Invader-mediated Targeting of Chromosomal DNA

A Thesis

Presented in Partial Fulfillment of the Requirements for the

Degree of Master 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 Master of Science with a Major in Chemistry and titled "Invader-mediated 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:	_____ Patrick J. Hrdlicka, Ph.D.	Date: _____
Committee Members:	_____ Eric B. Brauns, Ph.D.	Date: _____
	_____ Peter B. Allen, Ph.D.	Date: _____
Department Administrator:	_____ Ray von Wandruszka, Ph.D	Date: _____

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 along with challenging experimental conditions requirements (homopurine targets; denaturing steps; low ionic strengths; short target sequences). The 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 for specific detection of chromosomal DNA in 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 labelled isolated nuclei, offering a potential advantage over the laborious microscope evaluation of detection (CHAPTER 3).

Acknowledgements

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

Dedication

To my friends and family that did not let me waver

Table of Contents

List of Tables

List of Figures

Statement of Contribution

Coauthors for Chapter 2 are Raymond G. Emehiser, Saswata Karmakar, and Patrick J. Hrdlicka. The work was conceptualized by my graduate advisor Patrick J. Hrdlicka. Experiments were designed by Raymond G. Emehiser, Patrick J. Hrdlicka and myself (Caroline P. Shepard). Synthesis of materials and experiments were completed by Raymond G. Emehiser, Saswata Karmakar, and me. Specifically, Saswata Karmakar synthesized most of the modified oligonucleotides, and conducted some of the thermal denaturation experiments. Raymond G. Emehiser was responsible for recording MALDI spectra of all modified oligonucleotides, performing some of the gel electrophoresis-based dose-response experiments, and carrying out statistical analysis of the data. All remaining experiments - including some thermal denaturation experiments, all determination of thermodynamic parameters, the vast majority of gel electrophoresis experiments, all synthesis and purification of optimized Invader probes, all fluorescent *in situ* hybridization experiments, and all microscopy work - were conducted by me. Data was analyzed jointly by Raymond G. Emehiser, Patrick J. Hrdlicka, and me. I prepared all figures and tables and wrote the original draft of this chapter. The chapter was reviewed and edited by Patrick J. Hrdlicka, Raymond G. Emehiser, and me.

Coauthors for Chapter 3 are Saswata Karmakar and Patrick J. Hrdlicka. This work was conceptualized jointly by graduate advisor Patrick J. Hrdlicka and myself (Caroline P. Shepard). Experiments were designed by me. All experimental work, including data work-up was completed by me. Data was analyzed by Patrick J. Hrdlicka and me. I prepared all figures and tables and wrote the original draft of this chapter. The chapter was reviewed and edited by Patrick J. Hrdlicka and me.

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 inside 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 that is translated into usable structure and function proteins.[1] For example, long fibrous keratin proteins are used to hold epithelial cells together and enzymatic proteins such as RNA polymerases are responsible for synthesizing RNA functional units.[1,2] The genetic flow from DNA to protein synthesis is entitled the Central Dogma of Molecular Biology as it represents the core foundation for life.

Accordingly, abnormalities occurring at any point along this path can be detrimental for living organisms. For these disease-causing abnormalities, 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. From commonly used beta-blockers targeting β1 receptor proteins[3] to lower blood pressure to a Carbidopa/levodopa combination allowing for targeting of dopamine receptors to combat the devastating effects of Parkinson's Disease,[4] there has been success in targeting proteins. However, 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.[5] Moreover, even if a protein can be isolated and studied, there is difficulty in designing drugs with the correct shape and binding complementarity. Even if this is accomplished, high doses of drugs are needed to bind to the tens of thousands of proteins per cell to achieve a therapeutic effect. This incentivizes moving focus 'upstream' to the nucleic acid level to increase the number of targets within DNA and RNA. Identification of lead compounds would be simplified as only the sequence of a diseases-causing gene is needed for design, rather than the complex shape of proteins. Additionally, the predictability of nucleic acid structures in following base-pairing rules (Figure 1.1) or binary recognition rules provides easier determination of targeting molecules in comparison to the complexity of protein structures that take on any number of conformations.

In the movement away from protein targeting, there has been extensive work aiming to develop drugs that target RNA.[6] For example, RNA-dominant diseases, such as myotonic dystrophy represent diseases, are not caused by malformed proteins but rather the formation of toxic RNA that wreak havoc on cellular processes (i.e. long, repetitive RNA inducing effects on the processing of pre-mRNA into mature RNA).[7] Work has been done to design RNA targeting molecules to bind to problematic RNAs to prevent their effects. However, far fewer efforts have been devoted to (and less success realized) developing drugs targeting specific sequences of DNA level, even though this could prove to be a highly advantageous approach. For RNA targeting agents, higher concentrations of probe must be utilized to match the hundreds to 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. For instance, instead of targeting the mutant, toxic huntington protein responsible for the motor and cognitive deficits in those

suffering from Huntington's disease, it may prove advantageous to target the expanded trinucleotide (CAG) repeat on human chromosome 4 to prevent the toxic protein from forming in the first place.[8] These possibilities bring 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). Each strand is comprised of repeating nucleotide units that are connected by a negatively charged phosphodiester backbone. Monomeric nucleotide units consist of a phosphate group, a 2'-deoxy-β-D-*erythro*-pentofuranose, and a nucleobase (A, C, G, and T) that is connected to the C1-position of the sugar ring. C and T are termed pyrimidines for their 6-membered ring structure. The bicyclic nucleobases, G and A, are classified as purines.[9] The sequence of nucleotides determines the 'code' of a DNA, which RNA polymerases transcribe into complementary RNA strands that are eventually translated into proteins or used as functional units themselves. RNAs as a group are very diverse in function and type that have uses beyond protein formation. For instance, there are many others such as non-coding RNAs (ncRNA) that can act to regulate transcription as repressors or activators.[10] They can bind to DNA or proteins to regulate the RNA levels within a cell. Two strands of complementary nucleotides can form an anti-parallel DNA duplex, stabilized by stacking interactions and specific hydrogen binding. 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).[9] These are known as Watson-Crick (WC) base pairs (bps). Strands hybridize in an antiparallel fashion, where the 5' end of one strand aligns with the 3'-end of the other strand. Knowledge of this structure renders DNA an excellent target. If the target sequence (i.e. order of nucleobases) is known, probes can be

designed to be complementary to the target sequence and interfere with processes at the nucleic acid level.

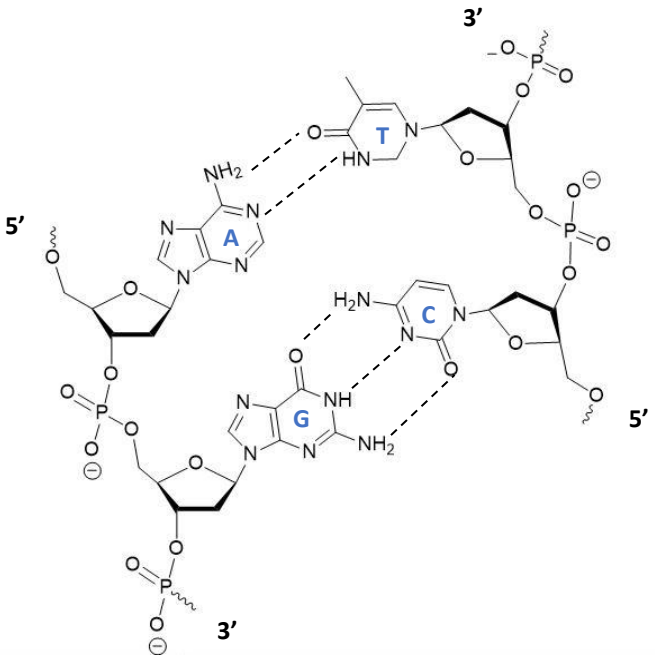


Figure 1.1. Structure of DNA and Watson-Crick base pairing (dashed lines denotes hydrogen bonds).

Duplexes can take on a myriad of geometries, which fall into three major classes termed *A*-type, *B*-type, and *Z*-type.[9,11] These three main helical geometries are double-stranded duplexes that form as a result of base complementarity, stabilization by intra- and inter-strand π-π-stacking of the planar nucleobases, hydration, and positive counterions minimizing repulsion between the negatively charged strands. They differ in the handedness of the helix (right or left), tightness of twist (i.e. length of turn), number of base pairs per turn of helix, and size and depth of minor and major grooves. Chromosomal DNA normally adopts the *B*-type geometry, which is characterized by a right-handed helix with narrow, deep minor grooves and wide, deep major grooves to give 10 bps per turn. The sugar ring plays a key role in determining the 3D structure. In *B*-type double-stranded helices, the

furanose ring adopts so-called *South* type conformations or ‘sugar pucker’ where the C2-atom moves out of the ring’s plane into an *endo* position.[11-13] Consideration of sequence and overall helical geometry is necessary when designing DNA-targeting probes capable of forming Watson-Crick base pairs or 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 (i.e. environmental conditions, delivery to cells) and diagnostic relevance (i.e. genes of interests related to disease).[14-17] This is no easy feat as i) chromosomal DNA is housed in the nucleus of mammalian cells, which requires the DNA-targeting molecule to cross the cell and nuclear membranes, ii) the duplex is supercoiled, wrapped around positively charged histones, and further compacted into chromatin that constitutes chromosomes, raising concerns regarding 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.[9,18] Would targeting the outside major and minor grooves prove sufficiently stable and specific? Could probes be designed to access and bind to the Watson-Crick faces of the DNA? A discussion of established DNA targeting probes follows in the subsequent section, outlining how these questions have been addressed, with a focus on DNA detection and imaging. Before DNA targeting probes are applied 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 long sought to design dsDNA targeting probes (Figure 1.2), often deriving inspiration from naturally occurring structures. It was noted in 1957 that naturally occurring polyadenylic acid (a single-stranded oligonucleotide consisting of only A RNA monomers) hybridizes with two equivalents of polyuridylic acid (a single-stranded oligonucleotide made up of only U RNA monomers) to form a 1:2 triplex.[19] Based on this observation, synthetic triplex forming oligonucleotides (TFOs) were designed to mimic this structure, binding as a third strand to polypurine segments in dsDNA.[20,21] Single-stranded in structure, TFOs form hydrogen bonds to 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. In the parallel orientation, there are two binding motifs: (T,C) and (G,T). This means that only T and C, or G and T base pairs will constitute the TFO sequence. Parallel TFOs will form T:AT, C+GC, and G:GC triplets, constituting the triplex structure. The anti-parallel orientation also has two binding motifs, sharing one with the parallel orientation: (G,T) and (G,A). These TFOs form reverse Hoogsteen hydrogen bonds with the polypurine target region, forming triplets (TFO:dsDNA target): A:AT, G:GC, and T:AT. Binding orientation and the polypurine sequence limitation are determined by the available hydrogen bond acceptor and donor groups present in the major groove by the target sequence. Pyrimidine base pairs lack available groups, requiring a dependence on purine base pairs to constitute the target sequence.

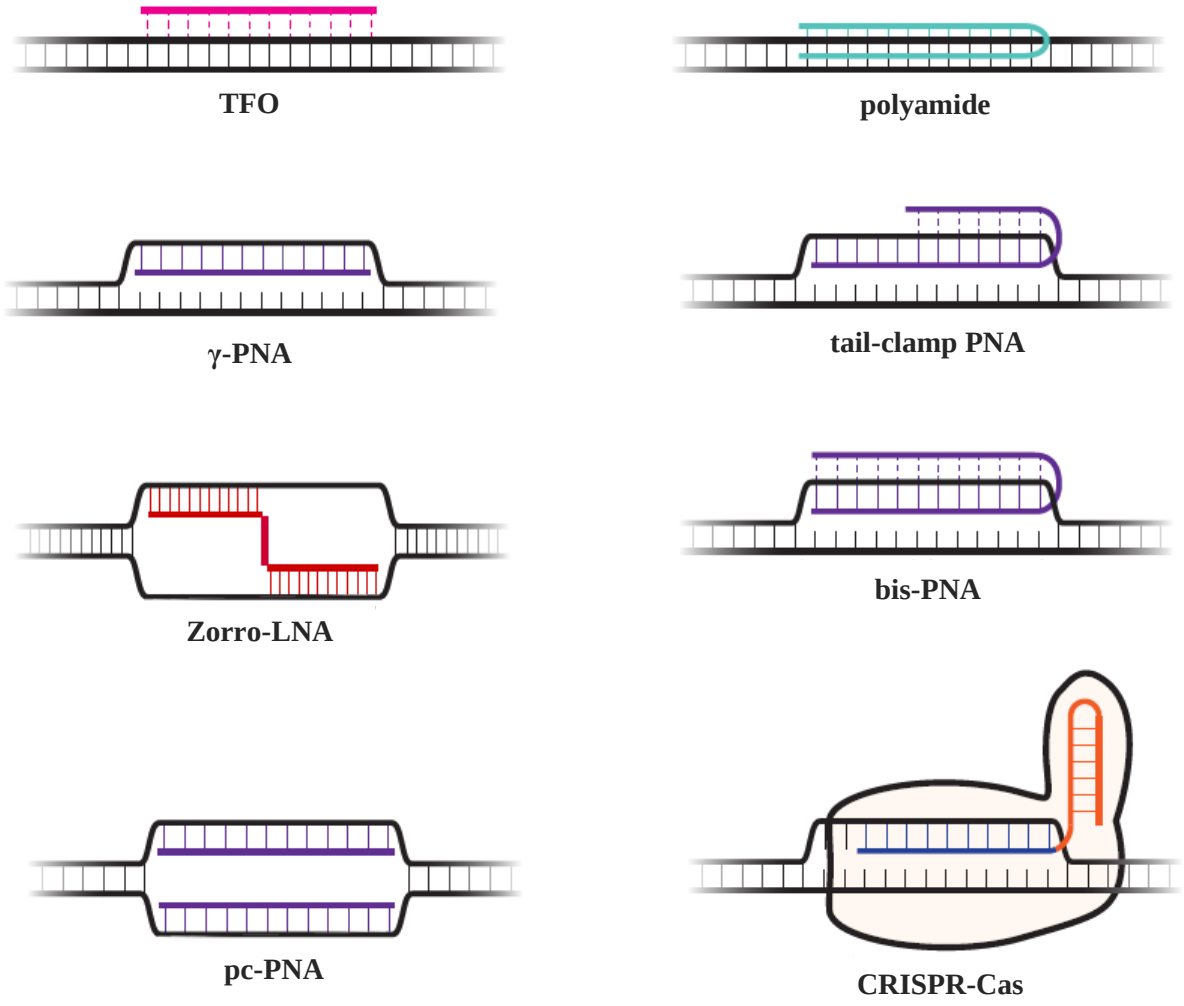


Figure 1.2. Illustrations of various dsDNA-targeting approaches.

There have been many successful applications of TFOs targeting chromosomal dsDNA, including mutagenesis and modulation of gene expression.[22,23] Moreover, TFOs have been used 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 on human and mouse chromosomes.[24]

Despite these successes, TFOs suffer many drawbacks. The polypurine sequence requirement severely limits the number of potential targets, which may preclude TFOs from

detecting most genes and sequences of interest. Additionally, the pH must be acidic (< 6.5) to ensure protonation of cytosines of TC-motif TFOs so they can form C+:CG triplets. This requirement renders TC-motif TFOs suboptimal for applications at physiological pH. Much work has been done to overcome these challenges, such as modifying TFOs with crossover moieties to recognize sequences with pyrimidine interruption sites in otherwise polypurine regions, development of 5-methylcytosine (mC) to negate the pH requirement of TC-motif TFOs, incorporation of cationic modified nucleotides to reduce charge repulsion between the negatively charged TFO and dsDNA strands, and increasing the overall target affinity through incorporation of affinity-enhancing modifications such as Locked Nucleic Acids (LNA).[25-28] While these advancements have led to improvements of traditional TFO designs, the polypurine sequence limitation largely remains, leaving a call for probes that enable sequence-unrestricted dsDNA-targeting.

To address sequence limitations, minor-groove binding polyamides were developed, which are comprised of *N*-methyl pyrrole (Py), hydroxy-*N*-methyl pyrrole (Hp), *N*-methyl imidazole (Im) units that are joined by peptide bonds (Figure 1.2). Each monomeric binds to a different DNA base pair (Py → A:T or C:G, Hp → T:A, Im → G:C) based on binding- and shape-complementarity.[29] Due to their unique binding mode, polyamides can recognize mixed sequences, offering an advantage relative to TFOs. Particularly, minor-groove binding polyamides have shown potential in detecting chromosomal DNA targets. For example, pericentromeric heterochromatin (important regions involved in cellular differentiation) in murine chromosomal DNA was detected using a fluorescent *in situ* hybridization (FISH) assay with FITC (Fluorescein isothiocyanate)-labeled pyrrole-imidazole polyamide probes, as well as telomeres.[30,31] However, this design leaves much to be desired as they cannot distinguish

between AT and TA base pairs, which poses concern regarding target fidelity and specificity. Moreover, they are most useful against short sequence targets (<8 bp). As the length of the target increases, shape complementarity is lost and binding capabilities are reduced.[17,29,32-34]

Peptide Nucleic Acid (PNA) probes (Figure 1.2) have been used to target chromosomal DNA for the purpose of manipulating gene expression, performing site-specific mutagenesis, and detecting specific dsDNA sequences.[15,34,35] Similar to TFOs, polypyrimidine PNAs can form triplexes with double-stranded polypurine DNA sequences. Nucleobases (A, C, G, T) are attached to an achiral, uncharged *N*-(2-aminoethyl)glycine backbone (Figure 1.3). The unique placement of these nucleobases provides the same distancing and optimal conformation to match the base-pairs along a DNA strand. The lack of charge results in decreased repulsion upon duplex formation with complementary DNA (cDNA); accordingly, fast association kinetics and high affinity towards cDNA is observed, offering an advantage over negatively charged oligonucleotide-based probes. 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.[36] However, TFO-like triplex formation, which is the preferred binding mode at physiological conditions and low PNA concentrations, is similarly limited to polypurine regions.[37]

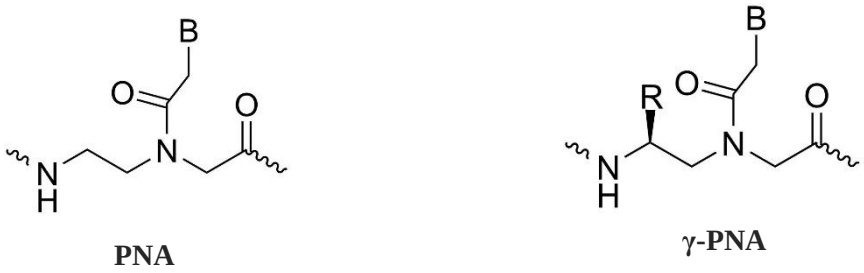


Figure 1.3. Structures of PNA and γ-PNAs. (B denotes nucleobases A, C, G, and T). First generation γ-PNAs utilized a methyl R group. Developments have been made in varying the R substituent to induce optimal structural conformations and increase solubility (i.e. ethylene glycol units).

However, PNAs display a diversity of binding modes.[37] Thus, alternative triplex structures can form at low salt conditions, which destabilize the target duplex, facilitating formation of a PNA:DNA:PNA triplex. In this triplex invasion mode, one PNA strand binds to a polypurine region from the major groove via Hoogsten base-pairing, which results in disruption of the underlying dsDNA duplex. This allows for invasion of a second PNA strand, i.e., disruption of the hydrogen bonds of the target duplex frees the nucleobases to form classical Watson-Crick hydrogen bonds with the nucleobases of the PNA probe. 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).[35] 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 mitochondrial and chromosomal DNA through combined incubation of a bis-PNA and an additional padlock probe.[38] Despite targeting and detection success, the polypurine sequence restriction persists as invasion is contingent on triplex formation.

Tail-clamp PNA bind similarly to bis-PNA; however, the invading PNA strand which forms Watson-Crick base pairs is extended beyond the polypurine region to improve binding

affinity and enable some level of mixed-sequence recognition (Figure 1.2).[39] Unfortunately, tail-clamp PNA still require the presence of polypurine regions for binding, although the polypurine region can be significantly shorter than for TFOs.

Efficient sequence-unrestricted recognition of dsDNA sequences is the ultimate goal for dsDNA-targeting probes. To improve on past single-stranded PNA designs, next-generation γ-PNAs were developed. Introduction of a chirality center through substitution at the γ carbon position of the glycine backbone with various R groups (i.e. methyl, ethylene glycol), enables the recognition of mixed dsDNA sequences (Figure 1.3).[40,41] As with all PNA probes, improved target affinity is facilitated by the uncharged backbone due to reduced electrostatic repulsion. However, the additional chiral group preorganizes the PNA strand to match the chiral right-handed conformation of chromosomal DNA, and therefore has increased affinity towards its cDNA stand that enables invasion. Upon invasion, the dsDNA target strand is displaced to form Watson-Crick base pairs with γ-PNA (Figure 1.2). This binding mode differs from triplex invasion inasmuch formation of an initial triplex is not needed. Several designs and optimizations have been deployed such as G-clamps (increase binding affinity by increasing the number of hydrogen bond between GC bps), substitution of the γ-methyl group for diethylene glycol (solubility improvements), addition of acridine units (positively charge intercalators that improves binding to negatively charged DNA target), and increasing probe length (improve free binding energy between PNA and DNA target).[42-44] These improvements have resulted in γ-PNAs capable of mixed-sequence DNA detection, e.g. labeling of telomeres.[45] However, current efforts are attempting to reducing aggregation, solubility issues, and poor cellular uptake to advance these probes into *in vivo* studies.

Another probe class that has been reported to result in mixed-sequence recognition of DNA are Zorro LNAs. The original design of Zorro LNAs entailed a short DNA duplex with two single-stranded DNA-targeting overhangs that form a “Z”-shaped probe (Figure 1.2). For the classic Zorro-LNA structure, a DNA duplex was designed to have single stranded overhangs that are complementary to a DNA target sequence. Affinity enhancing LNA monomers have been incorporated into the duplex region and the single-stranded overhangs to ensure the construct remains intact and to provide the necessary energetic gradient for dsDNA invasion. 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.[46] These designs require special attention to synthesis in order to achieve an inversion of polarity so that the single stranded overhangs can bind antiparallel to target sequence. 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 potential issues of self-hybridization due to the affinity enhancing LNA monomers, and there is evidence that the binding mechanism might require initial triplex formation via Hoogsteen base-pairing of one of the single-stranded overhangs to facilitate invasion.

The most relevant type of dsDNA-targeting binding mode for the research discussed later herein, is double-duplex invasion. It differs from duplex invasion, inasmuch each strand of a double-stranded probe is used to form Watson-Crick base pairs with complementary strands encompassed in the dsDNA target. 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 bind to dsDNA via double-duplex invasion (Figure 1.2).[47,48] Thus this design entails a labile PNA duplex. To reduce the stability of the PNA

duplex and increase each PNA strand's affinity for cDNA, A and T nucleobases are replaced with 2,6-diaminopurine (D) and 2-thiouracil (S) (Figure 1.4). This modification of A and T cause the two probe strands have a higher affinity towards cDNA than their complementary pcPNA partner as D forms three hydrogen bonds with thymine, and S form two hydrogen bonds with adenine. When these monomers are placed across from each other on opposing strands of the double-stranded probe, the 2-amino group of the D monomer and the 2-thio group of the S monomer clash, resulting in weakened base-pairing. This more labile probe is 'activated' to follow a thermodynamic gradient (i.e. unstable to stable) 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.[49,50] However, this design requires a high level of modification per strand (i.e. many D and S units) and low ionic strength to increase access to the target site and decrease probe stability, especially when targeting GC-rich dsDNA, to facilitate recognition.

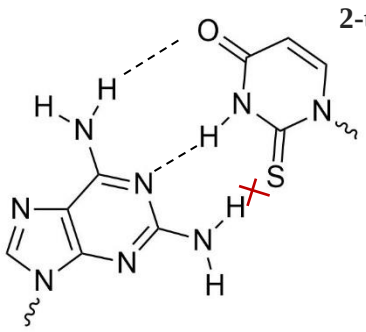


Figure 1.4. Pseudocomplementary (pc) PNA base pair.

Notable recent dsDNA-targeting approaches include CRISPR Cas (Clustered Regularly Interspaced Short Palindromic Repeats with associated Cas protein), which has

taken the world by storm as a potential tool for gene detection and manipulation (Figure 1.2). It is a cluster of regulatory interspaced short palindromic repeats within CRISPR associated proteins which have an RNA-guided nuclease function. Once viral genetic material is incorporated into these repeat regions, guide RNA capable of recognizing the virus is made and incorporated into the protein complex. Once this nucleoprotein unit is bound to its complementary target, the nuclease CAS protein can induce cuts to perform gene edits, removing the viral genome. Utilized by bacteria as part of virus "immune" responses, this construct can be preprogrammed and redesigned to target chromosomal DNA for the purposes of gene manipulation and detection.[51,52] For example, a CASFISH assay was developed by deactivating the nuclease function of Cas9 protein and fusing it EGFP (Enhanced Green Fluorescent Protein) for fluorescent visualization. With this design, pericentromere, centromere, G-rich telomere, and different coding genes in genomic DNA were fluorescently labeled and detected.[53] While CRISPR-Cas-based constructs have been very successful in targeting mixed-sequence dsDNA *in vitro*, they face many challenges such as inadequate specificity, limited targets (e.g. binding of PAM unit of CAS can limit targeting regions through a one target every eight base pair limitation), major *in vivo* challenges, including stimulation of immunogenic responses and problematic delivery[54,55] (i.e. plasmids coding for the CRISPR Cas protein units must be delivered to the cell and subsequently transcribed and translated. This proves difficult due to complexity and ability to cross cellular membranes).

1.3 Introduction to Invader Probes

Given the sequence restrictions and environmental challenges posed with other DNA targeting probes, there is a need to design probe technologies that accurately recognize mixed sequence dsDNA target regions at physiological relevant conditions for diagnostic and

therapeutic purposes. To meet this need, our lab has been working on developing and optimizing double-stranded probes, termed Invader probes. Like pcPNA, Invader-mediated recognition of dsDNA is based on double-duplex invasion although the way the driving force for recognition is generated is fundamentally different (Figure 1.5). Placement of intercalator-functionalized monomers like 2'-*O*-(pyren-1-yl)methyl RNA monomers in +1 interstrand zippers produces results in a destabilized probe duplex that is activated for target recognition. By arranging these monomers in +1 interstrand zippers (also termed energetic hotspots), the intercalating pyrene moieties are forced to occupy the same space, resulting in a violation of the neighbor exclusion principle (NEP).[56] The ingenuity of this design is the use of pyrenes as DNA intercalators. They slide in between base pairs and interact with them through π-π stacking interactions. According to the NEP, intercalators are limited to binding every second base pair as there is a limit to the extent that the DNA helix can locally unwind and expand to accommodate intercalators (every intercalation event results in an elongation of the DNA duplex by ~3.4 Å).Violation of NEP lead to an easily denaturing duplex. Once apart, both probe strands display increased affinity towards cDNA due to stacking interactions by the pyrene moieties and neighboring base pairs, which generates the driving force for double-duplex invasion.[56-58]

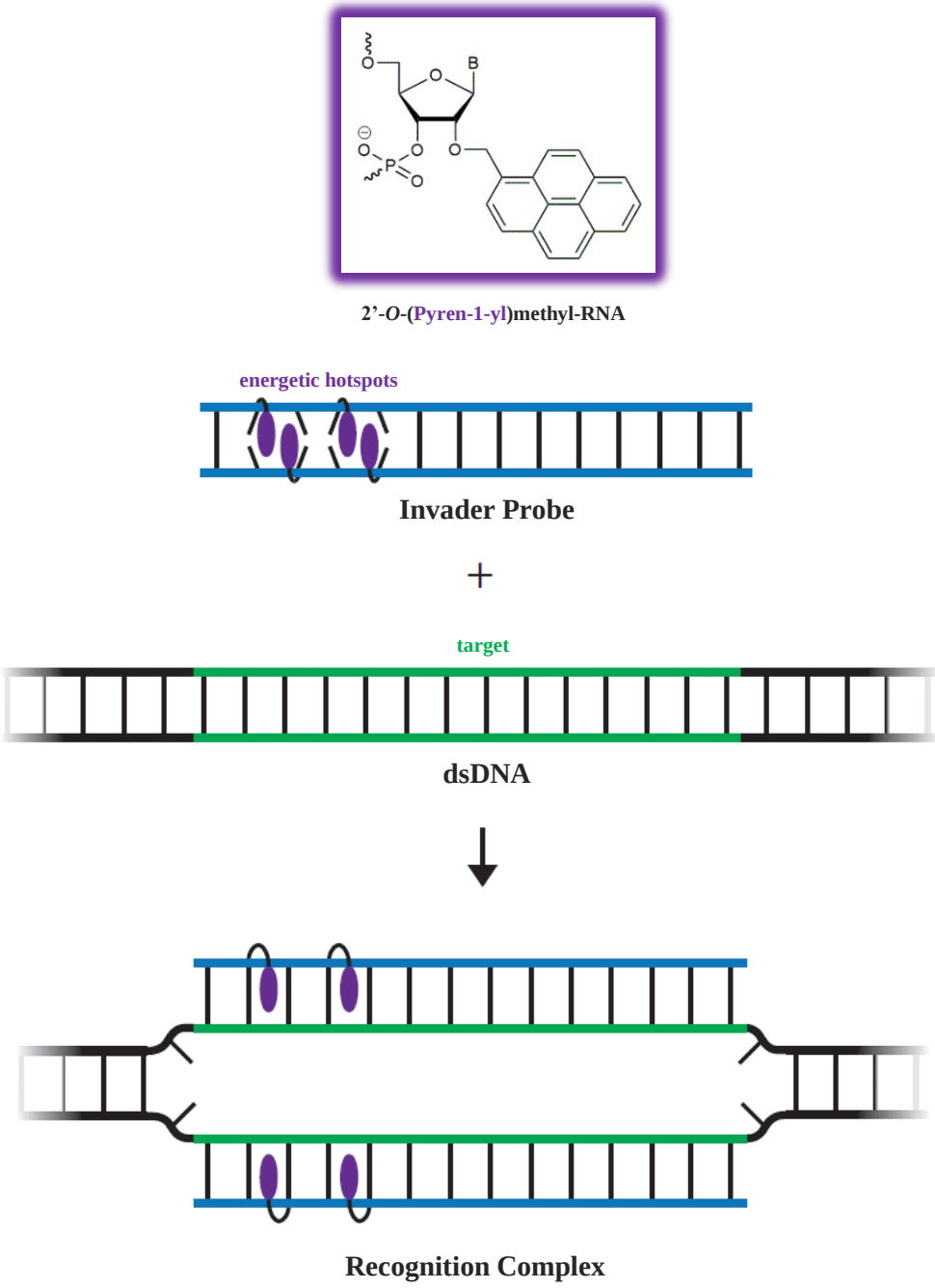


Figure 1.5. Double duplex invasion mechanism of Invader probes.

