Optimizing Invader Probe Architectures for Sequence-Unrestricted Recognition of dsDNA Targets

A Dissertation

Presented in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

with a

Major in Chemistry

in the

College of Graduate Studies

University of Idaho

by

Shiva P. Adhikari

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

Committee Members: Peter B. Allen, Ph.D.; Kristopher Waynant, Ph.D.;

Ching-An Peng, Ph.D.

Department Administrator: Ray von Wandruszka, Ph.D.

August 2021

Authorization to Submit Dissertation

This dissertation of Shiva P. Adhikari, submitted for the degree of Doctor of Philosophy with a Major in Chemistry and titled "**Optimizing Invader Probe Architecture for Sequence-Unrestricted Recognition of dsDNA Targets**," has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: _____ Date: _____

Patrick J. Hrdlicka, Ph.D.

Committee Members: _____ Date: _____

Peter Allen, Ph.D.

_____ Date: _____

Kristopher Waynant, Ph.D.

_____ Date: _____

Ching-An Peng, Ph.D.

Department

Administrator: _____ Date: _____

Ray von Wandruszka, Ph.D.

Abstract

Oligonucleotides capable of sequence-specific recognition of double-stranded (ds) DNA have tremendous potential as tools in the areas of diagnostics, genome editing, and molecular therapy. One of the envisioned advantages of gene-targeting modalities is the prospect for lower drug doses since genetic diseases often emanate from a single gene. Conventional DNA-targeting modalities include pyrrole-imidazole polyamides, triplex-forming oligonucleotides (TFOs), peptide nucleic acids (PNAs), and engineered proteins. While proof-of-principle has been demonstrated for these strategies, they exhibit shortcomings, e.g., TFOs require extended polypurine target regions, while PNAs need low ionic strengths for efficient dsDNA-recognition. The emergence of CRISPR/Cas, an RNA-guided protein-based DNA-targeting technique, shows considerable promise albeit being marred by significant off-target and delivery challenges. There is, hence, an unmet need for oligonucleotide-based chemical probes capable of recognizing mixed-sequence dsDNA at physiological conditions. Towards this end, our laboratory has designed so-called Invader probes, i.e., short DNA duplexes modified with one or more +1 interstrand zipper arrangements of intercalator-modified nucleotides such as $2'$-O-(pyren-1-yl)methyl-RNA monomers. These monomer arrangements result in a destabilized and energetically activated duplex, where each strand exhibits very high affinity towards complementary DNA regions since recognition results in highly favorable stacking interactions between intercalators and flanking base pairs. Invader probes enable recognition of mixed-sequence dsDNA regions with excellent specificity at physiological conditions. The double-stranded probes must dissociate easily for this approach to be effective. This renders GC-rich and long target regions challenging since the probes are high-melting. We aim to develop optimized Invader probe architectures that display efficient recognition of challenging regions. Specifically, the work in this dissertation describes three novel architectures; i) spermine-bulged Invader probes where non-nucleotidic spermine bulges are introduced in Invader probes to destabilize the duplex, ii) toehold Invader probes in which the probes are trimmed to shorten the double-stranded region of the probe to expose single-stranded overhangs that can be additionally modified with affinity-enhancing modifications, and iii) nicked Invader probes in which nicks are introduced in longer probes to facilitate dissociation. The ability of probes to recognize mixed-sequence dsDNA was evaluated using an electrophoretic mobility shift assay in which either a dsDNA or a hairpin was used as a model target comprising regions

complementary to the probes. Dose-response profiles were determined to compare the efficiency between different probes. Lastly, the probes were evaluated for their ability to recognize the corresponding *DYZ-1* target region in the context of fluorescence in situ hybridization (FISH) assays. Overall, this dissertation work advances DNA-targeting knowledge and will aid in the selection of the best probe architecture to improve DNA-targeting for future applications in molecular biology, biotechnology, and biomedicine.

Acknowledgments

I would like to express my deep appreciation and gratitude to Dr. Patrick Hrdlicka for always being supportive, pushing me forward to investigate and understand the basics of the experiments, contributing a good chunk of time to guide me to improve on my research, presentation and writing skills. If it weren't for his guidance and supervision, I would not be the oligonucleotide scientist I am today. I am also extremely grateful for my dissertation committee members, Dr. Kristopher Waynant, Dr. Peter B Allen and, Dr. Ching-An Peng for providing friendly support and guidance throughout my doctoral training. I'd like to express my gratitude to Dr. Richard Williams and Dr. Dan Stelck for always being accessible to address any issues regarding the research problems, work challenges and career counseling.

I am deeply indebted to chemistry department, University of Idaho for providing me an opportunity for my doctoral study. I would like to thank Dr. Ray von Wandruszka for continuous support and guidance as a Department Chair. I'm also grateful to Dr. Lee Deobald for his invaluable assistance in mass spectrometry. Many thanks to Deb Cissell for the support and assistance.

A very special thanks to Dr. Saswata Karmakar for synthesizing the Invader phosphoramidites. Additionally, I would also like to thank my fellow group members, Dr. Raymond G. Emehiser, Caroline Shepard, Michaela Brown, and Ibrahim-Al Janabi for being helpful in project collaborations and always extending a helping hand. Lastly, I cannot begin to express my thanks to all the members of the Chemistry department, who directly or indirectly helped me.

Dedication

To my parents - the driving force for who I've become now. To my brother and sisters - for believing in me. To my family, and friends, for their kind and unwavering support.

Table of Contents

List of Tables

List of Figures

adjusting the exposure. Nuclei were viewed at 60X magnification using a Nikon Eclipse Ti-S inverted microscope. The amount of **DYZ-NIP** was chosen based on optimization studies (Fig.

CHAPTER 1: Overview of DNA-targeting Strategies

1.1 Importance of DNA as a target

Conventional drugs mostly include small molecules or antibodies which have been utilized to target disease-associated proteins that are either expressed endogenously or by infectious organisms. However, the requirement for the shape and binding complementarity of these drugs vis-à-vis their targets introduces a risk for non-specific binding to unrelated biomolecules that may result in adverse effects. In addition, the limited druggable universe for small molecule drugs necessitates the development of alternative strategies, as only ~2% of the human genome is translated into proteins and only ~15% of these are druggable.[1] In order to expand the druggable universe and target diseases at their genetic origins, alternative therapeutic approaches utilizing oligonucleotide (ON) chemistries have been developed. Thus, chemically modified ONs have been designed to target disease-associated DNA and - especially - RNA sequences via specific base-pairing between the target strand and a chemically modified ON, thereby impacting the expression level of the nucleic acids. Currently, there are 14 FDA/EMA-approved RNA-targeting oligonucleotides available for the treatment of diseases,[2] which establishes that RNA-based oligonucleotide therapeutics are viable in the clinic. Alternatively, diseases can be envisioned treated at the DNA level by gene-targeting modalities. The central dogma of molecular biology highlights the importance of DNA, deoxyribonucleic acid, which carries the genetic blueprint needed for all known life forms. DNA can be replicated, or transcribed into RNA, which, in turn, can be translated into amino acids and proteins (Fig. 1.1). Next-generation sequencing has made it possible to routinely sequence and analyze a whole human genome within a short time-frame, allowing for the identification of sequences related to ailments, structural variation, or regulatory and functional elements at a genomic scale. Such sequences might constitute potential sites for gene targeting to modulate gene expression. Besides, with the knowledge of the whole genome, targets of high homology can be avoided to minimize non-specific bindings. Considering the recent technology developments and advancements in RNA therapeutics, there is an unmet need to develop a general approach to targeting DNA.
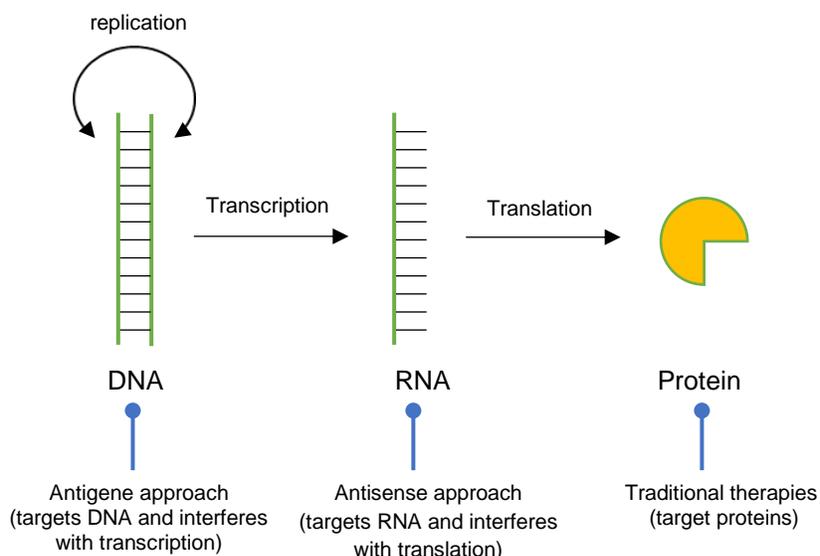
Figure 1.1. Illustration showing antigene, antisense, and protein targeting approaches.

1.2 Structure of nucleic acids

The basic structural units of nucleic acids must be understood to develop and design an approach for gene targeting. Briefly, nucleotides are the monomeric units of nucleic acids and consist of a nucleobase, attached to the 1'-position of a 5-membered furanose sugar ring via a glycosidic bond, and a phosphodiester linkage that connects two nucleoside units via 3' and 5'-positions to form a polymeric ON strands. The sugar-phosphate constitutes the backbone of the nucleic acid strands (Fig. 1.2). Two complementary strands can self-assemble into a right-handed antiparallel double-helical structure that twists around a helical axis with nucleobases being buried in the hydrophobic duplex core.[3,4] Depending on the nature of the 2'-substituent, nucleic acids are further classified as deoxyribonucleic acid (DNA), which has hydrogen at the 2'-down position of the sugar, and ribonucleic acids (RNA), which have a hydroxyl group. The glycosidic bonds that connect a nucleobase to the sugar ring in a nucleotide are not symmetrically opposed to each other in a base-pair[5] – which gives rise to major and minor grooves running along with the double-helical structure (Fig. 1.2).[3,4]
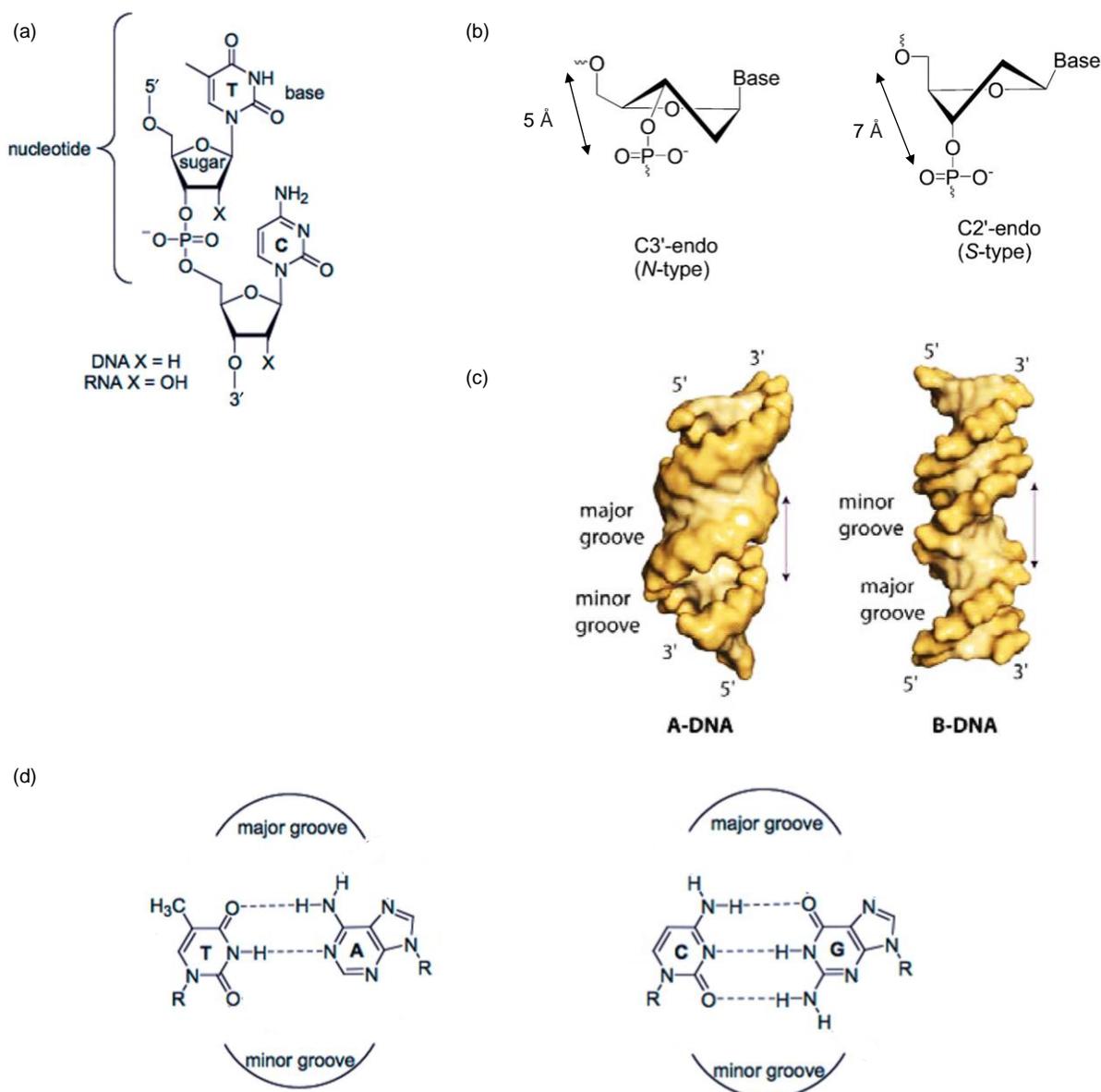
Figure 1.2. Structure of nucleic acid (a) structures showing sugar-phosphate backbone of a DNA molecule. (b) N-type and S-type of sugar puckering in nucleic acids. Reprinted with permission from (H. Kaur, B. R. Babu, and S. Maiti, *Chem. Rev.*, 2007, **107**, 4672–4697). Copyright (2007), American Chemical Society. (c) A-type and B-type double helix of nucleic acids showing major and minor grooves (d) Watson-Crick base pairing between nucleobases; adenine (A): thymine (T) and cytosine (C): guanine(G). Reprinted with permission from (M. Duca, P. Vekhoff, K. Oussedik, L. Halby and P. B. Arimondo, *Nucleic Acids Res.*, 2008, **36**, 5123–5138). Copyright (2008), Oxford University Press.

 The nucleic acid duplex is held together via supramolecular bonds i.e., non-covalent bonds, such as hydrogen bonds, and π-π stacking between the aromatic nitrogenous bases. The Watson-Crick (wc) base-pairing guided double-helical structures of nucleic acids exhibit conformational diversity to give rise to allomorphic groups that comprise right-handed A-type and B-type helices (Fig. 1.2).[6] The differences in these structures of double-helix mainly arise from different ring puckering of the sugar unit. In A-type helical structures, the sugar adopts a so-called C3'-*endo* (N-type) conformation, in which the C3' is above the plane described by the remaining ring atoms, while in B-type helical structures, the sugar adopts C2'-endo (S-type) conformations, where C2' is above the plane described by the remaining sugar ring atoms (Fig. 1.2).[3] A consequence of the S-type sugar pucker is that it influences some of the sugar-phosphate backbone torsion angles, which establishes the distance between the 5'-P and 3'-P as 7 Å, whereas, N-type sugar pucker impacts some of the sugar-phosphate backbone torsion angles which establishes 5'-P and 3'-P to lie in closer proximity with each other (5.9 Å). RNA:RNA duplexes adopt A-type helical geometry, characterized by wide and shallow minor grooves and narrow and deep major grooves. DNA:DNA duplexes adopt B-type helical geometry, characterized by wide and deep major grooves, and narrow and shallow minor grooves (Fig. 1.2). The major groove edge of a base-pair holds a large number of H-bond donors and acceptors compared minor groove edge.[7] The major and minor grooves of DNA duplex are extensively used to access the target biding site by many ligands, small molecules, and DNA-targeting probes.

1.3 Introduction of double-stranded DNA targeting probes

Double-stranded DNA (dsDNA) is a challenging target, both from a biological and chemical perspective. In cells, dsDNA is tightly wrapped around cationic proteins called histones, which compacts the DNA into chromosomes within the nucleus of the cell. Any molecule designed to target specific dsDNA sequences must be correctly distributed to the target tissue, cross the plasma membrane of target cells, and traverse the nuclear membrane to reach the nucleus. Throughout this process, the ON-based drug must be resistant to adverse physiological conditions such as nucleolytic enzymes. Assuming the ON-based drug reaches the nucleus, it must gain access to sequence-specific recognition elements that are buried deep within the duplex. The stability of wc base-pairing renders it challenging to construct probes capable of

disrupting the duplex. Various DNA-targeting approaches have been developed over the past thirty years, including minor-groove binding polyamides, major groove binding triplex-forming oligonucleotides (TFOs), and strand invading probes as peptide nucleic acids (PNAs), bisPNA, tail-clamp PNA (tcPNA), gamma PNA (γPNA), pseudocomplementary PNA (pcPNA) Zorro LNA, and more recently, RNA-guided protein-based DNA-targeting technology (CRISPR/Cas).

1.3.1 Minor-groove binding polyamides

The Watson-Crick base pairs contain H-bonding elements in the minor groove i.e., lone pairs of N3 of purines and O2 of pyrimidines (H-bond acceptor), and the 2-amino group of guanine (H-bond donor). These H-bond elements are recognized by minor-groove binding polyamides. The most extensively studied minor-groove binding polyamides are polymers of *N*-methylpyrrole (Py), *N*-methylhydroxypyrrole (Hp), and *N*-methylimidazole (Im), which contact complementary nucleobases from the minor groove through hydrogen bonds and stacking of the aromatic unit (Fig. 1.3).[8] The polyamides bind to the minor groove of dsDNA as a 2:1 ligand-DNA complex, where the two antiparallel polyamides are covalently connected by γ-aminobutyric acid (γ-turn), which results in improved affinity and specificity compared to unlinked dimers. The pairing rules of pyrrole-Imidazole polyamides were determined based on the interactions of polyamide dimer to dsDNA in a sequence-specific manner where, Im/Py pair discriminates GC from CG, while Py/Im pair discriminates CG from a GC. Similarly, Py/Hp discriminates AT from TA and, Hp/Py discriminates TA from AT. Moreover, the Im unit preferentially binds to G, while Hp recognizes a T over other nucleobases. However, Py does not have preferential recognition over A and C, which triggers the issue of specificity.[8,9] The minor groove hydrogen bonding elements of A and C both contain a hydrogen bond acceptor and look alike, thus when an H-bond donor motif is accessing it from the minor groove site, it cannot discriminate between A and C (Fig. 1.3). Moreover, Hp is unstable in both the solid and solution which limits its usefulness.

Polyamides have been reported to enter nuclei of live cells without any transfecting agents,[10] and have been investigated for several biological applications e.g., regulation of gene expression by inhibiting transcription binding factor[11], detection of dsDNA specific sequences (telomeres) using fluorescent derivatives, etc.[12] Minor-groove binding ligands need to have a

crescent shape and preserve an 'isohelical' conformation in order to fit the minor groove,[13] and thus minor groove binding polyamides are confined to recognition of short sequences as shape complementarity is lost with longer target sequences. [9,14]

(a)

Py/Im targets C•G

Py/Hp targets A•T

Hp/Py targets T•A

Im/Py targets G•C

(b)

G •• H •• C

C •• H •• G

A •• •• •• T

T •• •• •• A
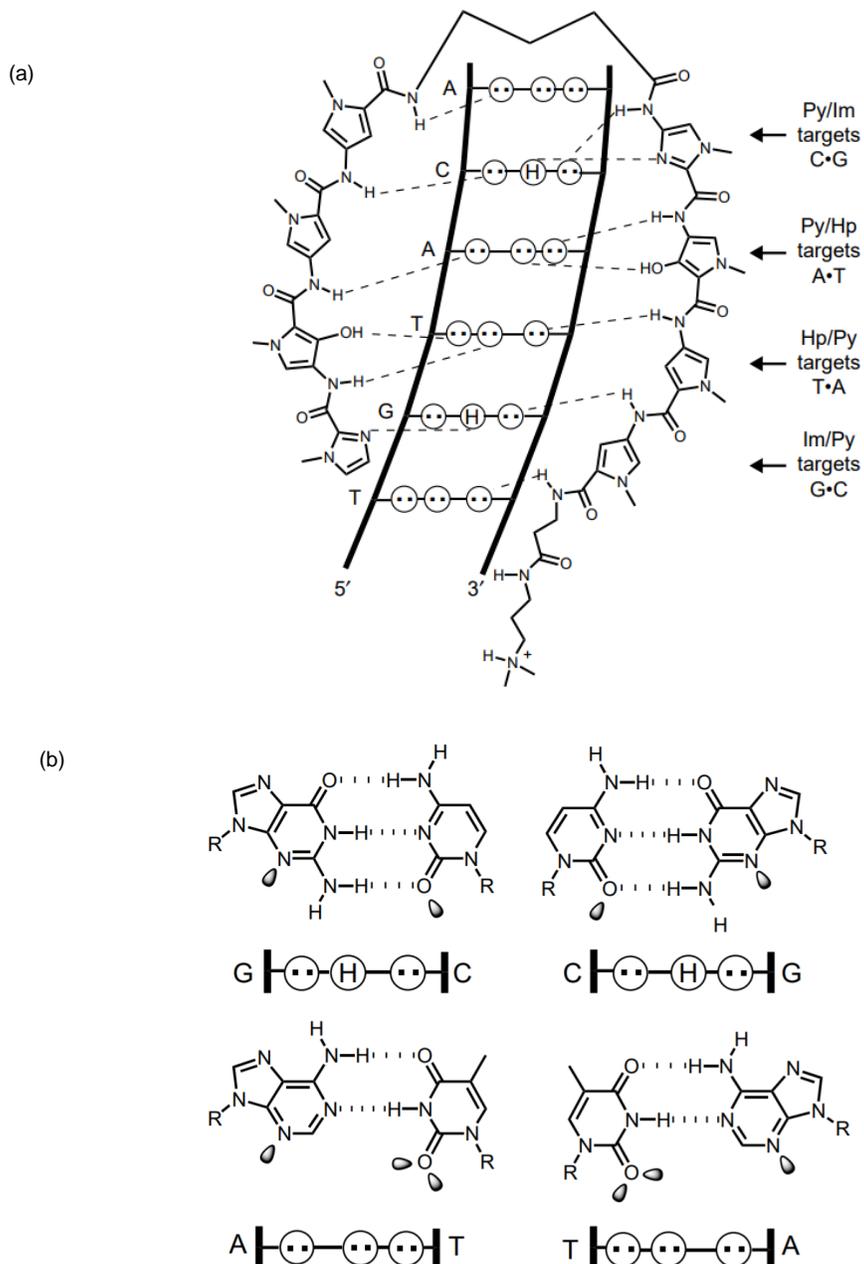
Figure 1.3. (a) Minor groove binding polyamide (ImHpPyPy-ImHpPyPy) binding to a 5'-TGTACA segment. The H-bonds between nucleobases and polyamides are shown in dash lines. (b) Minor groove hydrogen-bonding patterns of Watson–Crick base pairs. Circles with dots represent lone pairs of N(3) of purines and O(2) of pyrimidines, and circles containing an H

represent the 2-amino group of guanine. The R group represents the sugar-phosphate backbone of DNA. Electron lone pairs projecting into the minor groove are represented as shaded orbitals. Reprinted with permission from (P. B. Dervan and B. S. Edelson, *Curr. Opin. Struct. Biol.*, 2003, **13**, 284–299). Copyright 2003 Elsevier Science Ltd.

1.3.2 Triplex forming oligonucleotides

In 1957, the formation of a three-stranded polynucleotide molecule was reported, in which a poly-U bound a poly-rA in a poly-rA:U duplex.[15] Later in 1987, it was realized that short sequences of polypyrimidines can bind to polypurine regions of dsDNA by forming a triple helix, and subsequently, sequence-specific recognition of dsDNA was reported by the formation of the triple helix.[16] TFOs can recognize specific sequences by aligning themselves in the major groove of dsDNA targets[3] however, the binding of a TFO to a target duplex is thermodynamically weaker than duplexes, which is partly because of the charge repulsion fostered among three neighboring ON strands in a triplex. Binding of TFO with the double-stranded target occurs via so-called Hoogsteen or reverse Hoogsteen base-pairing in parallel (eg. pyrimidine TC-motif) or antiparallel orientation (eg. purine GA-motif) relative to a polypurine target region. Hoogsteen base-pairing offers an additional pairing geometry to wc-base pairing but is only possible from the major groove and requires one of the two strands of the duplex to have a long stretch of purines. The Hoogsteen bonding between a TC-motif TFO and a polypurine-rich target region takes place as T of TFO binds to an AT- base-pair of the dsDNA forming T-A-T, while C^+ binds GC forming C^+-G-C base triplexes (Fig. 1.4). Similarly, in GA-motif TFOs, G binds to GC base-pairs forming G-G-C base triplets and A binds to an AT-pair forming A-A-T base triplets. In order to increase target affinity, overcome charge repulsion, increase nuclease resistance, and prevent secondary structure formation (e.g., G-quadruplexes and GA-homoduplexes), chemical modifications were introduced in TFOs.

One of the limitations displayed by parallel binding TC-motif TFOs, the most extensively studied class of TFOs, is reduced target affinity at physiological pH because the N3 of cytosines of the TFO must be protonated to bind to the Hoogsteen face of G (Fig. 1.4).[17] Therefore, low pH (< 6) is required for optimal binding. To overcome this requirement of low pH, C of the TFO strand was replaced with 5-methyl-C (^{m}C) or pseudoisocytosine (J), which remains protonated at a broad physiological pH and, allows for triplex formation.[18] Further,

incorporation of modified monomers e.g., 2'-O,4'-C-aminoethylene-bridged monomers in pyridine motif (TC-motif) TFOs promotes triplex formation at physiological pH,[19] hence eliminating the requirement of low pH.
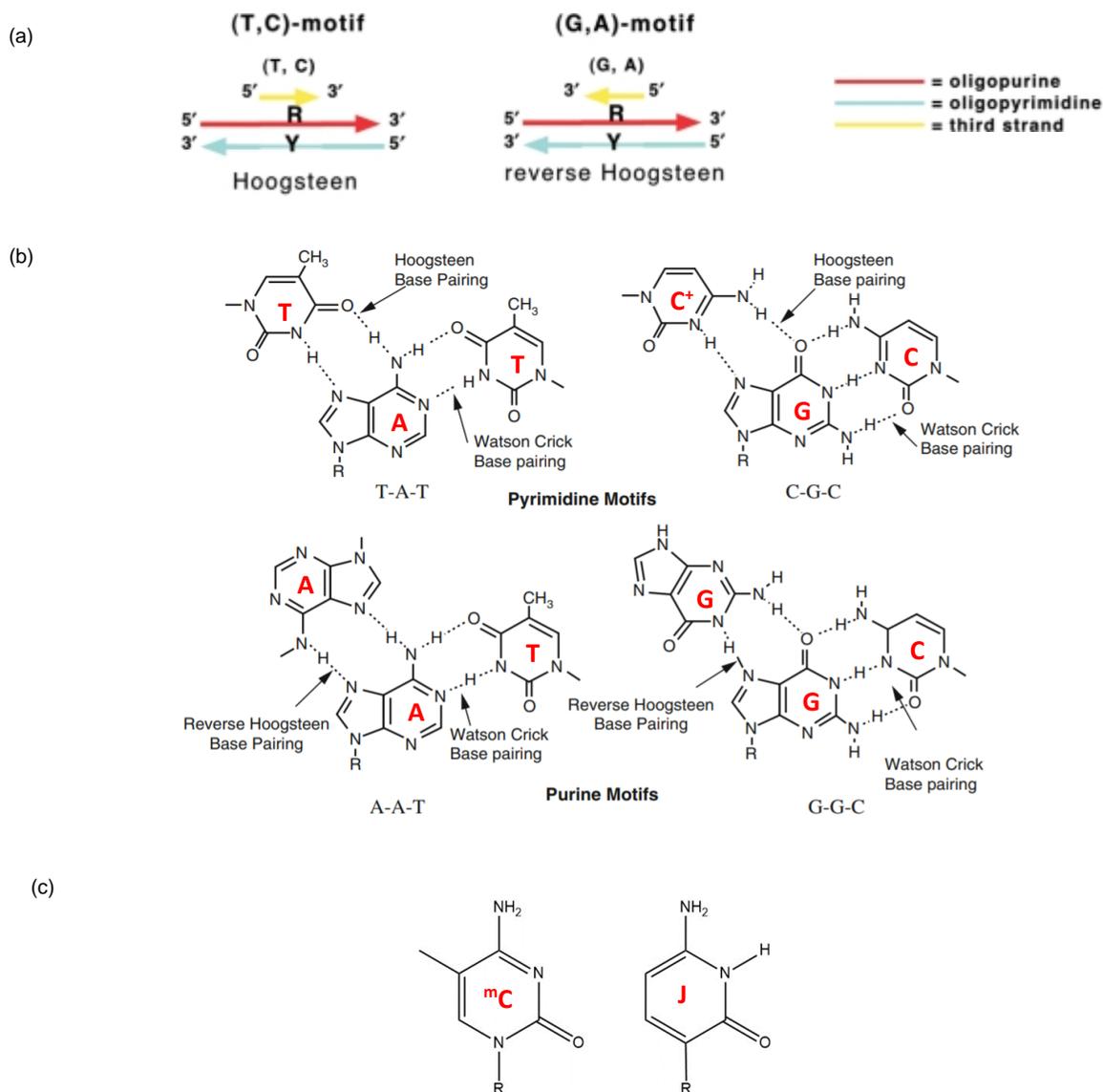
Figure 1.4. Pyrimidine TC-motif TFOs. (a) Illustration showing parallel Hoogsteen base pairing in pyrimidine TC-motif TFO. The figure also contrasts the Watson-crick and Hoogsteen base-pairing faces. Reprinted with permission from (M. Duca, P. Vekhoff, K. Oussedik, L. Halby and P. B. Arimondo, *Nucleic Acids Res.*, 2008, **36**, 5123–5138). Copyright (2008), Oxford University Press. (b) Motifs for triple helix formation- pyrimidine motif, where the third strand binds parallel to the purine strand of DNA via Hoogsteen bonds, and purine motif, where the third strand binds antiparallel to the purine strand via reverse Hoogsteen hydrogen

bonds. The canonical base triplets are shown for each motif. Reprinted with permission from (P. B. Dervan and B. S. Edelson, *Curr. Opin. Struct. Biol.*, 2003, **13**, 284–299). Copyright (2003) Elsevier Science Ltd.

Triplex formation is more stable with RNA-TFOs than DNA-TFOs, because the C3'-endo conformation adopted by N-type sugar rings results in less perturbation of the underlying duplex thus contributing higher affinity. The replacement of the phosphorodiester backbone with phosphothioates not only increases the stability of TFOs against nucleases but also displayed improved pharmacokinetic properties because of its interaction with serum proteins. TFOs are explored for various biological applications such as inducing triplex mediated genomic modification. TFOs have been also used to restrict helicase activity of DNA polymerase, detect specific targets using fluorophores, and modulate gene expression. Psoralen conjugated TFOs were employed to cause damage to the DNA of a mammalian cell (Psoralean crosslinks to thymines when exposed to UV light).[20] Statistically, polypurine-rich TFO binding sites are over-represented in the genome and are often located near promoters.[21] However, in order for stable Hoogsteen base-pairs to develop in triplex-forming probes, long polypurine runs are required, limiting the target scope.

1.3.3 Peptide nucleic acids

Peptide nucleic acids (PNAs) are synthetic mimics of DNA,[22-24] in which the negatively charged sugar-phosphate backbone of DNA is replaced by repetitive units of nucleobases on an uncharged and achiral *N*-(2-aminoethyl) glycine backbone (Fig. 1.5). While structurally very different from DNA, the PNA backbone maintains the necessary distance between nucleobases to ensure binding to complementary DNA strands. PNA can form wc base pairing with a complementary PNA-strand (PNA:PNA), which is extremely stable. PNA has proven advantageous relative to other probe chemistries as its structure renders it unrecognizable by nucleases and proteases. Moreover, the duplex formation between PNA and complementary DNA (PNA:DNA) is very stable due to minimal electrostatic repulsion between the strands as the PNA strand is neutrally charged, which also enables it to have a faster association with the target. PNAs bind to target dsDNA by forming triplexes or invading dsDNA duplexes by disrupting wc base-pairs of the target duplex. Different binding modes of PNA includes - i)

triplex formation, ii) triplex invasion, iii) single strand invasion, and iv) double duplex invasion (Fig. 1.5). Similar to TFOs, PNA can form triplexes through hoogsteen base-pairing with dsDNA targets but as with TFOs, the target must contain a polypurine-rich region (Fig. 1.5).[25,26] Polypyrimidine PNA can also undergo triplex invasion, when two appropriately designed PNA strands form a PNA$_2$:DNA triplexes (one of the PNA strand bind via wc base pairing to the polypurine rich target strand, while other strand binds through hoogsteen base-pairing with the same target strand) with a polypurine-rich strand of dsDNA target,[27] resulting in the displacement of the nontarget strand know as a displacement loop (D-loop). This binding mode has been perfected in bis-PNAs where two PNA strands are joined by a loop, enabling the Hoogsteen-arm of bis-PNA to first bind to a polypurine-rich region as a TFO, followed by the wc-binding arm folding back and binding with the existing wc-complementary strand of dsDNA, thereby displacing the non-target strand of DNA in the process (Fig. 1.5). The tethering of two strands of PNA minimizes the entropic cost upon binding.[28] The binding ability of bisPNA is improved by introducing positively charged groups (e.g., lysine) in the linker.[28] Further, conjugation of PNA with intercalators (e.g., acridine) further enhances dsDNA strand invasion efficiency.[29]

Along similar lines, tcPNA are bis PNA structures in which the wc arm is extended enabling recognition of a mixed-sequence region adjacent to a polypurine-rich region (Fig. 1.5).[30] The tail provides additional binding energy by decreasing the dissociation rate of the probe-target complex[30] and, enables the recognition of mixed-sequence (purines and pyrimidines) by the elongated tail, thus, broadening the scale of the accessible target region. Moreover, introducing positive charges along the tail enhances binding efficiency[30] and, tcPNA conjugated with positively charged cationic peptide has been used to inhibit transcription by phage T7 polymerase.[31] However, tcPNA still requires the presence of 8-10 polypurine-rich regions near the target site, which limits the potential target sites across the genome.[32] Furthermore, the tail displayed a poor mismatch discrimination ability.[30]

Single-stranded (ss) PNAs have also been shown to recognize dsDNA by forming wc base pairs with cDNA region and displacing the other strand (Fig. 1.5). The probe-target complex is formed via single-stranded invasion,[33] because neutral ssPNA has a high affinity towards cDNA as it will have minimal electrostatic repulsion. However, this process required denaturing

conditions,[34] the hybridization is less efficient and only favored within AT-rich target regions that are comparably unstable.[25,35]
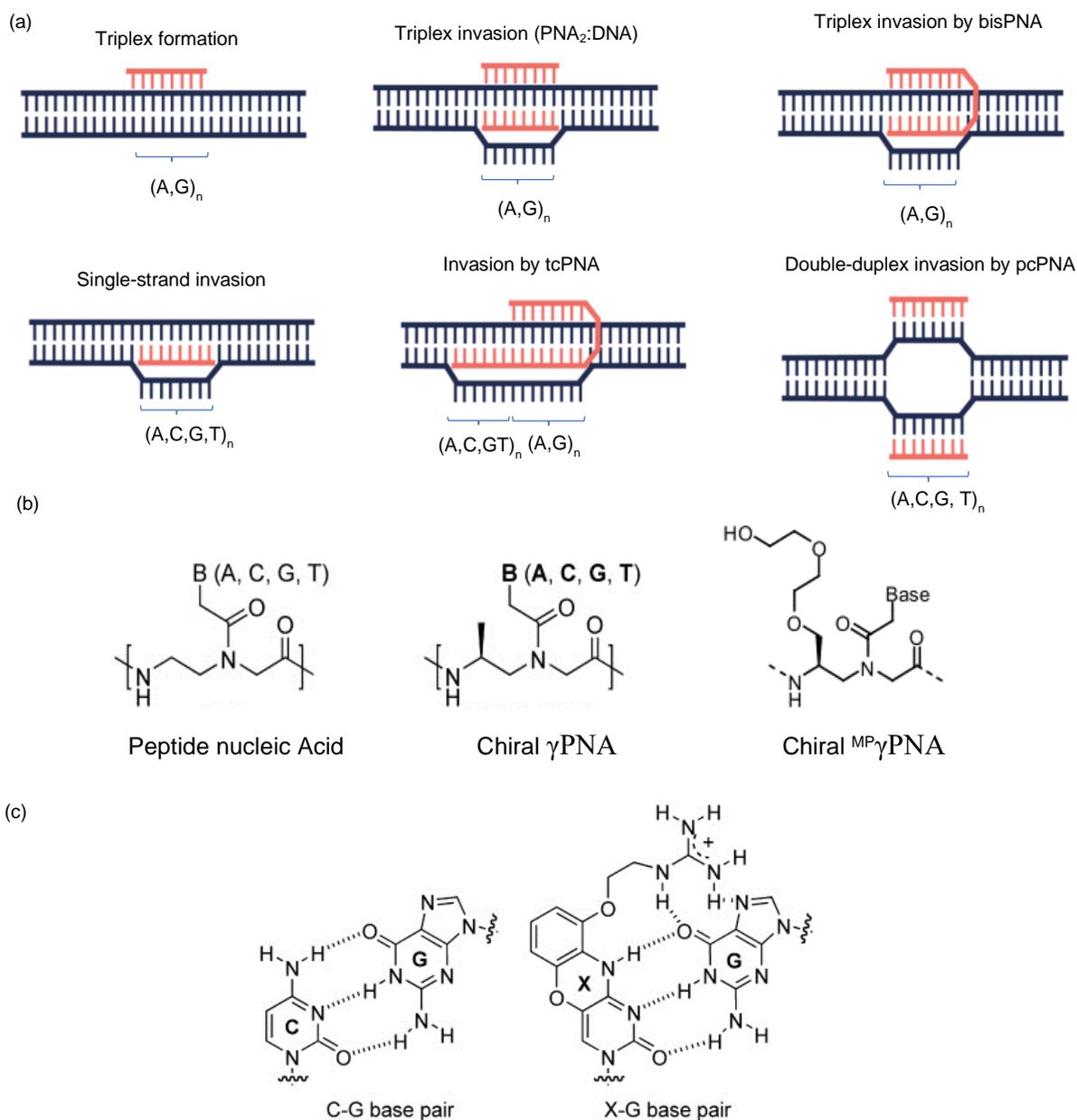
Figure 1.5. Peptide nucleic acids: (a) Different modes of targeting DNA by PNA. Reprinted with permission from (P. Muangkaew and T. Vilaivan, *Bioorg. Med. Chem. Lett.*, 2020, **30**, 127064.). Copyright (2020) Elsevier Science Ltd. (b) Structures of PNA, γPNAs, and pcPNA base-pairs. (c) C-G and X-G base pairing, Bold letters indicate γ-backbone modifications and, X = G-clamp. Reprinted with permission from (V. Chenna, S. Rapireddy, B. Sahu, C. Ausin,

E. Pedroso and D. H. Ly, *ChemBioChem*, 2008, **9**, 2388–2391). Copyright (2008) John Wiley and Sons.

To access the mixed-sequence target region with enhanced invasion efficiency, high-affinity chiral PNA has been developed, which contains a methyl group at the γ-position, termed as γPNAs (Fig. 1.5).[36,37] The single-stranded γPNA is pre-organized into a right-handed helix with a complementary chirality to that of ssDNA. This likely reduces entropic cost and increases the binding affinity for complementary ssDNA. Single-stranded γPNAs have proven useful for recognition of mixed-sequence dsDNA by single-strand invasion.[26] Initially, positively charged acridine was incorporated in short 10-mer γPNA, to invade mixed-sequence dsDNA. Acridine-linked γPNA displays significant binding and excellent mismatch specificity in low ionic conditions.[38] Later, unconjugated 10-mer γPNA, where cytosine was replaced by G-clamps, were designed to facilitate recognition of mixed-sequence double-helical B-form DNA. G-clamp is an affinity enhancing modification that forms three regular H-bonds plus the guanidino group associates to the 4th and 5th Hoogsteen bonding (Fig. 1.5). Thus, the G-clamp binds very strongly to a G nucleobase with five bonds. γPNA conjugated with two G-clamps displays complete recognition of target when used in 3-fold molar excess at low ionic concentration. However, at physiological concentrations of KCl, the recognition dropped significantly,[36] which could be likely because of minimized breathing of base-pairs at high salt concentrations. Subsequently, a longer 15 to 20-mer γPNA without any nucleobase substitution also displays invasion of dsDNA with excellent mismatch discrimination. Because of their low water solubility, inclination to form aggregates, and adhesion to surfaces and macromolecules including target nonspecifically, the applications of the first generation of γPNA investigated were restricted to certain chosen.[39] Modified γPNA monomers viz., MiniPEG-γPNA ([MP]γPNA) were synthesized to address this problem, as MiniPEGs are reported to increase water solubility.[40] [MP]γPNAs are mimics of γPNA,[41,42] where the methyl group at the γ-position of γPNA is replaced by (R)-diethylene glycol (MiniPEG or MP) moiety which still retain the preorganized structure to bind to DNA (Fig. 1.5). Using [MP]γPNA, sequence-unrestricted recognition of dsDNA was achieved ranging from 0-100 % GC content, and the aggregation-related nonspecific binding was diminished. [MP]γPNA exhibits a very high affinity, higher than γPNA without the need for any base modification.[43] On an *ex-vivo* exploration for gene editing in mixed-sequence dsDNA target, 0.8% gene editing frequency (low gene editing frequency)

was reported in mouse bone marrow cells, while in vivo treatment via IV injection resulted in the gene editing frequency of 0.1% without traceable toxicity.[44] $^{MP}\gamma$PNA displays high affinity and specificity to the cDNA targets and hence are an attractive class of molecular tools for gene targeting as they hybridize to their targets with the simple rule of wc base-pairing. However, precaution must be taken while designing the strands to avoid self-complementarity as these probes have high affinity and self-hybridization might result in the probes with very high melting temperature, which will negatively impact to target binding. Further, they require low ionic concentrations for optimal binding, where the target dsDNA is more relaxed, and, the enantioselective synthesis of the chiral monomers is challenging.[45]

Pseudocomplementary PNA (pcPNA) are double-stranded probes that can invade dsDNA to form double-duplex invasion complexes (Fig. 1.6), where the wc base-apiring in target dsDNA is disrupted and both strands of the probes recognize the cDNA strands through wc base-pairing forming two probe-target duplexes.[26] The driving force is the result of a destabilized double-stranded probe due to steric clashes between the 2-amino of 2,6-diamino purine (D) and the 2-thio of 2-thiouracil (U^s) which replace A and T (Fig. 1.6). In contrast, the base-pairing between D:U^s and U^s:A in probe-target duplexes is stable and thus the probe strands display increased affinity towards cDNA.[32] The higher stability of the probe-target recognition complex is attributed to two factors: i) the increased stacking interaction of 2-thiouracil with adjacent nucleobases promoted by the sulfur in the 2-position, and ii) formation of three H-bonds between D- U^s base-pairs rendering it more stable than A-T base pairs which form only two H-bonds.[46] While considering the effective design of pcPNA probes, D- U^s can effectively replace A-T, but presently effective replacement of G-C with pseudocomplementary base-pairs is unavailable. Attempts have been made to find suitable pseudocomplementary base-pairs to replace G-C e.g., N6-methoxy-2,6-diaminopurine:N4-benzoylcytosine (K*-CBz) pcbase pairs, where K*-and CBz recognizes a C and a G respectively (Fig. 1.6). As per the requirement to form pcPNA base pairs, there would be steric clash between the overextending benzoyl group of CBz and methoxy group of K*, rendering K*-CBz base pairs unstable. However, the K*- C and G-CBz base pairs, formed in probe-traget duplexes, are unstable. Further, K* failed to display mismatch specificity over C and T, pairing equally with both. Therefore, the mixed dsDNA target sequence generally must have a low GC content in order for pcPNA to recognize efficiently.
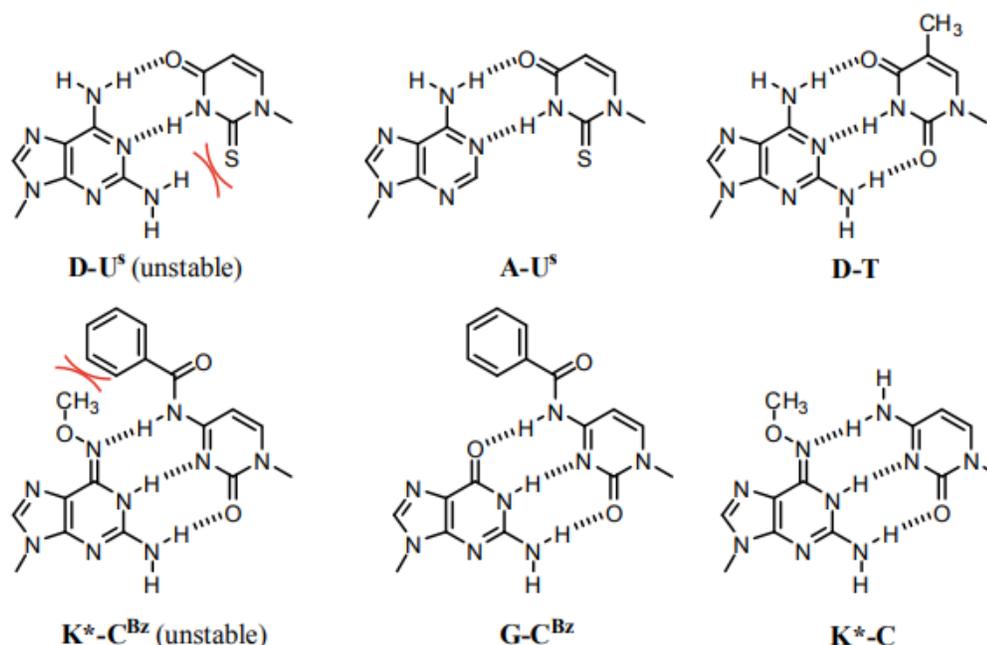
Figure 1.6. Pseudocomplementary base pairs. Reprinted with permission from (P. Muangkaew and T. Vilaivan, *Bioorg. Med. Chem. Lett.*, 2020, **30**, 127064.). Copyright (2020) Elsevier Science Ltd.

Alternatively, an amalgamation of pcPNA approach with chiral (D) PNA monomers was explored, where a positive charge was introduced in the backbone to promote invasion.[47] Here, positively charged lysine was added to form a positively charged chiral backbone that would create electrostatic repulsion between two positively charged probe duplexes and electrostatic attraction between PNA-DNA duplexes. Positively charged chiral pcPNA modified with L-lysine lacks recognition efficiency while D-isomers promote recognition even in GC-rich sequence context by destabilization of probe duplex and stabilization of PNA-DNA duplexes.[47] Later, pcPNAs were shown to facilitate dsDNA in physiological salt concentration under molecular crowding conditions.[48] PEG was used in the reaction buffer to mimic the molecular crowding conditions of the cells which contains a myriad of biomolecules. Recently, an ss-pcPNA strand was reported to efficiently invade dsDNA at 25 - 50 °C when the strand was partially modified when D and U^s in place of A and T, which could be attributed to higher stability of probe-target duplexes fostered by D-T and A-S base pairs.[46]

When dsDNA is targeted by ss-probes, it forms a D-loop in the recognition complex, while the benefit of a double-stranded probe is that it avoids the formation of a D-loop and thereby increases mismatch discrimination due to stringency clamping effects.[49,50] The binding specificity is the result of multiple influences including stringency clamping effects, i.e., greater stability differences between matched vis-à-vis mismatched recognition complexes seen with structured probes, avoidance of energetically unfavorable formation of multiple double-stranded segments with mismatched base-pairs.[51] Overall, the invasion of dsDNA targets by pcPNA is generally only possible at low ionic strength and elevated temperatures,[25] and are very difficult to synthesize compared to nucleoside building blocks.[25]

1.3.4 Locked Nucleic Acids (LNAs)

LNAs are conformationally restricted nucleotides that contain an oxymethylene group that links the 2'- and 4'- positions (Fig. 1.7).[52] Not only does the incorporation of LNA monomers improve the nuclease stability of ONs, but its preorganized conformation also renders the ONs with a significantly improved affinity towards the complementary target sequences. The oxymethylene linker constrains the sugar, forcing it to adopt N-type sugar puckering or, more simply, it locks the sugar ring in 3'-endo conformation,[52] which is structurally similar to RNA. Because of the conformational restriction (entropy-driven) of these LNA-modified ONs, they exhibit exceptionally high affinity and thermal stability toward RNA/DNA targets..[53] LNA-based oligonucleotides (LNA-ONs) hybridize ssDNA/ssRNA efficiently through wc base pairing. Incorporation of a single LNA monomer results in stable probe-target duplexes which display increases in T_m of up to 8 °C against DNA and up to 10 °C against RNA.[53] Alternately, LNA-modified oligonucleotides have been used as TFO (LNA-TFO) to form a triplex against a polypurine dsDNA duplex via Hoogsteen bonding. Incorporation of LNA monomers in TFOs increases the stability of the triplex. TFOs modified with an LNA monomer at the central position formed a triplex which was stabilized by more than 10 °C. The stability of the triplex increases with increasing LNA content in TFO, up to 5 °C per modification.[53] Along these lines, α-L-LNA, the α-anomer of enantio-LNA (L-LNA) with the inverted stereochemistry at C2', C3', and C4' compared to parent LNA, exhibits very high stability while forming triplexes at physiological pH conditions (pH 6.8).[53] Moreover, bisLNA, which contains a Hoogsteen arm and a wc arm tethered together by a linker, displayed 30% invasion of supercoiled DNA

duplexes, providing proof-of-concept for potential applications of bisLNA for interfering cellular genes.[54] The analogous LNA-TFO was unable to bind to the target at all, which underlines the importance of bisLNA structure. The optimal invasion of supercoiled DNA by bisLNA was achieved in physiological pH but necessitate lower salt concentrations.[54] Further, replacement of some LNA monomers for 2'-glycylamino-LNA enabled the construction of bisLNA featuring a positive charge,[55] which resulted in successful targeting of plasmids inside a bacteria.[56]
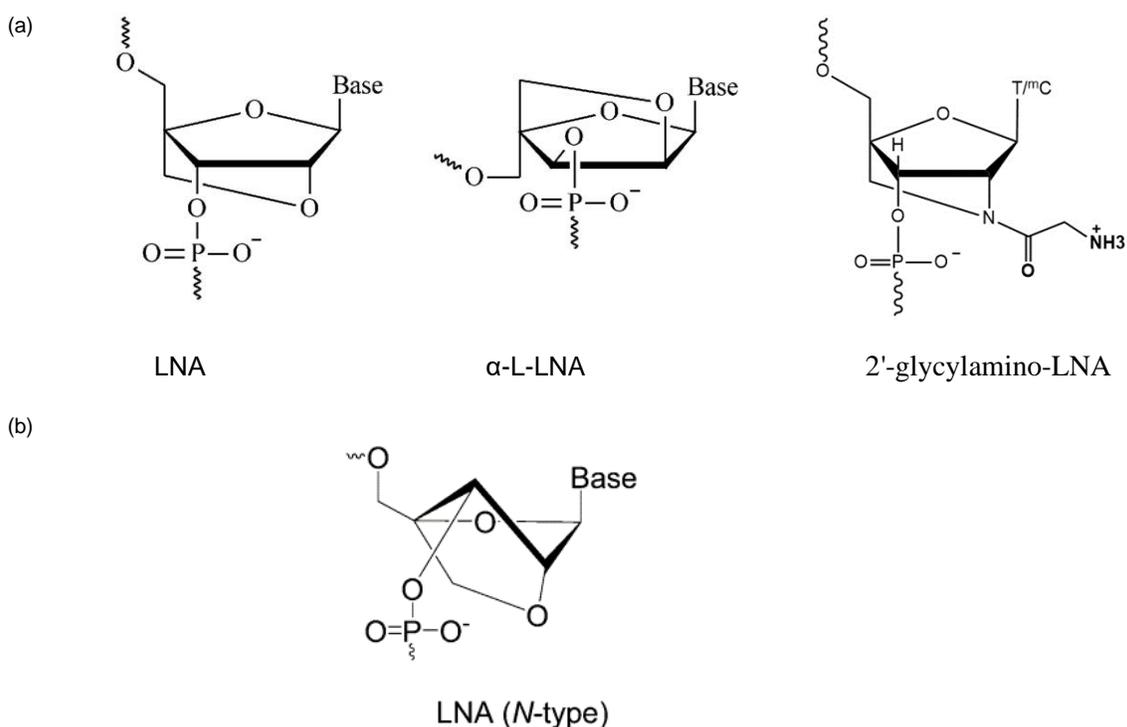
Figure 1.7. (a) Structure of LNA and its analogs, (b) Molecular structure of locked nucleic acid (LNA), which shows the locked C3'-endo sugar conformation. Reprinted with permission from (H. Kaur, B. R. Babu, and S. Maiti, *Chem. Rev.*, 2007, **107**, 4672–4697). Copyright (2007), American Chemical Society.

LNA chemistry is compatible with other monomers e.g., 2'-*O*-Methyl-RNA, phosphorothioate, DNA, RNA, and phosphodiester linkages. So, LNA-modified oligonucleotides have been explored as antisense and antigene targeting agents, where fully-modified LNA-oligonucleotides or partially modified mixmers are used to tune affinity, stability, and permeability of the probes in the biological system. Locked nucleic acids have been explored

in the regulation of gene expression,[52] gene silencing,[57], and RNA targeting. However, a strand of LNA-modified ON has a very high affinity against a complementary LNA-ON, which should be taken into consideration while designing fully modified LNA-ON or LNA mixmers to avoid the problem of self-hybridization.

1.3.5 Zorro-LNA

Zorro-LNAs are LNA-based constructs, consisting of two highly LNA-modified strands that form a short dsDNA region with two single-stranded overhangs.[58] Initially, the design consisted of a 7-mer dsDNA region and long 3'-overhangs regions which would form wc base pairing with complementary target strands (Fig. 1.8). The first generation of Zorro-LNA was designed in such a way that one of the two 3'-overhangs strands of the probe would hybridize via wc binding with one strand of the target while the other 3'-overhang strands of the probe bind with another strand of the probe via wc base pairing, the binding regions lying adjacent to each other. The resulting recognition complex contains two ss-loop. These Zorro-LNAs inhibit RNA polymerase-dependent transcription from plasmids *in vivo*[59] and induces sequence-specific gene silencing in mammalian cells.[58] One of the pros of Zorro-LNA is that it supposedly allows for sequence-unrestricted dsDNA-recognition.
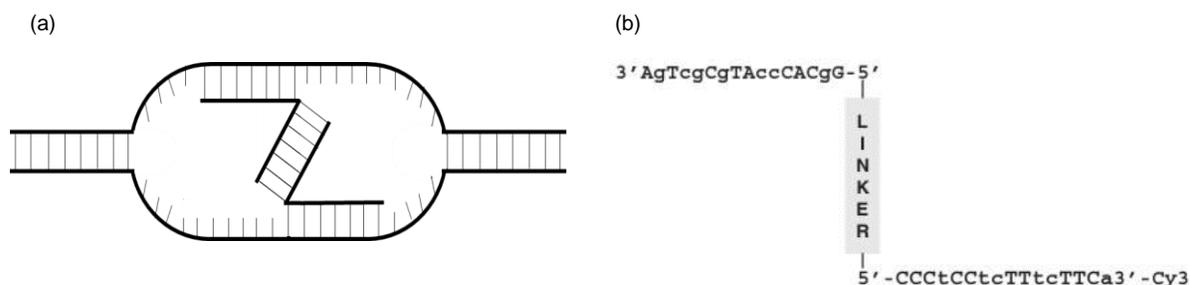
Figure 1.8. (a) Illustration of the first generation of Zorro LNA forming a recognition complex, and (b) second generation of Zorro-LNA construct where two LNA strands are linked together by a linker. Reprinted with permission from (E. M. Zaghloul, A. S. Madsen, P. M. D. Moreno, I. I. Oprea, S. El-Andaloussi, B. Bestas, P. Gupta, E. B. Pedersen, K. E. Lundin, J. Wengel and C. I. E. Smith, *Nucleic Acids Res.*, 2011, **39**, 1142–1154). Copyright (2011), Oxford University Press.

The two strands of the probes require a pre-annealing step for hybridization to create the Zorro-LNA construct. The use of two high affinity enhancing strands in the probe design could induce intramolecular binding among the bases in the overhangs and the linker region. The second generation of Zorro-LNA viz., ssZorroLNA was constructed to avoid such disadvantage, by replacing the 7-mer dsDNA region with non-nucleotide linkers of various lengths and hydrophobicity (Fig. 1.8).

The ssZorro-LNA displayed enhanced dsDNA invasion efficiency and kinetics vis-à-vis the first generation constructs.[60] The linker length and type donot contribute significantly to invasion efficiency and kinetics of probes, except for long alkyl or aromatic hydrophobic linkers that reduces binding efficiency and kinetics. This detrimental effect fostered by hydrophobic linkers were attributed to form potential aggregation by such highly hydrophobic molecules. In the same study, the binding mechanism of ZorroLNA was explored. The binding of ZorroLNA to dsDNA target gives rise to ss-stretches, or loop, in the recognition complex which indicates a double-stranded binding mechanism for ZorroLNA hybridization. However, formation of a triplex with the target by one of the arm of ZorroLNA was also reported in the same study. In this regard, a study involving the exploration of binding modes might enlighten the controversies of triplex or strand invasion binding modes. Further, high affinity-enhancing LNA-modified overhangs may self-hybridize, if they contain self-complemantary regions impeding invasion of dsDNA.

