93, and -59 kJ/mol) in comparison to the unoptimized versions (-22,

-52, and -19 kJ/mol) (Table 2.4-6), indicating very favorable binding energetics that should lead to improved chromosomal DNA binding.

Recognition of mixed-sequence model DNA hairpin targets by Optimized Invader Probes

The same hairpin screen and dose-responses experiments conducted above were used to evaluate the optimized probes (Figure 2.2-8 and Figure 2.2-9). **OPT8** and **OPT9** reached nearly full recognition of hairpin incubated with hairpin at 100x excess for 15 h – a vast improvement from the unoptimized probes (compare Table 2.2-2 and Table 2.2-6). Peculiarly, **OPT6** had a negligible increase in recognition in comparison to **INV6**, despite its additional hotspots, and did not reach 50% recognition even at high levels of excess (Figure 2.2-9). **OPT8** and **OPT9** had favorable reductions in C_{50} values (0.2 μM and 0.3 μM) in comparison to unoptimized **INV8** and **INV9** (0.6 μM and 1.5 μM) (Table 2.2-2 and Table 2.2-6). As low C_{50} values at 15 h are highly correlated to recognition of chromosomal DNA under non-denaturing conditions, this first screen of potential binding suggested at least 2 out of 3 of the Optimized Invader probes should result in chromosomal DNA recognition at high levels of coverage.

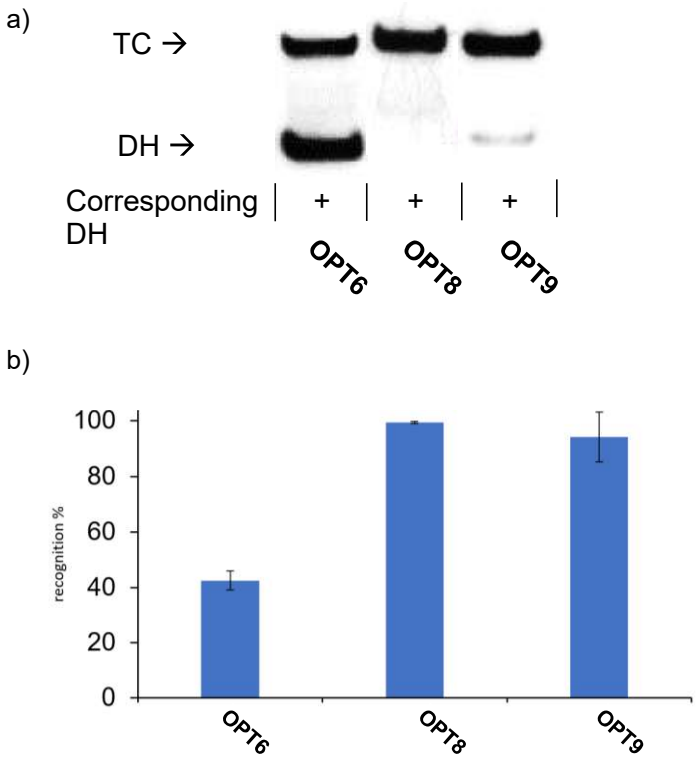


Figure 2.2-8. a) Representative gel electrophoretograms from recognition experiments at 15 h between a 100-fold molar excess of Optimized Invader probes **OPT6**, **OPT8**, and **OPT9** and their corresponding DNA hairpin targets **DH6**, **DH8**, and **DH9**. b) Histograms depict averaged results from at least three recognition experiments with error bars representing standard deviation. TC = ternary complex. DIG-labeled **DH6**, **DH8**, and **DH9** (34.4 nM, sequences shown in Table SX) were incubated with the corresponding Invader probe in HEPES buffer (50 mM HEPES, 100 mM NaCl, 5 mM MgCl2, pH 7.2, 10% sucrose, 1.44 mM spermine tetrahyrdochloride) for 15 h at 37 °C. Incubation mixtures were resolved on 12% non-denaturing TBE-PAGE slabs (~70 V, ~4 °C, ~1.5 h).

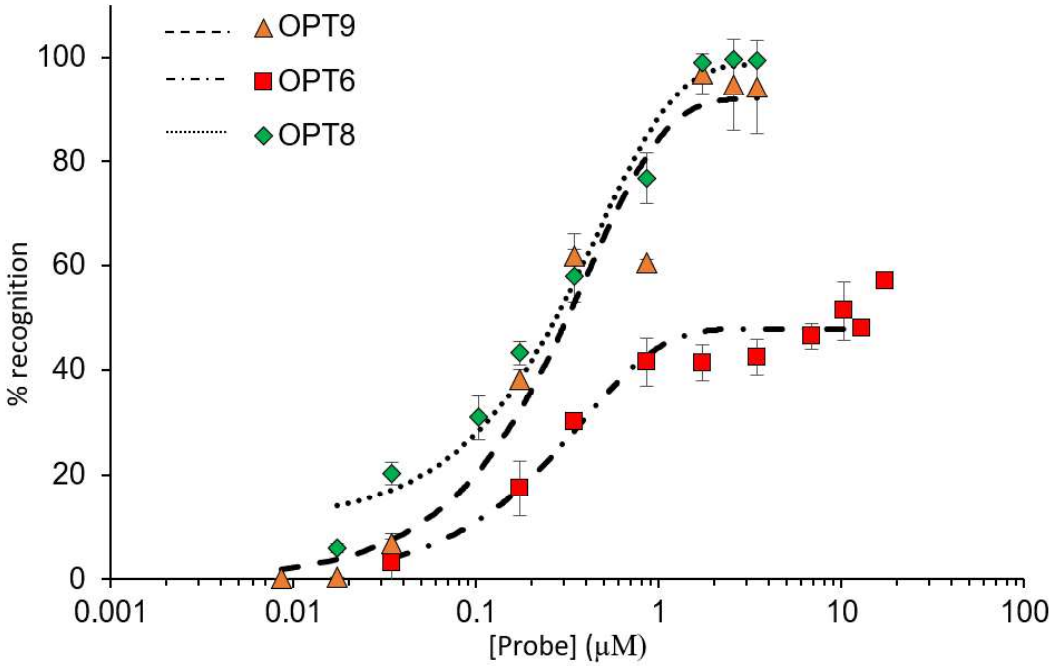


Figure 2.2-9. Dose-response curves for **OPT6, OPT8, and OPT9** at 37 °C for 15 h. Experimental conditions are as described in Figure 2.2-8, except for variable probe concentrations

Table 2.2-6. C_{50} values at 15 h for Invader probes studied herein.[a]

Probes	C_{50} (μM)	Rec_{100X} (%)
OPT6	ND	42 ± 3.4
OPT8	0.2	99 ± 0.4
OPT9	0.3	99 ± 8.9

[a] Rec_{100} = level of DNA hairpin recognition using 100-fold molar probe excess. C_{50} values were determined from dose-response curves shown in Figure 2.2-9.

Detection of chromosomal DNA by Optimized Invader probes

With results confirming their binding potential, the optimized probes were evaluated in our FISH assay under denaturing and non-denaturing conditions (Figure 2.2-10). To verify our predictions that percent modification is a main factor in facilitating chromosomal DNA recognition, **OPT6**, **OPT8**, and **OPT9** were incubated with the male bovine nuclei under the same non-denaturing conditions as the original ten Invader screen. Based on T_m values, thermodynamic properties, and hairpin binding, **OPT8** behaved as expected, nearly doubling

the number of nuclei presenting clear signal under non-denaturing conditions (Table 2.2-7). Unexpectantly, **OPT6** had the highest levels of coverage out of the three optimized probes, under both denaturing or non-denaturing conditions (~90%, Table 2.2-7), despite its modest recognition of hairpin (Table 2.2-6). Even more surprising was **OPT9**, which despite having favorably small C_{50} values (Table 2.2-6), did not reach high coverage percentages like **OPT6** and **OPT9** (Table 2.2-7). Although the optimization did not result in perfect recognition of chromosomal DNA for all optimized probes, each optimized probe outperformed its unoptimized predecessor under both FISH Assay conditions. Particularly **OPT9**, demonstrates the predictive power of our statistical analysis. Unoptimized **INV9** was unable to detect chromosomal DNA in our nd-FISH assay; however, upon increasing the modification density of the probe, **OPT9** was able to give clear signal in 20-30% of nuclei (Table 2.2-7). By increasing the affinity of the Invader probe, **OPT9** was able to acquire the adequate energetic driving force needed to detect its chromosomal target at physiologically relevant conditions. Overall these results confirm our Spearman Rank Correlation analysis prediction that an increase in modification density is correlated with an increase in chromosomal dsDNA recognition.

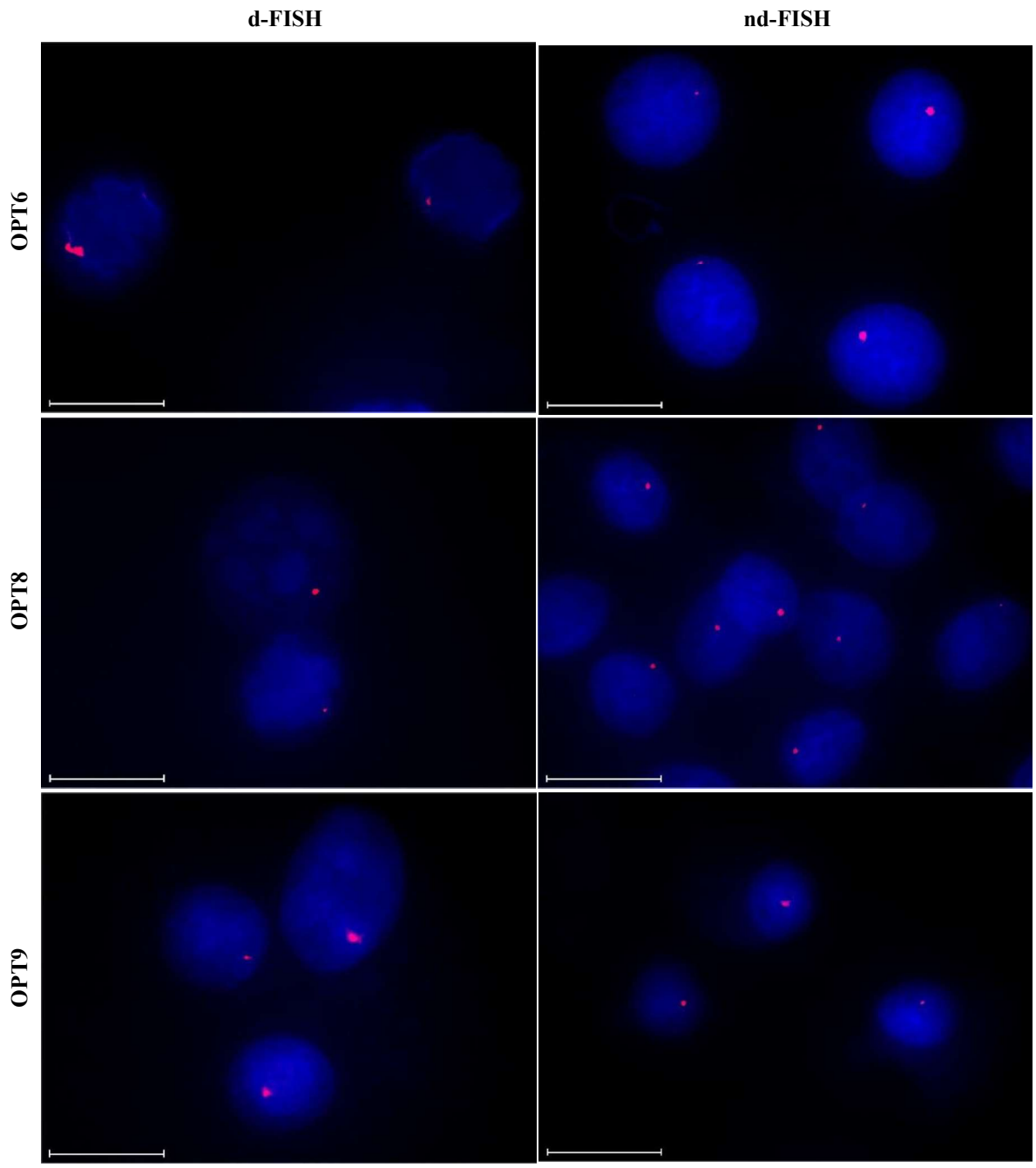


Figure 2.2-10. Images from FISH experiments using DYZ1-targeting **OPT6**, **OPT8**, and **OPT9** under denaturing (80 °C, 5 min) (left), or non-denaturing (3 h, 38.5 °C) conditions (right). Incubation and imaging specifications are described in Figure 2.2-5 Scale bar represents 16 µm.

Table 2.2-7 Percentages of nuclei presenting clear signal (coverage) in d-FISH and nd-FISH assays using *DYZ1*-targeting Optimized Invader probes.

Probe	d-FISH	nd-FISH
OPT6	90%	80-90%
OPT8	90%	70-80%
OPT9	70-80%	20-30%

2.3 Conclusions

To summarize, this studied screened ten Invader probes functionalized with multiple energetic hotspots composed of 2'-*O*-(pyren-1-yl)methyl RNA monomers for potential recognition of different chromosomal DNA targets. Based on T_m data, these Invader probes displayed the expected and necessary stability characteristics for double-strand invasion (i.e., probe duplexes were more labile than the duplexes between individual probe strands and cDNA resulting in positive *TA* values). FISH studies revealed probes capable of detecting chromosomal DNA. Based on these results, a Spearman Rank Correlation analysis revealed that high levels of modification (i.e. incorporation of many energetic hotspots) correlates with increased signaling of chromosomal DNA under non-denaturing conditions. Subsequently, three of the ten Invader probes were optimized by incorporating 2-3 additional hotspots and then were evaluated in our FISH assay.

Although utilization of the Spearman Rank Correlation gave mixed results in the level of chromosomal recognition under non-denaturing conditions, all optimized probes resulted in improved chromosomal DNA detection. In the case of **OPT9**, increasing its modification density by 13.3% gave 20-40% coverage in the nd-FISH assay - a vast improvement over the unoptimized **INV9**, which was unable to produce signal. Optimization of **INV8** to **OPT8** nearly double the number of signals in the nd-FISH assay. Moreover, signal was quadrupled in the optimization of **INV6** to **OPT6** under non-denaturing conditions. While these results

confer with our statistically derived hypothesis, there is variability between probes. Our Spearman coefficients showed that low C_{50} values at 15 h correlate with high levels of nd-FISH signaling. Based on C_{50} values, **OPT9** should have gave superb coverage while **OPT6** should have given relatively less coverage. Our nd-FISH results for optimized probes revealed that this is not the case. This can be explained from a multitude of perspectives. First, while our hairpin assay is a decent predictor model, it cannot mimic the exact conditions of the chromosomal DNA target. Chromosomal DNA is supercoiled and interacts with different proteins,[33] whether it be for structural (i.e. histones) or functional (i.e. polymerases) purposes, which will affect the accessibility of target sequence.[35] The hairpin assay cannot account for this type of variance and thus our Spearman Correlation that utilizes the hairpin assay results cannot perfectly predict binding to chromosomal DNA. Second, our Spearman Correlation only considers the results of a small sample size. Perhaps with a larger sample size, different factors that contribute to chromosomal binding would be identified. For example, there were several parameters that showed a modest correlation to chromosomal binding, (i.e. *TA*, ΔG^{310}_{rec}, unmodified stretches) that may be augmented and prove to be statistically significant with additional data. Additionally, incorporating additional probes with different targets may highlight a sequence dependence (i.e. location of GC or AT steps, number of GC or AT bp in a row, etc). While there many factors to consider for future work, this study emphasizes that future probes must be densely modified to target mixed-sequence chromosomal DNA, and will serve as a reference point for all future Invader probes.

2.4 Supplementary Material

Protocol - Synthesis and purification of probe strands

Individual Invader strands - i.e., oligodeoxyribonucleotides (ONs) modified with 2′-*O*-(pyren-1-yl)methyl-RNA monomers - were synthesized on an Expedite DNA synthesizer (0.2 μmol scale) using columns packed with long-chain alkylamine-controlled pore glass (LCAA-CPG) solid support with a pore size of 500 Å. Standard protocols were used for incorporation of DNA phosphoramidites. The 2′-*O*-(pyren-1-yl)methyl-RNA phosphoramidites were prepared as previously described for U monomer[36] and C/A monomers[28] and incorporated into ONs via extended hand-couplings (15 min, ~45-fold molar excess, 0.02 M in anhydrous acetonitrile, using 0.01 M 4,5-dicanoimidazole as the activator) and oxidation (45 s) resulting in coupling yields of at least 85%. Cy3-labeling of Invader strands was accomplished by incorporating a commercially available Cy3 phosphoramidite (Glen Research) into ONs by hand-coupling (4,5-dicyanoimidazole, 3 min, anhydrous CH3CN). Treatment with 32% aq. ammonia (55 °C, 17 h) ensured deprotection and cleavage from solid support. DMT-protected ONs were purified via ion-pair reverse phase HPLC (XTerra MS C18 column: 0.05 M triethyl ammonium acetate and acetonitrile gradient) followed by detritylation (80% acetic acid, 20 min) and precipitation (NaOAc, $NaClO_4$, acetone, -18 °C, 16 h). The purity and identity of the synthesized ONs was verified using analytical HPLC (Figure 2.4-5) and MALDI-MS analysis (Tables 2.4-1 and 2.4-2 and Figures 2.4-1, 2.2-2, 2.4-3, 2.4-4) recorded on a Quadrupole Time-of-Flight (Q-TOF) mass spectrometer using 3-hydoxypicolinic acid matrix.

Protocol - thermal denaturation experiments.

ON concentrations were estimated using the following extinction coefficients (OD_{260}/μmol): G (12.01), A (15.20), T (8.40), C (7.05), pyrene (22.4)[37] and Cy3 (4.93)[38]. Thermal denaturation temperatures (T_ms) of duplexes (1.0 μM final concentration of each strand) were

determined on a Cary 100 UV/VIS spectrophotometer equipped with a 12-cell Peltier temperature controller and measured as the maximum of the first derivative of thermal denaturation curves (A_{260} *vs. T*) recorded in medium salt buffer (T_m buffer: 100 mM NaCl, 0.2 mM EDTA, and pH 7.0 adjusted with 10 mM Na_2HPO_4 and 5 mM Na_2HPO_4). Strands were mixed in quartz optical cells with a path length of 1.0 cm and annealed by heating to 85 °C (2 min) followed by cooling to the starting temperature of the experiment. The temperature of the denaturation experiments ranged from at least 15 °C below the T_m to 15 °C above the T_m (though not above 95 °C). A temperature ramp of 1.0 °C/min was used in all experiments. Reported T_ms are averages of two experiments within ± 1.0 °C.

Protocol - Electrophoretic mobility shift assays

The nd-PAGE assay was performed essentially as previously described.[27] DNA hairpins (DH) were obtained from commercial sources and used without further purification. Hairpins were 3'-labelled with digoxigenin (DIG) using the 2nd generation DIG Gel Shift Kit (Roche Applied Bioscience) as recommended by the manufacturer. Briefly, 11-digoxigenin-ddUTP was incorporated at the 3′-end of the hairpin (100 pmol) using a recombinant terminal transferase. The reaction mixture was quenched through addition of EDTA (0.05 M), diluted to 68.8 nM, and used without further processing. Invader probes (concentrations as specified) were heated (90 °C, 2 min) and cooled to room temperature over approximately 30 min, and subsequently incubated with the corresponding DIG-labeled DNA hairpin (final concentration 34.4 nM) in HEPES buffer (50 mM HEPES, 100 mM NaCl, 5 mM $MgCl_2$, pH 7.2, 10% sucrose, 1.44 mM spermine tetrahydrochloride) at 37 °C for the specified time. Following incubation, loading dye (6X) was added and the mixtures were loaded onto 12% non-denaturing TBE-PAGE slabs (45 mM tris-borate, 1 mM EDTA; acrylamide:bisacrylamide

(19:1)). Electrophoresis was performed using constant voltage (~70 V) at ~4 °C for ~1.5 h. Bands were subsequently blotted onto positively charged nylon membranes (~100 V, 30 min, ~4 °C) and cross-linked through exposure to UV light (254 nm, 5 x 15 W bulbs, 5 min). Membranes were incubated with anti-digoxigenin-alkaline phosphatase F_{ab} fragments as recommended by the manufacturer and transferred to a hybridization jacket. Membranes were incubated with the chemiluminescence substrate (CSPD) for 10 min at 37 °C, and chemiluminescence of the formed product was captured on X-ray films. Digital images of developed X-ray films were obtained using a BioRad ChemiDoc™ MP Imaging system, which also was used for densitometric quantification of the bands. The percentage of dsDNA recognition was calculated as the intensity ratio between the recognition complex band and unrecognized hairpin. An average of three independent experiments is reported along with standard deviations (±). Shown electrophoretograms may be composite images of lanes from different runs. Non-linear regression was used to fit data points from dose–response experiments. A script written for the "Solver" module in Microsoft Office Excel was used to fit data points from dose–response experiments to the following equation: y = C + A (1 - e-kt) where C, A and k are fitting constants. The resulting equation was used to calculate C_{50} values by setting y = 50 and solving for t.