1.4 Development of Optimized Invader Probes

The original Invader probes used N2'-pyrene-functionalized-2'-amino- α-L-LNA monomers as the key building blocks (Figure 1.6).[59] The 2′-C,4′-C-oxymethylene linker conformationally restricts the sugar ring into a C2'-*endo* sugar conformation. The use of a conformationally restricted nucleotide enhances the affinity of the probe strands as the intercalators are preorganize to intercalate, resulting in prominent destabilization of the probe duplexes and greatly increased affinity towards cDNA of the individual probe strands. While Invader probes based on these modifications result in very efficient recognition of model dsDNA targets (one modification results in a melting temperature increase of +7.0 to +16.0 ºC of probe-target duplex in comparison to unmodified duplex), the synthesis of the monomers is synthetically challenging, which generated a call for functional analogs that are synthetically more accessible. Though more flexible than their LNA-based predecessors, the 2'-*O*-(pyren-1-yl)methyl-RNA monomers (Figure 1.5), have less synthetic difficulty and retain adequate hybridization properties (i.e. one modification yields probe-target duplex melting temperature +13 ºC in comparison to unmodified duplex).[57] The 2'-*O*-(pyren-1-yl)methyl-RNA monomers were accordingly selected as next-generation Invader building blocks (Figure 1.6).

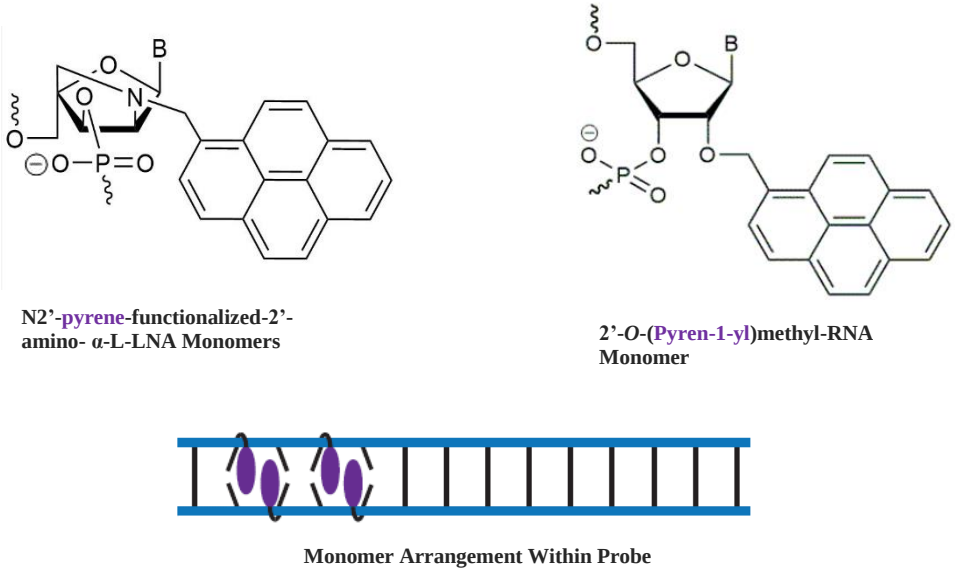


Figure 1.6. Structures of N2'-pyrene-functionalized-2'-amino- α-L-LNA and 2'-*O*-(Pyren-1-yl)methyl-RNA monomers and the +1 interstrand arrangement within the probe duplex.

Having identified synthetically accessible next-generation Invader building blocks, it became important to establish design rules and determine optimal probe architectures. First, the impact of the nucleobase (B) of the Invader monomer was established and the 3'- flanking base pairs was studied as this would establish preferred sequence contexts for dsDNA-targeting. To this end, 2'-*O*-(pyren-1-yl)methyl-RNA monomers **A**, **T**, **C**, and **G** were incorporated into Invader probe strands.[60] 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 duplexes between probe strands and cDNA are more stabilized when the monomers are flanked by a 3'-purine. As a consequence hereof, Invader probes modified with **U** and **C** monomers showed greater potential for dsDNA recognition. These studies laid the groundwork for future

Invader probe designs, which favored the use of **U** and **C**, with limited use of **A**, and avoidance of **G** monomers. Additionally, incorporated monomers should by flanked by a 3'-purine.

Next, the optimal modification level was studied 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. Invader probes with a low modification density result in low dsDNA-targeting efficiency, whereas a very high modification density results in decreased binding specificity.[61,62] The studies revealed that probes that are 15-30% modified display adequate affinity and specificity towards their dsDNA targets.

1.5 Applications of Invader Probes

DNA targeting tools have many potential diagnostic application. Recent studies have shown the potential of Invader probes for such purposes. Thus, they can detect different pathogenic bacteria that are common food contaminates through use of sandwich assay.[56] Invader probes were immobilized on the surface of a 96 well plate. Fragments of dsDNA modeling bacterial DNA sequences were 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 was quantified. This assay might be advantageous evaluate potential microbial contamination in food preparation as Invader probes have mixed-sequence recognition capabilities (i.e. target sequence is not a limitation), occurs isothermally, and is 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 Invader probes for detection of chromosomal DNA in non-denaturing fluorescent *in situ* hybridization (nd-FISH) assays. Towards this end, Invader probes were labeled with Cy3 fluorophores in order to be visualized by fluorescence microscopy. These fluorophore-labeled probes were incubated with fixed nuclei on microscope slides under physiological relevant conditions. Following incubation, nuclei were DAPI counterstained to signal the location of the nucleus and overlaid with the Cy3 signal from the probe. Using this assay, Invader probes have been shown to effectively target telomeric DNA and specific targets on Y chromosomes of bovine cells.[61,62] Control experiments verified that binding is specific through relevant mismatch studies, incubation with female bovine endothelial cells (displayed no signal), pre-treatment with DNase (disappearance of signal), and pre-treatment with RNase or proteinase (signal is maintained). This represents an advancement in DNA targeting probes as Invader probes can recognize mixed-sequence chromosomal DNA under non-denaturing conditions, underscoring that recognition of chromosomal DNA at physiological conditions is within reach.

1.6 Conclusion

Development of probes capable of recognizing specific mixed-sequence double-stranded DNA targets has remained largely elusive. Such nucleic acid targeting probes would have widespread applications in diagnostics and gene expression studies. Invader probes utilize 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 usage in diagnostics and detection as well as biological applications.

1.7 References

1. T. Schneider-Poetsch and M. Yoshida, *Annu. Rev. Biochem*, 2018, **87**, 391-420.
2. D. Depianto and P. Coulombe, *Experimental Cell Research*, 2004, **301**, 68–76.
3. J.C. Davies, *J. R. Coll. Gen. Pract*., 1976, **26**, 219-226.
4. R. Dhall and D. L. Kreitzman, 2016, *Neurology*, **86**, S13–S24.
5. M. Pertea, *Genes*, 2012, **3**, 344-360.
6. S. T. Crooke, J. L. Witztum, C. F. Bennet and B. F. Baker, *Cell Metabolism*, 2018, **27**, 714-739.
7. R. J. Osborne and C. A. Thornton, Osborne, *Human Molecular Genetics*, 2006, **15**, R162–R169.
8. M. Jackson, L. Marks, G. May and J. B. Wilson, *Essays Biochem.*, 2018, **62**, 643-723.
9. A. Travers and G. Muskhelishvili, *FEBS Journal*, 2015, **282**, 2279-2295.
10. J. A. Goodrich and J. F. Kugel, *Nat. Rev. Mol. Cell Biol*, 2006, **7**, 612–616.
11. R. Dickerson, H. Drew, B. Connor, R. Wing, A. Fratini, and M. Kopka, *Science*, 1982, **216**, 475-485.
12. A. Rich, *Nat. Struct. Mol. Biol*, 2003, **10**, 247-249.
13. S. Arnott. *Trends in Biochemical Sciences*, 2006, **31**, 349-354.
14. I. Ghosh, C. I. Stains, A. T. Ooi and D. J. Segal, *Mol. Biosyst*., 2006, **2**, 551-560.
15. P. E. Nielson, *Biodiv*., 2010, **7**, 786.
16. Y. Aiba, J. Sumaoka and M. Komiyama, *Chem. Soc. Rev*., 2011, **40**, 5657-5668.
17. T. Vaijayanthi, T. Bando, G. N. Pandian and H. Sugiyama, *ChemBioChem*., 2012, **13**, 2170-2185.
18. M. Radman-Livaja and O. J. Rando, *Dev. Biol.*, 2010, **339**, 258-266.

19. G. Felsen, D. R. Davies and A. Rich. *J. Am. Chem. Soc*., 1957, **79**, 2023-2024.

20. M. Duca, P. Vekhoff, K. Oussedik, L. Halby and P. B. Arimondo, *Nucleic Acids Res*., 2008, **36**, 5123-5138.

21. S. Buchini, *Current Opinion in Chemical Biology*, 2003, **7**, 717-736.

22. K. M. Vasquez, L. Narayanon and P. M. Glazer, *Science*, 2000, **290**, 530-533.

23. K. R. Fox, D. A. Rusling, V. J. Broughton-Head and T. Brown, *Current Chemical Biology*, 2008, **2**, 1-10.

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

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

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

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

28. 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.

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

30. 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.

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

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

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

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

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

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

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

38. A. I. Yaroslavsky, I. V. Smolina, *Chemistry and Biology*, 2013, **20**, 445-453.

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

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

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

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

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

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

45. 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.

46. 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.

47. 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.

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

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

50. 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.

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

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

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

54. A. Pickar-Oliver and C. A. Gersbach, *Nat. Rev. Mol. Cell. Biol.*, 2019, **20,** 490–507.

55. M. Adli, *Nat Commun*, 2018, **9**, 1911.

56. B.A. Didion, S. Karmakar, D. C. Guenther, S. P. Sau, J. P. Verstegen and P. J Hrdlicka, *ChemBioChem*, 2013, **14**, 1534-1538.

57. 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.

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

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

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

61. 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.

62. 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

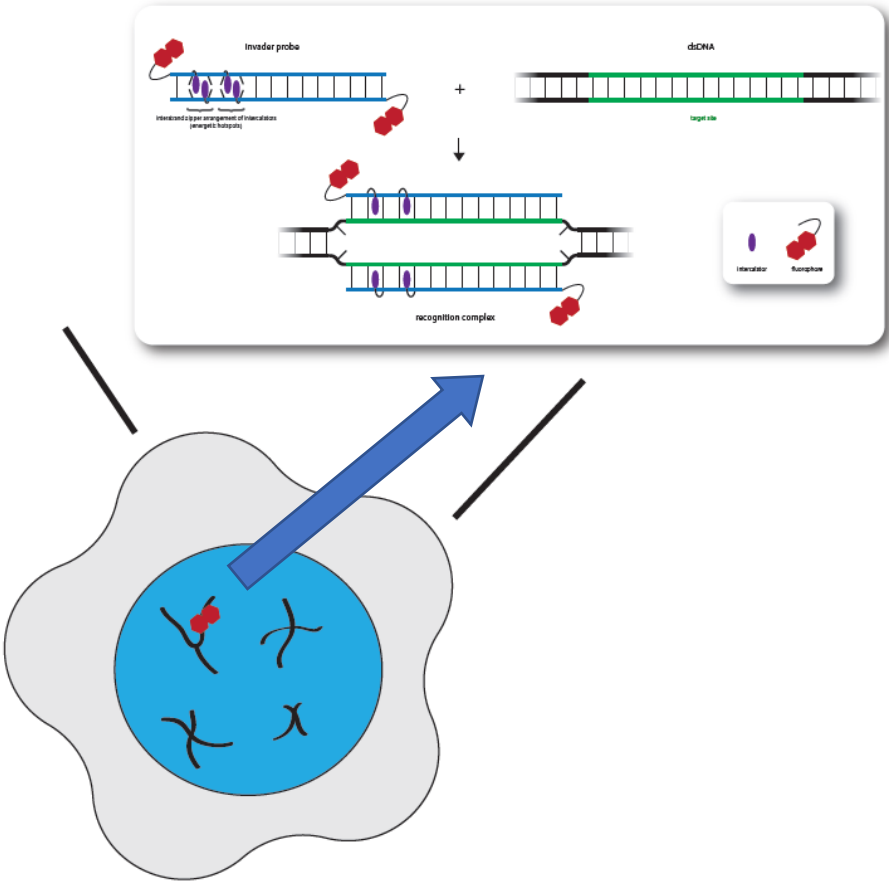


Figure 2.1. Schematic of Invader probe fluorescent labeling of chromosomal DNA.

Abstract

Future advancements in chemical biology will benefit from 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 or peptide nucleic acids (PNAs), and minor-groove-binding polyamides only bind DNA under

limited experimental conditions (homopurine targets; low pH; short target sequences), whereas probes capable of binding with the Watson Crick faces of the target region (y-PNAs, pc-PNAs, etc) require denaturing steps, low ionic strengths, or high levels of modification. The more recently discovered CRISPR-Cas9 approach requires transfection of plasmids encoding CRISPR-Cas9 components and have specificity issues. 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 complementary double-stranded DNA (dsDNA) regions (Figure 2.2). 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 demonstrated specific binding of several of the Invader probes, resulting in high signal intensity. A subsequent Spearman rank correlation analysis of properties and design features of Invader probes revealed that derived C_{50} values from our model hairpin assay and binding efficiency in denaturing FISH experiments are predictive of probe targeting of chromosomal DNA under near physiological conditions. Moreover, this analysis revealed that increased probe modification density results in increased recognition of chromosomal DNA. Accordingly, three Invader probes were optimized by increasing modification incorporations and were subsequently evaluated. Results confirmed the predictions from the statistical analysis and displayed improved recognition properties.

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, the traditional peptide nucleic acid (PNA) design[3,4] and 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 the number of suitable 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 orientation to interact with TFOs. 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 eight nucleotides as structural shape compatibility with the minor groove is gradually lost with increasing length.[10,11] Additionally, Py-Im probes do not distinguish between AT and TA base pairs, which reduces their specificity.[12,13] Strand invasion probes have been studied to overcome limitations of groove binding. For instance, PNAs have nucleobases attached to an uncharged *N*-[2-aminoethyl)glycine backbone which reduces charge repulsion associated with duplex formation.[14,15] In specific sequence contexts, this facilities Hoogsteen base pairing, opening the target regions for formation of very stable duplexes with cDNA *via* Watson-Crick base pairing. Due to the need for Hoogsteen base pairing for invasion, PNAs are limited to targeting sequences with a short stretch of polypurines. Different designs have been employed

to overcome these limitations, such as bis-PNA and tail-clamp PNA that utilize Hoogsteen base pairing to perturb the target duplex to facilitate Watson-Crick hybridization between the cDNA and PNA strands.[15,16] While these designs allow for formation of more stable recognition complexes and some mixed-sequence hybridization, they are still woefully restricted due to the necessity of a polypurine region to form the Hoogsteen base pairs to gain access to the Watson-Crick base pair faces. Next generation single stranded γ-PNA has been used to reduce sequence limitations by its significant increase in binding affinity.[17,18] Introduction of a chiral center at the γ-position preorganizes the probe strand, thereby reducing entropic costs and increasing cDNA affinity. Although γ-PNA can recognize mixed-sequence dsDNA target regions, they require low ionic conditions to destabilize the target region and/or affinity enhancing incorporations (i.e. G-clamps), and addition of substituents that increase their solubility (i.e. ethylene glycol) in order to facilitate duplex invasion,[19,20] rendering them less ideal synthetically and for physiologically relevant ionic strengths. The recent discovery of CRISPR-Cas9 and its programmability to introduce 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.[21,22] Unfortunately, these abilities are limited to *ex vivo* and *in vitro* studies due to the need for plasmid transfection.[22,23]

To overcome restriction in the choice of target sequences and experimental conditions, we have developed modified double-stranded DNA probes termed Invader probes, which have been shown to target mixed-sequence chromosomal DNA through double-duplex invasion (Figure 2.2a).[24-26] The driving force for dsDNA-recognition is the energy difference between the less stable reactants (i.e. the double-stranded probe and dsDNA target) and the highly stable

products (i.e. the two probe-target duplexes formed as part of the recognition complex). The Invader probe is destabilized or "activated" due to 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.[24-26] The hotspots are comprised of 2'-*O*-(pyren-1-yl)methyl-RNA monomers (Figure 2.2b).[27] 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 2.2a). We have previously shown that Invader probes can be designed to recognize different targets such as: DNA fragments specific to different food pathogens (28-mer mixed sequence dsDNA fragments in a sandwich assay using),[25] telomeric DNA,[28] and specific regions of Y chromosomes in isolated nuclei from male bovine kidney cells under non-denaturing conditions.[29] Encouraged by these results, In order to fundamentally improve our understanding of targetable regions and how Invader probes should be designed for efficient recognition, we herein set out to investigate the detection chromosomal DNA by a library of Invader probes with different energetic hotspot arrangements and targeting various sequences in a *DYZ*-1 model system. In this region, ~1175 base-pair repeat region was targeted in order to specifically label the Y chromosome of bovine cells. Thermal denaturing temperatures, thermodynamic parameters associated with duplex formation, and dose-response experiments conducted with a model double-stranded hairpin were used as measures of the physical properties and invasion potential for each of the ten probes. This was followed by an evaluation

of their ability to target chromosomal DNA by deploying denaturing and non-denaturing FISH assays. A Spearman rank correlation analysis was subsequently performed to identify biophysical attributes that correlate with chromosomal DNA targeting efficiency under physiologically relevant, non-denaturing conditions. Informed by these initial studies, Invader probes were re-designed and evaluated to establish the predictive power of the design rules derived from our statistical analysis.

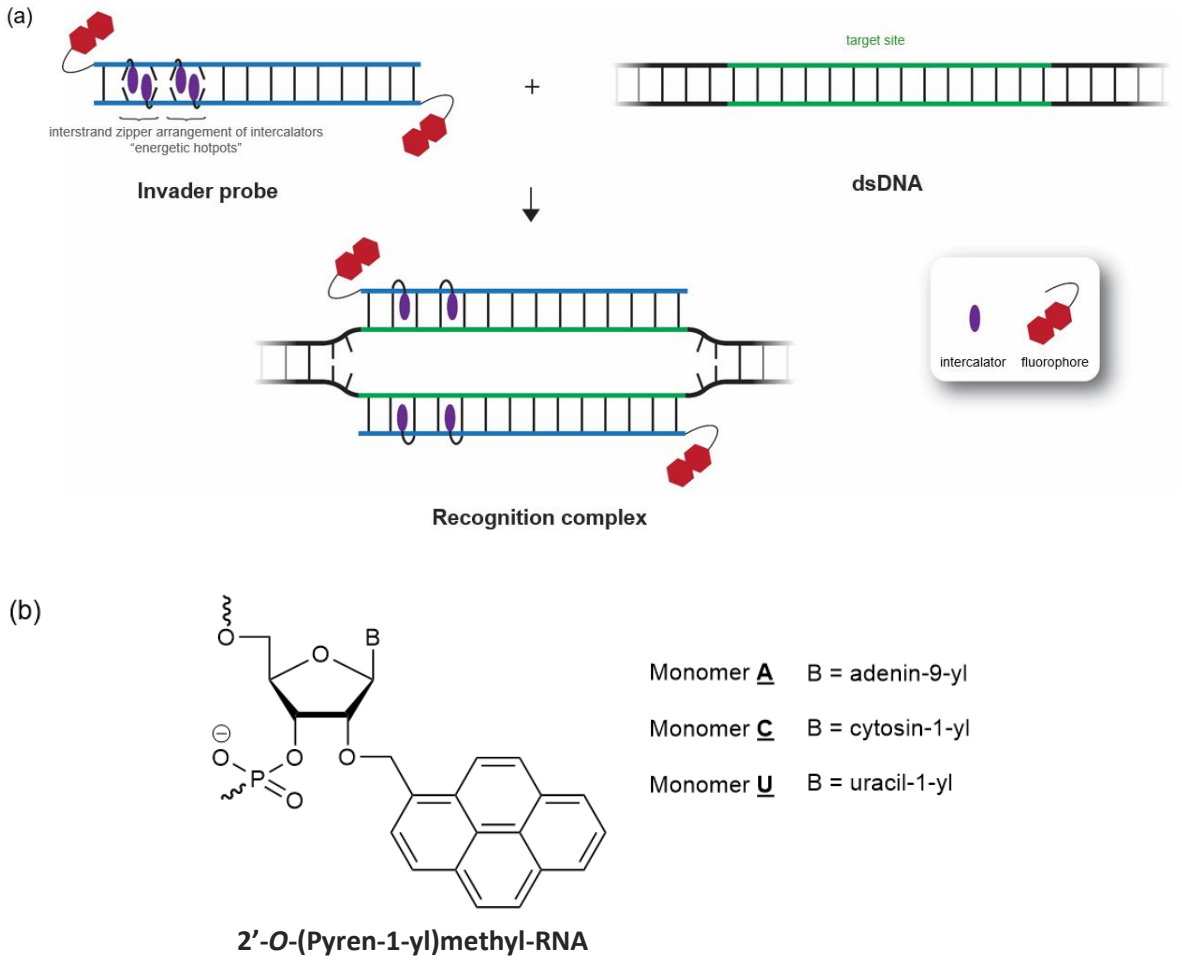


Figure 2.2. (a) Illustration of the recognition process for Invader-mediated recognition of dsDNA via a double-duplex invasion process. (b) Structure of monomers used herein.

2.2 Results and Discussion

Experimental design of Invader probes

Ten oligodeoxyribonucleotide (ON)-based Invader probes (Table 2.1) - varying in length (14-16 base pairs (bps)), number of energetic hotspots (3-4), and GC-content (GC%) (31-71%) - were 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 2.18).[30] The corresponding individual probe strands were obtained using previously established machine-assisted solid-phase DNA

synthesis protocols.[28] Each Invader probe targets one unique region within the ~1175 bp repeat with exception of **INV4** whose target sequence is present six times within each repeat. The identity and purity of the modified ONs was established by MALDI-TOF (Table 2.10 and Figures 2.13, 2.14, and 2.15) 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 2.1) was chosen based on our prior studies[28,29] which indicated that Invader probes with a hotspot content of 20-30% strike a favourable balance between displaying high binding affinity and ensuring satisfactory binding specificity. Less densely modified probes (<15%) display reduced binding affinity, whereas more densely modified probes (>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 2.1). Invader probes are slightly destabilized, displaying T_ms that on average are ~1.6 °C lower than the T_ms of the corresponding unmodified DNA duplexes **DNA1**-**DNA10**, although the ΔT_m values vary widely (-18.5 °C to +9.5 °C, Table 2.1). Utilizing a Spearman rank statistical analysis (i.e. correlating each parameters rank on a 1 to 10 scale instead of using the determined numerical values), there are no obvious correlations between ΔT_m of Invader probe and their length, GC-content, nor modification density, suggesting that there is a threshold to the instability derived by Invader modification incorporation. Probes of length 14-16 of variable GC-content will not be destabilized linearly by increasing the modification density. 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 2.1). 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 modifications levels of >20% and ≤20%, respectively).[A] These observations follow our expectations that Invader probes are more labile[24,31] as the neighbor exclusion principle is violated due to high localized intercalator densities,[32,33] 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.[25,31] A Spearman rank correlation analysis, supports these observations as a significant positive correlation between modification densities and ΔT_ms of probe-target duplexes is observed (5'-**ON**:cDNA, r_s = 0.774, p value = 0.009) and 3'-**ON**:cDNA (r_s = 0.661, p value = 0.037) (Table 2.2).

[A] 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.[30] 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 2.1. 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'-ON: cDNA	3'-ON: 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	20.0%	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[b] [-18.5]	66.0 [+13.0]	66.0 [+13.0]	+44.5	+25.5	25.0%	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	20.0%	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	25.0%	31%
INV8	5'-Cy3-TT**C**ACAGCCC**U**G**U**GC-3' 3'- AAG**U**GTCGGGA**C**A**C**G-Cy3-5'	58.5[b] [-1.5]	70.5 [+10.5]	74.5 [+14.5]	+26.5	+6.0	20.0%	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	20.0%	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 = denaturation curves recorded at $A_{230-280}$ did not display clear transitions. ND = not determined. For structures of **A**, **C** and **U**, see Figure 2.2. [b] Denoting a broad transition in thermal denaturation curve.

Table 2.2. Spearman Rank Correlation with Modification Density (%mod) Results.

Variables	Correlation (r_s)	p-value
Probe Duplex	-0.306	0.423
Probe Duplex ΔT_m	-0.037	0.924
5'-**INV**:cDNA T_m	0.410	0.239
5'-**INV**:cDNA ΔT_m	**0.774**	**0.009**
3'-**INV**:cDNA T_m	-0.032	0.931
3'-**INV**:cDNA ΔT_m	**0.661**	**0.037**
T_m cDNA	-0.447	0.195
TA_{isq}	**0.662**	**0.052**
TA_{DH}	**0.662**	**0.052**
GC%	-0.412	0.237
%mod	-	-
#mod	**0.828**	**0.003**
Probe Length	-0.350	0.321
Unmodified Stretch	**-0.921**	**0.000**
Probe Duplex ΔG	0.244	0.561
Probe Duplex $\Delta\Delta G$	-0.325	0.432
5'-**INV**:cDNA ΔG	-0.063	0.862
5'-**INV**:cDNA $\Delta\Delta G$	**-0.600**	**0.067**
3'-**INV**:cDNA ΔG	0.000	1.000
3'-**INV**:cDNA $\Delta\Delta G$	**-0.663**	**0.037**
ΔG cDNA	0.269	0.452
ΔG_{rec} of DH	-0.511	0.195
ΔG_{rec} of cDNA	-0.439	0.276
C_{50} (2.5 h)	-0.619	0.056
Rec_{100X} (2.5 h)	0.573	0.084
C_{50} (15 h)	**-0.850**	**0.008**
Rec_{100X} (15 h)	**0.699**	**0.025**
d-FISH Coverage	**0.713**	**0.021**
nd-FISH Coverage	**0.738**	**0.015**

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'-**ON**:cDNA) + T_m (3'-**ON**:cDNA) - T_m (probe duplex) - T_m (dsDNA) (Table 2.1). Large positive TA_{isq} values indicate that a probe is activated for recognition of isosequential (isq) dsDNA targets, i.e., 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 that these probes have significant 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. A Spearman rank correlation analysis confirmed a positive correlation between TA_{isq} and modification density (r_s = 0.662, *p* value = 0.052) (Table 2.2).

To more fully understand the driving force for dsDNA-recognition, thermodynamic parameters associated with duplex formation were determined *via* the van't Hoff method.[34] This is achieved by baseline fitting of the thermal denaturation curves of each of the ten Invader probes. The available free energy for dsDNA-recognition can be described as $\Delta G_{rec}^{310} = \Delta G^{310}$ (5'-**ON**:cDNA) + ΔG^{310} (3'-**ON**:cDNA) – ΔG^{310} (probe duplex) – ΔG^{310} (dsDNA) (Table 2.12). Highly negative values indicate a probe with a strong thermodynamic driving force for dsDNA-recognition. In agreement with the T_m-based trends, Invader probes are prominently activated for dsDNA-recognition (ΔG_{rec}^{310} between -7 to -56 kJ/mol) which is a consequence of the low stability of the Invader probes ($\Delta\Delta G^{310}$ ranges between -1 kJ/mol and +25 kJ/mol, averaging ~12 kJ/mol) and high stability of the probe-target duplexes ($\Delta\Delta G^{310}$ ranges between +19 kJ/mol and -33 kJ/mol, averaging ~ -7 kJ/mol) (Table 2.12). The favorable driving force for dsDNA-recognition is a result of highly favorable changes in enthalpy ($\Delta H_{rec}^{310} << 0$ kJ/mol, except for **INV9**), further underscoring the role of intercalation as a driving force of dsDNA-recognition (Table 2.13).

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

The dsDNA-recognition potential of the ten *DYZ1*-targeting Invader probes was initially evaluated in an electrophoretic mobility shift assay[29] using 3'-digoxigenin (DIG)-labelled DNA hairpins (**DH**) as model targets (Figure 2.3). DNA hairpins **DH1**-**DH10** are comprised of a double-stranded stem that is complementary to the corresponding Invader probe

and 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 only denature at high temperatures, rendering them challenging dsDNA targets (T_ms for **DH1**-**DH10** between 62-82 °C, Table 2.18). 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), as appearing as lower mobility bands on non-denaturing (nd) PAGE gels.