1.3.6 CRISPR/Cas9

CRISPR/Cas (Clustered Regulatory Interspaced Short Palindromic Repeats) is an archael/bacterial immune system that has recently been adapted for eukaryotic gene-targeting applications.[61] It consists of a guide RNA that matches the target gene and a CRISPR associated (Cas) protein, which is an endonuclease that breaks double-stranded DNA.[63] CRISPR/Cas9 is an easily programmable enzyme technology that cleaves dsDNA in a eukaryotes cell (Fig. 1.9).[62] Primarily, CRISPR/Cas9 was discovered as a defense mechanism in prokaryotes such as bacteria and archaea against viruses and phages. These prokaryotes develop a cellular memory by incorporating DNA sequences that are identical to infectious agents. All of the CRISPR/Cas systems follow three steps for the defense mechanism (Fig. 1.9) - i) adaptation where foreign nucleic acid spacers are incorporated into bacterial DNA, ii) expression where

transcription and processing produce CRISPR RNAs (crRNA) and, iii) interference where the genetic material from the invading organism are cleaved by crRNA-guided Cas enzymes.[64] Originally the CRISPR/Cas9 system utilized two natural RNA strands to form crRNA:tracrRNA complex, which recruits a single Cas9 endonuclease protein to form active ribonucleoprotein (RNP) complex.[64] Cas9 proteins are RNA guided endonucleases that can cleave the DNA strand of DNA-RNA duplexes formed in the Cas RNP.
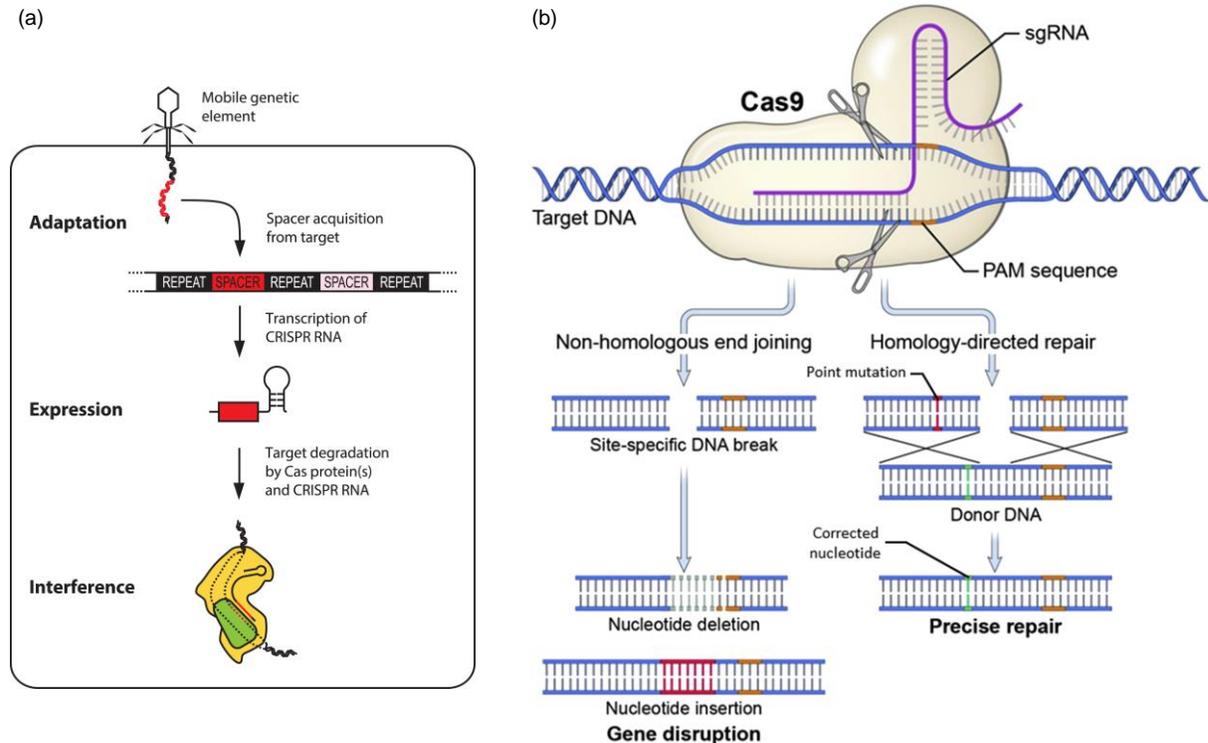
Figure 1.9. (a) Steps for the defense mechanism of CRISPR/Cas systems, (b) DSB, NHEJ, and HDR. Reprinted with permission from (J. R. Guitart, J. L. Johnson and W. W. Chien, *J. Invest. Dermatol.*, 2016, **136**, e87–e93.). Copyright (2016) The Authors. Published by Elsevier, Inc. on behalf of the Society for Investigative Dermatology.

CRISPR/Cas system has been adapted for biotechnological uses in eukaryotes by fusing the crRNA and tracrRNAs into a single guide RNA (gRNA), which directs RNP to search for complementary sequences in foreign DNA. Upon target recognition, double-stranded break (DSB) at specific sites in a genome are initiated thus allowing for gene editing by the addition of new genetic material. Eukaryotic cells are known to have the ability to repair DSB by employing one of the two DNA repair pathways- i) Non-Homologous end joining (NHEJ) or

homology directed repair (HDR).[64,65] NHEJ is error-prone as it causes random insertion, deletion, or substitution of base-pairs in DSB sites introducing mutations in DNA. However, the HDR pathway allows for precise repair of DSB sites in the presence of a homologous repair template or donor DNA. HDR repair method provides the advantage of performing precise gene modifications, knock-in, deletion, correction, or mutagenesis.[64]

Genome editing necessitates the use of gRNA and the Cas protein. Cas protein and gRNA can be produced as DNA plasmids, RNA, or RNA/protein complexes and delivered into cells by viral and non-viral means. The CRISPR/Cas9 technology has been investigated for zygote editing with the goal of eradicating a genetic illness from an entire family. This was accomplished by inserting CRISPR/cas9 components into a zygote embryo, which results in the modification of a gene in all cells of that organism. Mice with a mutated Crygc gene that causes cataracts were rescued (i.e. back to normal vision) via homology-directed repair using CRISPR/Cas9 with minial off-target effects.[66] The CRISPR/cas9 system was also found to facilitate the clearance of the intrahepatic HBV template to cure hepatitis B virus (HBV) infection.[67] CRISPR/Cas9 based gene therapy has also been applied *ex vivo* for treating β-thalassemia and HIV infections.[68]

By taking advantage of cellular DNA repair pathways, CRISPR-Cas9 can be employed to generate targeted mutations and insert a sequence of interest in DSB sites, however, creating DSB at a site of interest is challenging because of deleterious mutations. CRISPR-Cas9 is beneficial if the DSB is repaired through the HDR repaired pathway, which was reported to 25% of total repair, while the remaining 75% repair may produce deleterious mutation. Moreover, a high frequency of off-targeting activity ($\geq$50%) was observed.[69] Lastly, along with its specificity, cellular/nuclear delivery remains challenging for *in vivo* applications.[70]

1.4 Overview of Invader probe approach

While recognition of dsDNA has been demonstrated for the aforementioned strategies, they do display several shortcomings - e.g., most of the TFO approaches requires low pH, most of the PNA approaches necessitate AT-rich target region and/or low salt conditions for optimal binding, while high-affinity probes like chiral PNA and LNA retain potential to form self-complementary duplexes. Most of the approaches either require a non-physiological condition

(low pH, low salt conditions) or display sequence consideration while designing the probes (e.g., TFO required polypurine target, affinity enhancing probes should be avoided with self-complementary regions) for optimal binding which limits the targetable sequences in biological conditions.

To address the need for a rapid, efficient, and site-specific DNA-targeting approach at physiological conditions, our laboratory recently introduced so-called Invader probes for recognition of mixed-sequence dsDNA targets. The double-stranded Invader probes are modified with one or more +1 interstrand zipper arrangements of intercalator-modified nucleotides (also known as energetic hotspots),[71] which activates them for dsDNA-recognition because it offers highly labile probe duplexes that induce excellent affinity towards cDNA targets.[72]. This particular monomer arrangement forces two intercalating moieties to compete for the same inter-base-pair region, resulting in localized unwinding[73] that leads to destabilization of the probe duplex[74] because the neighbor exclusion principle – stating that intercalation is anti-cooperative at adjacent sites[75] – is violated.[76] In contrast. each individual strand exhibits high affinity towards complementary target DNA strands because duplex formation results in highly favorable stacking interactions between the intercalators of an Invader strand and the adjacent flanking base pairs.[77,78] Thus, the driving force for recognition of complementary dsDNA regions via double-duplex invasion is due to the greater stability of the two probe-target duplexes in the recognition complex (i.e., between individual Invader strands and complementary regions) relative to the Invader probe and dsDNA target duplexes (Fig. 1.10). Invader probe strands are synthesized via solid-phase ON chemistry and can be used to target mixed sequence dsDNA target regions. For instance, Invader probes have been used to recognize mixed-sequence target regions on Y-chromosomes or telomeric DNA in fixed nuclei from male bovine kidney cells, [51,77] under otherwise non-denaturing physiological conditions.[79] Moreover, Invader probes were used for sequence-unrestricted detection of dsDNA sequences specific to food pathogens.[80] This underscores the potential applications of Invader probes in molecular biology, diagnostics, and therapeutics.

Optimization of Invader probes for enhanced dsDNA recognition have been processed via two tactics – i) optimization of building blocks ii) optimization of probe architecture. Over the years, a series of intercalator-modified nucleotide monomers have been studied to identify building

blocks that are straightforward to synthesize, provide access to nucleobase monomers (analogs of A, G, C, and, U/T), and facilitate optimization of hotspots content to improve dsDNA targeting efficiency and specificity.[73,81] Once the optimal monomer has been identified, the probe architecture then can be tuned for enhanced dsDNA targeting efficiency of the probes. Exploration of alternative probe architecture has been done e.g., introducing non-nucleotidic bulges in the backbone,[82,83] using phosphorothioated backbone,[71] combining pseudocomplementary base-pairing with Invader approach, and the developing so-called chimeric Invader probes.[84, 85]
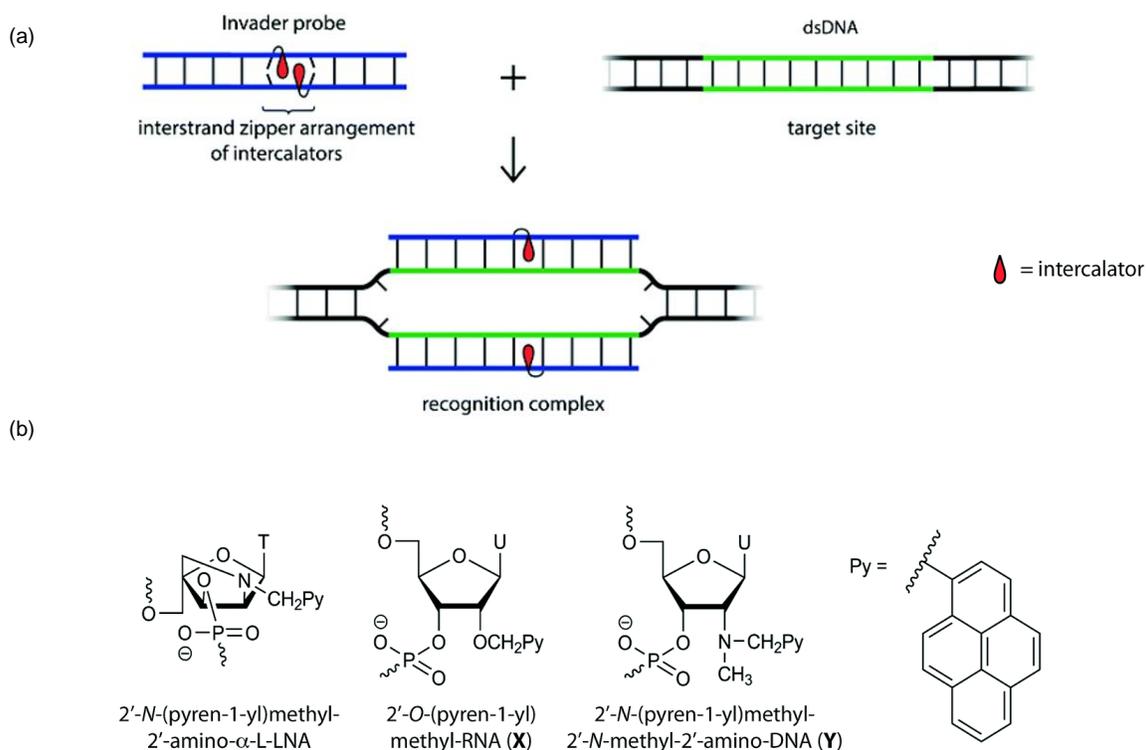
Figure 1.10. (a) Illustration showing Invader probes recognizing dsDNA target by double-strand invasion. Reprinted with permission from (S. Karmakar, T. Horrocks, B. C. Gibbons, D. C. Guenther, R. Emehiser and P. J. Hrdlicka, *Org. Biomol. Chem.*, 2019, **17**, 609–621.) Copyright (2019) Royal Society of Chemistry. (b) Structure of first and second-generation building blocks. Reprinted with permission from (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.) Copyright (2015) Royal Society of Chemistry.

1.4.1 *First and second-generation Invader building blocks.*

The first-generation of Invader probes consisted of N2'-pyrene-functionalized 2'-amino-α-L LNA monomers (Fig. 10), which are locked in a C2'-endo conformation.[72] The building blocks are restricted in a beneficial conformation, resulting in enhanced affinity towards complementary DNA strands as the intercalators are preorganized to intercalate and engage in π-π stacking with flanking nucleobases. However, due to the cumbersome synthesis of first-generation building blocks (more than 18 steps), alternative monomer chemistries with analogous properties were explored. Two types of building blocks, i.e., 2′-*N*-(pyren-1-yl)methyl-2′-*N*-methyl-2′-RNA and 2'-*O*-(pyren-1-yl)-methyl RNA, emerged as promising second-generation building blocks, despite being more conformationally flexible than the first-generation monomers given the absence of the O2',C4'-azamethylene linker (Fig. 1.10). [71,73] The influence of the length of linkers, as well as intercalator size, connectivity, and substitution pattern, was evaluated to discover optimized conditions that favor intercalation.[81,86,87] The introduction of short linkers in building blocks displayed stabilizing effects in the resulting probe-target duplexes vis-à-vis longer linkers, likely because of the entropic penalty associated with intercalation of intercalators connected via longer linkers. Moreover, larger intercalators, such as perylene and coronene, resulted in more stable Invader probes,[86] which is not desirable as a labile prone duplex is a prerequisite for dsDNA-targeting. In a study conducted to explore dsDNA recognition efficiency of Invader probes modified with monomers with varying intercalator size, probes based on the O2′-pyrene-functionalized monomers were found to display more efficient dsDNA recognition vis-à-vis corresponding triphenylene- or coronene-functionalized monomers.[86]

The relative orientation of the intercalating pyrene moiety relative to the sugar ring depends on which pyrene carbon is used to attach the linker (Fig. 1.11 (a)), which in turn impacts binding properties.[74] Probes modified with 2'-*O*-(pyren-2-yl)methyl uridines (**Y**) or 2'-*O*-(pyren-1-yl)methyl uridines (**X**) recognize dsDNA targets more efficiently than 2'-*O*-(pyren-4-yl)methyl uridine (**Z**) monomers, since the latter results in inefficient stacking of the intercalator.[88]

Moreover, oligonucleotides modified with O2′-alkylated uridine monomers featuring substituted pyrene moieties viz., 2′-*O*-(7-neo-pentylpyren-1-yl)methyl-uridine (V) and 2′-*O*-(7-tert-butyl-1-methoxypyren-5-yl)methyl-uridine monomer (**W**), were explored as potential building blocks of

Invader probes (Fig. 1.11(b)), assuming that the steric bulk on position 7 of the pyrene intercalator will foster further destabilization in the probe duplex while maintaining the stability of the probe-target duplex, increasing the thermodynamic driving force for dsDNA recognition.
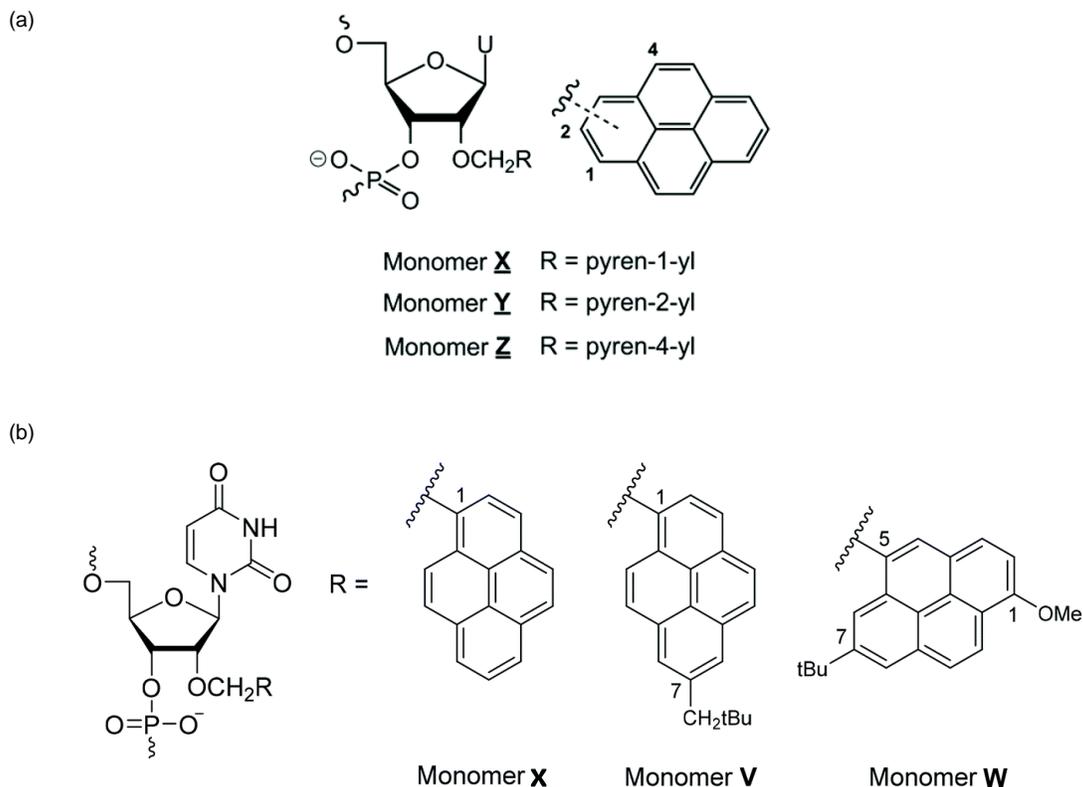
Figure 1.11. (a) Structures of O2′-pyrene-functionailized uridine monomers (**X, Y and Z**) , where the relative orientation of the intercalating pyrene moiety relative to the sugar ring are different. (b) Structures of O2′-pyrene-functionailized uridine monomers (**X, V and W**) where, V and W are substituted pyrene monomers. Reprinted with permission from (S. Karmakar, T. Horrocks, B. C. Gibbons, D. C. Guenther, R. Emehiser and P. J. Hrdlicka, *Org. Biomol. Chem.*, 2019, **17**, 609–621.) Copyright (2019) Royal Society of Chemistry.

The stability of duplex is determined by estimating the thermal denaturation temperature (T_m). Invader duplexes with +1 interstrand zipper arrangement, formed by ONs containing **V-** and **Y-** monomers are labile vis-à-vis invader duplexes with unsubstituted pyrene monomers **X** ($T_m = 26.5$ °C), where **W**-monomers introduces higher destabilization ($T_m = 22.0$ °C) than **V**-monomers ($T_m = 25.0$ °C). ONs functionalized with monomer-**V** display higher cDNA affinity indicated by increase in stabilization of probe-target duplexes (T_m as high as 10.5 °C compared

to unmodified DNA) while, **Y**-monomer functionalized ONs show lower cDNA affinity (decrease in T_m by as low as -6.5 °C compared to unmodified DNA). However, the stability of probe:cDNA duplexes (duplexes between probe strand and complementary DNA target strand) formed by **V**- and **W**-modified ONs is reduced relative to probe:cDNA duplexes formed by ONs with unsubstituted monomers. Thus, the substituted monomers display a less prominent thermodynamic driving force than unsubstituted monomers. In agreement with the fact that substituted monomers have lower affinity to cDNA target than unsubstituted ones, probes containing both of the O2′-alkylated uridine monomers featuring substituted pyrene moieties (**V** and **W**) displayed lower recognition of dsDNA targets compared to probes with monomers containing unsubstituted pyrene (**X**).[89]

1.4.2 Optimization of hotspots

The hands-on availability of all four natural nucleobase versions of the optimized 2'-O-(pyren-1-yl)methyl RNA building blocks, allowed for a study of sequence impact on dsDNA recognition efficiency. Probe duplexes, in which the hotspots were generated using pyrimidine monomers were found to display more efficient recognition of dsDNA than probes in which the hotspots were generated using purine monomers.[90] Invader modified probe strands display higher affinity towards cDNA vis-à-vis unmodified strands.[90] The stabilizing effect of the monomers are more pronounced when pyrene-functionalized monomers are 3'-flanked by nucleobases with large aromatic surfaces, i.e., A or G. The thermodynamic driving force follows the trend; **U**A:A**U** > **C**G:G**C** > **U**T:A**A** >> **G**C:C**G**, where bold letters indicate 2'-*O*-(pyren-1-yl)methyl monomers. This knowledge resulted in design rules in which pyrimidine monomers are preferred, A monomers are acceptable, and G monomers should be avoided.

Next, a series of Invader probes with varying numbers, the distance between and/or positions of energetic hotspots were evaluated to determine the impact of these parameters on dsDNA recognition efficiency.[71] Probes with a greater number of hotspots (1 < 2 < 4) display greater affinity towards cDNA,[71] most likely because of the increased number of intercalations by aromatic pyrenes in the probe-target duplex. To this end, the study suggested that probes containing ~ 30% of modification are advantageous for dsDNA targeting as they demonstrate significant affinity and specificity towards dsDNA targets.[79]

1.4.3 Pseudocomplementary Invader probes

An alternative design was explored in which the concept of pseudocomplementary base pairs and violation of the nearest neighbor exclusion principle (i.e., forced intercalation, Invader approach) were combined. In this probe design, there are two destabilizing factors offered by energetic hotspots and pseudocomplementary D:S base-pairs. It was expected that pseudocomplementary Invader probes would display additional disruption of the probe duplex, while displaying increased target affinity, thusly increasing the overall thermodynamic driving force towards dsDNA recognition. Two different approaches were explored in the study- i) the intercalator was integrated as a part of pseudocomplementary nucleotide, ii) separating the two duplex destabilizing elements viz., energetic hotspots and pseudocomplementary nucleotide in the probe design (Fig. 1.12).
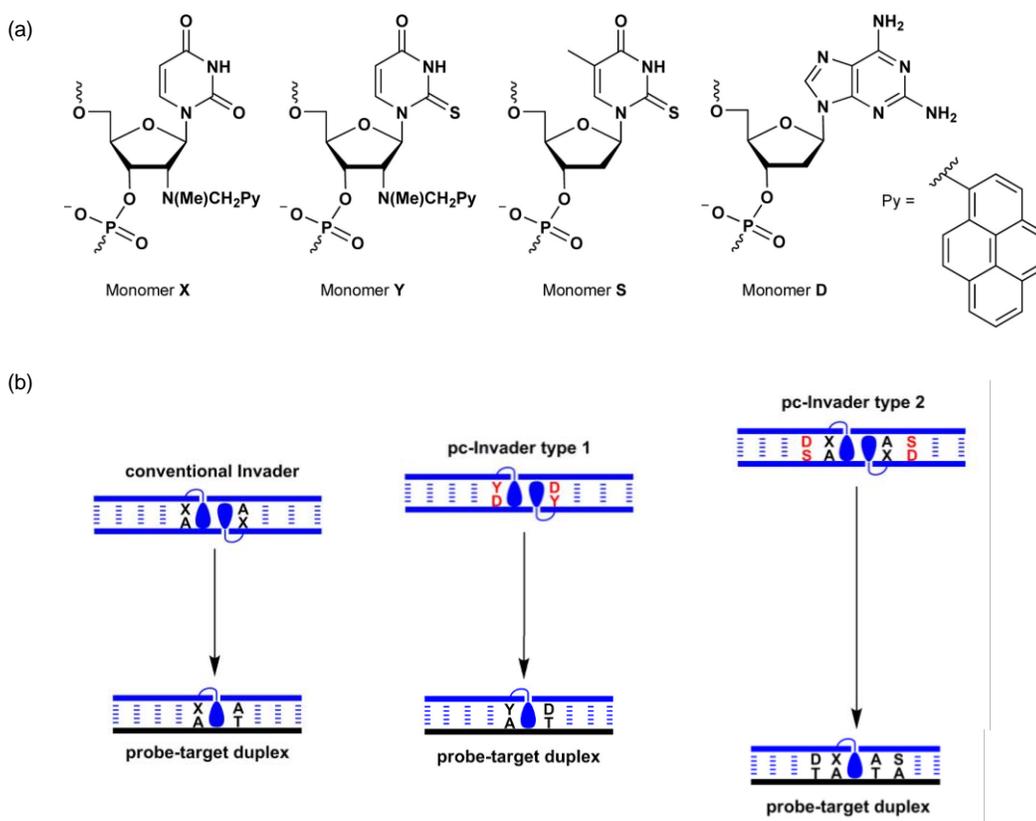
Figure 1.12. (a) Structures of monomers (X, Y, S, and D). (b) Illustrations of probe duplexes and probe-target duplexes formed by DY probes (pc-Invader type 1), and DSX probes (pc-Invader type 2). Pseudocomplementary base-pairs are shown in red. Droplets denote intercalators. (only one probe-target duplex is shown). Reprinted with permission from (B. A.

Anderson and P. J. Hrdlicka, *J. Org. Chem.*, 2016, **81**, 3335–3346.) Copyright (2016) American Chemical Society. Further permissions related to the material excerpted should be directed to the ACS (https://pubs.acs.org/doi/10.1021/acs.joc.6b00369).

The first approach was pursued by using 2′-*N*-(pyren-1-yl)methyl-2′-*N*-methyl-2′-amino-2-thiouridine monomer (Y), which contains both of the key structural features i.e., pyrene intercalator as well as pseudocomplementary base. This monomer (Y) is positioned opposite to a 2,6-diaminopurine DNA (D), and the +1 interstrand zipper arrangement is maintained in the Invader design, thus producing a probe with pseudocomplementary hotspots (DY probes). The second approach is pursued by separating pc-basepair (D-S) and energetic hotspots maintained by 2′-*N*-(pyren-1-yl)methyl-2′-*N*-methyl-2′-aminouridine monomers (X) in +1 interstrand zipper arrangement, while keeping them alongside each other (probe DSX).

On a head-to-head comparison to conventional Invaders in the 13-mer sequence context, DSX probes strands show higher affinity towards cDNA whereas, ONs of DY probes show slightly lower affinity towards cDNA. Both probe types are more labile than conventional Invaders, withDSX having the highest lability. Because the DY probes form weaker probe-target duplexes vis-à-vis probe-target duplexes formed by conventional probes, this design suggests that the close proximity of the two structural features is not a beneficial design, which is attributed to marginally compatible binding modes for the pyrene and 2-thiouracil moieties.[84] On the other hand, when the two structural destabilizing motifs were separated (DSX probes), probe-target duplexes are formed that are more stable relative to those by conventional Invaders. Thus, DSX probes are strongly activated for recognition of dsDNA targets which is in agreement with the results of recognition experiments of model dsDNA targets, where DSX probes display improved recognition efficiency vis-à-vis conventional Invaders. Conventional Invaders were more efficient in recognizing targets than DY probes. In sum, pseudocomplementary Invader probes in which the two destabilizing elements separated are more labile than conventional Invaders, and form relatively stable probe-target duplexes, which enhances the recognition efficiency. This strategy is interesting not only because the appropriately designed probes offer stable probe-target duplexes and highly labile probes but also it offers excellent mismatch discrimination specificity against a singly mismatched hairpin target.[84]

1.4.4 Chimeric γPNA-Invader and chimeric LNA-Invader probes.

A study was conducted comparing probes of different chemistries (Invader probes, ^{MP}γPNAs, and LNA) to compare the efficiency and specificity to target mixed-sequence dsDNA (see Fig. 1.13 for structures of ^{MP}γPNAs, and LNA) .[51] In an AT-rich 13-mer sequence context, the Invader probe was more efficient and specific for dsDNA targeting vis-à-vis ^{MP}γPNAs, or LNA probes. Affinity-enhancing ss-probes like ^{MP}γPNA and LNA adopt secondary structures e.g., a hairpin or undergo dimerization if they contain self-complementary regions. This potential to form secondary structures renders it challenging for these high-affinity ss-probes to bind dsDNA as it will impede the process of forming a probe-target recognition complex. In contrast, double-stranded probes like Invader probes are tuned to form easily denaturing duplexes, irrespective of the sequence context of the target region. Thus the limitation of avoiding self-complementary sequence in ss-probes design no longer applies in ds-probes. The thermal denaturation study revealed that all of the probes (^{MP}γPNAs, LNA, and Invader probes) display high affinity towards complementary target strands (cDNA) compared to analogous unmodified DNA. The stability of duplexes, formed by probe strand and cDNA strand decreasess in the order: ^{MP}γPNA:cDNA >> Invader:cDNA > LNA:cDNA. The ability of these probes to form recognition complexes with the   dsDNA-targets was compared. Invader probes display efficient recognition of hairpin targets (~75 % recognition), while recognition was moderate (<15 - ~ 60 % recognition) with ^{MP}γPNAs, and low (<15 - ~ 25% recognition) with LNA probes. Despite having a very high affinity towards cDNA, ^{MP}γPNA probes suffered from poor target recognition. Similarly, LNA probes also display poor recognition despite having the potential to form highly stable probes:cDNA duplexes. The poor efficiency of ss-probes (LNA/PNA) is because of the formation of secondary structures and self-hybridization. Moreover, Invader probes display excellent mismatch discrimination whereas, ^{MP}γPNA displays minimal discrimination against singly mismatched DNA targets. Thus, ^{MP}γPNAs despite being high-affinity enhancing modification proved to be less efficient while forming recognition complexes with target hairpin, due to self-dimerization of γPNAs as the target includes substantial self-complementary AT-rich regions.[51]
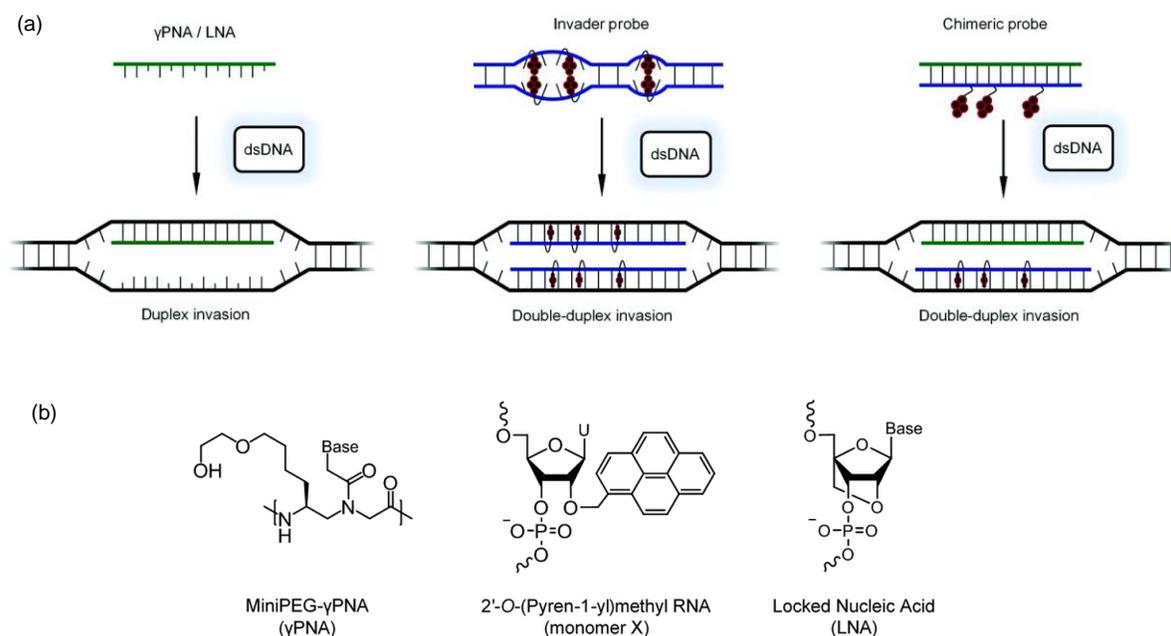
Figure 1.13. (a) Illustration of recognition of dsDNA by single-stranded γPNA or LNA probes, and double-stranded Invader and chimeric probes. (b) Structures of modifications and nucleic acid mimics were used in this study. Reprinted with permission from (R. G. Emehiser and P. J. Hrdlicka, *Org. Biomol. Chem.*, 2020, **18**, 1359–1368.) Copyright (2020) Royal Society of Chemistry.

To reduce tendency to self-pair or form secondary hairpin-like structures in such sequence context which contains substantial self-complementary AT-rich regions, chimeric probes between γPNA and Invader strands were designed (Fig. 1.13). The motivation behind the chimeric design is based upon recent work, where it was reported that PNA:DNA duplexes lack the accommodative flexibility to retain intercalators ensuing easily denaturing probe duplexes.[91,92] The chimeric probes formed from duplex between fully modified γPNA strand and complementary Invader strand which contains pyrene moiety would thus form a chimeric γPNA-Invader duplex, which will readily denature and recognize complementary target cDNA strands via double-duplex invasion. Thus, readily labile ds-probes (chimeric γPNA-Invader probes) will not only be strongly activated for target recognition, but presumably they also will minimize dimerization of γPNA strands and improve the mismatch specificity as well.

The double-stranded chimeric γPNA-Invader probes are labile (T_ms up to 21 °C lower than the corresponding γPNA-DNA duplexes), which is likely due to a PNA-DNA duplex not having

sufficient flexibility to accommodate intercalators.[92] Perturbation of pyrene-nucleobase stacking in chimeric PNA-Invader duplexes is indicated by the hypsochromically shifted pyrene absorption bands, as has been observed in Invader probe duplexes.[88] In a comparison study, where a hairpin model target was used to study the recognition efficiency of the probes, chimeric γPNA–Invader probes displayed enhanced recognition efficiency vis-à-vis single-stranded γPNA (C_{50} = 45-165 nM vs > 250 nM), which are prone to forming self-complementary duplexes. The C_{50} value is the concentration of probes required to attain 50% recognition; lower C_{50} values of probes indicate more efficient target recognition. Thus, chimeric γPNA–Invader probes prevent γPNA from forming self-hybridizing duplexes. However, corresponding Invader probes (C_{50} ~ 40-50 nM) result in more efficient recognition than γPNA–Invader probes.

Likewise, in the same study, chimeric LNA-Invader probes were explored. The motivation of designing chimeric LNA-Invader probes is based on the characteristic of intercalator-modified ONs being poorly housed in A-type probe duplexes adopted by LNA (or RNA), thus, producing a readily denaturing chimeric LNA-Invader probe duplexes,[93,94] Not only is an easily denaturing duplex formed, but chimeric LNA-Invader duplexes will also reduce dimerization of LNA strands, which will likely improve the binding efficiency and specificity. Contrary to the hypothesis, the chimeric LNA-Invader probe (where LNA strand contains ~31 % of LNA monomers) is more stable than the corresponding LNA:cDNA duplexes (ΔT_m of up to 17 °C), which is likely because the pyrene intercalator of Invader strand is adopting altered binding modes for better accommodation in the LNA-Invader duplex.[85]   The observed minor bathochromic shift of the pyrene absorption indicates better accommodation of the pyrene moiety in chimeric-LNA probes.[88,85] The high stability of the chimeric LNA-Invader probes likely precluded efficient recognition of the model hairpin targets.[85]

1.4.5 Bulged Invader probes

The probe architecture of Invader probes has been additionally explored further by introducing non-nucleosidic nonyl (C_9) bulges (Fig. 1.14).[83] The presence of such bulges is known to allow for retention of B-type duplex geometry while interrupting the π-stacking and promoting the dissociation of the duplex.[82] A series of near-terminally C_9-modified 13-mer Invader probes with two hotspots at the center was assembled and the influence of the nonyl bulges on probe

destabilization and cDNA affinity was studied relative to regular Invader probes. The probe designs contain C_9 bulges in varying numbers in the Invader duplex.
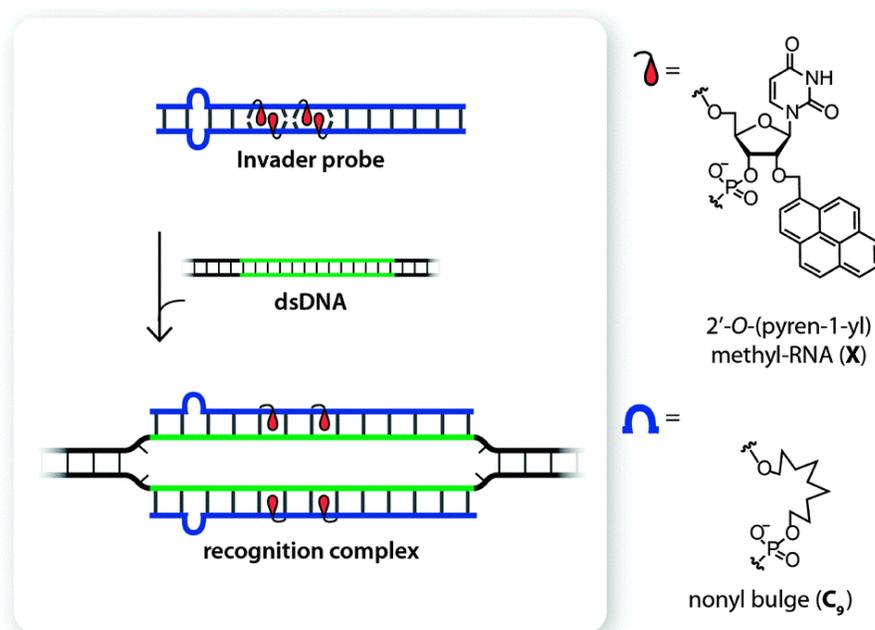
Figure 1.14. Illustration showing bulge Invader probe targeting dsDNA and structures of intercalator-modified monomers and C_9-bulge. Reprinted with permission from (D. C. Guenther, S. Karmakar, and P. J. Hrdlicka, *Chem. Commun. Camb. Engl.*, 2015, **51**, 15051–15054). Copyright (2015) Royal Society of Chemistry.

The insertion of a single C_9-bulge destabilizes the 13-mer Invader duplex (T_m ~ 28.5 -35 °C) vis-à-vis conventional probes (T_m ~ 45 °C). The probe-target duplexes resulting from modified Invader strand and cDNA also face the destabilizing effect of the C_9-bulge as the T_m reduces by 9 - 27 °C relative to the corresponding probe:cDNA duplexes (T_m ~ 55°C) formed by conventional Invader strands and cDNA. The driving force for dsDNA recognition is approximated by the term thermal advantage (*TA*), which is the difference between the T_ms of products (probe-target duplexes) and reactants (probe duplex and dsDNA target). Singly C_9-modified Invader probes display marginally greater driving force (*TA* increases by up to 4.5 °C) relative to conventional probes. Probes with two C_9-bulges opposite of each other display a more favorable driving force (*TA* >35°C) since the probe duplex is highly destabilized (T_m <15 °C), whereas the resulting probe-target duplexes only are slightly destabilized. In agreement with the estimated *TA* values, Invader probes with opposing C_9 bulges resulted in enhanced

recognition of hairpin targets relative to conventional probes (C_{30} ~1.0 μM vs ~9.7 μM). Furthermore, Invader probes with opposing C_9 bulges display ~4-fold faster recognition kinetics.[83]

Later, to explore the impact of bulge length, 18-mer probes with non-nucleotidic two, four, or nine atom linkers in the central region were studied (Fig. 1.15).[95] This probe architecture offers maximal disruption from the centrally modified non-nucleotidic bulges and the near-terminally positioned hotspots, which encourage end-fraying.
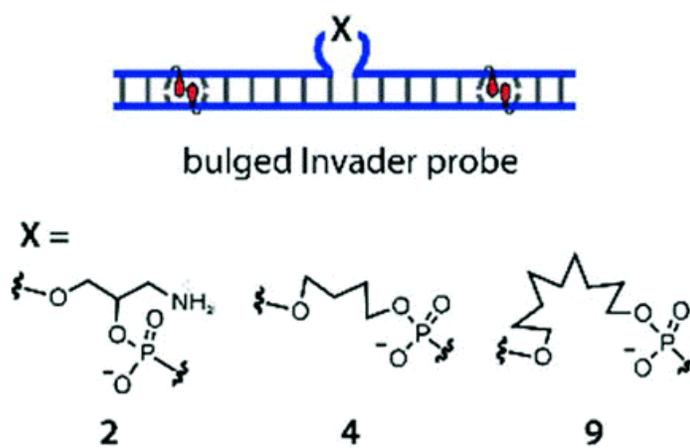
Figure 1.15. Illustration of bulged Invader probes and structure of bulge modifications. Reprinted with permission from D. C. Guenther, R. G. Emehiser, A. Inskeep, S. Karmakar, and P. J. Hrdlicka, *Org. Biomol. Chem.*, 2020, **18**, 4645–4655) Copyright (2020) Royal Society of Chemistry.

As expected, with the increasing size of the bulge, the probe destabilization became more pronounced as base-pair cooperativity was disturbed further. Singly C_9-bulged Invader probes proved favorable for sequence-unrestricted recognition of dsDNA targets, displaying enhanced recognition and excellent discrimination of mismatched dsDNA targets. In summary, non-nucleotidic C_9-modified bulged Invader probes exhibit faster reaction kinetics, enhanced recognition, and excellent mismatch specificity of dsDNA targets at non-denaturing conditions.

1.5. Conclusion

The gene-targeting strategies offer the promise of treatment of diseases at their source (the genetic information embedded in DNA), while each of the strategies features some advantages with some

limitations. Extensive biological and chemical exploration of these methodologies will help us to understand the perks and underlying limitations that need to be improved intending for enhanced recognition of dsDNA targets with excellent mismatch specificity – that ensures the platform for the development of DNA targeting novel therapeutic tools.

1.6 References:

1 K. D. Warner, C. E. Hajdin and K. M. Weeks, *Nat. Rev. Drug Discov.*, 2018, **17**, 547–558.

2 H. Xiong, R. N. Veedu and S. D. Diermeier, *Int. J. Mol. Sci.*, 2021, **22**, 3295.

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

4 S. Harteis and S. Schneider, *Int. J. Mol. Sci.*, 2014, **15**, 12335–12363.

5 J. M. Berg, J. L. Tymoczko and L. Stryer, *Biochem. 5th Ed.*

6 E. Lescrinier, M. Froeyen and P. Herdewijn, *Nucleic Acids Res.*, 2003, **31**, 2975–2989.

7 D. Kahne, *Chem. Biol.*, 1995, **2**, 7–12.

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

9 W. L. Walker, E. M. Landaw, R. E. Dickerson and D. S. Goodsell, *Proc. Natl. Acad. Sci.*, 1998, **95**, 4315–4320.

10 J. M. Belitsky, S. J. Leslie, P. S. Arora, T. A. Beerman and P. B. Dervan, *Bioorg. Med. Chem.*, 2002, **10**, 3313–3318.

11 L. A. Dickinson, J. W. Trauger, E. E. Baird, P. Ghazal, P. B. Dervan and J. M. Gottesfeld, *Biochemistry*, 1999, **38**, 10801–10807.

12 K. Maeshima, S. Janssen and U. K. Laemmli, *EMBO J.*, 2001, **20**, 3218–3228.

13 A. Lauria, A. Montalbano, P. Barraja, G. Dattolo and A. M. Almerico, *Curr. Med. Chem.*, 2007, **14**, 2136–2160.

14 S. Bhaduri, N. Ranjan and D. P. Arya, *Beilstein J. Org. Chem.*, 2018, **14**, 1051–1086.

15 G. Felsenfeld, D. R. Davies and A. Rich, *J. Am. Chem. Soc.*, 1957, **79**, 2023–2024.

16 H. E. Moser and P. B. Dervan, *Science*, 1987, **238**, 645–650.

17 R. Bahal, A. Gupta and P. M. Glazer, in *Genome Editing: The Next Step in Gene Therapy*, eds. T. Cathomen, M. Hirsch and M. Porteus, Springer, New York, NY, 2016, pp. 93–110.

18 M. Egholm, L. Christensen, K. L. Deuholm, O. Buchardt, J. Coull and P. E. Nielsen, *Nucleic Acids Res.*, 1995, **23**, 217–222.

19 H. Torigoe, S. M. A. Rahman, H. Takuma, N. Sato, T. Imanishi, S. Obika and K. Sasaki, *Chem. – Eur. J.*, 2011, **17**, 2742–2751.

20 Z. Sandor and A. Bredberg, *Nucleic Acids Res.*, 1994, **22**, 2051–2056.

21 J. R. Goñi, J. M. Vaquerizas, J. Dopazo and M. Orozco, *BMC Genomics*, 2006, **7**, 63.

22 A. Gupta, A. Mishra and N. Puri, *J. Biotechnol.*, 2017, **259**, 148–159.

23 J.-C. Wu, Q.-C. Meng, H.-M. Ren, H.-T. Wang, J. Wu and Q. Wang, *Acta Pharmacol. Sin.*, 2017, **38**, 798–805.

24 F. Pellestor and P. Paulasova, *Eur. J. Hum. Genet.*, 2004, **12**, 694–700.

25 K. Kaihatsu, B. A. Janowski and D. R. Corey, *Chem. Biol.*, 2004, **11**, 749–758.

26 P. Muangkaew and T. Vilaivan, *Bioorg. Med. Chem. Lett.*, 2020, **30**, 127064.

27 P. E. Nielsen, M. Egholm and O. Buchardt, *Bioconjug. Chem.*, 1994, **5**, 3–7.

28 M. C. Griffith, L. M. Risen, M. J. Greig, E. A. Lesnik, K. G. Sprankle, R. H. Griffey, J. S. Kiely and S. M. Freier, *J. Am. Chem. Soc.*, 1995, **117**, 831–832.

29 T. Bentin and P. E. Nielsen, *J. Am. Chem. Soc.*, 2003, **125**, 6378–6379.

30 T. Bentin, H. J. Larsen and P. E. Nielsen, *Biochemistry*, 2003, **42**, 13987–13995.

31 K. Kaihatsu, R. H. Shah, X. Zhao and D. R. Corey, *Biochemistry*, 2003, **42**, 13996–14003.

32 A. S. Ricciardi, E. Quijano, R. Putman, W. M. Saltzman and P. M. Glazer, *Mol. Basel Switz.*, , DOI:10.3390/molecules23030632.

33 P. E. Nielsen and L. Christensen, *J. Am. Chem. Soc.*, 1996, **118**, 2287–2288.

34 P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, *Science*, 1991, **254**, 1497–1500.

35 T. Ratilainen, A. Holmén, E. Tuite, P. E. Nielsen and B. Nordén, *Biochemistry*, 2000, **39**, 7781–7791.

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

37 W. He, M. J. Crawford, S. Rapireddy, M. Madrid, R. R. Gil, D. H. Ly and C. Achim, *Mol. Biosyst.*, 2010, **6**, 1619–1629.

38 S. Rapireddy, G. He, S. Roy, B. A. Armitage and D. H. Ly, *J. Am. Chem. Soc.*, 2007, **129**, 15596–15600.

39 R. Bahal, B. Sahu, S. Rapireddy, C.-M. Lee and D. H. Ly, *Chembiochem Eur. J. Chem. Biol.*, 2012, **13**, 56–60.

40 B. Sahu, I. Sacui, S. Rapireddy, K. J. Zanotti, R. Bahal, B. A. Armitage and D. H. Ly, *J. Org. Chem.*, 2011, **76**, 5614–5627.

41 A. Gupta, E. Quijano, Y. Liu, R. Bahal, S. E. Scanlon, E. Song, W.-C. Hsieh, D. E. Braddock, D. H. Ly, W. M. Saltzman and P. M. Glazer, *Mol. Ther. - Nucleic Acids*, 2017, **9**, 111–119.

41 S. Dutta, B. A. Armitage and Y. L. Lyubchenko, *Biochemistry*, 2016, **55**, 1523–1528.

43 R. Bahal, N. A. McNeer, D. H. Ly, W. M. Saltzman and P. M. Glazer, *Artif. DNA PNA XNA*, 2013, **4**, 49–57.

44 B. Raman, Q. Elias, A. M. Nicole, L. Yanfeng, C. B. Dinesh, L.-G. Francesco, J. F. Rachel, M. S. William, H. L. Danith and M. G. Peter, *Curr. Gene Ther.*, 2014, **14**, 331–342.

45 A. Manna, S. Rapireddy, G. Sureshkumar and D. H. Ly, *Tetrahedron*, 2015, **71**, 3507–3514.

46 N. Shigi, Y. Mizuno, H. Kunifuda, K. Matsumura and M. Komiyama, *Bull. Chem. Soc. Jpn.*, 2019, **92**, 330–335.

47 T. Ishizuka, J. Yoshida, Y. Yamamoto, J. Sumaoka, T. Tedeschi, R. Corradini, S. Sforza and M. Komiyama, *Nucleic Acids Res.*, 2008, **36**, 1464–1471.

48 J. Sumaoka and M. Komiyama, *Chem. Lett.*, 2014, **43**, 1581–1583.

49 V. V. Demidov and M. D. Frank-Kamenetskii, *Trends Biochem. Sci.*, 2004, **29**, 62–71.

50 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.

51 R. G. Emehiser, E. Hall, D. C. Guenther, S. Karmakar and P. J. Hrdlicka, *Org. Biomol. Chem.*, 2019, **18**, 56–65.

52 J. S. Jepsen, M. D. Sørensen and J. Wengel, *Oligonucleotides*, 2004, **14**, 130–146.

53 H. Kaur, B. R. Babu and S. Maiti, *Chem. Rev.*, 2007, **107**, 4672–4697.

54 P. M. D. Moreno, S. Geny, Y. V. Pabon, H. Bergquist, E. M. Zaghloul, C. S. J. Rocha, I. I. Oprea, B. Bestas, S. E. Andaloussi, P. T. Jørgensen, E. B. Pedersen, K. E. Lundin, R. Zain, J. Wengel and C. I. E. Smith, *Nucleic Acids Res.*, 2013, **41**, 3257–3273.

55 J. T. Nielsen, P. C. Stein and M. Petersen, *Nucleic Acids Res.*, 2003, **31**, 5858–5867.

56 S. Geny, P. M. D. Moreno, T. Krzywkowski, O. Gissberg, N. K. Andersen, A. J. Isse, A. M. El-Madani, C. Lou, Y. V. Pabon, B. A. Anderson, E. M. Zaghloul, R. Zain, P. J. Hrdlicka,

P. T. Jørgensen, M. Nilsson, K. E. Lundin, E. B. Pedersen, J. Wengel and C. I. E. Smith, *Nucleic Acids Res.*, 2016, **44**, 2007–2019.

57 J. S. Jepsen and J. Wengel, *Curr. Opin. Drug Discov. Devel.*, 2004, **7**, 188–194.

58 R. Ge, J. E. Heinonen, M. G. Svahn, A. J. Mohamed, K. E. Lundin and C. I. E. Smith, *FASEB J.*, 2007, **21**, 1902–1914.

59 R. Ge, M. G. Svahn, O. E. Simonson, A. J. Mohamed, K. E. Lundin and C. I. E. Smith, *J. Gene Med.*, 2008, **10**, 101–109.

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

61 D. M. Thurtle-Schmidt and T.-W. Lo, *Biochem. Mol. Biol. Educ. Bimon. Publ. Int. Union Biochem. Mol. Biol.*, 2018, **46**, 195–205.

62 T. Gaj, C. A. Gersbach and C. F. Barbas, *Trends Biotechnol.*, 2013, **31**, 397–405.

63 M. Redman, A. King, C. Watson and D. King, *Arch. Dis. Child. Educ. Pract. Ed.*, 2016, **101**, 213–215.

64 D. Rath, L. Amlinger, A. Rath and M. Lundgren, *Biochimie*, 2015, **117**, 119–128.

65 J. R. Guitart, J. L. Johnson and W. W. Chien, *J. Invest. Dermatol.*, 2016, **136**, e87–e93.

66 Y. Wu, D. Liang, Y. Wang, M. Bai, W. Tang, S. Bao, Z. Yan, D. Li and J. Li, *Cell Stem Cell*, 2013, **13**, 659–662.

67 S.-R. Lin, H.-C. Yang, Y.-T. Kuo, C.-J. Liu, T.-Y. Yang, K.-C. Sung, Y.-Y. Lin, H.-Y. Wang, C.-C. Wang, Y.-C. Shen, F.-Y. Wu, J.-H. Kao, D.-S. Chen and P.-J. Chen, *Mol. Ther. - Nucleic Acids*, , DOI:10.1038/mtna.2014.38.

68 N. Savić and G. Schwank, *Transl. Res.*, 2016, **168**, 15–21.

69 X.-H. Zhang, L. Y. Tee, X.-G. Wang, Q.-S. Huang and S.-H. Yang, *Mol. Ther. - Nucleic Acids*, 2015, **4**, e264.

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

71 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.

72 P. J. Hrdlicka, T. S. Kumar and J. Wengel, *Chem. Commun.*, 2005, 4279–4281.

73 S. P. Sau, A. S. Madsen, P. Podbevsek, N. K. Andersen, T. S. Kumar, S. Andersen, R. L. Rathje, B. A. Anderson, D. C. Guenther, S. Karmakar, P. Kumar, J. Plavec, J. Wengel and P. J. Hrdlicka, *J. Org. Chem.*, 2013, **78**, 9560–9570.

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

75 D. M. Crothers, *Biopolymers*, 1968, **6**, 575–584.

76 S. P. Adhikari, R. G. Emehiser, S. Karmakar and P. J. Hrdlicka, *Org. Biomol. Chem.*, 2019, **17**, 8795–8799.

77 B. A. Didion, S. Karmakar, D. C. Guenther, S. P. Sau, J. P. Verstegen and P. J. Hrdlicka, *Chembiochem Eur. J. Chem. Biol.*, 2013, **14**, 1534–1538.

78 S. P. Sau, T. S. Kumar and P. J. Hrdlicka, *Org. Biomol. Chem.*, 2010, **8**, 2028–2036.

79 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.

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

81 S. Karmakar, T. Horrocks, B. C. Gibbons, D. C. Guenther, R. Emehiser and P. J. Hrdlicka, *Org. Biomol. Chem.*, 2019, **17**, 609–621.

82 D. V. Pyshnyi, A. A. Lomzov, I. A. Pyshnaya and E. M. Ivanova, *J. Biomol. Struct. Dyn.*, 2006, **23**, 567–579.

83 D. C. Guenther, S. Karmakar and P. J. Hrdlicka, *Chem. Commun. Camb. Engl.*, 2015, **51**, 15051–15054.

84 B. A. Anderson and P. J. Hrdlicka, *J. Org. Chem.*, 2016, **81**, 3335–3346.

85 R. G. Emehiser and P. J. Hrdlicka, *Org. Biomol. Chem.*, 2020, **18**, 1359–1368.

86 B. A. Anderson, J. J. Onley and P. J. Hrdlicka, *J. Org. Chem.*, 2015, **80**, 5395–5406.

87 M. E. Østergaard and P. J. Hrdlicka, *Chem. Soc. Rev.*, 2011, **40**, 5771–5788.

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

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

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

91 H. Asanuma, R. Niwa, M. Akahane, K. Murayama, H. Kashida and Y. Kamiya, *Bioorg. Med. Chem.*, 2016, **24**, 4129–4137.

92 P. Wittung, S. K. Kim, O. Buchardt, P. Nielsen and B. Norden, *Nucleic Acids Res.*, 1994, **22**, 5371–5377.

93 V. V. Filichev, B. Vester, L. H. Hansen and E. B. Pedersen, *Nucleic Acids Res.*, 2005, **33**, 7129–7137.

94 V. Marin, H. F. Hansen, T. Koch and B. A. Armitage, *J. Biomol. Struct. Dyn.*, 2004, **21**, 841–850.

95 D. C. Guenther, R. G. Emehiser, A. Inskeep, S. Karmakar and P. J. Hrdlicka, *Org. Biomol. Chem.*, 2020, **18**, 4645–4655.

CHAPTER 2: Recognition of mixed-sequence DNA targets using spermine-modified Invader probes

Shiva P. Adhikari,[a]   Raymond G. Emehiser,[a]   Saswata Karmakar[a] and Patrick J. Hrdlicka[a]

[a]Department of Chemistry, University of Idaho, Moscow, ID-83844, USA.

Abstract

Double-stranded oligodeoxyribonucleotides with +1 interstrand zipper arrangements of 2'-*O*-(pyren-1-yl)methyl-RNA monomers are additionally activated for highly specific recognition of mixed-sequence DNA targets upon incorporation of non-nucleotidic spermine bulges.