Protocol - cell culture and nuclei preparation

The male bovine kidney cell line (MDBK, ATCC: CCL-22, Bethesda, MD) and female fibroblast cell line (Minitube, Verona, WI) were maintained in DMEM with GlutaMax (Gibco, 10569-010) and 10% fetal bovine serum (Invitrogen). Cells were cultured in separate 25 mL or 75 mL flasks at 38.5 °C in a 5% CO_2 atmosphere for 72-96 h to achieve 70-80% confluency. At this point, colcemid (65 µL per 5 mL of growth media) was added and the adhered cells

were incubated at 37 °C and 5% CO_2 for an additional 20 min. At this point, the medium was replaced with pre-warmed 0.05% Trypsin-EDTA in DMEM to detach adherent cells (37 °C, up to 8 min). The cell suspension was transferred to a tube and centrifuged (10 min, 1000 rpm). The supernatant was discarded and the dislodged cell pellet was incubated with a hypotonic KCl solution (5-8 mL, 0.075 M, 20 min), followed by addition of fixative solution (10 drops, MeOH:AcOH, 3:1), and further incubation with gentle mixing (10 min, room temperature). The suspension was centrifuged (1000 rpm, 10 min), the supernatant discarded, and additional fixative solution (5-8 mL) added to the nuclei suspension. This was followed by gentle mixing and incubation (30 min, room temperature). The centrifugation/resuspension/incubation steps in fixative solution were repeated three additional times. The final pellet – containing somatic nuclei – was resuspended in methanol and glacial acetic acid (3:1, v/v) and stored at -20 °C until use.

Protocol - preparation of slides for FISH assays

The prepared nuclei suspension was warmed to room temperature and resuspended in fresh fixative solution. Glass microscope slides (Fisher Scientific) were dipped in distilled water to create a uniform water layer across the slide. An aliquot of the nuclei suspension (3-5 µL or enough to cover the slide) was dropped onto the slide, while holding the slide at a 45° angle, and allowed to run down the length of the slide. Slides were then allowed to dry at a ~20° angle in an environmental chamber at 28 °C and a relative humidity of 38%.

Protocol - fluorescence in situ hybridization

An aliquot of labeling buffer (~200 µL) consisting of 30 ng of Cy3-labeled Invader probe per 200 µL of PCR buffer (20 mM Tris, 100 mM KCl, pH 8.0) was added to each slide. This 200

μL solution (termed 1x in Table 2.4-12) was placed on each slide. Invader probe concentrations were varied between 2x-4x for the optimization studies described in Table SX.

For use in denaturing FISH assays, slides with labeling buffer were placed on a heating block (5 min, 80 °C) and covered with a lid to prevent evaporation of the labeling buffer. For use in non-denaturing FISH assays, slides with labeling buffer were placed in a glass culture disk, covered with a lid, and incubated in an oven (3 h, 37.5 °C). Following incubation, slides were washed (3 min, 37.5 °C) in a chamber with TE Buffer (10 mM Tris, 1 mM EDTA, pH 8.0) and allowed to dry at room temperature. Once dried, Gold SlowFade plus DAPI (3 μL, Invitrogen) was placed directly on each slide and a round glass coverslip was mounted for fluorescence imaging.

A Nikon Eclipse Ti-S Inverted Microscope, equipped with Cy3 and DAPI filter sets, was used to visualize nuclei at 60x magnification. Images of fluorescently labeled nuclei were captured and processed with NIS-Elements BR 4.20 software.

In experiments aiming at optimizing signaling output in nd-FISH assays, the concentration of the Invader probe was varied from 1x to 4x. The coverage of signal (i.e., percentage of nuclei with representative signals) was estimated by manually evaluating 100+ nuclei per Invader probe, for each d-FISH and nd-FISH assays. The quality of the signal and level of background noise seen with each Invader probe was assessed on a scale between 1-3 (Table 2.4-12) across ~200 nuclei. Thus, signals of low, medium, or high intensity were scored 1, 2 or 3, respectively, while high, medium, or low backgrounds were scored 1, 2, or 3, respectively. An intensity score of “3” reflects a single signal that can be clearly discerned, whereas an intensity score of “1” reflects no discernable signal. An intensity score of “2” is given to signals falling between these two extremes. A background score of “nd” represents high levels of non-specifically

bound probes (i.e., multiple, scattered spots) and/or residual probe masking potential signal (i.e., a strong but diffuse signal haze), whereas a background score of "3" is given if minimal non-specific binding/residual signal is observed. A background score of "2" is given to background signals falling between these two extremes. The representative images shown in Figures 2.4-23, 2.4-24, and 2.4-25 reflect the "dFISH" and "[1x]" rows of Table 2.4-12.

DNase, RNase, and proteinase control experiments required pre-treatment of MDBK cells fixed nuclei prior to incubation with Invader probes. DNase pre-treatment: Cloned DNase I (RNase Free) (Takara N101 JF) was mixed in 1x Reaction Buffer (diluted 10x Cloned DNase I Buffer II (Takara A301)) per manufacturer's recommendation. The solution was pipetted onto slides with fixed nuclei in 50 mL amounts. DNase I was incubated at 37.5 °C for 20 minutes and then rinsed with TE Buffer. RNase pre-treatment: 1 µL of RNase A (Fisher Reagents BP2539-100) (5 mg/mL) in 100 µL of buffer (10 mM Tris-HCl, pH 6.5) was placed on slides and incubated for 15 min at 37.5 °C and then rinsed with TE buffer. Proteinase Pre-treatment: 1 µL of Proteinase K (Fisher BioReagents BP1700-100) (6.25 µg/mL) was added to 200 µL of buffer (10 mM Tris-HCl, pH 7.5) and incubated with fixed nuclei for 10 min at 37.5 °C and then rinse with TE buffer.

Protocol – Spearman Rank Correlation calculations

Spearman Rank Correlation coefficients were calculated with the XRealStat function add-on for Microsoft ® Excel ®. Invader Probes were rank 1 to 10 for each parameter and these rankings were compared to determine correlation between two parameters. For example, low C_{50} values, high Rec_{100X} (%), and more negative ΔG^{310}_{rec} would rank closer to 1 while high C_{50} values, low Rec_{100X} (%), and relatively more positive ΔG^{310}_{rec} values would rank closer to

10. Statistical significance of correlation values was determined with p-values less than the α-value of 0.05.

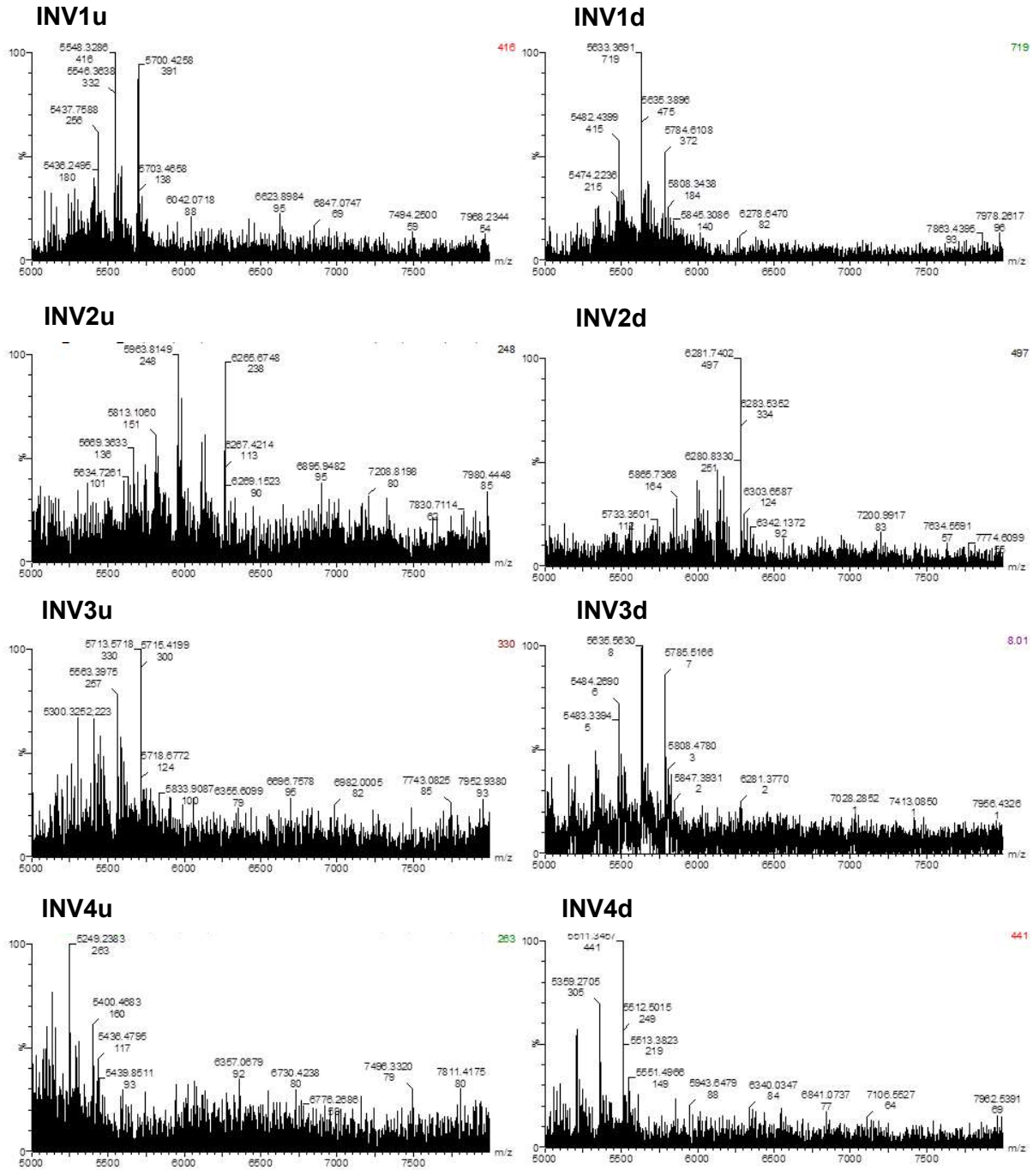


Figure 2.4-1. MALDI-MS Spectra of Individual Strands (up and down) of Invader Probes **INV1**, **INV2**, **INV3**, and **INV4**

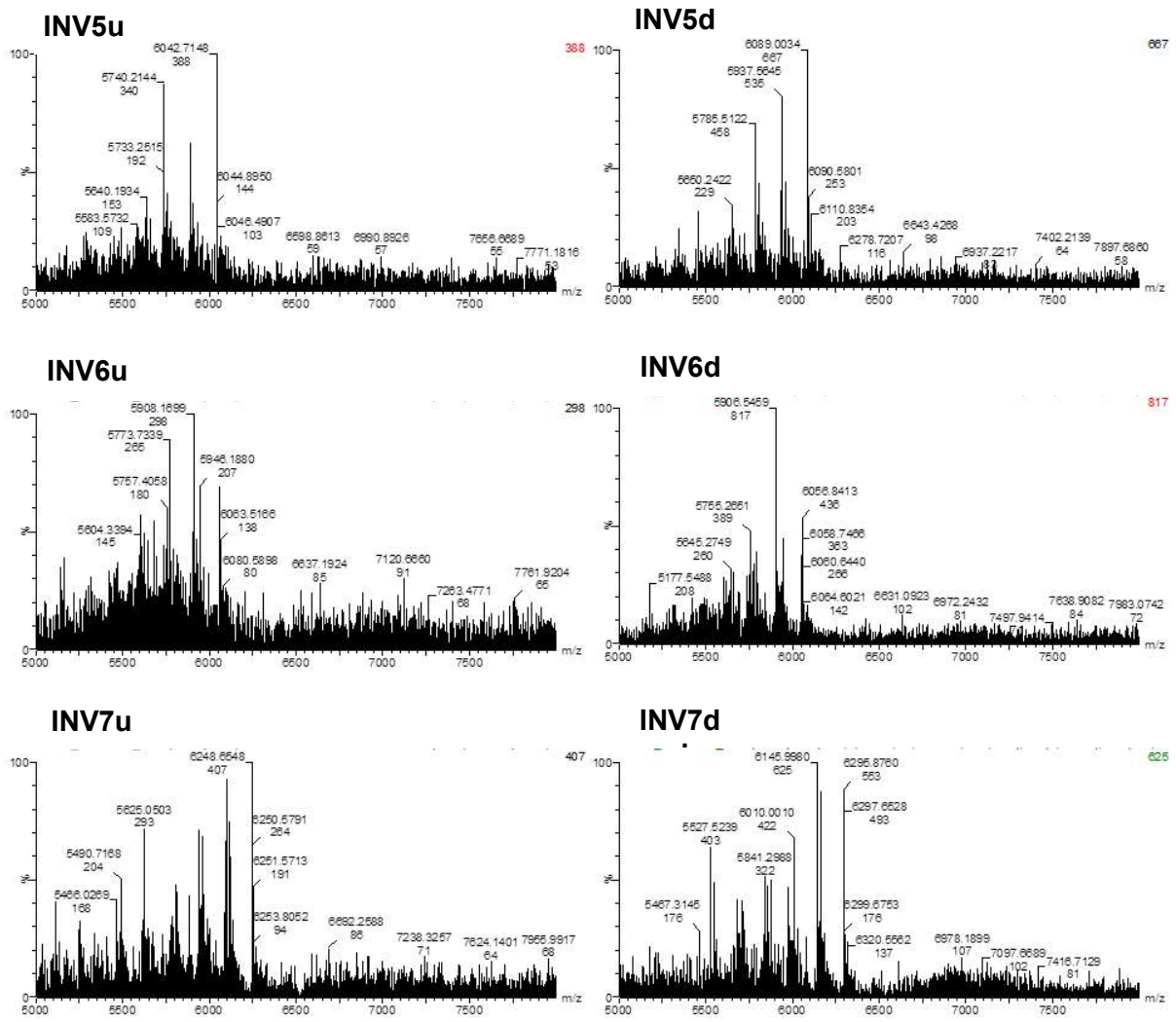


Figure 2.4-2. MALDI-MS Spectra of Individual Strands (up and down) of Invader Probes **INV5**, **INV6**, and **INV7**

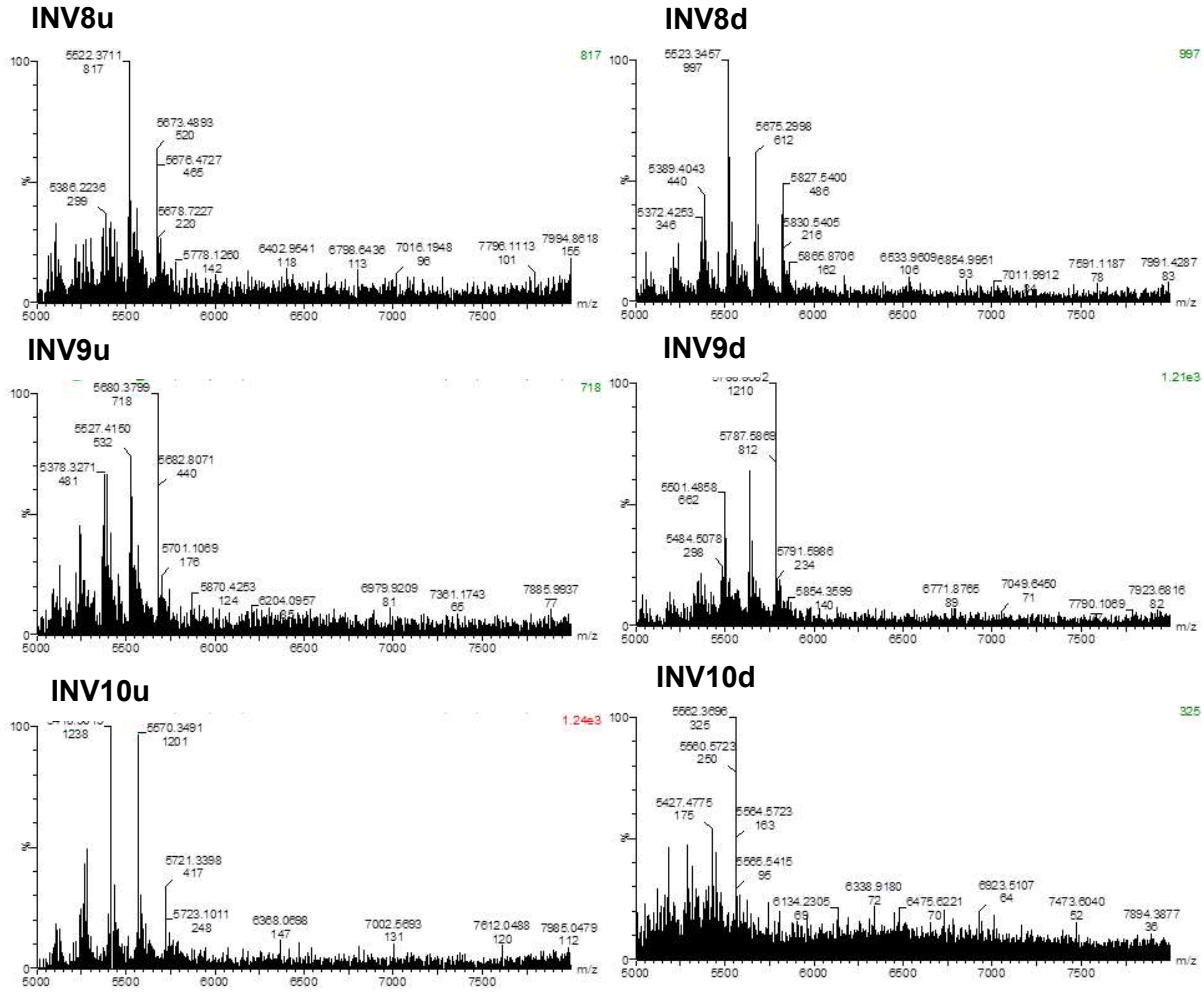


Figure 2.4-3. MALDI-MS Spectra of Individual Strands (up and down) of Invader Probes **INV8**, **INV9**, and **INV10**

Table 2.4-1. MALDI-MS of individual Invader probe strands denoted up (u) or down (d).