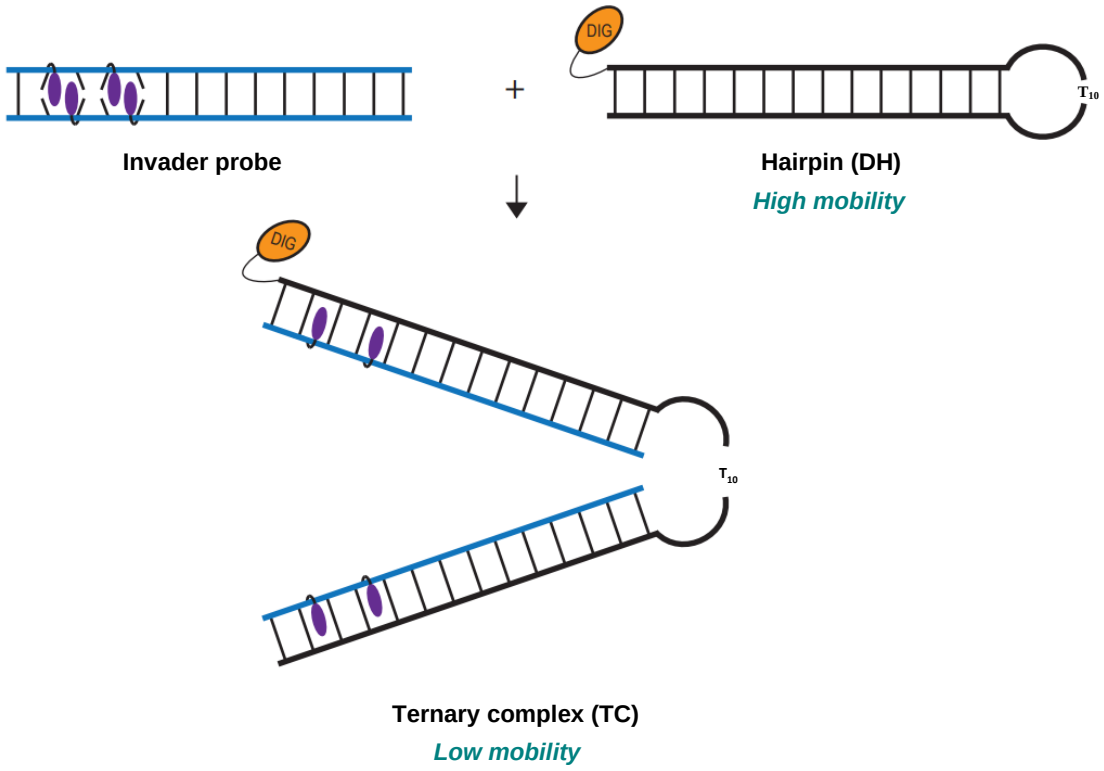


Figure 2.3. Principle of hairpin assay used to evaluate Invader probes.

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 2.4) and the degree of hairpin recognition quantified as the amount of ternary

complex formed relative to the hairpin. An incubation time of 15 hours was chosen to ensure that recognition had reached equilibrium. **INV2**, **INV7**, **INV8**, and **INV10** resulted in complete dsDNA-recognition (>95%) while **INV1**, **INV3**, and **INV9** resulted in moderate recognition (60-70%)(Figure 2.4 and Table 2.3). **INV6** and **INV5** resulted in lower levels of recognition (41% and 39% respectively), while, paradoxically, no dsDNA-recognition was observed with **INV4** in this screen. This is surprising considering its favorable binding in the previously report nd-FISH.[29] This discrepancy can be attributed to its unique sequence characteristics within chromosomal DNA that are not present in the hairpin model. This will be fully explored in the FISH experiments conducted herein

TA_{DH} values – defined as $TA_{DH} = T_m$ (5'-**ON**:cDNA) + T_m (3'-**ON**:cDNA) - T_m (probe duplex) - T_m (**DH**) – were determined to estimate how activated Invader probes are for recognition of their corresponding DNA hairpins (Table 2.1). Similar trends are observed for TA_{isq} and TA_{DH} values, i.e., higher values for more highly activated probes. However, since the DNA hairpins denature at higher temperatures than their unlinked (stem only) counterparts and thus represent a more challenging dsDNA target $TA_{DH} << TA_{isq}$.. For example, a positive TA_{isq} (+13.5 °C) is observed for **INV6**-mediated recognition of its corresponding isosequential dsDNA target, whereas a negative (unfavorable) TA_{DH} is observed for recognition of **DH6** (-3.5 °C) (Table 2.1). This indicates **INV6** that is not strongly activated for hairpin recognition, which is agreement with its moderate recognition of **DH6** (Table 2.3).

The high recognition efficiency of the quadruply modified **INV2**, **INV7**, and **INV10** probes was expected given their favorable ΔG^{310}_{rec}, TA_{Isq} and TA_{DH} values and their relatively low T_m values (Table 2.12 and Table 2.1), which is expected to facilitate probe separation. The high level of recognition observed with **INV8** is surprising considering that is less energetically

activated (TA_{DH} = +6.0), less densely modified (20%) and more GC-rich (60%) than its counterparts (Table 2.12 and Table 2.1). It is worth noting that recognition after 15 h is at least moderately correlated to ΔT_m (5'-**ON**:cDNA r_s = 0.818, *p* value = 0.004; 3'-**ON**:cDNA r_s = 0.442, *p* value = 0.200) and $\Delta\Delta G^{310}$ (3'-**ON**:cDNA r_s = 0.511, *p* value = 0.132; 5'-**ON**:cDNA r_s = -0.838, *p* value 0.002) of probe-target duplexes. High levels of recognition displayed by **INV8** could be explained by positive ΔT_m (average of +12.5 °C) (i.e. more stable probe-target duplexes and more negative $\Delta\Delta G^{310}$ (averaging -12.5 kJ/mol) (i.e. more thermodynamically favorable probe-target duplexes over the target duplex) (Table 2.1, Table 2.12). This rationale can be extrapolated to the low-to-moderate levels of recognition seen by **INV1**, **INV3**, **INV5**, **INV6** and **INV9** as these probes have less favorable formation of probe-target duplexes ($\Delta\Delta G^{310}$ values averaging at +4, -4.5, -2, -4, and -10 kJ/mol respectively) (Table 2.12). Moreover, these probes are less densely modified (~20%) and thus less energetically activated for DNA recognition (TA_{DH} = +3.5, ND, +11.5, -3.5, 0 respectively), which could rationalize the lower levels of recognition relative to the more highly modified **INV2**, **INV7**, and **INV10**. Although ΔG^{310}_{rec} could not be determined for **INV1** and **INV3**, the low levels of recognition displayed by **INV5**, **INV6**, and **INV9** might also explained by a less prominent ΔG^{310}_{rec} (between -29 kJ/mol and -19 kJ/mol) (Table 2.12 and Table 2.1).The absence of recognition by **INV4** may be attributed to low energetic activation (TA_{DH} ,+1.5 °C), and a highly GC-rich sequence (71%), which results in high probe and target T_ms (T_m = 61.5 °C and 82 °C, Tables 2.1 and 2.18). Spearman correlation analyses lend support for these rationales as the degree of recognition at 15 h is highly correlated to TA_{DH} (r_s = 0.707, *p* value = 0.033), GC-content (r_s = -0.640, *p* value = 0.046), modification density (r_s = 0.699, *p* value = 0.025), and ΔG^{310}_{rec} of **DH** (r_s = -0.872, *p* value = 0.005). Thus, the results from the initial screen suggest that probes

must be highly modified to ensure sufficient thermodynamic activation for dsDNA-recognition.

An additional recognition screen, in which a 100-fold molar excess of each Invader probe was incubated with its corresponding DNA hairpin target at 37 °C for a shorter period of time (2.5 h) revealed that dsDNA-recognition is slow and incomplete (Figure 2.23). Thus, while **INV2** and **INV10** display high levels of recognition (>80%), **INV7**-**INV9** result in intermediate levels of recognition (30-60%). **INV1** and **INV3**-**INV6** result in no or low levels of recognition (<25%) (Table 2.19). Hence, the results from this assay highlights that while some Invader probes allow for fast recognition of model DNA hairpin targets, longer incubation times are likely needed for maximal recognition.

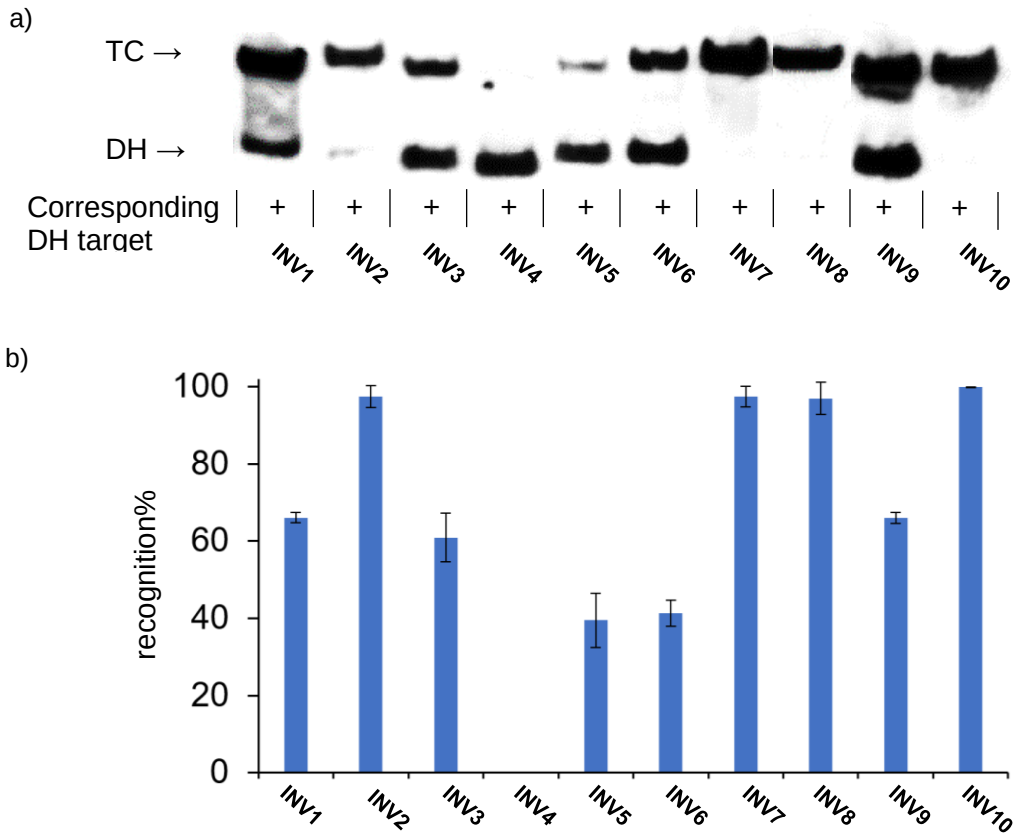


Figure 2.4. a) Representative gel electrophoretograms from recognition experiments between a 100-fold molar excess of Invader probes **INV1**-**INV10** and their corresponding DNA hairpin targets **DH1**-**DH10** following incubation for 15 h. b) Histograms depict averaged results from at least three recognition experiments with error bars representing standard deviation. TC = ternary complex. DIG-labeled **DH1**-**DH10** (34.4 nM, sequences shown in Table 2.18) 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, using incubation times of 15 h, were performed for Invader probes displaying more than 40% recognition in the preliminary screen (Figure 2.5) and used to determine C_{50} values, i.e., the probe concentrations that results in 50% recognition of a corresponding DNA hairpin target. **INV2** and **INV10** display the most efficient recognition of their DNA hairpin targets (C_{50} ~ 0.2 µM), followed by **INV7** and **INV8** (0.6-

0.7 μM; Table 2.3). Complete recognition is observed for all of these probes at high probe concentrations. **INV1**, **INV3**, **INV5**, and **INV9** display moderately efficient recognition of their hairpin targets (C_{50} values between 1.3 μM and 4.1 μM, Table 2.3), while **INV6** displays negligible levels of recognition. Spearman correlation analyses reveal that C_{50} values are strongly correlated with modification density (r_s = -0.850, p value = 0.008) and ΔG^{310}_{rec} of **DH** (r_s = 0.852, p value = 0.031). As C_{50} values follow a similar trend to the initial 100-fold screen, similar rationales related to ΔG^{310}_{rec} and modification density of each probe can be applied to explain their resulting dose response efficiency. Overall, the dose-response experiments further support the rationale that dense modification of probes is necessary for efficient dsDNA-recognition. For dose-response experiments from 2.5 h of incubation, see Table 2.19.

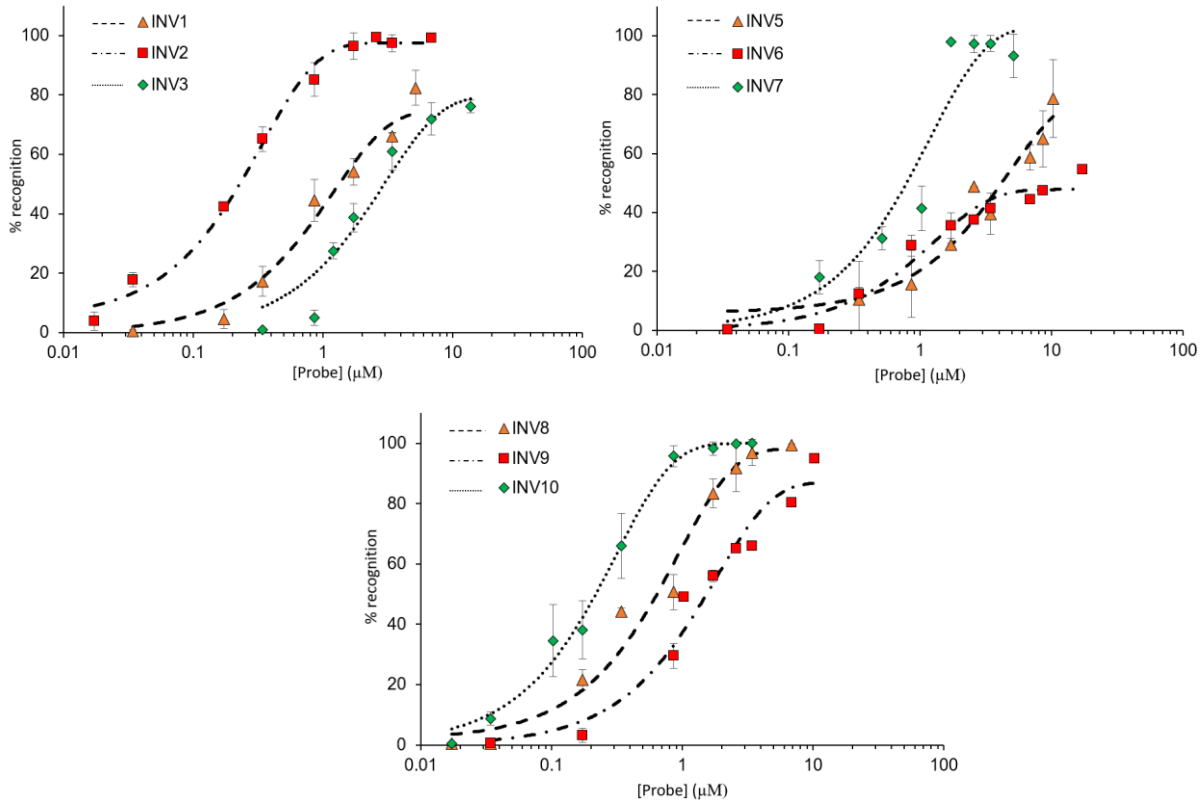


Figure 2.5. Dose-response curves for **INV1-INV3** (upper left panel), **INV5-INV7** (upper right panel), and **INV8- INV10** (lower panel) at 37 °C and 15 h incubation. Experimental conditions are as described in Figure 2.4, except for variable probe concentrations.

Table 2.3. 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	NR	<5
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_{100X} = level of DNA hairpin recognition using 100-fold molar probe excess. C_{50} values were determined from dose-response curves shown in Figure 2.5. (NR = no recognition observed).

Table 2.4. Spearman Rank Correlation with C_{50} Results.

Variables	Correlation (r_s)	p-value
Probe Duplex	0.032	0.945
Probe Duplex ΔT_m	0.010	0.983
5'-**INV**:cDNA T_m	-0.557	0.152
5'-**INV**:cDNA ΔT_m	**-0.782**	**0.022**
3'-**INV**:cDNA T_m	-0.141	0.738
3'-**INV**:cDNA ΔT_m	-0.566	0.144
T_m cDNA	0.373	0.363
TA_{isq}	-0.598	0.156
TA_{DH}	-0.598	0.156
GC%	0.323	0.435
%mod	**-0.850**	**0.008**
#mod	**-0.732**	**0.039**
Probe Length	0.282	0.499
Unmodified Stretch	**0.735**	**0.038**
Probe Duplex ΔG	-0.187	0.723
Probe Duplex $\Delta\Delta G$	0.138	0.795
5'-**INV**:cDNA ΔG	0.510	0.197
5'-**INV**:cDNA $\Delta\Delta G$	0.604	0.113
3'-**INV**:cDNA ΔG	0.131	0.757
3'-**INV**:cDNA $\Delta\Delta G$	**0.749**	**0.032**
ΔG cDNA	0.108	0.799
ΔG_{rec} of DH	**0.852**	**0.031**
ΔG_{rec} of cDNA	**0.817**	**0.047**
C_{50} (2.5 h)	**0.810**	**0.015**
Rec_{100X} (2.5 h)	**-0.835**	**0.010**
C_{50} (15 h)	-	-
Rec_{100X} (15 h)	**-0.972**	**0.000**
d-FISH Coverage	**-0.853**	**0.007**
nd-FISH Coverage	**-0.710**	**0.049**

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 to give a 100-fold excess of probe relative to the target (e.g., **INV1** was incubated with a mixture of **DH2**-**DH10**). Gratifyingly, none of the Invader probes recognized the mixture of non-target hairpins (Figure 2.6). In contrast, when individual Invader probes were incubated with a mixture of **DH1**-**DH10** at the same concentrations, i.e., a mixture containing the complementary DNA hairpin target as well as the nine non-target hairpins, recognition was

observed, indicating that the presence of non-target hairpins did not impeded with an Invader probe's ability to bind its target (Figure 2.32).

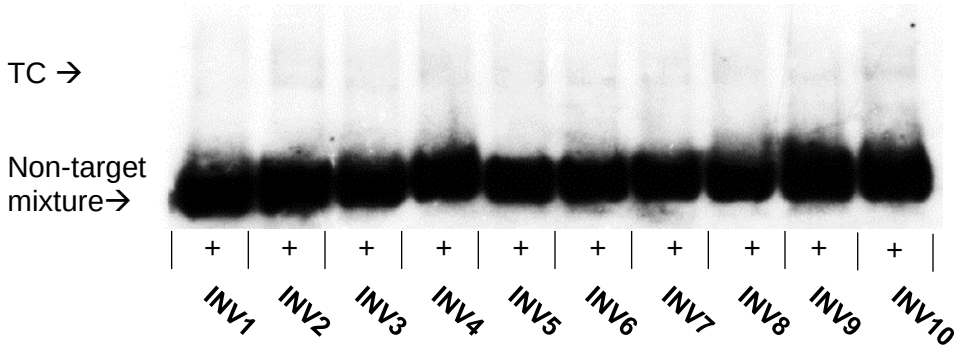


Figure 2.6. Binding specificity. Representative electrophoretograms for experiments in which the specified Invader probe (3.44 μM) was incubated at 37 °C for 2.5 h with a mixture of nine non-target DNA hairpins (each hairpin present at 34.4 nM). For example, **INV1** was incubated with **DH2**-**DH10**, while **INV10** was incubated with **DH1**-**DH9**. Conditions are otherwise as described in Figure 2.4.

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. Denaturing FISH experiments assay are expected to yield information about the maximal recognition capability of each Invader probe since access to the chromosomal DNA target regions is facilitated by the high experimental temperatures. In contrast, nd-FISH experiments are expected to reveal if a certain Invader probe can recognize its chromosomal DNA target at physiologically relevant conditions.

The d-FISH assays revealed that **INV2**, and **INV10** recognize their chromosomal DNA targets particularly efficiently as single localized high quality Cy3-signals – consistent with

DYZ1-recognition - are observed in ~90% of the analyzed nuclei with only a minimal level of background signal (Figure 2.7 and Figure 2.8 – left column; Table 2.5). Table 2.5 reflect percent ranges to better describe the variation of number of specifically labeled nuclei between different trials (consisting of ~100 nuclei) conducted on different microscope slides. 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. Paradoxically, excellent signal is observed with **INV4**, which did not result in detectable binding in the DNA hairpin assay and only was weakly activated for dsDNA-recognition. Other weakly activated probes, **INV3** and **INV6**-**INV9** display clear signals in 30-70% of the analyzed nuclei at denaturing conditions (Figures 2.35, 2.36, 2.37 – left column; Table 2.5), a trend that reflects the more moderate recognition levels seen for these probes in the hairpin assay. Conversely, **INV1** results in the formation of multiple signal blotches, indicative of non-specific binding (Figure 2.35). **INV5**, which was predicted to be a weak binder from the DNA hairpin assay, did not result in signals of any kind. These results indicate that most of the designed Invader probes allow for recognition of chromosomal DNA target under denaturing conditions and that the hairpin assay has good, albeit not perfect, predictive power in identifying efficient probes. This rationale is supported by a Spearman rank correlation analysis, showing that d-FISH signal coverage is highly correlated to C_{50} values derived from experiments conducted at 15 h (r_s = -0.853, p value = 0.007) (Table 2.4).

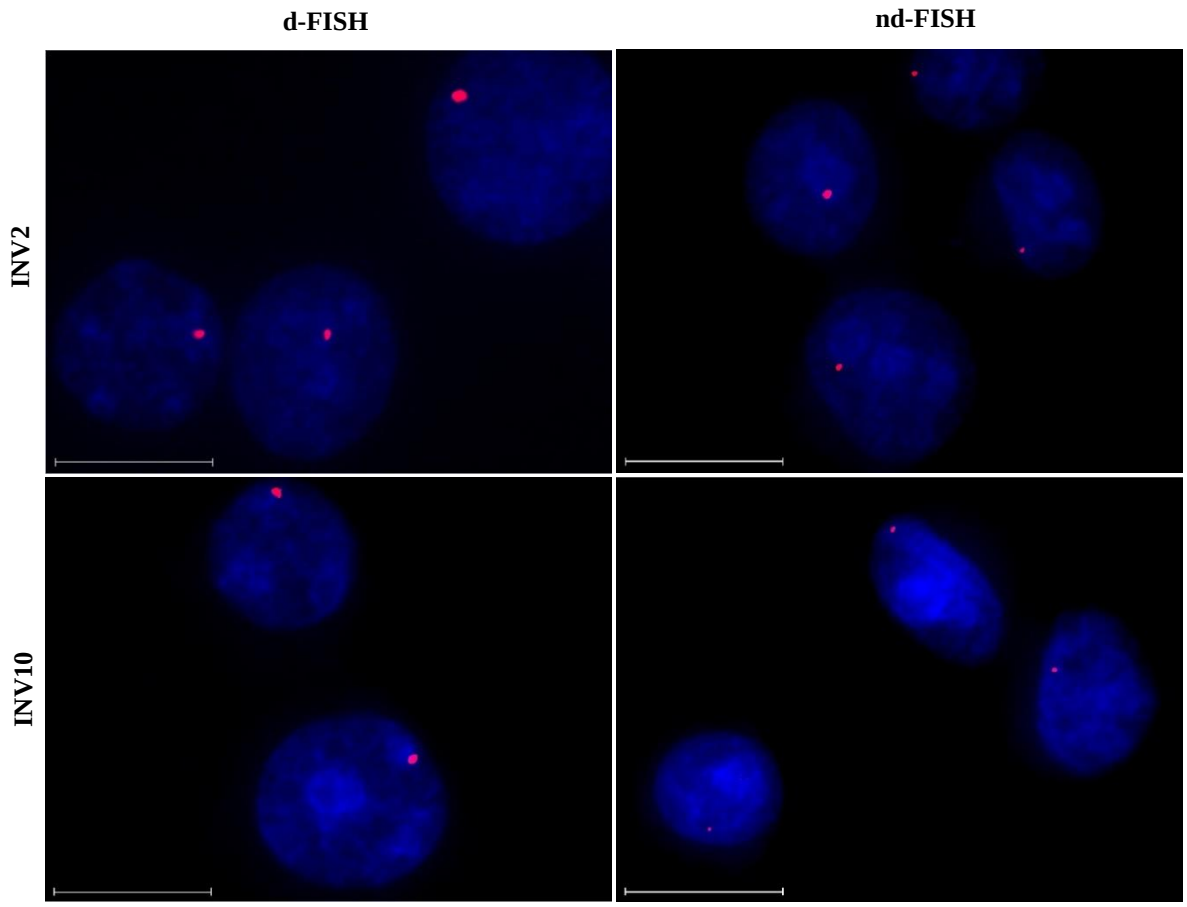


Figure 2.7. Images from FISH experiments using *DYZ1*-targeting Invader probes **INV2** and **INV10** under denaturing (5 min, 80 °C) (left), or non-denaturing (3 h, 37.5 °C) conditions (right). Images are representative of the signal intensity and background, and the size and morphology of all analyzed nuclei (~200 nuclei per probe). Fixed isolated nuclei from male bovine kidney cells were incubated with probes in a Tris buffer (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 2.34, 2.35, 2.36, and 2.37.

Gratifyingly, most of the Invader probes also enabled recognition of their mixed-sequence chromosomal DNA targets under non-denaturing conditions. Thus, **INV2**, **INV4**, and **INV10** yielded intense signals without noticeable background in 80-90% of nuclei (Figures 2.7, 2.8) and Table 2.5). Moderately intense signals were observed for **INV3** and **INV6**-**INV8** (Figure 2.35, 2.36), albeit in a lower percentage of nuclei than in the d-FISH assay (20-40%,

Table 2.5). Interestingly, **INV9**, which yielded distinct intense signals in 50-70% of nuclei in d-FISH assays, did not produce discernable signals in nd-FISH assays (Table 2.5). The lack of signal indicates that the target region is not accessible under non-denaturing conditions, suggesting that, while there are sufficient thermodynamics to bind in d-FISH, the activation energy for recognition is too high under nd-FISH conditions. The moderate level of signal coverage observed with **INV7** is surprising as well considering it is a highly modified probe (25%) that displays favorable binding in the hairpin assay (Table 2.3 and Table 2.5). Recognition efficiency under nd-FISH is correlated to modification density and factors correlated to modification density may indicate why **INV7,** despite being highly modified, does not result in high levels of detection under nd-FISH conditions. $\Delta\Delta G^{310}$ values are high-to-moderately correlated to modification ($\Delta\Delta G^{310}$ 3'-**ON**:cDNA r_s = -0.663, p value = 0.037, $\Delta\Delta G^{310}$ 5'-**ON**:cDNA r_s = -0.600, p value = 0.067) (Table 2.2). **INV7** has only moderately favorable $\Delta\Delta G^{310}$ values in comparison to other highly modified, efficient Invader probes such as **INV10** (-17 vs -27 kJ/mol average of both probe-target duplexes for each respective probe). This suggest that the high levels of modification of **INV7** is not resulting in a great enough energy difference between target duplex and recognition complex to facilitate binding at its chromosomal targeting region. Another surprising result was the low signal coverage by **INV8** is surprising considering its low C_{50} value (a factor which is statistically correlated to nd-FISH results). Low signal coverage could be rationalized similar to **INV7.** Due to the low modification density of **INV8** (20%) which results in lower average $\Delta\Delta G^{310}$ of +15.5 kJ/mol (a conclusion that is supported by the correlation between percent modification and $\Delta\Delta G^{310}$) (Table 2.2), the formation of recognition complex over target duplex is not favorable enough to give high signaling of chromosomal DNA (Table 2.12). **INV1** and **INV5** also exhibit less

than favorable ΔT_m and $\Delta\Delta G^{310}$, supporting the idea that the resulting stability and favorability of the recognition complex is not enough to drive to invasion of chromosomal DNA (Table 2.1, 2.12). 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.[35]

As discussed above, **INV2** and **INV10** are strongly activated for dsDNA-recognition (Table 2.1), thus providing a compelling rationale for their performance in the nd-FISH assay. The signaling efficiency of **INV4** is perplexing given that the probe is much less strongly activated (TA_{isq} = +23.0 °C, ΔG^{310}_{rec} = -29 kJ/mol, TA_{DH} = +1.5 °C, Table 2.1) than some other Invader probes (which display poor signaling characteristics), has a high GC-content (~71%), and does not result in recognition of its corresponding DNA hairpin (Figure 2.4). A unique feature of **INV4** - and, therefore, its corresponding target region, is that it features two GGG/CCC-tracts. We speculate that this may render the target region uniquely accessible due formation of non-canonical secondary structures.[36] An alternative explanation for the surprising signaling characteristics of **INV4** is that the corresponding target region is present six times within a single 1175 bp long *DYZ-1* repeat, which in turn is repeated ~6 x 10^4 times (Figure 2.18).[29] In contrast, the target regions of the other probes are only present once per repeating unit. The higher number of target sites may yield stronger signals and higher proportion of nuclei presenting signal.

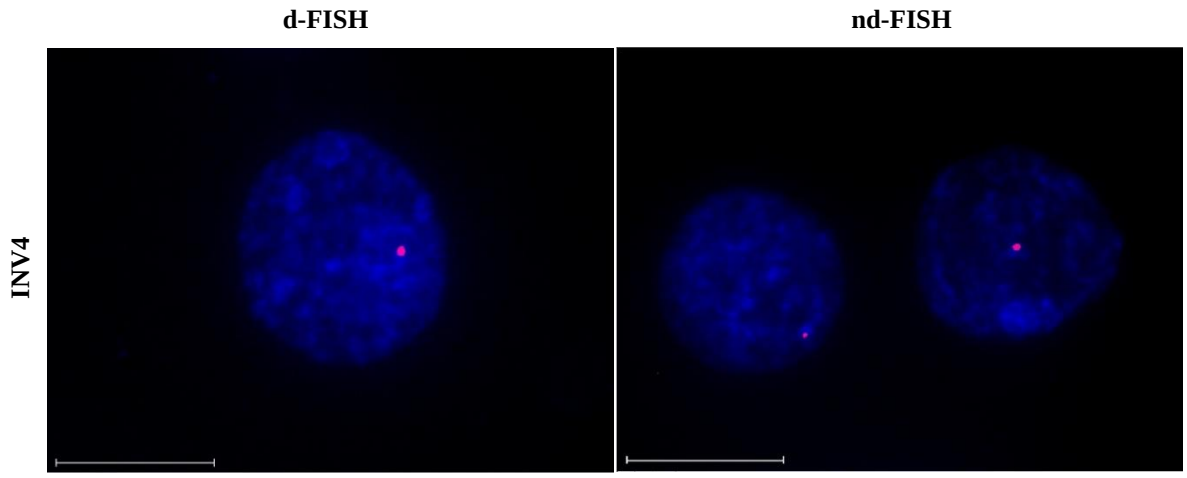


Figure 2.8. Images from FISH experiments using *DYZ1*-targeting Invader probes **INV4** under denaturing (5 min, 80 °C) (left), or non-denaturing (3 h, 37.5 °C) conditions (right). Incubation and imaging specifications are described in Figure 2.7. Scale bar represents 16 μm.