2.1 Introduction

Development of constructs capable of recognizing specific sequences of double-stranded DNA (dsDNA) continues to be an aspirational goal that is fuelled by the promise of tools that will enable site-specific regulation, detection, and manipulation of genomic DNA. Early technologies, such as triplex forming oligonucleotides and peptide nucleic acids (PNAs),[1,2] have proven robust but are limited to recognition of extended polypurine targets. Many alternative nucleic acid mimics have been developed[3–15] but recognition of mixed-sequence dsDNA sequences at physiologic conditions remains challenging. Even CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-associated (Cas) nucleoprotein constructs,[16] which have received much recent attention, face many challenges that remain to be resolved, including reducing off-target binding and editing activities, and improving cellular delivery.[17]

We have previously introduced Invader probes as a potential solution toward specific sequence-unrestricted recognition of dsDNA.[18] Invader probes are short DNA duplexes that are modified with one or more +1 interstrand zipper arrangements of intercalator-modified nucleotides like 2′-*O*-(pyren-1-yl)methyl-RNA **(**Fig. 2.1**).**[19] This particular monomer arrangement forces two intercalating moieties to compete for the same inter-base-pair region, resulting in localized

unwinding and probe destabilization[20,21] as the nearest-neighbor exclusion principle[22] – which asserts that intercalation is anti-cooperative at adjacent sites[23] – is violated. Each individual strand of an Invader probe displays exceptionally high affinity toward complementary DNA (cDNA) as duplex formation results in strongly stabilizing stacking interactions between the intercalators and flanking base-pairs.[18–21] The greater stability of the duplexes in the formed recognition complex vis-à-vis the Invader probe and dsDNA target region, provides the driving force for recognition of complementary dsDNA regions, which can be of mixed sequence composition (Fig. 2.1). Invader probes have, for example, been used to recognize mixed-sequence target regions on Y-chromosomes in fixed nuclei from male bovine kidney cells under otherwise non-denaturing conditions.[19]
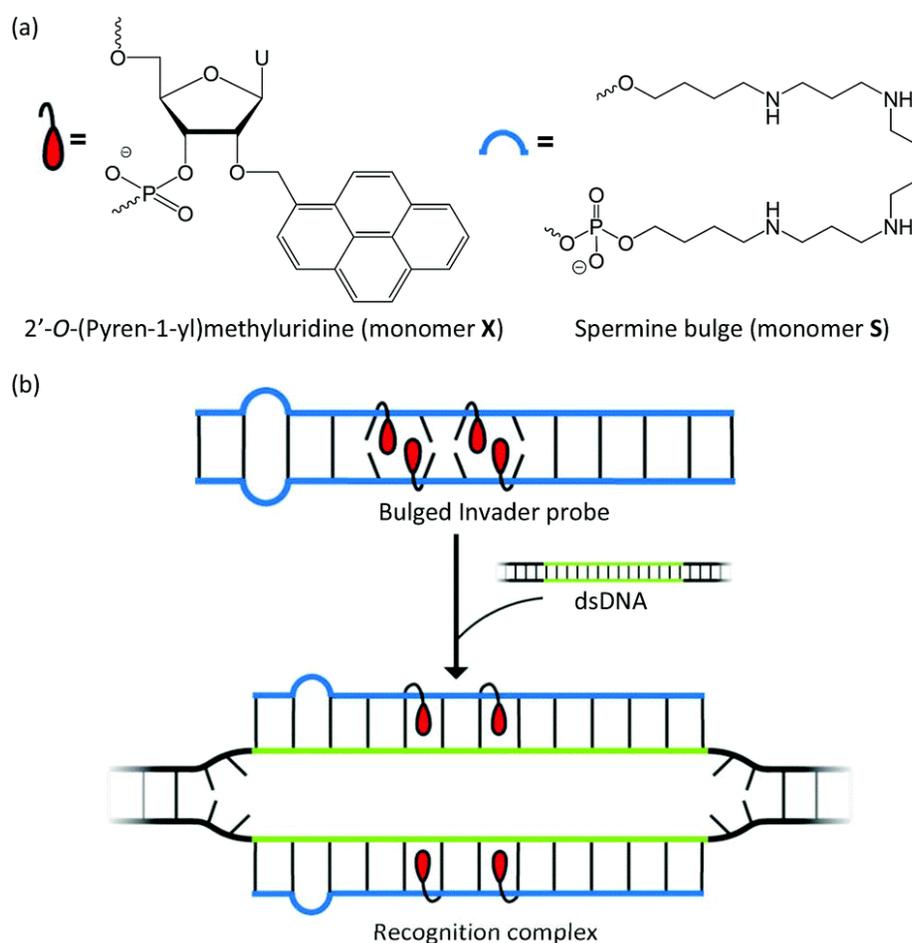
Figure 2.1. (a) Structures of 2′-O-(pyren-1-yl)methyluridine and spermine monomers used herein. (b) Principle of sequence-unrestricted dsDNA-recognition using bulged Invader probes.

We recently introduced Invader probes that are additionally modified with nonyl (C_9) bulges,[24] expecting that the non-nucleotidic bulges would promote further denaturation and destabilization of the probe duplex. This, in turn, was expected to increase the availability of the probe's Watson–Crick face for binding and accelerate nucleation with, and invasion of, dsDNA targets. Indeed, we observed that incorporation of C_9 bulges at certain positions of Invader probes results in more efficient (>5-fold) and faster (>4-fold) dsDNA-recognition relative to conventional Invader probes at certain conditions.[24]

Motivated by these findings, we set out to study Invader probes with spermine bulges, which are considerable larger than the C_9 bulges (Fig. 2.1). We were particularly eager to study probes with opposing spermine bulges due to the prospect for additional destabilization and energetic activation for dsDNA-recognition stemming from increased duplex disruption and electrostatic repulsion from the positively charged[25] spermine bulges. Towards this end, we synthesized a series of oligodeoxyribonucleotides (ONs) modified with commercially available spermine monomer S^{25} and 2′-O-(pyren-1-yl)methyluridine monomer X,[19] allowing us to evaluate a series of 13-mer spermine-containing Invader and control probes (Table 2.1 – see supplementary data for full details). The probe design features two consecutive +1 interstrand zipper arrangements of X monomers (i.e., two energetic hotspots) in the probe center and one or two spermine monomers at one or both termini.

2.2 Results and discussion

Thermal denaturation temperatures (T_mS) were determined for probe duplexes and duplexes between individual probe strands and cDNA (Table 2.1). As previously reported,**[24]** conventional Invader probe ON1:ON2 is moderately stabilized relative to the corresponding unmodified DNA duplex (ΔT_m = 7.5 °C, equivalent to an increase of ~1.9 °C per modification (mod)), whereas duplexes between individual probe strands and cDNA are extraordinarily stabilized (ΔT_m = 18 °C, ΔT_m/mod = 9 °C). The resulting thermodynamic driving force for dsDNA-recognition can be assessed by the term thermal advantage given as TA = T_m (upper probe vs. cDNA) + T_m (lower probe vs. cDNA) − T_m (probe duplex) − T_m (dsDNA).**[19]** As expected, conventional probe ON1:ON2 is strongly activated (TA = 28.5 °C, Table 2.1).

Table 2.1. Sequences of probes used in this study, T_ms of probe duplexes and duplexes between individual probe strands and cDNA, and thermal advantages (TAs) of probes.[a]

Probe	Sequence	T_m [ΔT_m] (°C) probe	upper strand vs cDNA	lower strand vs cDNA	TA (°C)
1:2[b]	5'-G G T A X A X A T A G G C 3'-C C A T A X A X A T C C G	45.0 [+7.5]	55.5 [+18.0]	55.5 [+18.0]	28.5
3:2	5'-G G T A X A X A T A G G C 3'-C C A T A X A X A T C C G	38.0 [+0.5]	48.0 [+10.5]	55.5 [+18.0]	28.0
1:4	5'-G G T A X A X A T A G G C 3'-C C A T A X A X A T C C G	43.0 [+5.5]	55.5 [+18.0]	51.0 [+13.5]	26.0
3:4	5'-G G T A X A X A T A G G C 3'-C C A T A X A X A T C C G	41.0 [+3.5]	48.0 [+10.5]	51.0 [+13.5]	20.5
7:2	5'-G G T A X A X A T A G G C 3'-C C A T A X A X A T C C G	23.0 [-14.5]	35.0 [-2.5]	55.5 [+18.0]	30.0
7:4	5'-G G T A X A X A T A G G C 3'-C C A T A X A X A T C C G	28.0 [-9.5]	35.0 [-2.5]	51.0 [+13.5]	20.5
3c:4c	5'-G G T A T A T A T A G G C 3'-C C A T A T A T A T C C G	30.0 [-7.5]	27.5 [-10.0]	31.5 [-6.0]	-8.5
3c:2	5'-G G T A T A T A T A G G C 3'-C C A T A X A X A T C C G	50.5 [+13.0]	27.5 [-10.0]	55.5 [+18.0]	-5.0
1:4c	5'-G G T A X A X A T A G G C 3'-C C A T A T A T A T C C G	52.5 [+15.0]	55.5 [+18.0]	31.5 [-6.0]	-3.0

[a] ΔT_m s are calculated relative to the corresponding unmodified dsDNA (T_m = 37.5 °C). Thermal denaturation curves were recorded in medium salt phosphate buffer ([Na+] = 110 mM, [Cl−] = 100 mM, pH 7.0 (NaH2PO4/Na2HPO4), [EDTA] = 0.2 mM), with each [ON] at 1.0 μM. See main text for the definition of TA. **ON7** is so numbered to facilitate comparison with C_9-modified Invader probes.[24]
[b] Data previously reported in ref. 24.

Introduction of a single spermine bulge, as in **ON3:ON2** and **ON1:ON4**, reduces the T_m of the probe duplex by 2–7 °C vis-à-vis the conventional Invader probe, but also decreases the T_ms of the corresponding duplexes between individual probe strands and cDNA by an equivalent amount (e.g., compare T_m for **ON3** and **ON1**vs. cDNA). Accordingly, the driving force for dsDNA-recognition remains largely unchanged (TA = 26 and 28 °C

for **ON1:ON4** and **ON3:ON2**, respectively). Similar trends were seen with the corresponding C_9 bulge-containing Invader probes, except that destabilization was even more pronounced (i.e., incorporation of a C_9 unit reduced the T_m by 9–13 °C).[24]

Invader probe ON7:ON2, in which two spermine monomers are introduced – one monomer near each end of a strand – is very labile ($T_m = 23$ °C). The stability of the duplex between spermine-modified strand ON7 and cDNA is reduced by a similar amount and consequently ON7:ON2 displays similar dsDNA-recognition potential (TA = 30 °C) as conventional Invader probe ON1:ON2.

Introduction of two spermine bulges opposite of each other does not result in additional probe destabilization relative to the corresponding single bulge probes (compare probe T_ms for **ON3:ON4**, **ON3:ON2**, and **ON1:ON4**). This is surprising for at least two reasons: (i) additional electrostatic repulsion between two proximal and positively charged spermine units, akin to the repulsion observed for pseudocomplementary PNA with two opposing lysine units,[13] could have been expected, and (ii) Invader probes with two opposing C_9 bulges are exceptionally destabilized.[24] Since incorporation of the spermine monomers decreases the cDNA-affinity of both probe strands, **ON3:ON4** displays less pronounced dsDNA-targeting potential (TA = 20.5 °C) than single bulge probes **ON3:ON2** and **ON1:ON4** or conventional Invader probe **ON1:ON2**. Control probe **ON3c:ON4c**, which lacks the two sequential energetic hotspots, is also less destabilized than expected (compare T_m for **ON3c:ON4c**, **ON3c:cDNA** and ON4c:cDNA). Along similar lines, **ON7:ON4**, which features one isolated and two opposing spermine monomers, denatures less easily ($T_m = 28$ °C) than **ON7:ON2**, resulting in reduced dsDNA-recognition potential (TA = 20.5 °C). Collectively, these observations suggest that opposing spermine monomers decrease electrostatic repulsion between probe strands, presumably because their overriding effect is reduction of the net negative charge of the strands rather than mutual interference. T_m measurements performed at low ionic strengths corroborate this conclusion (Table 2.4). Control experiments entailing Invader probes with small PEG bulges could be carried out to further study the impact of electrostatics vis-à-vis solvation.

As expected, negative controls **ON3c:ON4c**, **ON3c:ON2**, and **ON1:ON4c** are not activated for dsDNA-recognition (TA values <−3.0 °C) as they lack the +1 interstrand zipper arrangements of X monomers that are necessary for probe activation.

An established electrophoretic mobility shift assay, utilizing a digoxigenin (DIG)-labelled DNA hairpin (**DH1**) as a model target,[19] was used to evaluate the dsDNA-targeting properties of the probes (Fig. 2.2a). **DH1** is comprised of a 13-mer double-stranded target segment and a T_{10} loop that covalently links the two stem strands at one end. Recognition of DH1 is expected to result in the formation of a slower-moving ternary recognition complex (RC) upon electrophoretic resolution of incubation mixtures on non-denaturing polyacrylamide gels (Fig. 2.2b).
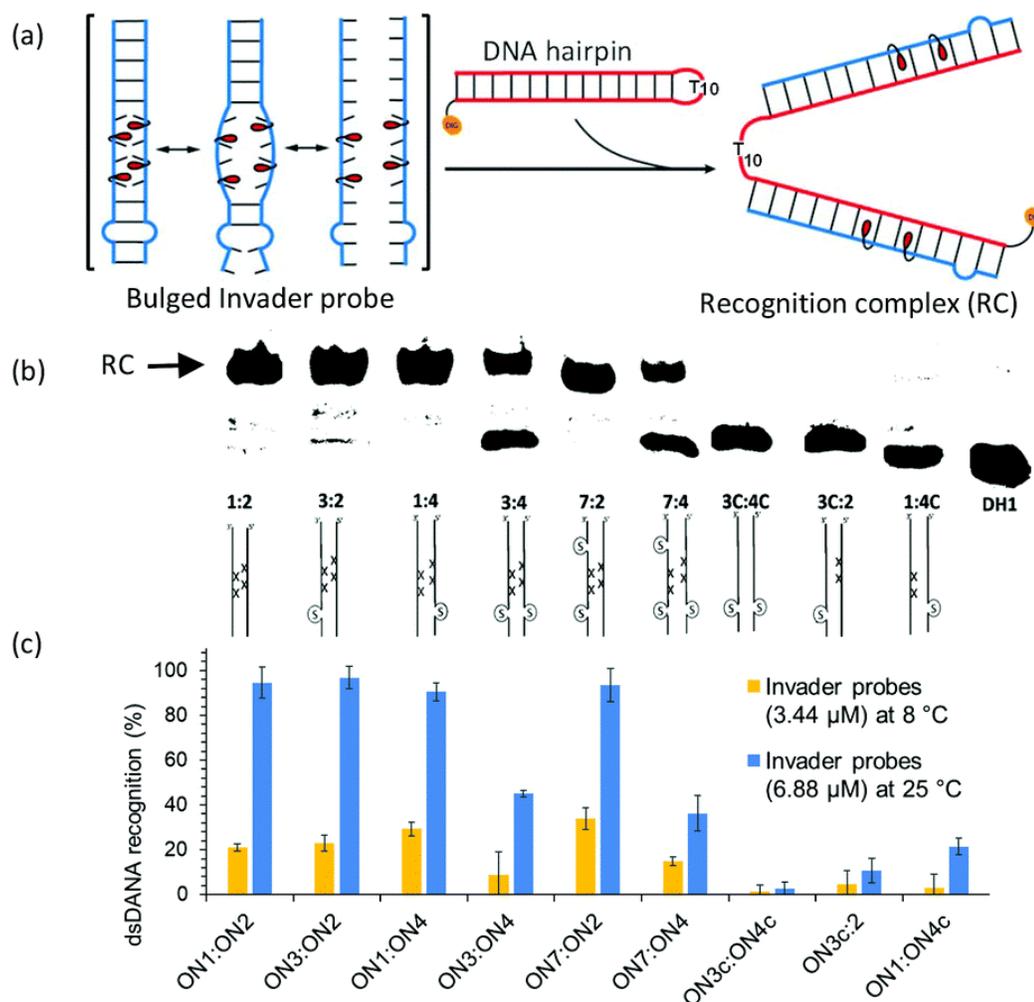
Figure 2.2 (a) Illustration of the electrophoretic mobility shift assay used to evaluate dsDNA-recognition of bulged Invader probes. (b) Representative gel electrophoretograms for recognition of model dsDNA target DH1 (34.4 nM) by various probes (6.88 µM; 25 °C). (c) Histogram depicting the average outcome of at least three recognition experiments at 8 °C or 25 °C using 100- or 200-fold probe excess, respectively (see Table 2.5 for tabulated data);

error bars represent standard deviation. DIG-labelled **DH1** (5′-GGTATATATAGGC-T_{10}-GCCTATATATACC-3′) (T_m = 58.5 °C) was incubated with pre-annealed probes in HEPES buffer (50 mM HEPES, 100 mM NaCl, 5 mM $MgCl_2$, pH 7.2, 10% sucrose, 1.44 mM spermine tetrahydrochloride) for 17 h.

An initial screen was performed in which DH1 was incubated with a 200-fold molar excess of probes at 25 °C. **ON3:ON2** and ON1:ON4 featuring a single spermine bulge, **ON7:ON2** with one spermine bulge near each end, and conventional Invader probe **ON1:ON2** result in highly efficient recognition of DH1 (Fig. 2.2b and c). In accordance with the observed TA values (Table 2.1), all other probes display moderate (**ON3:ON4** and **ON7:ON4**) or minimal recognition of **DH1** (**ON3c:ON2**, **ON1:ON4c**, and **ON3c:ON4c**) (see also Table 2.5). A subsequent screen in which a 100-fold molar probe excess was incubated with DH1 at 8 °C (Fig. 2.15), revealed, remarkably, that dsDNA-recognition also is possible at low experimental temperatures, albeit being less efficient. Spermine-containing Invader probes **ON3:ON2**, **ON1:ON4**, and **ON7:ON2** result in similar to slightly improved recognition of DH1vis-à-vis conventional Invader probe **ON1:ON2** (23–34% vs. 20%, respectively), whereas all other probes displayed minimal or no recognition (Fig. 2.2c and Table 2.5).

Recognition of **DH1** is more efficient at 25 °C (C_{50} values: 0.5 μM, 0.5 μM, and 1.2 μM, for **ON1:ON2**, **ON3:ON2**, and **ON7:ON2**, respectively, Table 2.2), presumably because localized probe denaturation is more pronounced at this experimental temperature (DH1 is still expected to be fully hybridized). Consistent with this, the improvement is most pronounced for the highest-melting probe **ON1:ON2** (C_{50} values of 5.6 μM and 0.5 μM at 8 °C and 25 °C, respectively), and least pronounced for the lowest-melting probe **ON7:ON2** (C_{50} values of 1.8 μM and 1.2 μM at 8 °C and 25 °C, respectively). Recognition of DH1 is less efficient at 37 °C (Table 2.2) presumably as the higher experimental temperature impacts the stability of the ternary recognition complexes in this particular sequence context. The trend is more pronounced for bulge-containing Invader probes, which display lower cDNA affinity than conventional Invader probe ON1:ON2 (compare T_ms for **ON3:cDNA**, **ON7:cDNA**, and **ON1:cDNA**, Table 2.1).
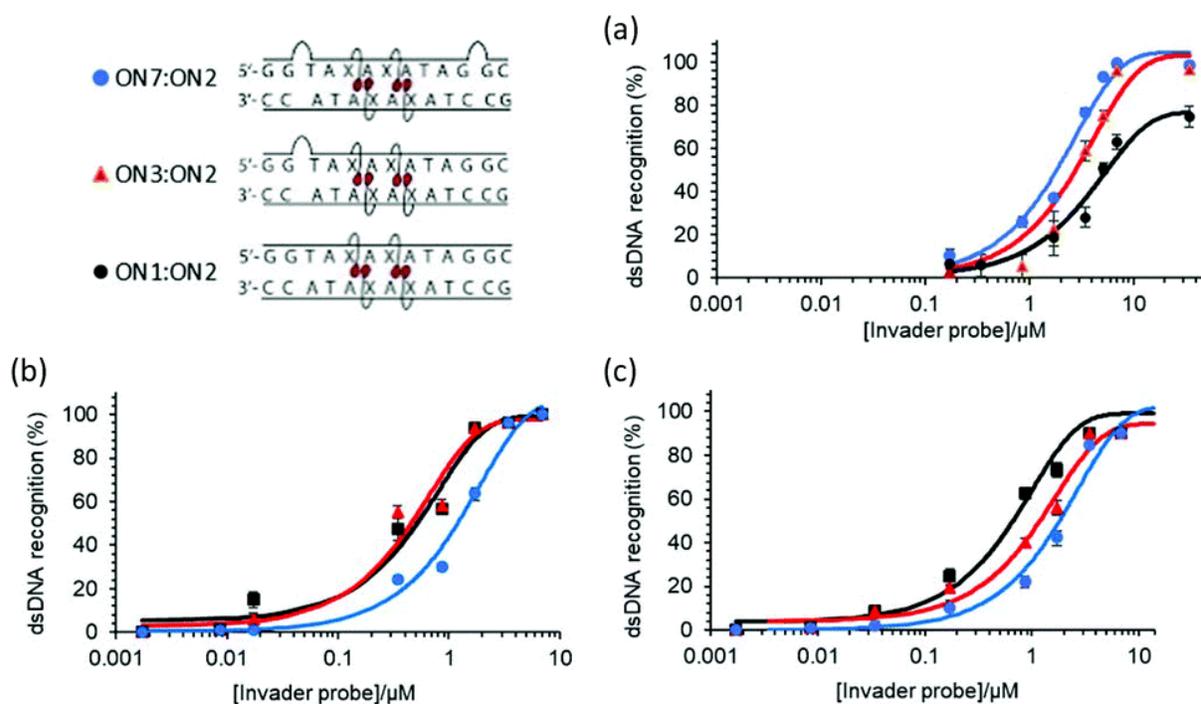
Figure 2.3. Dose-response curves for recognition of **DH1** by Invader probes **ON1:ON2**, **ON3:ON2** and **ON7:ON2** at (a) 8 °C, (b) 25 °C, and (c) 37 °C. Experimental conditions are outlined in Fig. 2.2**.** Bars denote standard deviation (three independent experiments).

Table 2.2 C_{50} values for recognition of DH1 by select Invader probes[a]

Probe	C_{50}, 8 °C (μM)	C_{50}, 25 °C (μM)	C_{50}, 37 °C (μM)
ON1:ON2	5.6	0.5	0.7
ON3:ON2	2.8	0.5	1.2
ON7:ON2	1.8	1.2	1.8

[a] Calculated from dose-response curves shown in Fig. 2.3.

Finally, the binding specificity of these three Invader probes was evaluated. Probes were incubated with DNA hairpins **DH2–DH7**, which have fully base-paired double-stranded stems that differ in sequence relative to the probes at either the 6- or 9-position (Fig. 2.4; for sequences and T_ms of **DH1–DH7** see Table 2.8). Remarkably, **ON3:ON2** and **ON7:ON2**, as well as, reference probe **ON1:ON2** display perfect discrimination of the singly mismatched dsDNA-targets at conditions that result in complete recognition of complementary target **DH1** (100-

fold probe excess, 25 °C, 17 h, Fig. 2.4). The double-stranded Invader probes likely display superb discrimination of mismatched dsDNA targets due to stringency clamping effects[26] and because binding to **DH2–DH7** would require the formation of recognition complexes with two energetically unfavorable mismatched duplexes.[27]
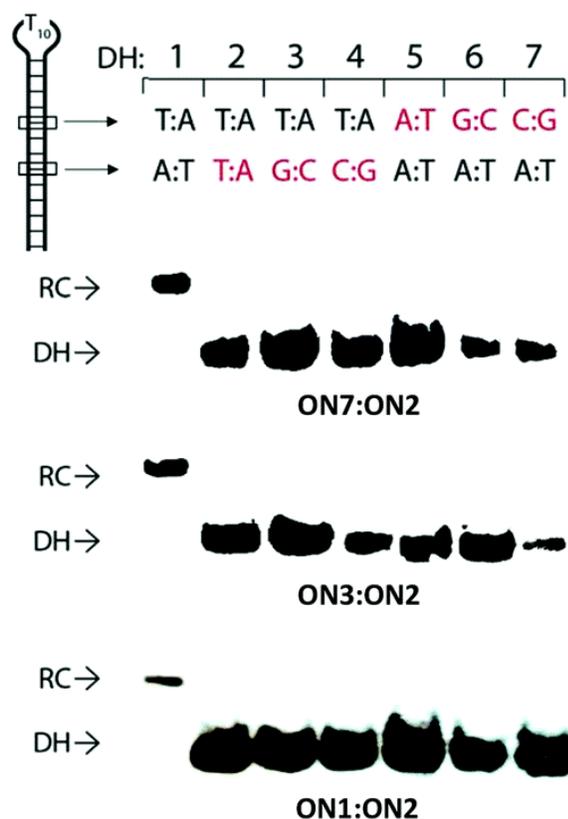
Figure 2.4 Binding specificity of spermine bulge-containing Invader probes. Top panel: Illustration of non-complementary targets **DH2–DH7** (see Table 2.6 for sequences). Other panels: Representative electrophoretograms from experiments in which a 100-fold excess of **ON7:ON2**, **ON3:ON2**, and **ON1:ON2** were incubated with **DH1–DH7** (34.4 nM) at 25 °C for 17 h. Experimental conditions are as outlined in Fig. 2.2.

2.3 Conclusion

In summary, Invader probes with one or two spermine bulges denature more readily than conventional Invader probes and result in improved and very specific recognition of mixed-sequence dsDNA targets at low incubation temperatures. Invader probes that are additionally activated through incorporation of spermine and other non-nucleotidic bulges are expected to

be particularly useful for recognition of high-melting DNA targets, i.e., extended and/or highly GC-rich regions, which currently represent challenging targets as the corresponding conventional Invader probes are very high-melting. Studies along these lines are ongoing.

2.4 Acknowledgement

2.5 Conflicts of interest

P.J.H. is an inventor on patents pertaining to Invader probes, which have been issued to the Univ. Idaho.

2.6 Supplementary data

Protocol - Synthesis and purification of ONs. Modified oligodeoxyribonucleotides (ONs) were synthesized on a computer-controlled DNA synthesizer (0.2 μmol scale) using long-chain alkyl amine-controlled pore glass (LCAA-CPG) solid support with a pore size of 500 Å. The corresponding DMTr-protected phosphoramidite of monomer **X** was prepared as previously described[28] and incorporated into ONs via hand-couplings (0.05 M in anhydrous acetonitrile; ~50-fold molar excess) using 0.01 M 4,5-dicyanoimidazole in anhydrous acetonitrile as the activator (15 min) and 0.02 M iodine in THF//H_2O/pyridine for extended oxidation (45 s). The DMTr-protected phosphoramidite of the spermine linker was obtained from a commercial source (Glen Research - **https://www.glenresearch.com/spermine-phosphoramidite.html**) and incorporated into ONs via hand-couplings using the abovementioned approach (3 min). Following synthesis, the columns were treated with 10% diethylamine in acetonitrile (5 min, room temperature) and rinsed with additional acetonitrile to remove the cyanoethyl protecting groups of the spermine monomers as recommended by the vendor to prevent acrylonitrile addition to the spermine units.[29] Subsequent treatment with 32% aq. ammonia (55 °C, 17 h) facilitated global deprotection and cleavage from the solid support. DMTr-protected ONs were purified via ion-pair reverse phase HPLC (XTerra MS C18 column: 0.05 M aq. triethyl ammonium acetate - acetonitrile gradient) and detrylated (80% aq. AcOH, 20 min) and

precipitated (NaOAc, NaClO$_4$, acetone, -18 °C, 16 h). The purity of the ONs was established using analytical HPLC (>85% purity), whereas the identity was verified by MALDI-MS (using 2,4,6-trihydroxyacetophenone as a matrix) or ESI-MS (modified ONs were dissolved in 2.5 M acetic acid, 25 mM imidazole and 25 mM piperidine in 80% aq. acetonitrile)[30] recorded on a Quadrupole Time-of-Flight (Q-TOF) mass spectrometer (Table 2.3). Raw ESI-signal peaks were deconvoluted using the Max Ent software provided with the spectrometer to obtain molecular ion peaks.

Table 2.3. MS data of ONs used in this study.[a]

ON	Sequence	Calculated *m/z* (M+H)$^+$	Observed *m/z* (M+H)$^+$
1	5'-GGTA**X**A**X**ATAGGC-3'	4445.5	4447.8
2	3'-CCATA**X**A**X**ATCCG-5'	4325.5	4326.9
3	5'-GG-**S**-TA**X**A**X**ATAGGC-3'	4855.5	4860.5
4	3'-CC-**S**-ATA**X**A**X**ATCCG-5'	4735.5	4735.1
3C	5'-GG-**S**-TATATATAGGC-3'	4423.0	4428.4
4C	3'-CC-**S**-ATATATATCCG-5'	4303.0	4302.0
7	5'-GG-**S**-TA**X**A**X**ATAG-**S**-GC-3'	5265.5	5268.8

[a] MALDI-MS was used to determine *m/z* for **ON1**, **ON2**, **ON4** and **ON4C**, whereas ESI-MS was used to determine *m/z* for **ON3**, **ON3C** and **ON7**.
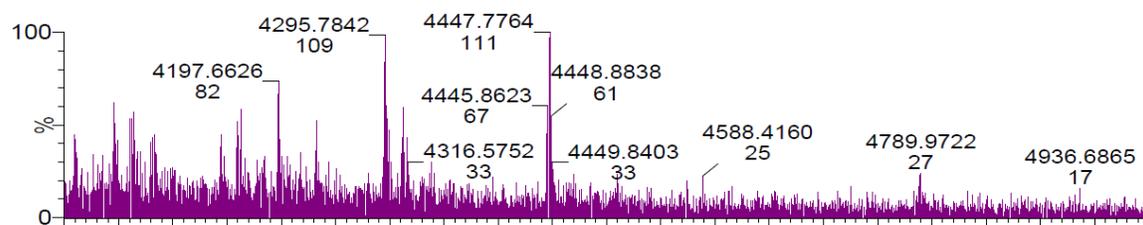
Figure 2.5. MALDI-MS spectrum of **ON1**.
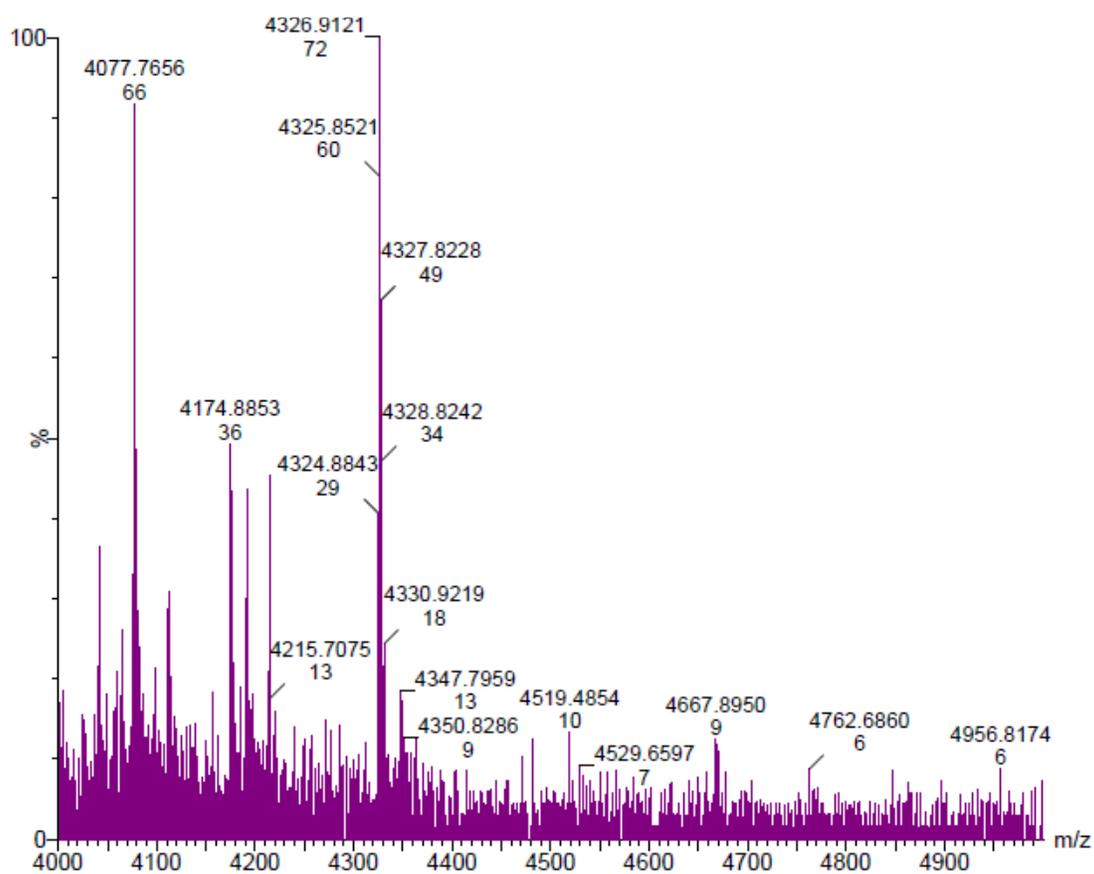
Figure 2.6. MALDI-MS spectrum of **ON2**.
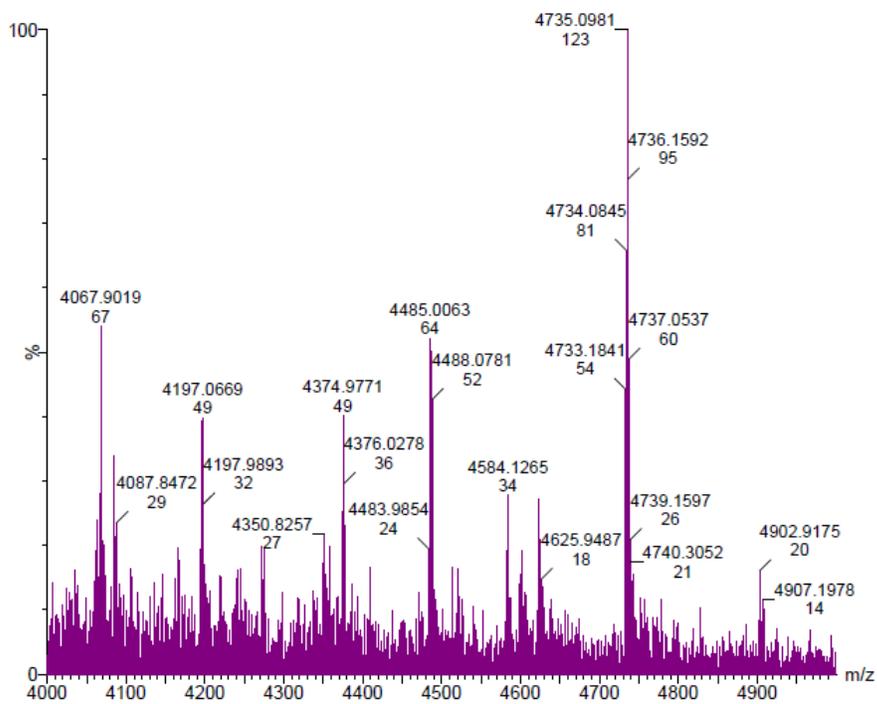
Figure 2.7. MALDI-MS spectrum of **ON4**.
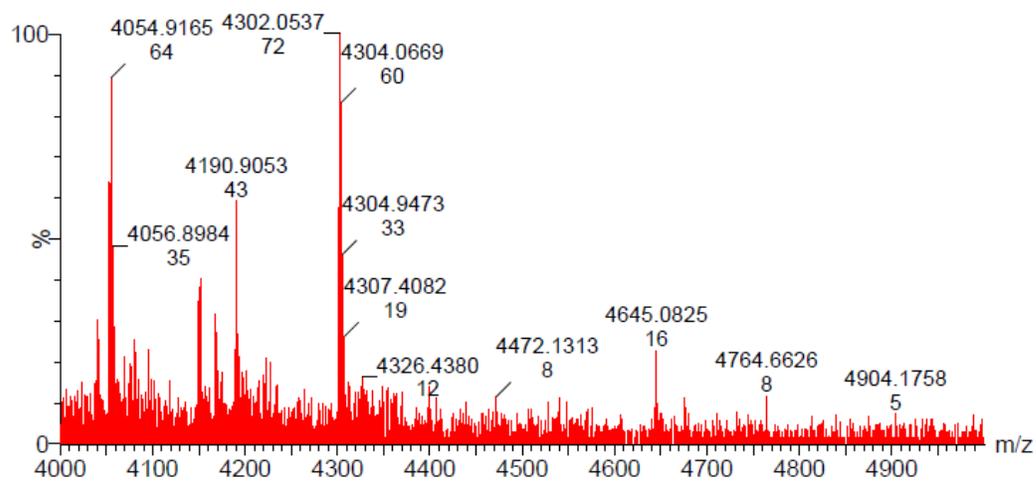
Figure 2.8. MALDI-MS spectrum of **ON4c**.
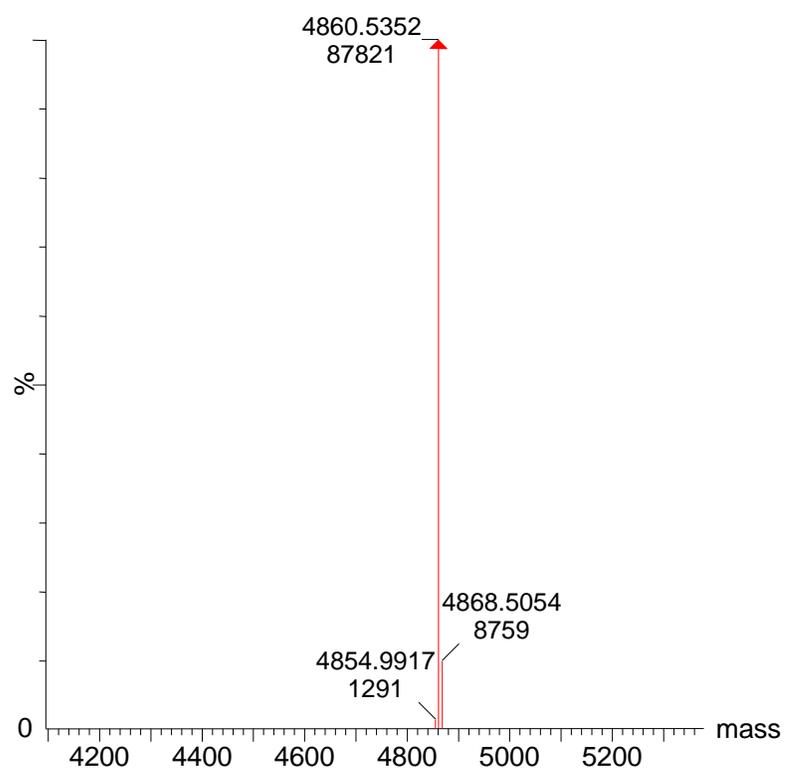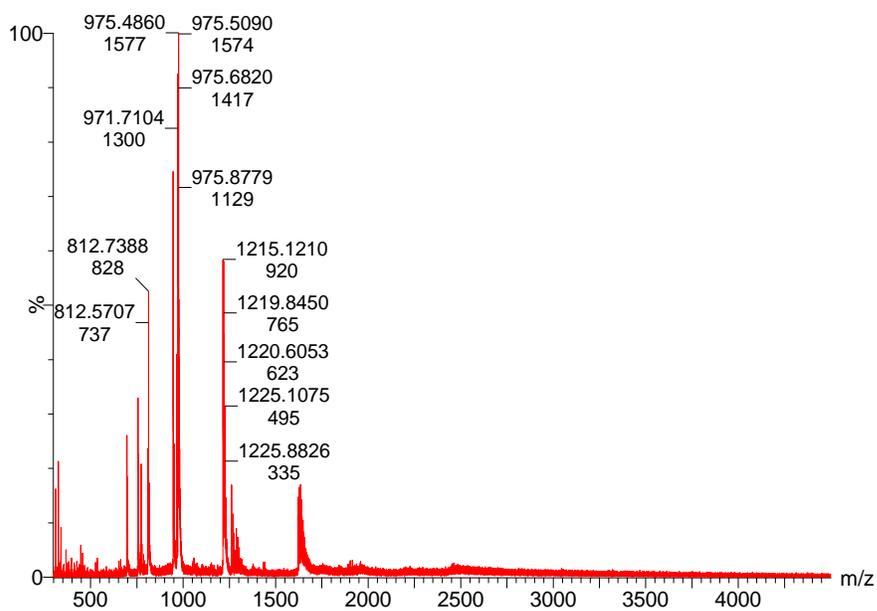
Figure 2.9. Unprocessed (upper panel) and deconvoluted (lower panel) ESI-MS spectrum of **ON3**.
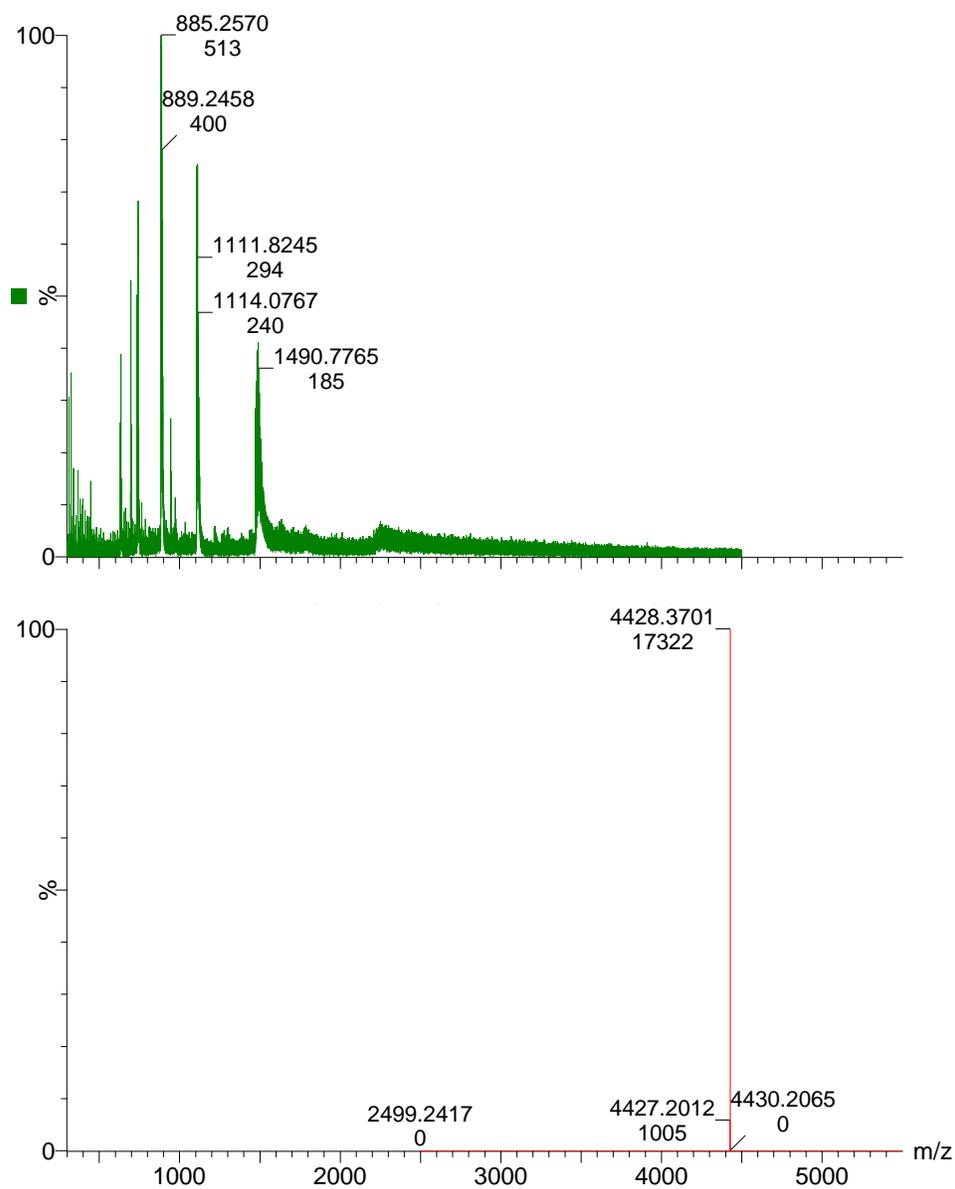
Figure 2.10. Unprocessed (upper panel) and deconvoluted (lower panel) ESI-MS spectrum of **ON3c**.
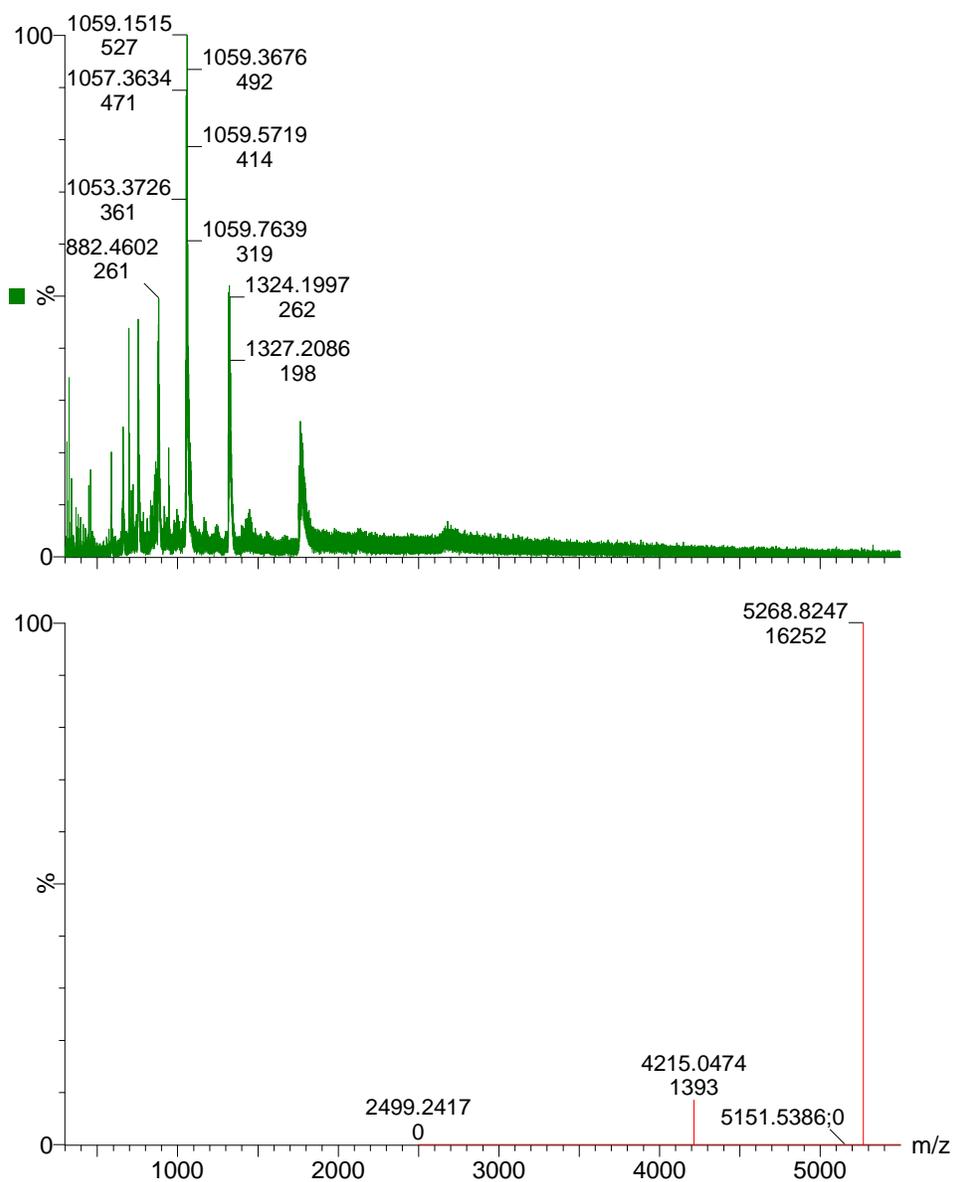
Figure 2.11. Unprocessed (upper panel) and deconvoluted (lower panel) ESI-MS spectrum of **ON7**.

ON1

ON3c

ON2

ON4c

ON3

ON7

ON4
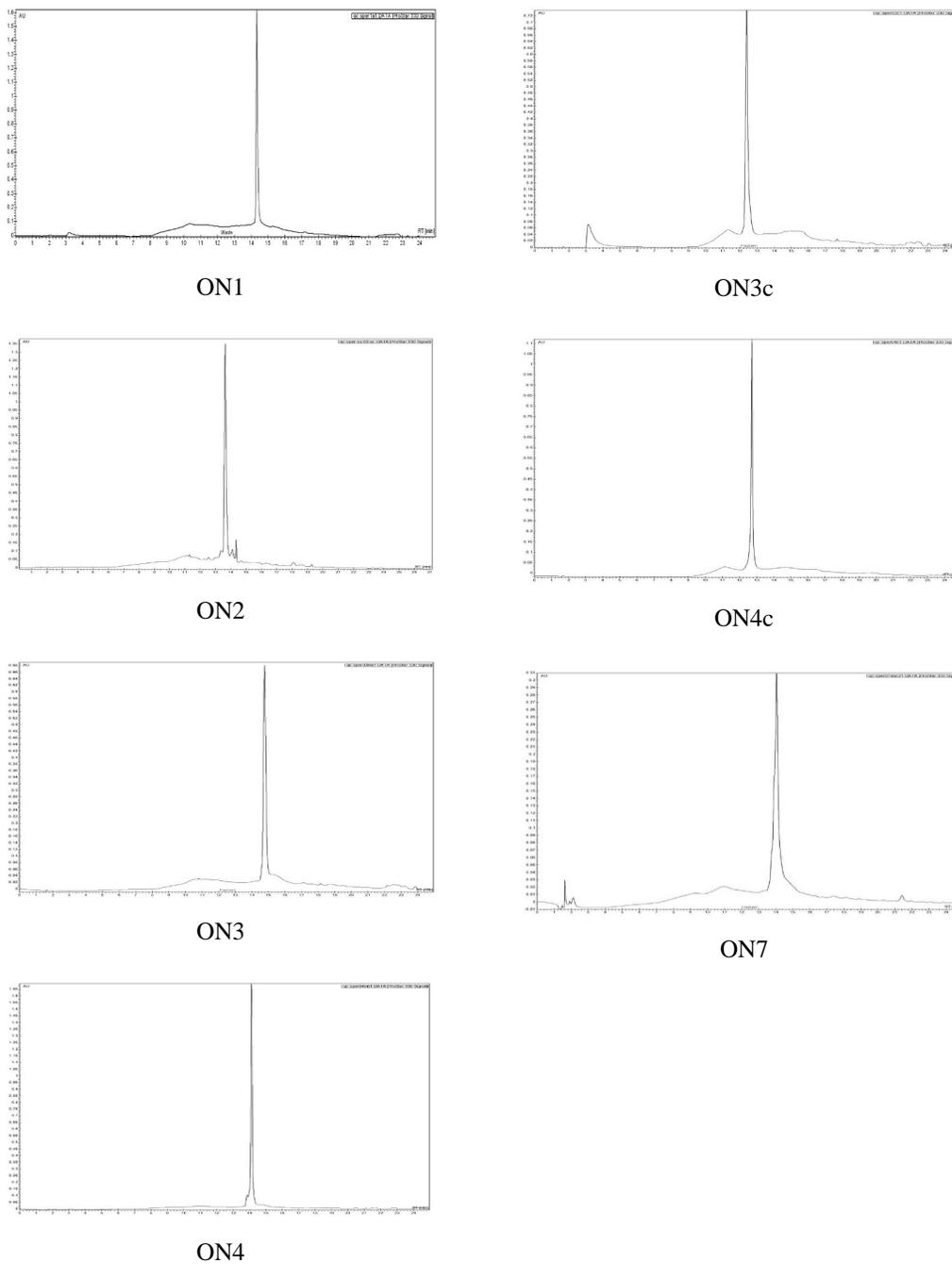
Figure 2.12. HPLC traces for modified ONs used in this study.

Protocol – thermal denaturation experiments. The concentrations of ONs were estimated using the following extinction coefficients (OD$_{260}$/µmol): G (12.01), A (15.20), T (8.40), C (7.05) and pyrene (22.4)[31]. Thermal denaturation temperatures (T_ms) of duplexes (1 µM final concentration of each strand) were measured on a Cary 100 UV/VIS spectrophotometer equipped with a 12-cell Peltier temperature controller and determined 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$_2$HPO$_4$ and 5 mM Na$_2$HPO$_4$). 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. A temperature range from 3 °C (low salt) or 10 °C (medium salt) to at least 20 °C above the duplex T_m was used, with T_ms determined as the average of two experiments within ±1.0 °C. A temperature ramp of 1 °C/min was used in all experiments.
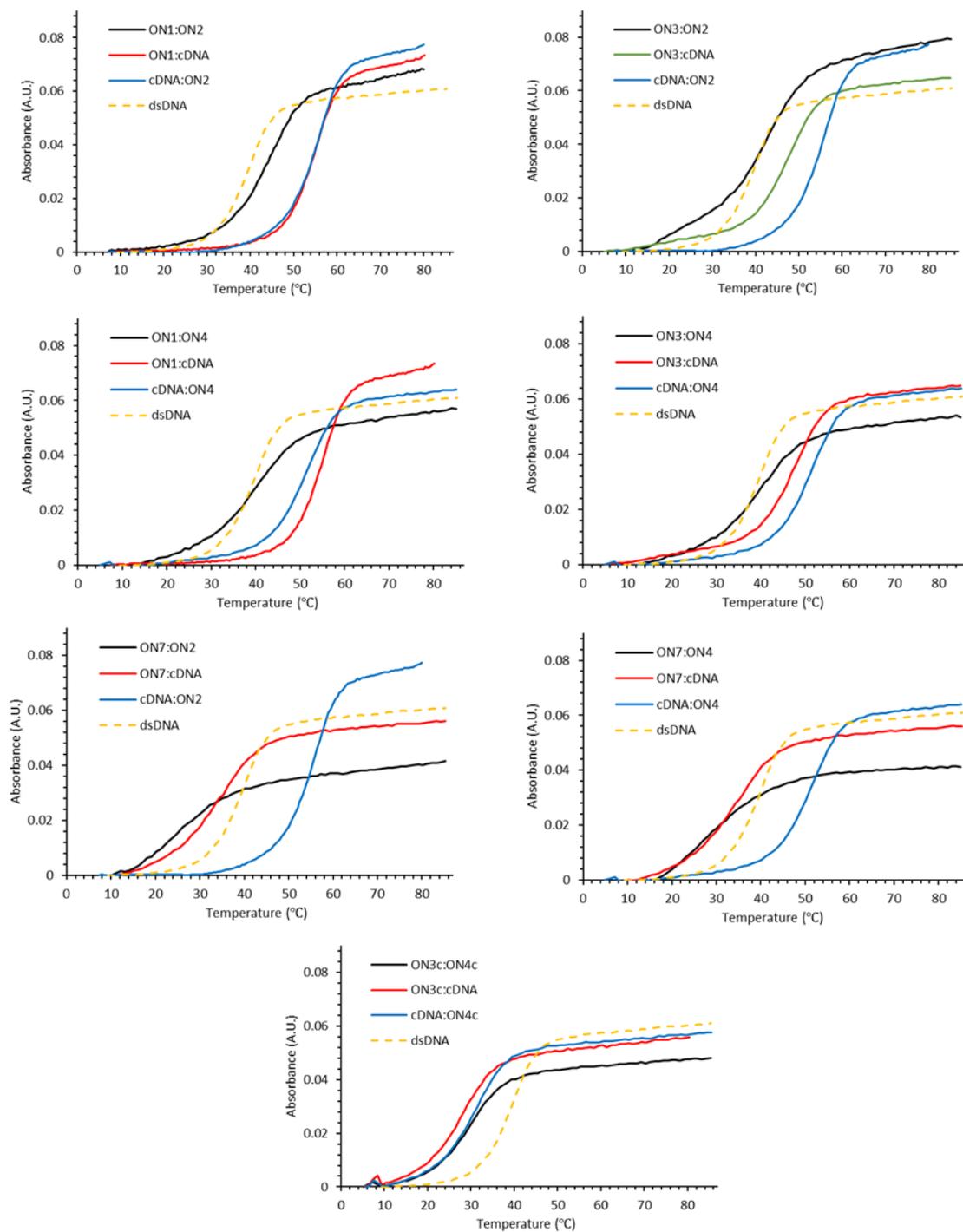
Figure 2.13. Representative thermal denaturation curves of Invader probes, duplexes between individual probe strands and cDNA, and unmodified reference duplexes recorded in *medium salt* buffer. Experimental conditions are described in Table 2.1.
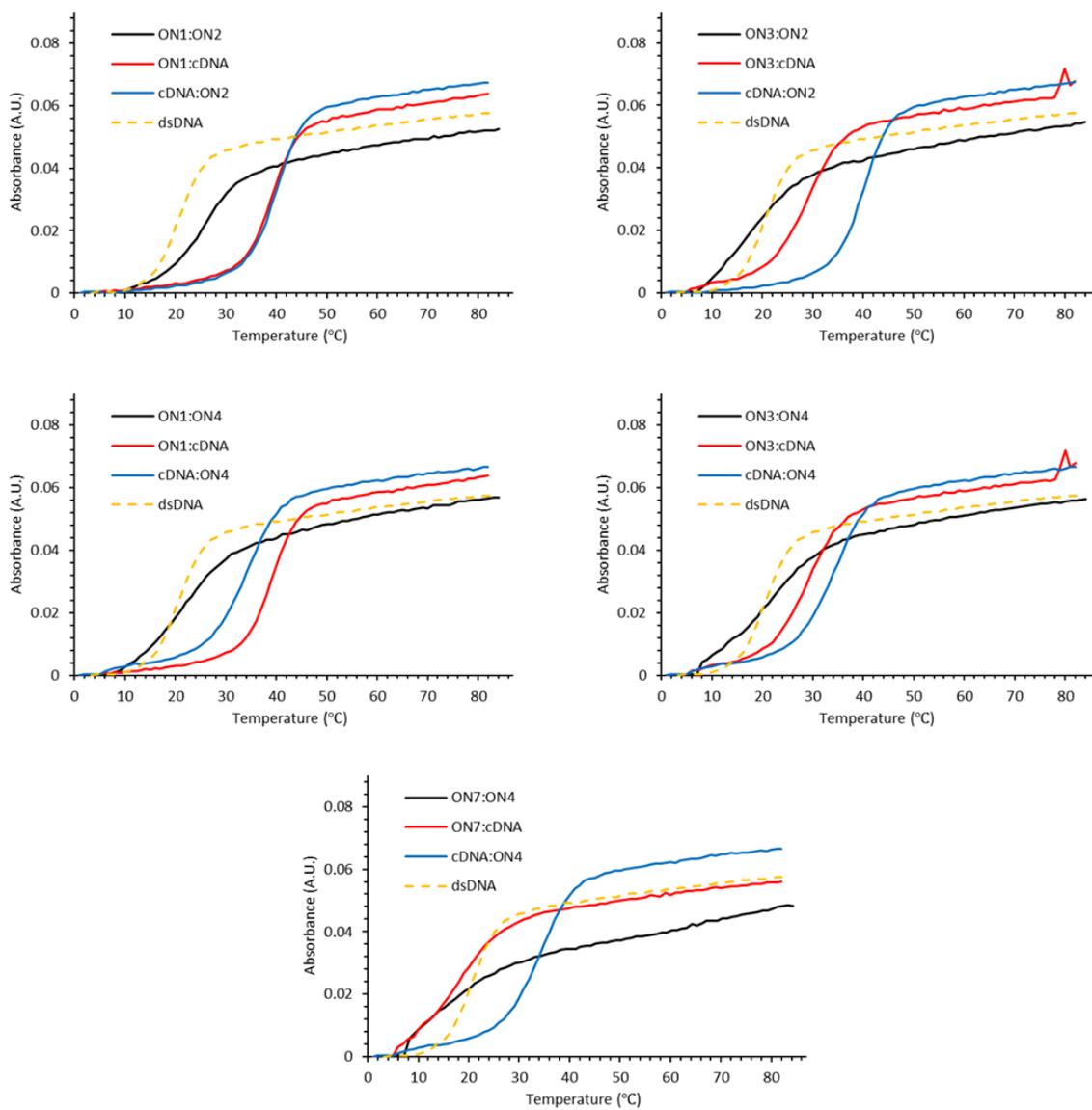
Figure 2.14. Representative thermal denaturation curves of Invader probes, duplexes between individual probe strands and cDNA, and unmodified reference duplexes recorded in *low salt* buffer. Experimental conditions are described in Table 2.4.