ON	Sequence	Obs. *m/z* $[M+H]^+$	Calc. *m/z* $[M+H]^+$
INV1u	5'-Cy3-T**U**ATCAGCAC**U**G**U**GC-3'	5700	5697
INV1d	3'- AA**U**AGTCGTGA**C**A**C**G-Cy3-5'	5784	5783
INV2u	5'-Cy3-A**U**AC**U**GGTTTG**U**G**U**TC-3'	6265	6264
INV2d	3'- TA**U**GA**C**CAAACA**C**A**A**G-Cy3-5'	6280	6279
INV3u	5'-Cy3-T**U**G**U**GCCCTGGC**A**AC-3'	5713	5712
INV3d	3'- AA**C**A**C**GGGACCGT**U**G-Cy3-5'	5785	5783
INV4u	5′-Cy3-**A**GCCC**U**GTGCC**C**TG-3'	5400[b]	5398
INV4d	3′- T**C**GGGA**C**ACGGG**A**C-Cy3-5'	5511[b]	5510
INV5u	5'-Cy3-G**A**TTTCAGCCA**U**G**U**GC-3'	6042	6040
INV5d	3'- CT**A**AAGTCGGTA**C**A**C**G-Cy3-5'	6089	6086
INV6u	5'-Cy3-C**U**G**U**GCAACTGGT**U**TG-3'	6063	6057
INV6d	3'- GA**C**A**C**GTTGACCAA**A**C-Cy3-5'	6056	6055
INV7u	5'-Cy3-C**U**G**U**GCAA**U**ATTT**U**GT-3'	6248	6247
INV7d	3'- GA**C**A**C**GTTA**T**AAAA**C**A-Cy3-5'	6295	6295
INV8u	5'-Cy3-TT**C**ACAGCCC**U**G**U**GC-3'	5673	5673
INV8d	3'- AAG**U**GTCGGGA**C**A**C**G-Cy3-5'	5827	5824
INV9u	5'-Cy3-T**U**A**U**ATGCTG**U**TCTC-3'	5680	5678
INV9d	3'- AA**U**A**U**ACGACA**A**GAG-Cy3-5'	5787	5787
INV10u	5'-Cy3-G**U**G**U**AGTG**U**A**U**ATG-3'	5721	5720
INV10d	3'- CA**C**A**U**CACA**U**A**U**AC-Cy3-5'	5560	5561

[b]MALDI-MS has been previously reported

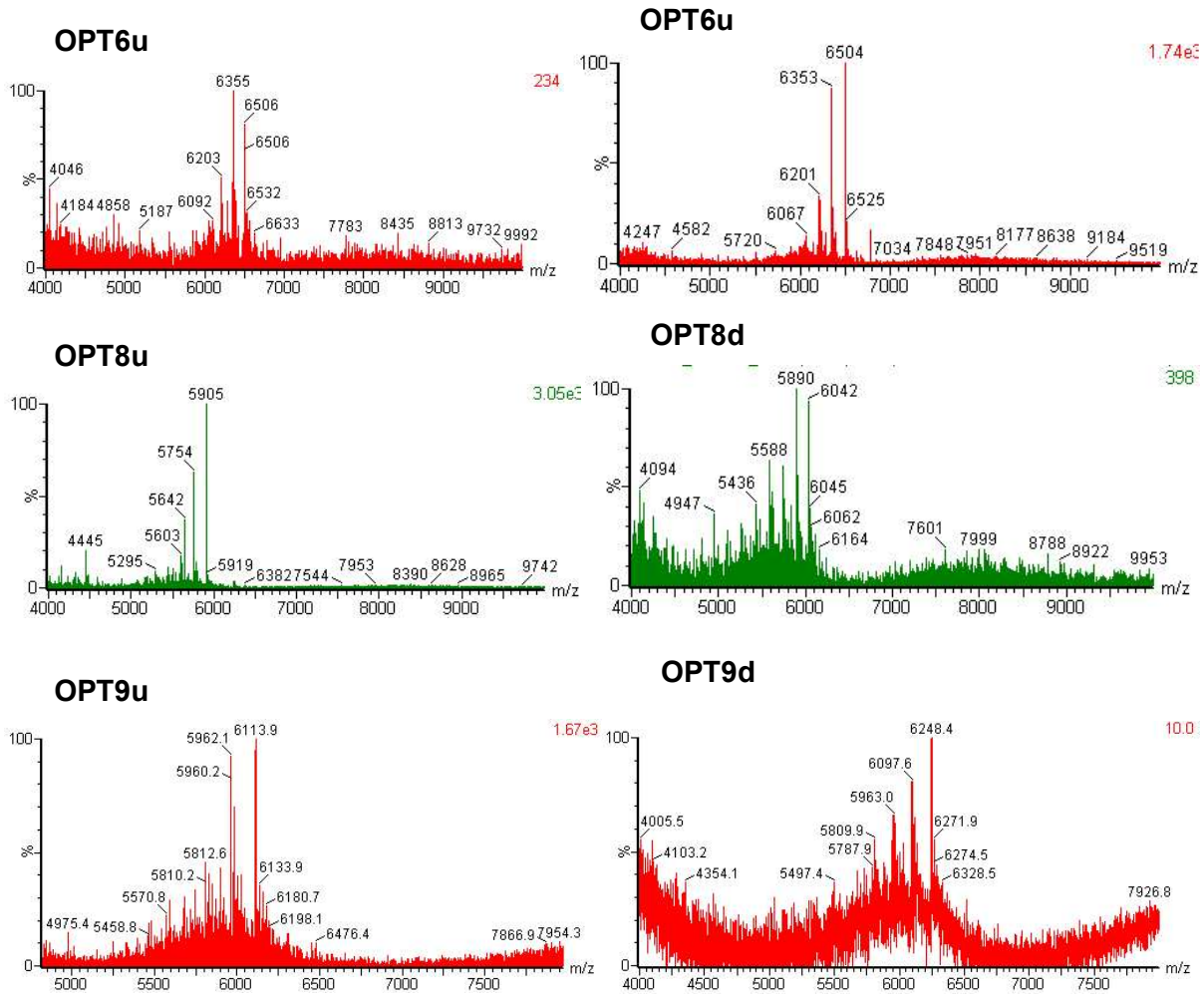


Figure 2.4-4. MALDI-MS Spectra of Individual Strands (up and down) of Invader Probes **OPT6**, **OPT8**, and **OPT9**

Table 2.4-2. MALDI-MS of individual Optimized Invader probe strands denoted up (u) or down (d).

ON	Sequence	Obs. *m/z* $[M+H]^+$	Calc. *m/z* $[M+H]^+$
OPT6u	5'-Cy3-C**U**G**U**G**C**AAC**U**GGT**U**TG-3'	6506	6504
OPT6d	3'- GA**C**A**C**G**U**TGA**C**CAA**A**C-Cy3-5'	6504	6502
OPT8u	5'-Cy3-TT**C**A**C**AGCCC**U**G**U**GC-3'	5905	5903
OPT8d	3'- AAG**U**G**U**CGGGA**C**A**C**G-Cy3-5'	6042	6041
OPT9u	5'-Cy3-T**U**A**U**A**U**GC**U**G**U**TCTC-3'	6113	6112
OPT9d	3'- AA**U**A**U**A**C**GA**C**A**A**GAG-Cy3-5'	6248	6247

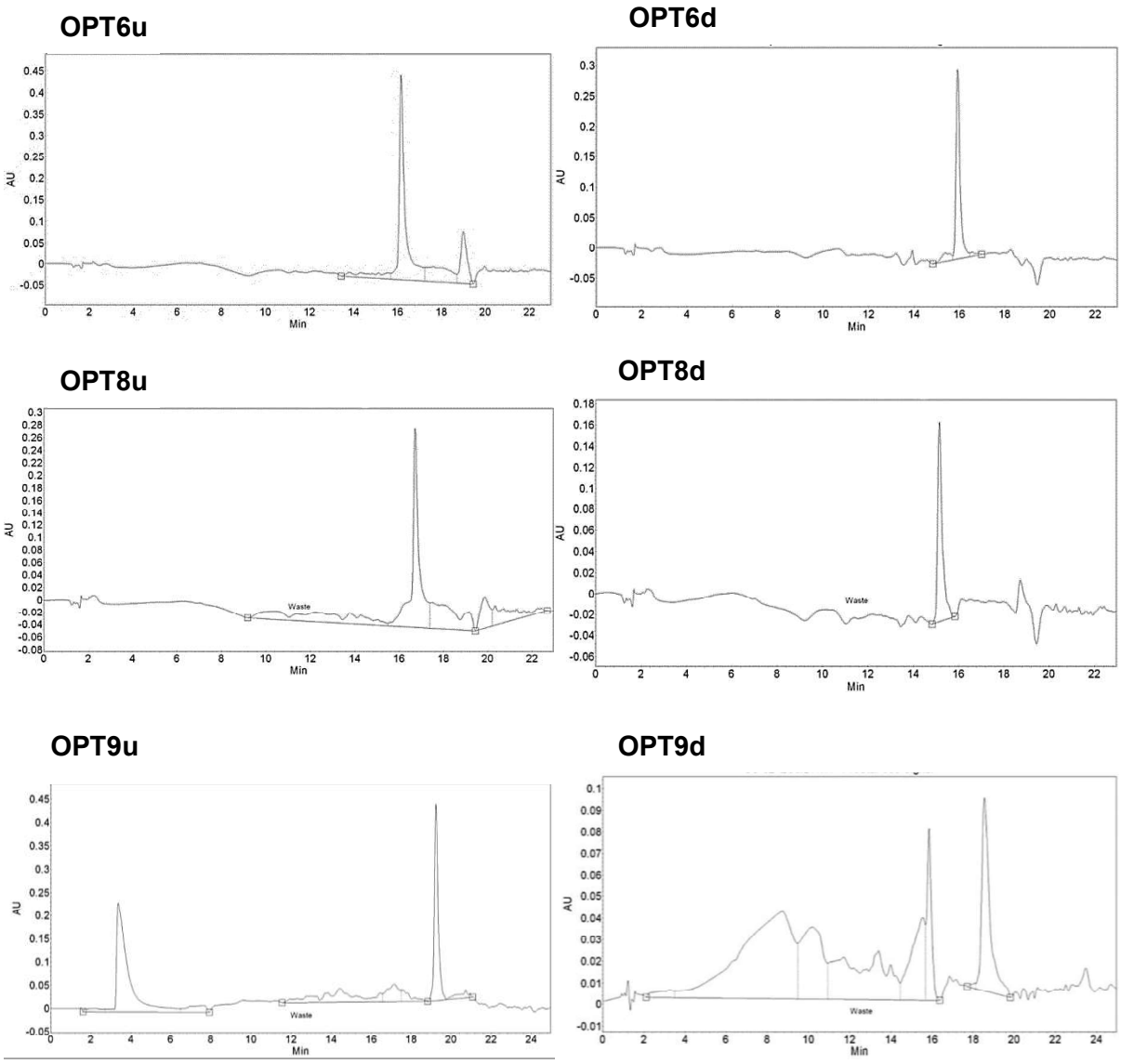


Figure 2.4-5. HPLC Spectra of Optimized Probes

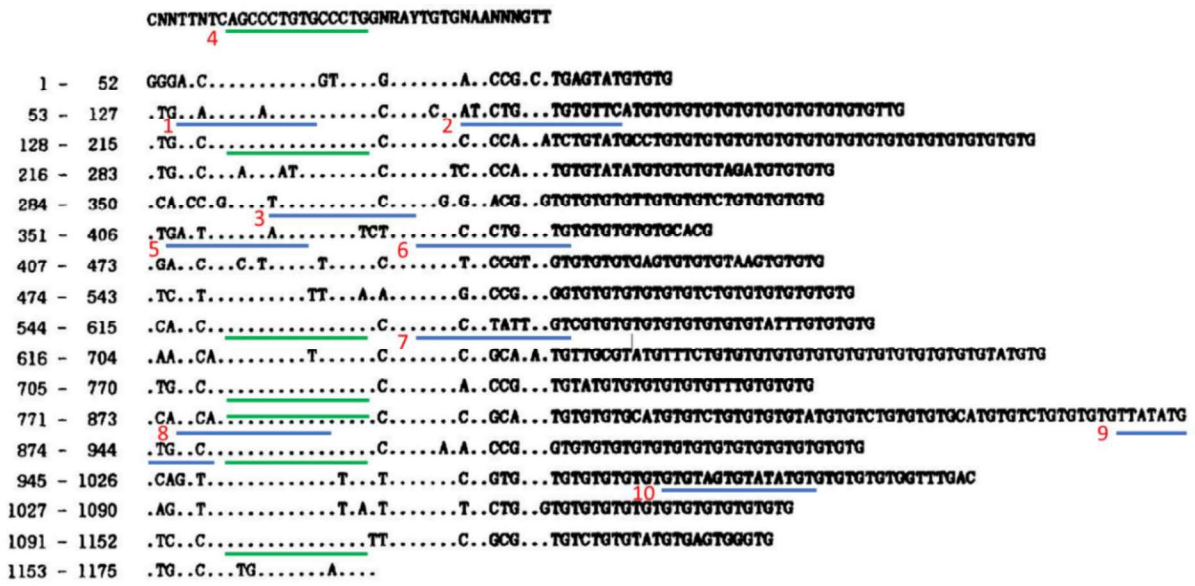


Figure 2.4-6. Position of sequences within the *DYZ-1* satellite gene on the bovine (*Bos taurus*) Y chromosome targeted by the different Invader probes.[26] The target sequence for **INV4** (shown as a green line) is presented six times within the tandem repeat (~6 x 10^4 tandem repeats); all other target sequences are present only once.

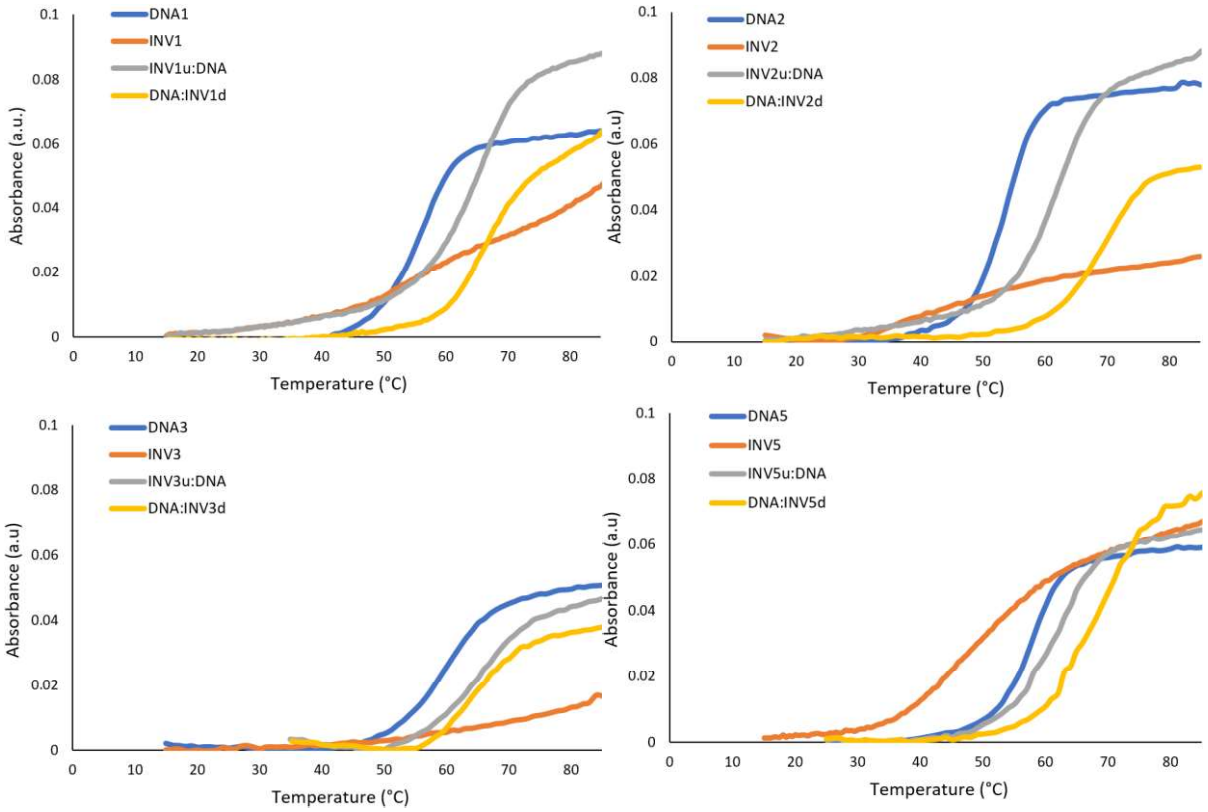


Figure 2.4-7. Representative thermal denaturation curves for Invader probes **INV1**, **INV2**, **INV3**, and **INV5** and the corresponding duplexes between individual probe strands and cDNA, and unmodified reference DNA duplexes (**DNA1**, **DNA2**, **DNA3**, **DNA5**).

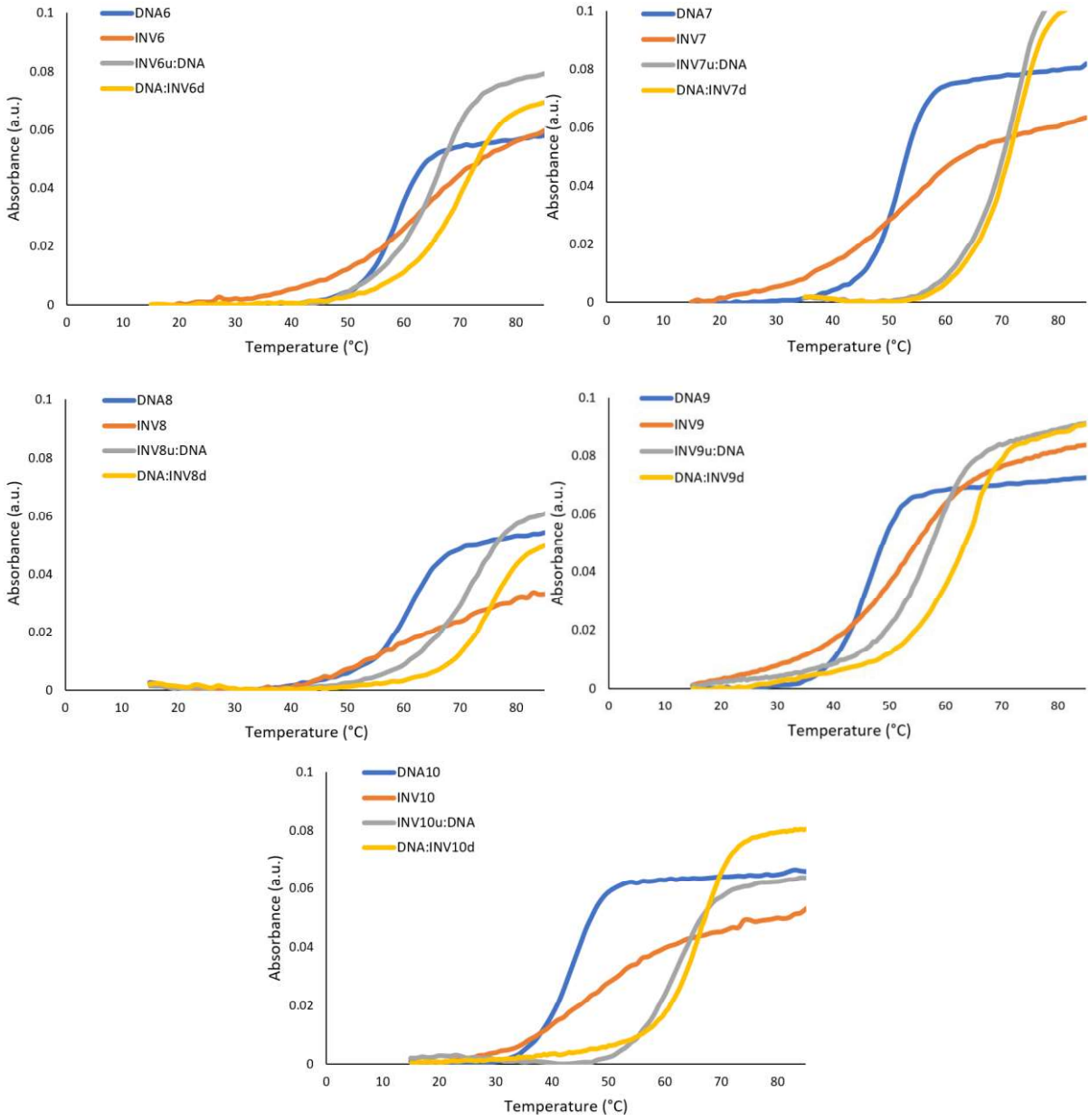


Figure 2.4-8. Representative thermal denaturation curves for Invader probes **INV6-INV10** and the corresponding duplexes between individual probe strands and cDNA, and unmodified reference DNA duplexes (**DNA6-DNA10**).

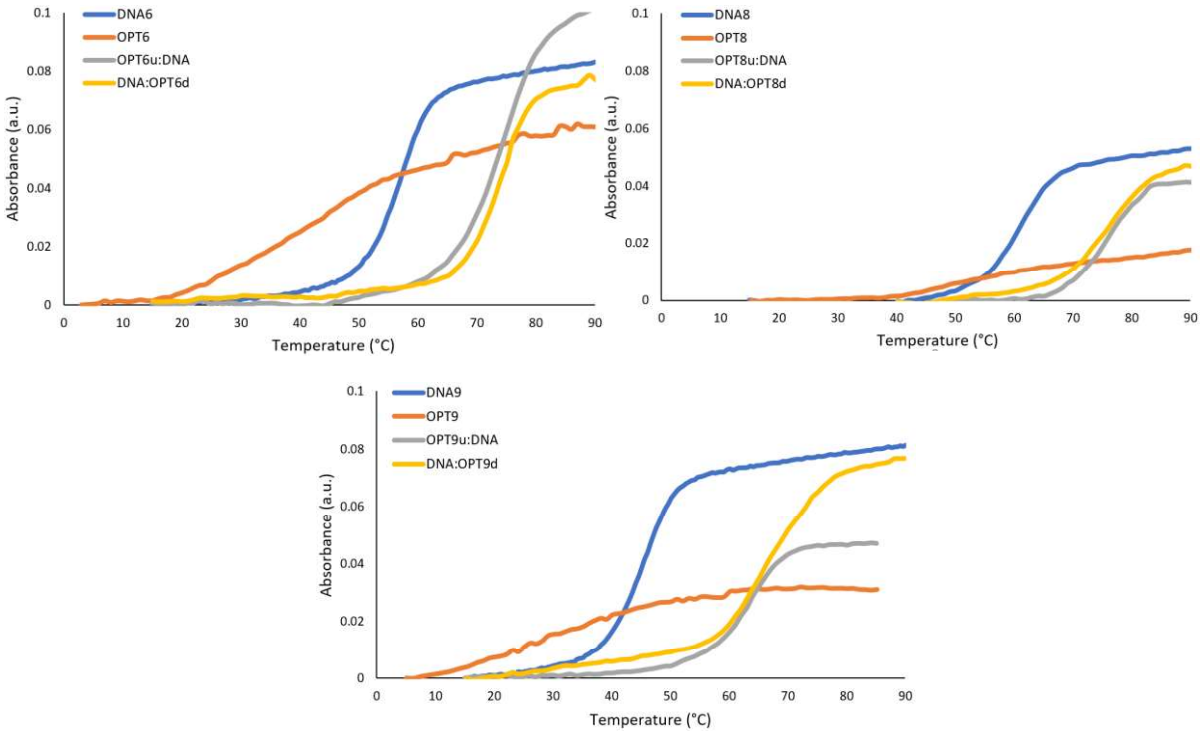


Figure 2.4-9. Representative thermal denaturation curves for Optimized Invader probes **OPT6**, **OPT8**, and **OPT9** and the corresponding duplexes between individual probe strands and cDNA, and unmodified reference DNA duplexes (**DNA6**, **DNA8**, and **DNA9**).