Table 2.5. Percentages of nuclei presenting clear signal (signal coverage) in d-FISH and nd-FISH assays when incubated with various *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%

Control experiments in which nuclei were pre-treated with either DNase I, RNase A, or Proteinase K prior to incubation with **INV2** or **INV10**, confirmed that Invader probes target chromosomal DNA, rather than binding to RNA or proteins. Thus, nuclei pre-treated with DNase I were completely devoid of signal (Figure 2.38), whereas pre-treatment with RNase A or Proteinase K still yielded signals with both **INV2** and **INV4**, albeit signals were less

pronounced and fainter (Figure 2.39), which likely is attributed to nuclei/chromosome loss during RNase A and Proteinase K treatment.

Moreover, incubation of the Y-chromosome-targeting **INV2** and **INV10** with a female bovine endothelial cell line failed to produce signals under denaturing condition (used to eliminate the barrier for recognition) failed to produce signals (Figure 2.9), further supporting the specificity of these Y-chromosome-targeting Invader probes. Previously reported **INV4** yielded similar results in these control experiments, further supporting that Invader probes are specifically targeting Y chromosome DNA.[28]

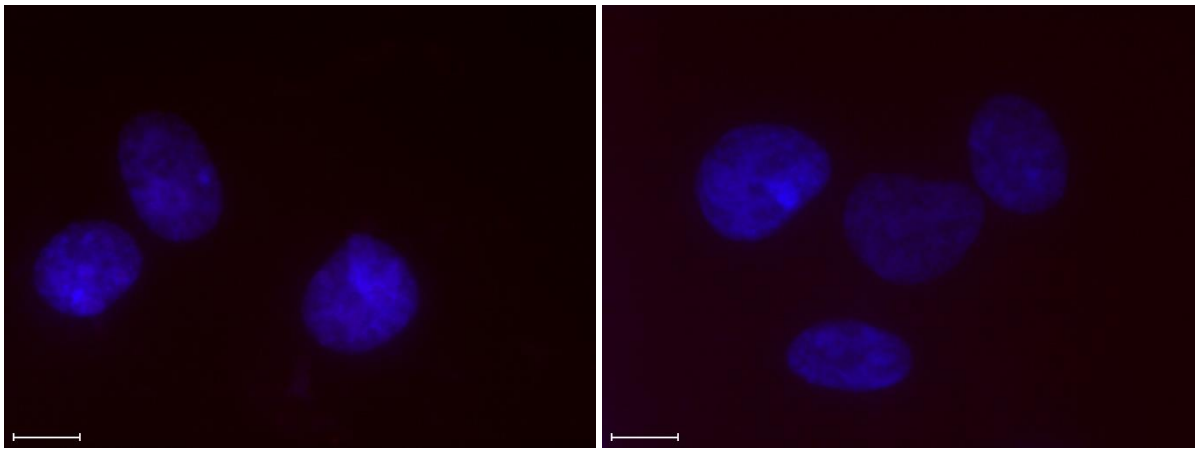


Figure 2.9. INV2 (left panel) and **INV10** (right panel) incubated with isolate female bovine endothelial nuclei under denaturing conditions. Note the absence of Cy3 signal. Incubation and imaging specifications are described in Figure 2.7. Scale bar represents 16 µm.

Spearman rank correlation analysis: rules of probe design

To identify the key parameters ensuring efficient recognition of chromosomal DNA in the nd-FISH assays, a Spearman rank correlation analysis was performed for a wide range of parameters including .. T_ms and ΔT_ms of the various duplexes (probe duplex T_m, probe duplex ΔT_m, 5'-**ON**:cDNA T_m, 5'-**ON**:cDNA ΔT_m, 3'-**ON**:cDNA T_m, 3'-**ON**:cDNA ΔT_m, cDNA T_m) estimates of probe activation (TA_{isq}, TA_{DH}, ΔG^{310}_{rec} of **DH**, ΔG^{310}_{rec} of cDNA, probe design features (GC%, %mod, # of modifications [#mod], probe length, longest stretch of nucleotides

between each hotspot [unmodified stretch]), thermodynamic parameters of the various duplexes (probe duplex Δ*G*, probe duplex ΔΔ*G*, 5'-**ON**:cDNA Δ*G*, 5'-**ON**:cDNA ΔΔ*G*, 3'-**ON**:cDNA Δ*G*, 3'-**ON**:cDNA ΔΔ*G*, cDNA Δ*G*), hairpin assay results (C_{50} and Rec_{100X} at 2.5 and 15 h), and d- and nd-FISH assay performance (percent of nuclei presenting signal) (Table 2.6). The results associated with each Invader probe were ranked from 1 to 10 and correlation between nd-FISH rank and ranking of all other parameters were determined. Correlations were considered significance with a *p* value < 0.05.

The strongest correlations to nd-FISH performance is with modification density (%mod) (r_s = 0.738, *p* value = 0.015), C_{50} values at 15 h (r_s = -0.710, *p* value = 0.049), and d-FISH performance (r_s = 0.789, *p* value = 0.006) (Table 2.6). First off, this suggests that the DNA hairpin assay has predictive power, although it may result in false negatives as with **INV4**. Secondly, the Spearman correlation analysis establishes that more densely modified probes are more likely to result in strong nd-FISH signals. Factors that correlate with percent modification reveal that an increase in Invader monomer incorporation increases chromosomal DNA recognition by forming a more stable and thermodynamically favorable recognition complex. This is supported by correlations between percent modification and ΔT_m of both probe-target duplexes, TA_{isq}, TA_{DH}, and ΔΔ*G* of probe-target duplexes. Along these lines, it is interesting to note that nd-FISH performance also correlates weakly with parameters estimating the level of probe activation, i.e., TA_{isq} (r_s = 0.505, *p* value 0.165), TA_{DH} (r_s = 0.505, *p* value = 0.165), and ΔG^{310}_{rec} of cDNA (r_s = -0.583, *p* value = 0.129) (Table 2.6). It is equally interesting to note the parameters that nd-FISH performance *does not* correlate with characteristics such as GC-content and probe length (Table 2.6). This suggests that (at least within the ranges tested in this work) stability derived for increase hydrogen bond interactions

by GC base pairs and increased number of base pairs (length) has little to no effect on the Invader probes' ability to target chromosomal DNA.

Table 2.6. Spearman Rank Correlation with nd-FISH Signal 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'-**ON**:cDNA T_m	0.413	0.236
5'-**ON**:cDNA ΔT_m	0.268	0.454
3'-**ON**:cDNA T_m	0.081	0.823
3'-**ON**: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'-**ON**:cDNA ΔG	0.094	0.797
5'-**ON**:cDNA $\Delta\Delta G$	-0.028	0.939
3'-**ON**:cDNA ΔG	-0.268	0.454
3'-**ON**: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

Having established that the modification density is the key critical factor for detection of chromosomal DNA under nd-FISH conditions, we set out experimentally verify the predictive power of this insight by optimizing the recognition characteristics of three Invader probes displaying moderate performance in nd-FISH assays, i.e., **INV6**, **INV8**, and **INV9**. These probes were selected for optimization since they are only moderately modified (~20%,

Table 2.1), offering opportunities for introduction of additional energetic hotspots using adenin-9-yl, cytosin-1-yl, and uracil-1-yl monomers, which are preferred as they are known to result in particularly favorable gains in thermodynamic driving force (which correlates weakly with nd-FISH performance)[30] and show moderate levels of recognition under denaturing and non-denaturing conditions, indicating that the corresponding chromosomal DNA targets are accessible for binding. Thus, 2-3 additional hotspots were introduced to yield probes to yield **OPT6**, **OPT8**, and **OPT9**, which have a modification density of 26.6-33.3% (Table 2.7).

Table 2.7. Thermal denaturation temperatures (T_ms), Thermal advantages (TA_{isq} and TA_{DH}), percent modification (Mod%), and 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'-ON: cDNA	3'-ON: 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[b] [-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[b] [-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_{isq} and TA_{DH}. NT = no clear transition (very broad) at wavelengths 230-280 nm. For structures of **A**, **C** and **U**, see Figure 2.2. [b] Denoting a broad transition in thermal denaturation curve

Thermal denaturation and thermodynamic properties of optimized Invader probes

Thermal denaturation temperatures were determined for the optimized probes and the corresponding duplexes between individual probe strands and cDNA (Table 2.7). In line with

our expectations, the optimized probes are significantly less stable (T_ms ~20 °C lower on average) than their less modified counterparts (compare T_ms for **OPT6/8/9** and **INV6/8/9**, Tables 2.7 and 2.1). Furthermore, the individual optimized probe strand form more stable duplexes with cDNA relative to their less extensively modified counterparts (T_ms ~5 °C higher on average). As a consequence, the driving force for dsDNA-recognition is markedly higher, with the three optimized probes displaying TA_{isq} values between 43.0-55.5 °C (average increase of ~31 °C), ΔG^{310}_{rec} values between -93 kJ/mol and -59 kJ/mol (average decrease of ~48 kJ/mol), and TA_{DH} values between 26-35 °C (average increase of ~31 °C) (Tables 2.7 and 2.15). Hence, the very favorable estimated binding energetics are expected to result in improved chromosomal DNA binding.

Recognition of mixed-sequence model DNA hairpin targets by optimized Invader probes

The optimized Invader probes were evaluated for dsDNA-recognition using the DNA hairpin assay that was used in the initial studies (Figure 2.3). The probes were first screened at 100-fold molar excess (Figure 2.10), and binding potential more fully evaluated in dose-response experiments (Figure 2.11). Gratifyingly, **OPT8** and **OPT9** resulted in near-complete recognition of their corresponding hairpin targets when incubated at 100-fold molar excess for 15 h, which represents a vast improvement over the unoptimized probes (compare Rec_{100x} values for **OPT8/9** and **INV8/9**, Tables 2.8 and 2.3). Conversely, **OPT6** did not result in improved recognition of **DH6** relative to **INV6**, indicating that target accessibility rather than binding thermodynamics, is the limiting factor. The dose-response experiments confirmed these trends as **OPT8** and **OPT9** display 3- and 5-fold reduction in their C_{50} values relative to their unoptimized counterparts **INV8** and **INV9** (Table 2.3 and Table 2.8). Additional dose response experiments for Optimized probes were conducted at 2.5 h and followed similar

trends to the 15 h study (Figure 2.28, Table 2.20). C_{50} between the 2.5 h and 15 h experiments were maintained for **OPT6** and **OPT9**, signifying that full recognition was reached by 2.5 h. The C_{50} value for **OPT8** decrease by 3.0 µM, highlighting the need for increase incubation times to ensure equilibrium is reached.

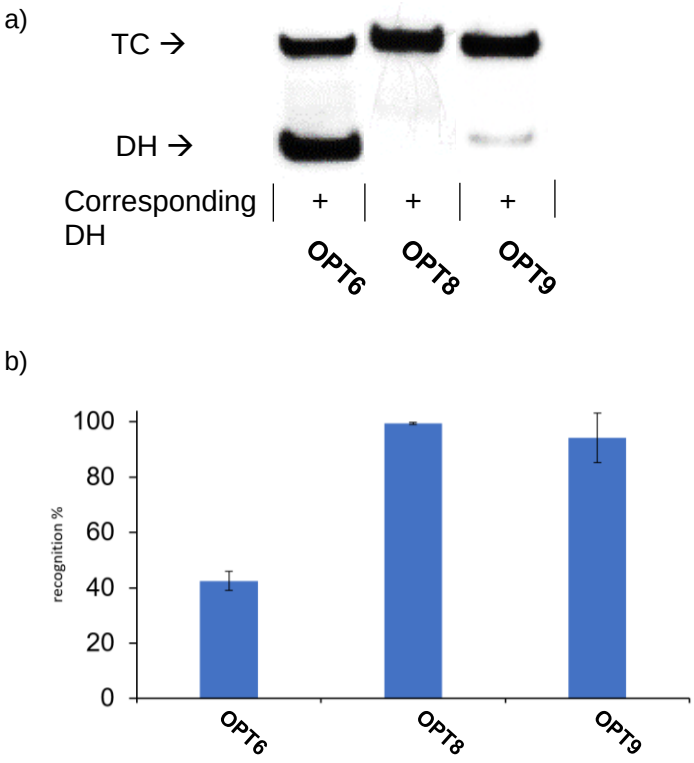


Figure 2.10. a) Representative gel electrophoretograms from recognition experiments between a 100-fold molar excess of optimized Invader probes **OPT6**, **OPT8**, and **OPT9** and their corresponding DNA hairpin targets **DH6**, **DH8**, and **DH9** following incubation for 15h. b) Histograms depict 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 2.18) 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).

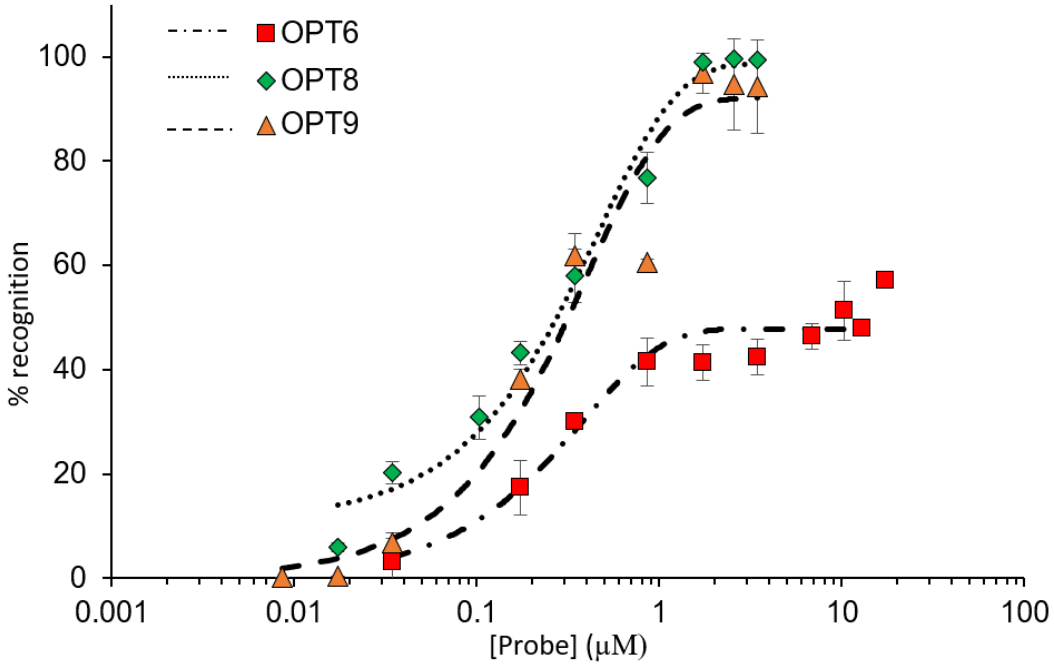


Figure 2.11. Dose-response curves for **OPT6, OPT8**, and **OPT9** at 37 °C and 15 h incubation. Experimental conditions are as described in Figure 2.10, except for variable probe concentrations.

Table 2.8. C_{50} values at 15 h for Invader probes studied herein.[a]

Probes	C_{50} (µM)	Rec_{100X} (%)
OPT6	>10	42 ± 3.4
OPT8	0.2	99 ± 0.4
OPT9	0.3	99 ± 8.9

[a] Rec_{100X} = level of DNA hairpin recognition using 100-fold molar probe excess. C_{50} values were determined from dose-response curves shown in Figure 2.11.

Detection of chromosomal DNA by Optimized Invader probes

With the DNA hairpin assays confirming their binding potential, the optimized probes were evaluated in the above described FISH assay under denaturing and non-denaturing conditions (Figure 2.12). Based on the increased driving force for dsDNA-recognition (represented in higher TA_{isq} and TA_{DH} values - derived from increased probe instability and probe-target stability - and favorable ΔG^{310}_{rec}) and improved C_{50} values, **OPT8** behaved as expected, resulting in a marked increase of nuclei presenting clear signals both under d-FISH

and nd-FISH conditions vis-à-vis **INV8** (Table 2.9). Somewhat surprisingly, **OPT6** resulted in the highest proportion of nuclei presenting signals out of the three optimized probes, under both denaturing or non-denaturing conditions (~90%, Table 2.9), despite displaying the least efficient recognition of its DNA hairpin target (Table 2.8). While a substantial increase in nuclei presenting signals was observed for **OPT9** relative to **INV9**, it did not reach the signaling performance of **OPT6** and **OPT9** (Table 2.9), despite displaying a low C_{50} value (Table 2.8). These results indicate that accessibility of the target region likely also is a key contributing factor in determining how well an Invader probe recognizes chromosomal DNA targets. Thus, while this study has elucidated how to chemically optimize Invader probes for recognition of complementary dsDNA targets (mainly by ensuring a high degree of modification of pyrimidine-based cytosin-1-yl, and uracil-1-yl RNA monomers), it will be necessary to screen a number of chemically optimized Invader probes in biological assays to identify accessible target sites. It is encouraging that out of six well-designed Invader probes (i.e., ≥25% modified), four resulted in efficient performance in nd-FISH assays, presumably as the corresponding targets of **INV7** and **OPT9** are difficult to access. Only one scarcely modified probe (**INV4**) resulted in efficient recognition of its target, presumably because it is highly abundant and/or easily accessible. Overall the results underscore that a more densely modified Invader probes is correlated with improved chromosomal dsDNA-recognition.

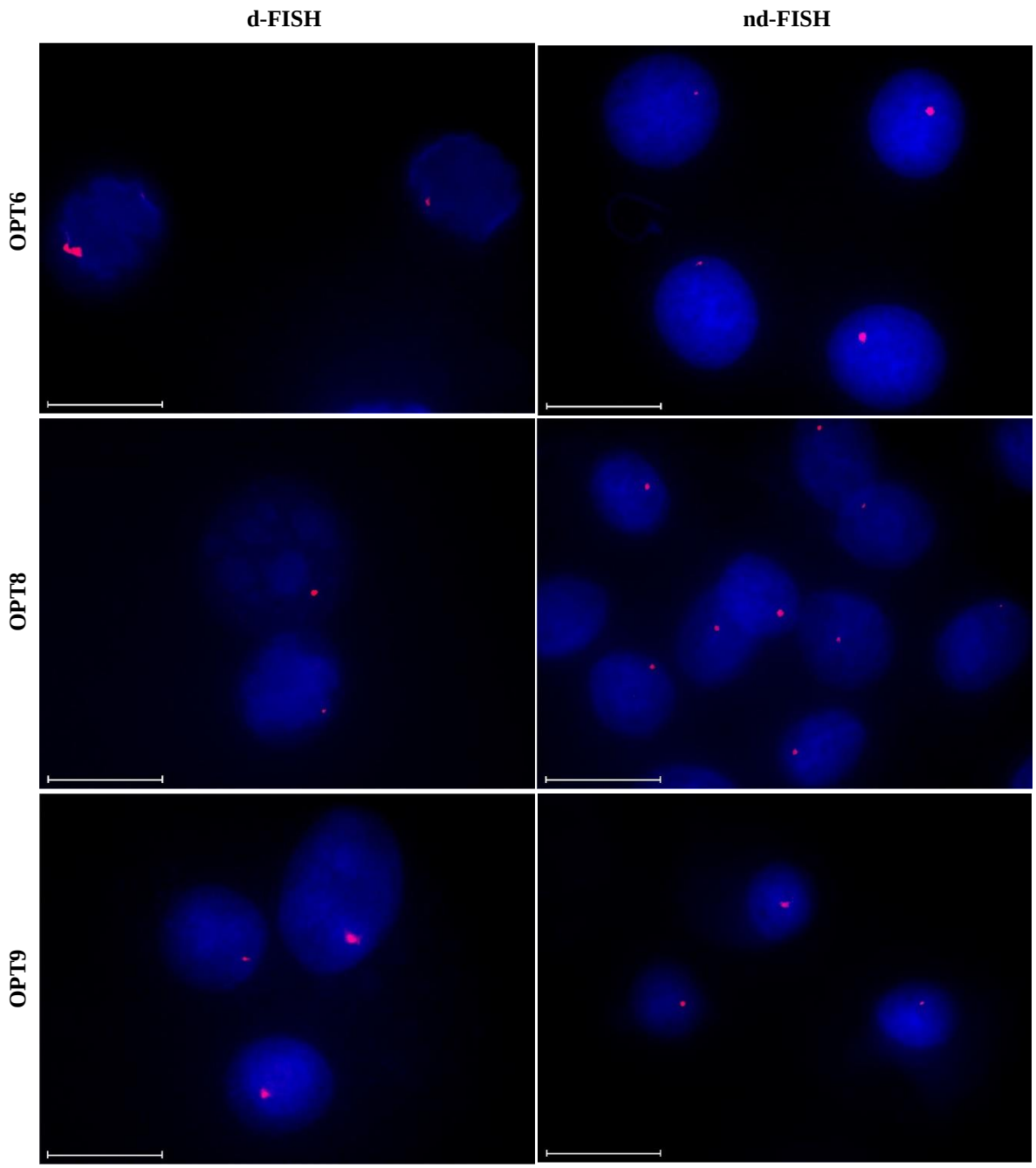


Figure 2.12. Images from FISH experiments using *DYZ1*-targeting **OPT6**, **OPT8**, and **OPT9** under denaturing (5 min, 80 °C) (left), or non-denaturing (3 h, 37.5 °C) conditions (right). Incubation and imaging specifications are described in Figure 2.7. Scale bar represents 16 µm.

Table 2.9. Percentages of nuclei presenting clear signal (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 Conclusion

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 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% signal 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 coverage was quadrupled in the optimization of **INV6** to **OPT6** under non-denaturing conditions. While

these results support 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 signal coverage while **OPT6** should have given relatively less signal 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,[35] whether it be for structural (i.e. histones) or functional (i.e. polymerases) purposes, which will affect the accessibility of target sequence.[37] 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_{isq}, TA_{DH}, 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[38] and **C**/**A** monomers[27] and incorporated into ONs via extended hand-couplings (15 min, ~45-fold molar excess at a concentration of 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 CH_3CN). 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 2.17) and MALDI-MS analysis (Tables 2.10 and 2.11, and Figures 2.13, 2.14, 2.15, 2.16) 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)[39] and Cy3 (4.93)[40].

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 non-denaturing (nd)-PAGE assay was performed essentially as previously described.[28] Thus, 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. Solutions of Invader probes (concentrations as specified) were 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 then 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.[41]

Protocol - cell culture and nuclei preparation

Male bovine kidney cells (MDBK, ATCC: CCL-22, Bethesda, MD) were maintained in DMEM with GlutaMax (Gibco, 10569-010) and 10% fetal bovine serum (Invitrogen). Female bovine endothelial cells (CPAE, ATCC: CCL-209) were maintained in Eagle's Minimum Essential Medium (ATTC, 30-2003) and 20% fetal bovine serum (Invitrogen) The

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, KaryoMax colcemid (Gibco, 15210-040) (65 μL per 5 mL of growth media) was added and the 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 incubated with a hypotonic KCl solution (5-8 mL, 75 mM, 20 min), followed by addition of fixative (10 drops, MeOH:AcOH, 3:1) to this solution, 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 with fixative solution steps 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 nuclei suspension was warmed to room temperature and resuspended in fresh fixative solution. Glass microscope slides 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 experiments

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 (termed 1x solution in Table 2.21) was placed on each slide. Preliminary assay optimization studies (for more information, see Table 2.21) revealed that a 1x concentration (~30 ng of probe per 200 μL of incubation buffer) afforded the best qualitative signal-to-background ratio for all Invader probes under denaturing and non-denaturing conditions, except for **INV4**, which was used at ¼ the concentration of the other Invader probes to reduce background levels. When used 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. When used 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). Slides were subsequently 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 a SOLA SMII LED light source system and Cy3 and DAPI filter sets, was used to visualize nuclei at 60x magnification in order to capture many nuclei in one image. Images of fluorescently labeled nuclei were captured using a 14-bit CoolSNAP HQ2 cooled CCD camera and processed with the NIS-Elements BR 4.20 software.

Protocol – image analysis

In experiments aiming at optimizing signaling output in nd-FISH assays, the concentration of the Invader probe was varied from 1x to 4x. The percentage of nuclei

presenting representative signals (i.e., signal coverage) was estimated by evaluating >100 nuclei per Invader probe, for each d-FISH and nd-FISH assays. The quality of signal and level of background noise was assessed on a qualitative scale between 1-3 across ~200 nuclei (Table 2.21). 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 is 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 (e.g., multiple scattered blotches) and/or distributed non-specific signal masking punctate signal (e.g., 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 2.34 2.35, 2.36, and 2.37 reflect the "dFISH" and "[1x]" rows of Table 2.21.

Control experiments in which fixed nucleic from MDBK cells were pre-treated with DNase, RNase, or proteinase were carried out as follows. DNase pre-treatment: 3 µL of loned DNase I (RNase Free) (Takara N101 JF) was mixed in 50 µL 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 µL amounts. The slides were incubated with the DNase I solution at 37.5 °C for 20 minutes and then rinsed with TE Buffer. RNase pre-treatment: 1 µL of RNase A (5 mg/mL, Fisher reagents BP2539-100,) in 100 µL of buffer (10 mM Tris-HCl, pH 6.5) was placed in 50 µL amounts 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 (6.25 µg/mL, Fisher BioReagents, BP1700-100) was added to 200 µL of buffer (10 mM Tris-HCl, pH 7.5).

Fixed nuclei were incubated with 50 µL of this solution for 10 min at 37.5 °C and then rinsed with TE buffer.

Protocol – Spearman Rank correlation analyses.

Spearman Rank Correlation coefficients were calculated with the XRealStat function add-on for Microsoft ® Excel ®. Invader Probes were ranked 1 to 10 for each studied parameter and these rankings were compared to determine correlation between parameter pairs. For example, high C_{50} values, high Rec_{100X} (%), and more negative ΔG_{rec}^{310} would rank closer to 1 while low C_{50} values, low Rec_{100X} (%), and relatively more positive ΔG_{rec}^{310} values would rank closer to 10. It was deemed that correlations with *p* values less than the α-value of 0.05 were statistically significant.

Table 2.10. 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'	5785	5783
INV2u	5'-Cy3-A**U**AC**U**GGTTTG**U**G**U**TC-3'	6266	6264
INV2d	3'- TA**U**GA**C**CAAACA**C**A**A**G-Cy3-5'	6282	6279
INV3u	5'-Cy3-T**U**G**U**GCCCTGGC**A**AC-3'	5714	5712
INV3d	3'- AA**C**A**C**GGGACCGT**U**G-Cy3-5'	5786	5783
INV4u	5′-Cy3-**A**GCCC**U**GTGCC**C**TG-3'	5400[a]	5398
INV4d	3′- T**C**GGGA**C**ACGGG**A**C-Cy3-5'	5511[a]	5510
INV5u	5'-Cy3-G**A**TTTCAGCCA**U**G**U**GC-3'	6043	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'	6064	6057
INV6d	3'- GA**C**A**C**GTTGACCAA**A**C-Cy3-5'	6057	6055
INV7u	5'-Cy3-C**U**G**U**GCAA**U**ATTT**U**GT-3'	6249	6247
INV7d	3'- GA**C**A**C**GTTA**T**AAAA**C**A-Cy3-5'	6296	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'	5828	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'	5789	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'	5562	5561

[a]MALDI-MS has been previously reported in reference 29.

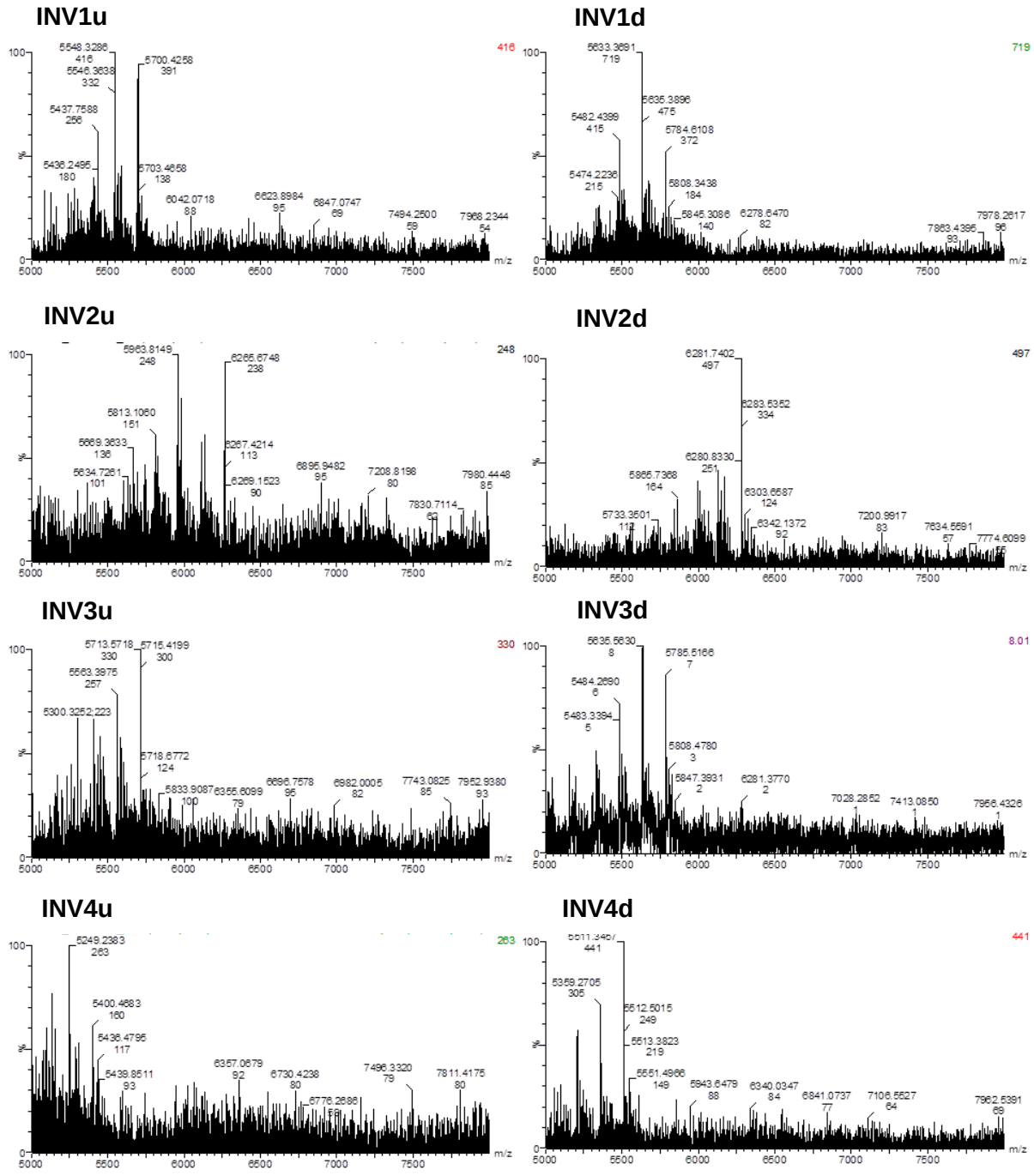


Figure 2.13. MALDI-MS Spectra of individual strands (up and down) of Invader probes **INV1**-**INV4**.