Trends of T_ms determined at low ionic strength. The T_ms of all duplexes are substantially lower at low ionic strength ([Na$^+$] = 10 mM, Table 2.4) than at medium ionic strength ([Na$^+$] = 110 mM, Table 2.1), as lower salt concentrations increase the electrostatic repulsion between polyanionic strands. Interestingly, probe duplexes are 2.5-6 °C more destabilized at low vis-à-vis medium ionic strength relative to the unmodified DNA reference duplex (e.g., ΔT_m for **ON1**:**ON2** = +3.5 °C vs +7.5 °C at low and medium strength, respectively), whereas duplexes between individual probe strands and cDNA duplexes are 0-3 °C more stable relative to the unmodified DNA reference duplex (e.g., ΔT_m for **ON1**:cDNA = +18 °C at either condition). Consequently, the probes are more strongly activated for dsDNA-recognition at low salt conditions (i.e., *TA* values are increased by 4.5-8.0 °C). The decrease in ΔT_m seen for probes at low ionic strength appears, in most part, to be an effect of the +1 interstrand zipper arrangements of **X** monomers rather than the spermine bulges. Thus, the relative effect of the change in salt concentration, calculated as ΔT_m (low salt) - ΔT_m (medium salt), is -2.5 °C for Invader probe **ON3**:**ON4** with two opposing spermine monomers and approximately -5.5 °C for single bulge Invader probes **ON3**:**ON2** and **ON1**:**ON4** and -4.0 °C for conventional Invader probe **ON1**:**ON2**. Evidently, **ON3**:**ON4** is far more stable than would be expected based on the T_m's of **ON3**:**ON2** and **ON1**:**ON4** suggesting that the overriding effect of the spermine bulges is a reduction of the net negative charge of the strands. We speculate that opposing spermine bulges are too far from each other to exert destabilizing electrostatic interactions (a fully extended protonated spermine residue spans a distance equivalent to ~5 base pairs;[25] an internal spermine bulge would, therefore, be expected to project a distance of up to 2-3 base pairs out from a duplex). We note that prior work, in which 5'-spermine-conjugated ONs were hybridized with complementary 3'-spermine-conjugated ONs, showed that opposing terminal spermine residues only have a limited effect on each other.[25]

Table 2.4. T_ms of probe duplexes and duplexes between individual probe strands and cDNA determined at *low ionic* strengths, as well as, thermal advantages (TAs) of probes.[a]

Probe	Sequence	T_m [ΔT_m] (°C) probe	upper strand vs cDNA	lower strand vs cDNA	TA (°C)
1:2	5'-G G T A X A X A T A G G C 3'-C C A T A X A X A T C C G	24.0 [+3.5]	38.5 [+18.0]	39.0 [+18.5]	33.0
3:2	5'-G G T A X A X A T A G G C 3'-C C A T A X A X A T C C G	16.0 [-4.5]	31.5 [+11.0]	39.0 [+18.5]	34.0
1:4	5'-G G T A X A X A T A G G C 3'-C C A T A X A X A T C C G	20.0 [-0.5]	38.5 [+18.0]	36.0 [+15.5]	34.0
3:4	5'-G G T A X A X A T A G G C 3'-C C A T A X A X A T C C G	21.5 [+1.0]	31.5 [+11.0]	36.0 [+15.5]	25.5
7:4	5'-G G T A X A X A T A G G C 3'-C C A T A X A X A T C C G	<15 [<-5.5]	21.0 [+0.5]	36.0 [+15.5]	>21.5
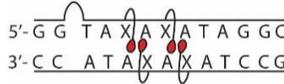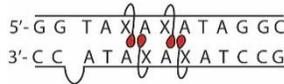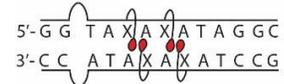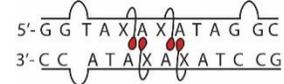
*ΔT_ms are calculated relative to the corresponding unmodified dsDNA (T_m = 20.5 °C). Thermal denaturation curves were recorded in low salt phosphate buffer ([Na$^+$] = 10 mM, pH 7.0 (NaH$_2$PO$_4$/Na$_2$HPO$_4$), [EDTA] = 0.2 mM) and each [ON] = 1.0 μM.

Protocol - electrophoretic mobility shift assay. DNA hairpins were obtained from commercial sources and were used without further purification. The DNA hairpins were DIG-labeled using the 2nd generation DIG Gel Shift Kit (Roche Applied Bioscience). Briefly, 11-digoxigenin-ddUTP was incorporated at the 3′-end of the hairpin (100 pmol) using a recombinant DNA terminal transferase. The reaction mixture was quenched through addition of EDTA (0.05 M), diluted to 68.8 nM in 2X HEPES buffer, and used without further processing. The recognition experiments were conducted essentially as previously reported.[19] Thus, Invader probes (variable 2X concentration in water) were annealed (90 °C for 2 min, followed by cooling to room temperature) and subsequently incubated with DIG-labeled DNA hairpins (34.4 nM final concentration in 1X HEPES buffer: 50 mM HEPES, 100 mM NaCl, 5 mM MgCl$_2$, pH 7.2, 10% sucrose, 1.44 mM spermine tetrahydrochloride) at either 8 °C ± 2 °C, 25 °C ± 2 °C, or 37 °C ± 2 °C for 17 hours. One microliter of loading dye (6X) was added and the reaction mixtures were loaded onto 12% non-denaturing TBE-PAGE gels (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 blotted onto positively charged nylon membranes (100 V, 30

min, ~4 °C) and cross-linked through exposure to UV light (254 nm, 5 × 15 W bulbs, 3 min). Membranes were incubated with anti-digoxigenin alkaline phosphatase F_{ab} fragments as recommended by the manufacturer and transferred to a hybridization jacket. Membranes were incubated with CSPD for 10 min at 37 °C, and chemiluminescence from the formed product was captured on X-ray films. Digital images of developed X-ray films were obtained using a Fluor-S MultiImager and quantified using appropriate software (Quantity One). The percentage of dsDNA recognition was calculated as the intensity of the recognition complex band relative to the total intensity of all bands. An average of three independent experiments is reported along with standard deviations (±). Non-linear S16 regression was used to fit data points from dose-response experiments. A script written for the "Solver" module in Microsoft Office Excel, S7 was used to fit the following equation to the data points: y = C + A (1 – e-kt) where C, A and k are constants. The resulting equation was used to calculate C50 values by setting y = 50 and solving for t.
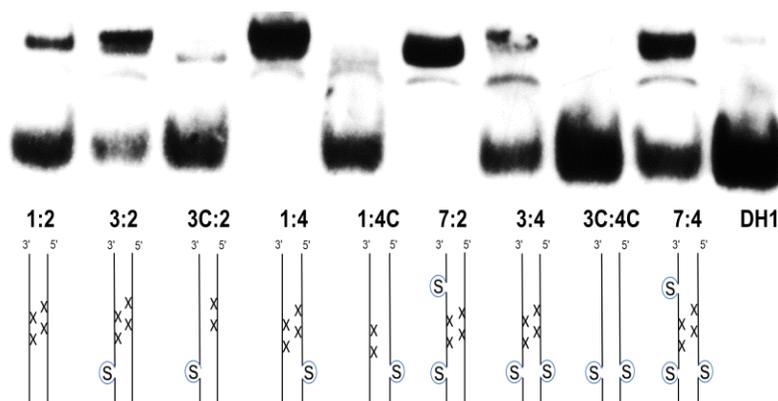
Figure 2.15. Representative electrophoretogram depicting recognition of model dsDNA target **DH1** using 100-fold molar excess (3.44 µM) of different probes at 8 °C. Experimental conditions are otherwise as described in Figure 2.2.

Table 2.5. Levels of recognition of DNA hairpin **DH1** using a 200-fold molar excess (at 25 °C) or 100-fold molar excess (at 8 °C) of various double-stranded probes.[a]

Probe	Sequence	Recognition (%)	
		8 °C	25 °C
1:2	5'-G G T A X A X A T A G GC 3'-C C A T A X A X A T C CG	20±2	>90
3:2	5'-G G T A X A X A T A G GC 3'-C C A T A X A X A T C CG	23±4	>90
1:4	5'-G G T A X A X A T A G GC 3'-C C A T A X A X A T C CG	29±3	>90
3:4	5'-G G T A X A X A T A G GC 3'-C C A T A X A X A T C CG	9±8	45±1
7:2	5'-G G T A X A X A T A G GC 3'-C C A T A X A X A T C CG	34±5	>90
7:4	5'-G G T A X A X A T A G GC 3'-C C A T A X A X A T C CG	15±2	36±8
3c:4c	5'-G G T A T A T A T A G GC 3'-C C A T A T A T A T C CG	<5	<5
3c:2	5'-G G T A T A T A T A G GC 3'-C C A T A X A X A T C CG	<5	11±5
1:4c	5'-G G T A X A X A T A G GC 3'-C C A T A T A T A T C CG	<5	21±3

[a] Experiments were performed in triplicate. Conditions are described in Figure 2.2.

Figure 2.16. Representative electrophoretograms illustrating dose-response experiment between dsDNA target **DH1** (34.4 nmol) using a variable molar excess of Invader probes **ON1**:**ON2**, **ON3**:**ON2**, and **ON7**:**ON2** at a) 8 °C, b) 25 °C, and c) 37 °C. Experimental conditions are otherwise as described in Figure 2.2.

Table 2.6. Seqeuences and T_ms of DNA hairpins used in this study.[a]

DH	Sequence	T_m (°C)
1	5'-GGTATATATAGGC 3'-CCATATATATCCG — T$_{10}$	58.5
2	5'-GGTAT*T*TATAGGC 3'-CCATA*A*ATATCCG — T$_{10}$	60.5
3	5'-GGTAT*G*TATAGGC 3'-CCATA*C*ATATCCG — T$_{10}$	63.5
4	5'-GGTAT*C*TATAGGC 3'-CCATA*G*ATATCCG — T$_{10}$	63.0
5	5'-GGTATATA*A*AGGC 3'-CCATATAT*T*TCCG — T$_{10}$	60.0
6	5'-GGTATATA*G*AGGC 3'-CCATATAT*C*TCCG — T$_{10}$	62.5
7	5'-GGTATATA*C*AGGC 3'-CCATATAT*G*TCCG — T$_{10}$	62.5

[a] For experimental conditions, see Table 2.1. Data previously published in reference S6.

2.7 References and notes.

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

2 K. Kaihatsu, B. A. Janowski and D. R. Corey, *Chem. Biol.*, 2004, **11**, 749-758.

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

4 Y. Hari, S. Obika and T. Imanishi, *Eur. J. Org. Chem.* 2012, 2875-2887.

5 D. A. Horne and P. B. Dervan, *J. Am. Chem. Soc.* 1990, **112**, 2435-2437.

6 V. V. Filichev, C. Nielsen, C. H. Bomholt and E. B. Pedersen, *Angew. Chem., Int. Ed.* 2006, **45**, 5311-5315.

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

8 K. Kaihatsu, R. H. Shah, X. Zhao and D. R. Corey, *Biochemistry*, 2003, **42**, 13996-14003.

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

10 P. R. Bohlander, T. Vilaivan and H. A. Wagenknecht, *Org. Biomol. Chem.*, 2015, **13**, 9223-9230.

11 I. V. Kutyavin, R. L. Rhinehart, E. A. Lukhtanov, V. V. Gorn, R. B. Meyer Jr and H. B. Gamper Jr, *Biochemistry*, 1996, **35**, 11170-11176.

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

13 T. Ishizuka, J. Yoshida, Y. Yamamoto, J. Sumaoka, T. Tedeschi, R. Corradini, S. Sforza and M. Komiyama, *Nucleic Acids Res.*, 2008, **36**, 1464-1471.

14 M. Hibino, Y. Aiba, Y. Watanabe and O. Shoji, *ChemBioChem*, 2018, **19**, 1601-1604.

15 V. V. Filichev, B. Vester, L. H. Hansen and E. B. Pedersen, *Nucleic Acids Res.*, 2005, **33**, 7129-7137.

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

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

18 P. J. Hrdlicka, T. S. Kumar and J. Wengel, *Chem. Commun.*, 2005, 4279-4281.

19 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.

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

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

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

23 C. Tsai, S. Jain and H. M. Sobell, *J. Mol. Biol.*, 1977, **114**, 301-315.

24 D. C. Guenther, S. Karmakar and P. J. Hrdlicka, *Chem. Commun.*, 2015, **51**, 15051-15054.

25 R. Noir, M. Kotera, B. Pons, J.-S. Remy, and J.-P. Behr, *J. Am. Chem. Soc.*, 2008, **130**, 13500-13505.

26 V. V. Demidov and M. D. Frank-Kamenetskii, *Trends Biochem. Sci*. 2004, **29**, 62-71.

27 S. X. Chen, D. Y. Zhang and G. Seelig, *Nature Chem*., 2013, **5**, 782-789.

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

29 Glen Report 24.11: **https://www.glenresearch.com/reports/gr24-11** - accessed July 30, 2019.

30 D. C. Muddiman, X. Cheng, H. R. Udseth and R. D. Smith, *J. Am. Soc. Mass Spectrom.*, 1996, **7**, 697-706.

31 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.

32 A. M. Brown, Comput. Methods Prog. Biol., 2001, 65, 191-200.

CHAPTER 3: Recognition of mixed-sequence dsDNA using toehold Invader probes

Shiva P. Adhikari,[a] Philip Vukelich,[a] Dale C. Guenther,[a] Saswata Karmakar[a] and Patrick J. Hrdlicka.[a]

[a] Department of Chemistry, University of Idaho, Moscow, ID-83844, USA.

Abstract

Chemically modified oligonucleotides enabling sequence-specific recognition of double-stranded (ds) DNA have tremendous potential as tools in diagnostics, gene editing, and molecular therapy. The Hrdlicka laboratory has designed so-called Invader probes, which are short DNA duplexes with +1 interstrand arrangements of O2'-intercalator-functionalized RNA monomers as the central design feature that activates Invader probes for sequence-unrestricted dsDNA recognition. The design relies on large stability differences between probe duplexes and recognition complexes to drive dsDNA recognition. These double-stranded probes exhibit avidity in binding to dsDNA via invasion modes, offer the promise of favorable binding affinity and specificity, and straightforward design. Probes with single-stranded overhangs called toeholds have been used extensively to facilitate enzyme-free strand displacement reactions. In the present study, we evaluated Invader probes with modified toeholds to improve mixed-sequence dsDNA recognition. Invader probes with appropriately designed toeholds display more efficient, and faster recognition of mixed-sequence dsDNA targets.

3.1. Introduction

Various attempts to target dsDNA using molecular tools led to the development of DNA targeting chemistry which mainly includes modification of nucleobases, sugar, and backbone.[1,2] The development of oligonucleotide therapeutics is focused on the modification of oligonucleotide chemistries for enhancing affinity, stability, and delivery of oligonucleotide probes to the nucleic acid target.[2,3,4] The beauty of oligonucleotide therapeutics is its potentiality to treat many rear and previously untreatable diseases. At present, complete human genome is available and DNA targeting approach could serve as an alternating useful tool in fundamental genomic research, diagnostics and oligonucleotide-based therapeutics.[5] Recent development in

chemical probes capable of detection, regulation and manipulation of genes has a great significance for targeting biological DNA. Several approaches have been introduced for targeting dsDNA, i.e., triplex forming oligonucleotides (TFOs), peptide nucleic acids (PNAs), and minor groove binding polyamides. However minor groove binding polyamides are limited for targeting short sequences (<8 bp) and are not useful for targeting long sequences of DNA as they lose shape complementarity.[6,7] TFOs form hoogsteen base pairing in the major groove of DNA but is limited to binding only with long polypurine regions.[8,9] PNAs have strong affinity to complementary sequences of DNA and can bind through Watson-crick and hoogsteen base pairing although it has same limitations as TFOs that requires long polypurine region for binding.[10,11] Conformationally restricted γ- PNAs are capable of invading mixed sequences of double stranded DNA (dsDNA), forming Watson crick base pairing with one of the sequences while the other sequence forms a D-loop, however, the enantioselective synthesis of chiral monomers is challenging.[12,13] Recently, CRISPR/Cas technology, RNA guided endonucleases, have gained enough attention because of its potentiality in targeting DNA for in-vivo applications albeit the challenges of PAM sequences restrictions, delivery and off-target binding and editing activities must be addressed in near future.[14,15]

An alternative class of dsDNA targeting oligonucleotide probes, Invader probes, was previously developed as a tool for precise sequence unrestricted recognition of mixed sequence of dsDNA targets in physiological conditions.[16] Invader probes are short DNA duplexes that are modified with one or more +1 interstrand zipper arrangements of intercalator-functionalized nucleotides like 2′-O-(pyren-1-yl)methyl-RNA. This unique monomer arrangement forces two intercalating moieties to compete for the same inter-base-pair region, resulting in localized unwinding[17] - and probe destabilization near these energetic hotspots[18] - as the nearest-neighbor exclusion principle, which asserts that intercalation is anti-cooperative at adjacent sites[19] - is violated. Individual strands of Invader probes illustrate exceptionally high affinity towards complementary DNA (cDNA) of target because of highly stabilizing stacking interactions between the pyrene intercalator and flanking base-pairs and the formation of highly stable recognition complex compared to probe duplex and target dsDNA serves as the driving force for invasion of mixed sequence dsDNA target.[18,20] 2′-O-(pyren-1-yl)methyl-RNA functionalized double-stranded DNA Invader probes exhibit avidity in binding to dsDNA via invasion modes, offer the promise of favorable binding thermodynamics, high binding

specificity, and straightforward design. Different modalities of Invader probes were designed, and invasion of dsDNA was examined eg. varying in the position, number and distance between energetic hotspots,[16] inserting bulky substituents in the pyrene moieties,[21] introducing non-nucleotidic bulges in the backbone,[22,23] phosphorothioated backbone,[16] and so on.
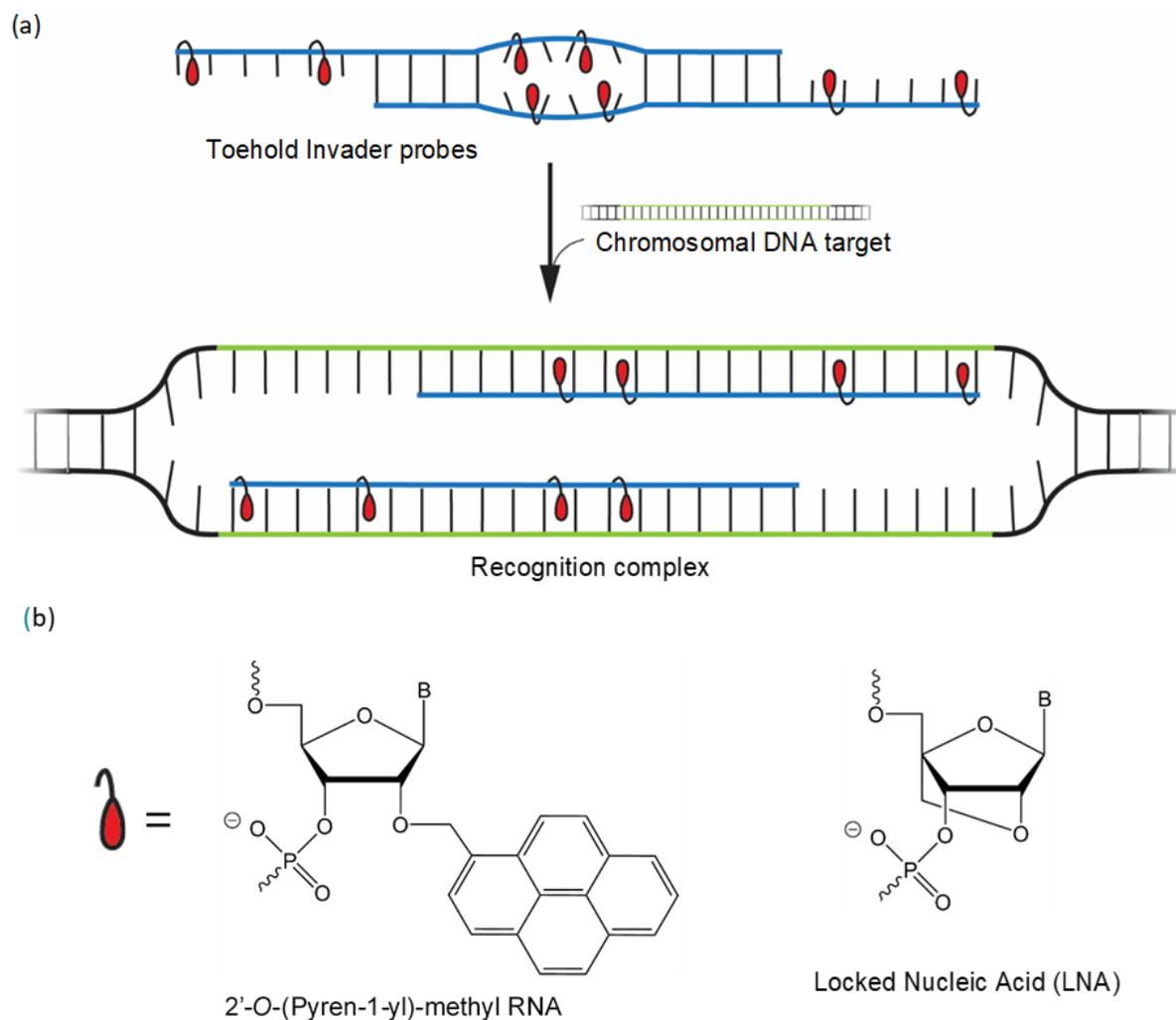
Figure 3.1. (a) Illustration of dsDNA-recognition principle using toehold Invader probes. (b) Structures of 2'-*O*-(pyren-1-yl)methyl-RNA and LNA monomers used herein.

DNA toeholds have been used extensively for non-enzymatic toehold-mediated strand displacement reactions between dsDNA and ssDNA.[24,25] The concept of single-stranded overhangs to increase binding affinity has been utilized in dsDNA-targeting probes like tcPNA and Zorro LNA.[26,27] However, there are drawbacks to this design as the LNA-modified

overhangs may self-hybridize, impeding invasion of dsDNA. Similarly, tcPNA requires the presence of a polypurine-rich region in the target which limits this probe for mixed dsDNA target recognition.[28] In the present study, we introduced nucleotide overhangs (toeholds or sticky overhangs) on Invader probes also called toehold Invader probes for recognition of longer dsDNA target region. The dangling ends of toeholds make the probe duplex stable by increasing the intercalation.[29] Further, in addition to the regular destabilizing effect of the energetic hotspots and their stabilizing effect in probe-target duplexes, introduction of single-stranded overhang regions to Invader probes is expected to increase the overall thermodynamic driving force and facilitate recognition of dsDNA targets, since the additional nucleotides of the overhang region of toehold Invaders will allow for formation of additional base-pairs, during dsDNA-recognition compared to conventional Invader probes (Fig. 3.1). Molecular constructs having overhanging region were used to improve the selectivity for recognition of dsDNA targets.[29] Invader probes with appropriately designed toeholds display more efficient recognition of mixed-sequence dsDNA targets vis-à-vis conventional Invader probes. Moreover, the toehold region modified with locked nucleic acid (LNA)-modified overhangs demonstrate improved binding specificity.

3.2. Results and discussion

Design and synthesis of probes. A series of oligodeoxyribonucleotides (ONs), modified with 2'-*O*-(pyren-1-yl)methyl-RNA monomers, were prepared as previously reported[30,31] (Tables 3.1 and 3.3). This enabled assembly of several double-stranded probes, i.e., a conventional, blunt-ended 13-mer Invader probe **ON1**:**ON2** featuring two central energetic hotspots, and extensions thereof with 3- or 6-mer single-stranded 5'-overhangs that either are unmodified or additionally modified with one or two 2'-*O*-(pyren-1-yl)methyl-RNA monomers (**ON3**:**ON4**-**ON9**:**ON10**). Additionally, a number of control probes were prepared, including double-stranded probes that lack the energetic hotspots but feature modified or unmodified 6-mer 5'-overhangs (**ON11**:**ON12** and **ON13**:**ON14**), as well as conventional blunt-ended 19-mer Invader probes with four energetic hotspots (**ON15**:**ON10** and **ON9**:**ON16**). The probes were designed to recognize complementary mixed-sequence regions embedded within a 33-mer model dsDNA.

Thermal denaturation properties. Thermal denaturation temperatures (T_ms) were determined for the double-stranded probes, as well as, duplexes between individual probe strands and the corresponding 33-mer DNA strand harboring the complementary region (Table 3.1). As expected from our previous work, conventional 13-mer Invader probe **ON1**:**ON2** is labile (T_m = 42.5 °C).[22,23] Double-stranded probes with single-stranded 5'-overhangs display comparable stability (T_ms = 42-44 °C for **ON3**:**ON4**, **ON5**:**ON6**, and **ON7**:**ON8**), except when the 6-mer overhang is modified, in which case moderate stabilization is observed (T_m = 49 °C for **ON9**:**ON10**). Presumably, the stabilization is a result of capping effects. The contribution of the energetic hotspots to duplex stability is minimal (compare T_m values for **ON9**:**ON10** and **ON11**:**ON12**). As expected, given the greater number of base pairs, conventional 19-mer Invader probes denature at higher temperatures (T_m ~ 66 °C for **ON15**:**ON10** and **ON9**:**ON16**).

Duplexes between the individual strands of conventional 13-mer Invader probes and the corresponding 33-mer single-stranded DNA harboring the complementary region, i.e., **ON1**:**DNA2** or **DNA1**:**ON2**, are more stable than the 13-mer probe duplex (T_m ~54.5 °C vs 42.5 °C) but less stable than the 33-mer DNA duplex (T_m ~54.5 °C vs 72 °C). The presence of 3- or 6-mer unmodified overhangs progressively increases probe-target duplex stability (T_ms between 53.5 and 67.0 °C for **ON3**:**DNA2**/**ON4**:**DNA1**/**ON7**:**DNA2**/**ON8**:**DNA1**), reflecting the increased number of base-pairs in the probe-target duplexes, i.e., 16 and 19 base-pairs (bps) vis-à-vis 13 bps as in **ON1**:**DNA2** and **DNA1**:**ON2**. The presence of 2'-*O*-(pyren-1-yl)methyl-RNA monomers in the overhangs further increases the stability of the duplexes (e.g., compare ΔT_ms for **ON9**:**DNA2** and **ON7**:**DNA2**). In fact – and rather remarkably – **ON9**:**DNA2** and **ON10**:**DNA1**, which comprise 19 bps and single-stranded overhangs on either side of the probe strand, are more stable than the unmodified 33-mer reference DNA duplex (ΔT_m = +3.5 °C). The well-known stabilizing effect of the 2'-*O*-(pyren-1-yl)methyl-RNA monomers is also evident upon comparing probe-target duplexes with increasing modification levels (e.g., T_m = 54.5, 64.0, and 75.5 °C, for **ON13**:**DNA2**, **ON11**:**DNA2**, and **ON9**:**DNA2** respectively).

Table 3.1. Sequences of probes used in this study, T_ms of probe duplexes and duplexes between individual probe strands and DNA targets, and thermal advantages (*TA*s) of probes.[a]

		T_m [ΔT_m] (°C)			
Probe	**Sequence**	**probe duplex**	**5'-probe vs DNA2**	**3'-probe vs DNA1**	*TA* (°C)
ON1	5'-GGTA<u>U</u>A<u>U</u>ATAGGC	42.5	54.5	54.0	
ON2	3'-CCATA<u>U</u>A<u>U</u>ATCCG	[-29.5]	[-17.5]	[-18.0]	-6.0
ON3	5'-ACA-GGTA<u>U</u>A<u>U</u>ATAGGC	42.0	53.5	64.0	
ON4	3'-CCATA<u>U</u>A<u>U</u>ATCCG-GCG	[-30.0]	[-18.5]	[-8.0]	3.5
ON5	5'-A<u>C</u>A-GGTA<u>U</u>A<u>U</u>ATAGGC	44.0	63.5	65.5	
ON6	3'-CCATA<u>U</u>A<u>U</u>ATCCG-G<u>C</u>G	[-28.0]	[-8.5]	[-6.5]	13.0
ON7	5'-TGCACA-GGTA<u>U</u>A<u>U</u>ATAGGC	44.0	60.0	67.0	
ON8	3'-CCATA<u>U</u>A<u>U</u>ATCCG-GCGTAT	[-28.0]	[-12.0]	[-5.0]	11.0
ON9	5'-<u>U</u>GCA<u>C</u>A-GGTA<u>U</u>A<u>U</u>ATAGGC	49.0	75.5	75.5	
ON10	3'-CCATA<u>U</u>A<u>U</u>ATCCG-G<u>C</u>GTA<u>U</u>	[-23.0]	[+3.5]	[+3.5]	30.0
ON11	5'-<u>U</u>GCA<u>C</u>A-GGTATATATAGGC	50.0	64.0	64.5	
ON12	3'-CCATATATATCCG-G<u>C</u>GTA<u>U</u>	[-22.0]	[-8.0]	[-7.5]	6.5
ON13	5'-TGCACA-GGTATATATAGGC	39.5	54.5	55.5	
ON14	3'-CCATATATATCCG-GCGTAT	[-32.5]	[-17.5]	[-16.5]	-1.5
ON15	5'-GGTA<u>U</u>A<u>U</u>ATAGGC-<u>C</u>GCA<u>U</u>A	66.0	76.5	75.5	
ON10	3'-CCATA<u>U</u>A<u>U</u>ATCCG-G<u>C</u>GTA<u>U</u>	[-6.0]	[+4.5]	[+3.5]	14.0
ON9	5'-<u>U</u>GCA<u>C</u>A-GGTA<u>U</u>A<u>U</u>ATAGGC	66.0	75.5	76.0	
ON16	3'- A<u>C</u>GTG<u>U</u>-CCATA<u>U</u>A<u>U</u>ATCCG	[-6.0]	[+3.5]	[+4.0]	13.5
ON17	5'-<u>t</u>GCA<u>c</u>A-GGTA<u>U</u>A<u>U</u>ATAGGC	47.5	71.5	73.0	
ON18	3'-CCATA<u>U</u>A<u>U</u>ATCCG-G<u>c</u>GTA<u>t</u>	[-24.5]	[-0.5]	[+1.0]	25.0
ON19	5'-<u>t</u>G<u>c</u>AC<u>a</u>-GG<u>U</u>A<u>U</u>ATA<u>U</u>AGGC	_b	78.0	79.5	
ON20	3'-CCA<u>U</u>A<u>U</u>ATA<u>U</u>CCG-<u>g</u>CG<u>t</u>A<u>t</u>		[+6.0]	[+7.5]	nd

[a]ΔT_m is calculated relative to the unmodified **DNA1**:**DNA2** duplex ($T_m = 72.0$ °C), where **DNA1** = 5'-AAGCTGCACAGGTATATATAGGCCGCATATGCA and **DNA2** = 3'-TTCGACGTGTCCATATATATCCGGCGTATACGT-5′. Thermal denaturation curves (see Figs. 3.29 – 3.30) were recorded in medium salt phosphate buffer ([Na⁺] = 110 mM, [Cl⁻] = 100 mM, pH 7.0 (NaH₂PO₄/Na₂HPO₄), [EDTA] = 0.2 mM), with each ON present at 0.5 µM concentration. For the definition of *TA*, see main text. For a discussion of *TA* values, see the ESI†. A = adenin-9-yl DNA monomer, C = cytosin-1-yl DNA monomer, G = guanin-9-yl DNA monomer, T = thymin-1-yl DNA monomer, <u>U</u> = uracil-1yl 2'-*O*-(pyren-1-yl)methyl-RNA monomer, <u>C</u> = cytosin-1-yl 2'-*O*-(pyren-1-yl)methyl-RNA monomer. LNA monomers are

designated by lower case letters in bold and underline (**c** = 5-methyl-cytosin-1-yl LNA monomer). For structures of modifications, see Fig. 3.1.

[b] The thermal denaturation curve of **ON19**:**ON20** displays two transitions at T_m = 34 °C and 55 °C (Fig. 3.30), which is consistent with the T_m values observed for **ON19** and **ON20** in absence of single-stranded DNA target (Fig. 3.34), suggesting that **ON19** and **ON20** adopt secondary structures that are more stable than the **ON19**:**ON20** duplex. nd = not determined.

Thermodynamic driving force for dsDNA-recognition. The available free energy for recognition of **DNA1**:**DNA2** at 310 K can be estimated as $\Delta G_{rec}^{310} = \Delta G^{310}$ (5'-probe:**DNA2**) + ΔG^{310} (3'-probe:**DNA1**) - ΔG^{310} (probe duplex) - ΔG^{310} (**DNA1**:**DNA2**), with large negative values signifying a strongly activated probe (Table 3.2).[16] The calculations assume that two separate probe-target duplexes, rather than a four-stranded recognition complex, are formed (i.e., 5'-probe:**DNA2** + 3'-probe:**DNA1** vis-à-vis 5'-probe:**DNA2**:**DNA1**:3'-probe). Experimental evidence in support of the former is presented later (*vide infra*). Thermodynamic parameters associated with duplex formation were estimated through baseline fitting of denaturation curves assuming bimolecular reactions, two-state melting behavior, and constant heat capacity.[32]

While conventional 13-mer Invader probe **ON1**:**ON2** is quite labile ($\Delta\Delta G^{310}$ = 44 kJ/mol), the corresponding duplexes between the individual probe strands and the 33-mer target strands are considerably more stable but still far less stable than **DNA1**:**DNA2** ($\Delta\Delta G^{310}$ = 22-35 kJ/mol), rendering recognition of **DNA1**:**DNA2** thermodynamically unfavorable (ΔG_{rec}^{310} = +13 kJ/mol). This is unsurprising given the short probe length vis-à-vis **DNA1**:**DNA2**. For similar reasons, probes **ON3**:**ON4** and **ON5**:**ON6**, which feature unmodified and modified 3-mer overhangs, respectively, also do not display favorable thermodynamics for recognition of the 33-mer dsDNA target (ΔG_{rec}^{310} = +23 and +10 kJ/mol, respectively).

ON7:**ON8**, which features unmodified 6-mer overhangs, displays moderately favorable thermodynamics for recognition of **DNA1**:**DNA2** (ΔG_{rec}^{310} = -20 kJ/mol), which is large part is due to the stability of the **ON8**:**DNA1** duplex ($\Delta\Delta G^{310}$ = -1 kJ/mol). Remarkably, the doubly modified 19-mer **ON8** displays similar affinity towards **DNA1** as the unmodified 33-mer strand **DNA2**.

ON9:**ON10**, which features modified 6-mer overhangs, displays very favorable thermodynamics for recognition of **DNA1**:**DNA2** (ΔG_{rec}^{310} = -87 kJ/mol). This is due to a combination of a strongly destabilized probe duplex ($\Delta\Delta G^{310}$ = +46 kJ/mol) and two highly stabilized probe-target duplexes ($\Delta\Delta G^{310}$ = -25 and -16 kJ/mol). The corresponding control probe **ON11**:**ON12** that lacks the central energetic hotspots, displays far less favorable thermodynamics (ΔG_{rec}^{310} = -25 kJ/mol), underscoring the importance of the energetic hotspots. Interestingly, the unmodified control probe **ON13**:**ON14** with 6-mer overhangs, is weakly activated for recognition of **DNA1**:**DNA2** (ΔG_{rec}^{310} = -10 kJ/mol). For a discussion of this observation, see the supplementary data.

Blunt-ended 19-mer Invader probes **ON15**:**ON10** and **ON9**:**ON16** display very favorable thermodynamics for dsDNA recognition (ΔG_{rec}^{310} = -76 to -70 kJ/mol) but are also quite stable (ΔG^{310} = -70 to -65 kJ/mol), which may impact binding kinetics.

The favorable thermodynamics for recognition of **DNA1**:**DNA2** displayed by double-stranded probes with 6-mer overhangs (**ON7**:**ON8**-**ON13**:**ON14**) and the 19-mer blunt-ended probes (**ON15**:**ON10**, and **ON9**:**ON16**) are the result of exceptionally favorable enthalpy changes (ΔH_{rec} between -654 kJ/mol and -227 kJ/mol, Table 3.3), that are only partially offset by unfavorable entropy changes (-$T\Delta S_{rec}^{310}$ between 216 kJ/mol and 567 kJ/mol Table 3.5). This, in turn, is a manifestation of enthalpically destabilized probe duplexes ($\Delta\Delta H$ as high as +321 kJ/mol, Table 3.4) and enthalpically stabilized duplexes between individual probe strands and 33-mer single-stranded DNA targets ($\Delta\Delta H$ as low as -228 kJ/mol, Table 3.6). These observations are in line with our expectations of probe duplexes being destabilized[17,18] due to the presence of multiple +1 interstrand zipper arrangements of 2'-*O*-(pyren-1-yl)methyl-RNA monomers resulting in violation of the neighbor exclusion principle,[19,33] and probe-target duplexes being stabilized by stacking interactions between intercalating pyrene moieties and neighboring base-pairs.[18]

The driving force for recognition of the 33-mer dsDNA model target by double-stranded probes can also be assessed by the term *thermal advantage*, which we define as $TA = T_m$ (5'-probe:**DNA2**) + T_m (3'-probe:**DNA1**) - T_m (probe duplex) - T_m (**DNA1**:**DNA2**), with large positive values indicative of a probe that is activated for dsDNA-recognition (Table 3.1).[16] The

TA-based conclusions corroborate the ΔG^{310}_{rec}-based conclusions (see discussion in supplementary data).

Table 3.2. Changes in Gibbs free energy at 310 K (ΔG^{310}) upon formation of probe duplexes and duplexes between individual probe strands and **DNA1** or **DNA2**. Also shown is the calculated change in free energy upon probe-mediated recognition of 33-mer target **DNA1:DNA2** (ΔG^{310}_{rec}).[a]

Probe	Sequence	ΔG^{310} [$\Delta\Delta G^{310}$] (kJ/mol)			ΔG^{310}_{rec} (kJ/mol)
		probe duplex	5'-probe vs DNA2	3'-probe vs DNA1	
ON1	5'-GGTA**U**A**U**ATAGGC	-50 [+44]	-72 [+22]	-59 [+35]	+13
ON2	3'-CCATA**U**A**U**ATCCG				
ON3	5'-ACA-GGTA**U**A**U**ATAGGC	-44 [+50]	-54 [+40]	-61 [+33]	+23
ON4	3'-CCATA**U**A**U**ATCCG-GCG				
ON5	5'-A**C**A-GGTA**U**A**U**ATAGGC	-46 [+48]	-66 [+28]	-64 [+30]	+10
ON6	3'-CCATA**U**A**U**ATCCG-G**C**G				
ON7	5'-TGCACA-GGTA**U**A**U**ATAGGC	-48 [+46]	-67 [+27]	-95 [-1]	-20
ON8	3'-CCATA**U**A**U**ATCCG-GCGTAT				
ON9	5'-**U**GCA**C**A-GGTA**U**A**U**ATAGGC	-48 [+46]	-119 [-25]	-110 [-16]	-87
ON10	3'-CCATA**U**A**U**ATCCG-G**C**GTA**U**				
ON11	5'-**U**GCA**C**A-GGTATATATAGGC	-52 [+42]	-80 [+14]	-91 [+3]	-25
ON12	3'-CCATATATATCCG-G**C**GTA**U**				
ON13	5'-TGCACA-GGTATATATAGGC	-46 [+48]	-75 [+19]	-75 [+19]	-10
ON14	3'-CCATATATATCCG-GCGTAT				
ON15	5'-GGTA**U**A**U**ATAGGC-**C**GCA**U**A	-65 [+29]	-119 [-25]	-110 [-16]	-70
ON10	3'-CCATA**U**A**U**ATCCG-G**C**GTA**U**				
ON9	5'-**U**GCA**C**A-GGTA**U**A**U**ATAGGC	-70 [+24]	-119 [-25]	-121 [-27]	-76
ON16	3'- A**C**GTG**U**-CCATA**U**A**U**ATCCG				
ON17	5'-**t**GCA**c**A-GGTA**U**A**U**ATAGGC	-55 [+39]	-103 [-9]	-108 [-14]	-62
ON18	3'-CCATA**U**A**U**ATCCG-G**c**GTA**t**				
ON19	5'-**t**G**c**A**Ca**-GG**U**A**U**ATA**U**AGGC	-[b]	-121 [-27]	-104 [-10]	nd
ON20	3'-CCA**U**A**U**ATA**U**CCG-**g**CG**t**A**t**				

[a] $\Delta\Delta G^{310}$ is calculated relative to the unmodified 33-mer target DNA duplex **DNA1:DNA2** (ΔG = -94 kJ/mol). For a definition of ΔG^{310}_{rec}, see the main text. For experimental conditions, see Table 3.1. nd = not determined. [b] See footnote b of Table 3.1.

Characterization of dsDNA-recognition by toehold Invader probes. The dsDNA-targeting properties of the developed probes were characterized using an electrophoretic mobility shift assay in which a doubly 3'-digoxigenin (DIG)-labeled version of **DNA1:DNA2** was used as a model dsDNA target (Fig. 3.2a). At the onset of the study, two outcomes were considered likely, i.e., recognition of **DNA1:DNA2** resulting in the formation of a) a four-stranded double-duplex invasion complex with lower mobility on non-denaturing polyacrylamide gels vis-à-vis **DNA1:DNA2**, or b) two separate duplexes between individual probe strands and the corresponding 33-mer DNA strands harboring the complementary region (i.e., two probe-target duplexes, PTDs) with greater mobility than **DNA1:DNA2** (Fig. 3.2a). In fact, the latter was observed when **DNA1:DNA2** was incubated with a 5-fold molar excess of certain probes at 37 °C as evidenced by the formation of a higher mobility band (Fig. 3.2b).

As expected from the unfavorable thermodynamics ($\Delta G_{\text{rec}}^{310} > 10$ kJ/mol, Table 3.2), blunt-ended 13-mer Invader probe **ON1:ON2** and probes **ON3:ON4** and **ON5:ON6** with unmodified or modified 3-mer overhangs, respectively, do not result in recognition of **DNA1:DNA2** (Figs. 3.2b and 3.2c, and Table 3.6). In contrast, **ON9:ON10**, which has modified 6-mer overhangs, results in ~45% recognition (Figs. 3.2b and 3.2c, and Table 3.6), which is consistent with its favorable $\Delta G_{\text{rec}}^{310}$ value (Table 3.2). **ON7:ON8**, which has unmodified 6-mer overhangs, and **ON11:ON12**, which has modified 6-mer overhangs but lacks the energetic hotspots, result in lower levels of recognition (~30% and ~39%, respectively, Figs. 3.2b and 3.2c, and Table 3.6), consistent with their less favorable $\Delta G_{\text{rec}}^{310}$ values. Unmodified 13-mer control probe **ON13:ON14** with 6-mer overhangs results in trace recognition of **DNA1:DNA2** (Figs. 3.2b and 3.2c, and Table 3.6). For supplemental discussion of the dsDNA-targeting properties of **ON7:ON8-ON13:ON14**, see Figs. 3.35 and 3.36 and the supplementary data.

Interestingly, blunt-ended 19-mer Invader probes **ON9:ON16** and **ON15:ON10** do not result in recognition of **DNA1:DNA2** despite displaying strongly favorable thermodynamics for dsDNA-recognition ($\Delta G_{\text{rec}}^{310} = -76$ to -70 kJ/mol, Table 3.2). To establish if the lack of recognition is due the high stability of these probe duplexes, additional experiments were performed in which **DNA1:DNA2** was mixed with a 5-fold molar probe excess under heat-shock conditions, which were expected to reduce kinetic barriers to recognition. Prominent formation of the corresponding probe-target duplexes was observed (~75 %), indicating that

recognition kinetics, indeed, are limited by the stability of the probes at 37 °C (Fig. 3.37 and Table 3.7). Incubation of **ON7**:**ON8** or **ON9**:**ON10** with **DNA1**:**DNA2** under heat-shock conditions also results in more prominent formation of probe-target duplexes (~38 % and ~67 %, respectively, Fig. 3.37 and Table 3.7).
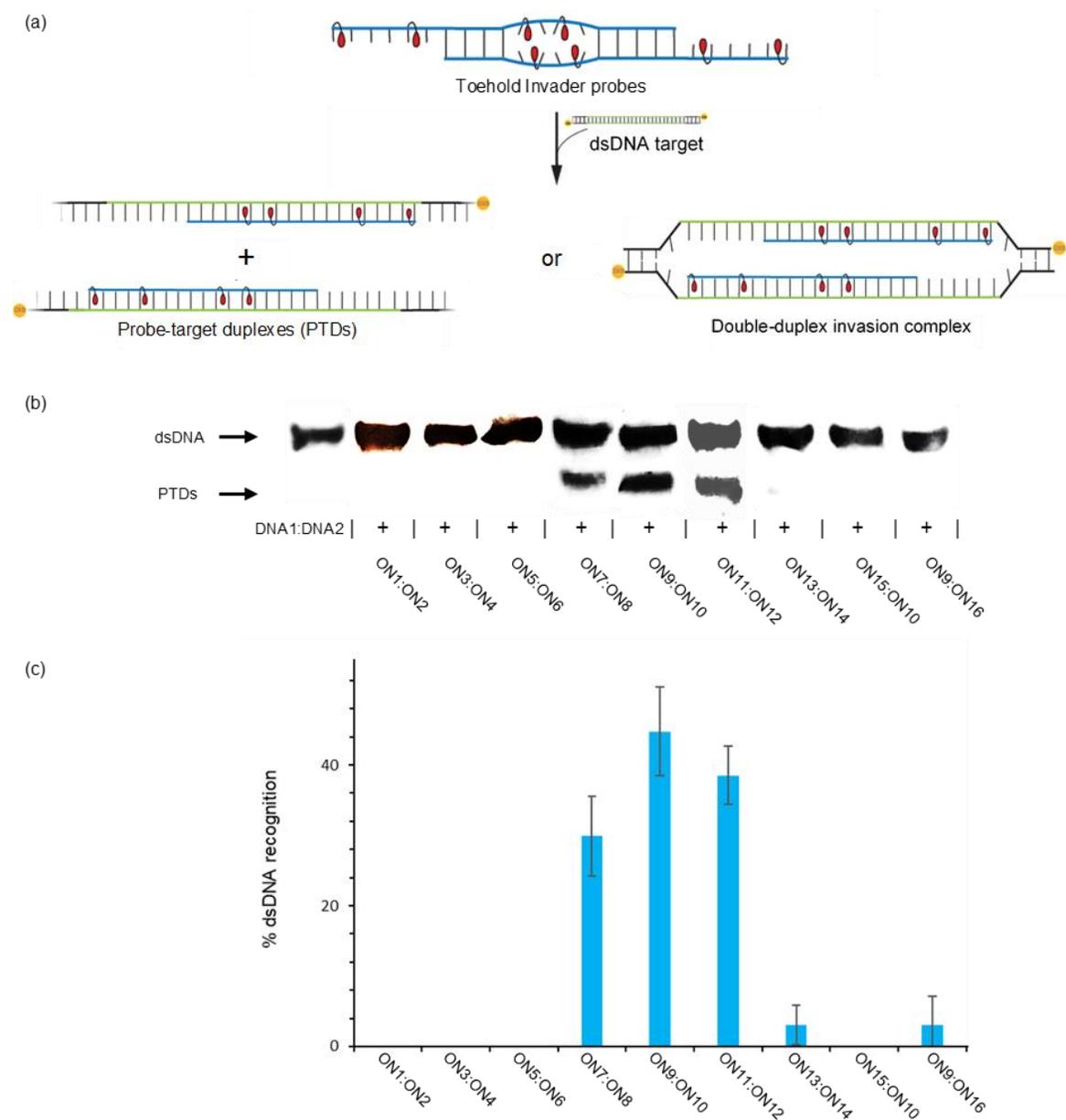
Figure 3.2. (a) Illustration of assay used to evaluate Invader-mediated recognition of dsDNA targets along with possible outcomes of the recognition process. (b) Representative gel electrophoretograms from recognition experiments in which **DNA1**:**DNA2** was incubated with

a 5-fold molar excess of different probes. PTDs = probe-target duplexes. (c) Histogram depicting averaged results from at least three independent recognition experiments with error bars representing standard deviation. Conditions: pre-annealed doubly 3'DIG-labeled **DNA1**:**DNA2** (50 nM) was incubated with a 5-fold molar excess of the specified pre-annealed probe 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 17 h. Mixtures were resolved on 16% nd-PAGE gels.

Another series of control experiments were carried out in which a 5-fold molar excess of individual probe strands **ON7**-**ON16** was incubated with **DNA1**:**DNA2** at 37 °C. Recognition was not observed with any of the single-stranded probes, underscoring that both probe strands are required to facilitate recognition of **DNA1**:**DNA2** (Fig. 3.38). Collectively, these initial observations indicate that Invader probes with modified single-stranded overhangs offer advantages for recognition of mixed-sequence dsDNA target regions vis-à-vis blunt-ended Invader probes.

Next, dose-response profiles for toehold Invader probes **ON7**:**ON8** and **ON9**:**ON10** were determined (Figs. 3.3 and 3.39). C_{30} values, i.e., the probe concentration resulting in 30% recognition of **DNA1**:**DNA2**, of ~30 nM and ~50 nM, were observed for **ON9**:**ON10** and **ON7**:**ON8**, respectively. Maximal recognition (~45%) was realized when probes were used at ~10-fold or greater excess. Recognition of **DNA1**:**DNA2** is fast (plateau reached within 30 min, Fig. 3.40).
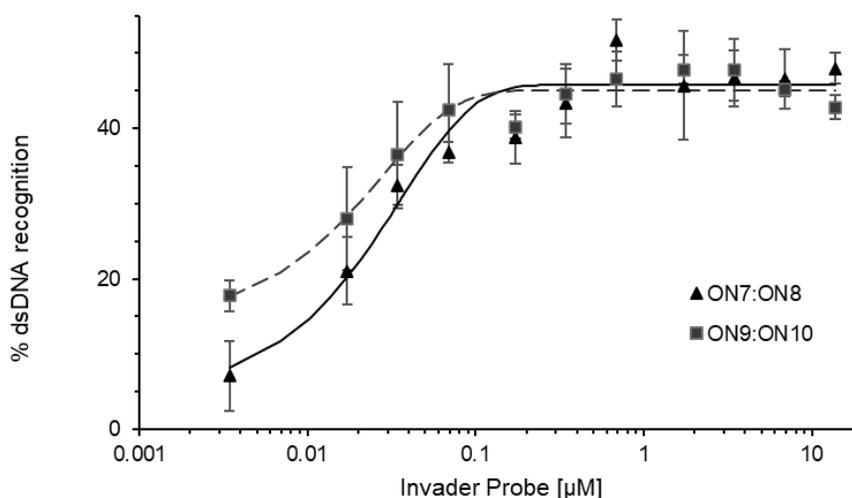
Figure 3.3. Dose-response curves for recognition of **DNA1**:**DNA2** by Invader probes **ON7**:**ON8** and **ON9**:**ON10** at 37 °C. Representative electrophoretograms are shown in Fig. 3.39. Experimental conditions are as described in Fig. 3.2.

Subsequently, the binding specificity of **ON7**:**ON8** and **ON9**:**ON10** was evaluated by incubating a 5-fold molar probe excess with:

i)   **DNA3**:**DNA4** (**MM1**), which differs in sequence at one position relative to **DNA1**:**DNA2** (corresponding to the position of the "left-most" energetic hotpot of the probes as drawn; recognition would result in the formation of two probe-target duplexes, each with one mismatched base-pair, Fig. 4a),

ii)   **DNA5**:**DNA6** (**MM2**), which differs in sequence at two positions relative to **DNA1**:**DNA2** (corresponding to the central position in each of the single-stranded overhangs of the probes; recognition would result in the formation of two probe-target duplexes, each with one mismatched base-pair, Fig. 4a) and,

iii)   **DNA7**:**DNA8** (**MM3**), which differs in sequence at three positions relative to **DNA1**:**DNA2** (one corresponding to the position of the "left-most" energetic hotpot of the probes as drawn and two corresponding to the central position in each of the single-stranded overhangs of the probes; recognition would result in the formation of two probe-target duplexes, each with two mismatched base-pairs, Fig. 4a).

Interestingly, **ON7**:**ON8**, which features unmodified 6-mer single-stranded overhangs, displays superior discrimination of all three mismatched targets compared to **ON9**:**ON10**, i.e., the higher-affinity probe with modified 6-mer single-stranded overhangs, which only results in efficient discrimination of **MM3** (Figs. 3.4 and 3.41).
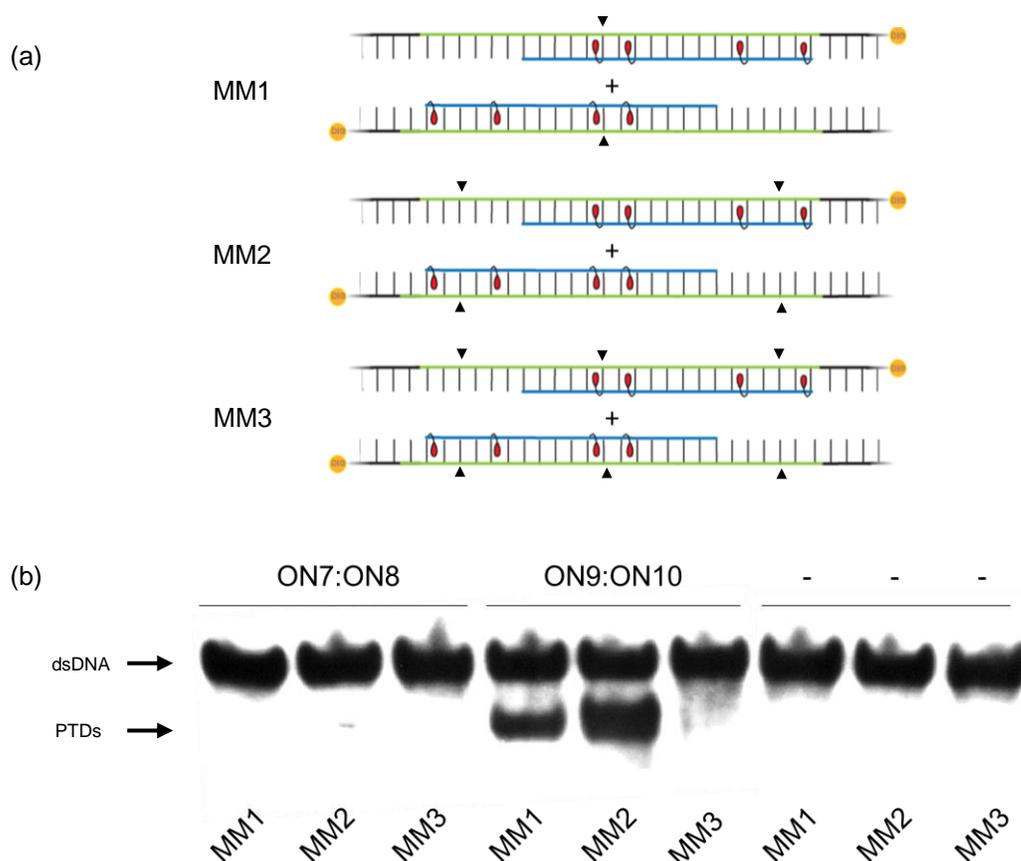
Figure 3.4. Binding specificity of toehold Invader probes. (a) Illustration of the mismatched probe-target duplexes that would ensue upon recognition of **MM1**-**MM3**, with black arrows indicating the position of mismatched base-pairs. (b) Representative electrophoretograms from experiments in which a 5-fold excess of **ON7**:**ON8** or **ON9**:**ON10** was incubated with non-complementary **MM1**-**MM3** targets (see Table 3.8 for sequences) as described in Figure 3.2. PTDs = probe-target duplexes.