Table 2.4-3. Change in Gibbs free energy (ΔG^{310}) at upon formation of duplexes and change in reaction free energy upon Invader-mediated recognition of dsDNA targets ($\Delta G^{310}_{\mathrm{rec}}$).[a]

Probe	Sequence	$\Delta G^{310}[\Delta\Delta G^{310}]$ (kJ/mol)			ΔG^{310}_{rec} (kJ/mol)
		Probe duplex	5'INV :cDNA	3'INV: cDNA	
INV1	5'-Cy3-T**U**ATCAGCAC**U**G**U**GC-3' 3'- AA**U**AGTCGTGA**C**A**C**G-Cy3-5'	N/A	-72 [-8]	-45 [+19]	-
INV2	5'-Cy3-A**U**AC**U**GGTTTG**U**G**U**TC-3' 3'- TA**U**GA**C**CAAACA**C**A**A**G-Cy3-5'	-39 [+25]	-72 [-8]	-81 [-17]	-50
INV3	5'-Cy3-T**U**G**U**GCCCTGGC**A**AC-3' 3'- AA**C**A**C**GGGACCGT**U**G-Cy3-5'	N/A	-66 [-3]	-69 [-6]	-
INV4	5'-Cy3-**A**GCCC**U**GTGCC**C**TG-3' 3'- T**C**GGGA**C**ACGGG**A**C-Cy3-5'	-65 [+7]	-54 [+18]	-90 [-18]	-29
INV5	5'-Cy3-G**A**TTTCAGCCA**U**G**U**GC-3' 3'- CT**A**AAGTCGGTA**C**A**C**G-Cy3-5'	-46 [+25]	-68 [+3]	-78 [-7]	-29
INV6	5'-Cy3-C**U**G**U**GCAACTGGT**U**TG-3' 3'- GA**C**A**C**GTTGACCAA**A**C-Cy3-5'	-58 [+14]	-73 [-1]	-79 [-7]	-22
INV7	5'-Cy3-C**U**G**U**GCAA**U**ATTT**U**GT-3' 3'- GA**C**A**C**GTTA**T**AAAA**C**A-Cy3-5'	-50 [+12]	-75 [-13]	-83 [-21]	-46
INV8	5'-Cy3-TT**C**ACAGCCC**U**G**U**GC-3' 3'- AAG**U**GTCGGGA**C**A**C**G-Cy3-5'	-50 [+21]	-74 [-3]	-99 [-28]	-52
INV9	5'-Cy3-T**U**A**U**ATGCTG**U**TCTC-3' 3'- AA**U**A**U**ACGACA**A**GAG-Cy3-5'	-53 [-1]	-62 [-10]	-62 [-10]	-19
INV10	5'-Cy3-G**U**G**U**AGTG**U**A**U**ATG-3' 3'- CA**C**A**U**CACA**U**A**U**AC-Cy3-5'	-47 [+2]	-70 [-21]	-82 [-33]	-56

[a]**$\Delta\Delta G^{310}$** is measured relative to the corresponding unmodified DNA duplex (**DNA1** = -64 kJ/mol, **DNA2** = -64 kJ/mol, **DNA3** = -63 kJ/mol, **DNA4** = -72 kJ/mol, **DNA5** = -71 kJ/mol, **DNA6** = -72 kJ/mol, **DNA7** = -62 kJ/mol, **DNA8** = -71 kJ/mol, **DNA9** = -52 kJ/mol, and **DNA10** = -49 kJ/mol)

Table 2.4-4. Change in enthalpy (ΔH) upon formation of duplexes and change in reaction enthalpy upon Invader-mediated recognition of dsDNA targets ΔH_{rec}.[a]

Probe	Sequence	ΔH[$\Delta\Delta H$] (kJ/mol)			ΔH_{rec} (kJ/mol)
		Probe duplex	5'INV: cDNA	3'INV: cDNA	
INV1	5'-Cy3-T**U**ATCAGCAC**U**G**U**GC-3' 3'- AA**U**AGTCGTGA**C**A**C**G-Cy3-5'	N/A	-414 [+40]	-345 [+109]	-
INV2	5'-Cy3-A**U**AC**U**GGTTTG**U**G**U**TC-3' 3'- TA**U**GA**C**CAAACA**C**A**A**G-Cy3-5'	-159 [+372]	-458 [+73]	-476 [+55]	-244
INV3	5'-Cy3-T**U**G**U**GCCCTGGC**A**AC-3' 3'- AA**C**A**C**GGGACCGT**U**G-Cy3-5'	N/A	-358 [+8]	-393 [+-27]	-
INV4	5'-Cy3-**A**GCCC**U**GTGCC**C**TG-3' 3'- T**C**GGGA**C**ACGGG**A**C-Cy3-5'	-295 [+174]	-203 [+266]	-472 [-3]	-95
INV5	5'-Cy3-G**A**TTTCAGCCA**U**G**U**GC-3' 3'- CT**A**AAGTCGGTA**C**A**C**G-Cy3-5'	-240 [+286]	-422 [+104]	-448 [+78]	-104
INV6	5'-Cy3-C**U**G**U**GCAACTGGT**U**TG-3' 3'- GA**C**A**C**GTTGACCAA**A**C-Cy3-5'	-267 [+253]	-422 [+98]	-441 [+79]	-76
INV7	5'-Cy3-C**U**G**U**GCAA**U**ATTT**U**GT-3' 3'- GA**C**A**C**GTTA**T**AAAA**C**A-Cy3-5'	-242 [+285]	-418 [+109]	-474 [+53]	-127
INV8	5'-Cy3-TT**C**ACAGCCC**U**G**U**GC-3' 3'- AAG**U**GTCGGGA**C**A**C**G-Cy3-5'	-209 [+262]	-373 [+98]	-567 [-96]	-260
INV9	5'-Cy3-T**U**A**U**ATGCTG**U**TCTC-3' 3'- AA**U**A**U**ACGACA**A**GAG-Cy3-5'	-275 [+188]	-386 [+77]	-329 [+134]	+23
INV10	5'-Cy3-G**U**G**U**AGTG**U**A**U**ATG-3' 3'- CA**C**A**U**CACA**U**A**U**AC-Cy3-5'	-240 [+275]	-425 [+90]	-488 [+27]	-158

[a]**$\Delta\Delta H$** is measured relative to the corresponding unmodified DNA duplex (**DNA1** = -454 kJ/mol, **DNA2** = -531 kJ/mol, **DNA3** = -366 kJ/mol, **DNA4** = -469 kJ/mol, **DNA5** = -526 kJ/mol, **DNA6** = -520 kJ/mol, **DNA7** = -527 kJ/mol, **DNA8** = -471 kJ/mol, **DNA9** = -463 kJ/mol, and **DNA10** = -515 kJ/mol).

Table 2.4-5. Change in entropy ($-T\Delta S^{310}$) upon formation of duplexes and change in reaction entropy upon Invader-mediated recognition of dsDNA targets ($-T\Delta S^{310}_{\mathrm{rec}}$).[a]

Probe	Sequence	$-T\Delta S^{310}$ [$\Delta(T\Delta S^{310})$] (kJ/mol)			$-T\Delta S^{310}_{rec}$ (kJ/mol)
		Probe duplex	5'INV: cDNA	3'INV: cDNA	
INV1	5'-Cy3-T**U**ATCAGCAC**U**G**U**GC-3' 3'- AA**U**AGTCGTGA**C**A**C**G-Cy3-5'	N/A	341 [-49]	300 [-90]	-
INV2	5'-Cy3-A**U**AC**U**GGTTTG**U**G**U**TC-3' 3'- TA**U**GA**C**CAAACA**C**A**A**G-Cy3-5'	119 [-347]	386 [-80]	209 [-257]	10
INV3	5'-Cy3-T**U**G**U**GCCCTGGC**A**AC-3' 3'- AA**C**A**C**GGGACCGT**U**G-Cy3-5'	N/A	292 [-11]	324 [+21]	-
INV4	5'-Cy3-**A**GCCC**U**GTGCC**C**TG-3' 3'- T**C**GGGA**C**ACGGG**A**C-Cy3-5'	229 [-168]	148 [-249]	382 [-15]	66
INV5	5'-Cy3-G**A**TTTCAGCCA**U**G**U**GC-3' 3'- CT**A**AAGTCGGTA**C**A**C**G-Cy3-5'	194 [-261]	354 [-101]	370 [-85]	75
INV6	5'-Cy3-C**U**G**U**GCAACTGGT**U**TG-3' 3'- GA**C**A**C**GTTGACCAA**A**C-Cy3-5'	209 [-239]	348 [-100]	362 [-86]	53
INV7	5'-Cy3-C**U**G**U**GCAA**U**ATTT**U**GT-3' 3'- GA**C**A**C**GTTA**T**AAAA**C**A-Cy3-5'	192 [-273]	343 [-122]	390 [-75]	76
INV8	5'-Cy3-TT**C**ACAGCCC**U**G**U**GC-3' 3'- AAG**U**GTCGGGA**C**A**C**G-Cy3-5'	160 [-240]	299 [-101]	467 [+67]	206
INV9	5'-Cy3-T**U**A**U**ATGCTG**U**TCTC-3' 3'- AA**U**A**U**ACGACA**A**GAG-Cy3-5'	222 [-189]	324 [-87]	262 [-149]	-47
INV10	5'-Cy3-G**U**G**U**AGTG**U**A**U**ATG-3' 3'- CA**C**A**U**CACA**U**A**U**AC-Cy3-5'	193 [-272]	355 [-110]	406 [-59]	103

[a] **$\Delta(T\Delta S^{310})$** is measured relative to the corresponding unmodified DNA duplex (**DNA1** = 390 kJ/mol, **DNA2** = 466 kJ/mol, **DNA3** = 303 kJ/mol, **DNA4** = 397 kJ/mol, **DNA5** = 455 kJ/mol, **DNA6** = 448 kJ/mol, **DNA7** = 465 kJ/mol, **DNA8** = 400 kJ/mol, **DNA9** = 411 kJ/mol, and **DNA10** = 465 kJ/mol).

Table 2.4-6. Change in Gibbs free energy (ΔG^{310}) for Optimized Invader probes at upon formation of duplexes and change in reaction free energy upon Invader-mediated recognition of dsDNA targets ($\Delta G^{310}_{\mathrm{rec}}$).[a]

Probe	Sequence	$\Delta G^{310}[\Delta\Delta G^{310}]$ (kJ/mol)			ΔG^{310}_{rec} (kJ/mol)
		Probe duplex	5'INV: cDNA	3'INV: cDNA	
OPT6	5'-Cy3-C**U**G**U**G**C**AAC**U**GGT**U**TG-3' 3'- GA**C**A**C**G**U**TGA**C**CAA**A**C-Cy3-5'	-44 [+28]	-95 [-23]	-105 [-33]	-84
OPT8	5'-Cy3-TT**C**A**C**AGCCC**U**G**U**GC-3' 3'- AAG**U**G**U**CGGGA**C**A**C**G-Cy3-5'	-46 [+25]	-101 [-31]	-104 [-33]	-93
OPT9	5'-Cy3-T**U**A**U**A**U**GC**U**G**U**TCTC-3' 3'- AA**U**A**U**A**C**GA**C**A**A**GAG-Cy3-5'	-38 [+41]	-74 [-22]	-75 [-23]	-59

[a]**$\Delta\Delta G^{310}$** is measured relative to the corresponding unmodified DNA duplex (**DNA6** = -72 kJ/mol, **DNA8** = -71 kJ/mol, **DNA9** = -52 kJ/mol)

Table 2.4-7. Change in enthalpy (ΔH) for Optimized Invader probes upon formation of duplexes and change in reaction enthalpy upon Invader-mediated recognition of dsDNA targets ΔH_{rec}.[a]

Probe	Sequence	$\Delta H[\Delta\Delta H]$ (kJ/mol)			ΔH_{rec} (kJ/mol)
		Probe duplex	5'INV: cDNA	3'INV: cDNA	
OPT6	5'-Cy3-C**U**G**U**G**C**AAC**U**GGT**U**TG-3' 3'- GA**C**A**C**G**U**TGA**C**CAA**A**C-Cy3-5'	-193 [+278]	-519 [-48]	-620 [-149]	-426
OPT8	5'-Cy3-TT**C**A**C**AGCCC**U**G**U**GC-3' 3'- AAG**U**G**U**CGGGA**C**A**C**G-Cy3-5'	-191 [+280]	-555 [-84]	-562 [-91]	-455
OPT9	5'-Cy3-T**U**A**U**A**U**GC**U**G**U**TCTC-3' 3'- AA**U**A**U**A**C**GA**C**A**A**GAG-Cy3-5'	-166 [+305]	-476 [-5]	-425 [+47]	-272

[a]**$\Delta\Delta H$** is measured relative to the corresponding unmodified DNA duplex (**DNA6** = -520 kJ/mol, **DNA8** = -471 kJ/mol, **DNA9** = -463 kJ/mol).

Table 2.4-8. Change in entropy ($-T\Delta S^{310}$) for Optimized Invader probes upon formation of duplexes and change in reaction entropy upon Invader-mediated recognition of dsDNA targets ($-T\Delta S^{310}_{rec}$).[a]

Probe	Sequence	$-T\Delta S^{310}$ [$\Delta(T\Delta S^{310})$] (kJ/mol)			$-T\Delta S^{310}_{rec}$ (kJ/mol)
		Probe duplex	5'INV: cDNA	3'INV: cDNA	
OPT6	5'-Cy3-C**U**G**U**G**C**AAC**U**GGT**U**TG-3' 3'- GA**C**A**C**G**U**TGA**C**CAA**A**C-Cy3-5'	149 [-299]	424 [-24]	515 [+67]	342
OPT8	5'-Cy3-TT**C**A**C**AGCCC**U**G**U**GC-3' 3'- AAG**U**G**U**CGGGA**C**A**C**G-Cy3-5'	150 [-298]	452 [+4]	458 [+10]	360
OPT9	5'-Cy3-T**U**A**U**A**U**GC**U**G**U**TCTC-3' 3'- AA**U**A**U**A**C**GA**C**A**A**GAG-Cy3-5'	147 [-301]	402 [-46]	349 [-99]	193

[a] **$\Delta(T\Delta S^{310})$** is measured relative to the corresponding unmodified DNA duplex (**DNA6** = 448 kJ/mol, **DNA8** = 400 kJ/mol, **DNA9** = 411 kJ/mol).

Table 2.4-9. Sequence and intramolecular T_m of the DNA hairpins used herein.[a]

Hairpin	Sequence	T_m °C
DH1		76.0
DH2		72.0
DH3		81.5
DH4		82.0
DH5		76.0
DH6		75.0
DH7		68.0
DH8		80.5
DH9		67.0
DH10		62.0

[a] T_m were determined as described in Table 1.

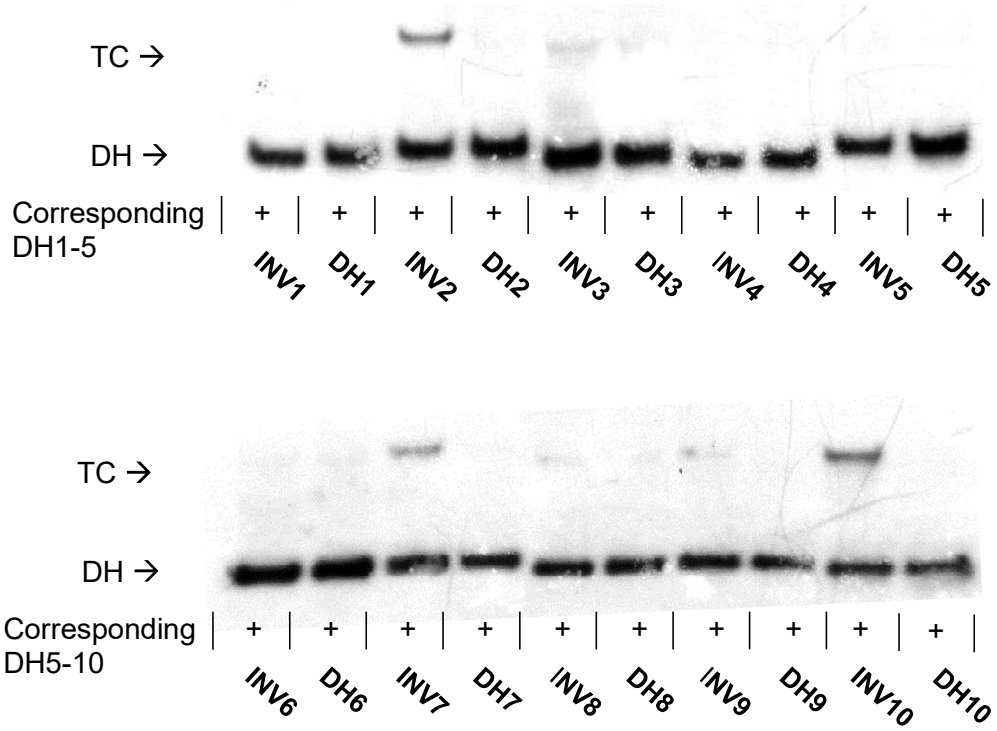


Figure 2.4-10. Invader and corresponding hairpins with the Invaders at a 5-fold molar excess to the hairpin. Shows that the hairpin in indeed the lower band. All other conditions are the same as Figure 2.4-12.

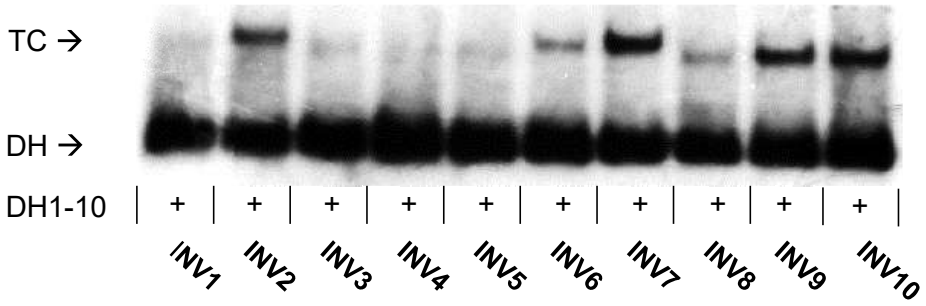


Figure 2.4-11. Screen of 100-fold of corresponding Invader probes against all 10 hairpin targets (one complementary nine non-complementary) in each lane. Incubation for 2.5 hours at 37 °C, all other conditions are as reported in Figure 2.4-12. Shows that recognition of complementary hairpin is not hindered by presence of all other hairpins.