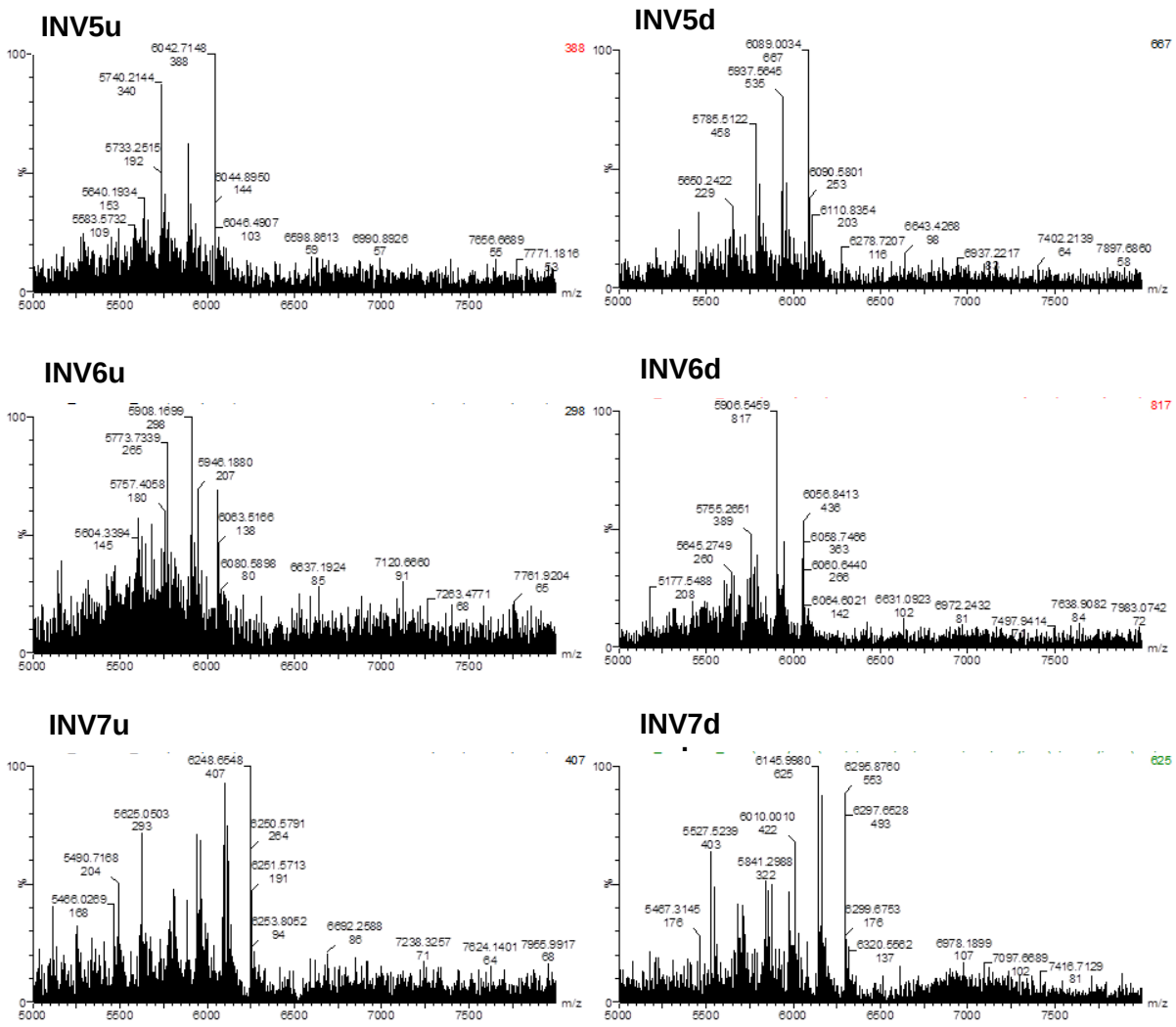


Figure 2.14. MALDI-MS Spectra of Individual Strands (up and down) of Invader Probes **INV5-INV7**.

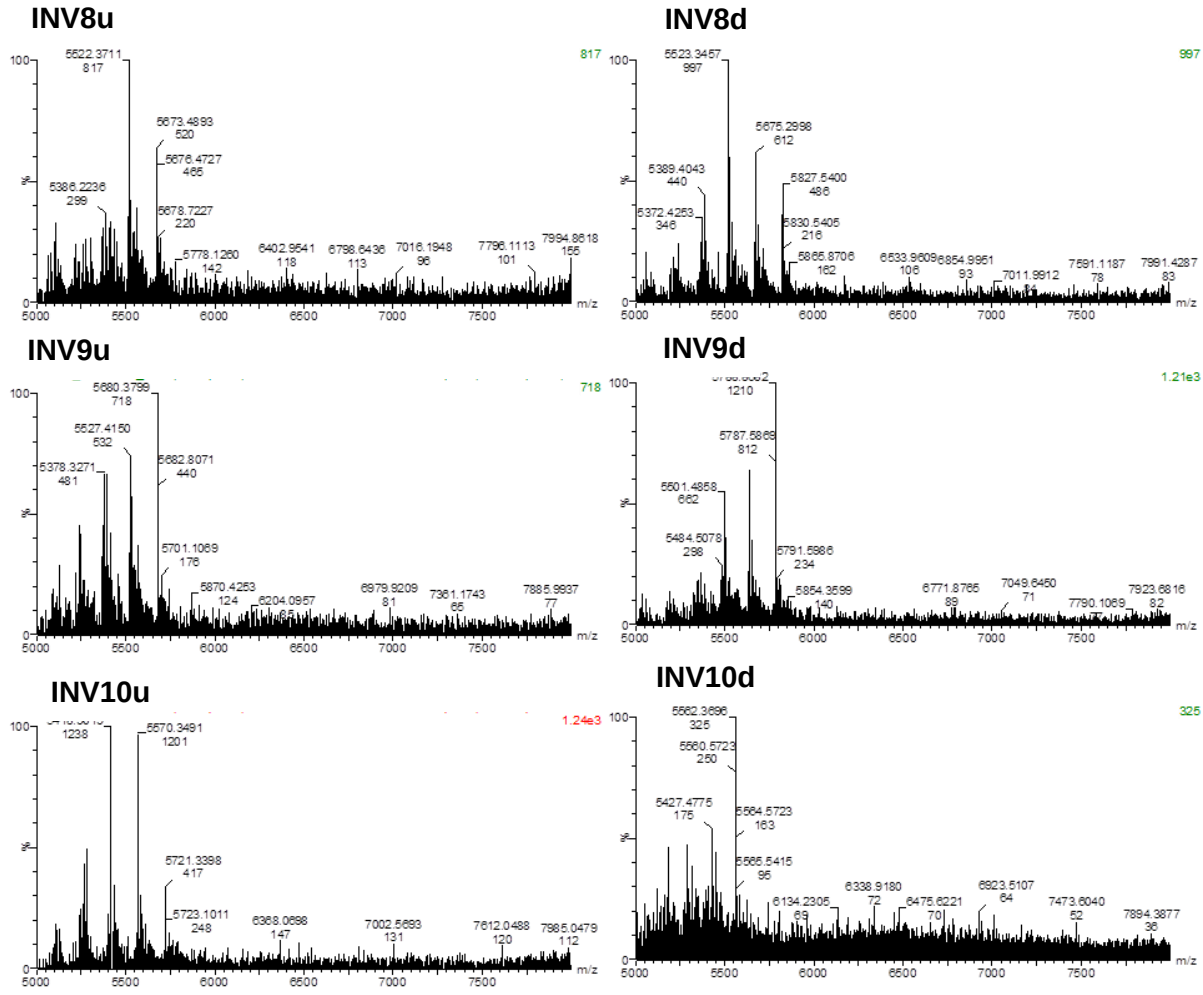


Figure 2.15. MALDI-MS Spectra of Individual Strands (up and down) of Invader Probes **INV8**-**INV10**.

Table 2.11. 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'	6114	6112
OPT9d	3'- AA**U**A**U**A**C**GA**C**A**A**GAG-Cy3-5'	6248	6247

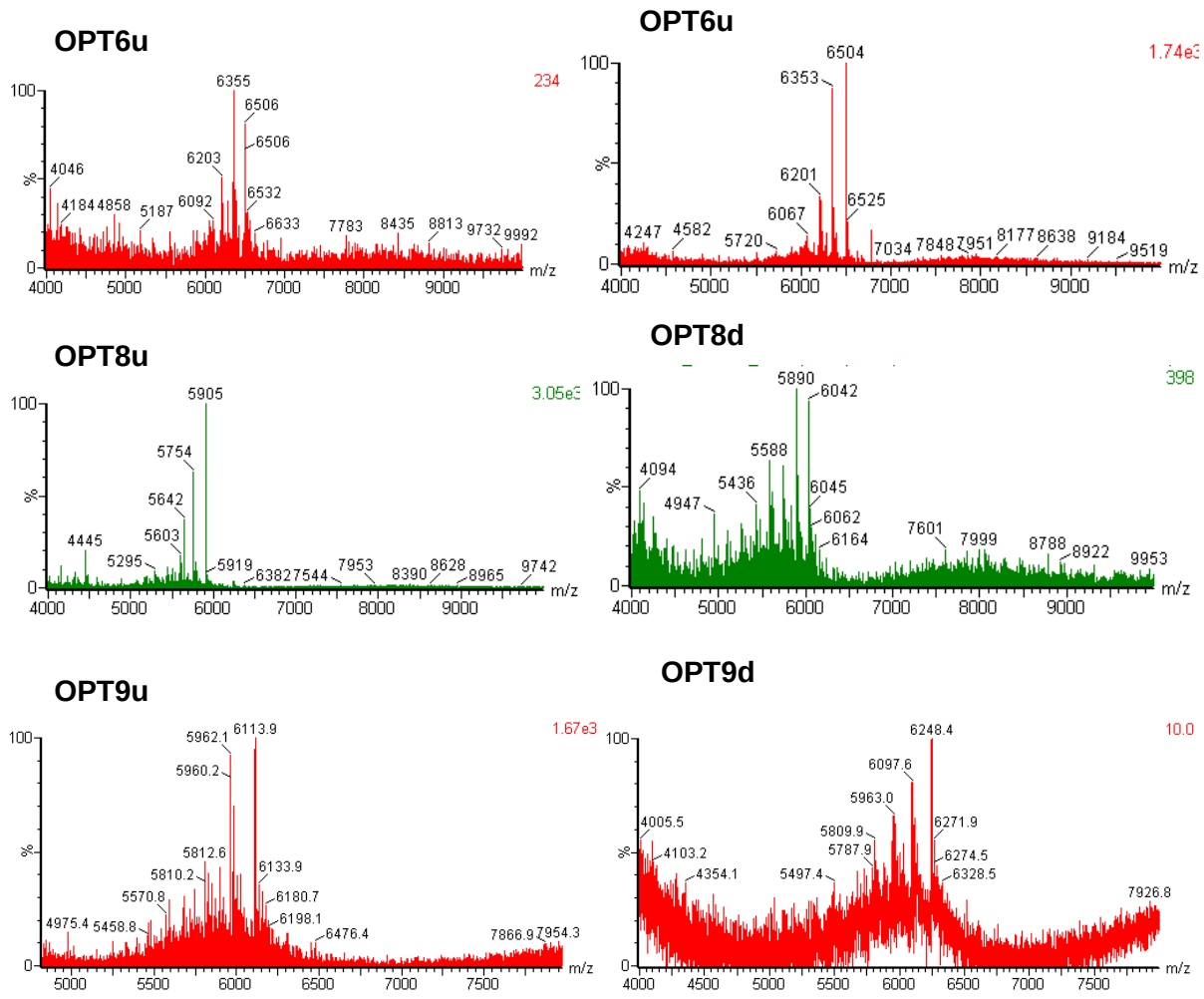


Figure 2.16. MALDI-MS Spectra of Individual Strands (up and down) of Invader Probes **OPT6**, **OPT8**, and **OPT9**.

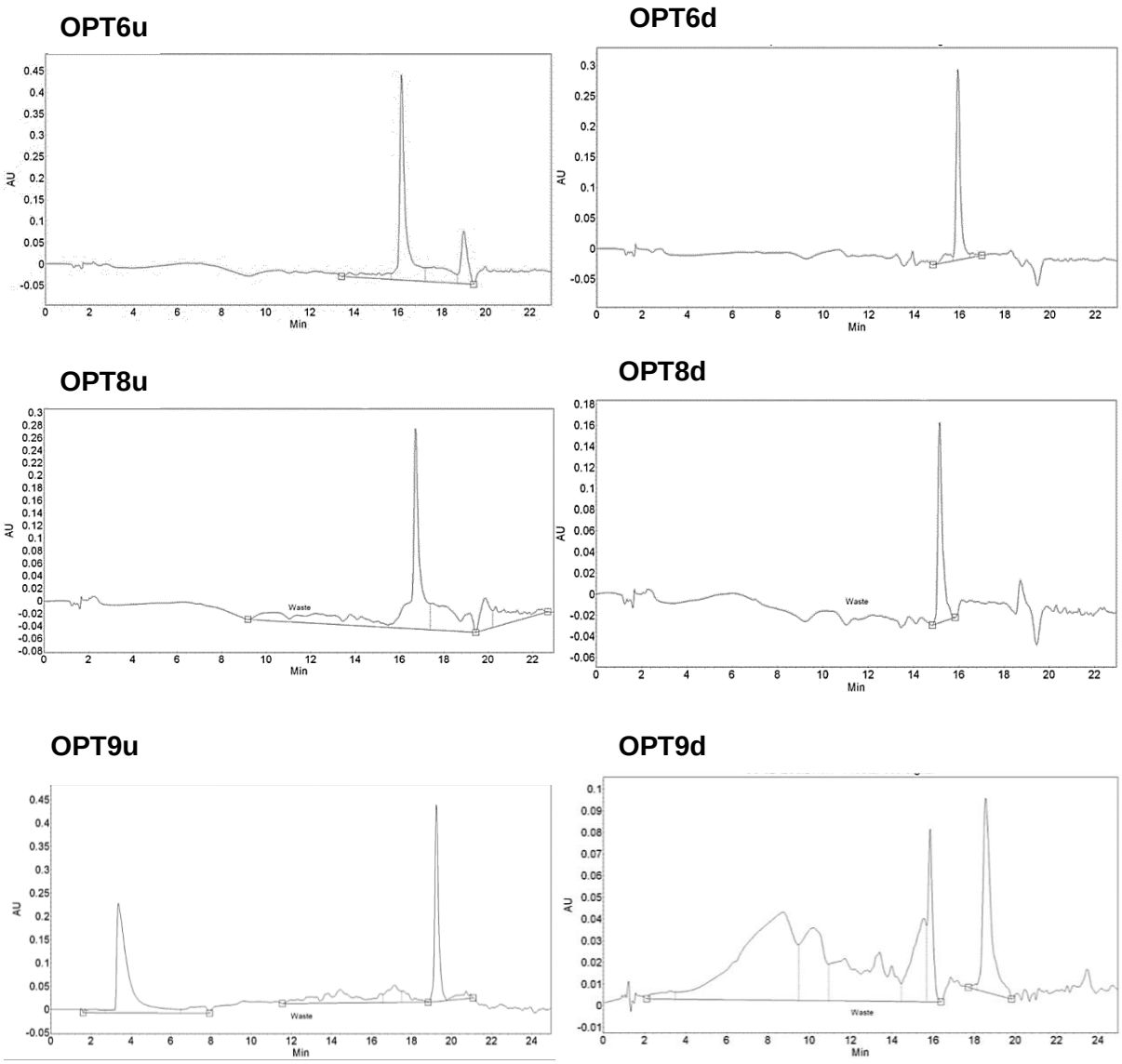


Figure 2.17. HPLC Spectra of optimized probes.

GGGATCTCAGCCCTGTGGTCTGGGAACTGGGAAACCGGCTTGAGTATGTGTGCTG1TTATCAGCACTGTGCCCTGGCAACTCTG
2ATACTGGTTTGTGTTCATGTGTGTGTGTGTGTGTGTGTGTTGCTGTTCTC4AGCCCTGTGCCCTGGCAACTGTGCAACCAGTATC
TGTATGCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGCTGTTCTCAACCCATTGCCCTGGCAACTGTTCAACCAGTTT
GTGTATATGTGTGTGTAGATGTGTGTGCCATCCTGAGCC3TTGTGCCCTGGCAACTGGGGAAACGGTGTGTGTGTGTTGTGTGTCT
GTGTGTGTGCT5GATTTCAGCCATGTGCCCATCTAA6CTGTGCAACTGGTTTGTGTGTGTGTGCACGCGATTCTCACCTCTGTGTCC
TGGCAACTGTGTAACCGTTTGTGTGTGTGAGTGTGTGTAAGTGTGTGCTCTTTTCAGCCCTGTTTCCTAGAAACTGTGGAACCGGTT
GGTGTGTGTGTGTGTCTGTGTGTGTGTGTGCCATTCTC4AGCCCTGTGCCCTGGCAA7CTGTGCAATATTTTGTCGTGTGTGTGTGTG
TGTGTATTTGTGTGTGCAATTCACAGCCCTGTTCCCTGGCAACTGTGCAAGCAGATTGTTGCGTATGTTTCTGTGTGTGTGTGTGTG
TGTGTGTGTGTGTATGTGCTGTTCTC4AGCCCTGTGCCCTGGCAACTGTGAAACCGGTTTGTATGTGTGTGTGTGTTTGTGTGTGCCA
8TTCAC4AGCCCTGTGCCCTGGCAACTGTGCAAGCAGTTTGTGTGTGCATGTGTCTGTGTGTGTATGTGTCTGTGTGTGCATGTGTCT
GTGTGTG9TTATATGCTGTTCTC4AGCCCTGTGCCCTGGCAACTGAGAAACCGGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGC
CAGTTTCAGCCCTGTGCCTTGGTAACTGTGCAAGTGGTTTGTGTGTGTGT10GTGTAGTGTATATGTGTGTGTGTGGTTTGACCAGTT
TTCAGCCCTGTGCCTTAGTAACTGTGTAACTGGTGTGTGTGTGTGTGTGTGTGTGTGTGCTCTTCTC4AGCCCTGTGCCCTGTTAACT
GTGCAAGCGGTTTGTCTGTGTATGTGAGTGGGTG

Figure 2.18. Position of sequences within the *DYZ-1* satellite gene on the bovine (*Bos taurus*) Y chromosome targeted by the different Invader probes.[30] The target sequence for **INV4** (highlighted grey) is presented six times within the tandem repeat (~6 x 10^4 tandem repeats); all other target sequences are present only once (underlined green).

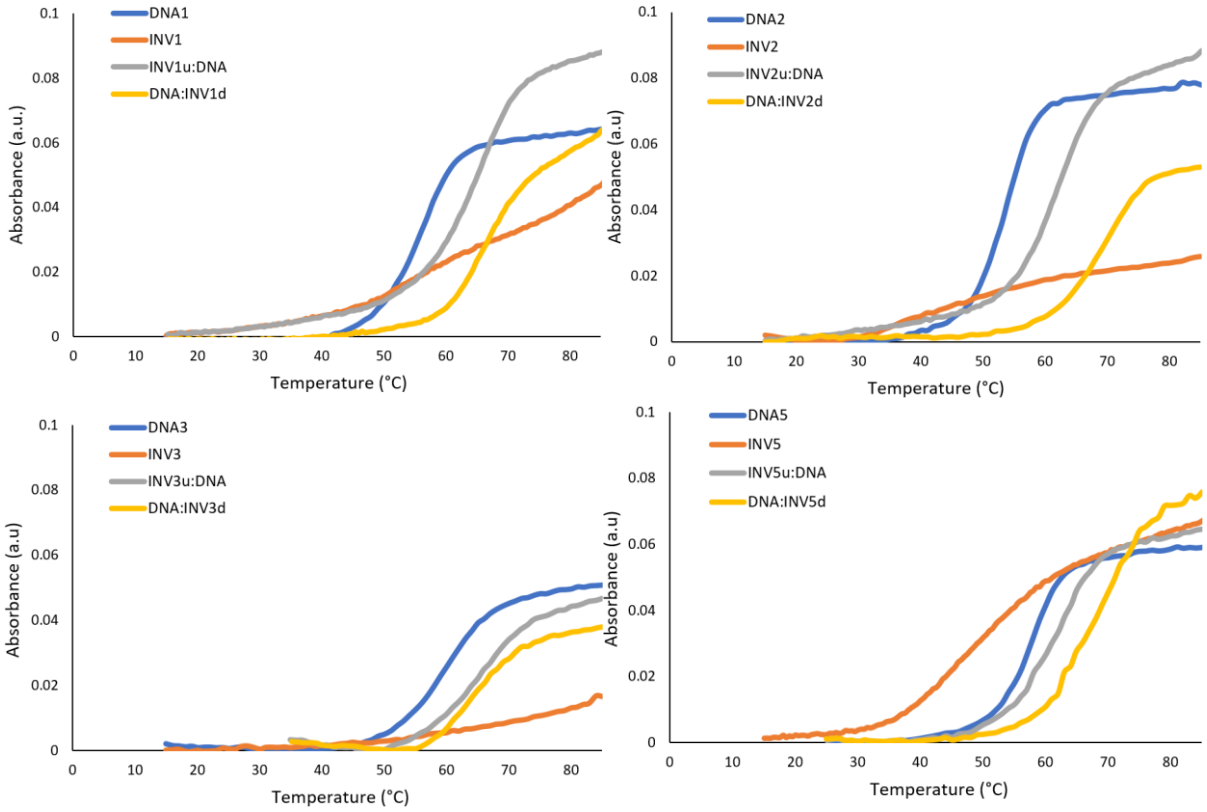


Figure 2.19. Representative thermal denaturation curves for Invader probes **INV1**-**INV3** and **INV5**, the corresponding duplexes between individual probe strands and cDNA, and unmodified reference DNA duplexes (**DNA1**- **DNA3** and **DNA5**).

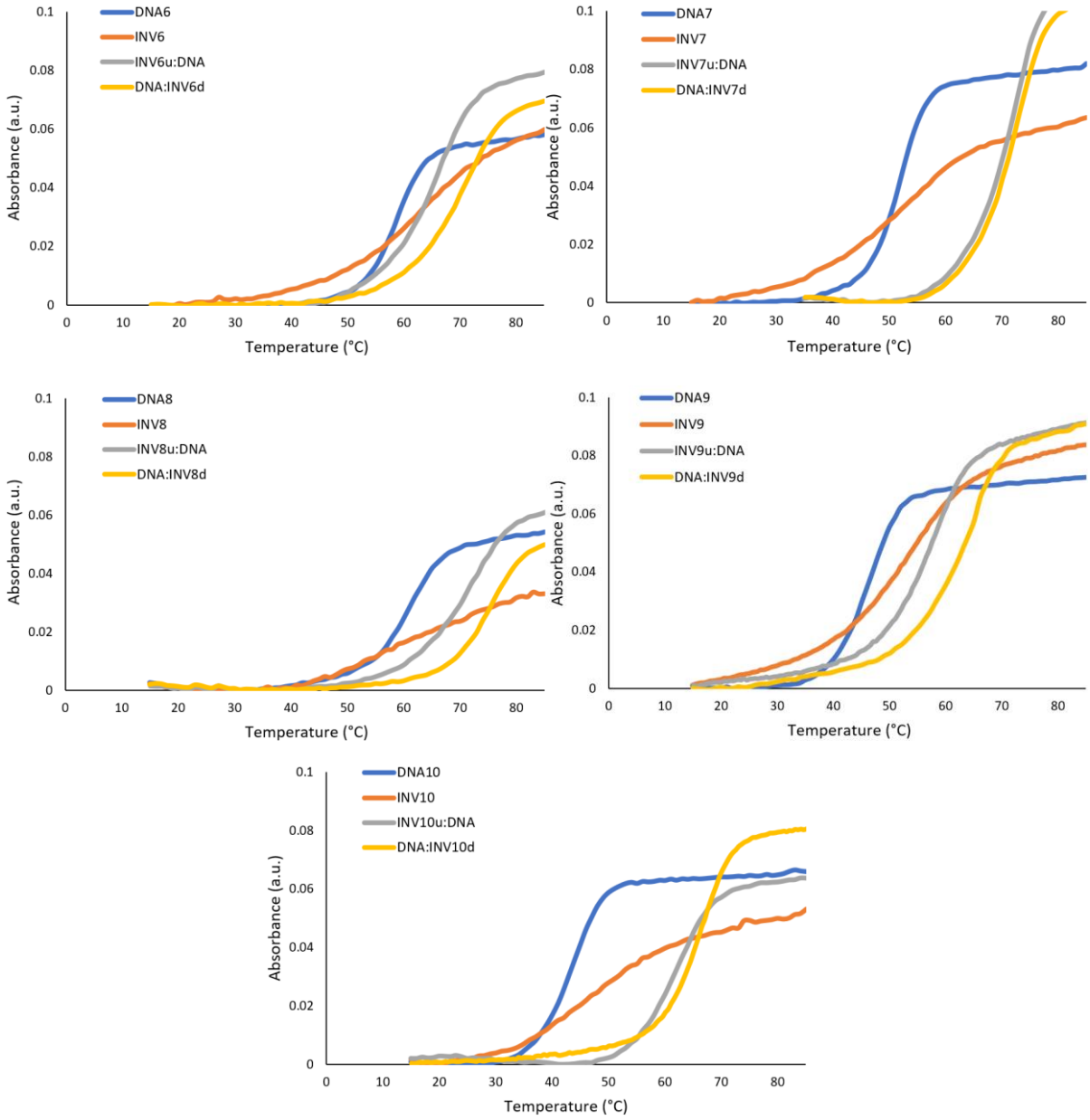


Figure 2.20. Representative thermal denaturation curves for Invader probes **INV6-INV10** and the corresponding duplexes between individual probe strands and cDNA, and unmodified reference DNA duplexes (**DNA6-DNA10**).

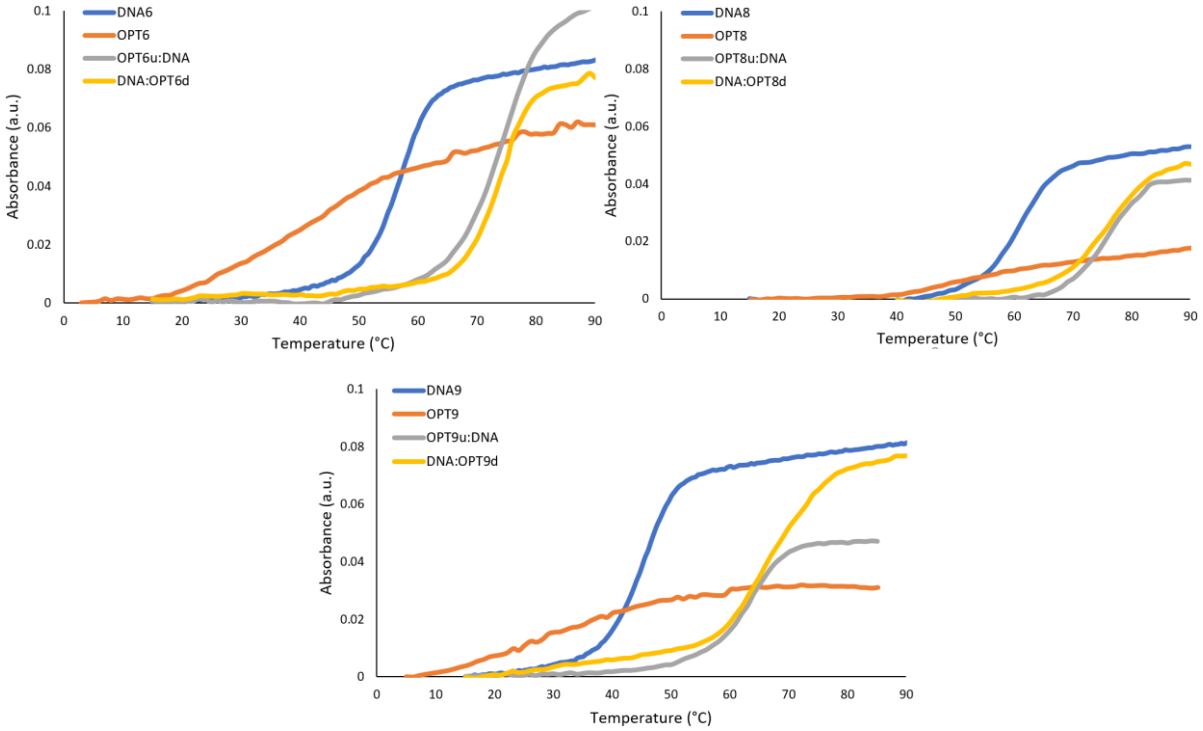


Figure 2.21. Representative thermal denaturation curves for Optimized Invader probes **OPT6**, **OPT8**, and **OPT9** and the corresponding duplexes between individual probe strands and cDNA, and unmodified reference DNA duplexes (**DNA6**, **DNA8**, and **DNA9**).