To rationalize the observed trends, we determined T_m values for **MM1**-**MM3**, as well as the duplexes between individual probe strands and the corresponding mismatched DNA strands,

allowing for calculation of *TA* values for recognition of **MM1**-**MM3** by **ON7**:**ON8** or **ON9**:**ON10** (Figs. 3.31 and Tables 3.8 and 3.9). As expected, the singly mismatched probe-DNA duplexes display slightly reduced T_ms vis-à-vis matched probe-DNA duplexes (e.g., T_m = 57.5 °C vs 60.0 °C for **ON7**:**DNA4** and **ON7**:**DNA2**, respectively), whereas doubly mismatched probe-DNA duplexes display substantially lower T_ms (e.g., T_m = 49.0 °C for **ON7**:**DNA8**). Consequently, probes are most strongly activated for recognition of matched dsDNA targets, while being least strongly activated for recognition of triply mismatched targets (e.g., *TA* values for **ON7**:**ON8** = 11.0 °C, 4.0 °C, 8.0 °C, and -5.0 °C for recognition of **DNA1**:**DNA2**, **MM1**, **MM2**, and **MM3**, respectively, Table 3.8). In agreement with the observed dsDNA-recognition trends (Figs. 3.4 and 3.41), **ON9**:**ON10** displays more prominent driving forces for recognition of **MM1**-**MM3** than **ON7**:**ON8** (*TA* values for **ON9**:**ON10** = 16.5 °C, 19.0 °C, and 9.5 °C, respectively, Table 3.9). Presumably, the less effective mismatch discrimination displayed by **ON9**:**ON10** is due to the presence of 2'-*O*-(pyren-1-yl)methyl-RNA monomers in the overhangs, which increase target affinity but reduce binding specificity.[30,31]

Invader probes with LNA-modified toeholds. As a logical next step, we set out to replace the 2'-*O*-(pyren-1-yl)methyl-RNA modifications in the single-stranded overhangs of **ON9**:**ON10** with a different affinity-enhancing modification, while maintaining the +1 interstrand zipper arrangements of 2'-*O*-(pyren-1-yl)methyl-RNA monomers in the double-stranded region to preserve the design feature that facilitates probe dissociation. Locked nucleic acids (LNAs) are conformationally restricted nucleotide monomers that are commercially available and well-known to increase the binding affinity and specificity of ONs.[36,37] In addition to synthesizing **ON17**:**ON18** as a direct analogue of **ON9**:**ON10**, we, in an attempt to develop an optimized probe that is even more strongly activated for specific dsDNA recognition, also synthesized **ON19**:**ON20**, which features three +1 interstrand zipper arrangements of 2'-*O*-(pyren-1-yl)methyl-RNA monomers in the double-stranded region and three LNA modifications in each of the single-stranded overhangs (Table 3.1).

Substitution of the 2'-*O*-(pyren-1-yl)methyl-RNA monomers for LNA monomers in the single-stranded overhangs results in a minimally more destabilized probe duplex (T_m = 47.5 °C vs T_m = 49.0 °C for **ON17**:**ON18** and **ON9**:**ON10**, respectively, Table 3.1) and probe-target duplexes (T_m = 71.5-73.0 °C vs 75.5 °C, Table 3.1). The more extensively modified **ON19**:**ON20** probe

does not form a stable duplex; instead, individual probe strands seemingly form intramolecular hairpin structures (Figs. 3.30 and 3.34). The corresponding probe-target duplexes denature at slightly higher temperatures than the equivalent duplexes involving **ON9** and **ON10** (i.e., T_m = 78.0-79.5 °C vs 75.5 °C).

ON17:**ON18** displays favorable thermodynamics for recognition of **DNA1**:**DNA2** (ΔG_{rec}^{310} = -62 kJ/mol) since the probe is destabilized ($\Delta\Delta G^{310}$ = +39 kJ/mol) and the two probe-target duplexes are stabilized ($\Delta\Delta G^{310}$ = -14 to -9 kJ/mol (Table 3.2). However, the driving force is less favorable than for **ON9**:**ON10**. Thermodynamic parameters were not determined for **ON19**:**ON20** due to the aforementioned formation of intramolecular hairpin structures. For additional discussion of the thermodynamic parameters, see Tables 3.10 and 3.11 and the supplementary data.

Dose-response profiles for **ON17**:**ON18** and **ON19**:**ON20** against the doubly 3'-DIG-labeled version of **DNA1**:**DNA2** were determined and compared to **ON9**:**ON10** (Figs. 3.42-3.44). The three probes display similar C_{30} values (~30 nM) but recognition reaches a plateau at ~55% with **ON17**:**ON18** and ~45% with **ON19**:**ON20** (Figs. 3.44) and is rapid (< 30 min, Fig. 3.45). Individual probe strands **ON17**-**ON20** does not result in recognition of **DNA1**:**DNA2** (Fig. 3.46), suggesting that both strands of the LNA-modified toehold Invader probes are necessary to render dsDNA-recognition thermodynamically feasible.

The binding specificities of **ON17**:**ON18** and **ON19**:**ON20** were evaluated as in the initial experiments using mismatched dsDNA targets **MM1**-**MM3**. Both LNA-modified toehold Invader probes - and **ON17**:**ON18** in particular - display improved discrimination of **MM2** vis-à-vis **ON9**:**ON10**, whereas the moderate discrimination of **MM1** and excellent discrimination of **MM3** remain unchanged (Figs. 3.5 and 3.41).
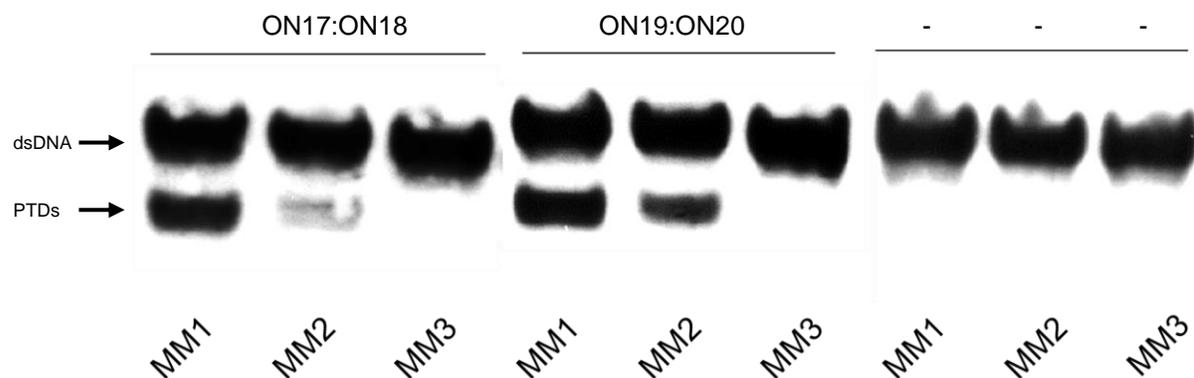
Figure 3.5. Binding specificity of toehold Invader probes. Representative electrophoretograms from experiments in which a 5-fold molar excess of **ON17**:**ON18** or **ON19**:**ON20** were incubated with non-complementary targets **MM1-MM3** at 37 °C for 17 h. For sequences of **MM1**-**MM3**, see Table 3.12. Pre-annealed 3'-DIG-labeled **MM1-MM3** (50 nM) was incubated with pre-annealed Invader probe at 37 °C for 17 h in HEPES buffer as outlined in Fig. 3.2. PTD = probe-target duplex.

These observations are in line with expectations as recognition of **MM2** would result in the formation of two probe-target duplexes with mismatched base-pairs in the vicinity of the LNA monomers (Fig. 3.4a), which are known to improve mismatch discrimination,[38] whereas recognition of **MM1** would result in two probe-target duplexes with mismatched base-pairs in the vicinity of the 2'-*O*-(pyren-1-yl)methyl-RNA monomers, and unaffected by the LNA monomers.

Besides, the binding specificity is the result of multiple effects, including stringency clamping effects, i.e., greater stability differences between matched vis-à-vis mismatched recognition complexes seen with structured probes,[34] avoidance of energetically unfavorable formation of multiple double-stranded segments with mismatched base-pairs.[35] The improved discrimination of **MM2** displayed by **ON17**:**ON18** vis-à-vis **ON9**:**ON10** is also reflected in a slightly lower *TA* value (*TA* = 17.0 °C vs 19.0 °C, respectively, Tables 3.9 and 3.12). Thus, the results suggest that it is possible to modulate the binding affinity and improve the specificity of dsDNA-recognition by incorporating LNA monomers in the overhangs of toehold Invader probes.

Next, we evaluated if toehold Invader probes can recognize target regions embedded within DNA hairpin structures. Towards this end, toehold Invader probes **ON9**:**ON10**, **ON17**:**ON18**, and **ON19**:**ON20** were incubated with a 3'-DIG-labeled DNA hairpin (**DH1**) analogue of **DNA1**:**DNA2**, i.e., **DH1**, which consists of a 33-mer double-stranded stem of the same sequence as **DNA1**:**DNA2** and a T_{10} loop that links one end of the hairpin. The unimolecular nature of **DH1** renders it as a higher-melting target compared to **DNA1**:**DNA2** ($T_m = 81$ °C vs 72 °C, Table 3.14 and Table 3.1, respectively). Remarkably, each of the three probes recognizes **DH1**, forming a ternary recognition complex as evidenced by the emergence of a slower moving band (Figs. 3.6 and 3.48). The extensively modified toehold Invader probe **ON19**:**ON20** result in the most efficient recognition of **DH1** with C_{25} values of ~2 μM as compared to ~6 μM for **ON17**:**ON18** and >25 μM for **ON9**:**ON10**.

(a)

dsDNA hairpin target

T_{10}

Recognition complex (RC)

(b)

RC →

Target →

DH1 | + | ON9:ON10 | ON17:ON18 | ON19: ON20

(c)

- ■ ON9:ON10
- ▲ ON17:ON18
- ● ON19:ON20

% dsDNA recognition
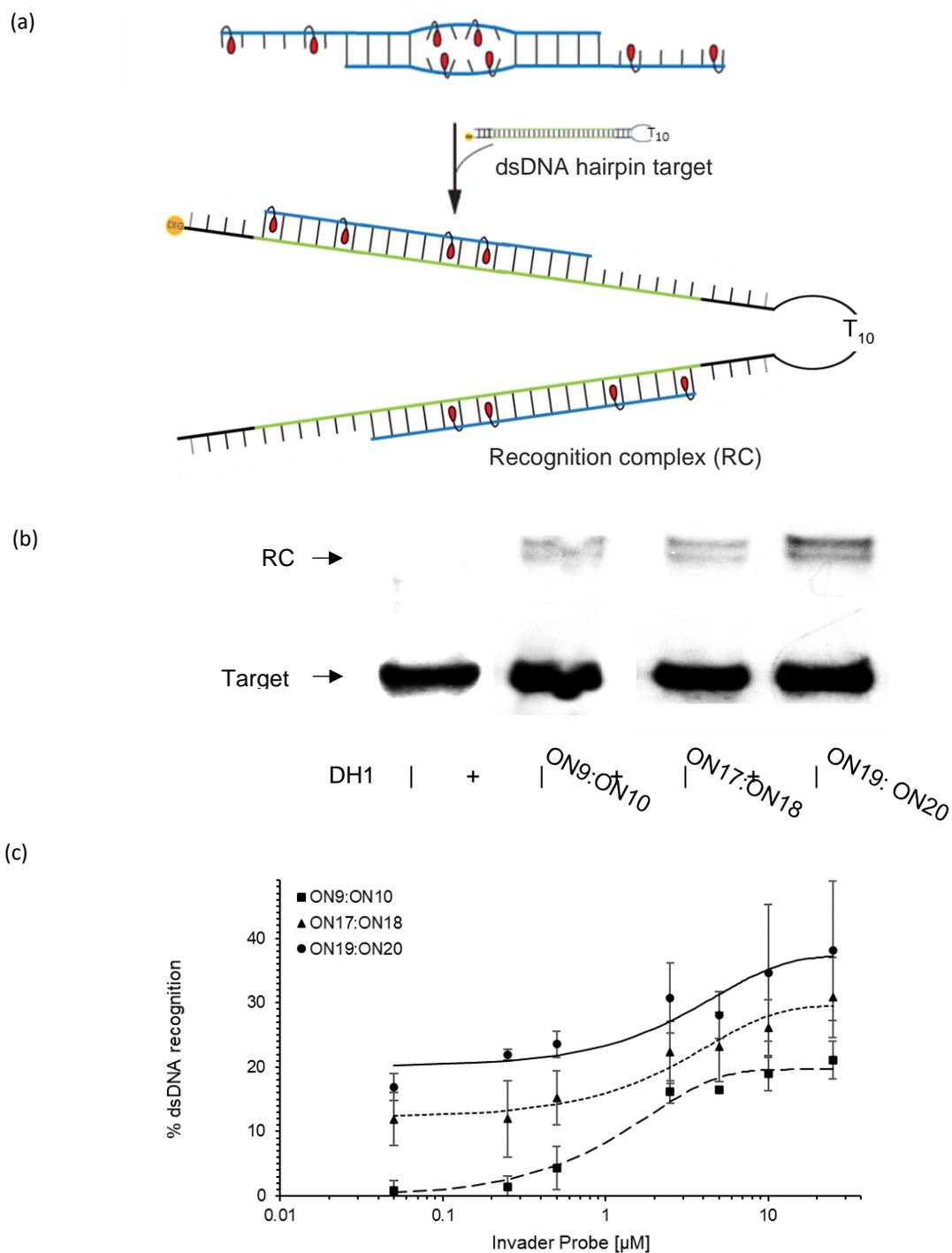
Invader Probe [µM]

Figure 3.6. (a) Assay used to evaluate Invader-mediated recognition of DNA hairpin targets. (b) Representative gel electrophoretogram from recognition experiments in which a 50-fold molar excess of toehold Invader probes was incubated with DNA hairpin **DH1** (50 nM). (c) Dose-response curves for recognition of **DH1** using toehold Invader probes **ON9:ON10**,

ON17:**ON18**, or **ON19**:**ON20**. Error bars represent standard deviation from at least three experiments. Experimental conditions are as stated in Fig. 2 except that the mixtures were resolved on 12% nd-PAGE gels.

Detection of chromosomal DNA in the context of FISH assays. Motivated by the above findings, we set out to demonstrate recognition of mixed-sequence chromosomal DNA targets using LNA-modified toehold Invader probes. We have previously reported[16] the use of blunt-ended Invader probes for recognition of a complementary, highly repeated region in the *DYZ-1* satellite gene (~6 x 10^4 tandem repeats of a ~1175 bp region) of the bovine (*Bos taurus*) Y chromosome (NCBI code: M26067)[39] under non-denaturing conditions. One probe, however, i.e., the 15-mer blunt-ended Invader probe **DYZ-REF** featuring three energetic hotspots (Fig. 7), has proven refractory to recognition of this region, presumably because the thermodynamic driving force is too small (unpublished results).[43] To overcome this shortcoming, we synthesized the analogous **DYZ-OPT** which features two doubly LNA-modified 6-mer overhangs (Fig. 7 and Table 3.15).

The denaturation curve of **DYZ-OPT** probe does not display a clear sigmoidal transition, indicating that the probe duplex either is unstable or highly distorted (Fig. 3.49). Conversely, duplexes between the individual probe strands of **DYZ-OPT** and 35-mer DNA strands within which the complementary target region is embedded, are as stable as the corresponding 35-mer reference duplex **DNA9**:**DNA10**[40] (ΔT_m = 0 to 3 °C, Table 3.15 and Fig. 3.49). The greater stability of the probe-target duplexes vis-à-vis the probe duplex indicates feasibility of dsDNA-recognition. Indeed, ~45% recognition is observed when a 5-fold molar excess of **DYZ-OPT** is incubated with **DNA9**:**DNA10** as evidenced by the presence of faster-moving probe-target duplex bands, whereas no recognition is observed for **DYZ-REF** (Fig. 3.7). The stability of the probe-target duplexes (T_m ~ 61-68 °C) formed by the blunt ended Invader probe **DYZ-REF** are comparable to stability of dsDNA target (T_m ~ 66 °C), while the probe duplex itself is too stable (T_m = 56 °C), suggesting the marginal thermodynamic potential for recognition of dsDNA target (TA = 7 °C), which is in agreement with the recognition assay.
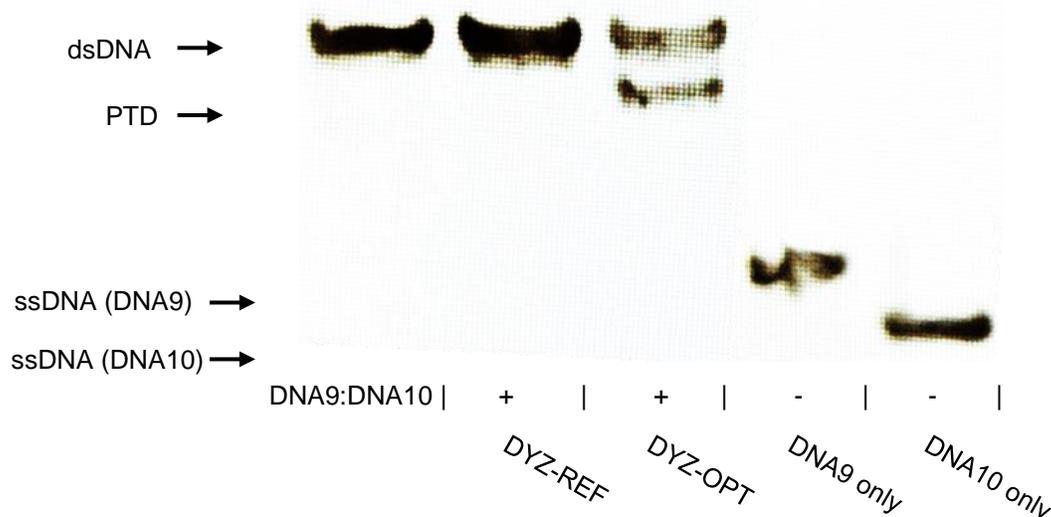
Figure 3.7. Representative gel electrophoretograms from recognition experiments in which pre-annealed 3'-DIG-labeled dsDNA target **DNA9**:**DNA10** (50 nM) was incubated with a 5-fold molar excess of **DYZ-REF** or **DYZ-OPT**. Condition**s** as outlined in Fig. 3.2. (**DYZ-REF** = 5'-T**UAU**ATGCTG**U**TCTC-3':3'-AA**UAU**ACGACA**A**GAG-5' and **DYZ-OPT** = 5'-Cy3-T**g**T**g**TG-T**UAU**ATGCTG**U**TCTC-3': 3'-AA**UAU**ACGACA**A**GAG-TC**g**G**g**A-Cy3-5').

The binding specificity of **DYZ-OPT** was evaluated using **DNA11**:**DNA12** (sequence shown in Table 3.16), i.e., a mismatched dsDNA target designed similarly to **MM3**. Thus, **DNA11**:**DNA12** differs in sequence at three positions vis-à-vis **DNA9**:**DNA10**. Successful recognition would require the formation of two probe-target duplexes, each with two mismatched base-pairs (Table 3.17). As expected, **DNA11**:**DNA12** is recognized less efficiently than **DNA9**:**DNA10** (Fig. 3.50).

Lastly, **DYZ-OPT** was evaluated for its ability to recognize the corresponding *DYZ-1* target region (NCBI code: M26067, positions: 861-887) in the context of fluorescence in situ hybridization (FISH) assays. Fixed interphase nuclei from a male bovine kidney cell line were incubated with **DYZ-OPT** under denaturing and non-denaturing FISH conditions. Unlike **DYZ-REF**, which does not generate a specific signal even at high probe concentrations, the LNA-modified toehold Invader probe **DYZ-OPT** recognizes the mixed-sequence chromosomal DNA target under non-denaturing FISH conditions as evidenced by the presence of localized punctate Cy3-signals in ~40 % of the fixed nuclei (Fig. 3.8). Punctate signals are also observed under denaturing conditions with **DYZ-OPT** (Fig. 3.52). Importantly, no signals were observed

when the Y-chromosome-targeting probe **DYZ-OPT** was incubated under non-denaturing conditions with fixed nuclei from a female bovine endothelial cell line, which lacks the *DYZ-1* target region (Fig. 3.53), suggesting that binding occurs with high specificity. Thus, these results demonstrate that LNA-modified toehold Invader probes can be utilized to bind to mixed-sequence chromosomal DNA targets that are refractory to recognition by conventional blunt-ended Invader probes.
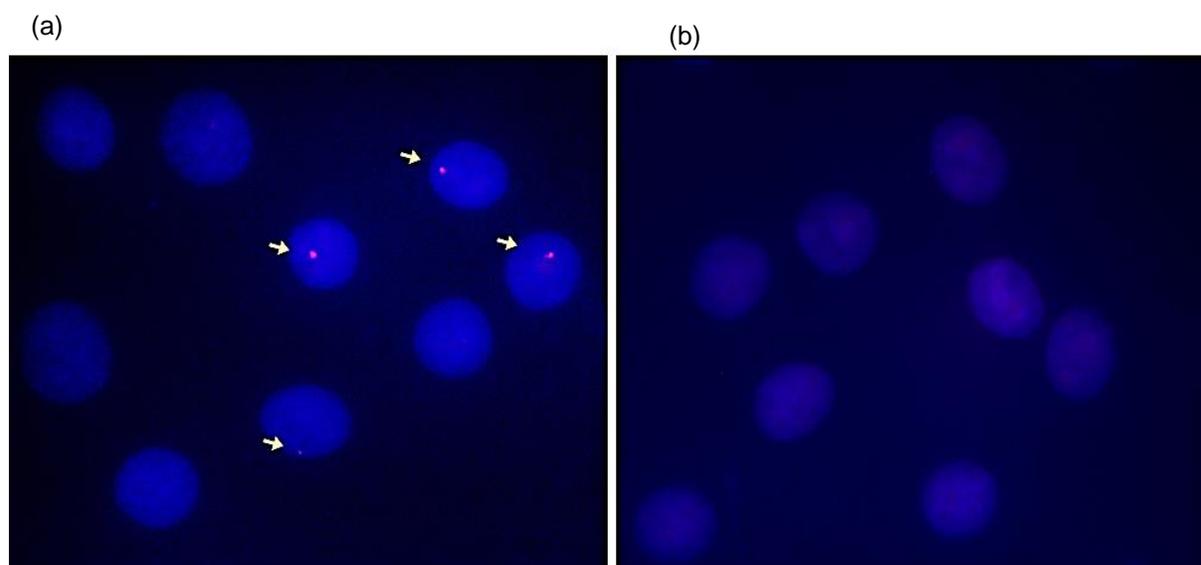
Figure 3.8. Images from FISH experiments using toehold Invader probes (a) **DYZ-OPT** (3 ng), and (b) **DYZ-REF** (15 ng) under non-denaturing conditions. Fixed isolated nuclei from male bovine kidney cells were incubated with probes for 3 h at 37.5 °C in a Tris buffer (20 mM Tris-Cl, 100 mM KCl, pH 8.0) and counterstained with DAPI. Images were obtained by overlaying images from Cy3 (red) and DAPI (blue) channels and adjusting the exposure. Nuclei were viewed at 60X magnification using a Nikon Eclipse Ti-S inverted microscope. 3 ng of Invader probes per 200µl PCR buffer (1X) was used based on initial optimization studies (Fig. 3.51).

3.3. Conclusion

Shorter toehold probes (3-nt overhangs) used herein displayed nominal or trace recognition of dsDNA target (**ON3:ON4/ON5:ON6** ~ <5% recognition) while probes with longer toeholds (6-nt overhangs) displayed efficient recognition of dsDNA target (**ON7:ON8/ON9:ON10**). Invader probes containing toeholds with affinity enhancing modification show greater

recognition of dsDNA (**ON9:ON10** ~ 45% recognition). This probe (**ON9:ON10**) completely discriminates non-complementary target (**MM3**) – recognition would result in two probe-target duplexes with two mismatched base-pairs in each but was unable to discriminate non-complementary targets (**MM1** and **MM2**) - recognition would result two probe-target duplexes with one mismatched base-pair in each. Fortunately, the LNA modified toehold Invader probe (**ON17:ON18**) demonstrated similar recognition as **ON9:ON10**, as well as better mismatch discrimination over **ON9:ON10**. Furthermore, toehold Invader with modified LNA overhangs (**ON17:ON18/ON19:ON20**) recognizes the hairpin target more efficiently than toehold Invader with invader-modified overhangs (**ON9:ON10**). Toehold Invaders with modified LNA overhangs (**DYZ-OPT**) effectively recognize a region of chromosomal DNA that blunt Invaders did not recognize before. While the mechanism governing recognition of complementary double-stranded DNA regions using toehold Invader probes is unknown, our results render a simple strand-exchange mechanism unlikely (e.g., all dsDNA targets used herein display high T_m values and are unlikely to be fully denatured at the experimental conditions. We speculate that the perturbation imparted by the energetic hotspots and/or the high cDNA-affinity of the single-stranded overhangs enables the two individual probe strands to gain access to complementary dsDNA regions that are partially denatured, be it due to base-pair breathing, high AT-content, and/or other factors, and unzip the helix via an energetically favorable double-duplex invasion process. Thus, toehold Invaders displayed enhanced recognition efficiency against complementary targets, and chromosomal DNA targeting further suggests that toehold Invaders have the possibility in therapeutic and diagnostic applications on biological targets.

3.4 Acknowledgments

3.5. Experimental section

Synthesis and purification of ONs. Modified ONs were synthesized on an automated DNA synthesizer (0.2 μmol scale) using a long chain alkyl amine-controlled pore glass (LCAA-CPG) solid support with a pore size of 500 Å. The corresponding phosphoramidites of monomer **U** and **C** (C[Bz]) were prepared as previously described[30,31] and incorporated into ONs via hand-

couplings (0.05 M in acetonitrile, ~50-fold molar excess) using 0.01 M 4,5-dicyanoimidazole as the activator (15 min) and 0.02 M iodine in THF/H_2O/pyridine for extended oxidation (45 s). Protected LNA phosphoramidites (A^{Bz}, $^{5-Me}C^{Bz}$, G^{dmf}, T) were obtained from a vendor (Exiqon, Vedbaek, Denmark) and incorporated into ONs via hand-coupling (3 min) using the abovementioned approach. 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). Subsequent treatment with 32% aq. ammonia (55 °C, 17 h) ensured deprotection and cleavage from the solid support of the DMTr-protected ONs, which were purified via ion-pair reverse phase HPLC (XTerra MS C18 column: 0.05 M triethylammonium acetate and acetonitrile gradient) followed by detritylation (80% aq. AcOH, 20 min) and precipitation (NaOAc, $NaClO_4$, acetone, -18 °C, 16 h). The purities of the synthesized ONs were verified using analytical HPLC (>85% purity, see Figs. 3.27 and 3.28), while identity was verified either by MALDI-MS (using 2,4,6-trihydroxy acetophenone as a matrix) or LC-ESI-MS analysis (Waters/Acquity C18 column; triethylammonium formate and acetonitrile gradient) recorded on a Quadrupole Time-of-Flight (Q-TOF) mass spectrometer (Table 3.3 and Figs. 3.9-3.26).

Thermal denaturation experiments. The concentrations of ONs were estimated using the following extinction coefficients (OD_{260}/µmol): G (12.01), A (15.20), T/U (8.40), C (7.05), Cy3 (4.93), and pyrene (22.4).[41] T_ms of duplexes (0.5 µM final concentration of each strand) were determined using a Cary 100 UV/Vis spectrophotometer equipped with a 12-cell Peltier temperature controller and determined 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 having 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. A temperature range from no more than 20 °C to at least 15 °C above the duplex T_m was used, with T_ms determined as the average of at least two experiments within ±1.0 °C. A temperature ramp of 1 °C/min was used in all experiments.

Determination of thermodynamic parameters. Thermodynamic parameters associated with duplex formation were estimated through baseline fitting of denaturation curves (van't Hoff method) using software provided with the UV-Vis spectrophotometer. Bimolecular reactions, two-state melting behavior, and constant heat capacity were assumed.[32] Two denaturation curves per duplex were analyzed at least three times to minimize errors arising from baseline choice.

Electrophoretic mobility shift assay. Unmodified DNA strands were obtained from commercial sources and used without further purification. Target strands were DIG-labeled using the 2nd generation DIG Gel Shift Kit (Roche Applied Bioscience). Briefly, 11-digoxigenin-ddUTP was incorporated at the 3′-end of the strand (100 pmol) using a recombinant DNA terminal transferase. The reaction mixture was quenched through the addition of EDTA (0.05 M), and then diluted to 100 nM in 2X HEPES buffer and used without further processing. The recognition experiments were conducted essentially as previously reported.[16] Thus, Invader probes (concentration as specified) were annealed (90 °C for 2-3 min, followed by cooling to room temperature) and subsequently incubated with separately pre-annealed DIG-labeled DNA (50 nM final concentration in 1X 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 ± 2 °C for 17 h.

Recognition experiments carried out under heat-shock conditions, the double-stranded probe and dsDNA target were briefly heated (3 min, 90 °C) and then cooled to 37 °C prior to incubation at 37 °C for 17 h.

In time-course experiments, aliquots were taken at the specified time-points, flash-frozen in liquid N$_2$, and stored in -20 °C until electrophoresis was performed.

Following incubation, loading dye (6X) was added and the mixtures were then loaded onto 12 % (DNA hairpin targets) or 16 % (linear dsDNA targets) non-denaturing TBE-PAGE gels (45 mM tris-borate, 1 mM EDTA; acrylamide:bisacrylamide (19:1)). Mixtures were resolved via electrophoresis, which was performed using constant voltage (70 V) at ~4 °C. Bands were blotted onto positively charged nylon membranes (100 V, 30 min, ~4 °C) and cross-linked through exposure to UV light (254 nm, 5 × 15W bulbs, 3 min). The membranes were incubated with anti-digoxigenin alkaline phosphatase F$_{ab}$ fragments as recommended by the manufacturer

and transferred to a hybridization jacket. Membranes were incubated with the chemiluminescence substrate (CSPD) for 10 min at 37 °C, and chemiluminescence was captured on X-ray films. Digital images of developed X-ray films were obtained using a BioRad ChemiDoc™ MP Imaging and used for densitometric quantification of the bands. The percentage of dsDNA recognition was calculated as the intensity ratio between the recognition band and the total lane. An average of at least three independent experiments is reported along with standard deviations (±). Electrophoretograms shown may be composite images 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,[42] was used to fit the following equation to the data points: $y = C + A (1 - e^{-kt})$ where C, A and k are constants. The resulting equation was used to calculate C_{30} values by setting $y = 30$ and solving for t.

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 suspension of nuclei. 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.

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 which allowed the suspension 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%.

Fluorescence in situ hybridization experiments. An aliquot of labeling buffer (~200 μL of a solution containing 1.5-30 ng of Cy3-labeled probes 1x PCR buffer (10 mM Tris, 50 mM KCl, pH 8.0) was placed on each slide. Preliminary optimization studies revealed that a probe quantity of ~3 ng per 200 μL of incubation buffer resulted in the most suitable signal-to-background ratio for toehold Invader probes under non-denaturing conditions (Fig. 3.51) whereas ~1.5 ng per 200 μL of incubation buffer was selected for FISH experiments under denaturing conditions (Fig. 3.52). 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 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 NISElements BR 4.20 software. The percentage of nuclei presenting representative signals (i.e., signal coverage) was estimated by evaluating >50 nuclei per Invader probe.

3.5. Supplementary data

Table 3.3. MS data of ONs used in this study.[a]

ON	Sequence	Calculated m/z (M+H)$^+$	Observed m/z (M+H)$^+$
1[b]	5'-GGTA<u>U</u>A<u>U</u>ATAGGC	4446.0	4447.5
2[b]	3'-CCATA<u>U</u>A<u>U</u>ATCCG	4326.0	4327.0
3	5'-ACA-GGTA<u>U</u>A<u>U</u>ATAGGC	5362.0	5362.0
4	3'-CCATA<u>U</u>A<u>U</u>ATCCG-GCG	5274.0	5274.0
5	5'-A<u>C</u>A-GGTA<u>U</u>A<u>U</u>ATA-GGC	5592.0	5591.0
6	3'-CCATA<u>U</u>A<u>U</u>ATCCG-G<u>C</u>G	5504.0	5505.5
7	5'-TGCACA-GGTA<u>U</u>A<u>U</u>ATAGGC	6284.5	6284.0
8	3'-CCATA<u>U</u>A<u>U</u>ATCCGGCG-TAT	6195.0	6195.0
9	5'-<u>U</u>GCA<u>C</u>A-GGTA<u>U</u>A<u>U</u>ATAGGC	6731.0	6731.5
10	3'-CCATA<u>U</u>A<u>U</u>ATCCG-G<u>C</u>GTA<u>U</u>	6642.0	6642.5
11	5'-<u>U</u>GCACA-GGTATATATAGGC	6298.5	6300.0
12	3'-CCATATATATCCG-G<u>C</u>GTA<u>U</u>	6209.5	6210.0
15	5'-GGTA<u>U</u>A<u>U</u>ATAGGC-<u>C</u>GCA<u>U</u>A	6731.0	6732.0
16	3'-A<u>C</u>GTG<u>U</u>-CCATA<u>U</u>A<u>U</u>ATCCG	6642.0	6642.5
17	5'- **t**GCA**c**A-GGTA<u>U</u>A<u>U</u>ATAGGC	6354.0	6354.0
18	3'- CCATA<u>U</u>A<u>U</u>ATCCG-G**c**GTA**t**	6265.5	6266.0
19	5'- **t**G**c**A**C**a-GG<u>U</u>A<u>U</u>ATA<u>U</u>AGGC	6598.5	6599.5
20	3'-CCA<u>U</u>A<u>U</u>ATA<u>U</u>CCG-**g**CG**t**A**t**	6495.5	6497.0
DYZ-OPTu	5'-Cy3-T**g**T**g**TGT<u>U</u>A<u>U</u>ATGCTG<u>U</u>TCTC-3'	7636.0	7637.0
DYZ-OPTd	3'- AA<u>U</u>A<u>U</u>ACGACA<u>A</u>GAGTC**g**G**g**A-Cy3-5'	7738.0	7737.5