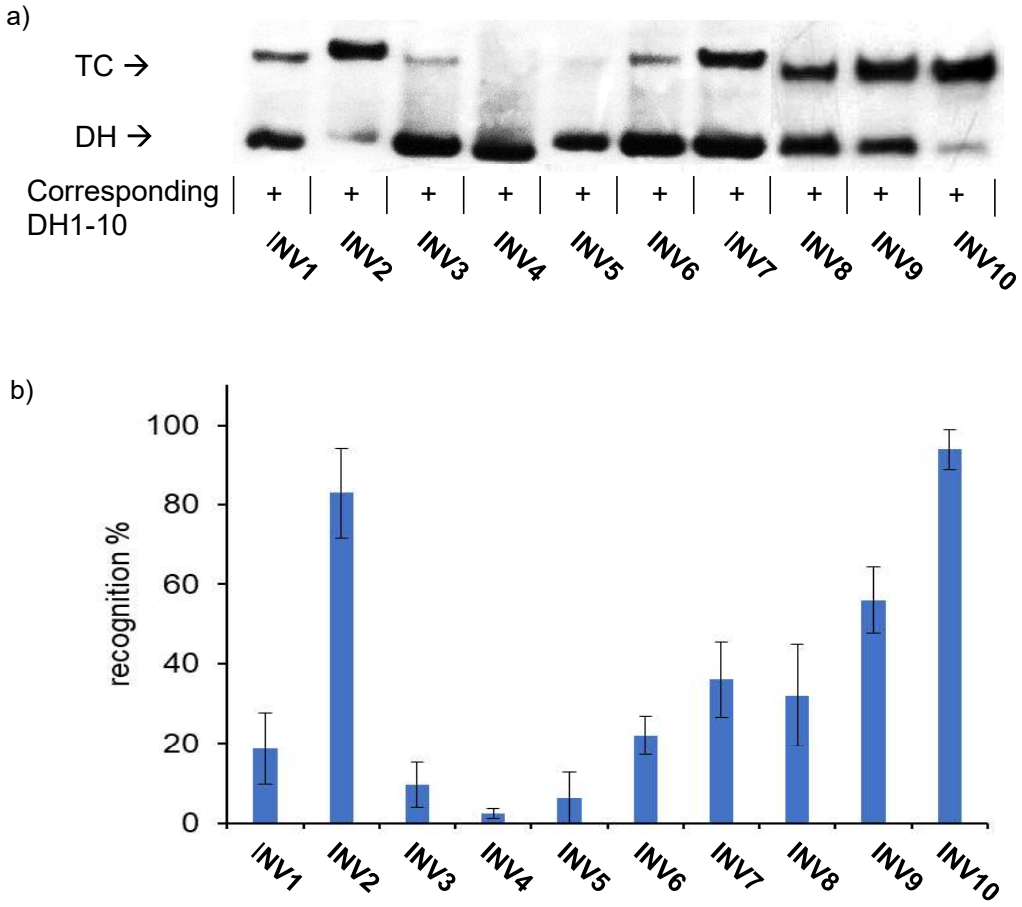


Figure 2.4-12. a) Representative gel electrophoretograms from recognition experiments at 2.5 h between a 100-fold molar excess of Invader probes **INV1**-**INV10** and their corresponding DNA hairpin targets **DH1**-**DH10**. b) Histograms depict averaged results from at least three recognition experiments with error bars representing standard deviation. TC = ternary complex. DIG-labeled **DH1**-**DH10** (34.4 nM, sequences shown in Table 2.4-9) were incubated with the corresponding Invader probe in HEPES buffer (50 mM HEPES, 100 mM NaCl, 5 mM $MgCl_2$, pH 7.2, 10% sucrose, 1.44 mM spermine tetrahyrdochloride) for 2.5 h at 37 °C. Incubation mixtures were resolved on 12% non-denaturing TBE-PAGE slabs (~70 V, ~4 °C, ~1.5 h).

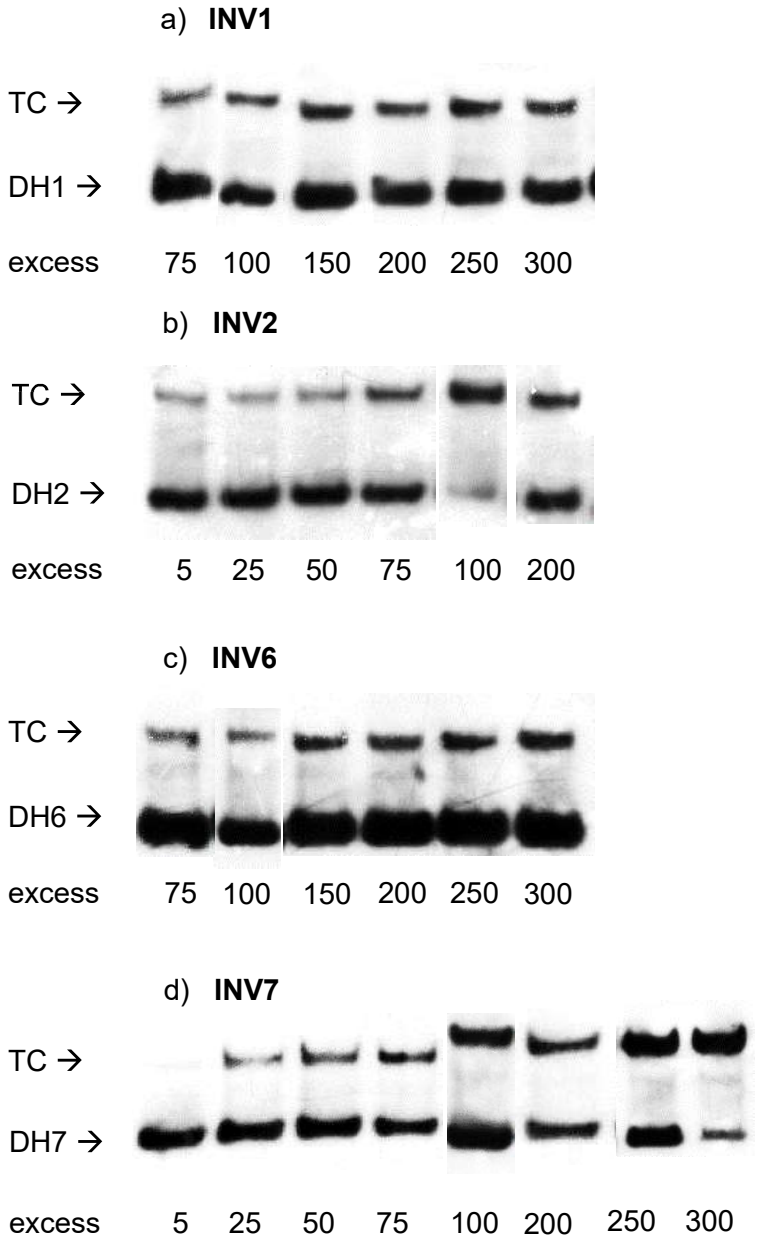


Figure 2.4-13. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) and different concentrations of Invader probes a) **INV1**, b) **INV2**, c) **INV6**, and d) **INV7** at 37 °C for 2.5 h. Experimental conditions are as specified in Figure 2.4-12. For dose-response curves, see Figure 2.4-15.

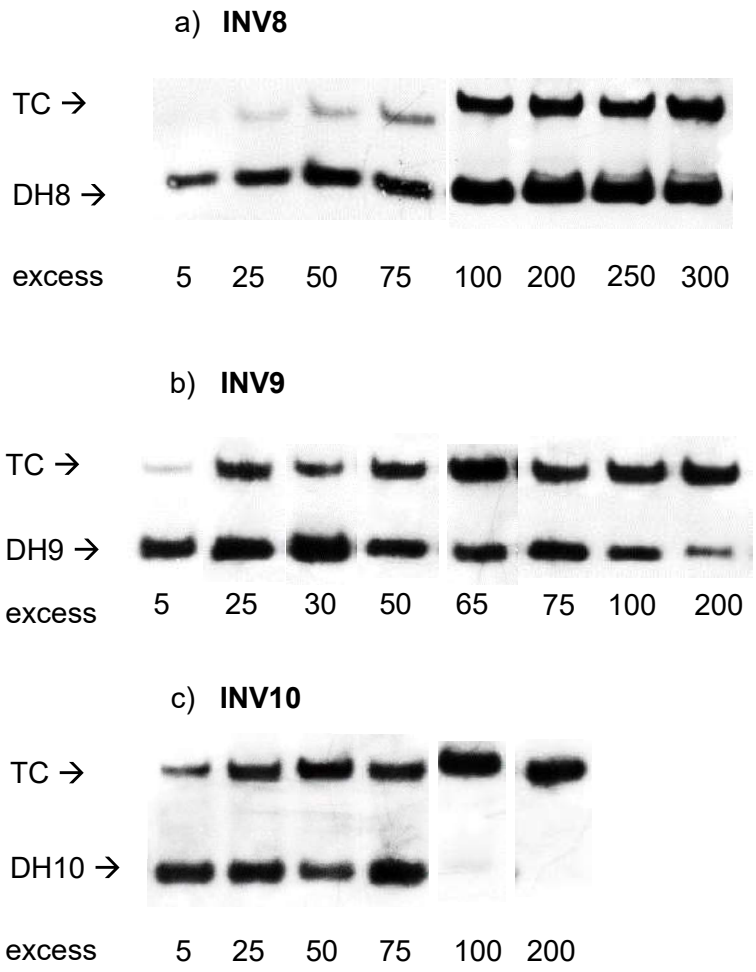


Figure 2.4-14. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) and different concentrations of Invader probes a) **INV8**, b) **INV9**, and c) **INV10** at 37 °C for 2.5 h.. Experimental conditions are as specified in Figure 2.4-12. For dose-response curves, see Figure 2.4-15.

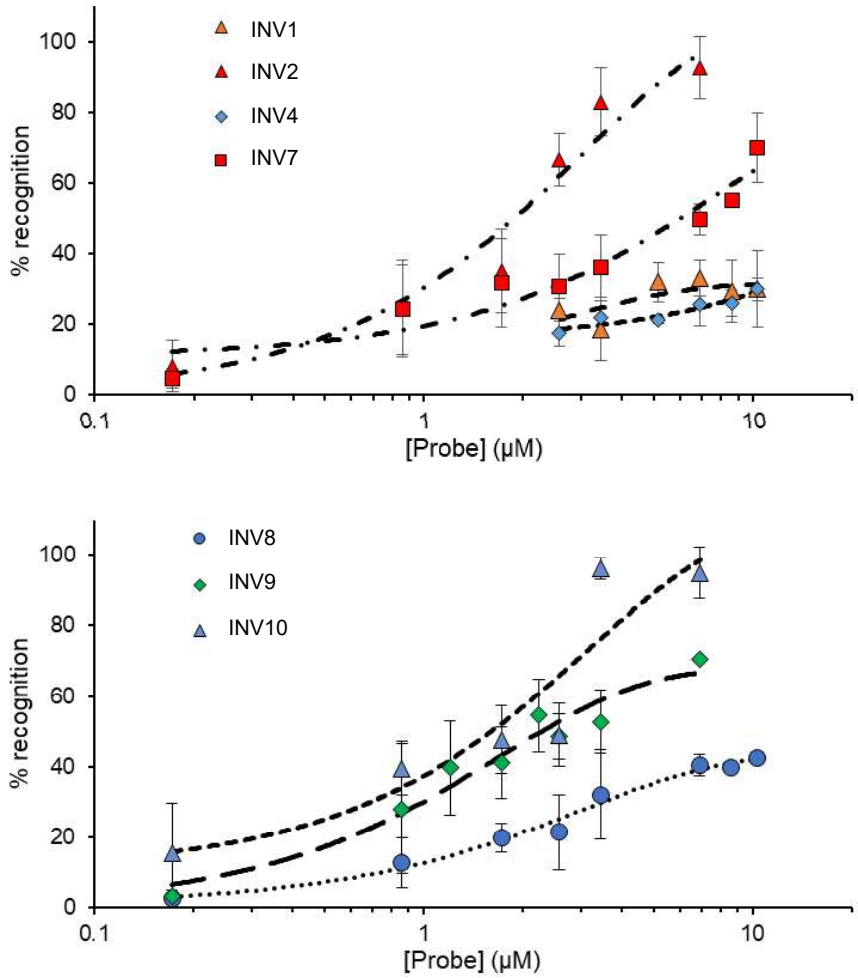


Figure 2.4-15. Dose-response curves for **INV1**, **INV2**, **INV6**, and **INV7** (upper panel), and **INV8**, **INV9**, and **INV10** (lower panel) at 37 °C for 2.5 h. Experimental conditions are as described in Figure 2.4-12, except for variable probe concentrations.

Table 2.4-10. C_{50} values at 2.5 h for Invader probes studied herein.[a]

Probes	C_{50} (μM)	Rec_{100X} (%)
INV1	>10	18 ± 8.9
INV2	1.9	82 ± 9.5
INV3	ND	9 ± 5.6
INV4	ND	2 ± 1.2
INV5	ND	6 ± 6.6
INV6	>10	22 ± 4.8
INV7	6.0	36 ± 9.3
INV8	>10	32 ± 12.7
INV9	2.2	56 ± 8.3
INV10	1.6	93 ± 5.0

[a] Rec_{100} = level of DNA hairpin recognition using 100-fold molar probe excess. C_{50} values were determined from dose-response curves shown in Figure 2.4-15.

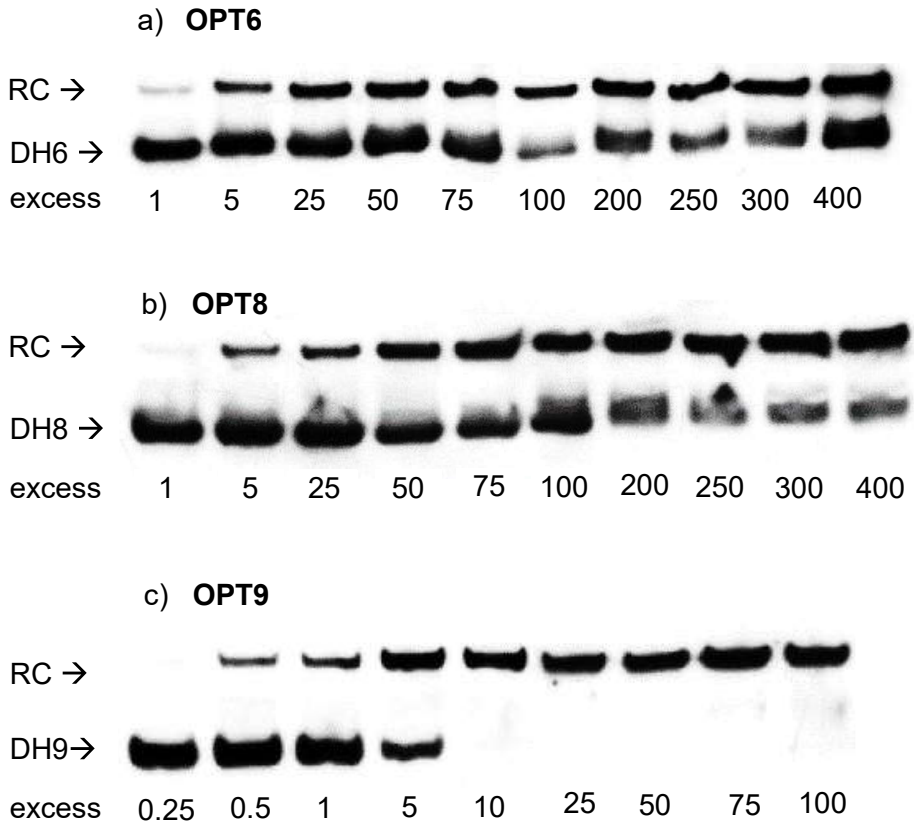


Figure 2.4-16. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) and different concentrations of Optimized Invader probes a) **OPT6**, b) **OPT8**, c) **OPT9** at 37 °C for 2.5 h. Experimental conditions are as specified in Figure 2.2-8. For dose-response curves, see Figure 2.4-17.

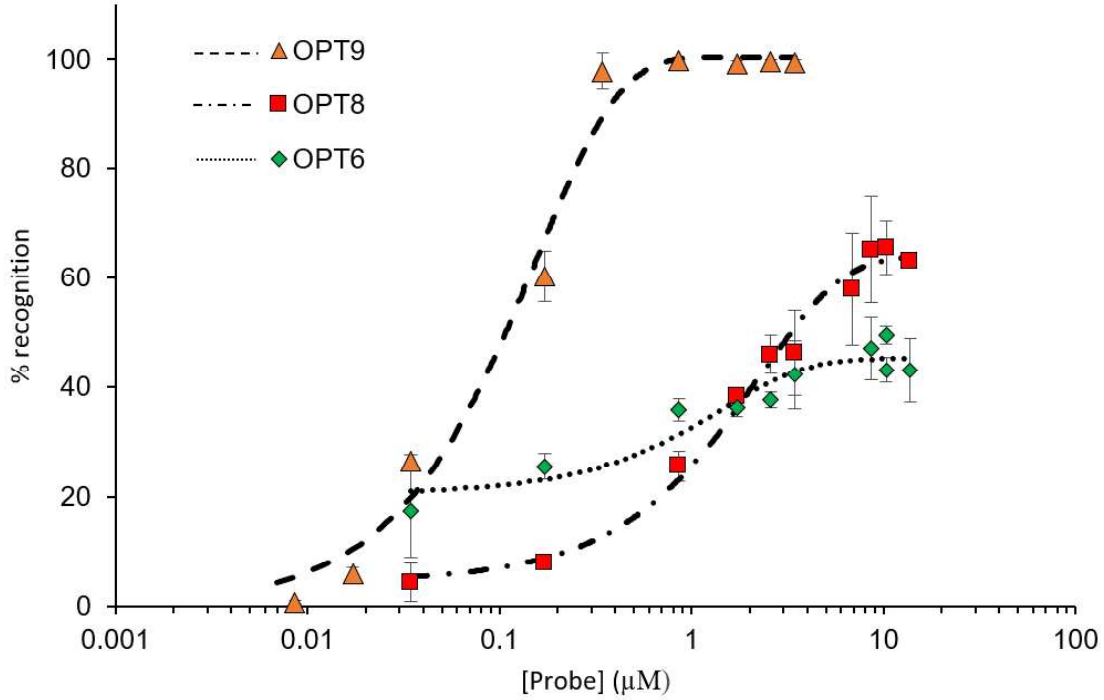


Figure 2.4-17. Dose-response curves for **OPT6, OPT8, and OPT9** at 2.5 h at 37 C. Experimental conditions are as described in Figure 2.2-8, except for variable probe concentrations.

Table 2.4-11. C_{50} values at 2.5 h for Optimized Invader probes studied herein.[a]

Probes	C_{50} (µM)	Rec_{100X} (%)
OPT6	1.3	66 ± 3.4
OPT8	1.5	66 ± 0.4
OPT9	0.2	99 ± 8.9

[a] Rec_{100} = level of DNA hairpin recognition using 100-fold molar probe excess. C_{50} values were determined from dose-response curves shown in Figure 2.4-17.

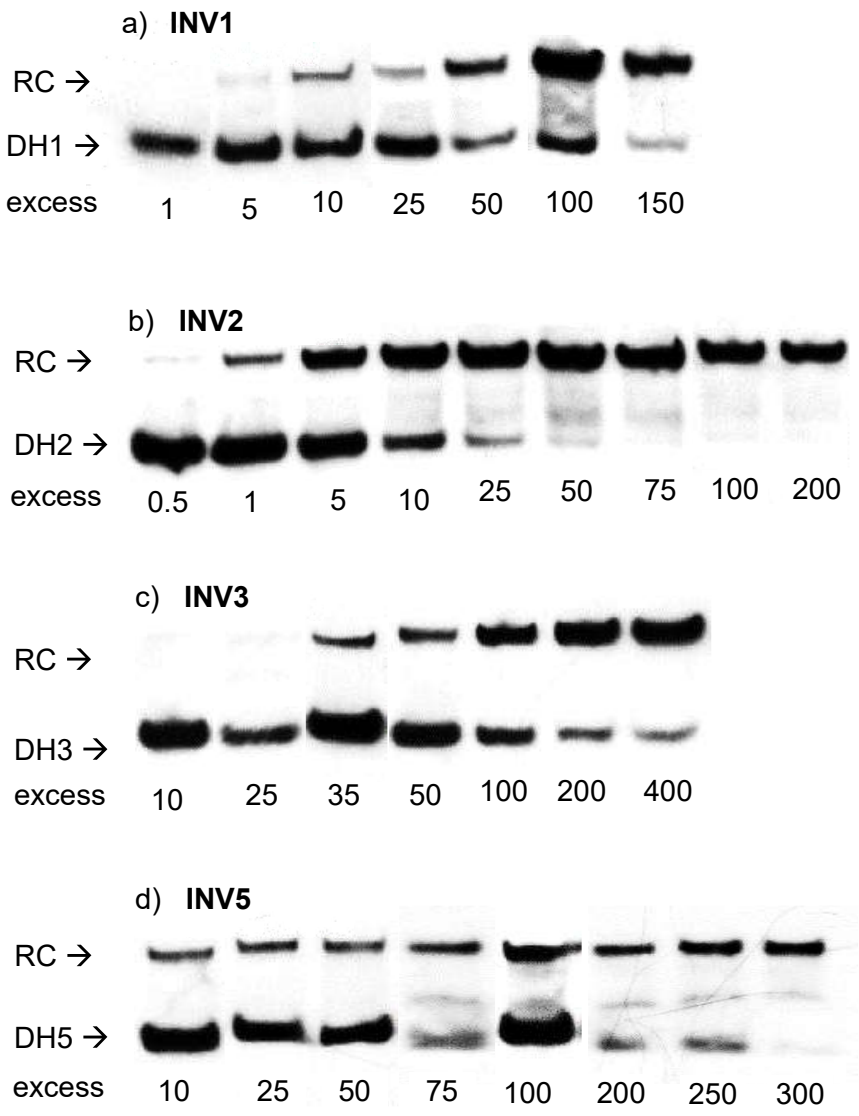


Figure 2.4-18. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) and different concentrations of Invader probes a) **INV1**, b) **INV2**, c) **INV3**, and d) **INV5** at 37 °C for 15 h. Experimental conditions are as specified in Figure 2.2-2. For dose-response curves, see Figure 2.2-3.