Table 2.12. Change in Gibbs free energy (ΔG^{310}) at upon formation of duplexes and change in reaction free energy upon Invader-mediated recognition of isosequential 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'ON :cDNA	3'ON: cDNA	
INV1	5'-Cy3-T**U**ATCAGCAC**U**G**U**GC-3' 3'- AA**U**AGTCGTGA**C**A**C**G-Cy3-5'	ND	-72 [-8]	-45 [+19]	ND
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'	ND	-66 [-3]	-69 [-6]	ND
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]	-7
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 duplexes (**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). ND represents thermal denaturation curves too broad for adequate baseline setting to determine thermodynamic parameters.

Table 2.13. Change in enthalpy (*ΔH*) upon formation of duplexes and change in reaction enthalpy upon Invader-mediated recognition of isosequential dsDNA targets ΔH_{rec}.[a]

Probe	Sequence	Δ*H*[ΔΔ*H*] (kJ/mol)			ΔH_{rec} (kJ/mol)
		Probe duplex	5'ON: cDNA	3'ON: cDNA	
INV1	5'-Cy3-T**U**ATCAGCAC**U**G**U**GC-3' 3'- AA**U**AGTCGTGA**C**A**C**G-Cy3-5'	ND	-414 [+40]	-345 [+109]	ND
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'	ND	-358 [+8]	-393 [+-27]	ND
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]	+89
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]	-123
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]ΔΔ*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). ND represents thermal denaturation curves too broad for adequate baseline setting to determine thermodynamic parameters.

Table 2.14. Change in entropy ($-T\Delta S^{310}$) upon formation of duplexes and change in reaction entropy upon Invader-mediated recognition of isosequential 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'ON: cDNA	3'ON: cDNA	
INV1	5'-Cy3-T**U**ATCAGCAC**U**G**U**GC-3' 3'- AA**U**AGTCGTGA**C**A**C**G-Cy3-5'	ND	341 [-49]	300 [-90]	ND
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]	394 [-257]	195
INV3	5'-Cy3-T**U**G**U**GCCCTGGC**A**AC-3' 3'- AA**C**A**C**GGGACCGT**U**G-Cy3-5'	ND	292 [-11]	324 [+21]	ND
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]	-96
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). ND represents thermal denaturation curves too broad for adequate baseline setting to determine thermodynamic parameters.

Table 2.15. Change in Gibbs free energy (ΔG^{310}) for optimized Invader probes upon formation of duplexes and change in reaction free energy upon Invader-mediated recognition of isosequential dsDNA targets (ΔG^{310}_{rec}).[a]

Probe	Sequence	$\Delta G^{310}[\Delta\Delta G^{310}]$ (kJ/mol)			ΔG^{310}_{rec} (kJ/mol)
		Probe duplex	5'ON: cDNA	3'ON: 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]	-88
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 [+14]	-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 2.16. Change in enthalpy (ΔH) for optimized Invader probes upon formation of duplexes and change in reaction enthalpy upon Invader-mediated recognition of isosequential dsDNA targets ΔH_{rec}.[a]

Probe	Sequence	$\Delta H[\Delta\Delta H]$ (kJ/mol)			ΔH_{rec} (kJ/mol)
		Probe duplex	5'ON: cDNA	3'ON: 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 [+378]	-519 [+1]	-620 [-100]	-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 [+297]	-476 [-13]	-425 [+38]	-272

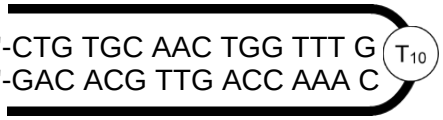
[a]$\Delta\Delta H$ is measured relative to the corresponding unmodified DNA duplex (**DNA6** = -520 kJ/mol, **DNA8** = -471 kJ/mol, **DNA9** = -463 kJ/mol).

Table 2.17. 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 isosequential 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'ON: cDNA	3'ON: 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 [-250]	452 [+52]	458 [+58]	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 [-264]	402 [-9]	349 [-62]	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 2.18. Sequence and intramolecular T_ms of DNA hairpins used herein.[a]

Hairpin	Sequence	T_m °C
DH1	5'-TTA TCA GCA CTG TGC 3'-AAT AGT CGT GAC ACG 	76.0
DH2	5'-ATA CTG GTT TGT GTT C 3'-TAT GAC CAA ACA CAA G 	72.0
DH3	5'-TTG TGC CCT GGC AAC 3'-AAC ACG GGA CCG TTG 	81.5
DH4	5'-AGC CCT GTG CCC TG 3'-TCG GGA CAC GGG AC 	82.0
DH5	5'-GAT TTC AGC CAT GTG C 3'-CTA AAG TCG GTA CAC G 	76.0
DH6	5'-CTG TGC AAC TGG TTT G 3'-GAC ACG TTG ACC AAA C 	75.0
DH7	5'-CTG TGC AAT ATT TTG T 3'-GAC ACG TTA TAA AAC A 	68.0
DH8	5'-TTC ACA GCC CTG TGC 3'-AAG TGT CGG GAC ACG 	80.5
DH9	5'-TTA TAT GCT GTT CTC 3'-AAT ATA CGA CAA GAG 	67.0
DH10	5'-GTG TAG TGT ATA TG 3'-CAC ATC ACA TAT AC 	62.0

[a] T_m were determined as described in Table 2.1.

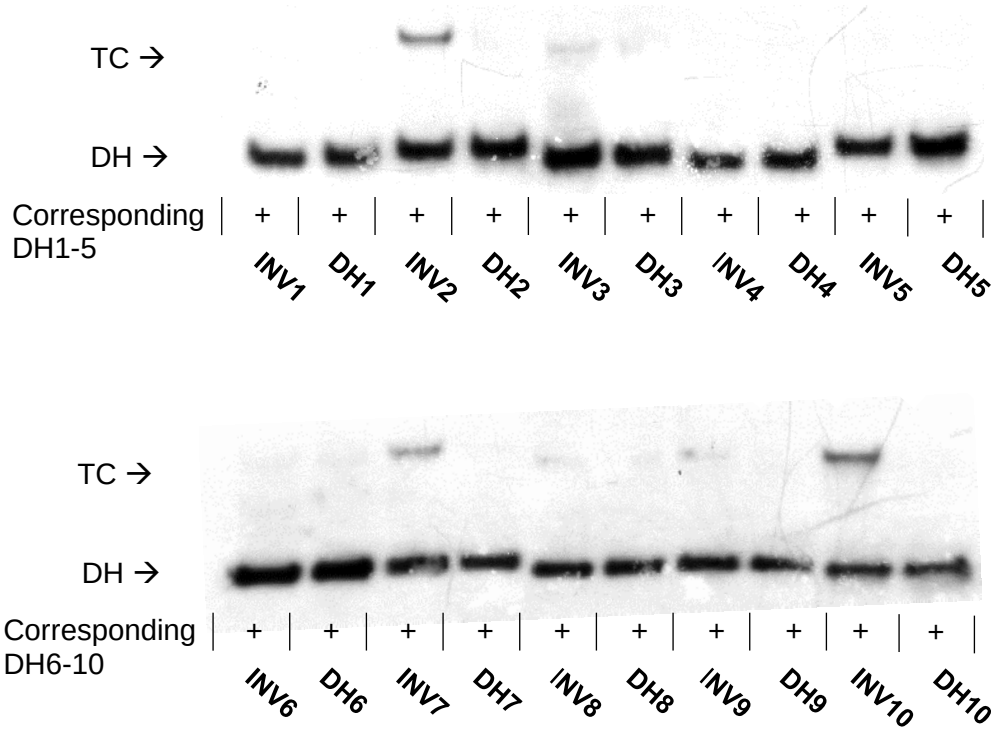


Figure 2.22. DNA hairpin (**DH1**-**DH10**) in absence or presence of a 5-fold molar excess of corresponding Invader probes. Shows that the hairpin in indeed the lower band. DIG-labeled **DH1**-**DH10** (34.4 nM, sequences shown in Table 2.18) were incubated with the corresponding Invader probe in HEPES 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).

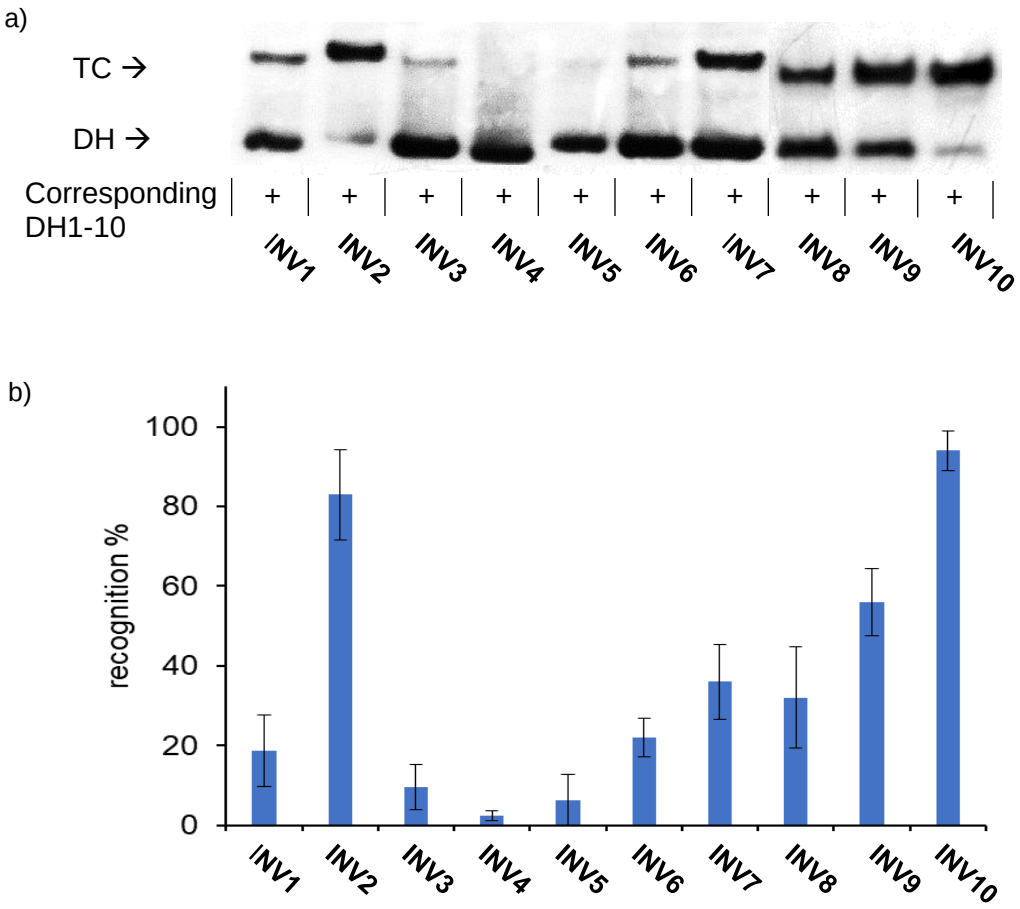


Figure 2.23. a) Representative gel electrophoretograms from recognition experiments between a 100-fold molar excess of Invader probes **INV1**-**INV10** and their corresponding DNA hairpin targets **DH1**-**DH10** following incubation for 2.5 h. b) Histograms depict averaged results from at least three recognition experiments with error bars representing standard deviation. TC = ternary complex. DIG-labeled DNA hairpins (34.4 nM, sequences shown in Table 2.18) 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).

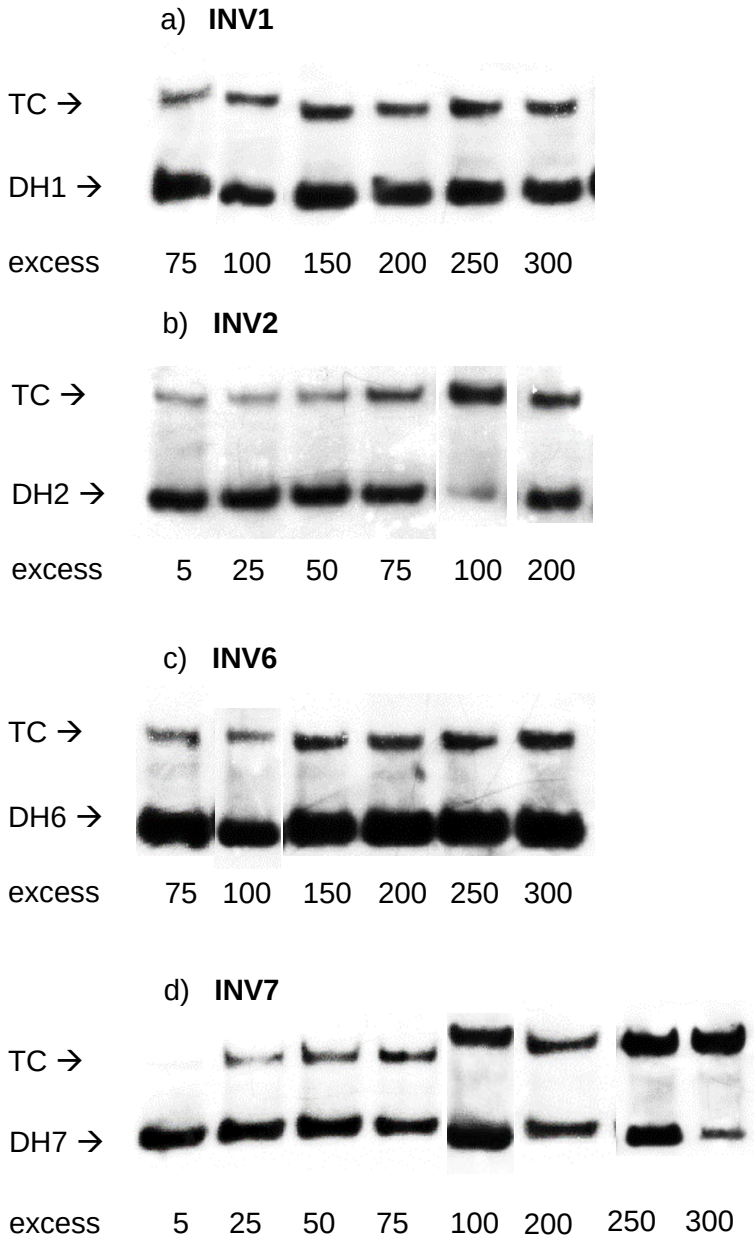


Figure 2.24. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) using different concentrations of Invader probes a) **INV1**, b) **INV2**, c) **INV6**, and d) **INV7** at 37 °C for 2.5 h. Experimental conditions are as specified in Figure 2.23. For dose-response curves, see Figure 2.26.

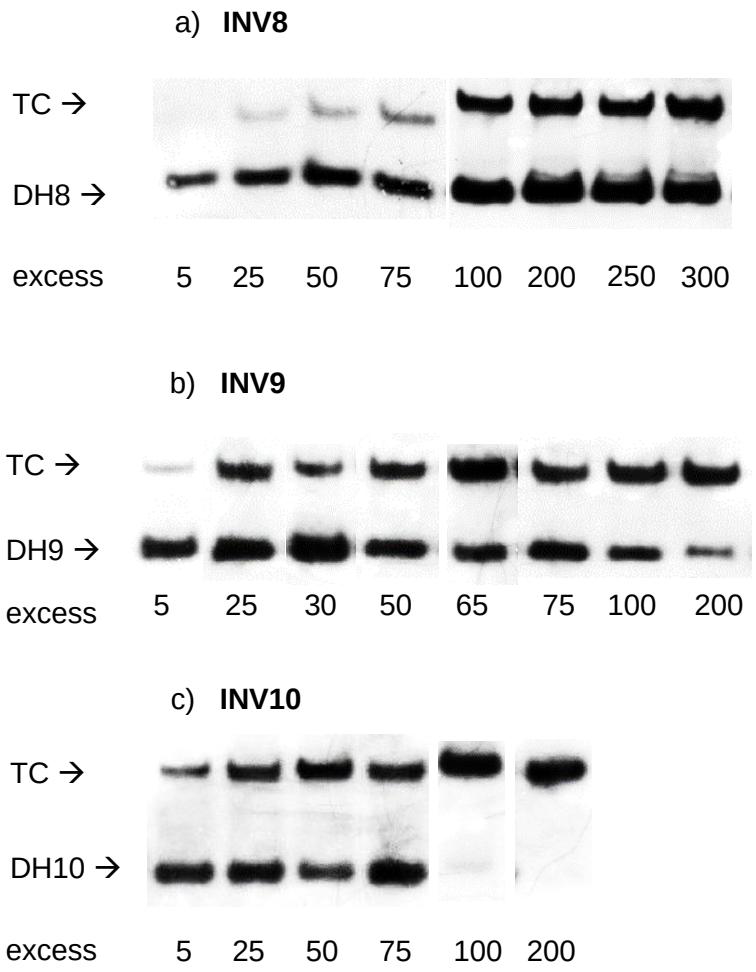


Figure 2.25. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) using different concentrations of Invader probes a) **INV8**, b) **INV9**, and c) **INV10** at 37 °C for 2.5 h.. Experimental conditions are as specified in Figure 2.23. For dose-response curves, see Figure 2.26.

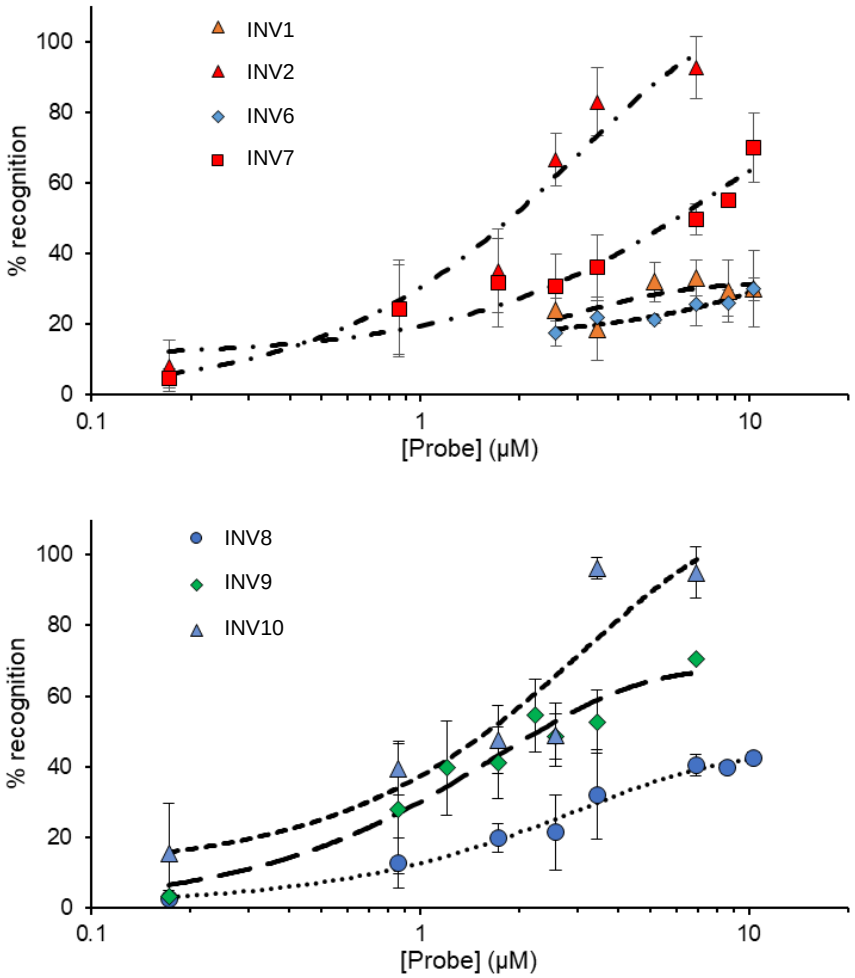


Figure 2.26. Dose-response curves for **INV1**, **INV2**, **INV6**, and **INV7** (upper panel), and **INV8**-**INV10** (lower panel) at 37 °C for 2.5 h. Experimental conditions are as described in Figure 2.23, except for variable probe concentrations.

Table 2.19. C_{50} values for Invader probes studied herein following 2.5 h of incubation.[a]

Probe	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 ± 13
INV9	2.2	56 ± 8.3
INV10	1.6	93 ± 5.0

[a] Rec_{100X} = level of DNA hairpin recognition using 100-fold molar probe excess. C_{50} values were determined from dose-response curves shown in Figure 2.26. ND = Not Determined due to low levels of recognition in the initial screen (See Figure 2.23)

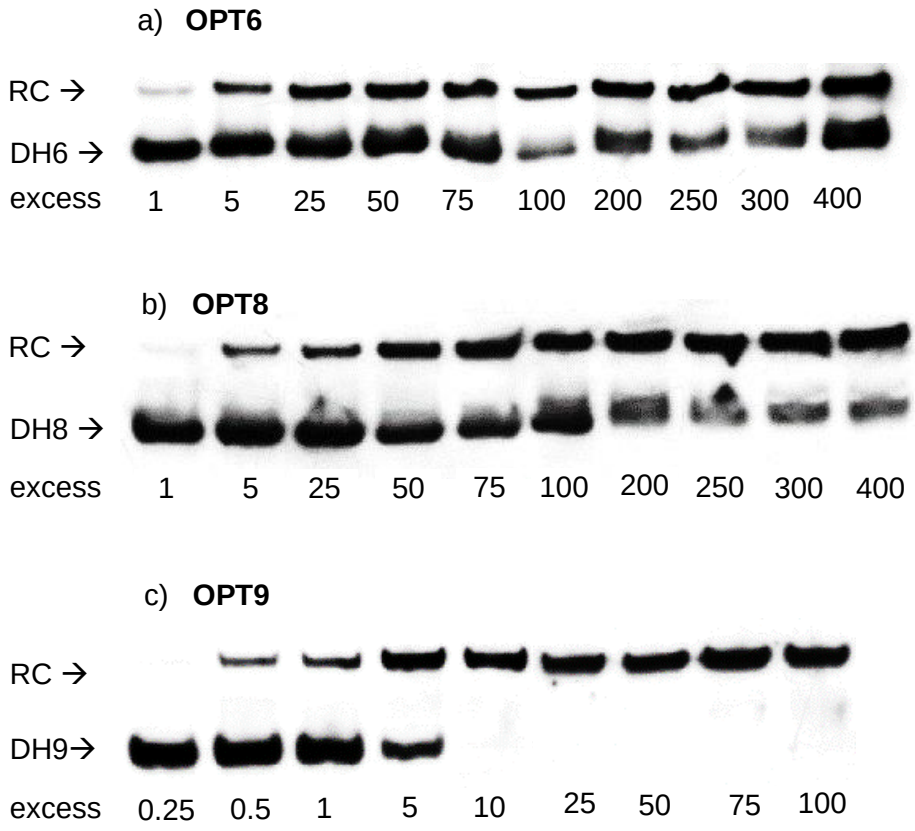


Figure 2.27. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) using different concentrations of optimized Invader probes a) **OPT6**, b) **OPT8**, c) **OPT9** at 37 °C following 2.5 h of incubation. Experimental conditions are as specified in Figure 2.10. For dose-response curves, see Figure 2.28.

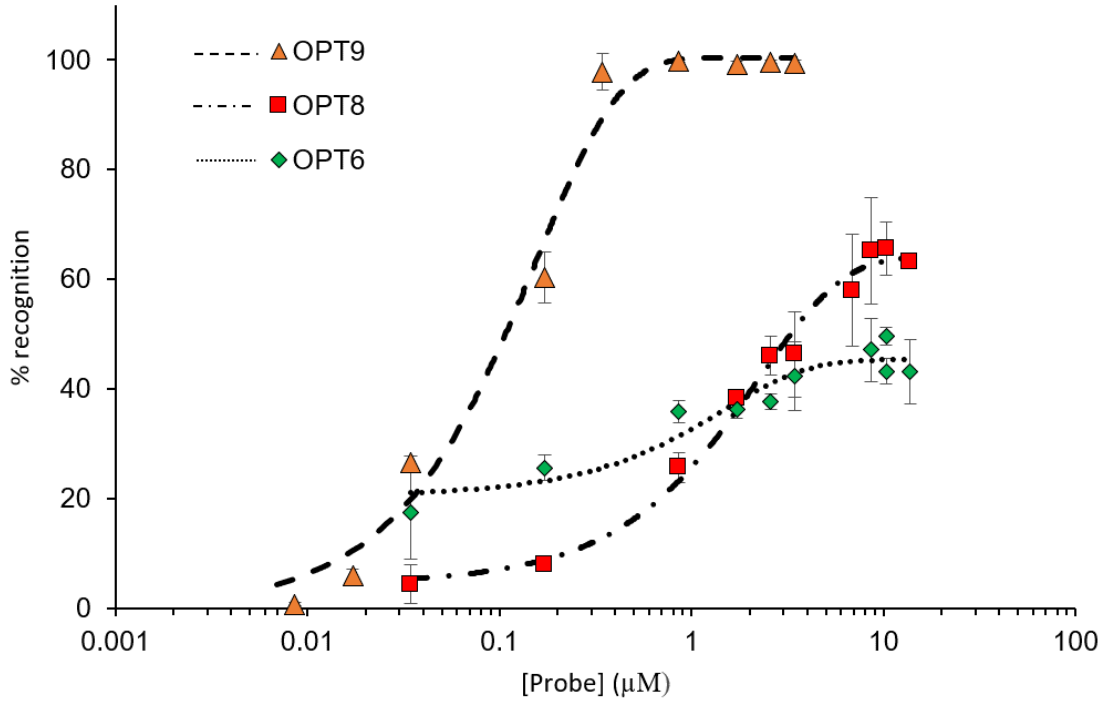


Figure 2.28. Dose-response curves for **OPT6**, **OPT8**, and **OPT9** at 37 °C following 2.5 h of incubation. Experimental conditions are as described in Figure 2.10, except for variable probe concentrations.

Table 2.20. C_{50} values for optimized Invader probes following 2.5 h of incubation.[a]

Probe	C_{50} (µM)	Rec_{100X} (%)
OPT6	ND	66 ± 3.4
OPT8	3.2	66 ± 0.4
OPT9	0.1	99 ± 8.9

[a] Rec_{100X} = level of DNA hairpin recognition using 100-fold molar probe excess. C_{50} values were determined from dose-response curves shown in Figure 2.28.

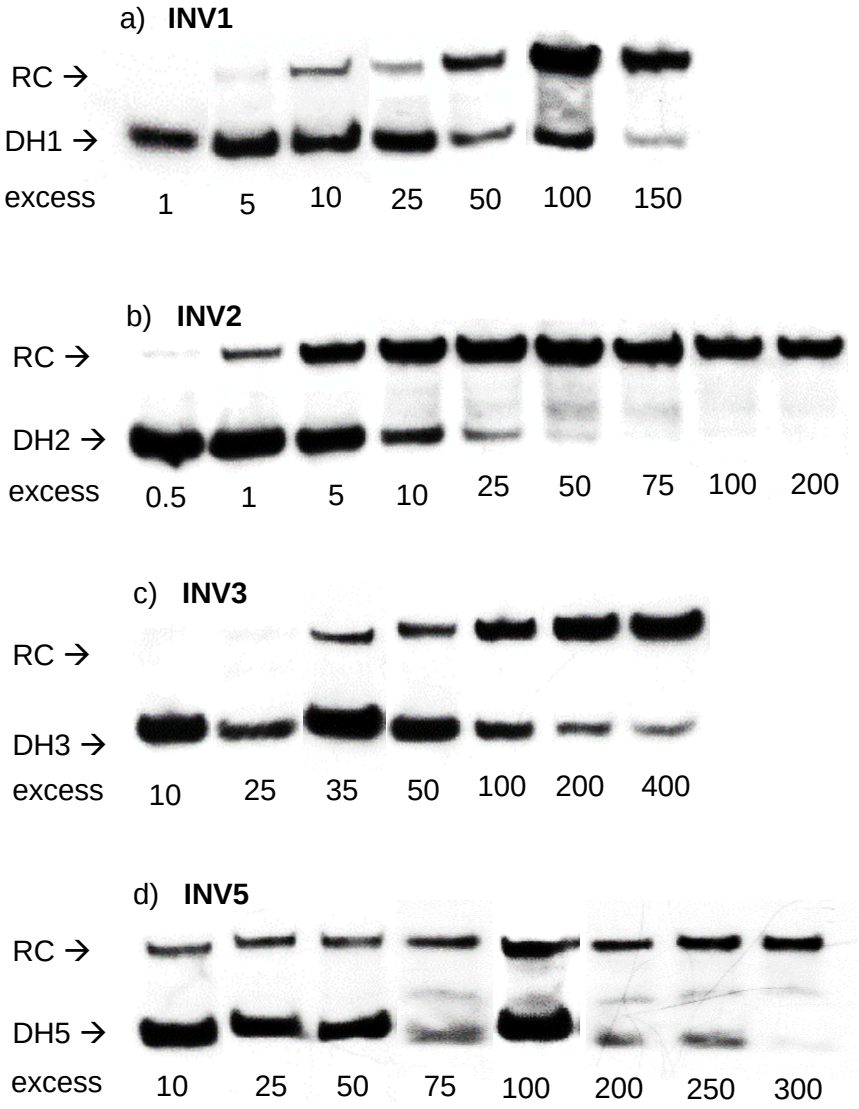


Figure 2.29. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) using different concentrations of Invader probes a) **INV1**, b) **INV2**, c) **INV3**, and d) **INV5** at 37 °C for 15 h. Experimental conditions are as specified in Figure 2.4. For dose-response curves, see Figure 2.5.