[a] All reported data are from MALDI-MS except for DYZ-OPTu and DYZ-OPTd (LC-ESI-MS).
[b] Data previously reported in reference 22.
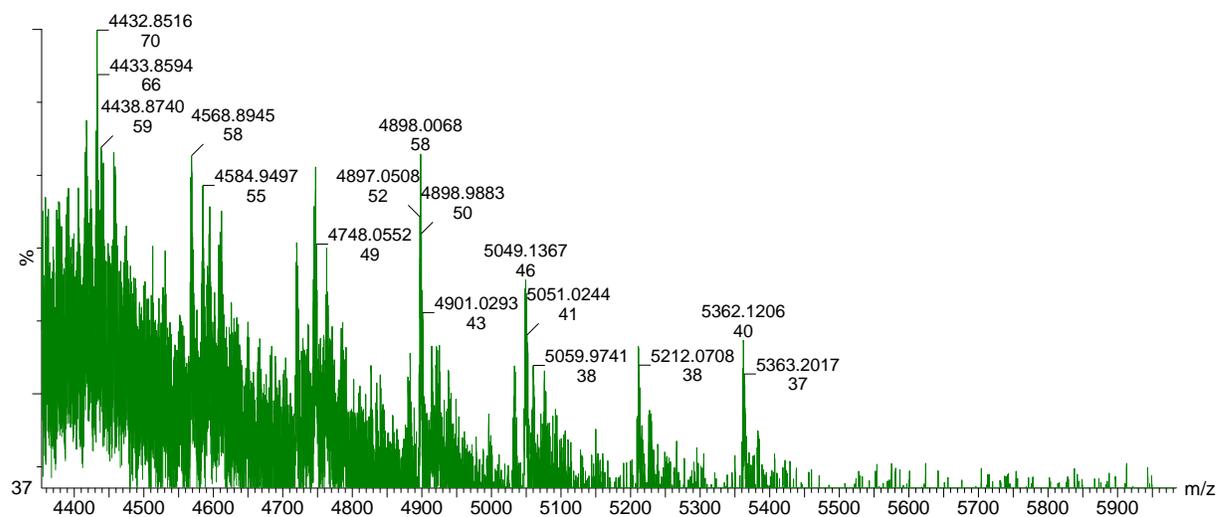
Figure 3.9. MALDI-MS spectrum of **ON3**.
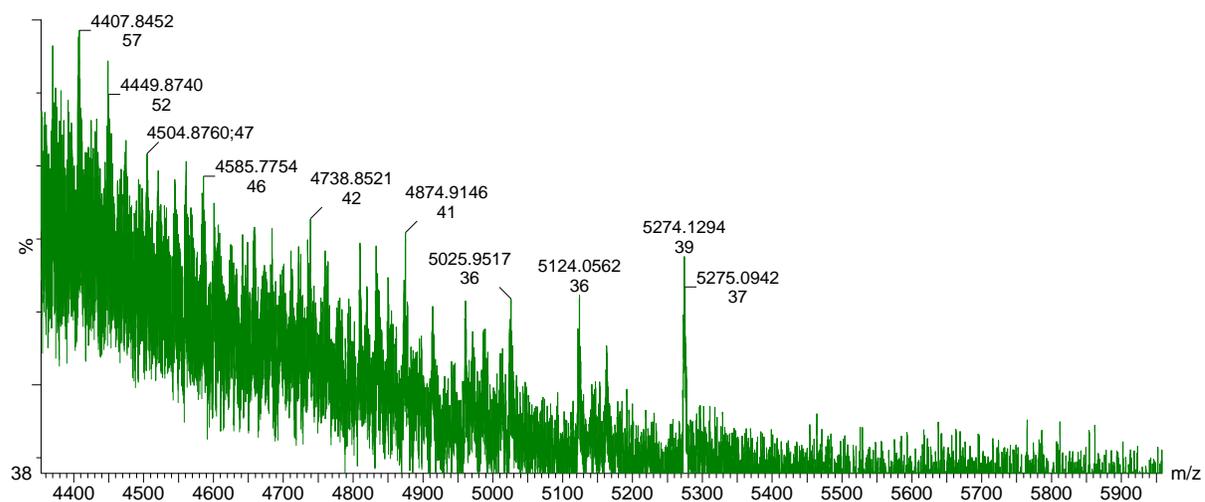
Figure 3.10. MALDI-MS spectrum of **ON4**.
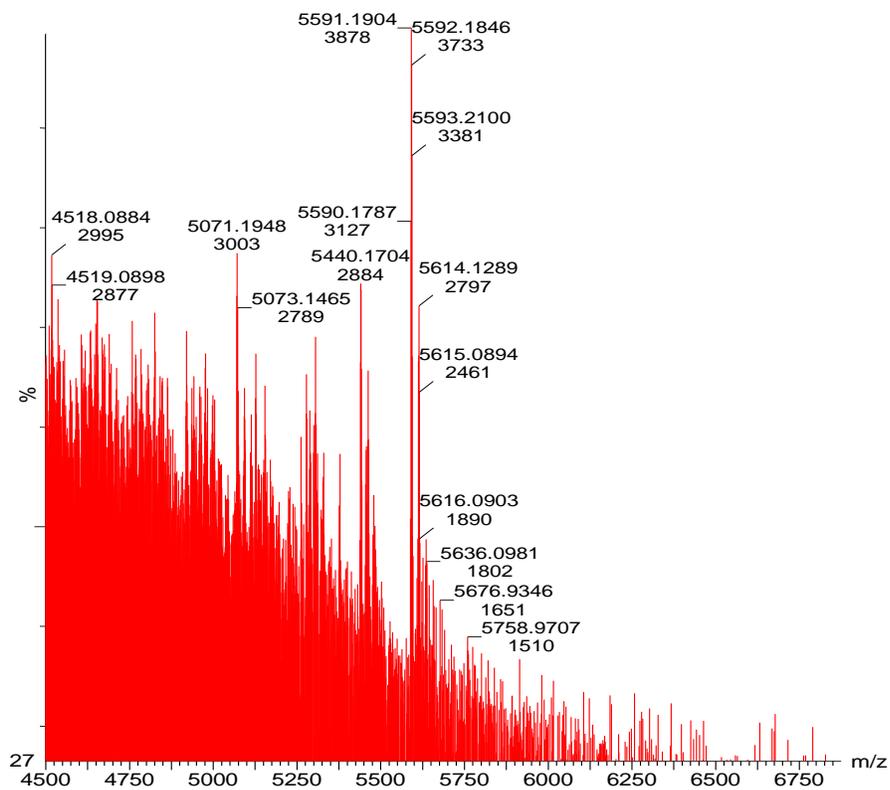
Figure 3.11. MALDI-MS spectrum of **ON5**.
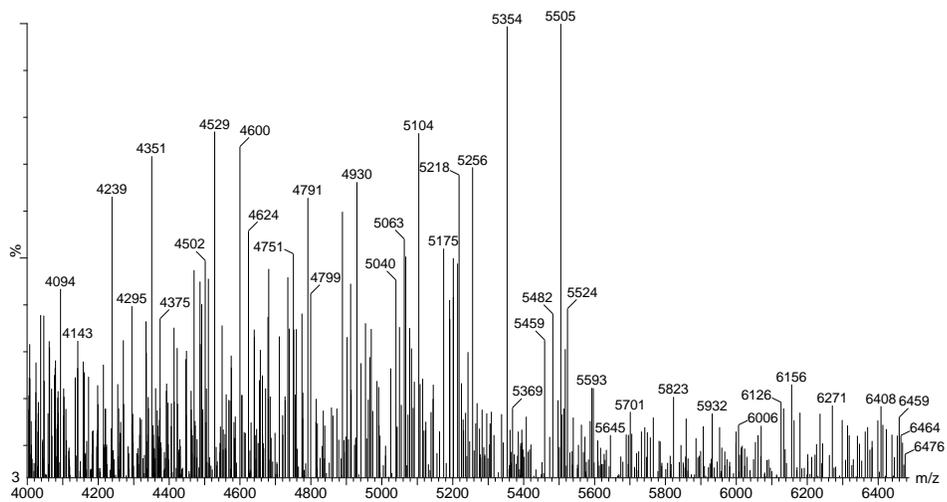
Figure 3.12. MALDI-MS spectrum of **ON6**.
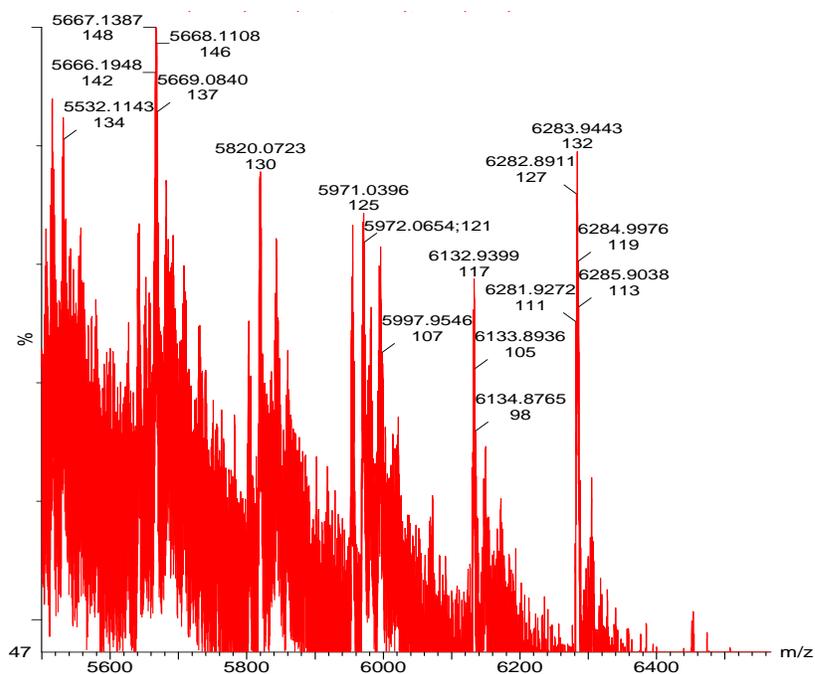
Figure 3.13. MALDI-MS spectrum of **ON7**.
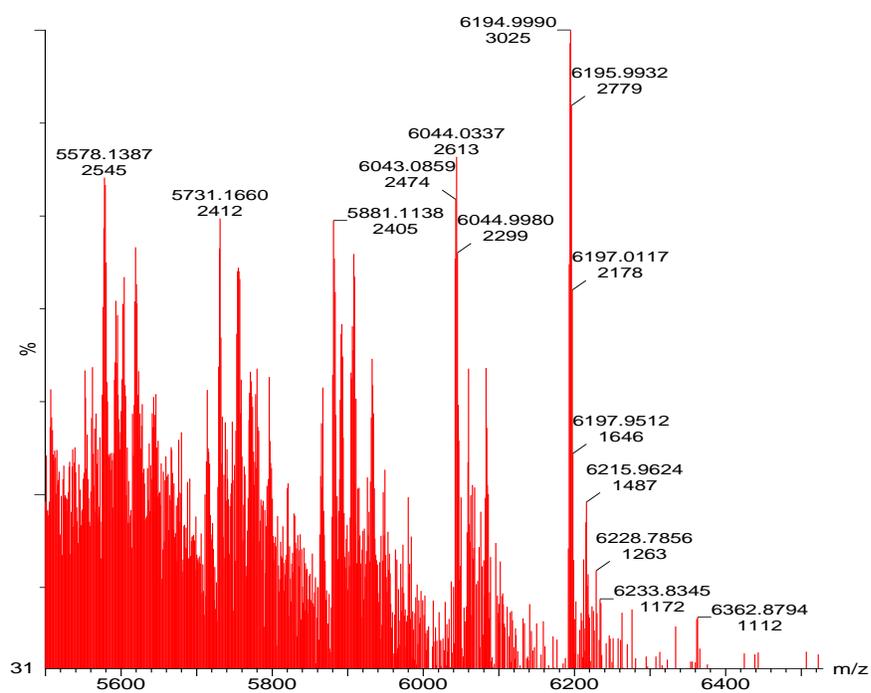
Figure 3.14. MALDI-MS spectrum of **ON8**.
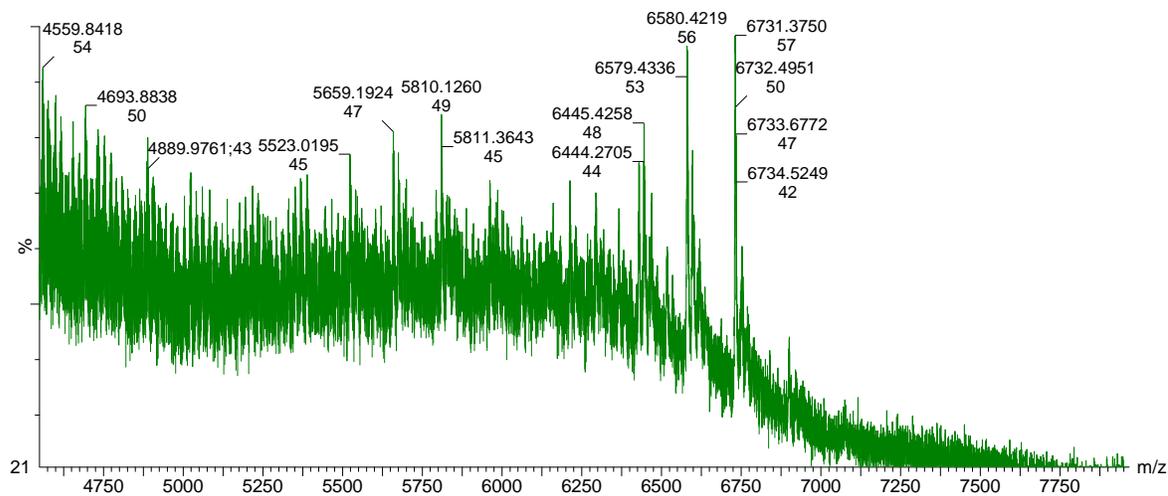
Figure 3.15. MALDI-MS spectrum of **ON9**.
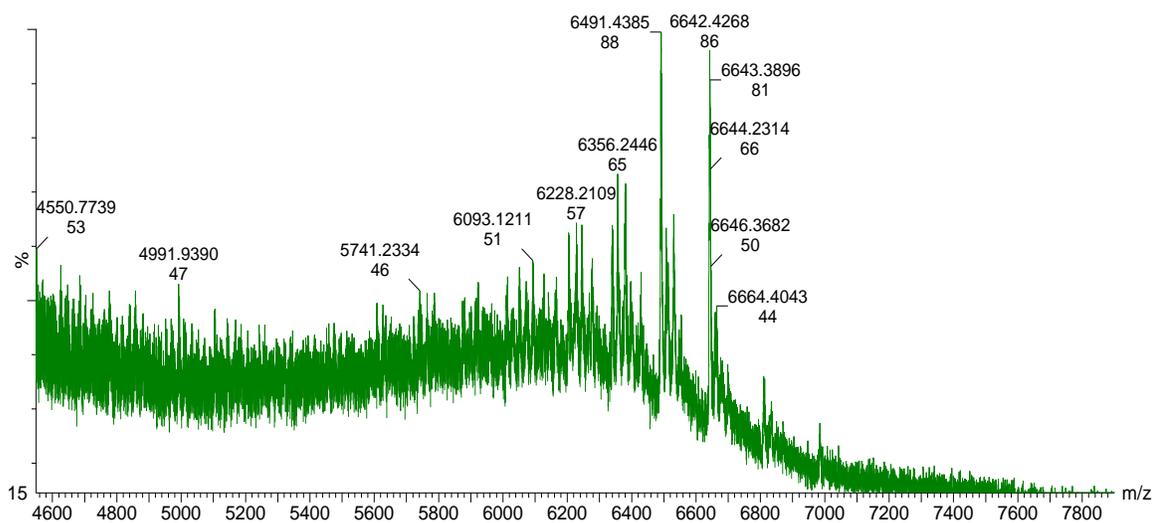
Figure 3.16. MALDI-MS spectrum of **ON10**.
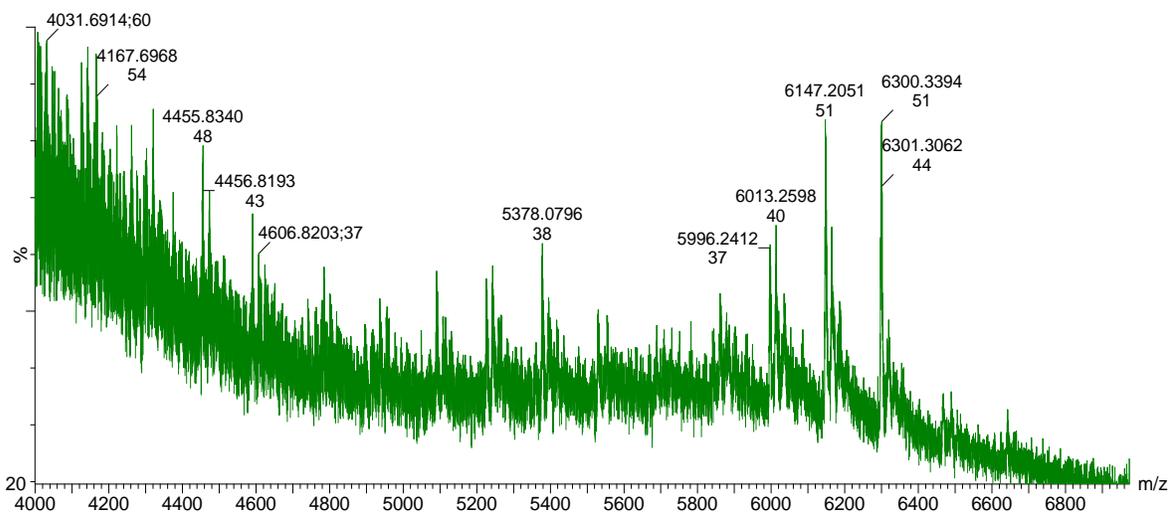
Figure 3.17. MALDI-MS spectrum of **ON11**.
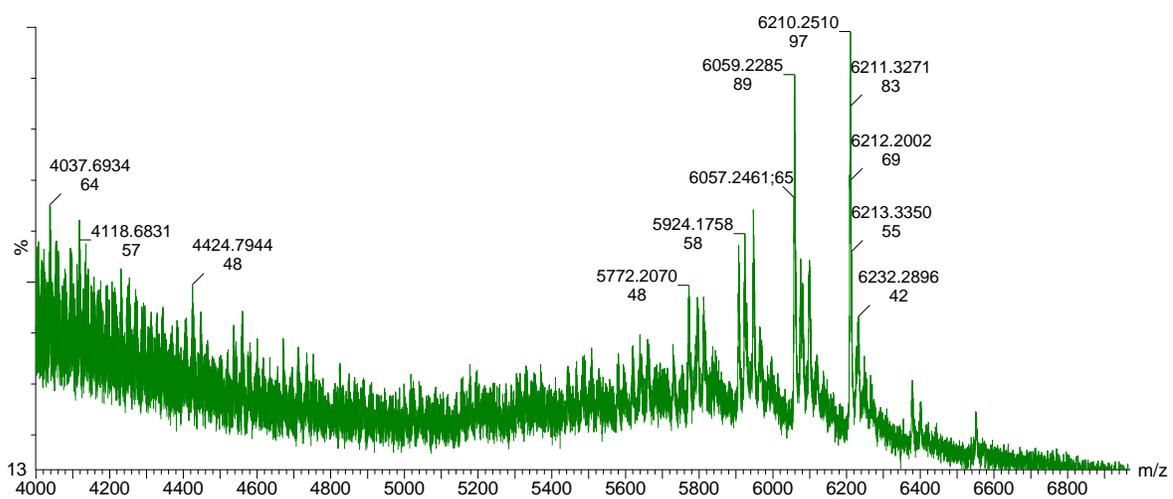
Figure 3.18. MALDI-MS spectrum of **ON12**.
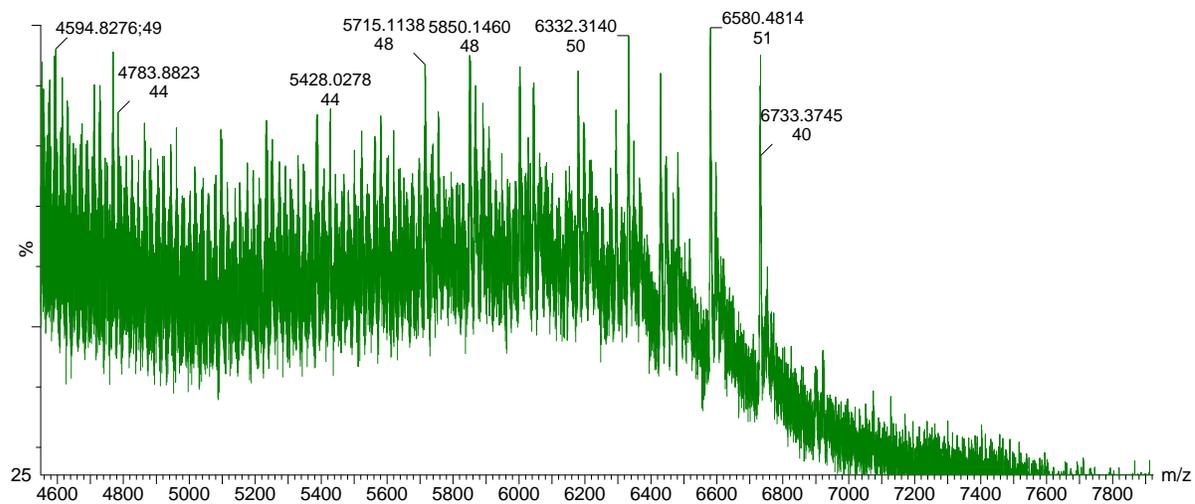
Figure 3.19. MALDI-MS spectrum of **ON15**.
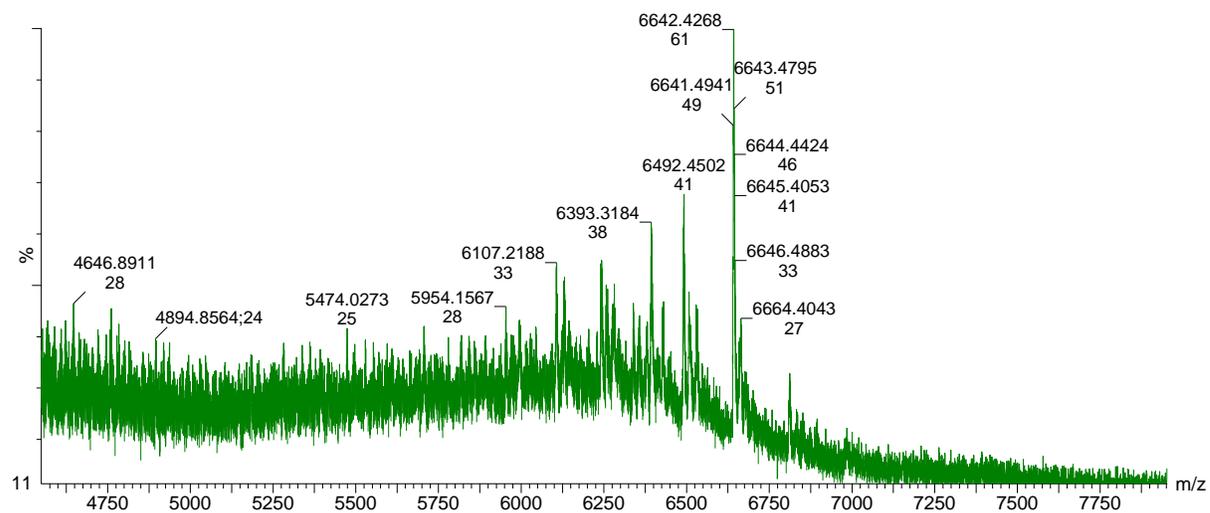
Figure 3.20. MALDI-MS spectrum of **ON16**.
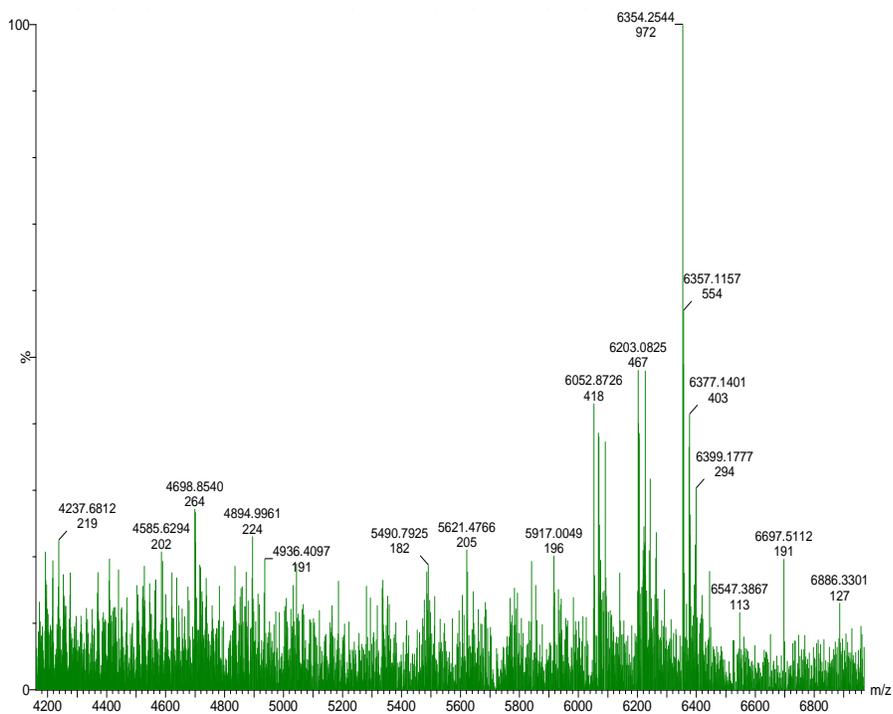
Figure 3.21. MALDI-MS spectrum of **ON17**.
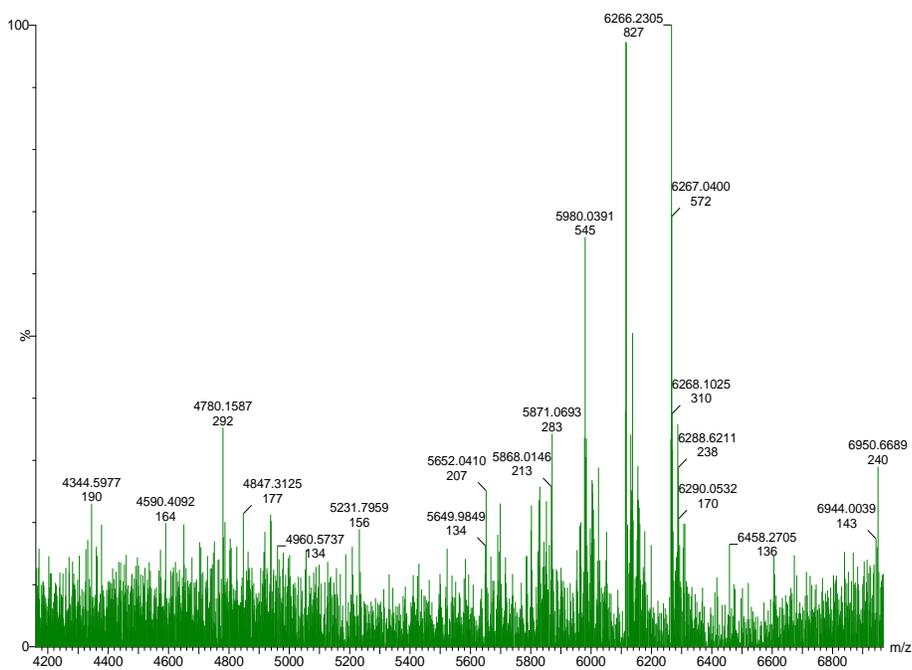
Figure 3.22. MALDI-MS spectrum of **ON18**.
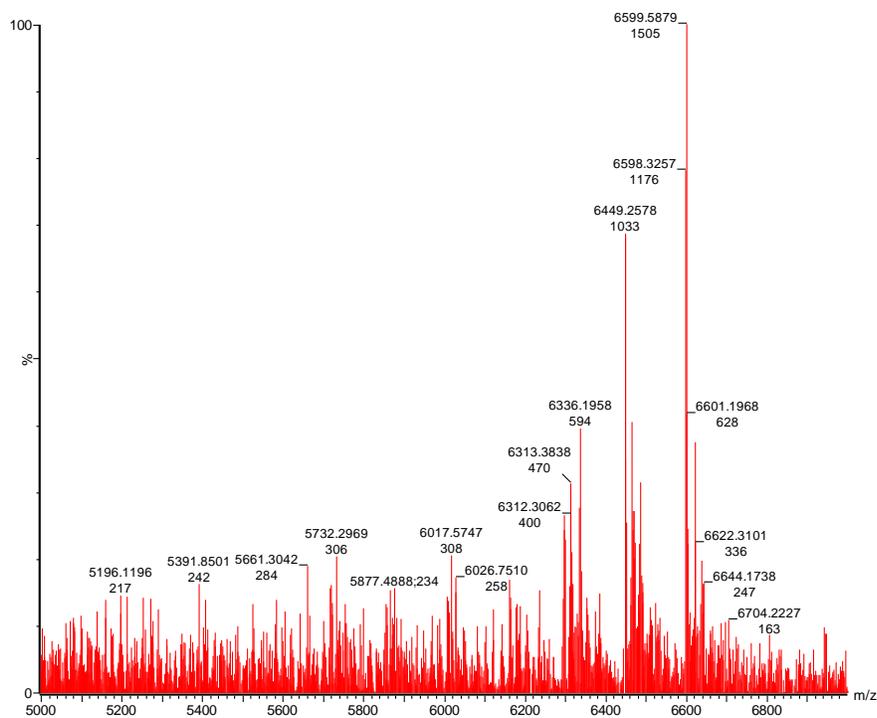
Figure 3.23. MALDI-MS spectrum of **ON19**.
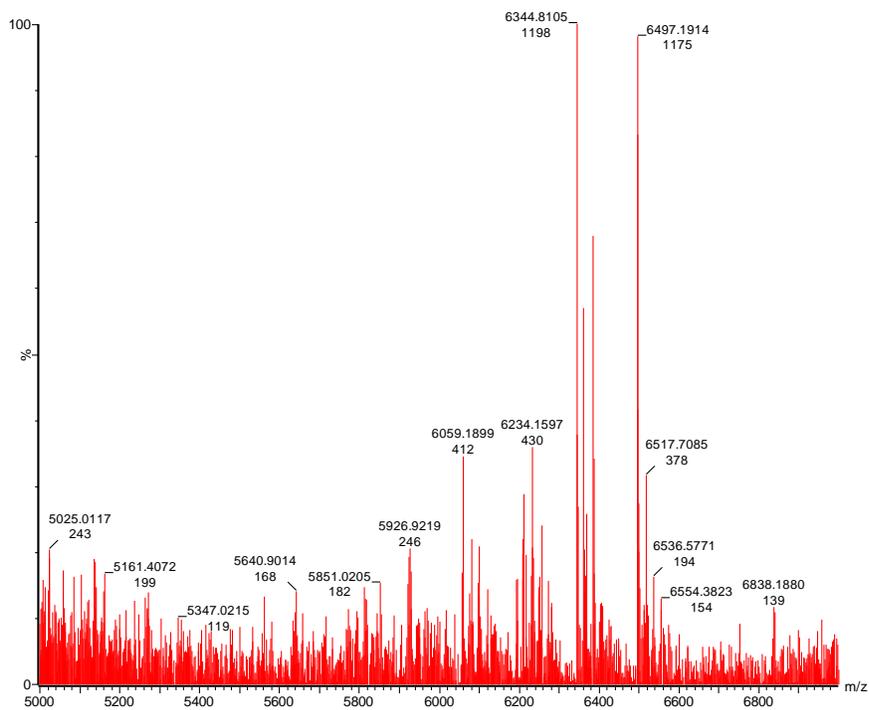
Figure 3.24. MALDI-MS spectrum of **ON20**.
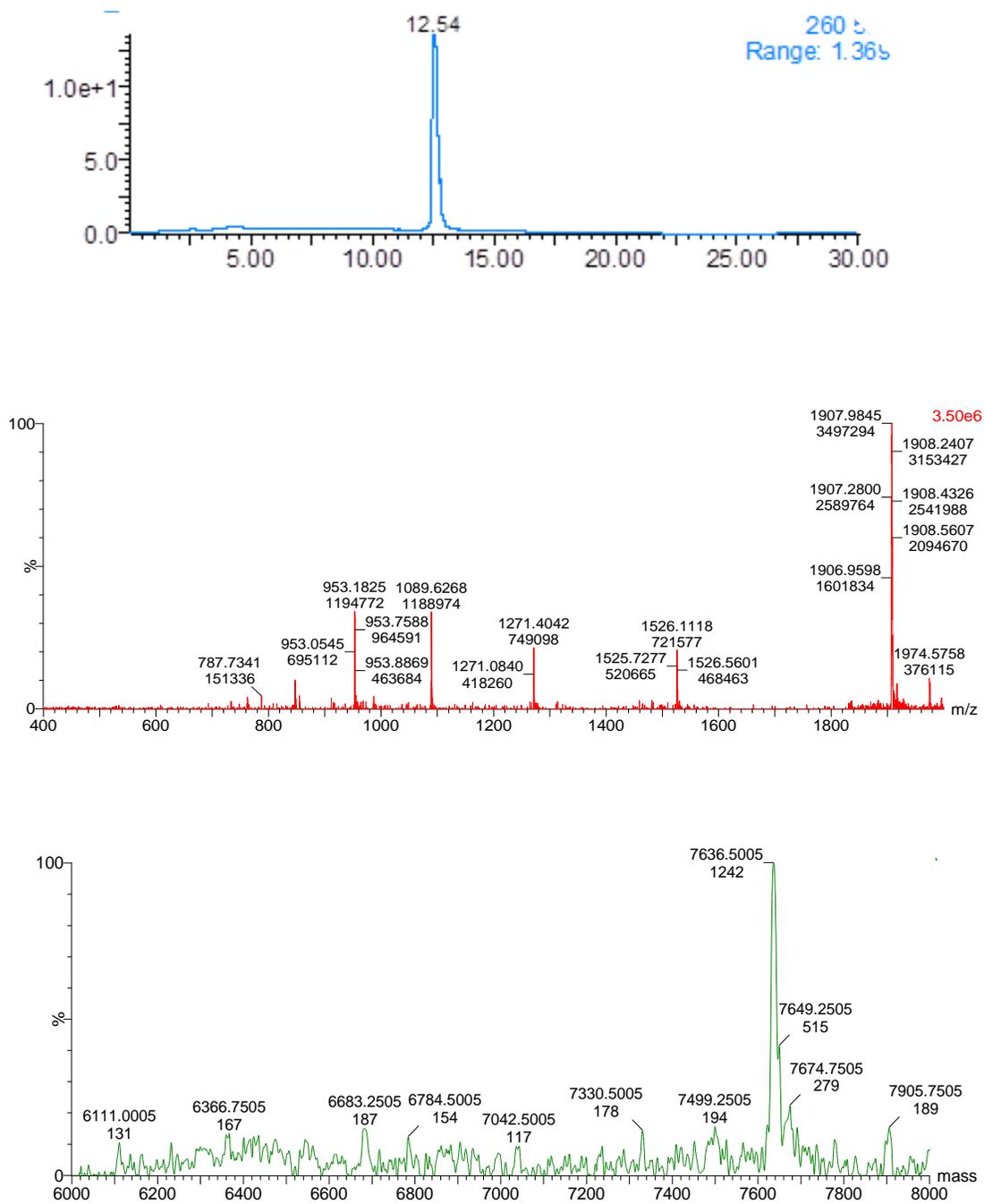
Figure 3.25. LC-ESI-MS analysis of **DYZ-OPTu.** LC-trace (upper panel), unprocessed (middle panel) and deconvoluted (lower panel) MS spectrum.
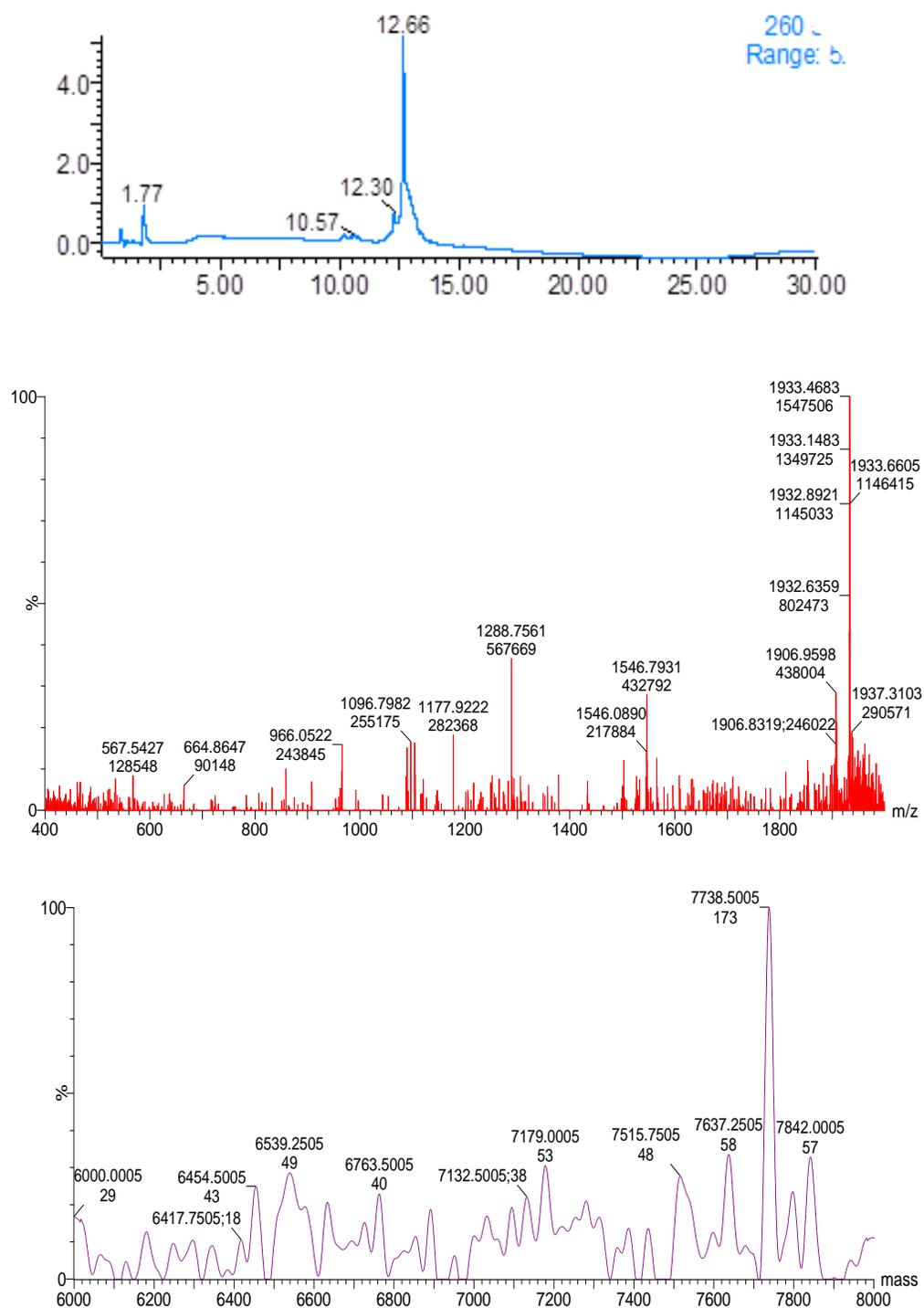
Figure 3.26. LC-ESI-MS analysis of **DYZ-OPTd**. LC trace (upper panel), unprocessed (middle panel) and deconvoluted (lower panel) MS spectrum.
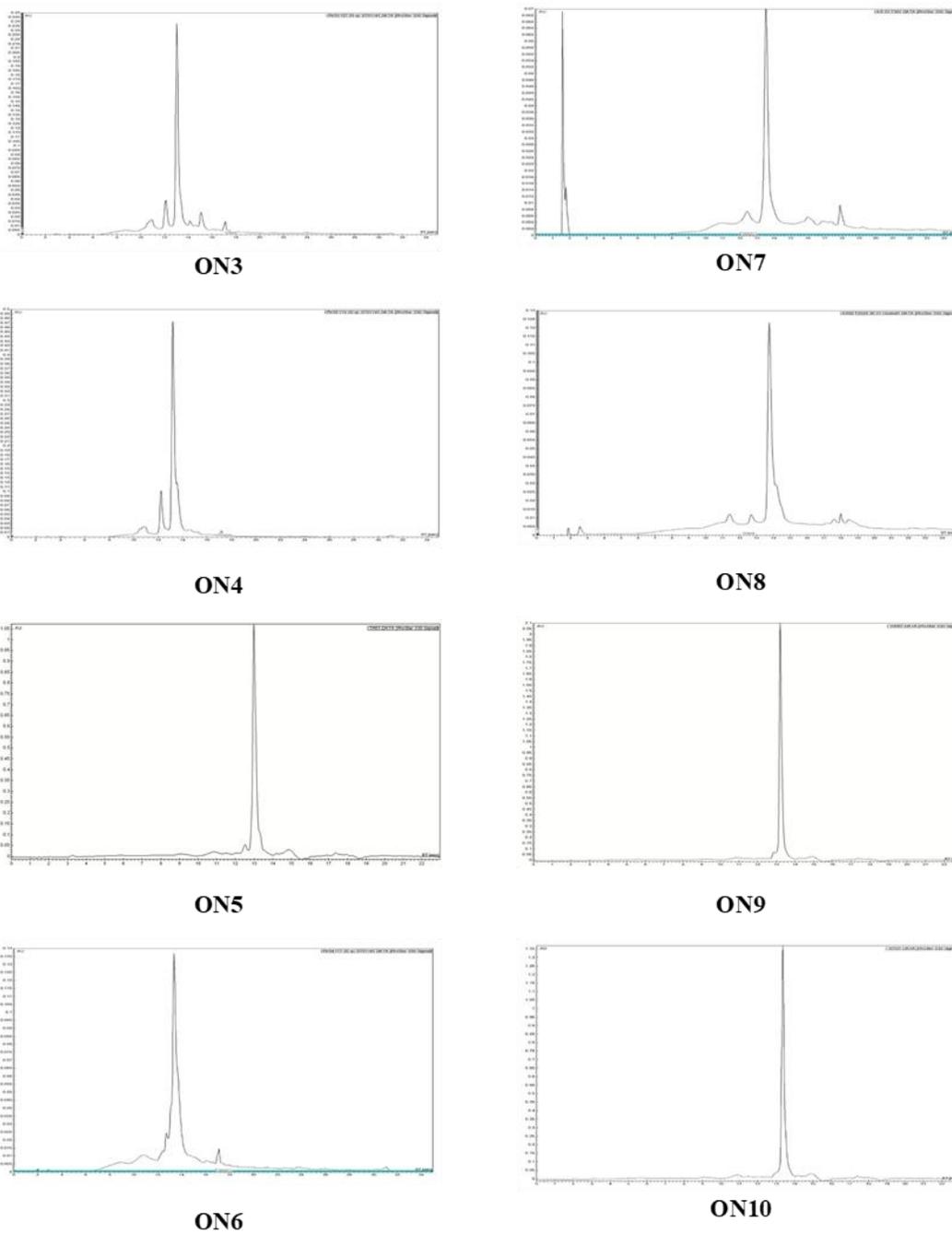
Figure 3.27. HPLC traces of **ON3-ON10**.
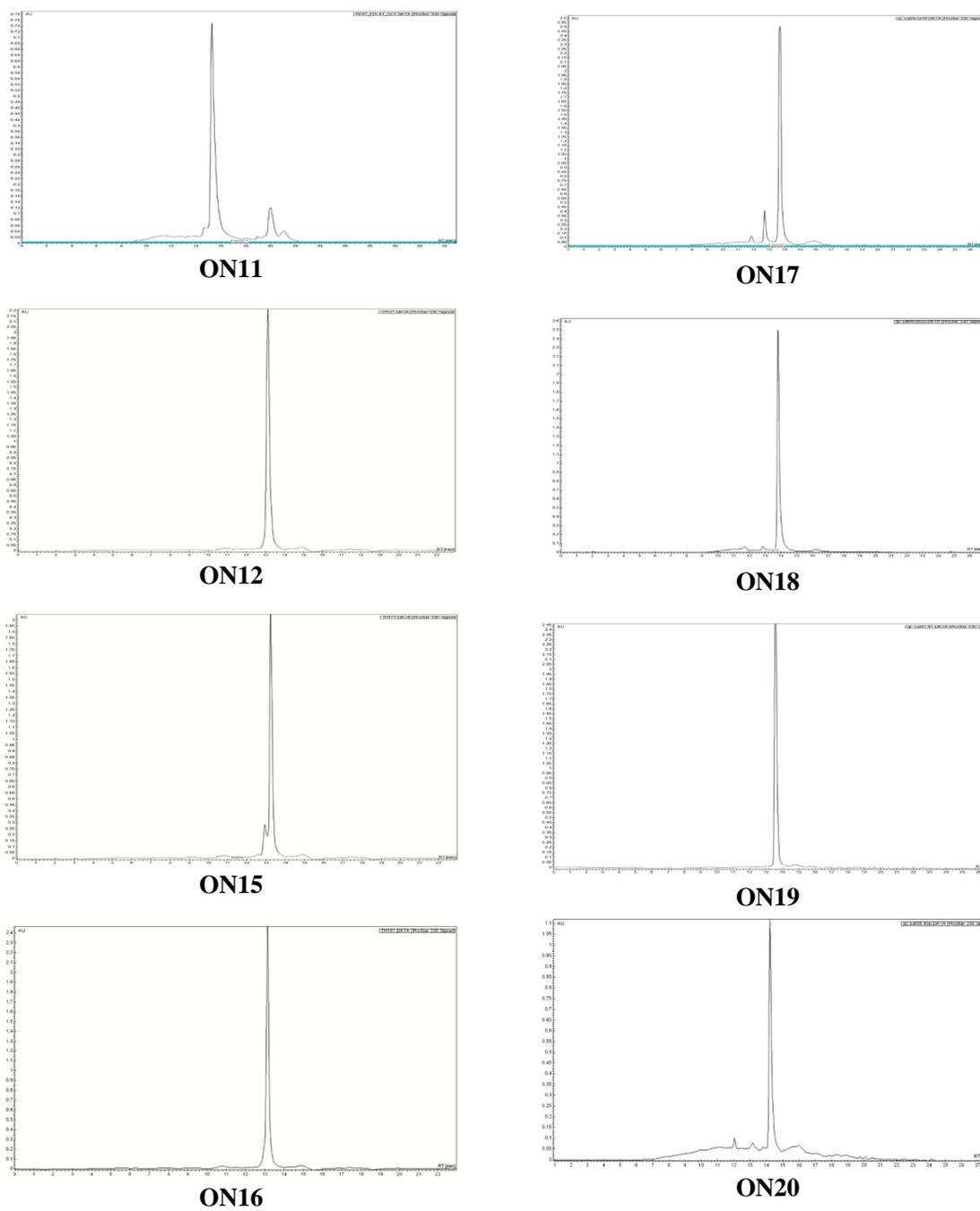
Figure 3.28. HPLC traces of **ON11, ON12, ON15, ON16, and ON17-ON20.**
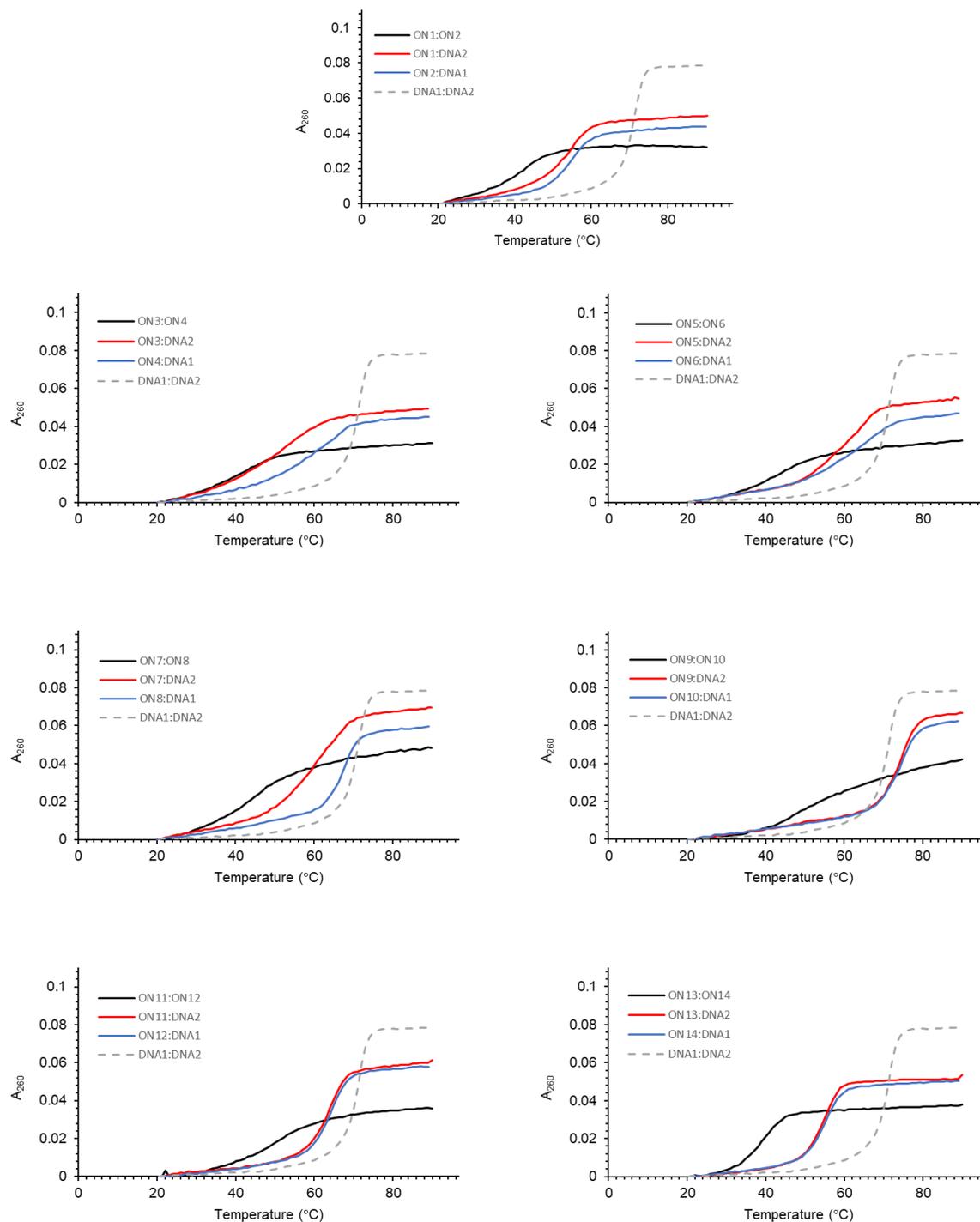
Figure 3.29. Representative thermal denaturation curves of double-stranded probes (**ON1**:**ON2**-**ON13**:**ON14**), duplexes between individual probe strands and single-stranded 33-mer DNA strands **DNA1** or **DNA2**, and model dsDNA target **DNA1**:**DNA2**. For experimental conditions, see Table 3.1.
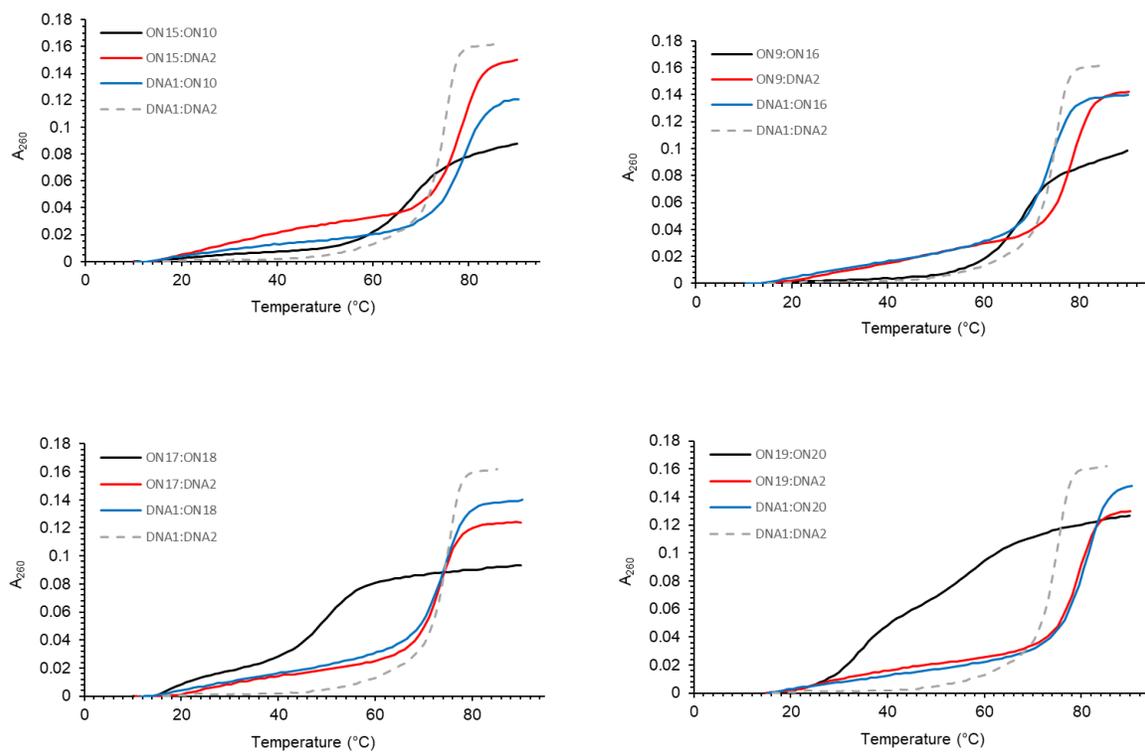
Figure 3.30. Representative thermal denaturation curves of Invader probes (**ON15**:**ON10**, **ON9:ON16**, **ON17**:**ON18**, and **ON19**:**ON20**), duplexes between individual probe strands and single-stranded 33-mer DNA strands **DNA1** or **DNA2**, and model dsDNA target **DNA1:DNA2** in medium salt buffer. For experimental conditions, see Table 3.1.
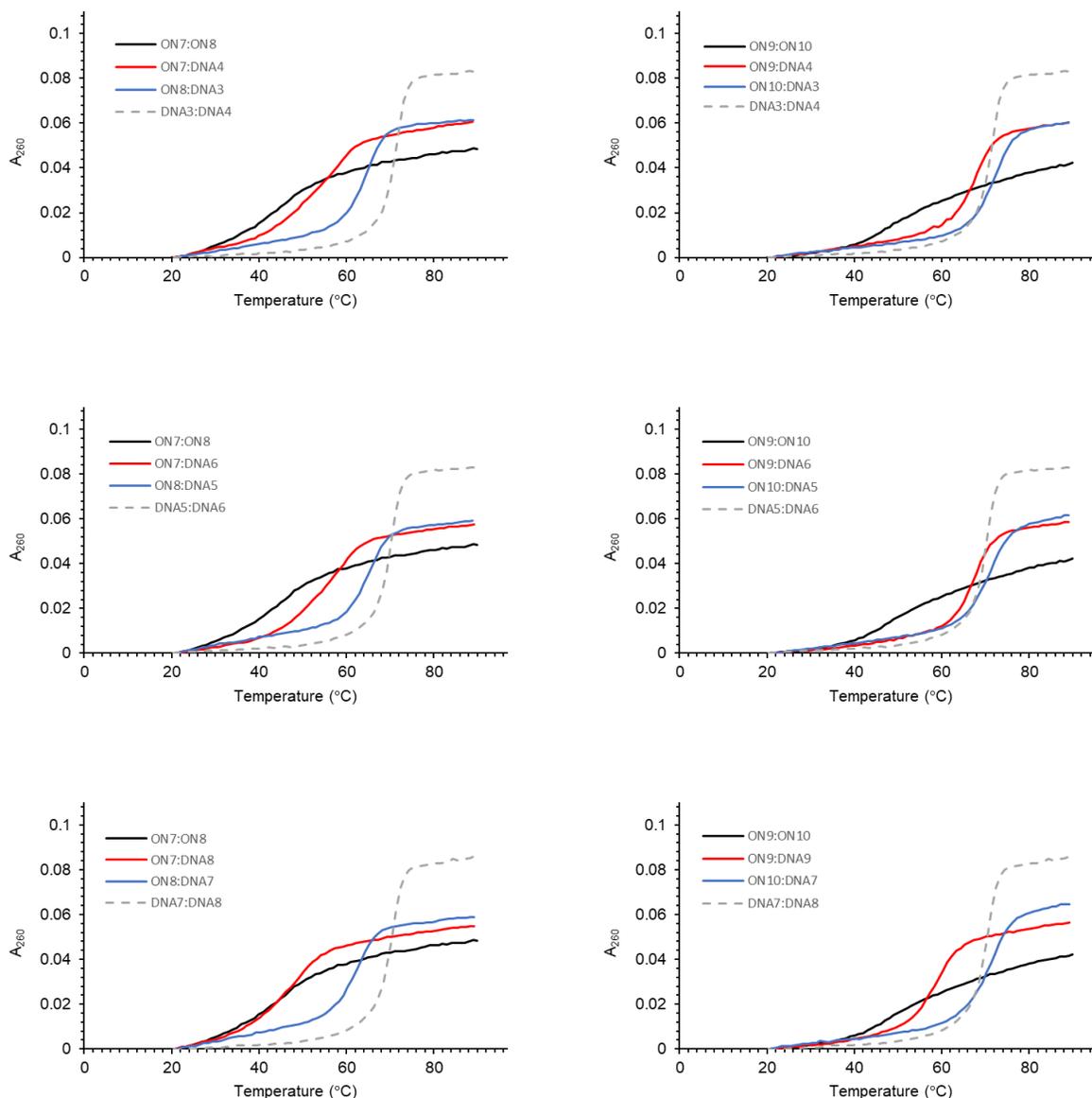
Figure 3.31. Representative thermal denaturation curves for Invader probes **ON7**:**ON8** and **ON9**:**ON10**, mismatched duplexes between individual probe strands and single-stranded non-target DNA strands, and **DNA3**:**DNA4**, **DNA5**:**DNA6** and **DNA7**:**DNA8**, which differ in sequence at one, two and three positions respectively, relative to **DNA1**:**DNA2** (for sequences of the mismatched dsDNA and tabulated data, see Tables 3.8 and 3.9). For experimental conditions, see Table 3.1.
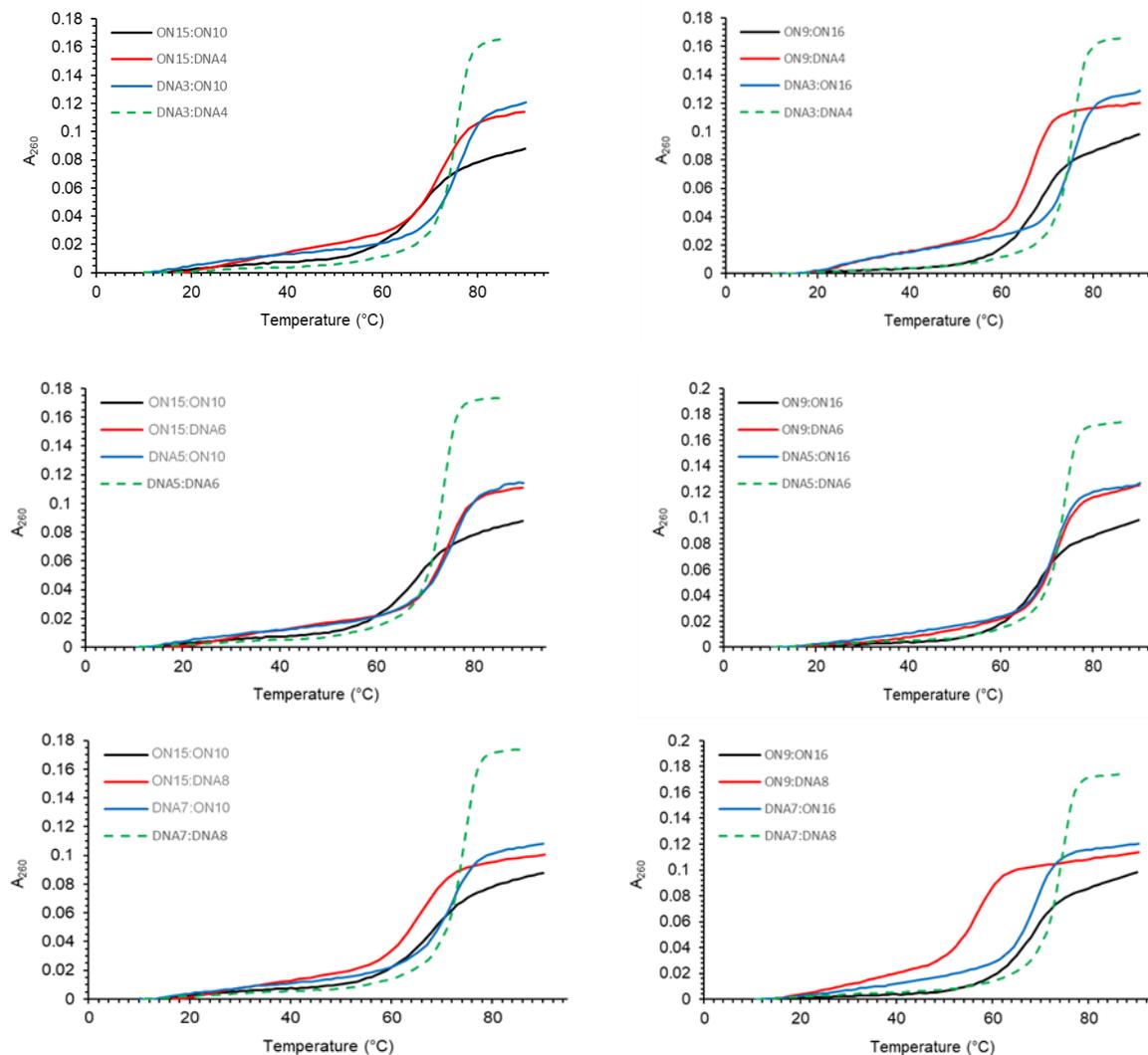
Figure 3.32. Representative thermal denaturation curves for Invader probes **ON15**:**ON10**, **ON9**:**ON16** and mismatched duplexes between individual probe strands and single-stranded non-targeted DNA strands, and **DNA3**:**DNA4**, **DNA5**:**DNA6** and **DNA7**:**DNA8**, which differ in sequence at one, two and three positions respectively, relative to **DNA1**:**DNA2** (for sequences of the mismatched dsDNA and tabulated data, see Tables 3.12 & 3.13). For experimental conditions, see Table 3.1.
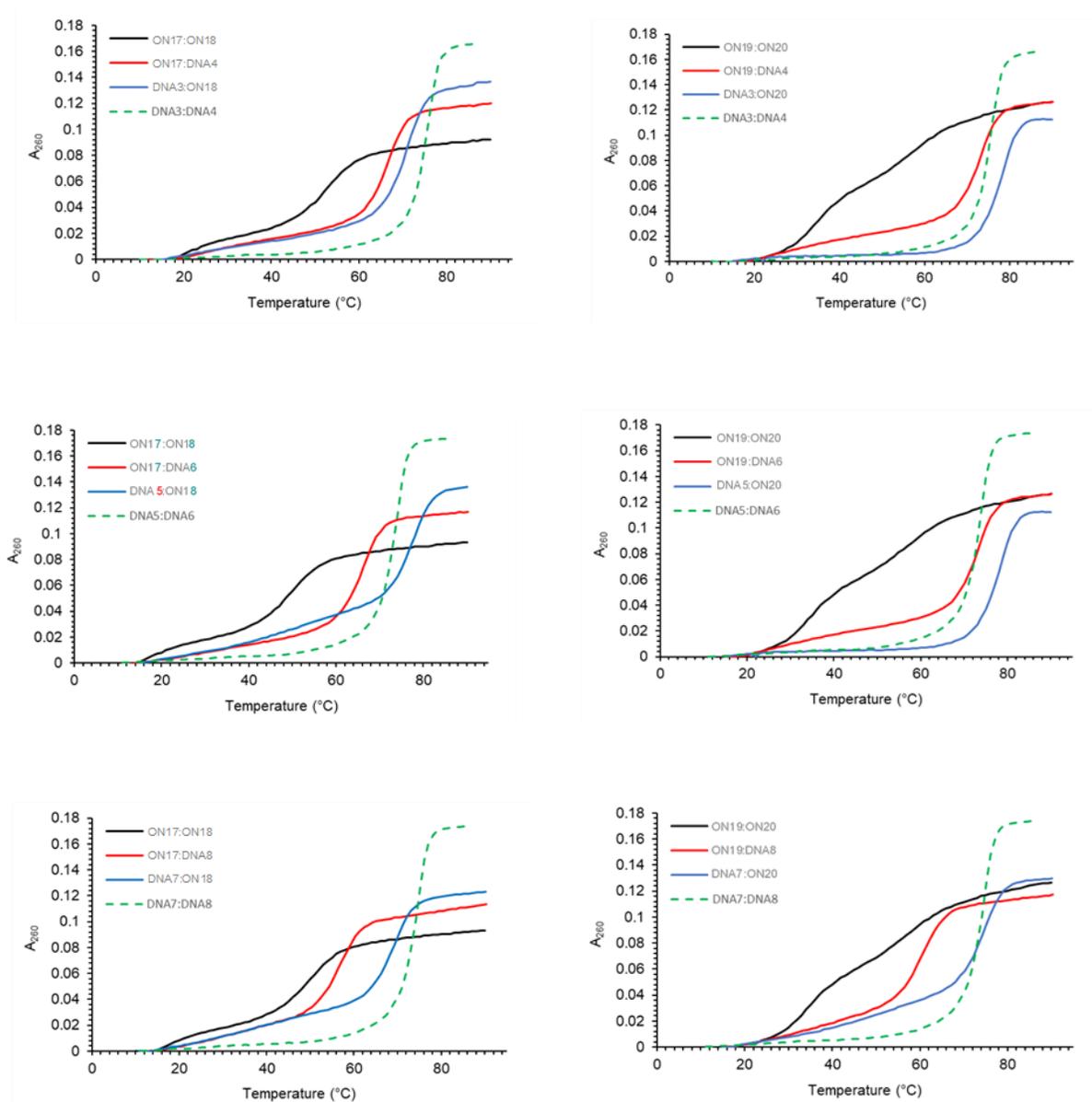
Figure 3.33. Representative thermal denaturation curves for Invader probes **ON17**:**ON18** and **ON19**:**ON20**, mismatched duplexes between individual probe strands and single-stranded non-targeted DNA strands, and **DNA3**:**DNA4**, **DNA5**:**DNA6** and **DNA7**:**DNA8**, which differ in sequence at one, two and three positions respectively, relative to **DNA1**:**DNA2** (for sequences of the mismatched dsDNA and tabulated data, see Tables 3.12 & 3.13). For experimental conditions, see Table 3.1.
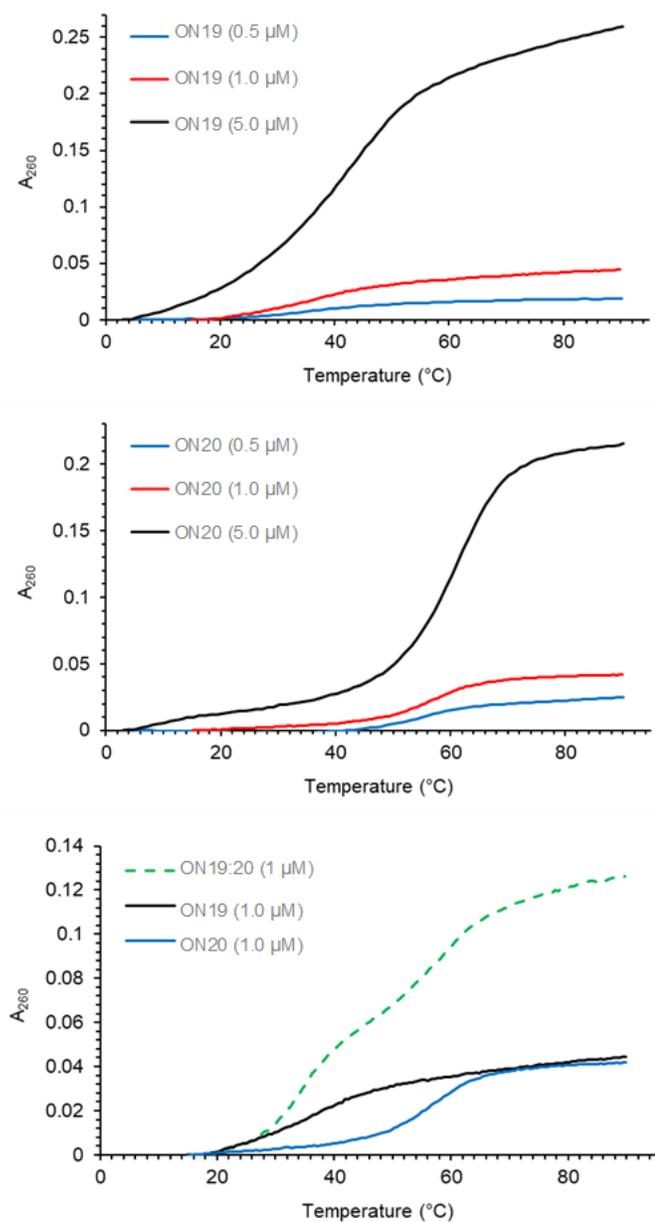
Figure 3.34. Representative thermal denaturation curves of individual Invader probe strands **ON19** and **ON20** at different concentrations in medium salt buffer (upper and middle panels) and of probe strands **ON19** and **ON20** as compared to **ON19:ON20** at 1 μM concentration (lower panel). For experimental conditions, see Table 3.1.

TA-based discussion of dsDNA-recognition potential of probes. The driving force for recognition of the 33-mer model dsDNA target **DNA1:DNA2** by double-stranded probes can be estimated by the term *thermal advantage*, defined as $TA = T_m$ (5'-probe vs **DNA2**) + T_m (3'-probe strand vs **DNA1**) - T_m (probe duplex) - T_m (**DNA1:DNA2**), with large positive values indicative of a probe that is strongly activated for dsDNA-recognition (Table 3.1).[16] The *TA*-based conclusions presented in the following generally align with the ΔG_{rec}^{310}-based conclusi\ons discussed in the main manuscript (Table 3.2). Thus, the 13-mer conventional Invader probe **ON1:ON2** is not sufficiently activated to facilitate recognition of the 33-mer model dsDNA target (*TA* = -6 °C, Table 3.1). Invader probe **ON3:ON4**, which features unmodified 3-mer overhangs, is minimally activated for DNA recognition (*TA* = 3.5 °C, Table 3.1), whereas Invader probes with modified 3-mer or unmodified 6-mer overhangs are moderately activated (*TA* = 11-13 °C, **ON5:ON6** and **ON7:ON8**, Table 3.1). Invader probe **ON9:ON10**, which features modified 6-mer overhangs, is strongly activated for dsDNA-recognition (*TA* = 30 °C, Table 3.1). This is the result of a highly destabilized probe duplex (ΔT_m = -23 °C, Table 3.1) and moderately stabilized probe-target duplexes (ΔT_m = 3.5 °C each, Table 3.1) as compared to the 33-mer dsDNA target. The impact of the +1 interstrand zipper arrangements of 2'-*O*-(pyren-1-yl)methyl-RNA monomers (i.e., energetic hotspots) in activating Invader probes for dsDNA-recognition is underscored by the markedly lower *TA* value for control probe **ON11:ON12** which lacks the energetic hotspots (*TA* = 6.5 °C, Table 3.1). Along similar lines, **ON13:ON14**, which lacks hotspots and features unmodified overhangs, is not activated for dsDNA-recognition (*TA* = -1.5 °C, Table 3.1). Conventional 19-mer Invader probes **ON15:ON10** and **ON9:ON16** are only moderately activated for DNA recognition (*TA* ~ 14 °C, Table 3.1), in large part due to the high probe stability.

*Enthalpic and entropic parameters associated with formation of double-stranded probes and duplexes between individual probe strands and **DNA1** or **DNA2**.* Formation of the double-stranded probes is less enthalpically favorable than formation of **DNA1:DNA2** (i.e., $\Delta\Delta H$ = 112-321 kJ/mol, Table 3.4). This is expected for two reasons, i.e., i) the double-stranded regions of the probes feature fewer base-pairs (bps) than **DNA1:DNA2** (i.e., 13 or 19 bps vis-à-vis 33 bps), and ii) the double-stranded regions of the all probes – except for **ON11:ON12** and **ON13:ON14** - feature two or four +1 interstrand zipper arrangements of 2'-*O*-(pyren-1-

yl)methyl-RNA monomers, which are known to destabilize duplexes[,17,18] due to violation of the neighbor exclusion principle.[19,33]

Formation of duplexes between short individual probe strands and the corresponding single-stranded 33-mer DNA strands is generally less enthalpically favorable than formation of **DNA1**:**DNA2** (e.g., $\Delta\Delta H$ for **ON2/ON4/ON6**:**DNA1** = 150-239 kJ/mol, Table 3.4). Conversely, formation of duplexes between long individual probe strands and the corresponding 33-mer single-stranded DNA target is generally more enthalpically favorable than formation of **DNA1**:**DNA2**, presumably due to the stabilizing effect from pyrene intercalation[18] (e.g., $\Delta\Delta H$ for **ON8/ON10/ON12/ON14/ON16**:**DNA1** between -228 and -109 kJ/mol, Table 3.4). Consequently, the calculated change in reaction enthalpy upon probe-mediated recognition of the 33-mer target **DNA1**:**DNA2** is enthalpically unfavorable for short Invader probes (e.g., see ΔH_{rec} for **ON3**:**ON4** and **ON5**:**ON6**, Table 3.4), while being enthalpically favorable for long Invader probes (e.g., ΔH_{rec} for **ON7**:**ON8/ON9**:**ON10/ON15**:**ON10/ON9**:**ON16** between -654 and -236 kJ/mol, Table 3.4).

Enthalpy-entropy compensation is observed, i.e., whenever formation of a given duplex is enthalpically favorable (Table 3.4), it is entropically unfavorable (Table 3.5). Accordingly, the calculated change in reaction entropy upon probe-mediated recognition of the 33-mer target **DNA1**:**DNA2** is entropically favorable for short Invader probes (e.g., see $-T\Delta S_{rec}^{310}$ for **ON3**:**ON4** and **ON5**:**ON6**, Table 3.5) and unfavorable for long Invader probes (e.g., $-T\Delta S_{rec}^{310}$ for **ON7**:**ON8/ON9**:**ON10** /**ON15**:**ON10/ON9**:**ON16** between +220 and +567 kJ/mol, Table 3.5).

Table 3.4. Change in enthalpy (ΔH) upon formation of probe duplexes and duplexes between individual probe strands and **DNA1** or **DNA2**. Also shown is the calculated change in reaction enthalpy upon probe-mediated recognition of 33-mer target **DNA1:DNA2** (ΔH_{rec}).[a]

| Probe | Sequence | ΔH [$\Delta\Delta H$] (kJ/mol) | | | ΔH_{rec} (kJ/mol) |
		probe duplex	5'-probe vs DNA2	3'-probe vs DNA1	
ON1	5'-GGTA<u>UA</u>UATAGGC	-406	-564	-368	-8
ON2	3'-CCATA<u>UA</u>UATCCG	[+112]	[-46]	[+150]	
ON3	5'-ACA-GGTA<u>UA</u>UATAGGC	-270	-264	-279	+245
ON4	3'-CCATA<u>UA</u>UATCCG-GCG	[+248]	[+254]	[+239]	
ON5	5'-A<u>C</u>A-GGTA<u>UA</u>UATAGGC	-261	-360	-299	+120
ON6	3'-CCATA<u>UA</u>UATCCG-G<u>C</u>G	[+257]	[+158]	[+219]	
ON7	5'-TGCACA-GGTA<u>UA</u>UATAGGC	-277	-361	-670	-236
ON8	3'-CCATA<u>UA</u>UATCCG-GCGTAT	[+241]	[+157]	[-152]	
ON9	5'-<u>U</u>GCA<u>C</u>A-GGTA<u>UA</u>UATAGGC	-197	-718	-647	-654
ON10	3'-CCATA<u>UA</u>UATCCG-G<u>C</u>GTA<u>U</u>	[+321]	[-200]	[-129]	
ON11	5'-<u>U</u>GCA<u>C</u>A-GGTATATATAGGC	-259	-506	-627	-356
ON12	3'-CCATATATATCCG-G<u>C</u>GTA<u>U</u>	[+259]	[+12]	[-109]	
ON13	5'-TGCACA-GGTATATATAGGC	-523	-639	-629	-227
ON14	3'-CCATATATATCCG-GCGTAT	[-5]	[-121]	[-111]	
ON15	5'-GGTA<u>UA</u>UATAGGC-<u>C</u>GCA<u>UA</u>	-307	-712	-647	-534
ON10	3'-CCATA<u>UA</u>UATCCG-G<u>C</u>GTA<u>U</u>	[+211]	[-194]	[-129]	
ON9	5'-<u>U</u>GCA<u>C</u>A-GGTA<u>UA</u>UATAGGC	-364	-718	-746	-582
ON16	3'- A<u>C</u>GTG<u>U</u>-CCATA<u>UA</u>UATCCG	[+154]	[-200]	[-228]	

[a] $\Delta\Delta H$ is calculated relative to the unmodified 33-mer target DNA duplex **DNA1:DNA2** (ΔH = -518 kJ/mol). ΔH_{rec} = ΔH (5'-probe:**DNA2**) + ΔH (3'-probe:**DNA1**) - ΔH (probe duplex) - ΔH (**DNA1:DNA2**). For experimental conditions, see Table 3.1.

Table 3.5. Change in entropy at 310 K ($-T\Delta S^{310}$) upon formation of probe duplexes and duplexes between individual probe strands and **DNA1** or **DNA2**. Also shown is the calculated change in reaction entropy upon probe-mediated recognition of 33-mer target **DNA1:DNA2** ($-T\Delta S^{310}_{rec}$).[a]

| Probe | Sequence | $-T\Delta S^{310}$ [$\Delta(T\Delta S^{310})$] (kJ/mol) | | | $-T\Delta S^{310}_{rec}$ |
		probe duplex	5'-probe vs DNA2	3'-probe vs DNA1	(kJ/mol)
ON1	5'-GGTA<u>UA</u>UATAGGC	+357	+493	+310	+22
ON2	3'-CCATA<u>UA</u>UATCCG	[-67]	[+69]	[-114]	
ON3	5'-ACA-GGTA<u>UA</u>UATAGGC	+226	+210	+218	-222
ON4	3'-CCATA<u>UA</u>UATCCG-GCG	[-198]	[-214]	[-206]	
ON5	5'-A<u>C</u>A-GGTA<u>UA</u>UATAGGC	+215	+294	+235	-110
ON6	3'-CCATA<u>UA</u>UATCCG-G<u>C</u>G	[-209]	[-130]	[-189]	
ON7	5'-TGCACA-GGTA<u>UA</u>UATAGGC	+229	+294	+579	+220
ON8	3'-CCATA<u>UA</u>UATCCG-GCGTAT	[-195]	[-130]	[+155]	
ON9	5'-<u>U</u>GCA<u>C</u>A-GGTA<u>UA</u>UATAGGC	+150	+599	+536	+567
ON10	3'-CCATA<u>UA</u>UATCCG-G<u>C</u>GTA<u>U</u>	[-274]	[+175]	[+112]	
ON11	5'-<u>U</u>GCA<u>C</u>A-GGTATATATAGGC	+207	+426	+535	+330
ON12	3'-CCATATATATCCG-G<u>C</u>GTA<u>U</u>	[-217]	[+2]	[+111]	
ON13	5'-TGCACA-GGTATATATAGGC	+477	+563	+554	+216
ON14	3'-CCATATATATCCG-GCGTAT	[+53]	[+139]	[+130]	
ON15	5'-GGTA<u>UA</u>UATAGGC-<u>C</u>GCA<u>UA</u>	+243	+593	+536	+462
ON10	3'-CCATA<u>UA</u>UATCCG-G<u>C</u>GTA<u>U</u>	[-181]	[+169]	[+112]	
ON9	5'-<u>U</u>GCA<u>C</u>A-GGTA<u>UA</u>UATAGGC	+294	+599	+623	+504
ON16	3'- A<u>C</u>GTG<u>U</u>-CCATA<u>UA</u>UATCCG	[-130]	[+175]	[+199]	

[a] $\Delta(T\Delta S^{310})$ is calculated relative to the unmodified 33-mer target DNA duplex **DNA1:DNA2** ($-T\Delta S^{310}$ = 424 kJ/mol). $-T\Delta S^{310}_{rec}$= $\Delta(T\Delta S^{310})$ (5'-probe:**DNA2**) + $\Delta(T\Delta S^{310})$ (3'-probe:**DNA1**) - $\Delta(T\Delta S^{310})$ (probe duplex). For experimental conditions, see Table 3.1.