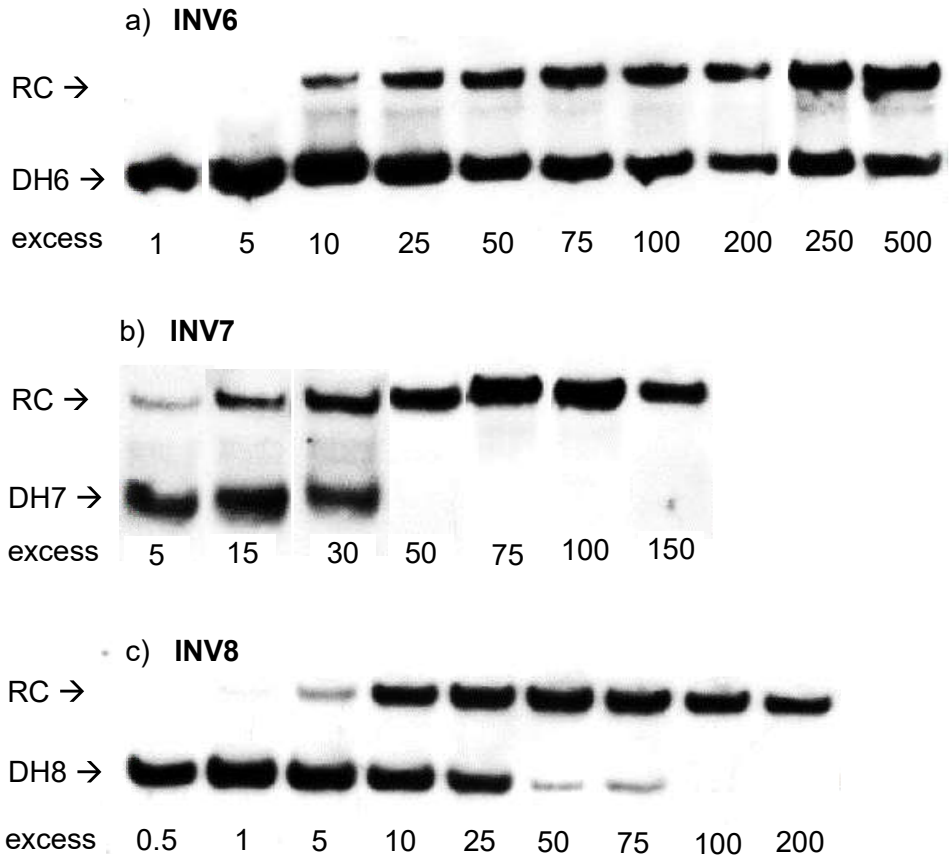


Figure 2.4-19. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) and different concentrations of Invader probes a) **INV6**, b) **INV7**, and c) **INV8** at 37 °C for 15 h. Experimental conditions are as specified in Figure 2.2-2. For dose-response curves, see Figure 2.2-3.

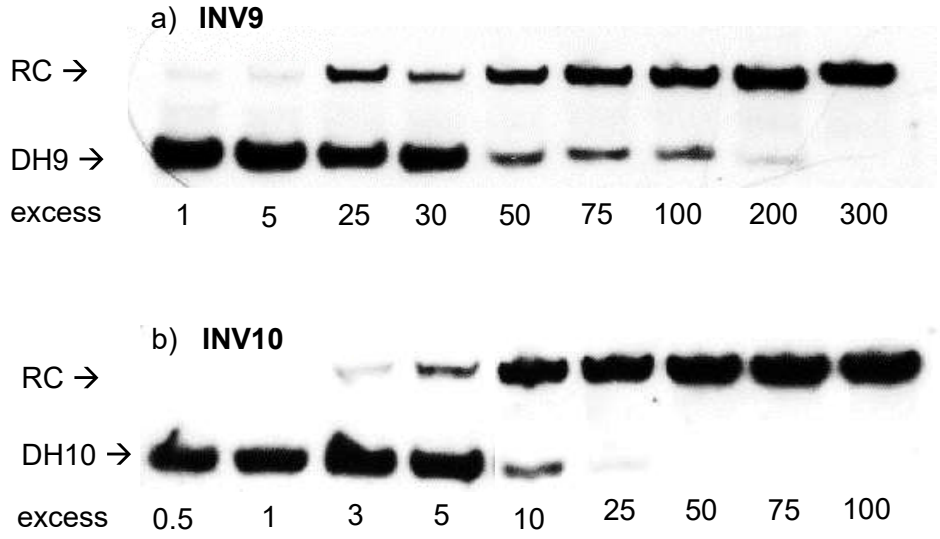


Figure 2.4-20. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) and different concentrations of Invader probes a) **INV9** and b) **INV10** at 37 °C for 15 h. Experimental conditions are as specified in Figure 2.2-2. For dose-response curves, see Figure 2.2-3.

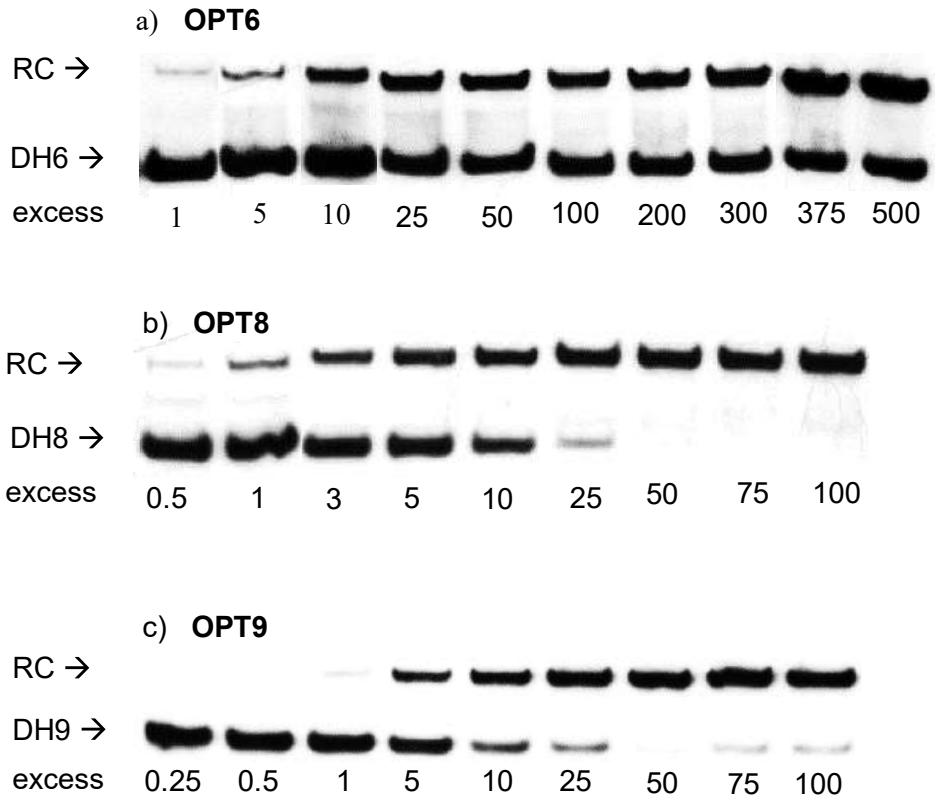


Figure 2.4-21. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) and different concentrations of Optimized Invader probes a) **OPT6**, b) **OPT8**, c) **OPT9** at 37 °C for 15 h. Experimental conditions are as specified in Figure 2.2-8. For dose-response curves, see Figure 2.2-9.

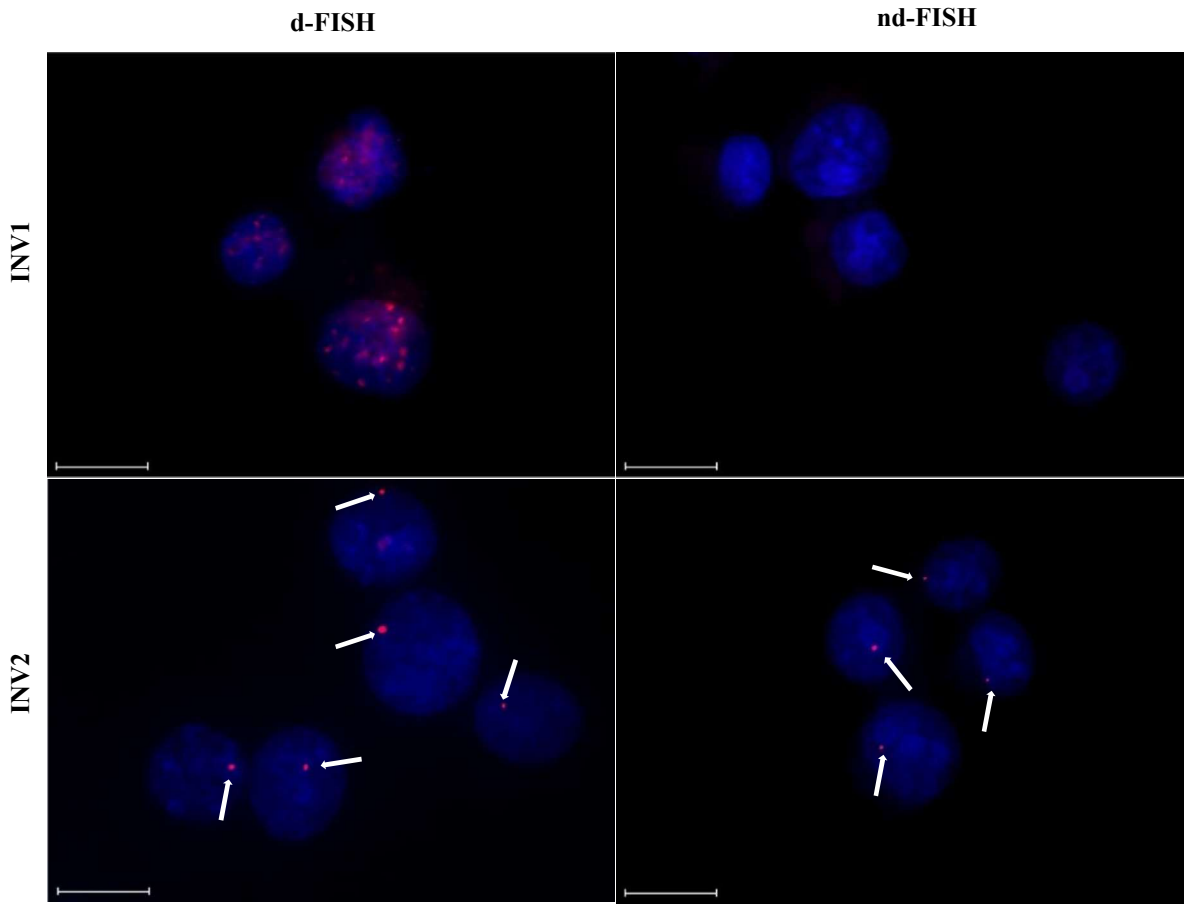


Figure 2.4-22. Representative images of **INV1** and **INV2** from FISH experiments using *DYZ1*-targeting Invader probes under denaturing (80 °C, 5 min) (left), or non-denaturing (3 h, 37.5 °C) conditions (right). Incubation and imaging specifications are described in Figure 2.2-5. Scale bar represents 16 μm.

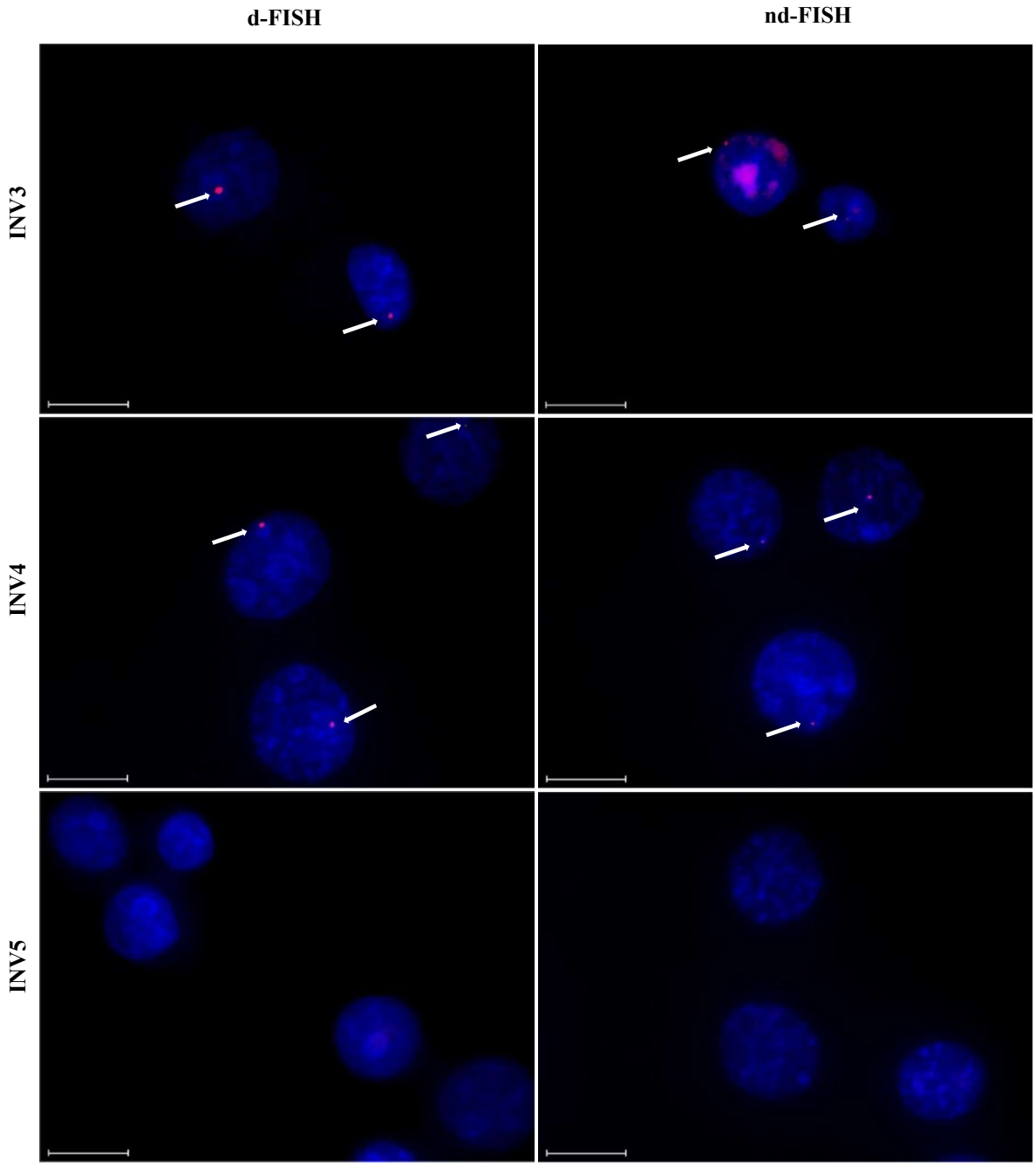


Figure 2.4-23. Representative images of **INV3**, **INV4** and **INV5** from FISH experiments using *DYZ1*-targeting Invader probes. Incubation and imaging specifications are described in Figure 2.2-5. Scale bar represents 16 μm.

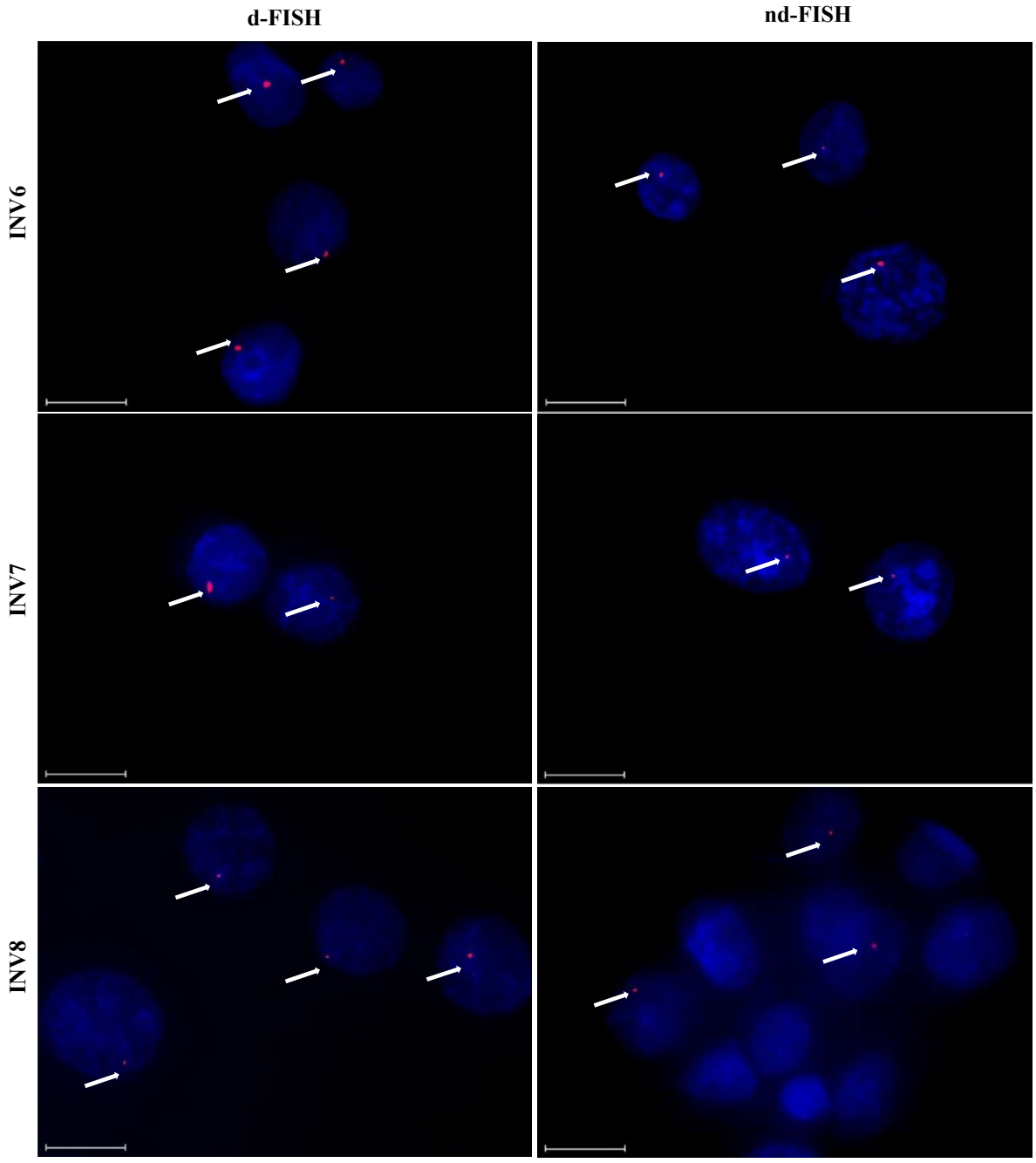


Figure 2.4-24. Representative images of **INV6**, **INV7** and **INV8** from FISH experiments using *DYZ1*-targeting Invader probes. Incubation and imaging specifications are described in Figure 2.2-5. Scale bar represents 16 μm.