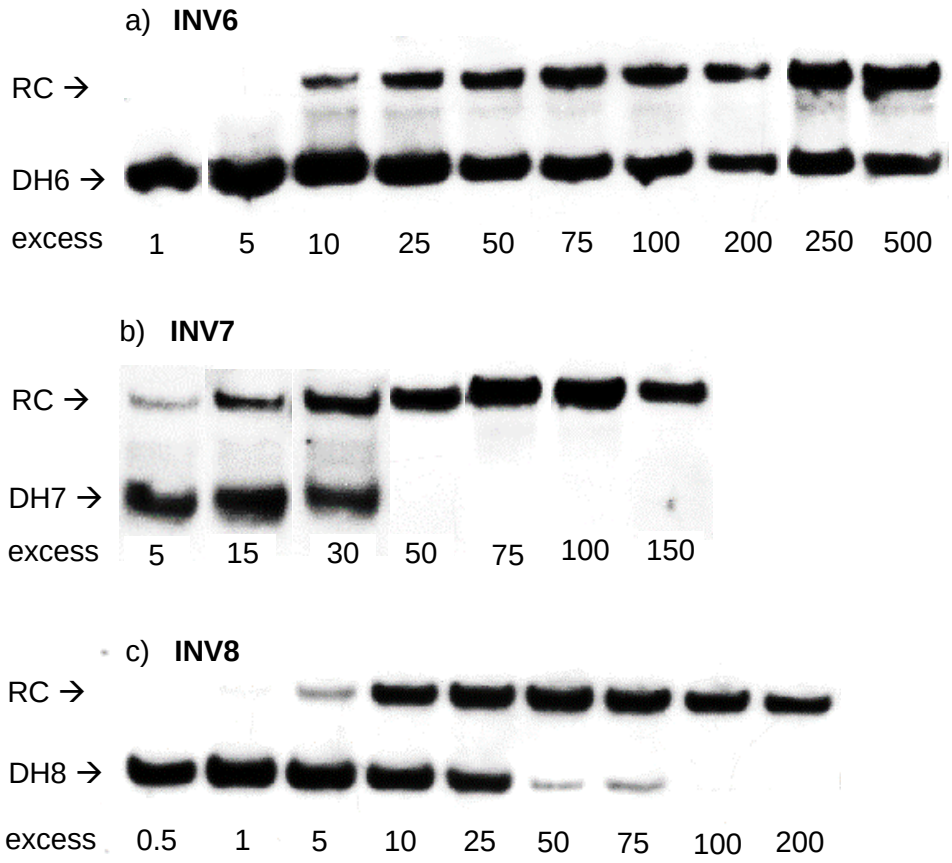


Figure 2.30. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) using different concentrations of Invader probes a) **INV6**, b) **INV7**, and c) **INV8** at 37 °C for 15 h. Experimental conditions are as specified in Figure 2.4. For dose-response curves, see Figure 2.5.

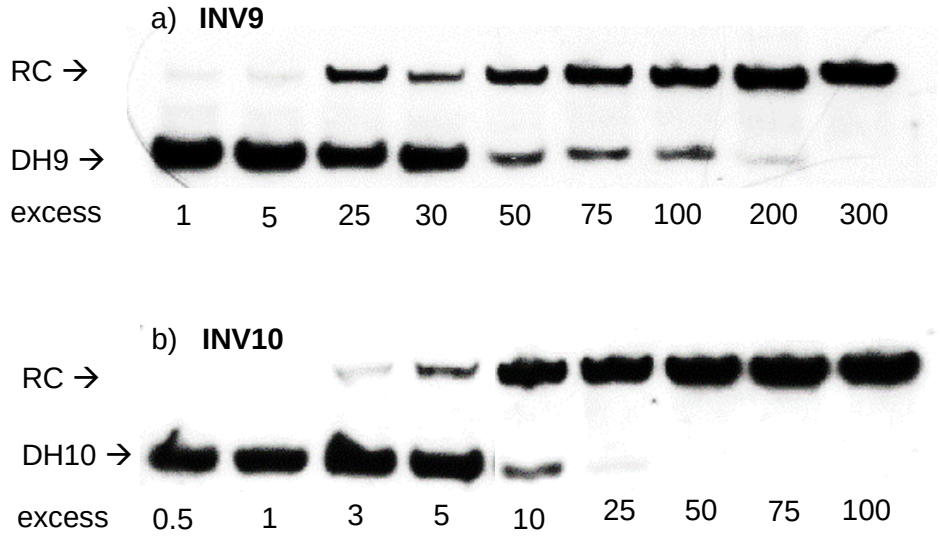


Figure 2.31. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) using different concentrations of Invader probes a) **INV9** and b) **INV10** at 37 °C for 15 h. Experimental conditions are as specified in Figure 2.4. For dose-response curves, see Figure 2.5.

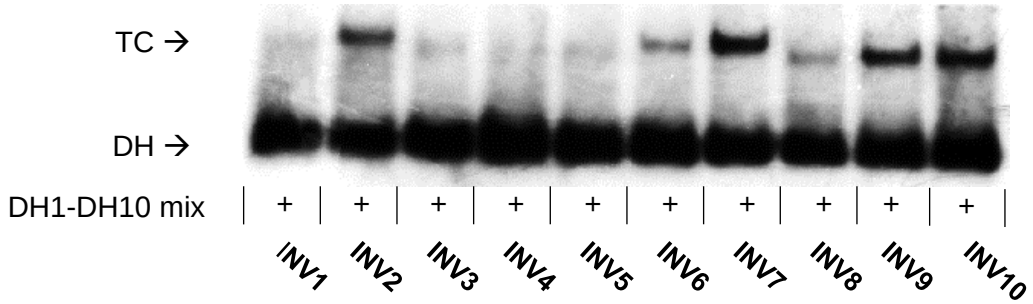


Figure 2.32. 100-fold excess of Invader probe (3.44 μM) was incubated at 37 °C for 2.5 h with a mixture of all ten DNA hairpins targets **DH1**-**DH10** (each hairpin present at 34.4 nM) (i.e., one complementary and nine scrambled targets) in each lane. Incubation for 2.5 hours at 37 °C, all other conditions are as reported in Figure 2.23. This shows that recognition of complementary DNA hairpin targets is not hindered by the presence of scrambled hairpin targets.

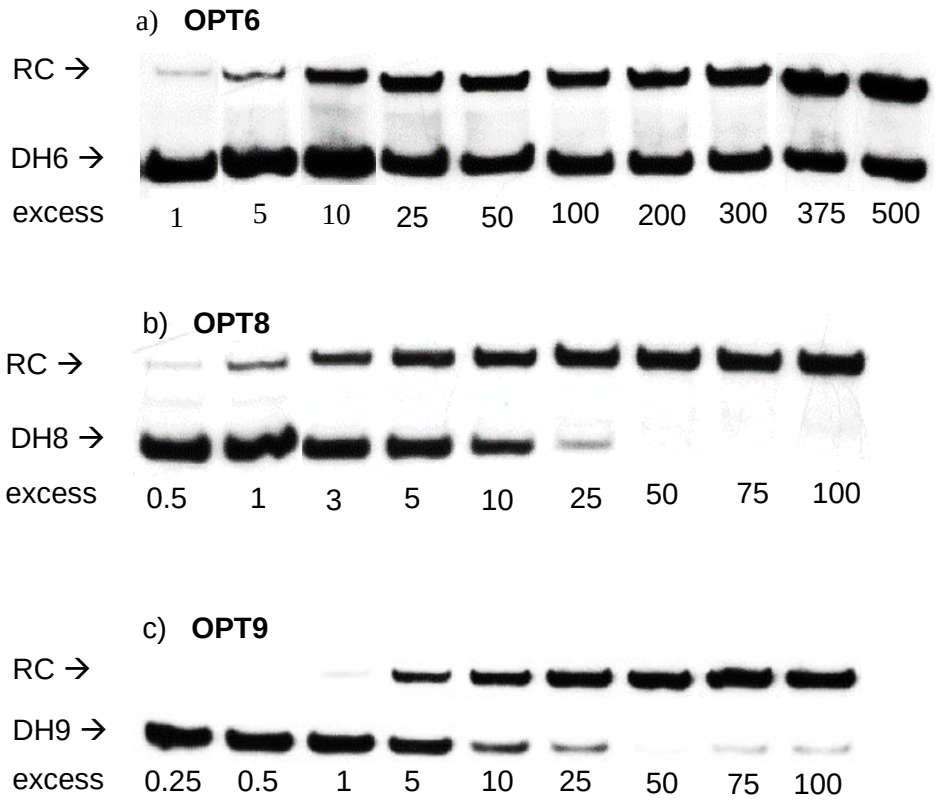


Figure 2.33. Representative electrophoretograms for recognition of model DNA hairpin targets (34.4 nM) using different concentrations of optimized Invader probes a) **OPT6**, b) **OPT8**, c) **OPT9** at 37 °C for 15 h. Experimental conditions are as specified in Figure 2.10. For dose-response curves, see Figure 2.11.

Discussion of imaging optimization

Following incubation with isolated nuclei, Cy3-labeled Invader probes display varying levels of signal strength, non-specific/background signal, and signal coverage, i.e., number of nuclei displaying one strong localized signal) (Figure 2.35, 2.36 and 2.37 and Table 2.21). 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 (5 min, 80 °C, 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 2.21). 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 2.21). 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.

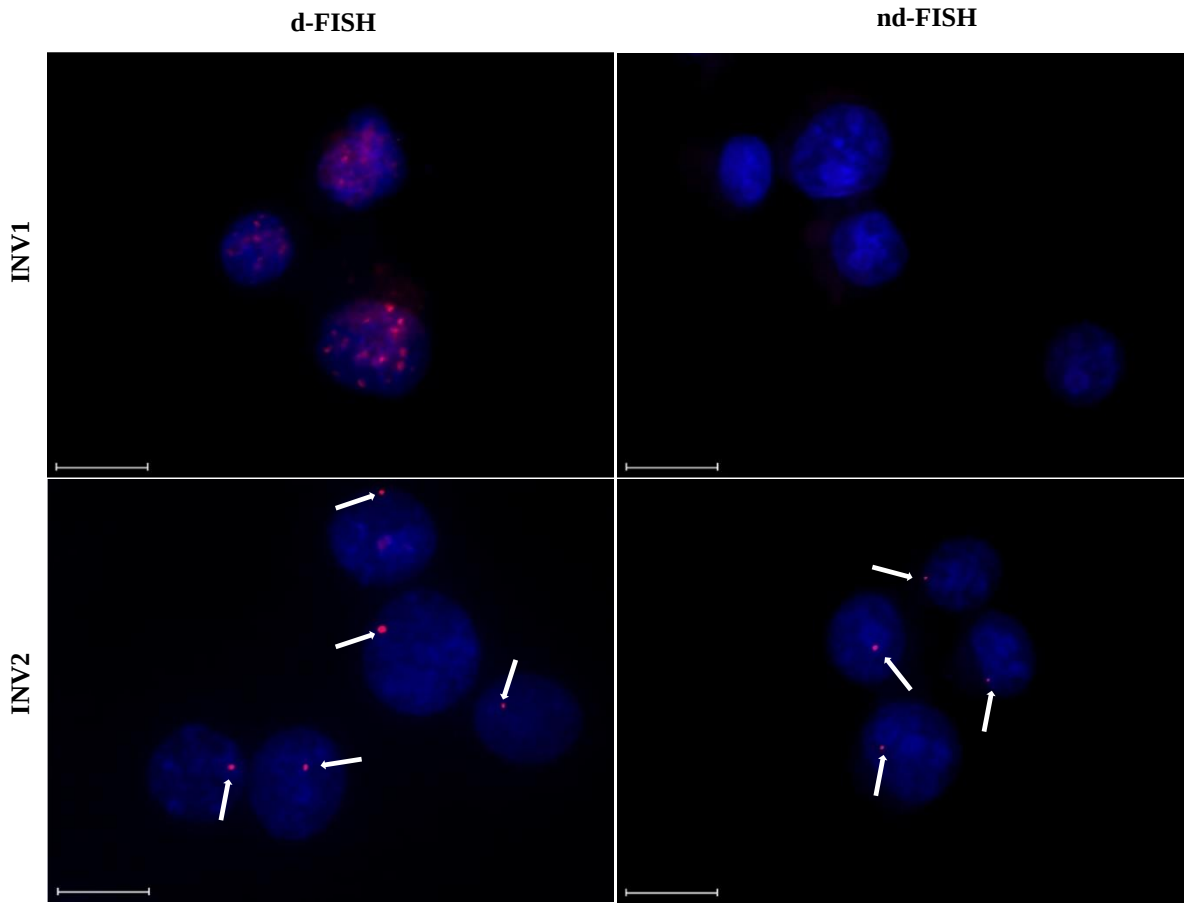


Figure 2.34. Representative images from FISH experiments using *DYZ1*-targeting Invader probes of **INV1** and **INV2** under denaturing (5 min, 80 °C) (left) or non-denaturing (3 h, 37.5 °C) conditions (right). Images are from representative of the signal intensity and background, and the size and morphology of all analyzed nuclei (~200 nuclei per probe). Fixed isolated nuclei from male bovine kidney cells were incubated with probes in a Tris buffer (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.

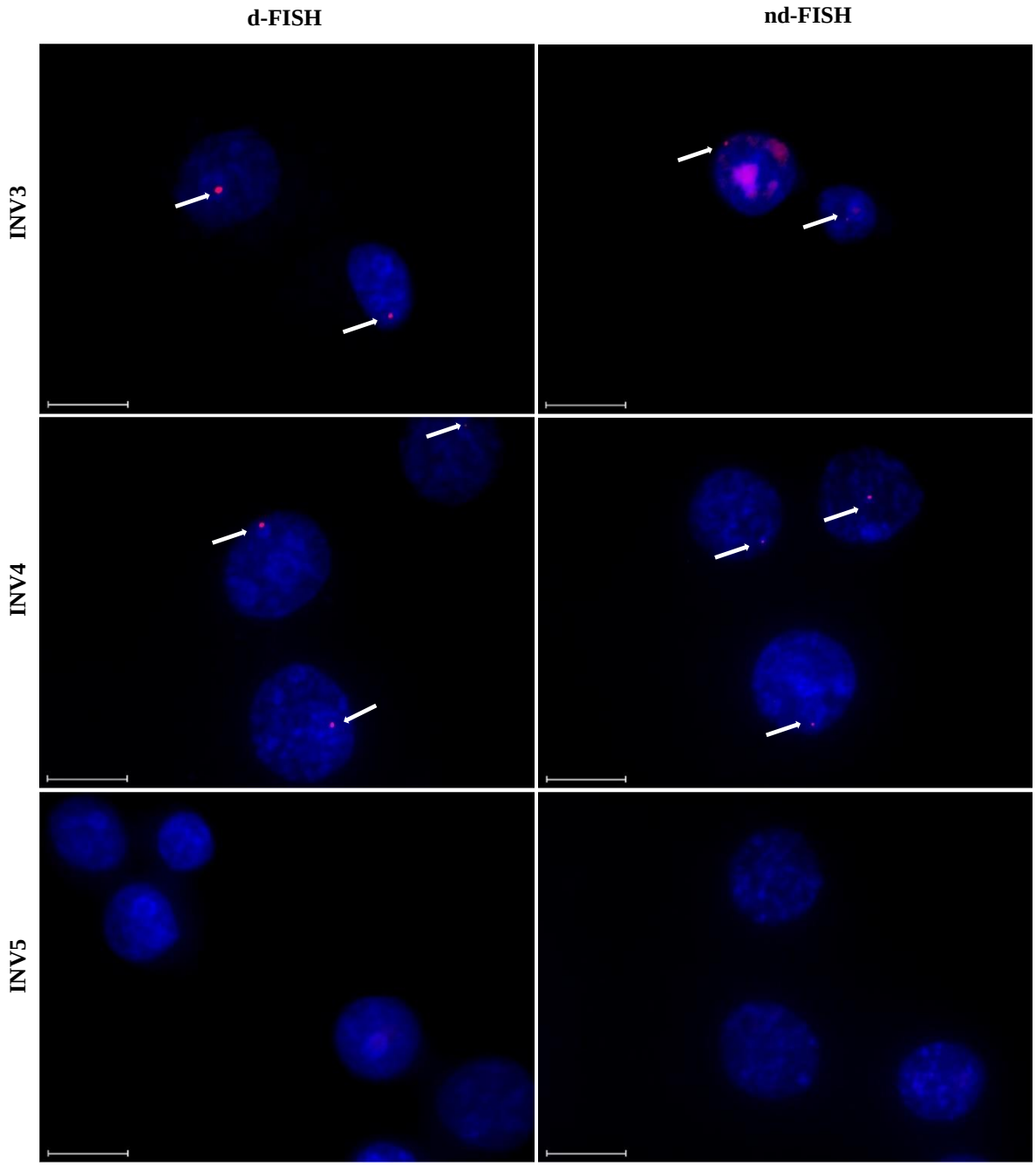


Figure 2.35. Representative images of **INV3**, **INV4**, and **INV5** from FISH experiments using *DYZ1*-targeting Invader probes. Incubation and imaging specifications are described in Figure 2.34. Scale bar represents 16 µm.

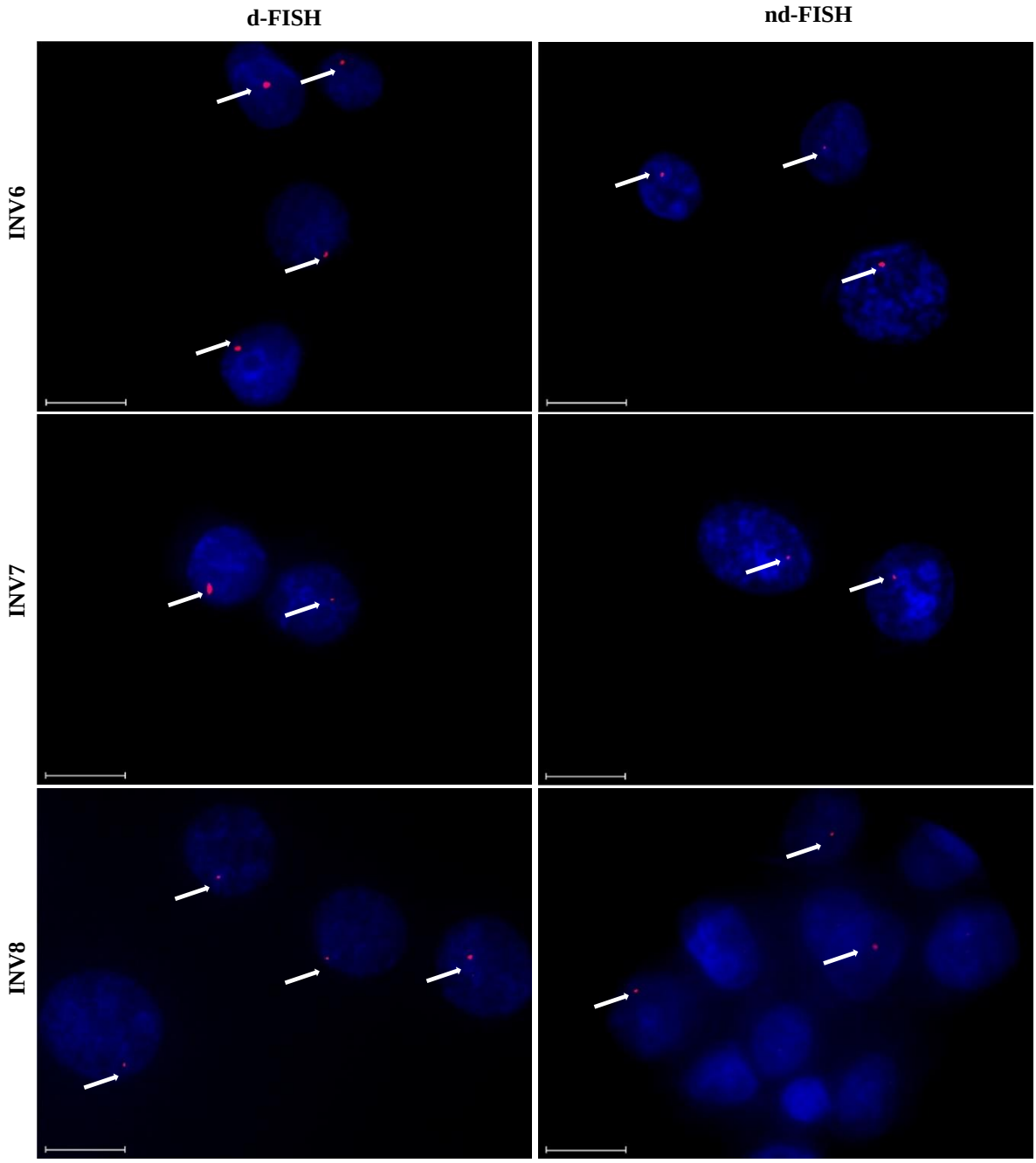


Figure 2.36. Representative images of **INV6**, **INV7** and **INV8** from FISH experiments using *DYZ1*-targeting Invader probes. Incubation and imaging specifications are described in Figure 2.34. Scale bar represents 16 µm.

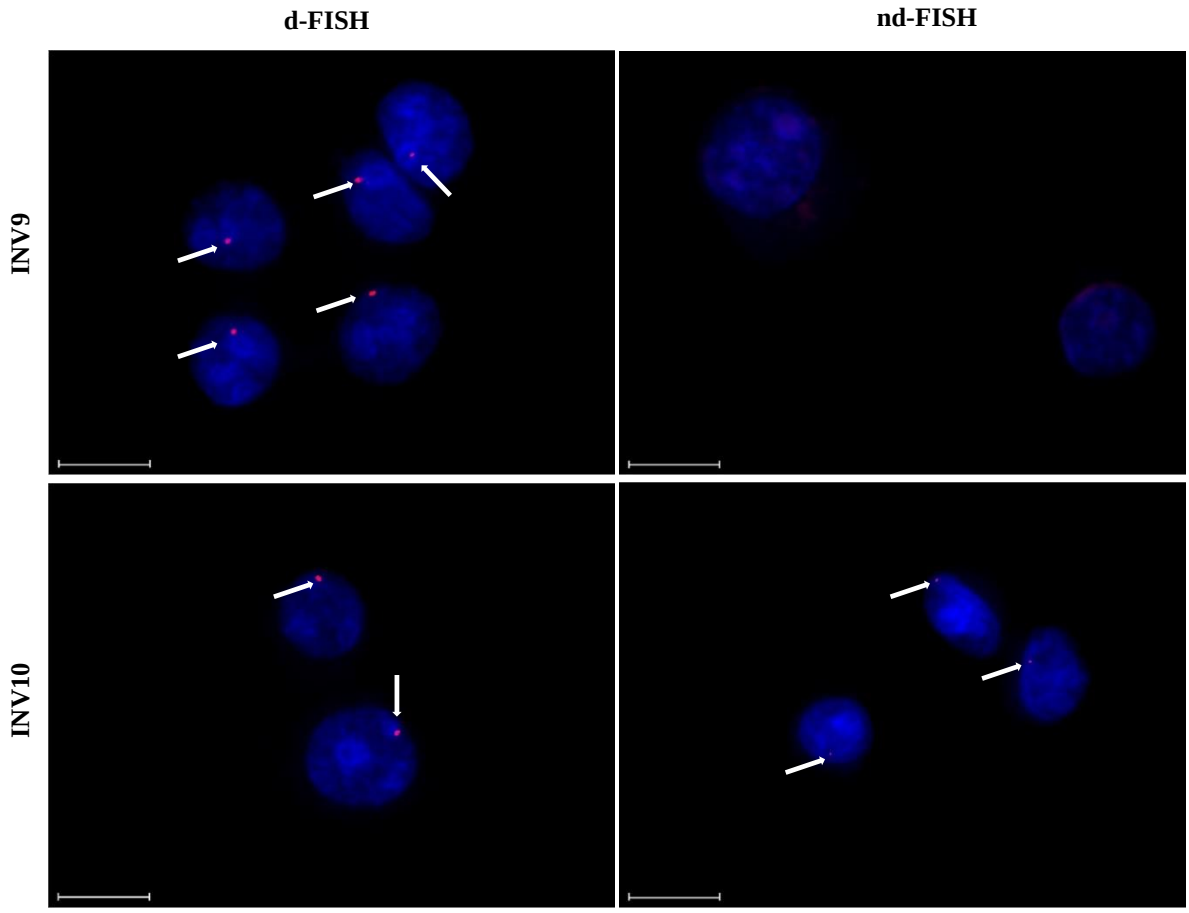


Figure 2.37. Representative images of **INV9** and **INV10** from FISH experiments using *DYZ1*-targeting Invader probes. Incubation and imaging specifications are described in Figure 2.34. Scale bar represents 16 μm.

Table 2.21. Qualitative assessment of signal intensity (I) and background (B) of representative images from d-FISH experiments in which Invader probes were used at 1x concentration 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	3	3	nd	3	3	3	3	3	3	3	3	3	3	3
[1x]	nd	3	3	3	2	3	1	1	nd	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. 1x concentration = 30 ng of probe per 200 µL of PCR buffer. **INV4** was not evaluated above 1x due to the presence of prominent background signal and multiple punctate dots. All other FISH experiments were conducted at 1/4x for **INV4** to yield acceptable signal-to-background signals. **INV5** and **INV6** displayed neither high signal coverage nor clear signal-to-background characteristics and were not evaluated at concentrations above 1x. **INV7-INV10** were not evaluated above 3x probe concentration due to poor signal-to-background characteristics at 3x probe concentration. nd = no discernable signal.

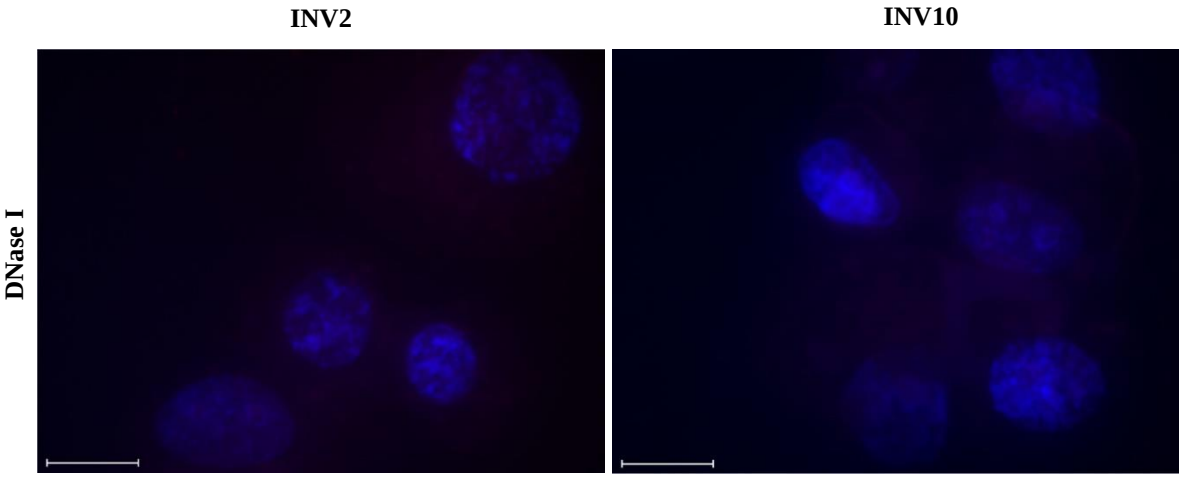


Figure 2.38. Representative Images for FISH experiments of nuclei pre-treated with DNase I prior to incubation with **INV2** and **INV10**. Note the absence of signal in the DNase I treatment indicating the Invader probes' targeting of the chromosomal DNA. Incubation and imaging specifications are described in Figure 2.34. Scale bar represents 16 µm.

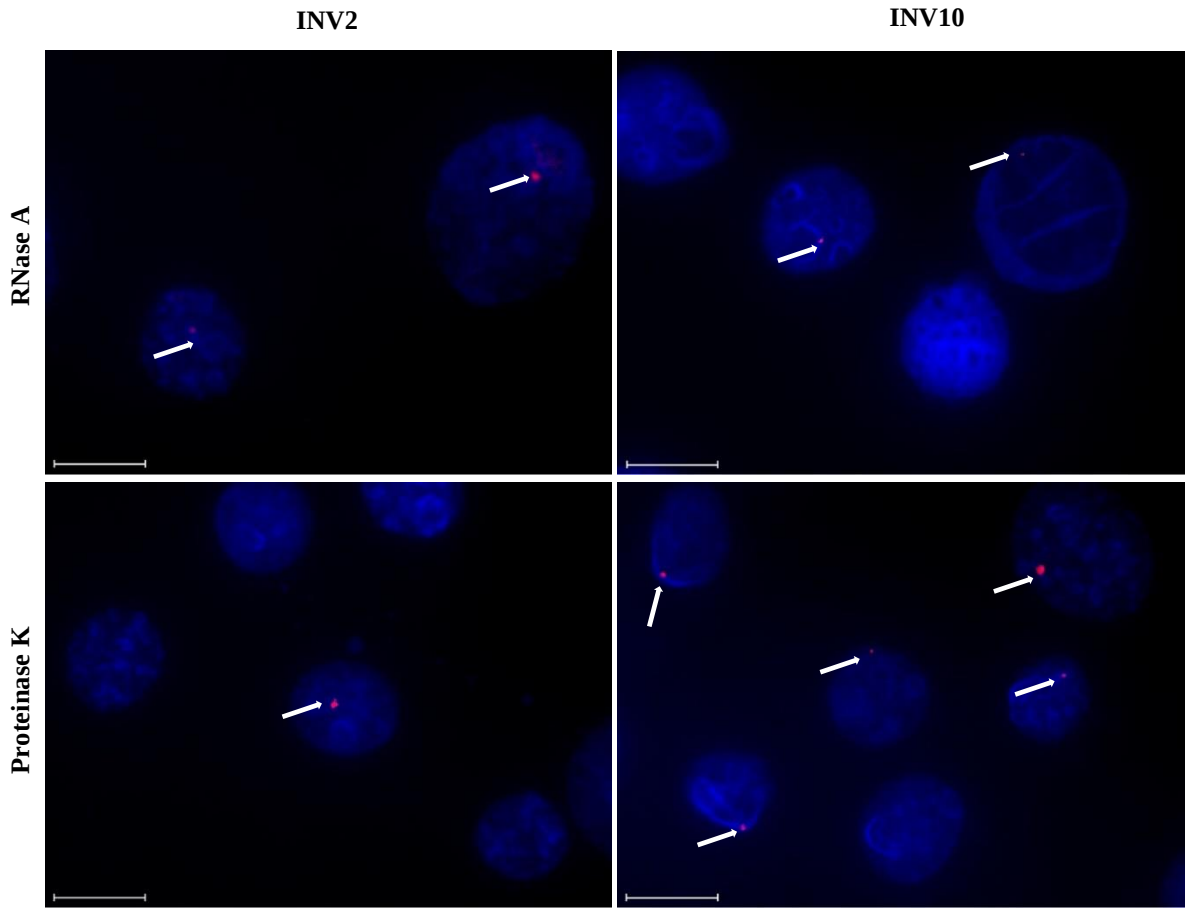


Figure 2.39. Representative Images for FISH experiments of nuclei pre-treated with RNase A and Proteinase K prior to incubation with **INV2** and **INV10**. Note the continued presence of signal in the RNase A and Proteinase K pre-treated nuclei (albeit there is a reduction in signal coverage which could be attributed to the loss of genetic material/number of nuclei resulting from the enzyme treatments, which is not seen in DNase treatment). Denaturing incubation and imaging specifications are described in Figure 2.34. Scale bar represents 16 μm.