Table 3.6. Quantification of **DNA1**:**DNA2**-recognition at 37 °C using a 5-fold molar excess of different double-stranded probes.[a]

Probe	Sequence	% Recognition
1:2	5'-GGTA**UA**U**A**TAGGC-3' 3'-CCATA**UA**U**A**TCCG-5'	<5
3:4	5'-ACA-GGTA**UA**U**A**TAGGC-3' 3'-CCATA**UA**U**A**TCCG-GCG-5'	<5
5:6	5'-A**C**A-GGTA**UA**U**A**TAGGC-3' 3'-CCATA**UA**U**A**TCCG-G**C**G-5'	<5
7:8	5'-TGCACA-GGTA**UA**U**A**TAGGC-3' 3'-CCATA**UA**U**A**TCCG-GCGTAT-5'	30 ± 6
9:10	5'-**U**GCA**C**A-GGTA**UA**U**A**TAGGC-3' 3'-CCATA**UA**U**A**TCCG-G**C**GTA**U**-5'	45 ± 6
11:12	5'-**U**GCA**C**A-GGTATATATAGGC-3' 3'-CCATATATATCCG-G**C**GTA**U**-5'	39 ± 4
13:14	5'-TGCACA-GGTATATATAGGC-3' 3'-CCATATATATCCG-GCGTAT-5'	<5
15:10	5'-GGTA**UA**U**A**TAGGC-**C**GCA**U**A-3' 3'- CCATA**UA**U**A**TCCG-G**C**GTA**U**-5'	<5
9:16	5'-**U**GCA**C**A-GGTA**UA**U**A**TAGGC-3' 3'-A**C**GTG**U**-CCATA**UA**U**A**TCCG-5'	<5
17:18	5'-**t**GCA**c**A-GGTA**UA**U**A**TAGGC-3' 3'-CCATA**UA**U**A**TCCG-G**c**GTA**t**-5'	46 ± 5
19:20	5'-**t**G**c**A**C**a-GG**UA**U**A**TA**U**AGGC-3' 3'-CCA**UA**U**A**TA**U**CCG-**g**CG**t**At-5'	43 ± 6

[a] Experiments were performed in triplicate. Conditions are described in Fig. 3.2.

Supplemental discussion of dsDNA-targeting properties of ON7:ON8-ON13:ON14. We were initially surprised to observe that **ON7**:**ON8**, **ON11**:**ON12** and – to a lesser extent **ON13**:**ON14** – result in recognition of **DNA1**:**DNA2** given that a net loss of eight base-pairs ensues, i.e., the 13 bps of the double-stranded probes and 33 bps of **DNA1**:**DNA2** denature to form two 19-mer probe-target duplexes, each featuring a 4-mer and a 10-mer overhang (Fig. 3.35 - left). We speculate that i) the 1-3 outermost base-pairs at each end of **DNA1**:**DNA2** are subject to end-fraying effects[44] and thus only contribute minimally to the overall duplex stability, whereas ii) the overhangs in the probe duplexes and probe-target duplexes reduce fraying due to capping effects, such that the actual energy loss upon recognition of **DNA1**:**DNA2** more closely corresponds to a loss of 2-6 bps. We further speculate that this net loss of base pairs is negated through the use of affinity-enhancing modifications and/or an excess of probe.

Analogously, recognition of **DNA1**:**DNA2** by **ON3**:**ON4** or **ON5**:**ON6** would entail a net loss of 14 bps, i.e., the 13 bps of the double-stranded probes and the 33 bps of **DNA1**:**DNA2** being denatured to form two 16-mer duplexes, each encompassing a 7-mer and a 10-mer single-stranded overhang (Fig. 3.35 - right). Taking end-fraying effects into account, the loss in energy likely corresponds more closely to 8-12 bps, which, in contrast, cannot be overcome through the use of a probe excess and/or by incorporating 2-3 affinity-enhancing modifications in each probe strand.

To test this hypothesis, **ON7**:**ON8**-**ON13**:**ON14** were incubated with an extended version of **DNA1**:**DNA2**, i.e., a 39-mer dsDNA target in which three additional base-pairs are added to each end (Fig. 3.36). Recognition is expected to result in a net loss of 14 bps (and 8-12 bps if end-fraying effects as discussed above are considered). Indeed, recognition of **DNA1**:**DNA2** is not observed with **ON7**:**ON8**, **ON11**:**ON12** or **ON13**:**ON14**, while less efficient recognition (compared to recognition of **DNA1**:**DNA2**) is observed with **ON9**:**ON10**, lending support to the hypothesis (Fig. 3.36).

= 2'-*O*-(pyren-1-yl)methyl-RNA monomers.
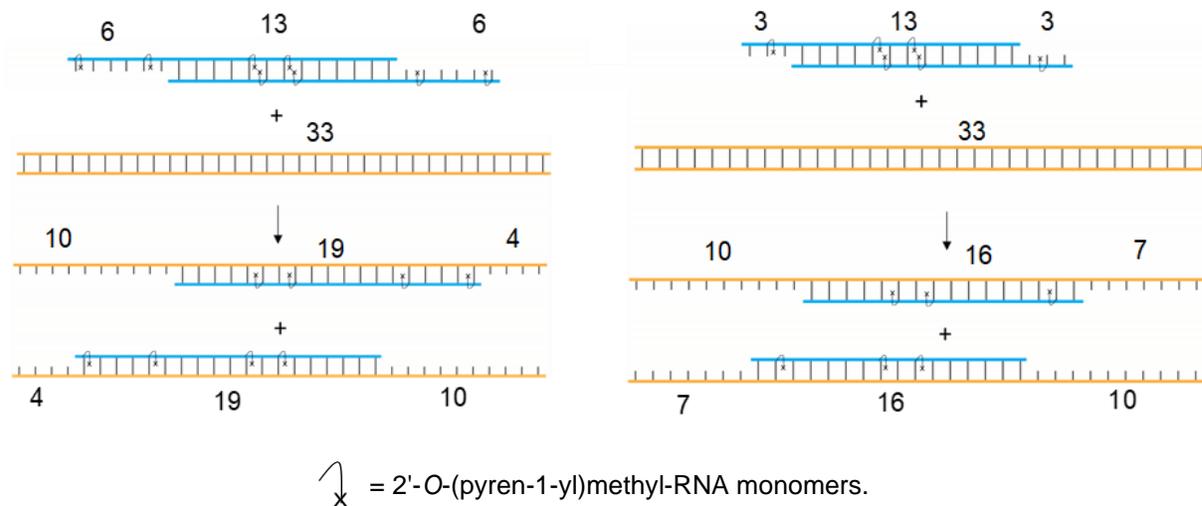
Net loss of base pairs:

13 + 33 -19 – 19 = 8 bps

Net loss of base pairs:

13 + 33 -16 – 16 = 14 bps

Figure 3.35. In-depth view on dsDNA-recognition using toehold probes.

(a)

(b)
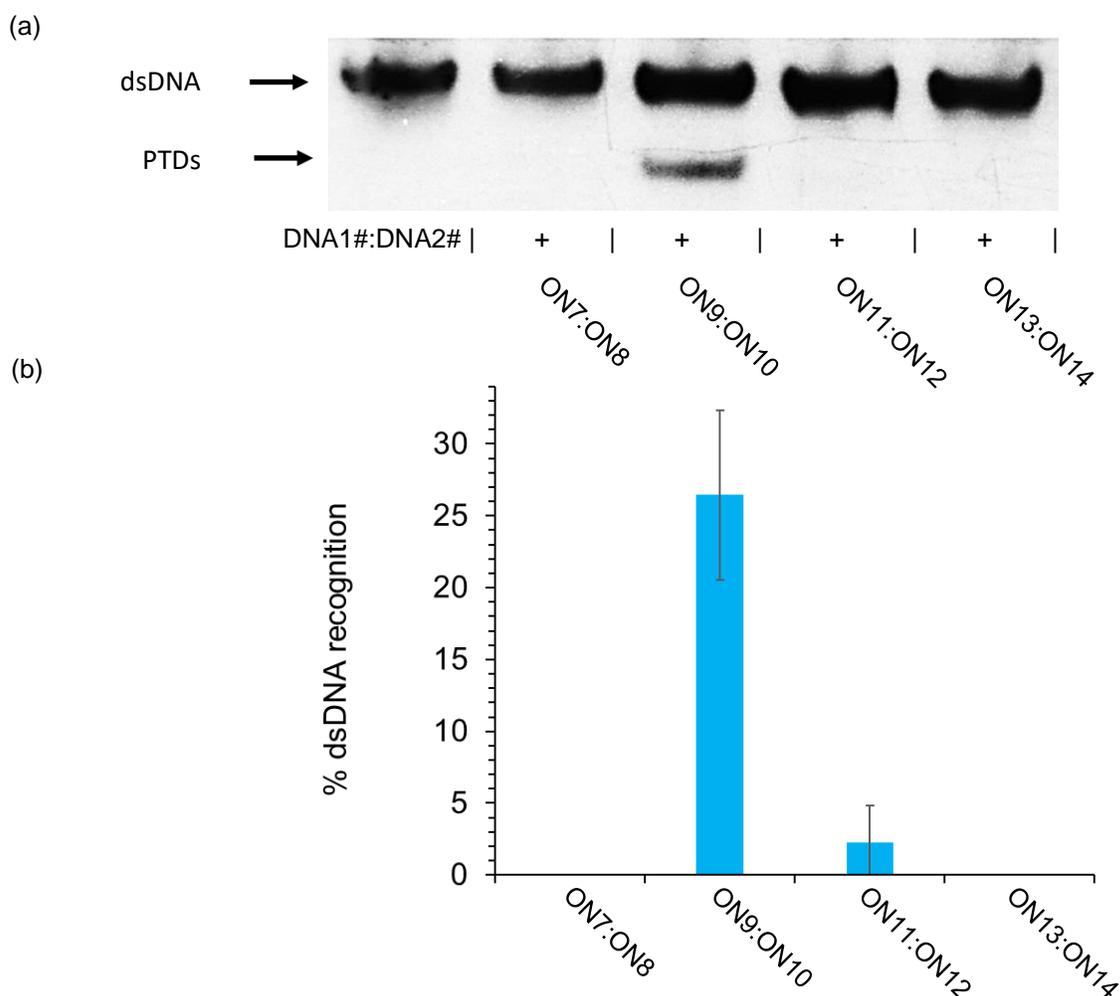
Figure 3.36. (a) Representative gel electrophoretograms from recognition experiments in which 39-mer dsDNA target **DNA1#:DNA2#** was incubated with a 5-fold molar excess of different probes. **DNA1#** = 5'-TCA AAG CTG CAC AGG TAT ATA TAG GCC GCA TAT GCA AGT and **DNA2#** = 3'-AGT TTC GAC GTG TCC ATA TAT ATC CGG CGT ATA CGT TCA. PTD = probe-target duplex. (b) Histogram depicting averaged results from at least three independent recognition experiments with error bars representing standard deviation. Conditions: 3'-DIG-labeled **DNA1#:DNA2#** (50 nM) was incubated with a 5-fold molar excess of the specified probe in HEPES buffer (50 mM HEPES, 100 mM NaCl, 5 mM $MgCl_2$, pH 7.2, 10% sucrose, 1.44 mM spermine tetrahydrochloride) for 17 h at 37 °C.
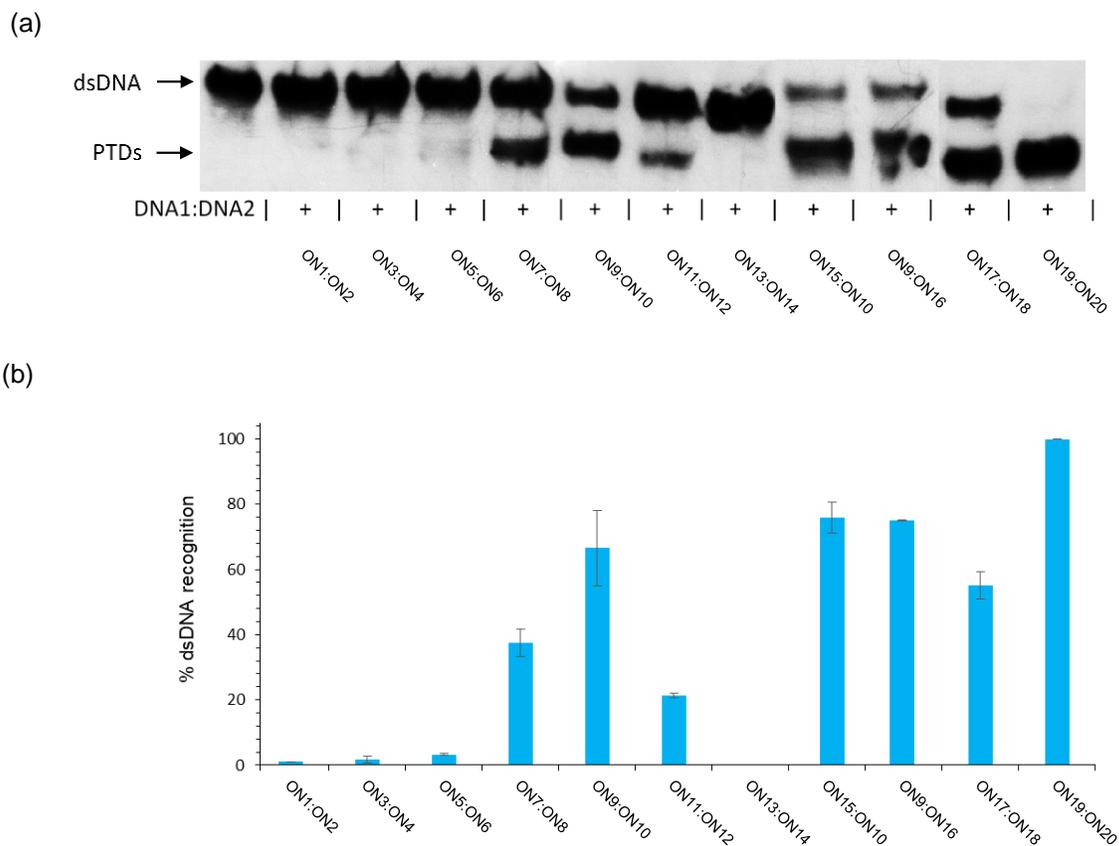
Figure 3.37. (a) Representative gel electrophoretograms from experiments in which **DNA1**:**DNA2** was incubated with different Invader probes under heat-shock conditions. PTD = probe-target duplexes. (b) Histogram depicting averaged results from at least three independent recognition experiments with error bars representing standard deviation. DIG-labeled 33-mer target **DNA1**:**DNA2** (50 nM) was mixed with a 5-fold molar excess of different double-stranded probes in HEPES buffer (100 mM NaCl, 5 mM MgCl$_2$, 10% sucrose, 1.4 mM spermine tetrahydrochloride, pH 7.2), briefly heated (3 min, 90 °C), then cooled to 37 °C and incubated at 37 °C for 17 h. Tabulated results are shown in Table 3.7.

Table 3.7. Quantification of **DNA1:DNA2**-recognition using a 5-fold molar excess of various double-stranded probes under heat-shock conditions.[a]

Probe	Sequence	% Recognition
1:2	5'-GGTA<u>U</u>A<u>U</u>ATAGGC-3' 3'-CCATA<u>U</u>A<u>U</u>ATCCG-5'	<5
3:4	5'-ACA-GGTA<u>U</u>A<u>U</u>ATAGGC-3' 3'-CCATA<u>U</u>A<u>U</u>ATCCG-GCG-5'	<5
5:6	5'-A<u>C</u>A-GGTA<u>U</u>A<u>U</u>ATAGGC-3' 3'-CCATA<u>U</u>A<u>U</u>ATCCG-G<u>C</u>G-5'	<5
7:8	5'-TGCACA-GGTA<u>U</u>A<u>U</u>ATAGGC-3' 3'-CCATA<u>U</u>A<u>U</u>ATCCG-GCGTAT-5'	38 ± 4
9:10	5'-<u>U</u>GCA<u>C</u>A-GGTA<u>U</u>A<u>U</u>ATAGGC-3' 3'-CCATA<u>U</u>A<u>U</u>ATCCG-G<u>C</u>GTA<u>U</u>-5'	67 ± 4
11:12	5'-<u>U</u>GCA<u>C</u>A-GGTATATATAGGC-3' 3'-CCATATATATCCG-G<u>C</u>GTA<u>U</u>-5'	22± 1
13:14	5'-TGCACA-GGTATATATAGGC-3' 3'-CCATATATATCCG-GCGTAT-5'	<5
15:10	5'-GGTA<u>U</u>A<u>U</u>ATAGGC-<u>C</u>GCA<u>U</u>A-3' 3'- CCATA<u>U</u>A<u>U</u>ATCCG-G<u>C</u>GTA<u>U</u>-5'	76± 5
9:16	5'-<u>U</u>GCA<u>C</u>A-GGTA<u>U</u>A<u>U</u>ATAGGC-3' 3'-A<u>C</u>GTG<u>U</u>-CCATA<u>U</u>A<u>U</u>ATCCG-5'	75± 1
17:18	5'-**t**GCA**c**A-GGTA<u>U</u>A<u>U</u>ATAGGC-3' 3'-CCATA<u>U</u>A<u>U</u>ATCCG-G**c**GTA**t**-5'	55 ± 4
19:20	5'-**t**G**c**AC**a**-GG<u>U</u>A<u>U</u>ATA<u>U</u>AGGC-3' 3'-CCA<u>U</u>A<u>U</u>ATA<u>U</u>CCG-**g**C**G**t**A**t-5'	100

[a]Experiments were performed in triplicate. Conditions are described in Fig. 3.37.
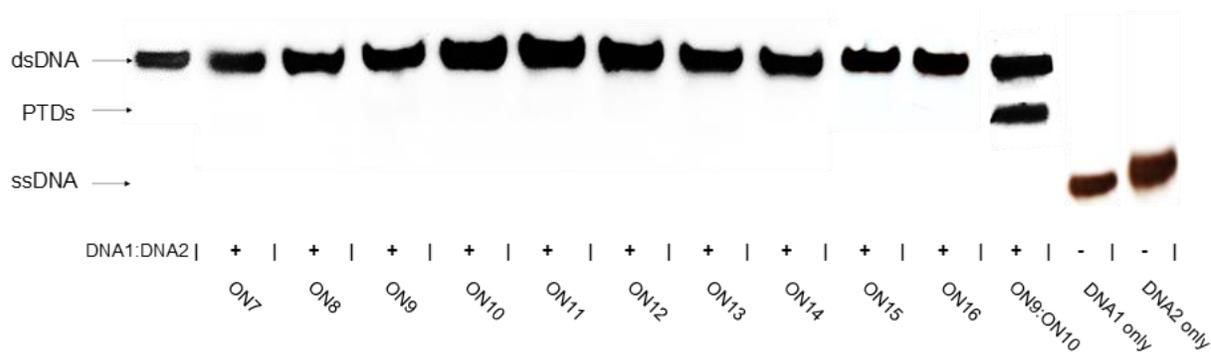
Figure 3.38. Representative gel electrophoretograms from recognition experiments in which **DNA1**:**DNA2** (dsDNA, 50 nM) was incubated with a 5-fold molar excess of individual probe strands at 37 °C. Experimental conditions are as described in Fig. 3.2. PTD = probe-target duplex.
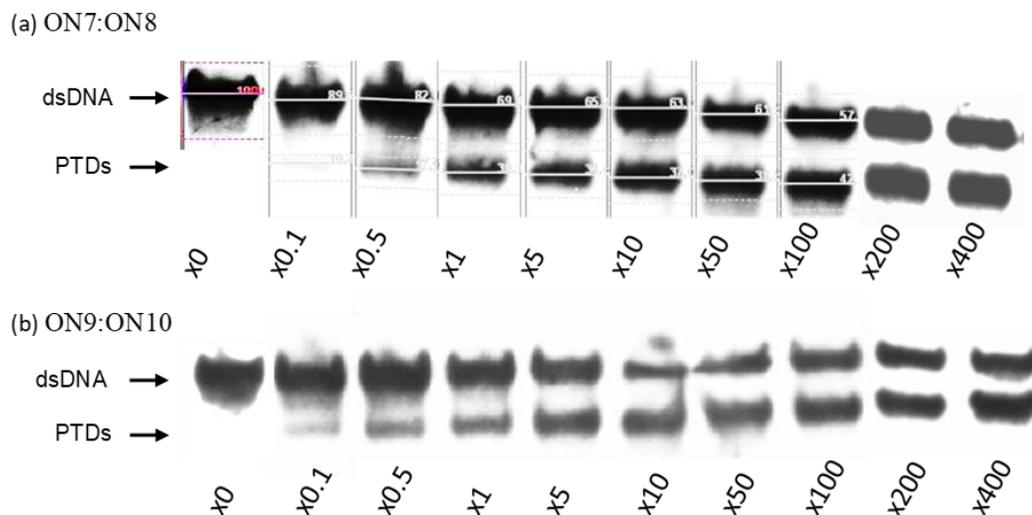
Figure 3.39. Representative electrophoretograms from dose-response experiments in which **DNA1**:**DNA2** (50 nM) was incubated with a variable excess of a) **ON7**:**ON8** or b) **ON9**:**ON10** at 37 °C. Experimental conditions are as described in Fig. 3.2. PTD = probe-target duplex.
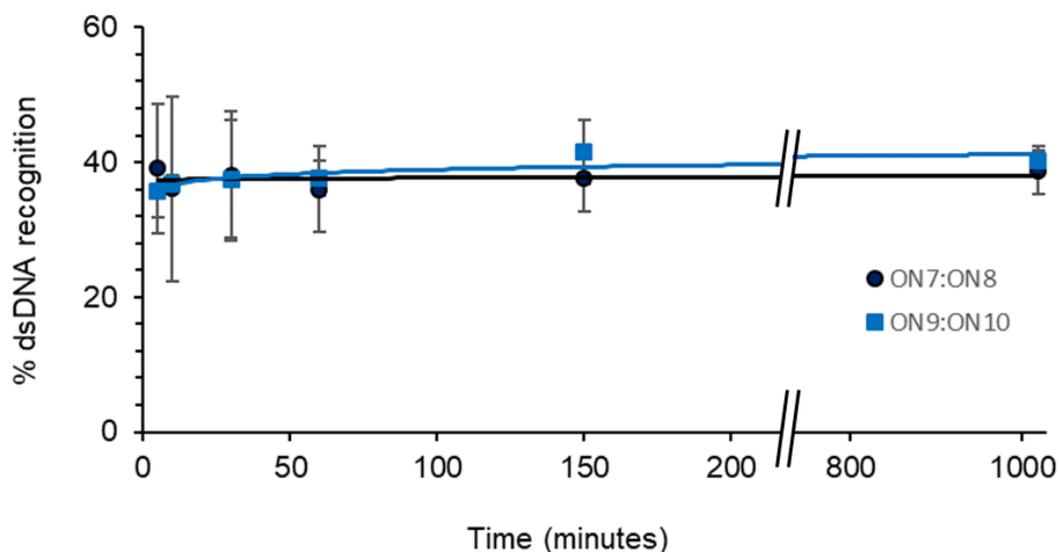
Figure 3.40. Time-course experiments in which **DNA1**:**DNA2** (50 nM) was incubated with a 5-fold excess of toehold Invader probes **ON7**:**ON8** or **ON9**:**ON10** at 37 °C in HEPES buffer (100 mM NaCl, 5 mM $MgCl_2$, 10 % sucrose, 1.44 mM spermine tetrahydrochloride, pH 7.2).

Aliquots were taken at the specified time-points, flash-frozen and stored in liquid N_2 until electrophoresis was performed as described in Fig. 3.2.
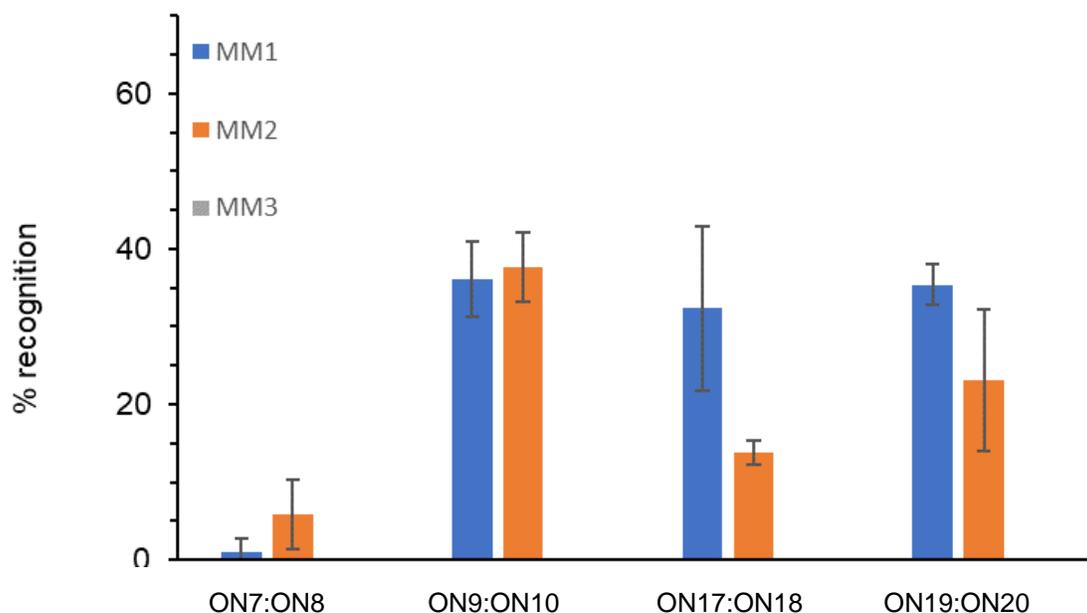
Figure 3.41. Histogram depicting averaged results from at least three experiments in which pre-annealed 3'-DIG-labelled **MM1**-**MM3** were incubated with a 5-fold molar excess of various pre-annealed Invader probes at 37 °C for 17 h in HEPES buffer as outlined in Fig. 3.2. Error bars represent standard deviation. For representative electrophoretograms, see Figs. 3.4 and 3.5. Sequences of **MM1**-**MM3** are shown in Table 3.8.

Table 3.8. Thermal denaturation temperatures (T_ms) of **MM1**-**MM3**, as well as, duplexes between individual probe strands **ON7** or **ON8** and the corresponding mismatched DNA strands. Also shown are *TA* values for recognition of **MM1**-**MM3** by **ON7**:**ON8**.[a]

DNA	Sequence	T_m (°C)			
		dsDNA	5'-DNA vs ON8	3'-DNA vs ON7	*TA* (°C)
DNA1	5'-AAGCTG CAC AGG TAT ATA TAG GCC GCA TATGCA-3'	72.0	67.0	60.0	11.0
DNA2	3'-TTCGAC GTG TCC ATA TAT ATC CGG CGT ATACGT-5'				
DNA3	5'-AAGCTG CAC AGG TAT TTA TAG GCC GCA TATGCA-3'	74.0	64.5	57.5	4.0
DNA4	3'-TTCGAC GTG TCC ATA AAT ATC CGG CGT ATACGT-5'				
DNA5	5'-AAGCTG GAC AGG TAT ATA TAG GCC GCT TATGCA-3'	72.0	66.0	58.0	8.0
DNA6	3'-TTCGAC CTG TCC ATA TAT ATC CGG CGA ATACGT-5'				
DNA7	5'-AAGCTG GAC AGG TAT TTA TAG GCC GCT TATGCA-3'	72.5	62.5	49.0	-5.0
DNA8	3'-TTCGAC CTG TCC ATA AAT ATC CGG CGA ATACGT-5'				

[a] Position of mismatched base-pairs relative to **ON7**:**ON8** highlighted in yellow. *TA* is calculated for **ON7**:**ON8** (T_m = 44.0 °C). Experimental conditions are as described in Table 3.1. For a definition of *TA*, see the main manuscript. **MM1** = **DNA3**:**DNA4**, **MM2** = **DNA5**:**DNA6**, and **MM3** = **DNA7**:**DNA8**. For the corresponding thermal denaturation curves, see Fig 3.31.

Table 3.9. Thermal denaturation temperatures (T_ms) of **MM1**-**MM3**, as well as, duplexes between individual probe strands **ON9** or **ON10** and the corresponding mismatched DNA strands. Also shown are *TA* values for recognition of **MM1**-**MM3** by **ON9**:**ON10**.[a]

DNA	Sequence	T_m (°C)			
		dsDNA	5'-DNA vs ON10	3'-DNA vs ON9	*TA* (°C)
DNA1	5'-AAGCTG CAC AGG TAT ATA TAG GCC GCA TATGCA-3'	72.0	75.5	75.5	30.0
DNA2	3'-TTCGAC GTG TCC ATA TAT ATC CGG CGT ATACGT-5'				
DNA3	5'-AAGCTG CAC AGG TAT TTA TAG GCC GCA TATGCA-3'	74.0	72.0	67.5	16.5
DNA4	3'-TTCGAC GTG TCC ATA AAT ATC CGG CGT ATACGT-5'				
DNA5	5'-AAGCTG GAC AGG TAT ATA TAG GCC GCT TATGCA-3'	72.0	72.0	68.0	19.0
DNA6	3'-TTCGAC CTG TCC ATA TAT ATC CGG CGA ATACGT-5'				
DNA7	5'-AAGCTG GAC AGG TAT TTA TAG GCC GCT TATGCA-3'	72.5	72.0	59.0	9.5
DNA8	3'-TTCGAC CTG TCC ATA AAT ATC CGG CGA ATACGT-5'				

[a] Position of mismatched base-pairs relative to probe highlighted in yellow. *TA* is calculated for **ON9**:**ON10** (T_m = 49.0 °C). Experimental conditions are as described in Table 3.1. For a definition of *TA*, see the main manuscript. **MM1** = **DNA3**:**DNA4**, **MM2** = **DNA5**:**DNA6**, and **MM3** = **DNA7**:**DNA8**. For the corresponding thermal denaturation curves, see Fig 3.31.

Additional discussion of enthalpic and entropic parameters observed for **ON17:ON18**. The favorable ΔG_{rec}^{310} value for recognition of **DNA1:DNA2** using probe **ON17:ON18** (Table 3.2) is a consequence of a highly favorable change in enthalpy (ΔH_{rec} = -342 kJ/mol), (Table 3.10), which only is partially offset by an unfavorable change in entropy ($-T\Delta S_{rec}^{310}$ +279 kJ/mol, Table 3.11).

Table 3.10. Change in enthalpy (ΔH) upon formation of probe duplexes and duplexes between individual probe strands and **DNA1** or **DNA2**. Also shown is the calculated change in reaction enthalpy upon probe-mediated recognition of 33-mer target **DNA1:DNA2** (ΔH_{rec}).[a]

Probe	Sequence	ΔH [$\Delta\Delta H$] (kJ/mol)			
		probe duplex	5'-probe vs DNA2	3'-probe vs DNA1	ΔH_{rec} (kJ/mol)
ON17	5'-**t**GCA**c**A-GGTA<u>UA</u><u>U</u>ATAGGC	-394	-608	-646	-342
ON18	3'-CCATA<u>UA</u><u>U</u>ATCCG-G**c**GTA**t**	[+124]	[-90]	[-128]	
ON19	5'-**t**G**c**AC**a**-GG<u>UA</u><u>U</u>ATA<u>U</u>AGGC	_[b]	-743	-524	nd
ON20	3'-CCA<u>UA</u><u>U</u>ATA<u>U</u>CCG-**g**CG**t**A**t**		[-225]	[-6]	

[a] $\Delta\Delta H$ is calculated relative to the unmodified 33-mer target DNA duplex **DNA1:DNA2** (ΔH = -518 kJ/mol). ΔH_{rec} = ΔH (5'-probe:**DNA2**) + ΔH (3'-probe:**DNA1**) - ΔH (probe duplex) - ΔH (**DNA1:DNA2**). For experimental conditions, see Table 3.1.

Table 3.11. Change in entropy at 310 K ($-T\Delta S^{310}$) upon formation of probe duplexes and duplexes between individual probe strands and **DNA1** or **DNA2**. Also shown is the calculated change in reaction entropy upon probe-mediated recognition of 33-mer target **DNA1:DNA2** ($-T\Delta S_{rec}^{310}$).[a]

Probe	Sequence	$-T\Delta S^{310}$ [$\Delta(T\Delta S^{310})$] (kJ/mol)			
		probe duplex	5'-probe vs DNA2	3'-probe vs DNA1	$-T\Delta S_{rec}^{310}$ (kJ/mol)
ON17	5'-**t**GCA**c**A-GGTA<u>UA</u><u>U</u>ATAGGC	+339	+504	+538	+279
ON18	3'-CCATA<u>UA</u><u>U</u>ATCCG-G**c**GTA**t**	[-85]	[+80]	[+114]	
ON19	5'-**t**G**c**AC**a**-GG<u>UA</u><u>U</u>ATA<u>U</u>AGGC	_[b]	+620	+420	nd
ON20	3'-CCA<u>UA</u><u>U</u>ATA<u>U</u>CCG-**g**CG**t**A**t**		[+196]	[-4]	

[a] $\Delta(T\Delta S^{310})$ is calculated relative to the unmodified 33-mer target DNA duplex **DNA1:DNA2** ($-T\Delta S^{310}$ = 424 kJ/mol). $-T\Delta S_{rec}^{310}$ = $\Delta(T\Delta S^{310})$ (5'-probe:**DNA2**) + $\Delta(T\Delta S^{310})$ (3'-probe:**DNA1**) - $\Delta(T\Delta S^{310})$ (probe duplex). For experimental conditions, see Table 3.1.
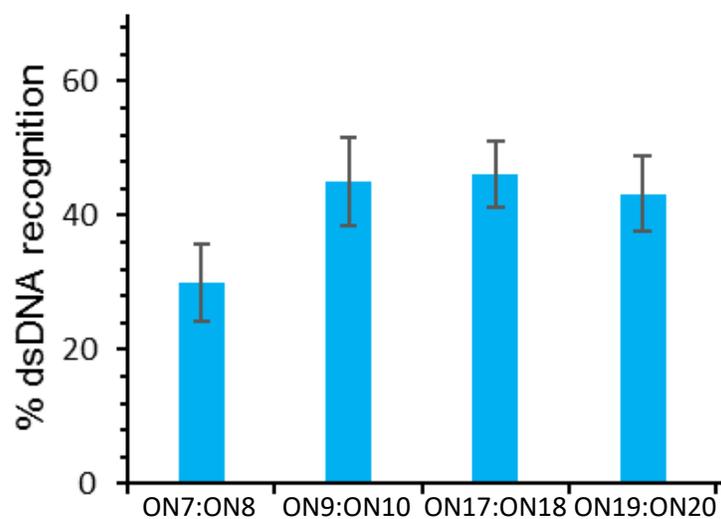
Figure 3.42. Histogram depicting the average result of at least three independent recognition experiments using model dsDNA target **DNA1**:**DNA2** (50 nM) and a 5-fold molar excess of various Invader probes. Error bars represent standard deviation. Experimental conditions are as outlined in Fig. 3.2.
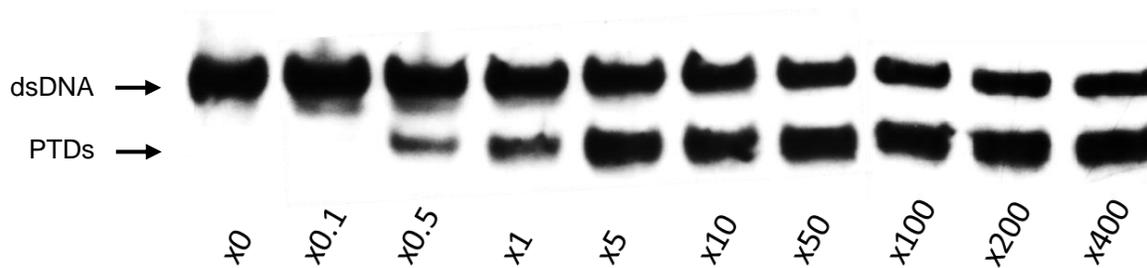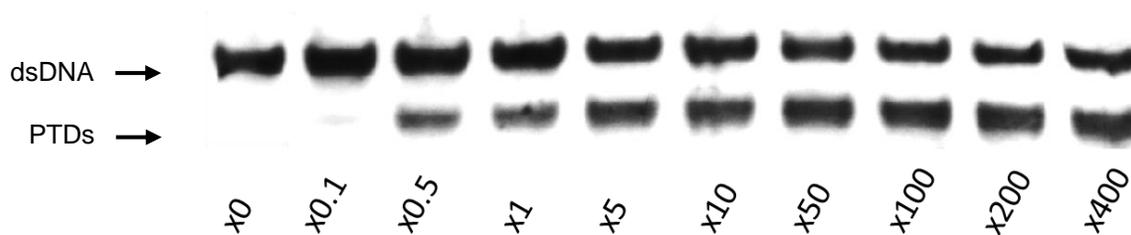
Figure 3.43. Representative electrophoretograms from experiments in which **DNA1:DNA2** (50 nM) was incubated with a variable excess of LNA-modified toehold Invader probes: a) **ON17:ON18** and b) **ON19:ON20**. Experimental conditions are as described in Fig. 3.2.
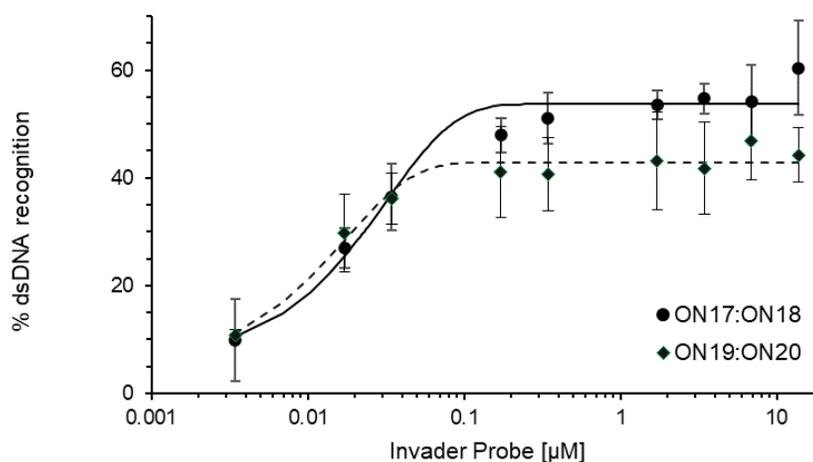
Figure 3.44. Dose-response curves for recognition of **DNA1**:**DNA2** by LNA-modified Invader probes **ON17**:**ON18** or **ON19**:**ON20**. Curves are constructed based on the electrophoretograms shown in Fig. 3.43. Experimental conditions are as in Fig. 3.2.
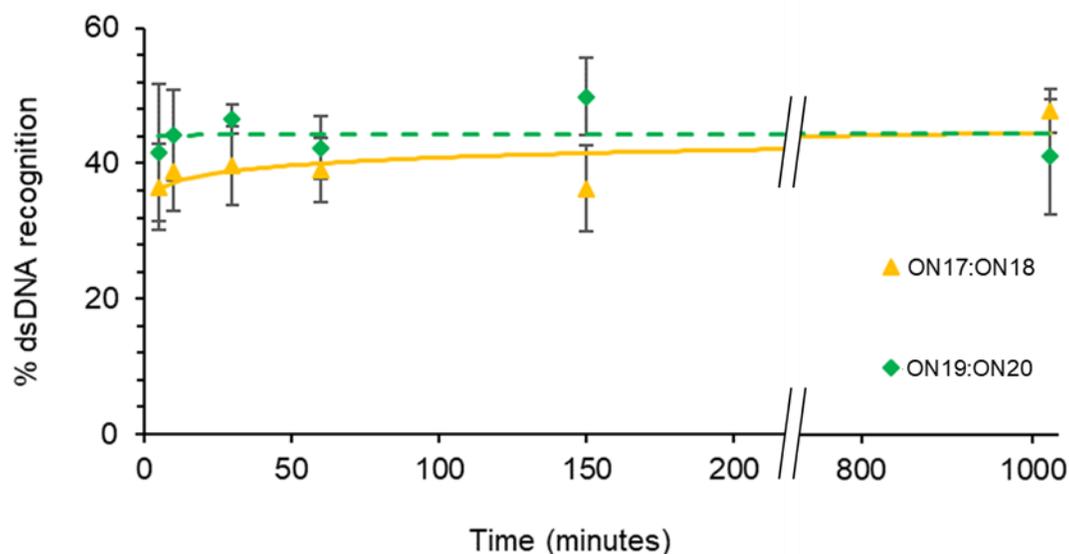
Figure 3.45. Time-course for recognition of **DNA1**:**DNA2** using LNA-modified toehold Invader probes. **DNA1**:**DNA2** (50 nM) was incubated with 5-fold molar excess of **ON17**:**ON18** or **ON19**:**ON20** at 37 °C in HEPES buffer (100 mM NaCl, 5 mM MgCl$_2$, 10 % sucrose, 1.44 mM spermine tetrahydrochloride, pH 7.2). Aliquots were taken at the specified time-points, flash-frozen, and stored in liquid N$_2$ until being resolved by gel electrophoresis as described in Fig. 3.2.
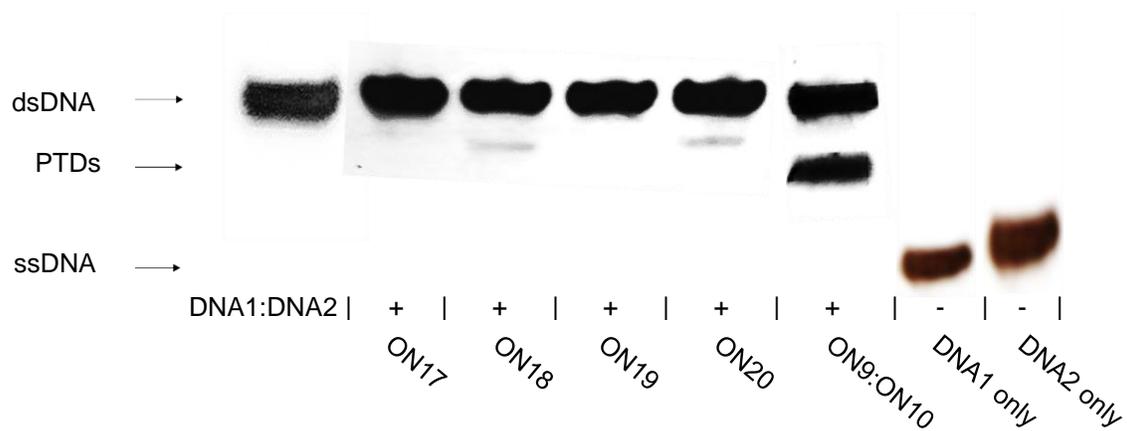
Figure 3.46. Representative gel eletrophoretograms from recognition experiments in which a 5-fold molar excess of single-stranded probes **ON17**-**ON20** was incubated with **DNA1:DNA2**. Experimental conditions are as described in Fig. 3.2. PTD = probe-target duplex.

Table 3.12. Thermal denaturation temperatures (T_ms) of **MM1**-**MM3** as well as duplexes between individual probe strands **ON17** or **ON18** and the corresponding mismatched DNA strands. Also shown are *TA* values for recognition of **MM1**-**MM3** by **ON17**:**ON18**.[a]

DNA	Sequence	T_m (°C)			
		dsDNA	5'-DNA vs ON18	3'-DNA vs ON17	*TA* (°C)
DNA1	5'-AAG CTG CAC AGG TAT ATA TAG GCC GCA TAT GCA -3'	72.0	73.0	71.5	25.0
DNA2	3'-TTC GAC GTG TCC ATA TAT ATC CGG CGT ATA CGT -5'				
DNA3	5'-AAG CTG CAC AGG TAT TTA TAG GCC GCA TAT GCA -3'	74.0	70.0	66.0	14.5
DNA4	3'-TTC GAC GTG TCC ATA AAT ATC CGG CGT ATA CGT -5'				
DNA5	5'-AAG CTG GAC AGG TAT ATA TAG GCC GCT TAT GCA -3'	72.0	71.5	65.0	17.0
DNA6	3'-TTC GAC CTG TCC ATA TAT ATC CGG CGA ATA CGT -5'				
DNA7	5'-AAG CTG GAC AGG TAT TTA TAG GCC GCT TAT GCA -3'	72.5	68.5	56.0	4.5
DNA8	3'-TTC GAC CTG TCC ATA AAT ATC CGG CGA ATA CGT -5'				

[a] Position of mismatched base-pairs relative to probe highlighted in yellow. *TA* values are calculated relative to **ON17**:**ON18** (T_m = 47.5 °C). Experimental conditions are as described in Table 3.1. For a definition of *TA*, see the main manuscript. **MM1** = **DNA3**:**DNA4**, **MM2** = **DNA5**:**DNA6**, and **MM3** = **DNA7**:**DNA8**.

Table 3.13. Thermal denaturation temperatures (T_ms) of **MM1**-**MM3** as well as duplexes between individual probe strands **ON19** or **ON20** and the corresponding mismatched DNA strands.[a]

DNA	Sequence	T_m (°C)		
		dsDNA	5'-DNA vs ON20	3'-DNA vs ON19
DNA1	5'-AAG CTG CAC AGG TAT ATA TAG GCC GCA TAT GCA -3'	72.0	79.5	78.0
DNA2	3'-TTC GAC GTG TCC ATA TAT ATC CGG CGT ATA CGT -5'			
DNA3	5'-AAG CTG CAC AGG TAT TTA TAG GCC GCA TAT GCA -3'	74.0	76.5	71.0
DNA4	3'-TTC GAC GTG TCC ATA AAT ATC CGG CGT ATA CGT -5'			
DNA5	5'-AAG CTG GAC AGG TAT ATA TAG GCC GCT TAT GCA -3'	72.0	76.5	68.5
DNA6	3'-TTC GAC CTG TCC ATA TAT ATC CGG CGA ATA CGT -5'			
DNA7	5'-AAG CTG GAC AGG TAT TTA TAG GCC GCT TAT GCA -3'	72.5	73.5	59.5
DNA8	3'-TTC GAC CTG TCC ATA AAT ATC CGG CGA ATA CGT -5'			

[a] Position of mismatched base-pairs relative to probe highlighted in yellow. Experimental conditions are as described in Table 3.1. For a definition of *TA*, see the main manuscript. *TA* values could not be determined calculated since **ON19**:**ON20** did not display a transition. **MM1** = **DNA3**:**DNA4**, **MM2** = **DNA5**:**DNA6**, and **MM3** = **DNA7**:**DNA8**.

Table 3.14. Sequence and T_m of DNA hairpin **DH1** used in this study.[a]

DNA target	sequence	T_m (°C)
DH1	5'- AAG CTG CAC AGG TAT ATA TAG GCC GCA TAT GCA $(T)_{10}$ 3'- TTC GAC GTG TCC ATA TAT ATC CGG CGT ATA CGT	81

[a] Experimental conditions are as stated in Table 3.1. Representative thermal denaturation curves are shown in Fig. 3.47.
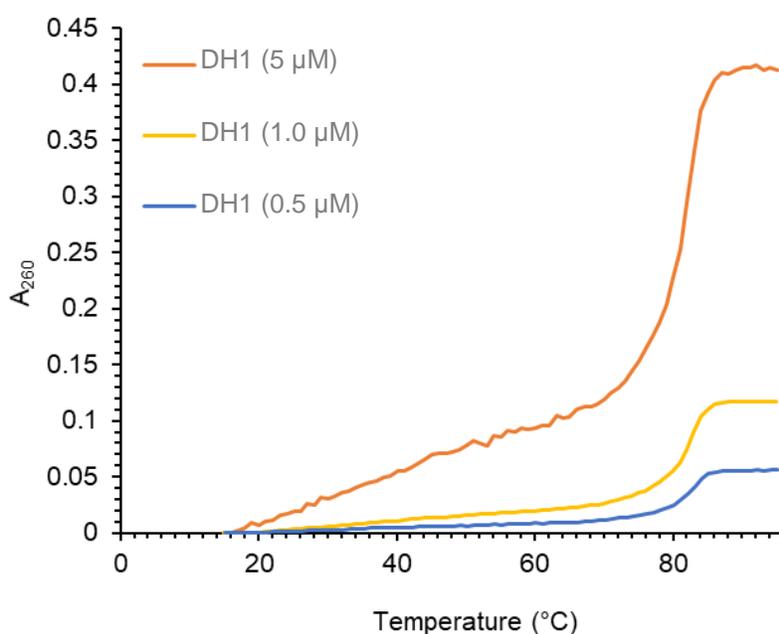
Figure 3.47. Representative thermal denaturation profiles of 33-mer dsDNA hairpin **DH1** recorded in medium salt buffer at strand concentrations of 0.5 μM, 1.0 μM and 5 μM. The observation of near-identical concentration-independent T_m values suggests that an intramolecular unimolecular hairpin structure is formed. Experimental conditions are as described in Table 3.1.

(a)   ON9:ON10

(b)   ON17:ON18
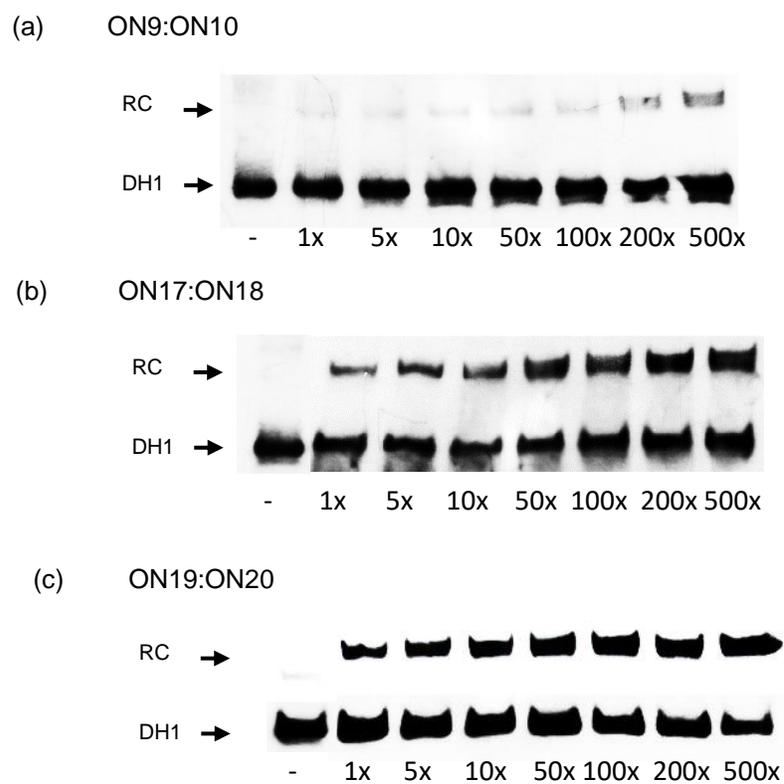
(c)   ON19:ON20

Figure 3.48. Representative electrophoretograms from recognition experiments in which **DH1** (50 nM) was incubated with a variable excess of a) **ON9:ON10**, b) **ON17:ON18** or c) **ON19:ON20.** Experimental conditions are as described in Fig. 3.2.

Table 3.15. Sequences of probes used in FISH experiments, as well as T_ms of probe duplexes and duplexes between individual probe strands and DNA targets.[a]

Probe	Sequence	T_m [ΔT_m] (°C)			
		probe duplex	5'-probe vs DNA10	3'-probe vs DNA9	T_A (°C)
DYZ-OPTu: DYZ-OPTd	5'-Cy3-TgTgTG-T<u>UA</u><u>U</u>ATGCTG<u>U</u>TCTC-3' 3'-AA<u>UA</u><u>U</u>ACGACA<u>**A**</u>GAG-TCgGgA-Cy3-5'	nt	66.0 [0.0]	69.0 [+3.0]	nd
DYZ-REFu: DYZ-REFd	5'-Cy3-T<u>UA</u> <u>U</u>AT GCT G<u>U</u>T CTC-3' 3'- AA<u>U</u> A<u>U</u>A CGA CA<u>**A**</u> GAG-Cy3-5'	56.0	61.0	68.0	+7.0

[a] ΔT_m is calculated relative to the unmodified **DNA9**:**DNA10** duplex (T_m = 66.0 °C). For the sequence of **DNA9**:**DNA10**, see Table 3.16. Conditions are as described in Table 3.1.
nd = not determinable; nt = no transition

Table 3.16. Thermal denaturation temperatures (T_ms) of complementary **DNA9**:**DNA10** and non-complementary target **DNA11**:**DNA12**, as well as duplexes between individual DNA strands and individual probe strands.[a]

DNA	Sequence	T_m (°C)		
		dsDNA	5'-DNA vs DYZ-OPTd	3'-DNA vs DYZ-OPTu
DNA9: DNA10	5'-TGAC *TGT GTG TTA TAT GCT GTT CTC AGC CCT* TGAC 3'-ACTG *ACA CAC AAT ATA CGA CAA GAG TCG GGA* ACTG	66.0	69.0	66.0
DNA11: DNA12	5'-TGAC *TGT GTC TTA TAT GGT GTT CTC TGC CCT* TGAC 3'-ACTG *ACA CAG AAT ATA CCA CAA GAG ACG GGA* ACTG	66.0	56.0	51.0

[a] Conditions are as described in Table 3.1. Here, the 27-mer mixed-sequence target region of **DNA9:DNA10** is italicized, extended by 4 bps in each side.

(a)

(b)
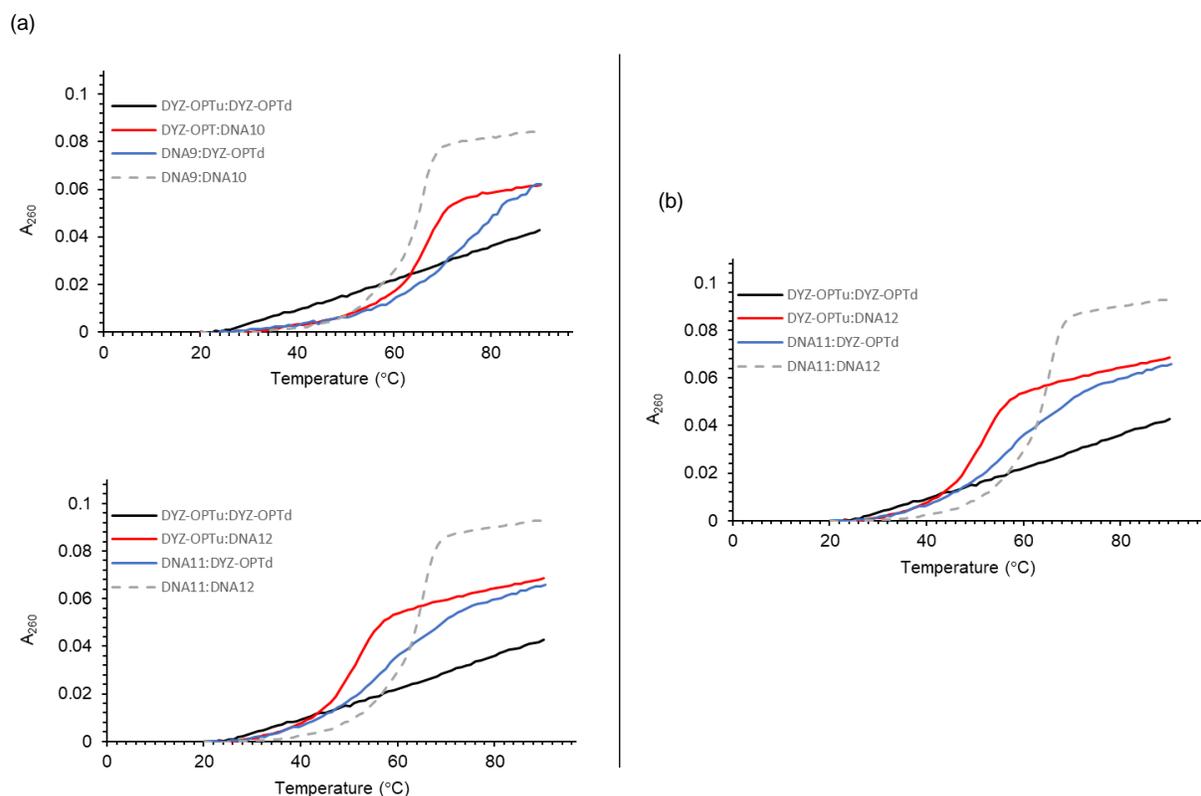
Figure 3.49. **(a)** Representative thermal denaturation curves for toehold Invader probe **DYZ-OPTu**:**DYZ-OPTd**, duplexes between individual probe strands and single-stranded complementary/non-complementary DNA targets, as well as **DNA9**:**DNA10** and **DNA11**:**DNA12** (the latter of which differs in sequence at three positions relative to **DNA9**:**DNA10**). (b) Representative thermal denaturation curves for toehold Invader probe **DYZ-REFu**:**DYZ-REFd**, duplexes between individual probe strands and single-stranded complementary DNA targets, **DNA9:DNA10.** For sequences see Table 3.16. Conditions are as outlined in Table 3.1.

Table 3.17. Illustration of matched or mismatched probe-target duplexes that would ensue upon recognition of complementary target **DNA9**:**DAN10** or triply-mismatched target **DNA11**:**DNA12** using **DYZ-OPT**. Positions highlighted in yellow denote the position of mismatched base-pairs.