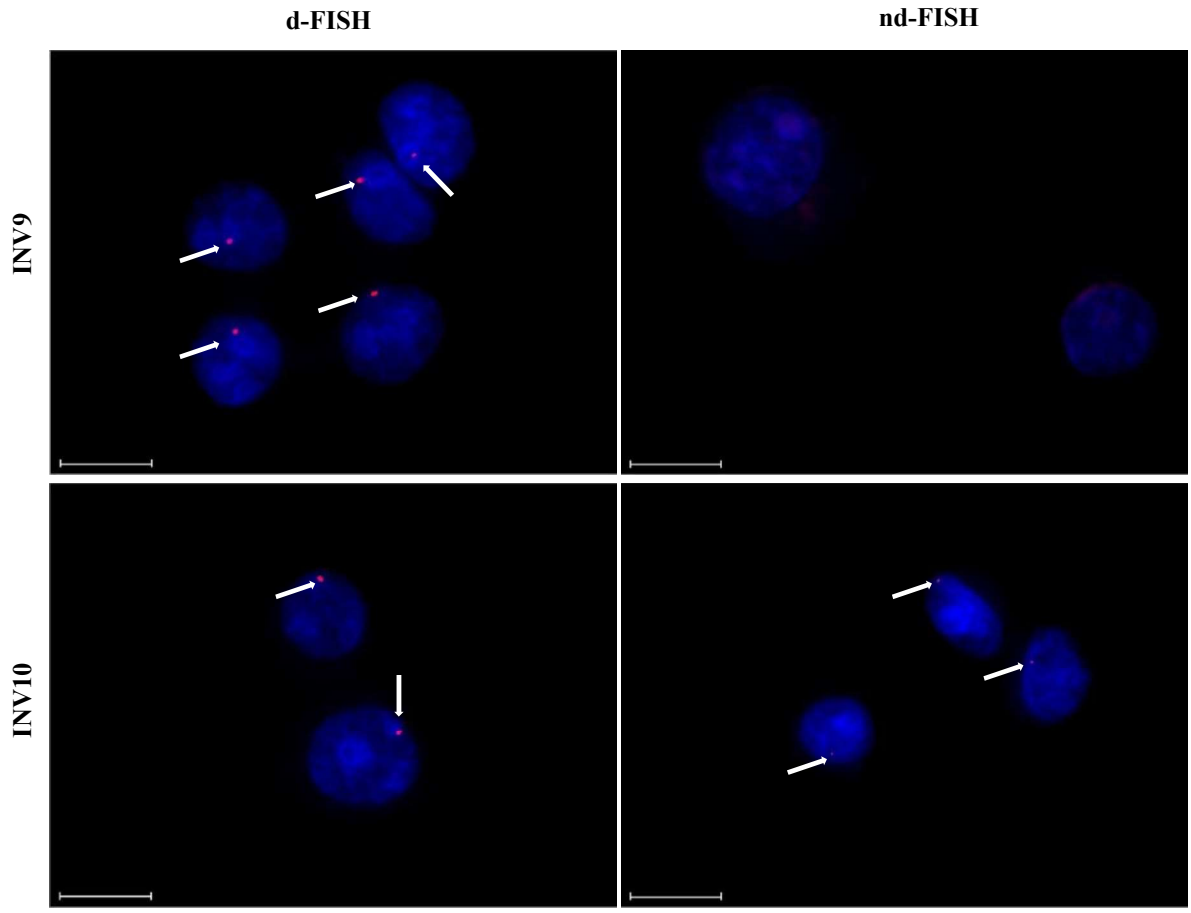


Figure 2.4-25. Representative images of **INV9** and **INV10** from FISH experiments using *DYZ1*-targeting Invader probes. Incubation and imaging specifications are described in Figure 2.2-5. Scale bar represents 16 μm.

Table 2.4-12. Quality of signal intensity (I) and background (B) in representative images from d-FISH experiments in which Invader probes were used at 1x concentration (30 ng of probe per 200 µL of PCR buffer), and nd-FISH experiments in which Invader probes were used at 1x-4x concentration.

	INV1		INV2		INV3		INV4		INV5		INV6		INV7		INV8		INV9		INV10	
Conc.	**I**	**B**	**I**	**B**	**I**	**B**	**I**	**B**	**I**	**B**	**I**	**B**	**I**	**B**	**I**	**B**	**I**	**B**	**I**	**B**
dFISH	**nd**	**3**	3	3	3	3	1	1	**1**	**3**	3	3	3	3	3	3	**3**	**3**	3	3
[1x]	nd	3	**3**	**3**	**2**	**3**	**1**	**1**	1	3	**2**	**3**	**2**	**3**	**2**	**3**	1	3	**3**	**3**
[2x]	nd	3	2	2	2	3	nd	nd	nd	nd	nd	nd	1	3	1	3	1	3	1	3
[3x]	nd	1	1	1	1	2	nd	nd	nd	nd	nd	nd	1	2	1	2	1	2	1	2
[4x]	nd	1	1	1	2	1	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd

[a] Signals of low, medium, or high intensity relative to background were scored 1, 2 or 3, respectively, while high, medium, or low backgrounds were scored 1, 2, or 3, respectively. **INV7-INV10** were not evaluated above 3x probe concentration due to poor signal-to-background characteristics at 3x probe concentration. **INV5** and **INV6** displayed neither high coverage or signal-to-background characteristics and were not evaluated at concentrations above 1x. **INV4** was not evaluated above 1x due to immense background binding and multiple potential signals. All other FISH experiments were conducted at 1/4x for **INV4** to yield one discernable signal. Nd = no discernable signal.

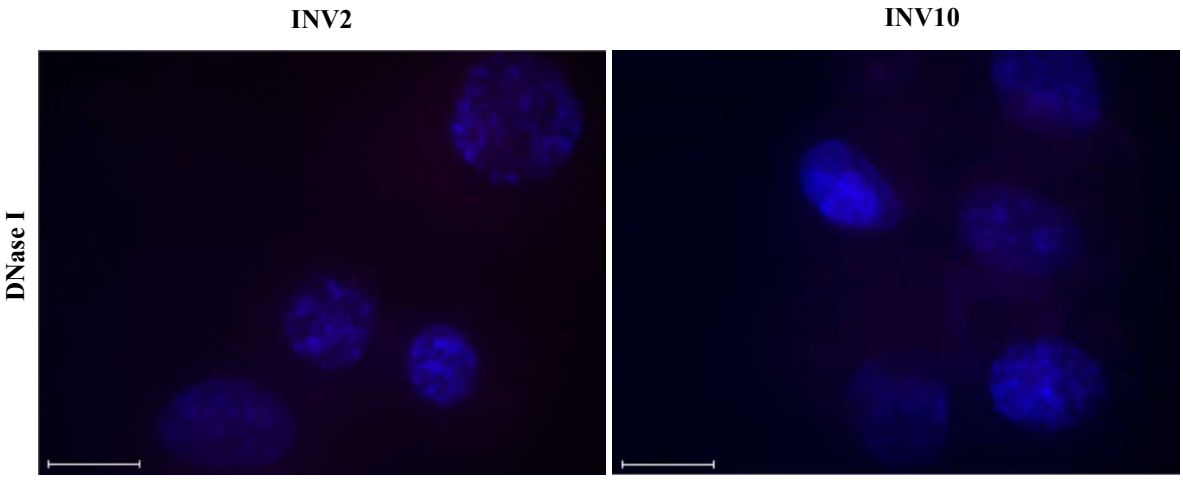


Figure 2.4-26. Representative Images for FISH experiments of nuclei pre-treated with DNase I prior to incubation with **INV2** and **INV10**. Note the absence of signal in the DNase I treatment indicating the Invader probes' targeting of the chromosomal DNA. Incubation and imaging specifications are described in Figure 2.2-5. Scale bar represents 16 µm.

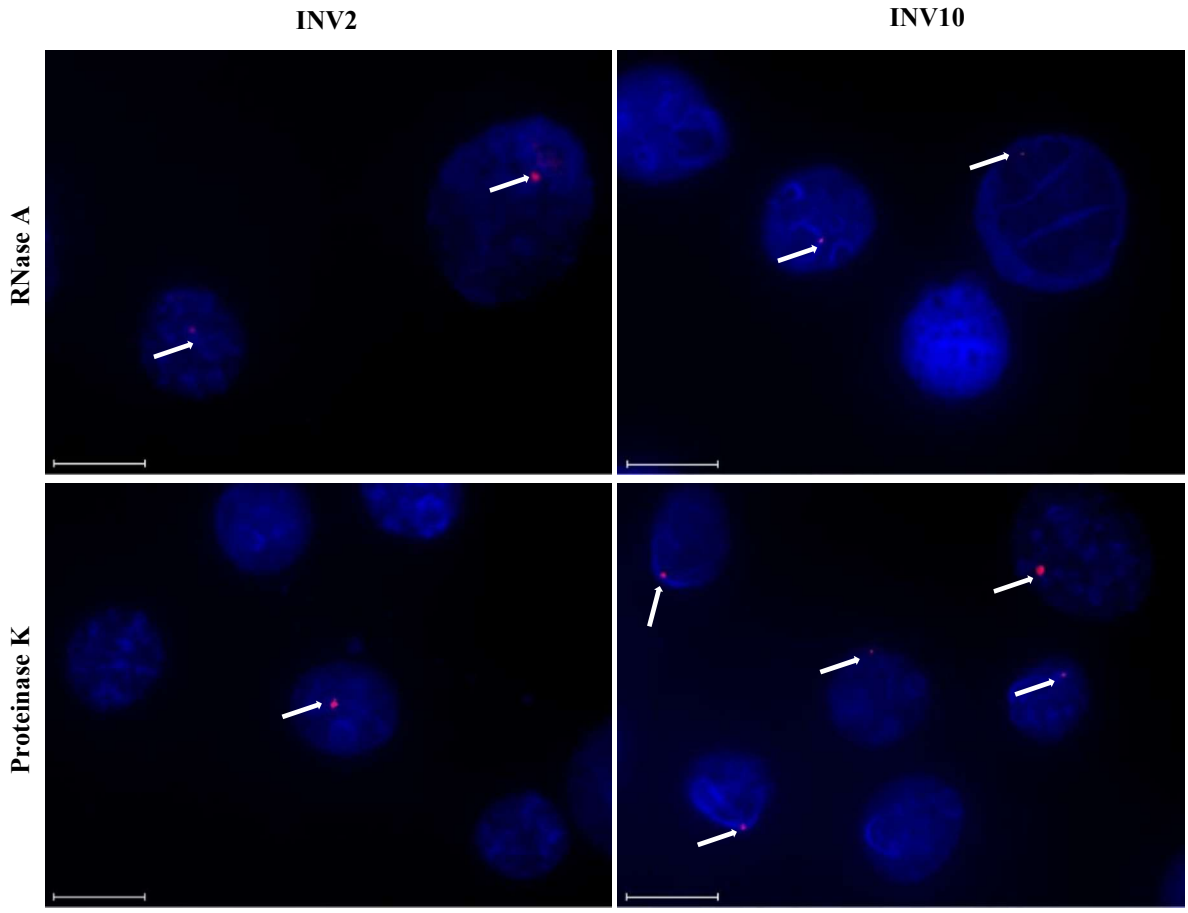


Figure 2.4-27. Representative Images for FISH experiments of nuclei pre-treated with RNase A and Proteinase K prior to incubation with **INV2** and **INV10**. Note the continued presence of signal in the RNase A and Proteinase K pre-treated nuclei (albeit there is a reduction in signal coverage which could be attributed to the loss of genetic material/number of nuclei during these enzyme treatments, not seen in DNase treatment). Incubation and imaging specifications are described in Figure 2.2-5 Scale bar represents 16 μm.

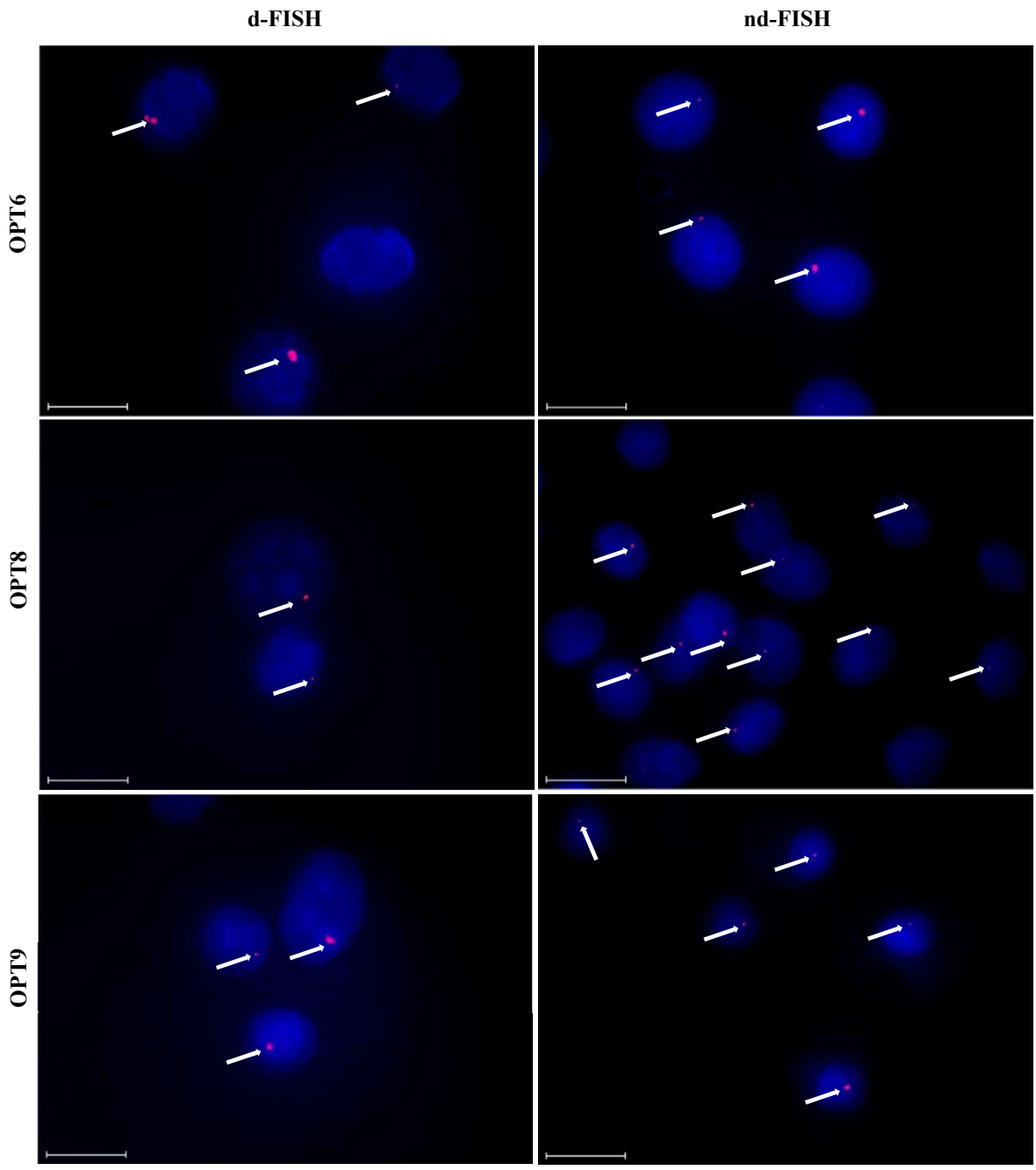


Figure 2.4-28. Images from FISH experiments using DYZ1-targeting **OPT6**, **OPT8**, and **OPT9** under denaturing (80 °C, 5 min) (left), or non-denaturing (3 h, 38.5 °C) conditions (right). Incubation and imaging specifications are described in Figure 2.2-5. Scale bar represents 16 µm.

2.5 References

1. R. Besch, C. Giovannangeli and K. Degitz, *Drug Targets*, 2004, **5**, 691-703.
2. F. A. Rogers, J. A. Lloyd and P. M. Glazer, *Curr. Med. Chem.: Anti-Cancer Agents*, 2005, **5**, 319-326.
3. I. Ghosh, C. I. Stains, A. T. Ooi and D. J. Segal, *Mol. Biosyst*., 2006, **2**, 551-560.
4. P. E. Nielson, *Biodiv*., 2010, **7**, 786.
5. A. Mukherjee and K. M. Vasquez, *Biochemie*., 2011, **93**, 1197-1208.
6. T. Vaijayanthi, T. Bando, G. N. Pandian and H. Sugiyama, *ChemBioChem*., 2012, **13**, 2170-2185.
7. B. Chen, L. A. Gilbert, B. A. Cimini, J. Schnitzbauer, W. Zhang, G. W. Li, J. Park, E. H. Blackburn, J. S. Wessman, L. S. Qi and B. Huang, *Cell*, 2013, ***155***, 1479-1491.
8. M. Duca, P. Vekhoff, K. Oussedik, L. Halby and P. B. Arimondo*, Nucleic Acids Res*., 2008, **36**, 5123-5138.
9. P. Vekhoff, A. Ceccaldi, D. Polverari, J. Pylouster, C. Pisano and P. B. Arimondo, *Biochemistry*, 2008, **47**, 12277-12289.
10. P. B. Dervan and B. S. Edelson, *Curr. Opin. Struct. Biol*., 2003, **13**, 284-299.
11. M. S. Blackledge and C. Melander, *Bioorg. Med. Chem*., 2013, **21**, 6101-6114.
12. Y. Kawamoto, T. Bando and H. Sugiyama, *Bioorg. Med. Chem*, 2018, **26**, 1393-1411.
13. S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird and P. B. Dervan, *Nature*, 1998, **391**, 468-471.
14. P. E. Nielsen, M. Egholm and O. Buchardt, *Science*, 1991, **254**, 1497-1500.
15. K. Kaihatsu, B. A. Janowski and D. R. Corey, *Chem. Biol*., 2004, **11**, 748-758.
16. T. Bentin, H. J. Larsen and P. E. Nielson, *Biochemistry*, 2003, **42**, 13987-13995.

17. R. Bahal, B. Sahu, S. Rapireddy, C. M. Lee and D. H. Ly, *ChemBioChem*, 2012, **13**, 56-60.

18. S. Rapireddy, R. Bahal and D. H. Ly, *Biochemistry*, 2011, **50**, 3913-3918.

19. B. Chen, L. A. Gilbert, B. A. Cimini, J. Schnitzbauer, W. Zhang, G. W. Li, J. Park, E. H. Blackburn, J. S. Wessman, L. S. Qi and B. Huang, *Cell*, 2013, ***155***, 1479-1491.

20. W. Deng, X. Shi, R. Tjian, T. Lionnet and R. H. Singer, *Proc. Natl. Acad. Sci.*, 2015, **112**, 11870-11875.

21. H. Ma, A. Naseri, P. Reyes-Gutierrez, S. A. Wolfe, S. Zhang and T. Pederson, *Proc. Natl. Acad. Sci.*, 2015, **112**, 3002-3007.

22. P. A. Sau, A. S. Madsen, P. Podbevsek, N. K. Anderson, T. S. Kumar, S. Anderson, R. L. Rathju, B. A. Anderson, D. C. Guenther, S. Karmakar, P. Kumar, J. Plavec, J. Wendel and P. J. Hrdlicka, *J. Org. Chem*., 2013, **78**, 9560-9570.

23. B. Denn, S. Karmakar, D. C. Guenther and P. J. Hrdlicka, *Chem. Commun*., 2013, **49**, 9851-9853.

24. B. A. Didion, S. Karmakar, D. C. Guenther, S. P. Sau, J. P. Verstegen and P. J Hrdlicka, *ChemBioChem*, 2103, **14**, 1534-1538.

25. D. C. Guenther, G. H. Anderson, S. Karmakar, B. A. Anderson, B. A. Didion, W. Guo, J. P. Verstegen and P. J. Hrdlicka, *Chem. Sci*., 2015, **6**, 5006-5015.

26. J. Perret, Y. Shia, R. Fries, G. Vassart and M. Georges, *Genomics,* 1990, **6**, 482-490.

27. R. G. Emehiser, E. Hall, D. C. Guenther, S. Karmakar and P. J. Hrdlicka, *Org. Biol. Chem.*, 2020, **18**, 56-65.

28. S. Karmakar, D. C. Guenther and P. J. Hrdlicka, *J. J. Org. Chem*., 2013, **78**, 12040-12048.

29. S. Karmakar, A. S. Madsen, D. C. Guenther, B. C. Gibbons and P. J. Hrdlicka*, Org. Biomol. Chem*., 2014, **12**, 7758-7773.

30. D. M. Crothers, *Biopolymers*, 1968, **6**, 575-584.

31. H. Ihmels and D. Otto, *Top*. *Curr*. *Chem*., 2005, **258**, 161-204.

32. J. L. Mergny and L. Lacroix, *Oligonucleotides*, 2003, **13**, 515-537.

33. M. Radman-Livaja and O. J. Rando, *Dev. Biol.*, 2010, **339**, 258-266.

34. P. Stadlbauer, P. Kührová, L. Wicherek, P. Banáš, M, Otyepla, L. Trantírek and J. Šponer, *Nucleic Acids Res*., 2019, **47**, 7276-7293.

35. J. Hei and T. Ha, *Proc. Natl. Acad. Sci.*, 2013, **110**, 17173-17174.

36. Karmakar, B. A. Anderson, R. L. Rathje, S. Andersen, T. Jensen, P. Nielsen and P. J. Hrdlicka, *J. Org. Chem*., 2011, **76**, 7119-7131.

37. N. N. Dioubankova, A. D. Malakhov, D. A. Stetsenko,M. J. Gait, P. E. Volynsky, R. G. Efremov and V. A. Korshun, *ChemBioChem*, 2003, **4**, 841–847.