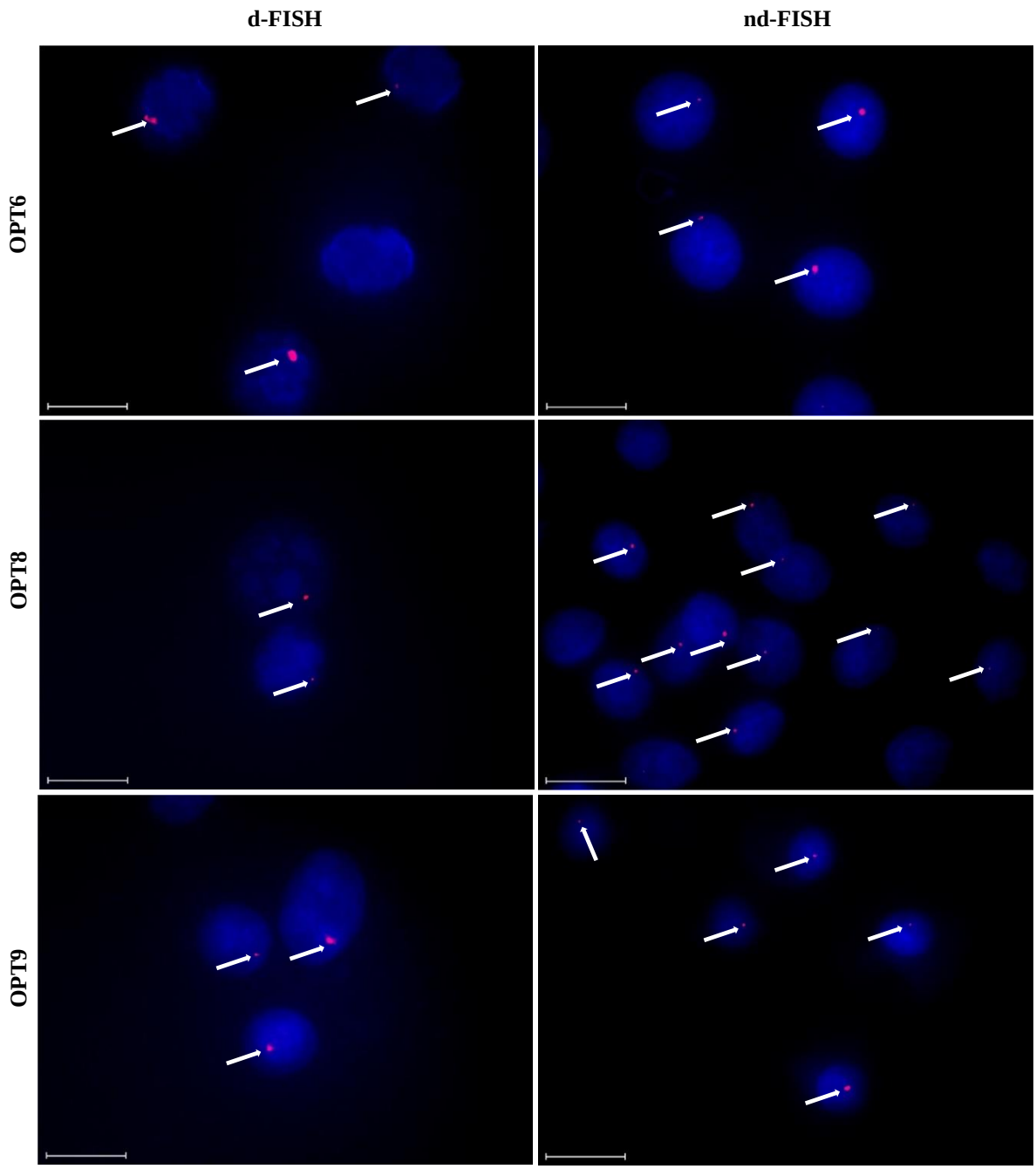


Figure 2.40. Images from FISH experiments using DYZ1-targeting **OPT6**, **OPT8**, and **OPT9** under denaturing (5 min, 80 °C) (left), or non-denaturing (3 h, 37.5 °C) conditions (right). Incubation and imaging specifications are described in Figure 2.34. Scale bar represents 16 µm.

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

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

21. 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.

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

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

24. 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.

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

26. B. A. Didion, S. Karmakar, D. C. Guenther, S. P. Sau, J. P. Verstegen and P. J Hrdlicka, *ChemBioChem*, 2013, **14**, 1534-1538.

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

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

29. 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.

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

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

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

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

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

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

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

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

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

39. 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.

40. M. A. Morgan, K. Okamoto, J. D. Kahn and D. S. English, Biophys. J., 2005, 89, 2588–2596.

41. A. M. Brown, *Comput. Meth. Prog. Biomed*., 2001, **65**, 191–200.

Chapter 3: Analysis of Select Invader Probes by Flow Cytometry

Caroline P. Shepard and Patrick J. Hrdlicka

3.1 Introduction

Fluorescent *in situ* hybridization (FISH) assays are powerful assays for detection of nucleic acids.[1] By labeling oligonucleotide probes with fluorophores, RNA and DNA can be visualized and quantified in many assay types (i.e. *in vitro*, *in vivo*, *in situ*, etc). Our laboratory has utilized Cy3-labeled Invader probes to detect specific target regions on the Y chromosome of isolated nuclei of bovine cells at near physiological conditions using a DNA FISH assay.[2] Invader probes were incubated with fixed nuclei on microscope slides and the signal/background quality and number of nuclei displaying the expected single punctate signal (signal coverage) was determined manually through visualization using an inverted fluorescent microscope. The FISH assay offers compelling proof-of-concept evidence for Invader-mediated detection of specific chromosomal DNA targets and hints at future diagnostic applications. However, the employed microscopy-based quantification is limited. Signals assessment is time- and labor-intensive as nuclei are evaluated one-by-one and quantification only reflects a small sample size (i.e. hundreds of nuclei). Additionally, the microscopy-based FISH assays to not provide information about important biological characteristics, such as cell cycle stages and telomere status. The use of flow cytometric assays overcomes the limitations of microscopy-based assays as it allows for high-throughput, multiparametric analysis. Recent studies have utilized flow cytometry to characterize the binding of fluorophore-labelled probe to mRNA and chromosomal DNA in cells and isolated nuclei.[3-6] Termed Flow-FISH by previous works, this assay utilizes the automation and detection power of flow cytometers to quantify fluorescence and shape characteristics of many isolated nuclei within a suspension.

Utilization of Flow-FISH assays for Invader-based detection of chromosomal DNA targets will enable: i) analysis of thousands of nuclei to quantify Invader probe binding to chromosomal DNA within a suspension, offering a first step to incubation with whole cells, ii) evaluation of the impact that cell behaviors (i.e. cycle stage) has on Invader-mediated recognition of chromosomal DNA targets, and iii) detection of multiple fluorophores simultaneously labeling through other biological molecules (i.e. proteins, nucleic acids, different genes, etc). To this end, we have developed a FISH assay for Flow Cytometry analysis to evaluate DNA-targeting Invader probes when incubated with nuclei suspensions. This proof-of-concept study demonstrates that: i) the experimental protocol yields isolated nuclei that can be analyzed with respect to cell cycle stages, and ii) Invader probes specifically target 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.

3.2 Results and Discussion

Experimental design

To render the chromosomal DNA accessible, nuclei were isolated from bovine cells the nuclear membrane was permeabilized. Removal of cellular debris and deconstruction of the nuclear membrane ensured that the Invader probes were not encumbered by cellular structural barriers, allowing the focus to be on recognition of chromosomal DNA. Invader probes **INV4** and **INV10** (Table 3.1) have been previously shown to target corresponding complementary sequences in the *DYZ-1* satellite gene (~6 x 10^4 tandem repeats of a ~1175 bp region) on the bovine (*Bos taurus*) Y chromosome (NCBI code: M2606).[2,7] The Cy3-labelled Invader probes were incubated with the nuclei in an estimated 30-fold excess of their target sequence for 3 h at 37.5 °C in a PCR buffer (20 mM Tris, 100 mM KCl, pH 8.0), which are the same incubation

conditions as used in the previously published FISH assays. These are non-denaturing conditions that are physiologically relevant, which will allow for extrapolation of results to cell studies (although efficient nuclear delivery must be realized). Following incubation, the nuclei were washed with 1xPBS to remove excess Invader probe and subsequently DAPI stained. Following a 10-minute DAPI labeling step, nuclei were washed once again by 1xPBS and analyzed by flow cytometry.

Table 3.1. Sequences of Invader probes used in Flow Cytometer FISH assays.

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'

[a] Invader modifications used are 2'-*O*-(pyren-1-yl)methyl RNA monomers with **A** (adenin-9-yl), **C** (cytosin-1-yl), and **U** (uracil-1-yl) nucleobases.

Gating strategies

A suspension of nuclei incubated with **INV10** was analyzed by flow cytometry to determine if the isolated nuclei and Cy3-labeled Invader probes could be detected. The DAPI stain allowed for gating (i.e., differentiation) of the nuclei population from cellular debris and possible unbound Invader probes present in solution. Debris is denoted as black dots and grey peaks in Figure 3.1. Using the acquisition software, a gate (i.e., the area within the orange box) was applied to isolate occurrences of intact nuclei as determined by high Forward Scatter intensity (FSC-H) and DAPI fluorescence (PB450-A). FSC is used as a measure of object size and DAPI fluorescence is a non-specific DNA stain. Large objects (i.e. high FSC) stained with DAPI (i.e. DNA) were determined to be the nuclei populations. This determination of nuclei populations was conferred by the histogram of PB450-A vs Count (number of nuclei) (Figure

3.1). The 2 high peaks (pink and blue) with a middle lower region (green) within the graph is typical of nuclei populations, with each color representing a cell cycle stage within interphase. This is possible since the amount of genetic material increases throughout interphase. Broken into stages G1 (cell components besides DNA is replicated), S (DNA is replicated), and G2 (DNA is fully replicated, and the genetic material has doubled from G1 stage), there is an increase in DAPI fluorescence in interphase caused by the increase in DNA. Nuclei populations can be gated into each stage (i.e. pink = G1, green = S, blue = G2) to compare fluorescent intensities. 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 these nuclei populations gated, Cy3 fluorescence intensities of the entire nuclei population and each interphase stage (G1, S, and G2) were quantified. This is depicted in Figure 3.1 as number of nuclei (i.e. count) displaying a particular Cy3 fluorescence intensity (PE-H). First, in comparing the Cy3 intensities for each stage, there is an overall increase in fluorescence from G1 to G2. For each Invader probe analyzed in this preliminary study, the Cy-3 intensity is double when comparing G1 to G2, with fluorescence intensities falling somewhere in between these values for S stage. This can be attributed to the fact that the amount of genetic material has double, including the target site. There are double the amount of target sites which means a doubling of Cy3 signal. This may indicate that there is little difference between targetability between G1, S, and G2. More studies will need to be conducted to see if this trend holds true for all Invader probes targeting different sequence. This initial survey has produced a usable gating strategy to determine Invader probe binding to the chromosomal DNA of isolated nuclei, allowing subsequent experiments to be conducted to determine the specificity of this binding.

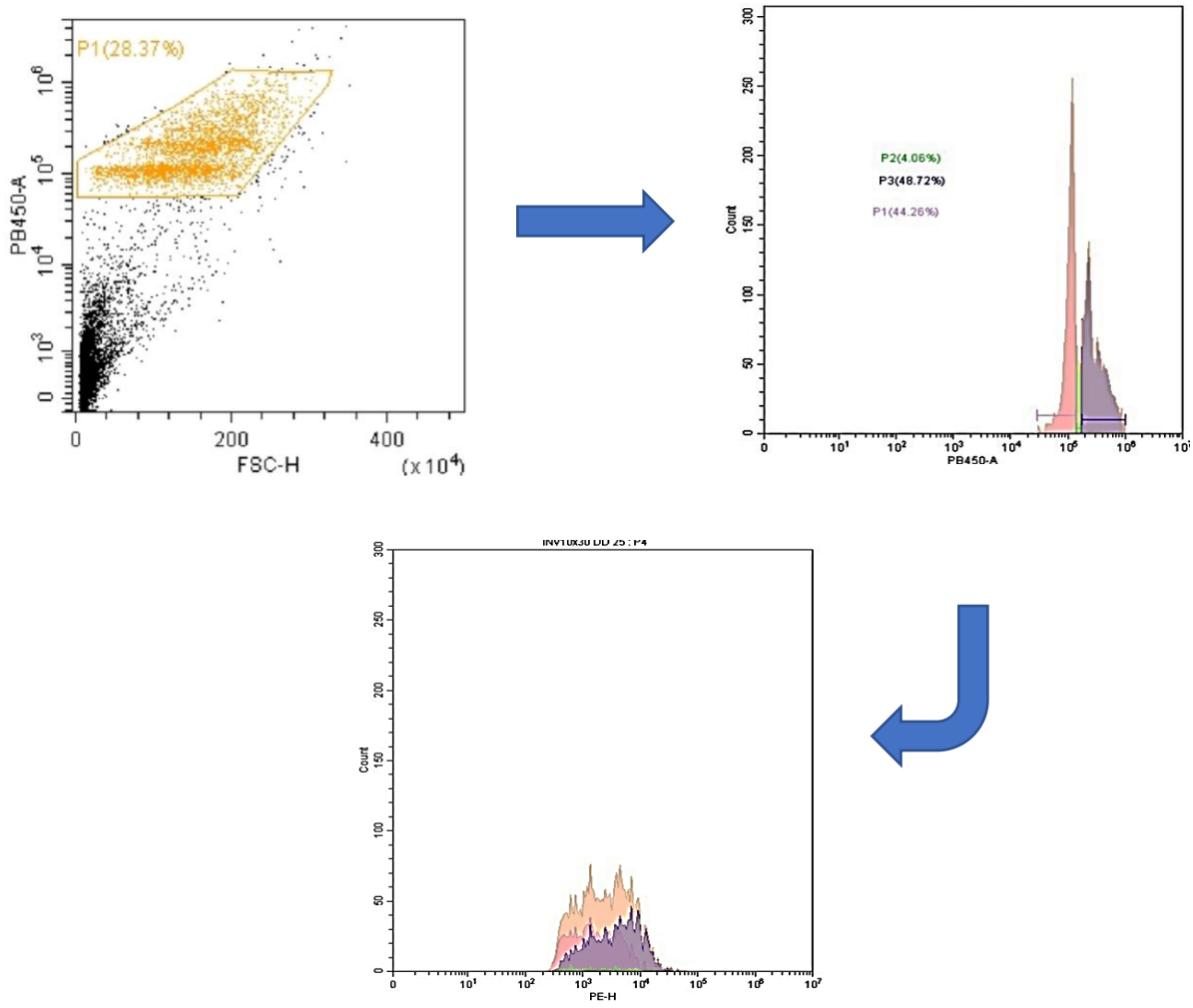


Figure 3.1. Gating strategy utilizing DAPI (PB450-A) and Cy3 (PE-H) channels. Anaylzed nuclei were incubated with **INV10** in 30-fold excess of target sequence for 3 h at 37.5 °C in PCR buffer (20 mM Tris, 100 mM KCl, pH 8.0). Utilizing acquisition software Cytflex, nuclei were gated for (upper right; isolate orange population). This population was further gated to isolate cell cycle stages of the nuclei (upper left; G1 = pink, S = green, G2 = blue). Cy3 intensities for the entire nuclei population (orange) and each cell cycle stage subpopulations (pink, green, blue) were determined.

Analysis of Invader probe targeting specificity with flow cytometer instrumentation

In order to determine the specificity of the Cy3 intensity values for Invader probe, **INV4** was chosen due to its ability to bind to chromosomal DNA under non-denaturing conditions.[4] Two controls were prepared to evaluate Cy3 intensity values. A Cy3-labeled

dsDNA probe with the same sequence as **INV4** but without any energetic hotspots (i.e., **INV4D**) was used to determine levels of non-specific binding, as negatively charged probes may bind postiively charged proteins or adhere to other cellular strucutures. High intensity of Cy3 fluorescence following incubation with **INV4D** would indicate non-specific binding that is unrelated to Watson-Crick binding, as **INV4D** is not energetically activated for dsDNA-recognition. Any observed Cy3 fluorescence with **INV4D** would need to be substracted when Invader probes are evaluated. A mis-match control Invader probe (**MM4**) was utilized as a negative control to evaluate base-pairing based non-specific binding. **MM4** differs in sequence at three positions relative to **INV4** (and the corresponding target region) and should, therefore, not bind to the target sequence if Invader-mediated binding is specific. Figure 3.2 show the fluorescent intensities of the nuclei population for each treatment.

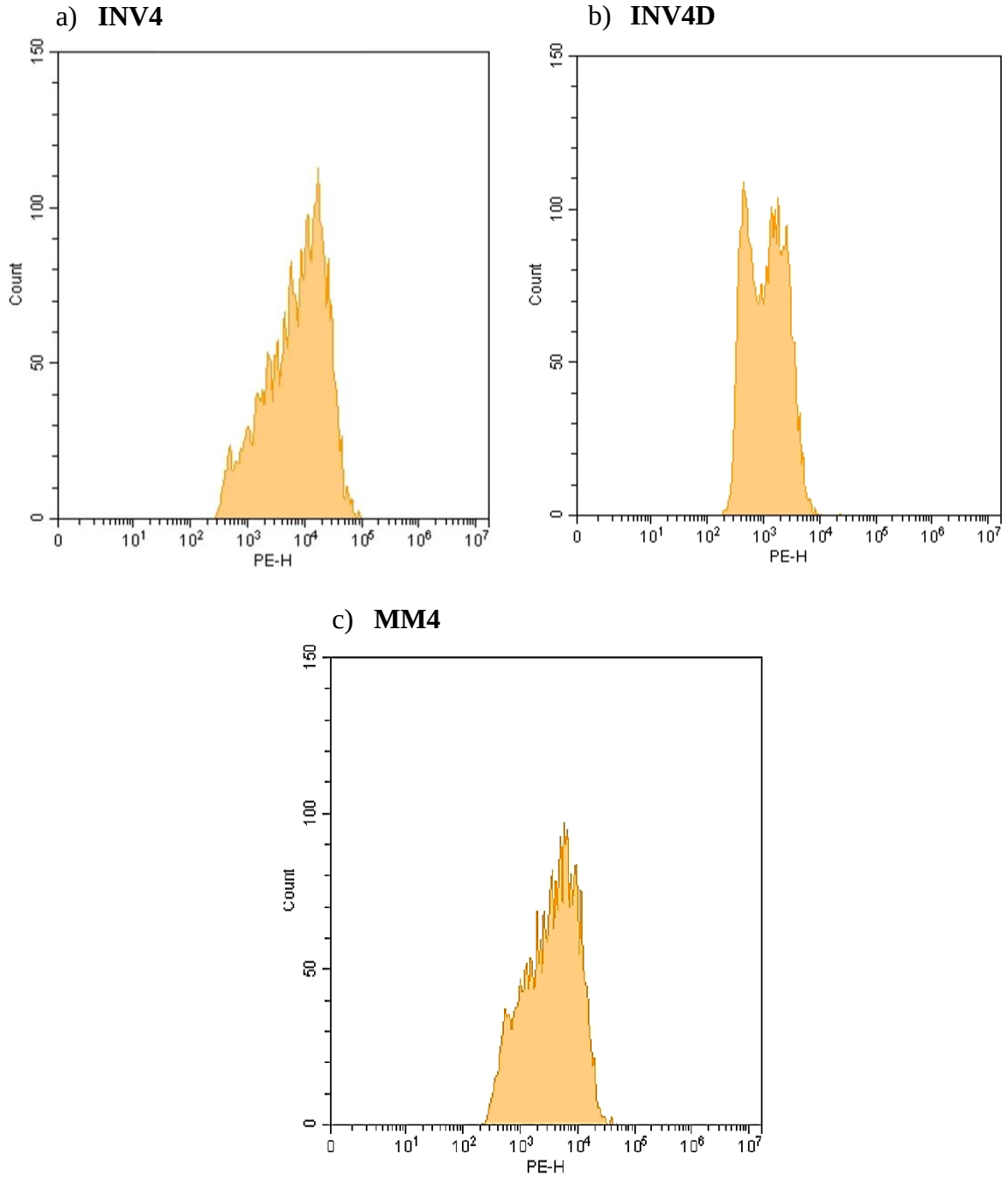


Figure 3.2. Probe Cy3 intensity (PE-H) values for nuclei populations (count) incubated with a) **INV4**, b) **INV4D**, and c) **MM4** utilizing the incubation conditions and gating strategy described in Figure 3.1.

Following flow cytometric analysis, mean intesity values for each fluorophore-labeled oligonucleotide **(INV4**, **INV4D**, and **MM4**) were determined from three separate nuclei populations including all interphase stages (G1, S, and G2) of ~3000 nuclei. The mean intesity

values were compared using one-way ANOVA tests to determine if there is a statisitically significant differences between the three groups (**INV4** vs **MM4** vs **INV4d**) (Figure 3.3). The analysis shows that incubation with **INV4** results in a significantly higher mean Cy3 fluorescence intensity than the two control groups. Thus, the observed Cy3 fluorescence observed with INV4 is, in all likelihood, due to specific binding to its corresponding DYZ-1 target region.

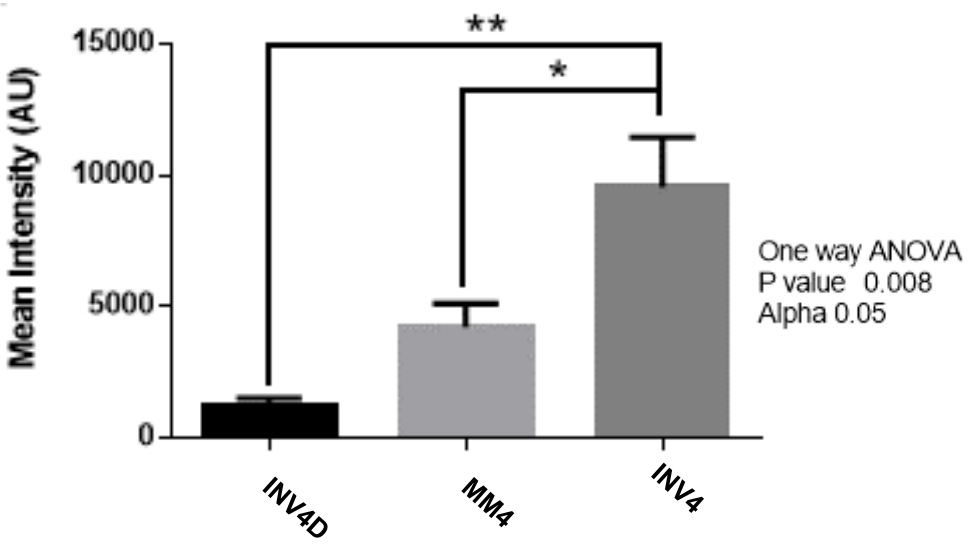


Figure 3.3. Comparison of Mean Intensities between Uumodified **INV4D**, mismatched **MM4**, and **INV4**. Mean intensity values for each probe were taken from three separate populations (~3000 nuclei) (example of mean intesnity values for one population shown in Figure 3.2) incubated with 30-fold fluorophore-labeled oligonucleotide excess in comparison target sequence. Statisitical significance was determine by a *p* value below the α value (0.05) (i.e. with a 95% confidence interval). There was no statistical significance between controls **INV4D** and **MM4**, while **INV4** was statisitically significant in comparsion to both controls (stars represent statistical significance between groups (i.e. one star = *p* value < 0.05, two stars = *p* value << 0.05)).

Having established the adequacy and differentiation of the two controls from **INV4,** the average fluorescent intensity of **INV4D** was subtracted from the mean intensities of **MM4** and **INV4** (i.e. values depicted in Figure 3.3). This allowed for statistical analysis between **MM4** and **INV4** to compare intensities resulting from specific and non-specific

Watson-Crick binding. Following an unpaired t-test analysis of the mean intensities of these two group, statistical significance was determined between **MM4** and **INV4** (Figure 3.4). This indicates that specific binding of Invader probe can be determined by flow cytometry. In future analysis to compare probes targeting different sequences, it might prove beneficial to subtract the intensity values of a mismatch control to give mean intensity values that reflect specific binding of Invader probe.

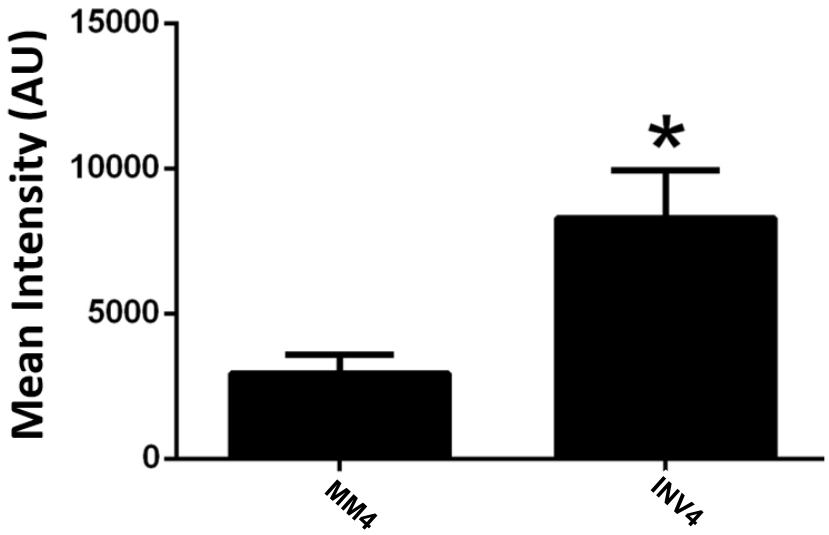


Figure 3.4. Comparison of Mean Intensities between mismatched **MM4** and **INV4** using an unpaired two-tailed t-test. Determination of incubation conditions and intensity values are described in Figure 3.2 and 3.3. Statisitical significance was determine by a *p* value below the α value (0.05) (i.e. with a 95% confidence interval). One star indicates significance between **MM4** and **INV4** intensities (*p* value = 0.03).

3.3 Conclusion

This study shows the potential use of flow cytomery to anaylze populations of nuclei incubated with Invader probes in order to detect sequences of chromosomal DNA. First, the nuclei preparation and Invader incubation protocol developed herein has yielded detectable nuclei populations (Figure 3.1). Nuclei can be analyzed as a whole population and/or as subpopulations (i.e. G1, S, G2) to determine Cy3 fluorescent intensity with our determined

gating strategy (Figure 3.1). Second, this strategy was successfully applied to determine the specific detection of chromosomal DNA by **INV4** through comparison of a mismatch control (Figure 3.4). Overall, this preliminary study sets the stage for future analysis of other Invader probes to detect specific sequences of chromosomal DNA utilizing this gating strategy and subsequent data processing to determine specific binding. Through the study of other Invader probes, correlations between cell behavior (i.e. phases, presence of other biomolecules, etc) and Invader probe binding may be drawn. Moreover, this may prove useful in determining efficiency of Invader probe binding by comparing intensity values of probes target different sequences and future cell delievery studies.

3.4 Acknowledgements

We would like to thank Dr. Ann Norton for her assistance with flow cytometric analysis and use of flow cytometer aquisition software. An additional thank you is extended to Saswata Karmakar for synthesis of Invader probes used herein.

3.5 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 (300 µL total) was added to permeabilize the nuclear membrane. After 10 min, the solution was centrifuged and the supernatant was removed. The pellet of nuclei was washed with 1 mL of cold 1x PBS and followed by an additional centrifuge/supernatant removal 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 the fixative solution, nuclei were suspended in a labeling buffer consisting of a double-stranded probe in PCR buffer (20 mM Tris, 100 mM KCl, pH 8.0). To yield optimal fluorescent signals, Invaders were added in 30-fold excess to its chromosomal target and nuclei were incubated for 3 h at 37.5 °C. Level of excess was calculated by determining the number of cells via a hemocytometer and multiplying by the number of target sequences per cell (i.e. **INV4** target sequence is present six times per cell). Invader probe concentrations were made to give 30 molecules of double-stranded probe to one molecule of target sequence. After incubation, he labeling buffer was then removed by centrifugation and removal of supernatant. The nuclei were rinsed with 1 mL of 1x PBS and incubated with 1 mL

of a DAPI solution (1 µg/mL) for 10 minutes. Follow a centrifugation 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 four 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) 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. Gates for determination of nuclei populations and their Cy3 intensity values (PE-H) were created using the Lasso tool with the analysis software. Events deemed to be nuclei were determined by size (high FSC intensity) and high DAPI fluorescence (high PB450-A). Acquisition software revealed Cy3 mean intensity values via the Statistics tool. Statistical analysis of intensities was conducted using GraphPad Prism 6 software.

3.6 References

1. C. Cui, W. Shu and P. Li, *Front. Cell Dev. Biol*., 2016, **4**, 1-11.
2. 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.
3. Rosner, K. Schipany and M. Hengstschläger, *Nature Protocols*, 2013, **8**, 602-626.
4. M. Wieser, G. Stadlera, E. Böhma, N. Bortha, H. Katingera, J. Grillariab and R. Voglauer, *Experimental Gerontology*, 2006, **4**, 230-235.

5. 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.
6. S. Roura, M. Fernández, E. Elchinova, I. Teubel, G. Requena, R. Cbanes, J. Lupón, A. Bayes-Genis, *Laboratory Investigation*, 2016, **96**, 1223-1230.
7. J. Perret, Y. Shia, R. Fries, G. Vassart and M. Georges, *Genomics,* 1990, **6**, 482-490.