38. M. A. Morgan, K. Okamoto, J. D. Kahn and D. S. English*, Biophys. J*., 2005, **89**, 2588–2598.

Appendix A. Analysis of Select Invader Probes by Flow Cytometry

Caroline P. Shepard, Saswata Karmakar, Patrick J. Hrdlicka

A.1 Introduction

Fluorescent *in situ* hybridization (FISH) assays are powerful tools for the detection of nucleic acids. By fluorescently labeling oligonucleotide probes, RNA and DNA can be visualized and quantified. Our laboratory has utilized a DNA FISH assay with Cy3-labeled Invader probes to detect the Y chromosome of isolated nuclei of bovine cells at near physiological conditions.[1] This is achieved by incubating Invader probes on microscope slides with fixed nuclei and evaluating coverage (i.e. the number of nuclei displaying one clear signal) by visualization through fluorescent microscopes. This FISH assay offers compelling evidence for future diagnostic and detection application; however, this microscopy-based quantification is limited. Signals are slowly counted one-by-one and quantification only reflects a small sample size (i.e. hundreds of nuclei). Additionally, there are biological characteristics, such as cell cycle stages and shortening of telomeres, that cannot be seen through a microscope. The use of flow cytometry overcomes the limitations of microscopy as it allows for high-throughput, multiparametric analysis. Current studies have utilized the flow cytometry to analyze FISH probe detection of mRNA and chromosomal DNA in cells and isolate nuclei.[2-5] Termed Flow-FISH, this assay utilizes the automation and detection power of the flow cytometer to quantify fluorescence and shape characteristics of many isolated nuclei within a suspension. Utilization of this instrument for Invader probe detection will allow for: i) analysis of thousands of nuclei to determine Invader probe binding to chromosomal DNA, ii) evaluation of characteristics such as cell cycle on Invader probe targeting of chromosomal DNA, and iii) determination of invasion properties with Invader

probes and target nuclei within a suspension, offering a first step to incubation with whole cells. To this end, we have developed an Invader Flow-FISH assay to incubate our Invader probes within a nuclei suspension that can be analyzed with flow cytometry. This study is a proof-of-concept as results show: i) our protocol gives isolated nuclei that can be analyzed for cell cycle stages, and ii) Invader probes specifically targeting their Y chromosome target. This sets the stage for future work to draw conclusions on the effects of cell behavior on Invader probe targeting and potential cell uptake studies.

A.2 Results and Discussion

Experimental design

To make the chromosomal DNA accessible, nuclei were isolated from bovine cells and subsequently, the nuclear membrane was permeabilized. Removal of cellular debris and deconstruction of the nuclear membrane ensured that our Invader probes were not encumbered by cellular structure barriers. This eliminated concerns regarding uptake of probes into cells and the nucleus, allowing the focus to be on binding to chromosomal DNA. Invader probes **INV10** and **INV4** (Table A.2-1) designed to target corresponding complementary sequences in the DYZ-1 satellite gene (~6 x 104 tandem repeats of the ~1175 bp region) on the bovine (Bos taurus) Y chromosome (NCBI code: M2606),[1,6] were incubated with nuclei in 30-fold excess of their target sequence for 3 h at 37.5 °C in PCR buffer (20 mM Tris, 100 mM KCl, pH 8.0), which are the same conditions used in our previously published non-denaturing FISH assays. These are non-denaturing conditions that are physiologically relevant (mimicking cellular conditions), allowing results from this experiment will be able to be extrapolated to cell studies. After the incubation period, the labeled nuclei were DAPI stained and analyzed by flow cytometry.

Table A.2-1. Sequences of Invader probes used for Flow-FISH Analysis

Probe	Sequence
INV4	5′-Cy3-**A**GCCC**U**GTGCC**C**TG-3' 3′- T**C**GGGA**C**ACGGG**A**C-Cy3-5'
INV10	5'-Cy3-G**U**G**U**AGTG**U**A**U**ATG-3' 3'- CA**C**A**U**CACA**U**A**U**AC-Cy3-5'

Gating strategies

A nuclei suspension after incubation with **INV10** was analyzed by flow cytometry to determine if the isolate nuclei and Cy3 signal could be detected. The DAPI stain allowed for gating of the nuclei population from other debris present in solution. Debris is shown as black dots and grey peaks in Figure A.2-1. Using the acquisition software, a gate (i.e. orange square) isolated the nuclei population as determined by high DAPI fluorescence (PB450-A). Within this orange square, cell cycle stages (G1, S, G2) of nuclei could be determined. This is achieved due to the increase of genetic material as DNA is replicated in interphase, causing an increase in the DAPI fluorescence. By comparing these intensities to changes in nuclei shape (quantified by the FSC-H channel), cell cycle stages of each nuclei could be evaluated. Between these two measurements, G1 (pink), S-phase (green), and G2 (blue) could be determined. Overall, this first set of gates confirms that the nuclei have retained their integrity in the isolation process through the incubation of Invader probes.

With this gated population isolated, the Cy3 fluorescent intensities of each nuclei could be quantified. This is depicted in Figure A.2-1 as number of nuclei (i.e. count) displaying a particular Cy3 fluorescent intensity (PE-H). Confident that our gating strategy has produced results depicting Invader probe binding to the chromosomal DNA of isolated nuclei, we conducted subsequent experiments to determine the specificity of this binding.

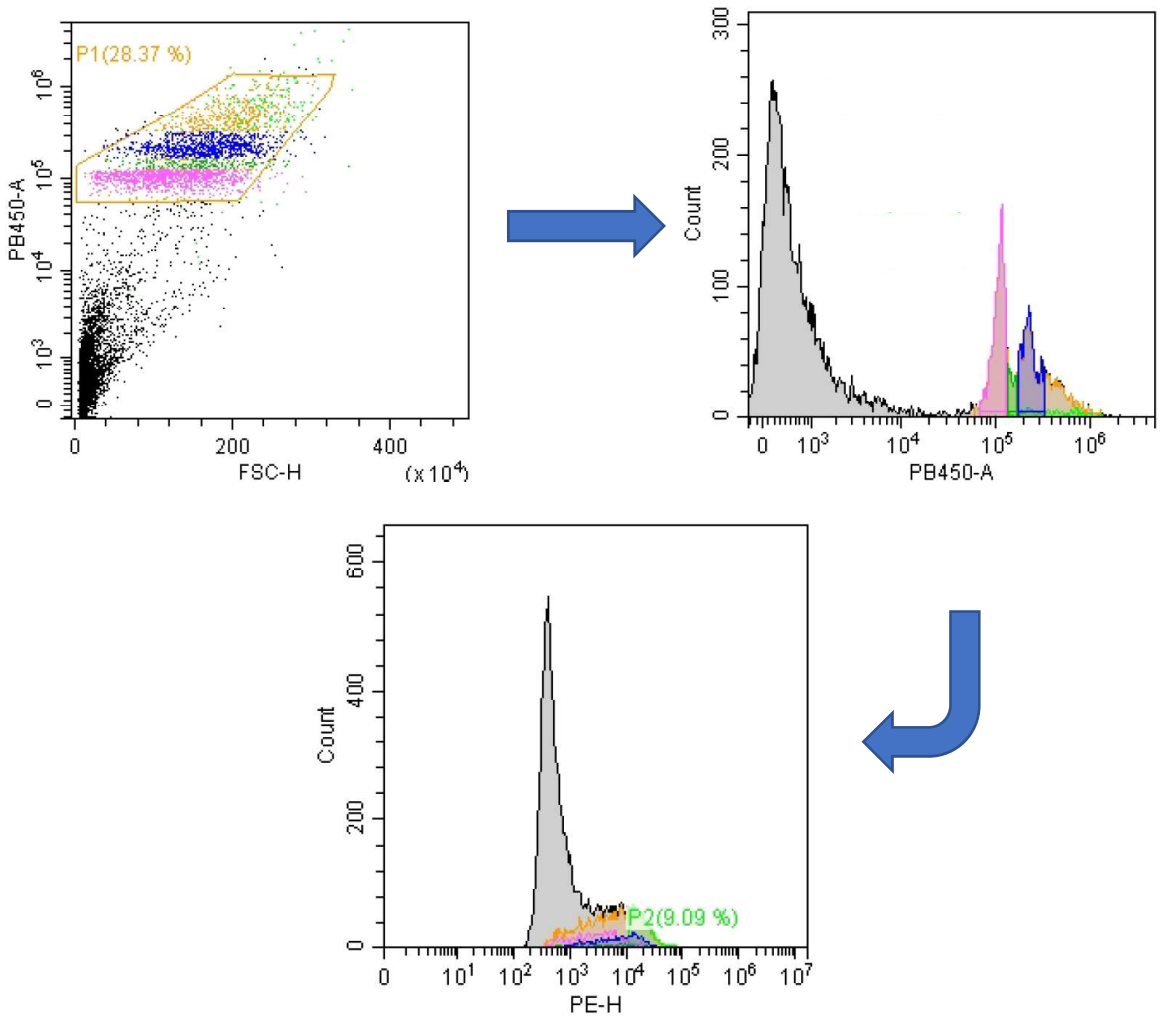
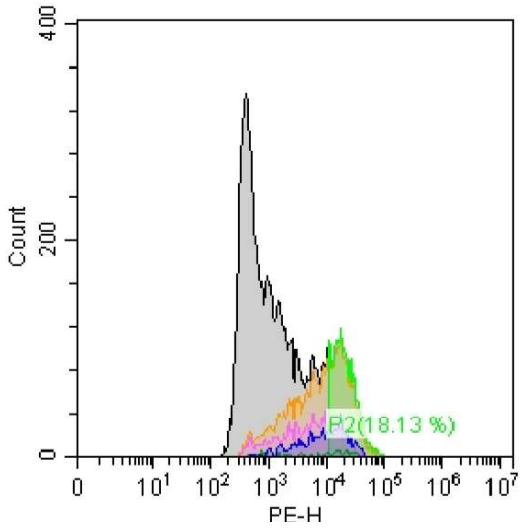


Figure A.2-1. Gating strategy utilizing DAPI (PB450-A) and Cy3 (PE-H) channels. Anaylzed nuclei (colored peaks) were incubated with **INV10** in 30-fold excess of target sequence.

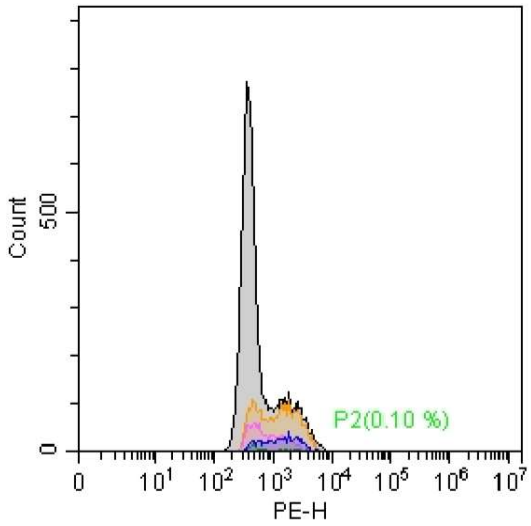
A fluorophore labeled **INV4** DNA strand (**INV4D**; i.e. **INV4** lacking modification/hotspots) was tested to determine levels of non-specific binding. During incubation, negatively charged probes could be wrapped around postiively charged proteins or adhered to other cellular strucutures. Fluorescent values after incubation with this oligonucleotide (**INV4D**) would indicate a non-specific binding baseline unrelated to Watson-Crick binding. Additionally, if its results matched **INV4** this would indicated **INV4** fluorescence is not due to specific hyrbidization to target sequence. A second mis-match control (**MM4**) was utilized as a

negative control. With three basepairs changed to be non-complementary to the target sequence in each Invader probe strand, the oligonucleotide should be unable to bind to the target sequence, and thus will not result in high fluorescent intensity. If intensity matches the results of **INV4**, this would indicate that **INV4** is nonspecifically binding to chromosomal DNA sequences. Figure A.3-2 show the fluorescent intensities of the nuclei population for each treatment.

a) **INV4**



b) **INV4D**



c) **MM4**

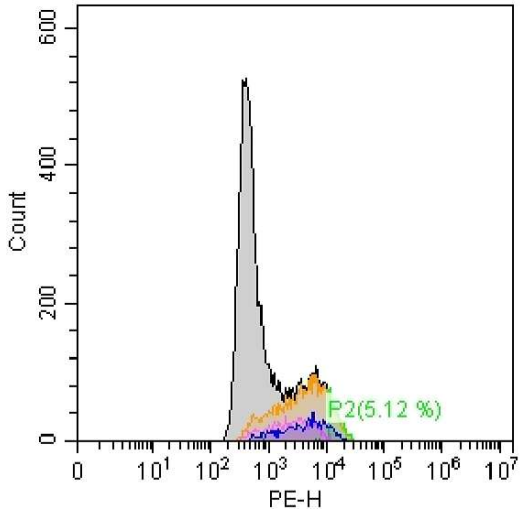


Figure A.2-2. Probe Cy3 intensity (PE-H) values for nuclei populations (count) incubated with a) **INV4**, b) **INV4D**, and c) **MM4**. Grey peak represents debris in sample. Colored peaks are isolated nuclei.

After flow cytometer analysis, mean intesity values for each run were taken for a total of three trials. These values were compared in a one way ANOVA test to determine if there is a statisitical significance between the three groups (**INV4** vs **MM4** vs **INV4d**) (Figure A.2-3). Results show that the **INV4** treatment has a significantly higher mean intensity compared to the two control groups. Additionally, there is no statistical significance between the two

negative controls, highlighting that i) specificy is not lost with the addition of high affintiy Invader monomers, necessitating a correct complementary sequence to invade chromosomal DNA, and ii) **INV4** resulting fluorescent intensity is due to invasion of target DNA.

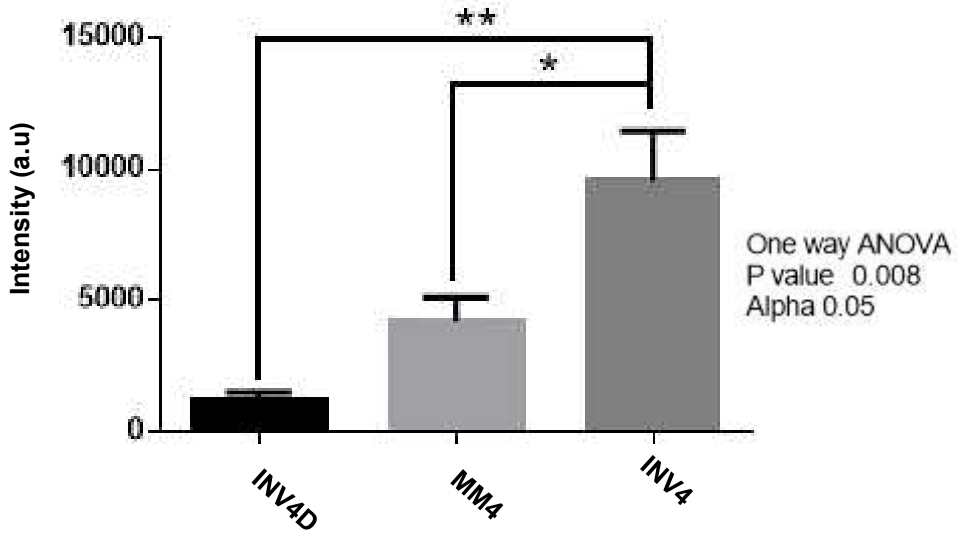


Figure A.2-3. Comparison of Mean Intensities between Unmodified **INV4D**, Mismatch **MM4**, and **INV**4. Mean intensity values for each probe were taken from three separate populations (~3000 nuclei) incubated in 30-fold excess of **INV4** target sequence.

A.3 Conclusions

This study shows the potential use of flow cytomery to anaylze Invader probe detection of chromosomal DNA. First, the protocol used herein to isolated nuclei and incubation Invader probes with their chromosomal targets in suspension has yielded detectable nuclei populations. These can be analyzed for cell cycle stages and Cy3 fluorescent intensity with our determined gating strategy. Sceond, this strategy was successfully applied to determine the specific detection of chromosomal DNA by **INV4**. Overall, these results set the stage for future discoveries and applications related to the Invader probe design.

A.4 Acknowledgements

We would like to thank our consultant Ann Norton for her assistance with flow cytometry use and analysis, and Dr. Onesmo Balemba for his continued support and access to flow cytometer instrumentation (Optical Imaging Core, University of Idaho).

A.5 Supplementary Material

Protocol – cell culture and isolation of nuclei

The male bovine kidney cell line (MDBK, ATCC: CCL-22, Bethesda, MD) were maintained in DMEM with GlutaMax (Gibco, 10569-010) and 10% fetal bovine serum (Invitrogen). Cells were cultured in separate 25 mL or 75 mL flasks at 38.5 °C in a 5% CO_2 atmosphere for 72-96 h to achieve 70-80% confluency. At this point, colcemid (65 µL per 5 mL of growth media) was added and the adhered cells were incubated at 37 °C and 5% CO_2 for an additional 20 min. At this point, the medium was replaced with pre-warmed 0.05% Trypsin-EDTA in DMEM to detach adherent cells (37 °C, up to 8 min). The cell suspension was transferred to a tube and centrifuged (10 min, 1000 rpm). The supernatant was discarded and the dislodged cell pellet was incubated with a hypotonic KCl solution (5-8 mL, 0.075 M, 20 min, 38.5 °C), followed by addition of fixative solution (10 drops, MeOH:AcOH, 3:1), and further incubation with gentle mixing (10 min, room temperature). The suspension was centrifuged (1000 rpm, 10 min), the supernatant discarded, and additional fixative solution (500 µL) added to the nuclei suspension. This was followed by gentle mixing and incubation (30 min, room temperature). After an additional centrifuge step and removal of supernatant, an ice cold solution of 200 µL of 0.1% triton-X-100 in 1x PBS with an additional 100 µL of 1x PBS was added to permeabilize the nuclear membrane. After 10 min, the solution was centrifuged and the supernatant was removed. The nuclei pellet was washed with 1 mL of cold 1x PBS and

followed by an additional centrifuge/removal of supernatant step. The final pellet – containing somatic nuclei – was resuspended in methanol and glacial acetic acid (3:1, v/v) and stored at -20 °C until use.

Protocol – Incubation with Invader probes and sample preparation

After removal of fixative solution, nuclei were suspended in a labeling buffer consisting of Invader probe in PCR buffer (20 mM Tris, 100 mM KCl, pH 8.0). To yield optimal fluorescent signaling, Invaders were added in 30-fold excess to its chromosomal target. Nuclei were incubated for 3 h at 37.5 °C. Labeling buffer was removed by centrifugation/removal of supernatant. Nuclei were rinsed with 1 mL of 1x PBS and incubated with 1 mL of DAPI solution (1 μg/mL) for 10 minutes. Follow a centrifuge step, the DAPI supernatant was removed and labeled nuclei were resuspended in 5-8 mL of 1x PBS.

Protocol – Flow Cytometry analysis

A Beckman Coulter CytoFLEX S flow cytometer outfitted with 4 lasers and able to characterize samples for side scatter, forward scatter and multiple fluorescent markers was used to analyze samples. Lasers able to excite Cy3 (PE) and DAPI (PB450) for fluorescent detection were used in combination with forward scattered light (FSC) to quantify and characterize labeled nuclei samples. Standard settings with the CytExpert for CytoFLEX Acquisition and Analysis software were utilized. The flow cytometer was set to read 10,000 events per run, yielding nuclei counts of ~3000. This software was used for gating strategies and determination of mean intensity values related to area and height of fluorescent signal. Statistical analysis of intensities were conducted with the Data Analysis Tool Pack add-on in Microsoft® Excel®.

A.6 References

1. D. C. Guenther, G. H. Anderson, S. Karmakar, B. A. Anderson, B. A. Didion, W. Guo, J. P. Verstegen and P. J. Hrdlicka, *Chem. Sci*., 2015, **6**, 5006-5015.
2. Rosner, K. Schipany and M. Hengstschläger, *Nature Protocols*, 2013, **8**, 602-626
3. M. Wieser, G. Stadlera, E. Böhma, N. Bortha, H. Katingera, J. Grillariab and R. Voglauer, *Experimental Gerontology*, 2006, **4**, 230-235.
4. R. Arrigucci, Y. Bushkin, F. Radford, K. Lakehal, P. Vir, R. Pine, D. Martin, J. Sugarman, Y. Zhao, G. S. Yap, A. A. Lardizabal and M. L. Gennaro, *Nature Protocols*, 2017, **12**, 1245-1260.
5. S. Roura, M. Fernández, E. Elchinova, I. Teubel, G. Requena, R. Cbanes, J. Lupón, A. Bayes-Genis, *Laboratory Investigation*, 2016, **96**, 1223-1230.
6. J. Perret, Y. Shia, R. Fries, G. Vassart and M. Georges, *Genomics,* 1990, **6**, 482-490.