DNA	Sequence
DNA9:	5'-TGAC *TGT GTG TTA TAT GCT GTT CTC AGC CCT* TGAC
DNA10	3'-ACTG *ACA CAC AAT ATA CGA CAA GAG TCG GGA* ACTG
DNA11:	5'-TGAC *TGT GTC TTA TAT GGT GTT CTC TGC CCT* TGAC
DNA12	3'-ACTG *ACA CAG AAT ATA CCA CAA GAG ACG GGA* ACTG
DYZ-OPTu:	5'-Cy3-T**g**T **g**TG-T**U**A **U**AT GCT G**U**T CTC
DNA10	3'-ACTG ACA CAC AAT ATA CGA CAA GAG TCG GGA ACTG
DNA9:	5'-TGAC TGT GTG TTA TAT GCT GTT CTC AGC CCT TGAC
DYZ-OPTd	3'-AA**U** A**U**A CGA CA**A** GAG-TC**g** G**g**A-Cy3-5'
DYZ-OPTu:	5'-Cy3-T**g**T **g**TG-T**U**A **U**AT GCT G**U**T CTC
DNA12	3'-ACTG ACA CAG AAT ATA CCA CAA GAG ACG GGA ACTG
DNA11:	5'-TGAC TGT GTC TTA TAT GGT GTT CTC TGC CCT TGAC
DYZ-OPTd	3'-AA**U** A**U**A CGA CA**A** GAG-TC**g** G**g**A-Cy3-5'

dsDNA →

PTDs →

DNA9:DNA10 | +DYZ-OPT   DNA11:DNA12 | +DYZ-OPT
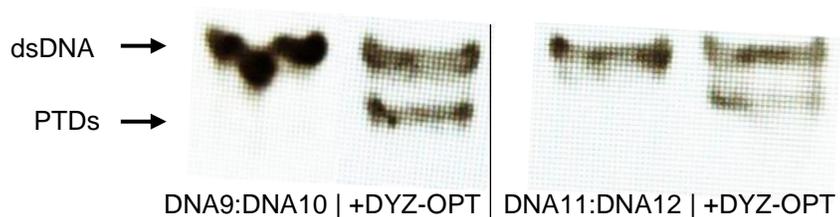
Figure 3.50. Representative electrophoretograms from experiments in which a 5-fold molar excess of **DYZ-OPT** was incubated with complementary **DNA9**:**DNA10** or non-complementary dsDNA target **DNA11**:**DNA12** (see Table 3.17 for sequences of **DNA9**:**DNA10**, and **DNA11**:**DNA12**). Pre-annealed 3'-DIG-labeled targets (50 nM) were incubated with pre-annealed Invader probes at 37 °C for 17 h in HEPES buffer as described in Fig. 3.2.
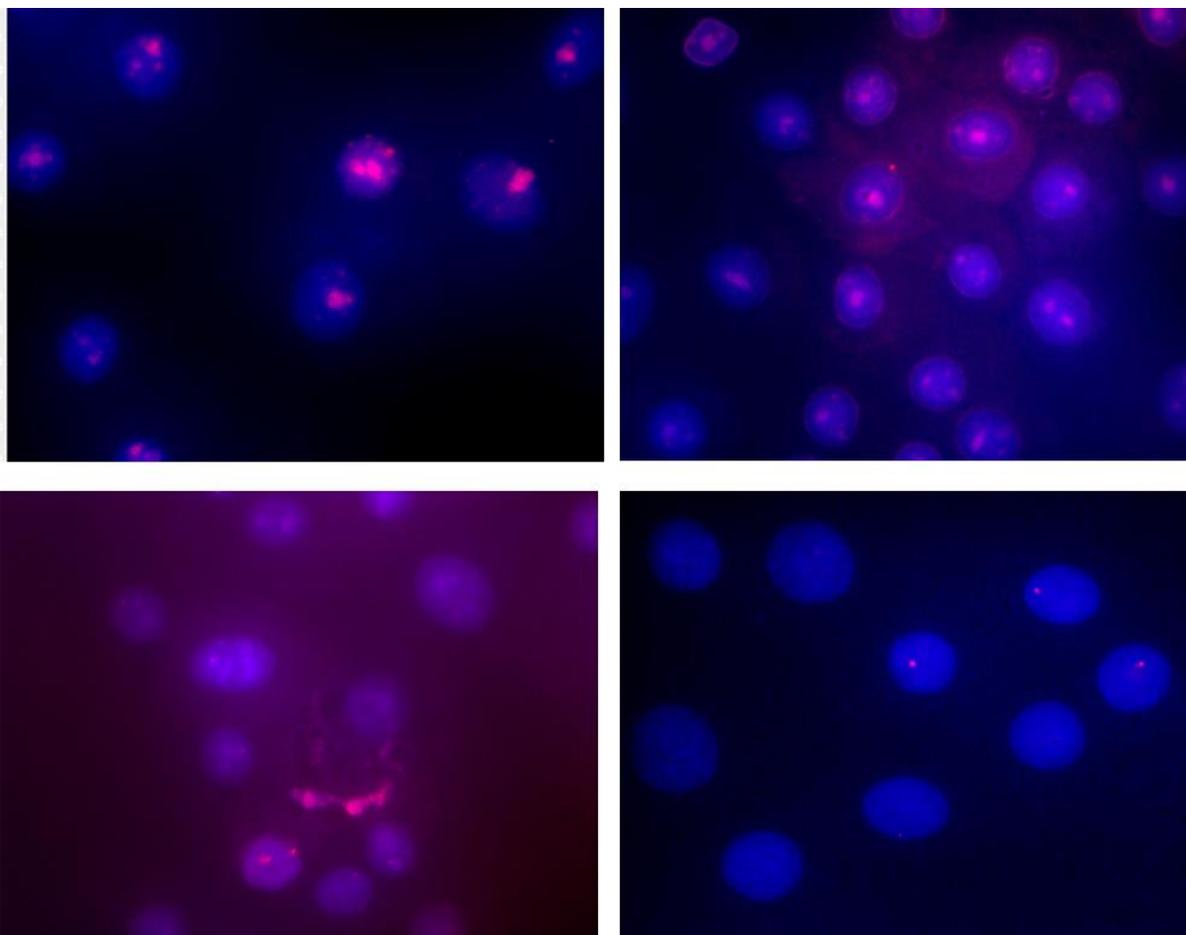
Figure 3.51. Concentration dependence of **DYZ-OPT**-produced FISH signals under non-denaturing conditions. Images from FISH experiments using fixed male bovine kidney cells and varying quantities of LNA-modified toehold Invader probe **DYZ-OPT**, i.e., 30 ng (top left), 15 ng (top right), 6 ng (bottom left), and 3 ng (bottom right) per 200 µl of 1X PCR buffer. Conditions are otherwise as specified in Fig. 3.8. Substantial amounts of background signals were observed when using 6-30 ng **DYZ-OPT**. Accordingly, 3 ng of **DYZ-OPT** per 200 µl of 1X PCR buffer was selected as an appropriate concentration.
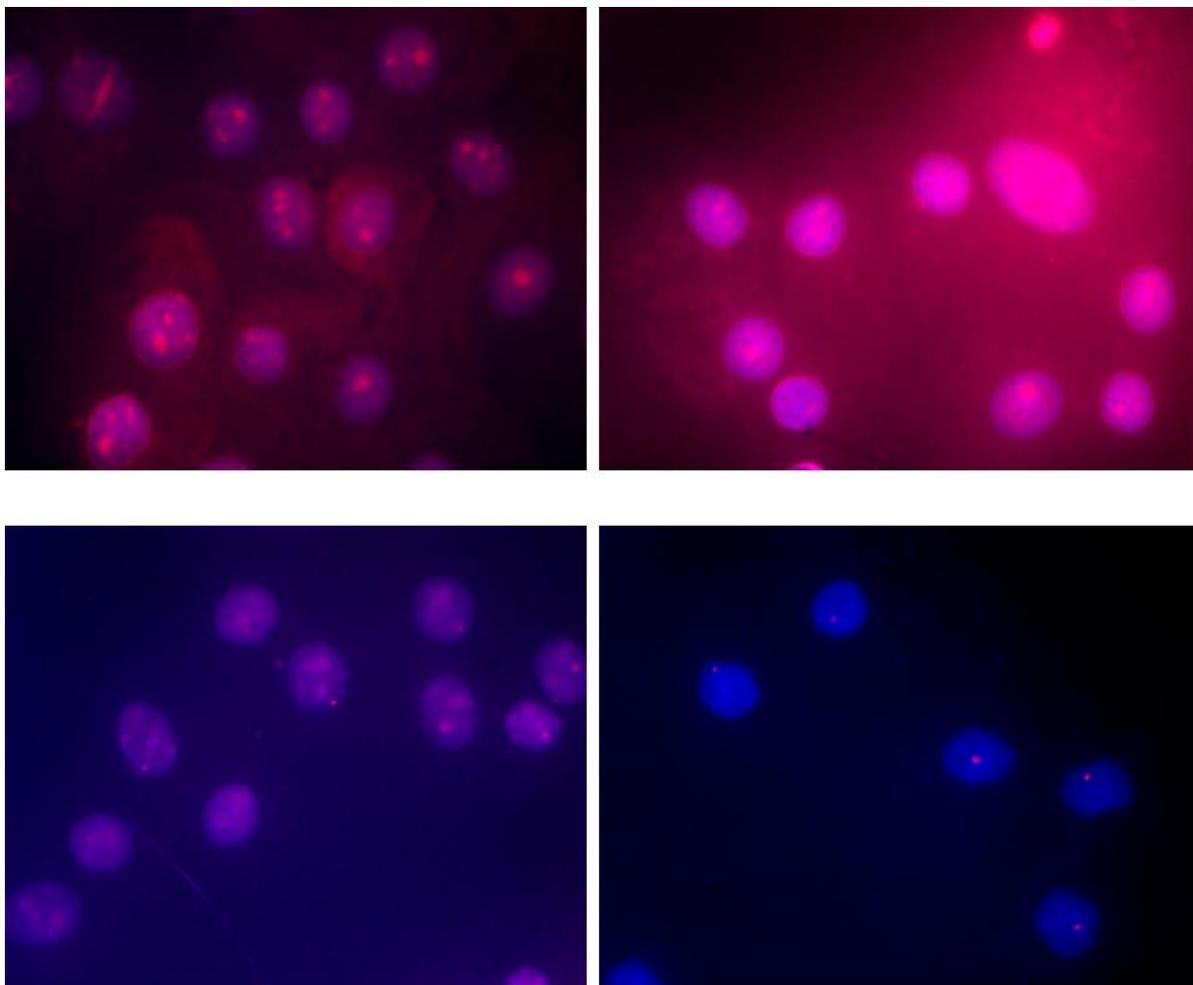
Figure 3.52. Concentration dependence of **DYZ-OPT**-produced FISH signals under denaturing conditions. Images from FISH experiments using fixed male bovine kidney cells and varying amounts of LNA-modified toehold Invader probe **DYZ-OPT**, i.e., 30 ng (top left), 15 ng (top right), 6 ng (bottom left), and 1.5 ng (bottom right) per 200 µl of 1X PCR buffer. Fixed isolated nuclei from male bovine kidney cells were incubated with **DYZ-OPT** for 5 min at 80 °C in a Tris buffer (20 mM Tris-Cl, 100 mM KCl, pH 8.0) and counterstained with DAPI. Images were obtained by overlaying images from Cy3 (red) and DAPI (blue) channels and adjusting the exposure. Nuclei were viewed at 60X magnification using a Nikon Eclipse Ti-S inverted microscope. Substantial levels of background signal was observed when **DYZ-OPT** was used in amounts ≥6 ng per 200 µl of 1X PCR buffer.
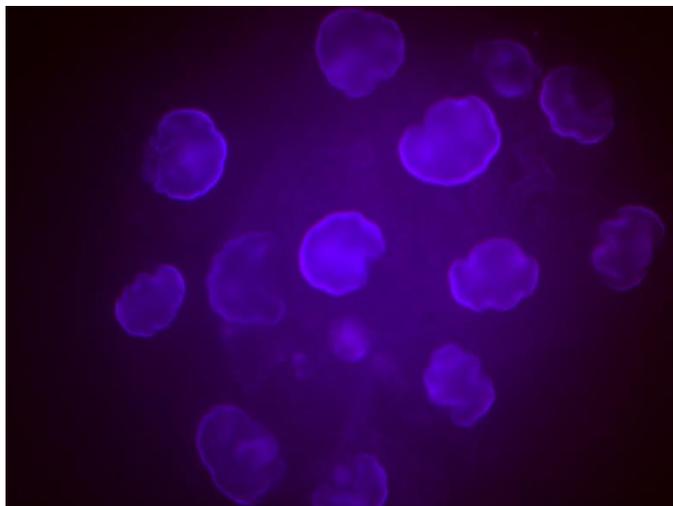
Figure 3.53. Images from FISH experiments in which LNA-modified toehold Invader probe **DYZ-OPT** was incubated with isolated nuclei from a female bovine endothelial cell line (which lacks the *DYZ-1* target region) under non-denaturing using 15 ng of **DYZ-OPT** per 200 µl of incubation buffer. Conditions are as specified in Fig. 3.8.

3.5 References and notes

1 V. K. Sharma, P. Rungta and A. K. Prasad, *RSC Adv.*, 2014, **4**, 16618–16631.

2 A. Khvorova and J. K. Watts, *Nat. Biotechnol.*, 2017, **35**, 238–248.

3 K. E. Lundin, O. Gissberg and C. I. E. Smith, *Hum. Gene Ther.*, 2015, **26**, 475–485.

4 R. L. Juliano, *Nucleic Acids Res.*, 2016, **44**, 6518–6548.

5 J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, et. al, *Science*, 2001, **291**, 1304–1351.

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

7 J. R. Goñi, J. M. Vaquerizas, J. Dopazo and M. Orozco, *BMC Genomics*, 2006, **7**, 63.

8 M. P. Knauert and P. M. Glazer, *Hum. Mol. Genet.*, 2001, **10**, 2243–2251.

9 K. M. Vasquez, J. M. Dagle, D. L. Weeks and P. M. Glazer, *J. Biol. Chem.*, 2001, **276**, 38536–38541.

10 A. Ray and B. Nordén, *FASEB J.*, 2000, **14**, 1041–1060.

11 K. Kaihatsu, B. A. Janowski and D. R. Corey, *Chem. Biol.*, 2004, **11**, 749–758.

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

13 A. Manna, S. Rapireddy, G. Sureshkumar and D. H. Ly, *Tetrahedron*, 2015, **71**, 3507–3514.

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

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

16 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.

17 S. P. Sau, A. S. Madsen, P. Podbevsek, N. K. Andersen, T. S. Kumar, S. Andersen, R. L. Rathje, B. A. Anderson, D. C. Guenther, S. Karmakar, P. Kumar, J. Plavec, J. Wengel and P. J. Hrdlicka, *J. Org. Chem.*, 2013, **78**, 9560–9570.

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

19 D. M. Crothers, *Biopolymers*, 1968, **6**, 575–584.

20 P. J. Hrdlicka, T. S. Kumar and J. Wengel, *Chem. Commun.*, 2005, 4279–4281.

21 S. Karmakar, T. Horrocks, B. C. Gibbons, D. C. Guenther, R. Emehiser and P. J. Hrdlicka, *Org. Biomol. Chem.*, 2019, **17**, 609–621.

22 S. P. Adhikari, R. G. Emehiser, S. Karmakar and P. J. Hrdlicka, *Org. Biomol. Chem.*, 2019, **17**, 8795–8799.

23 D. C. Guenther, S. Karmakar and P. J. Hrdlicka, *Chem. Commun. Camb. Engl.*, 2015, **51**, 15051–15054.

24 D. Y. Zhang and E. Winfree, *J. Am. Chem. Soc.*, 2009, **131**, 17303–17314.

25 B. Yurke, A. J. Turberfield, A. P. Mills, F. C. Simmel and J. L. Neumann, *Nature*, 2000, **406**, 605–608.

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

27 T. Bentin, H. J. Larsen and P. E. Nielsen, *Biochemistry*, 2003, **42**, 13987–13995.

28 A. S. Ricciardi, E. Quijano, R. Putman, W. M. Saltzman and P. M. Glazer, *Molecules*, 2018, **23**, 632.

29 T. D. Canady, A. S. Berlyoung, J. A. Martinez, C. Emanuelson, C. A. Telmer, M. P. Bruchez and B. A. Armitage, *Molecules*, 2020, **25**, 970.

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

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

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

33 H. Ihmels, K. Faulhaber, D. Vedaldi, F. Dall'Acqua and G. Viola, *Photochem. Photobiol.*, 2005, **81**, 1107–1115.

34 V. V. Demidov and M. D. Frank-Kamenetskii, *Trends Biochem. Sci.*, 2004, **29**, 62–71.

35 Y.-J. Chen, B. Groves, R. A. Muscat and G. Seelig, *Nat. Nanotechnol.*, 2015, **10**, 748–760.

36 H. Kaur, B. R. Babu and S. Maiti, *Chem. Rev.*, 2007, **107**, 4672–4697.

37 Y. You, B. G. Moreira, M. A. Behlke and R. Owczarzy, *Nucleic Acids Res.*, 2006, **34**, e60.

38 R. Owczarzy, A. V. Tataurov, Y. Wu, J. A. Manthey, K. A. McQuisten, H. G. Almabrazi, K. F. Pedersen, Y. Lin, J. Garretson, N. O. McEntaggart, C. A. Sailor, R. B. Dawson and A. S. Peek, *Nucleic Acids Res.*, 2008, **36**, W163–W169.

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

40 Note: **DNA9**:**DNA10** is a 35-mer model dsDNA target in which the mixed-sequence target region of **DYZ-OPT** is located centrally and flanked by 4 base-pairs on either side.

41 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.

42 A. M. Brown, *Comput. Methods Prog. Biol.,* 2001, **65**, 191-200.

43 Unpublished data, C. Shepard, Thesis - Invader-mediated targeting of chromosomal DNA.

44 D. Andreatta, S. Sen, J. L. P. Lustres, S. A. Kovalenko, N. P. Ernsting, C. J. Murphy, R. S. Coleman and M. A. Berg, *J. Am. Chem. Soc.*, 2006, **128**, 6885–6892.

CHAPTER 4: Nicked Invader probes: Enhanced mixed-sequence recognition of double-stranded DNA targets

Shiva P. Adhikari,[a] Saswata Karmakar[a] and Patrick J. Hrdlicka.[a]

[a] Department of Chemistry, University of Idaho, Moscow, ID-83844, USA

Abstract

Sequence-specific recognition of double-stranded (ds) DNA at physiological conditions has a wide range of applications in diagnostic and therapeutic fields. Our laboratory has designed so-called Invader probes, which are short DNA duplexes featuring +1 interstrand arrangements of O2'-intercalator-functionalized RNA monomers as the central design feature that activates Invader probes for specific sequence-unrestricted dsDNA recognition. The design relies on large stability differences between probe duplexes and recognition complexes to drive dsDNA recognition. Invader probes exhibit avidity in binding to dsDNA via invasion modes, offer the promise of favorable binding affinity and specificity, and straightforward design. In the present study, we set out to optimize the dsDNA-recognition properties of Invader probes through the introduction of nicks (i.e., nicked Invader probes, NIPs). Appropriately designed nicked Invader probes display significantly greater recognition of model mixed-sequence dsDNA targets vis-à-vis corresponding toehold and conventional Invader probes at physiological conditions.

4.1 Introduction

Chemical probes capable of detecting biological DNA are needed for gene identification, regulation, and editing.[1] Triplex forming oligonucleotides (TFOs),[2] minor-groove binding polyamides,[3] and peptide nucleic acids (PNAs)[4] have all made substantial advances in this direction, albeit they all have significant limitations, for instance, TFOs require long polypurine target regions, Polyamide necessitates short target region (< 6-8 nucleotides), while PNAs require low ionic strengths to recognize dsDNA efficiently.[1-5] Alternative PNA-based methods have been developed with less stringent sequence constraints, where the use of conformationally restricted γ-PNAs has been shown to increase cDNA binding affinity, possibly due to strand preorganization and decreased entropic penalties, although invasion is

inefficient at physiological ionic strengths.[6] Moreover, CRISPR/Cas, an RNA-guided protein-based DNA-targeting technology,[7] has shown great potential, despite presenting major off-target and delivery issues.[8] As a result, there is a need for oligonucleotide-based chemical probes that can recognize mixed-sequence dsDNA under physiological conditions. In this regard, for recognizing dsDNA, we have previously designed so-called Invader probes,[9] which are based on energetically activated double-stranded oligodeoxyribonucleotide (ON) probes modified with one or more +1 interstrand zipper arrangements of intercalator-modified nucleotides like 2′-O-(pyren-1-yl)methyl-RNA,[5] because the nearest neighbor exclusion principle[10] - which states that intercalation is anti-cooperative at adjacent sites[11] - is violated, when the intercalators are pushed into the same inter-base-pair region of the probe duplex, resulting in unwinding and probe destabilization,[12,13] hence, the name hotspots.[5] Since they form probe-target duplexes accompanied by strongly stabilizing stacking interactions between intercalators and surrounding nucleobases, each of the two strands of energetically activated probes has a very high affinity for cDNA. [12,13] The driving force for dsDNA recognition is the energy difference between the double-stranded probe duplex and the probe-target duplexes.[1,5] To enhance dsDNA recognition, we recently introduced affinity enhancing overhangs in Invader probes, known as toehold Invader probes.[14] To facilitate dsDNA identification, we recently created affinity boosting overhangs Invader probes in Invader probes, dubbed toehold Invader probes. We set out to investigate the impact of introducing complementary auxiliary strands in the toehold region after being inspired by the toehold project, which showed promise for efficient recognition of dsDNA targets vs blunt ended Invaders.[14] This probe design would mimic a nicked version of an otherwise long blunt Invader probe, thereby named as Nicked Invader probes. Introducing a nick likely increases flexibility and lability of the probe duplex, which should increase the likelihood of partial probe denaturation, facilitating recognition of dsDNA target. The coaxial adjacent base stacking at the nicked region will provide additional stability to form DNA duplexes since base stacking is the dipole−dipole interaction between the planar aromatic bases in two adjacent nucleotides, which contribute to the major forces for DNA duplex stability.[15] Aside from the destabilization offered by the centrally located hotspot pairs as in conventional probes, nicked probes are expected to exhibit greater lability assisted by nicks and presence of additional hotspots in auxiliary segments (Fig. 4.1). Different designs of nicked Invader probes were designed and evaluated to compare the recognition efficiency

relative to equivalent toehold Invader probes and blunt-ended conventional Invader probes (Fig. 4.2).
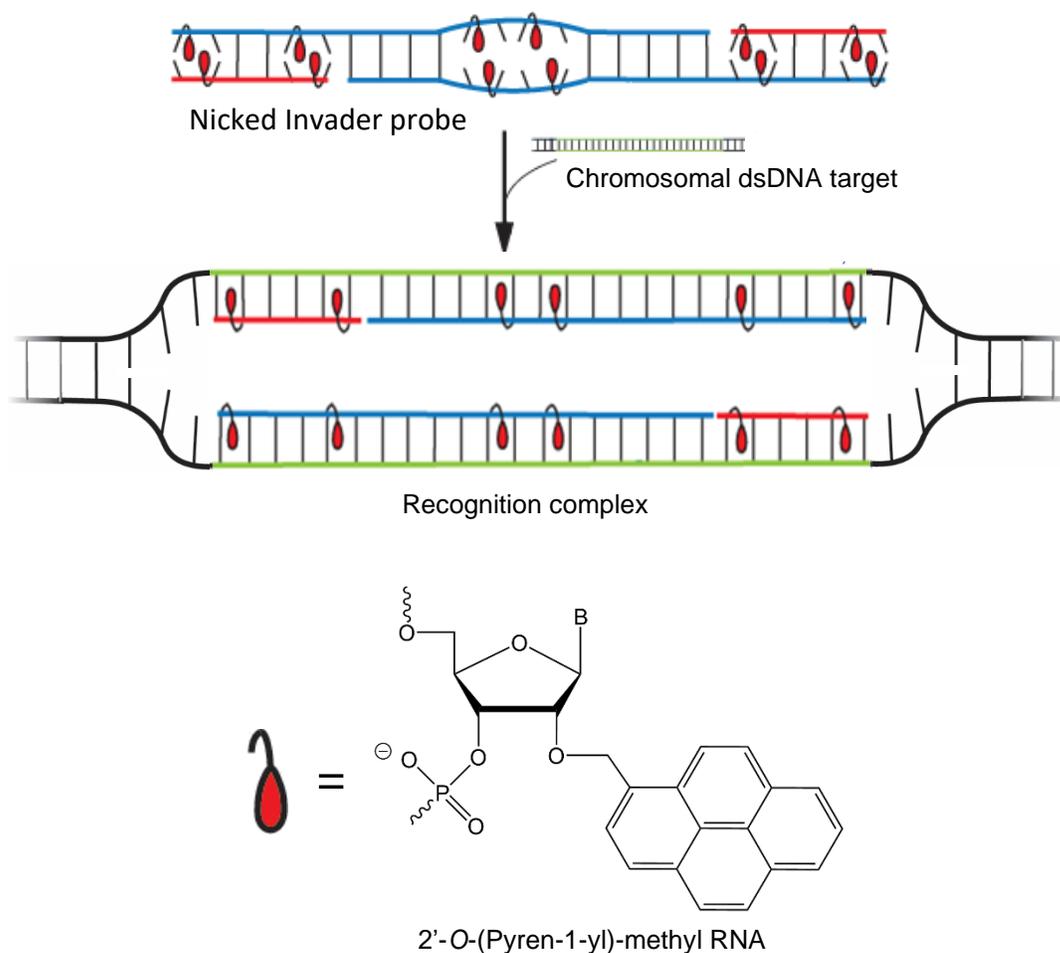
Figure 4.1. Schematic of the dsDNA-recognition process using nicked Invader probes (NIPs) and structure of 2'-*O*-(pyren-1-yl)methyl RNA monomer used to generate energetic hotspots.

4.2 Results and discussion

Design and synthesis of probes. The 2'-*O*-(pyren-1-yl)methyl-RNA-modified oligodeoxyribonucleotides (ONs) used herein were synthesized using established protocols or available from prior studies (Tables 4.3).[16,17] Access to these ONs enabled assembly of four nicked Invader probes (**NIP1-NIP4**), designed to recognize a 25-base-pair (bp) mixed-sequence DNA region embedded within longer targets (Fig. 4.2). Each nicked Invader probe is comprised of four strands and features a total of six identically positioned energetic hotspots. The location

of the nicks varies among the probes. Each nicked Invader probe can be viewed as comprising three double-stranded segments, i.e., a central segment that is flanked on either side by an auxiliary segment. Accordingly, nicked Invader probes can be described according to the number of base-pairs in each segment. Thus, **NIP1**-**NIP4** can be referred to as 6-13-6, 8-9-8, 10-5-10, and 12-1-12 constructs, respectively (Fig. 4.2). Alternatively, nicked Invader probes can be regarded as toehold Invader probes (i.e., probe duplexes with single-stranded 5'-overhangs) that are hybridized to two auxiliary strands. For example, **NIP1** can be regarded as a toehold Invader probe comprised of two 19-mer main strands and two auxiliary 6-mer strands that are hybridized to the 5'-single-stranded overhangs of the toehold probe. Thus, the auxiliary strands of **NIP1**-**NIP4** are designed to hybridize to 6-, 8-, 10- and 12-mer single-stranded overhangs of the corresponding toehold probes, respectively.

(a)

| 6-bp left auxiliary segment | 13-bp central segment | 6-bp right auxiliary segment |

Nicked Invader probe **NIP1** (6-13-6 construct)

(b)

Probe	Construct	Sequence
NIP1	6-13-6	5'- UGC ACA GGT AUA UAT AGG CCG CAU A -3' 3'- ACG TGU CCA TAU AUA TCC GGC GTA U -5'
NIP2	8-9-8	5'- UGC ACA GGT AUA UAT AGG CCG CAU A -3' 3'- ACG TGU CCA TAU AUA TCC GGC GTA U -5'
NIP3	10-5-10	5'- UGC ACA GGT AUA UAT AGG CCG CAU A -3' 3'- ACG TGU CCA TAU AUA TCC GGC GTA U -5'
NIP4	12-1-12	5'- UGC ACA GGT AUA UAT AGG CCG CAU A -3' 3'- ACG TGU CCA TAU AUA TCC GGC GTA U -5'

Figure 4.2. (a) Illustration of a representative nicked Invader probe and its different duplex segments. (b) Sequences of nicked Invader probes used herein. **U** and **C** denote 2′-*O*-(pyren-1-yl)methyluridine and 2′-*O*-(pyren-1-yl)methylcytidine monomers, respectively.

Thermal denaturation properties. Given the four-stranded nature of nicked Invader probes and their three distinct double-stranded segments, it is not feasible to determine thermal denaturation temperatures (T_ms) for the fully assembled probes, as overlapping denaturation would ensue. Instead, T_ms were estimated for each of the three segments by determining the T_ms for the duplexes formed by the i) 5'-main probe and 3'-auxiliary probe (estimating the stability of the "left" auxiliary segment), ii) 5'-main probe and 3'-main probe (estimating the stability of the central segment), and iii) 5'-auxiliary probe and 3'-main probe (estimating the stability of the "right" auxiliary segment) (1st-3rd T_m columns, Table 4.1). As expected for a 13-bp double-stranded segment with two energetic hotspots, the central segment of **NIP1** is labile as indicated by the moderate T_m value for **ON1**:**ON4** (T_m = 49.0 °C). Unsurprisingly, the value decreases as the number of base-pairs in the central segment is reduced in steps of four from thirteen to one (notice the T_m trend on going from **ON1**:**ON4** to **ON5**:**ON8**, **ON9**:**ON12** and **ON13**:**ON16**, 2nd T_m column, Table 4.1). Along similar lines, the stability of the "left" and "right" auxiliary segments increases as the number of base-pairs increases in steps of two from six to twelve (i.e., notice the T_m trend on going from **ON1**:**ON3** to **ON5**:**ON7**, **ON9**:**ON11** and **ON13**:**ON15**, 1st T_m column, Table 4.1, and on going from **ON2**:**ON4** to **ON6**:**ON8**, **ON10**:**ON12** and **ON14**:**ON16**, 3rd T_m column). Only the central segments of **NIP1** and **NIP2**, and the auxiliary sections of **NIP2**-**NIP4**, are expected to be stable at the experimental temperatures used herein (37 °C, *vide supra*). Hence, only **NIP2** is expected to be largely assembled at 37 °C.

To estimate the stability of the four probe-target duplex segments that would form upon successful dsDNA-recognition by nicked Invader probes (Fig. 4.1), we determined the T_m values for duplexes between individual probe strands and 33-mer single-stranded (ss) DNA targets that harbor the complementary regions (4th-7th T_m columns, Table 4.1). Unsurprisingly, the T_m values decrease as the length of the probe strand – and, thus, the number of formed base-pairs – decreases (e.g., note the decreasing T_ms for duplexes between **DNA2** and **ON1**, **ON5**, **ON9** or **ON13**, 4th T_m column, or – equivalently – the increasing T_ms for duplexes between **DNA2** and **ON2**, **ON6**, **ON10** or **ON14**, 5th T_m column). These results suggest that recognition of complementary dsDNA regions will result in the formation of four stable probe-target duplex segments when using **NIP2**-**NIP4** (i.e., all probe-target segments display T_ms $\geq$ 45 °C), whereas

the probe-target duplex segments entailing the 6-mer auxiliary strands of **NIP1** are labile ($T_m <$ 15 °C and $T_m = < 25.5$ °C, for **ON2**:**DNA2** and **ON3**:**DNA1**, respectively).

Table 4.1. Sequences of probes used in this study, as well as T_ms of probe duplexes and duplexes with DNA targets, and thermal advantages (*TA*s) of probes.[a]

ONs	Sequence	T_m (°C)							*TA* (°C)
		5'-main probe vs 3'-aux probe	5'-main probe vs 3'-main probe	5'-aux probe vs 3'-main probe	5'-main probe vs DNA2	5'-aux probe vs DNA2	3'-aux probe vs DNA1	3'-main probe Vs DNA1	
1/2 3/4	5'-<u>U</u>GCA<u>C</u>AGGTA<u>U</u>A<u>U</u>ATAGGC<u>C</u>GCA<u>U</u>A 3'-A<u>C</u>GTG<u>U</u>CCATA<u>U</u>A<u>U</u>ATCCGG<u>C</u>GTA<u>U</u>	<15.0	49.0[b]	<15.0	75.5	<15.0	<25.5	75.5	nd
5/6 7/8	5'-<u>U</u>GCA<u>C</u>AGGTA<u>U</u>A<u>U</u>ATAGG<u>C</u>CGCA<u>U</u>A 3'-A<u>C</u>GTG<u>U</u>CCATA<u>U</u>A<u>U</u>ATCCGG<u>C</u>CGTA<u>U</u>	36.0	37.0	36.0	73.0	57.0	46.0	74.5	69.5
9/10 11/12	5'-<u>U</u>GCA<u>C</u>AGGTA<u>U</u>A<u>U</u>ATAGGC<u>C</u>GCA<u>U</u>A 3'-A<u>C</u>GTG<u>U</u>CCATA<u>U</u>A<u>U</u>ATCCGG<u>C</u>GTA<u>U</u>	45.5	<15.0	52.0	72.5	66.0	57.5	74.0	>85.5
13/14 15/16	5'-<u>U</u>GCA<u>C</u>AGGTA<u>U</u>A<u>U</u>ATAGGC<u>C</u>GCA<u>U</u>A 3'-A<u>C</u>GTG<u>U</u>CCATA<u>U</u>A<u>U</u>ATCCGG<u>C</u>GTA<u>U</u>	54.0	<15.0	62.5	72.0	66.5	65.5	70.0	>70.5
17 4	5'-GGTA<u>U</u>A<u>U</u>ATAGGC<u>C</u>GCA<u>U</u>A 3'-CCATA<u>U</u>A<u>U</u>ATCCGG<u>C</u>GTA<u>U</u>	-	66.0	-	76.5	-	-	75.5	14.0
1 18	5'-<u>U</u>GCA<u>C</u>AGGTA<u>U</u>A<u>U</u>ATAGGC 3'-A<u>C</u>GTG<u>U</u>CCATA<u>U</u>A<u>U</u>ATCCG	-	66.0	-	75.5	-	-	76.0	13.5
19 20	5'-GGTA<u>U</u>A<u>U</u>ATAGGC 3'-CCATA<u>U</u>A<u>U</u>ATCCG	-	42.5	-	54.5	-	-	54.0	-6.0

[a] **NIP1** = **ON1**+**2**:**ON3**+**4**, **NIP2** = **ON5**+**6**:**ON7**+**8**, **NIP3** = **ON9**+**10**:**ON11**+**12** and **NIP4** = **ON13**+**14**:**ON15**+**16**. **DNA1** = 5'-AAGCTGCACAGGTATATATAGGCCGCATATGCA and **DNA2** = 3'-TTCGACGTGTCCATATATATCCGGCGTATACGT-5'. T_m (**DNA1**:**DNA2**) = 72.0 °C). Thermal denaturation curves were recorded in medium salt phosphate buffer ([Na$^+$] = 110 mM, [Cl$^-$] = 100 mM, pH 7.0 (NaH$_2$PO$_4$/Na$_2$HPO$_4$), [EDTA] = 0.2 mM), with each strand at a 0.5 μM concentration. A = adenin-9-yl DNA monomer, C = cytosin-1-yl DNA monomer, G = guanin-9-yl DNA monomer, T = thymin-1-yl DNA monomer. For structures of **U** and **C**, see Fig. 4.1. nd = not determinable; nt = no transition; "-" = N/A. For the definition of *TA*, see main text. [b] Data previously reported in reference 14.[14]

Moreover, as we have previously reported for conventional and toehold Invader probes,[5,17] duplexes between individual probe strands and complementary ssDNA regions are more stable than the double-stranded probes (e.g., T_m = 73.0 °C, 74.5 °C, and 37.0 °C, for **ON5**:**DNA2**, **ON8**:**DNA1** and **ON5**:**ON8**, respectively), because the probe duplex violate exclusion

principle, while the probe-target duplexes are stabilized by pyrene intercalation. Accordingly, dsDNA-recognition using nicked Invader probes is expected to be energetically favorable since the four probe-target duplex segments formed are more stable than the double-stranded probes and the dsDNA target. The driving force for recognition of **DNA1:DNA2**, i.e., a 33-mer model dsDNA target harboring the complementary 25-bp region, can be estimated by the term *thermal advantage* (*TA*) defined as $TA = T_m$ (5'-main probe *vs.* **DNA2**) + T_m (5'-aux probe *vs.* **DNA2**) + T_m (3'-main probe *vs.* **DNA1**) + T_m (3'-aux probe *vs.* **DNA1**) − T_m (5'-main probe *vs.* 3'-aux probe) − T_m (5'-main probe *vs.* 3'-main probe) − T_m (5'-aux probe *vs.* 3'-main probe) − T_m (dsDNA target), with large positive values indicating a strongly activated probe. Indeed, **NIP2**-**NIP4** display *TA* values in excess of 70 °C, which suggests that they are far more activated for recognition of **DNA1:DNA2** than conventional 13-mer Invader probe **ON19:ON20** featuring two energetic hotspots (*TA* = -6 °C, Table 4.1), conventional 19-mer Invader probes **ON1:ON18** and **ON17:ON2** featuring four energetic hotspots (*TA* ~ 14 °C, Table 4.1), or the toehold versions of **NIP1**-**NIP4**, i.e., **ON1:ON4** (*TA* ~ 30 °C), **ON5:ON8** (*TA* ~ 38.5 °C), **ON9:ON12** (*TA* > 59.5 °C) and **ON13:ON16** (*TA* > 57 °C) (for toehold Invader probes, the *TA* calculation simplifies to $TA = T_m$ (5'-probe *vs.* **DNA2**) + T_m (3'-probe *vs.* **DNA1**) − T_m (toehold Invader probe) − T_m (dsDNA target)).

Characterization of dsDNA-targeting properties of the nicked Invader probes.

Preliminary screen. The dsDNA-targeting properties of **NIP1**-**NIP4** were evaluated using an electrophoretic mobility shift assay in which a 3'-digoxigenin (DIG)-labeled hairpin **DH1** was used as a model dsDNA target (Fig. 4.3a). **DH1** comprises a 33-mer double-stranded stem – of the same sequence as **DNA1:DNA2** used in the denaturation studies – and a decameric thymidine loop (T_{10}) that links the two stem strands at one end (Table 4.4). The 25-bp region that is complementary to the nicked Invader probes is embedded within the stem of **DH1**. The unimolecular nature of **DH1** renders it as a high-melting target (T_m = 81 °C).[14]

Recognition of **DH1** is expected to result in the formation of a five-stranded recognition complex that has lower mobility on non-denaturing polyacrylamide gels than **DH1**. Indeed, when a 50-fold molar excess of pre-annealed **NIP1**-**NIP4** was incubated with **DH1** at 37 °C, low-mobility bands were observed (Fig. 4.3b). Varying levels of **DH1** recognition are observed, ranging from minor (~15% with **NIP4**) to moderate recognition (~40% with **NIP2**) (Table 4.5).

These results indicate that nicked Invader probes with a near-tiling architecture (12-1-12 construct) and/or duplex segments of unequal stability (6-13-6 or 10-5-10 constructs), recognize **DH1** less efficiently than a nicked Invader probe with duplex segments of comparable stability (8-9-8 construct), irrespective of the thermodynamic driving forces.

Interestingly, nicked Invader probes result in more efficient recognition of the model dsDNA target **DH1** than the corresponding toehold probes, with differences being more pronounced for nicked Invader probes constructed using longer auxiliary strands (e.g., compare recognition using **NIP2** relative to **ON5**:**ON8**, Fig. 4.3b and Table 4.5).

Toehold Invader probes with shorter overlaps between the probe strands, which accordingly target longer regions, result in more efficient recognition of **DH1**. For example, **ON1**:**ON4**, leading to the formation of two 19-bp probe-target duplexes, results in more efficient recognition than **ON5**:**ON8**, leading to the formation of two 17-bp probe-target duplexes (~19% and ~14%, respectively, Fig. 4.3b and Table 4.5). However, the modest differences among the probe-target duplexes (e.g., compare T_m for **DNA2** and **ON1**, **ON5**, **ON9** or **ON13**, 4th T_m column, Table 4.1), suggest that probe stability also is an important factor. Indeed, recognition of **DH1** is more efficient with increasing stability of the toehold probes (correlate data in 2nd T_m column of Table 4.1 with efficiency of **DH1**-recognition shown in Fig. 4.3b and Table 4.5), which indicates that a stable double-stranded segment is needed to ensure efficient dsDNA-recognition.

Conventional 19-mer Invader probes **ON1**:**ON18** and **ON17**:**ON2** only result in low levels of **DH1**-recognition (<15%) even though two 19-bp probe-target duplex segments are formed as with **ON1**:**ON4** (Fig. 4.3b and Table 4.5). The less efficient dsDNA-recognition is likely linked to the high probe stability (T_m ~ 66 °C, Table 4.1), which renders dissociation kinetically unfeasible.

Conventional 13-mer Invader probe **ON1**:**ON2**, which can be regarded as a truncated version of **NIP1**, does not result in appreciable levels of **DH1**-recognition due to unfavorable binding energetics (Fig. 4.3b and Table 4.5).

To sum up, the preliminary screen suggests that appropriately designed nicked Invader probes offer advantages with respect to dsDNA-recognition relative to toehold and conventional Invader probes.
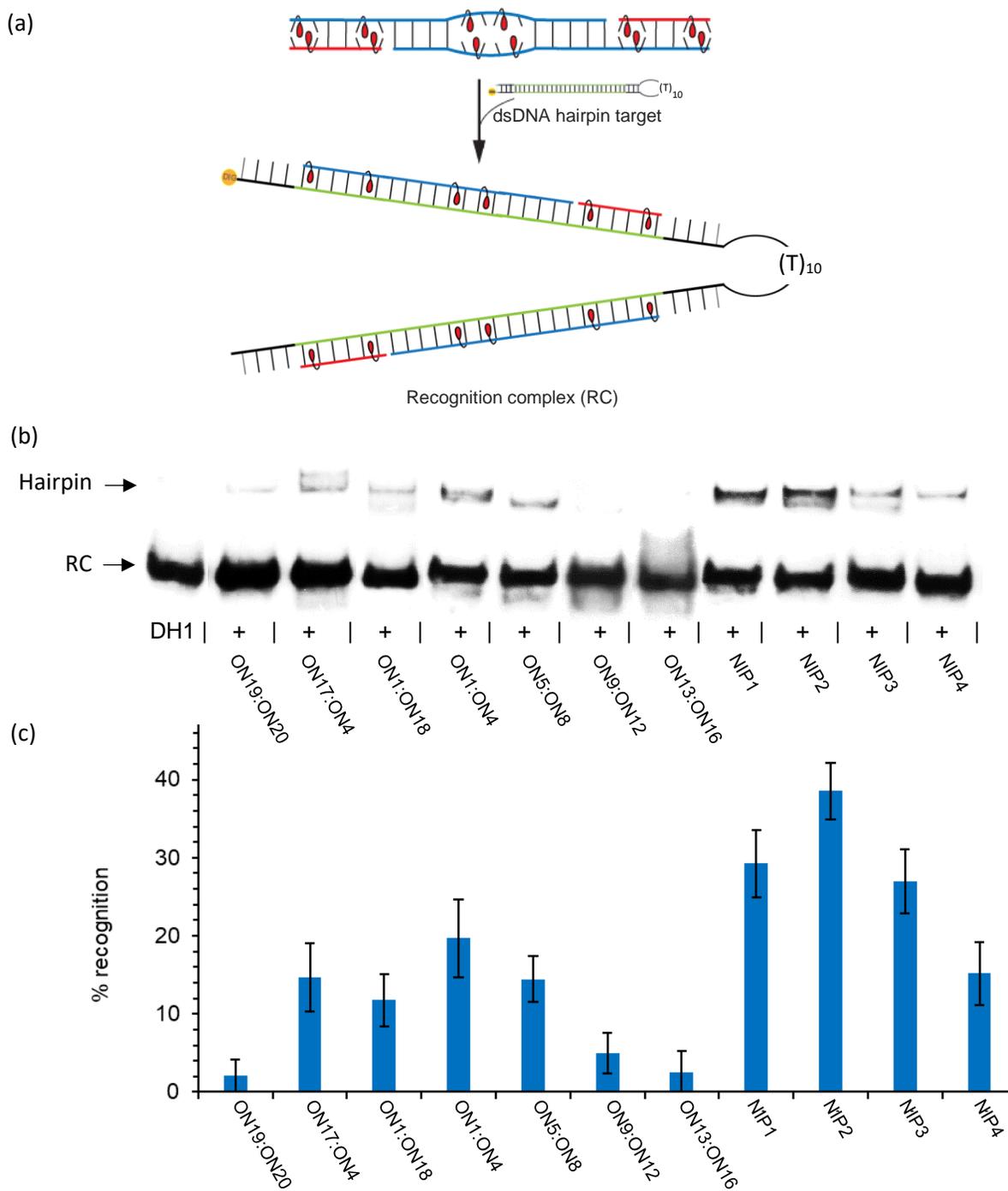
Figure 4.3. (a) Illustration of the electrophoretic mobility shift assay used to evaluate dsDNA-recognition of Invader probes. (b) Representative gel electrophoretograms from recognition experiments in which **DH1** was incubated with a 50-fold molar excess of different probes. RC = recognition complex. (c) Histogram depicting averaged results from at least three independent experiments with error bars representing standard deviation. Conditions: DIG-labeled **DH1** (50

nM) was incubated with a 50-fold molar excess of the specified probe in HEPES buffer (50 mM HEPES, 100 mM NaCl, 5 mM $MgCl_2$, pH 7.2, 10% sucrose, 1.44 mM spermine tetrahydrochloride) for 17h at 37 °C. Sequence of **DH1** shown in Table 4.4.

Dose-response profiles. Based on the encouraging preliminary results, we set out to determine dose-recognition profiles for nicked Invader probes **NIP1-NIP4** and the corresponding toehold probes (Figs. 4.4 and 4.36 - 4.37). In agreement with the results from the preliminary screen, **NIP2** displays the lowest C_{30} value (i.e., probe concentration resulting in 30% recognition of **DH1**), whereas **NIP3** and **NIP4** display C_{30} values that are nearly seven- and fifteen-fold higher, respectively (~0.5 µM vs ~3.4 µM, and 7.4 µM). Recognition of **DH1** reaches a plateau of ~ 28% when using **NIP1**, ~35% when using **NIP4**, and 40-45% when using **NIP2** or **NIP3**. Less than 20% recognition of **DH1** is observed with the corresponding toehold probes (Fig. 4.38).
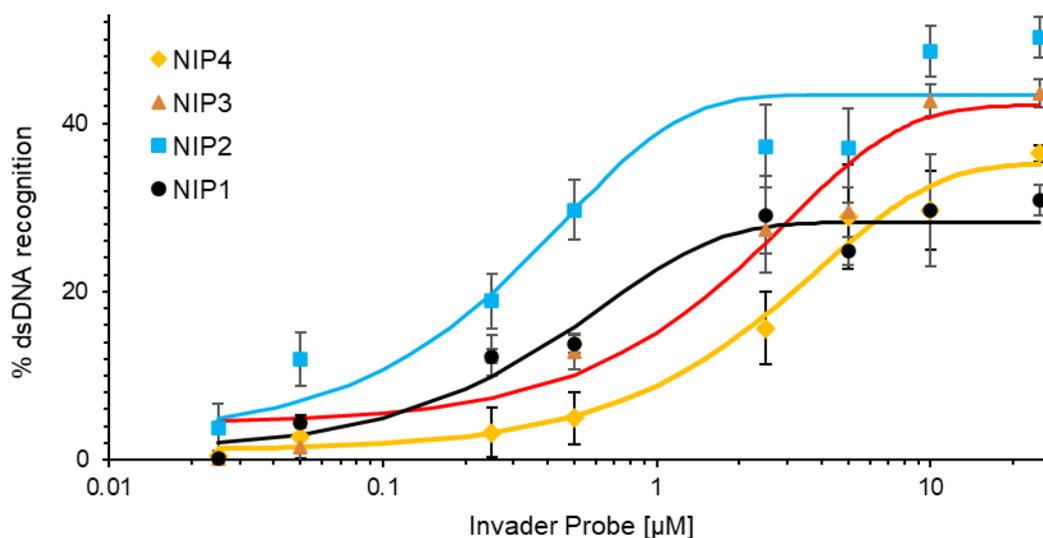
Figure 4.4. Dose-response profiles for recognition of **DH1** by nicked Invader probes **NIP1-NIP4** at 37 °C. Curves are constructed based on the electrophoretograms shown in Figs. 4.36 and 4.37. Experimental conditions are as described in Fig. 4.3.

Binding specificity. Next, the binding specificities of the different nicked Invader probes were evaluated by incubating a 50-fold molar probe excess with:

iv)   **MM1**, which differs in sequence at one position relative to **DH1** (corresponding to the position of the third energetic hotspot of the probes as drawn; binding would result in the formation of a recognition complex, in which two of the four formed probe-target duplex segments would harbor one mismatched base-pair each, Fig. 4.5)

v)   **MM2**, which differs in sequence at two positions relative to **DH1** (corresponding to the positions between the 1st and 2nd , and 5th and 6th energetic hotspot of the probes as drawn; binding would result in the formation of a recognition complex, in which all four of the formed probe-target duplex segments would harbor one mismatched base-pair, Fig. 4.5), and

vi) **MM3**, which differs in sequence at three positions relative to **DH1** (corresponding to the position of the third energetic hotpot, and between the 1st and 2nd , and 5th and 6th energetic hotspot of the probes as drawn; binding would result in the formation of a recognition complex, in which two of the four formed probe-target duplex segments would harbor two mismatched base-pairs each, while the other two probe-target duplex segments would harbor one mismatched base-pair each, Fig. 4.5).

Each of the nicked Invader probes discriminate the singly, doubly, and triply mismatched DNA hairpin targets with extraordinary efficiency (Figs 4.5, 4.39 and 4.40). In fact, the binding specificity is even greater than with the corresponding toehold Invader probes.[14] Presumably, the remarkable binding specificity is the result of multiple effects, including stringency clamping effects, i.e., greater stability differences between matched vis-à-vis mismatched recognition complexes seen with structured probes,[18,19] avoidance of energetically unfavorable formation of multiple double-stranded segments with mismatched base-pairs,[20] and the fact that shorter (nicked) oligonucleotides generally display improved specificity. Thus, these results strongly suggest that it is possible to improve both the binding affinity and binding specificity of conventional and toehold Invader probes by using a nicked probe design.
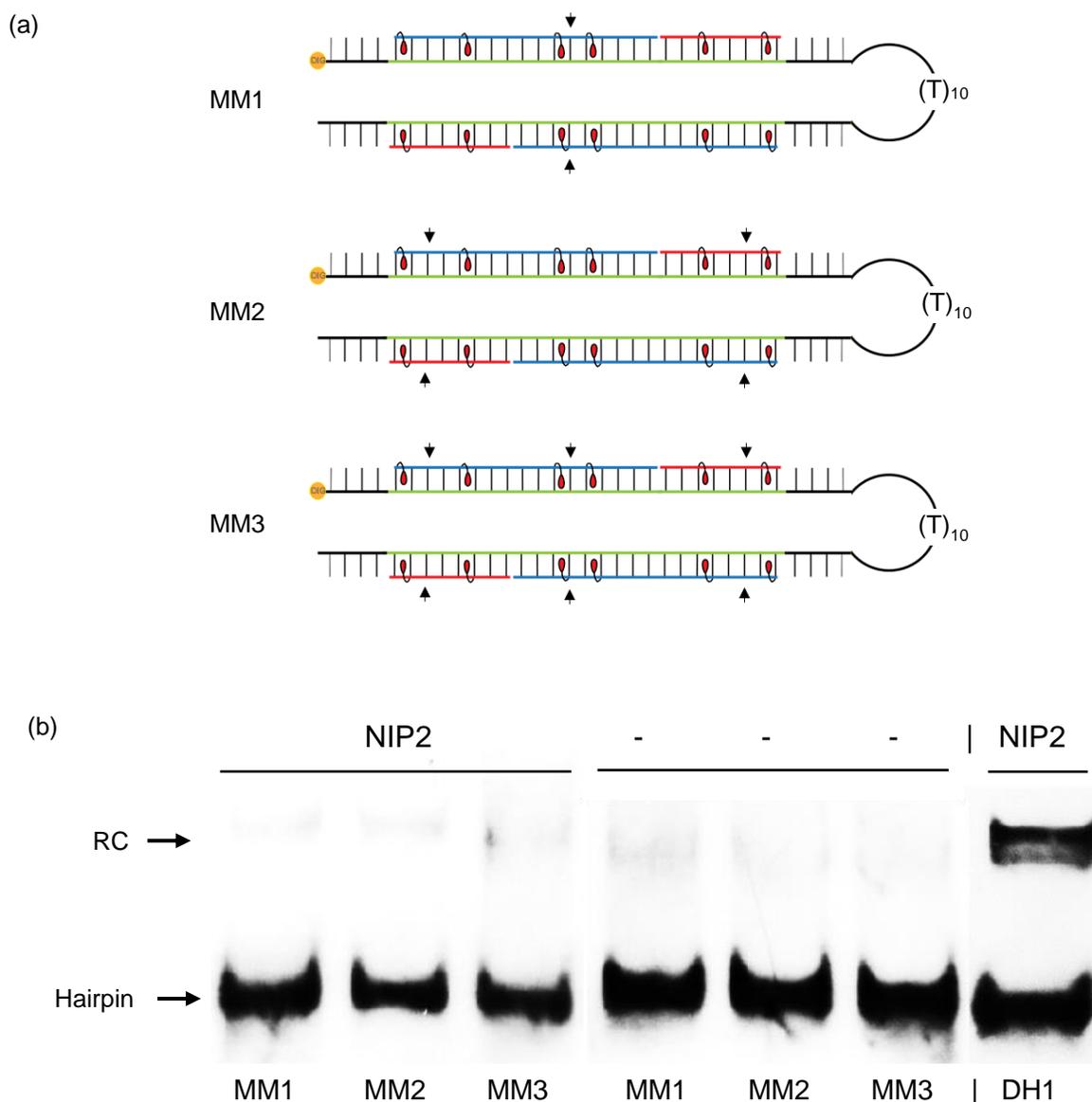
Figure 4.5. Binding specificity of nicked Invader probe **NIP2**. (a) Illustration of the mismatched recognition complexes that would ensue upon recognition of **MM1**-**MM3** by **NIP2**; arrows indicate position of mismatched base-pairs. For sequences of **MM1**-**MM3**, see Table 4.4. (b) Representative electrophoretograms from experiments in which **NIP2** was incubated with non-complementary targets **MM1**-**MM3** or complementary target **DH1**. Pre-annealed 3'-DIG-labelled hairpins (50 nM) were incubated with a 50-fold molar excess of pre-annealed probe at 37 °C for 17 h in HEPES buffer as outlined in Fig. 4.3.

Detection of chromosomal DNA using nicked Invader probes. Encouraged by the above findings, we set out to demonstrate the use of nicked Invader probes for recognition of mixed-sequence chromosomal DNA regions. We have previously demonstrated the feasibility of conventional[5] and toehold[14] Invader probes for recognition of a highly repeated region 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: M26067)[21] in the context of non-denaturing fluorescence in situ hybridization (nd-FISH) experiments. One probe, however, i.e., the conventional 15-mer Invader probe DYZ-REF (sequence shown later in Fig. 4.6, Table 4.2), has proven refractory to recognition of the *DYZ-1* region.[22] To overcome this limitation, we constructed the Cy3-labeled nicked Invader probe **DYZ-NIP** (**ON23/24**:**ON25/26**, Table 4.2), i.e., a 8-9-8 construct with six energetic hotspots that targets positions 862-886 of the *DYZ-1* gene.

Table 4.2. Sequences of probes used in FISH experiments and T_ms of probe duplexes and duplexes between individual probe strands and DNA targets.[a]

ONs	Sequence	5'-main probe vs 3'-aux probe	5'-main probe vs 3'-main probe	5'-aux probe vs 3'-main probe	5'-main probe vs DNA2	5'-aux probe vs DNA2	3'-aux probe vs DNA1	3'-main probe Vs DNA1	*TA* (°C)
23/24 **25/26**	5'-Cy3-TG**U**GT**U**ATA**U**GC**U**GTTCT**C**AGCC**C**T 3'-ACA**C**AA**U**ATA**C**GA**C**AAGAG**U**CGGG**A**-Cy3	<20.0	nt	nt	63.0	35.5	34.5	68.0	nd
21 **22**	5'-Cy3-T**U**A **U**AT GCT G**U**T CTC 3'- AA**U** A**U**A CGA CA**A** GAG-Cy3	-	56.0	-	61.0	-	-	68.0	+8.0

[a] **DNA3**:**DNA4** duplex is a model dsDNA target (T_m = 65.0 °C), where **DNA3** = 5'-ACTGTGTGTTATATGCTGTTCTCAGCCCTACTG and **DNA4** = 5'-CAGTAGGGCTGAGAACAGCATATAACACACAGT. Experimental conditions are as stated in Table 4.1. 'nt'= no clear sigmoidal transition. **ON23/24**:**ON25/26** = **DYZ-NIP, ON21:ON22 = DYZ-REF**.

None of the three double-stranded segments of **DYZ-NIP** display transitions above 20 °C (1st-3rd T_m columns, Table 4.2). For the left and central sections (estimated by **ON23**:**ON25** and **ON23**:**ON26**) this is likely related to their low GC-content, whereas the presence of a G-triplet might impact the stability of the right segment (estimated by **ON24**:**ON26**). In high salt buffers, transitions at 36-37 °C were observed for the left and central sections, while there was a continued absence of transitions for the right segment (1st-3rd T_m columns, Table 4.6). In concert, these observations suggest that **DYZ-NIP** is a labile probe. Each of the four **DYZ-**

NIP probe strands form stable duplexes with 33-mer single-stranded DNA targets that harbor complementary regions (4th-7th T_m columns, Table 4.2). The high affinity of the two main strands **ON23** and **ON26** towards their ssDNA target regions is particularly noteworthy ($T_m =$ 63 °C and 68 °C, respectively, Table 4.2). This greater stability of the probe-target duplexes vis-à-vis the double-stranded probes, suggests that recognition of the dsDNA target should be thermodynamically favorable when using **DYZ-NIP**. The stability of the probe-target duplexes ($T_m \sim$ 61-68 °C) formed by the blunt ended Invader probe **DYZ-REF** are comparable to stability of dsDNA target ($T_m \sim$ 65 °C), while the probe duplex itself is too stable ($T_m =$56 °C), suggesting the marginal thermodynamic potential for recognition of dsDNA target ($TA = 8$ °C).

The aforementioned assay was used to evaluate the ability of **DYZ-NIP** to recognize a complementary dsDNA region embedded within a DNA hairpin. Indeed, incubation of a 50-fold molar excess of **DYZ-NIP** with **DH2** (designed in an equivalent fashion as **DH1**) results in ~15% recognition, whereas the corresponding toehold and conventional probes (i.e., **ON23**:**ON26** and **DYZ-REF**, respectively) do not result in observable levels of recognition (Fig. 4.6).

(a)
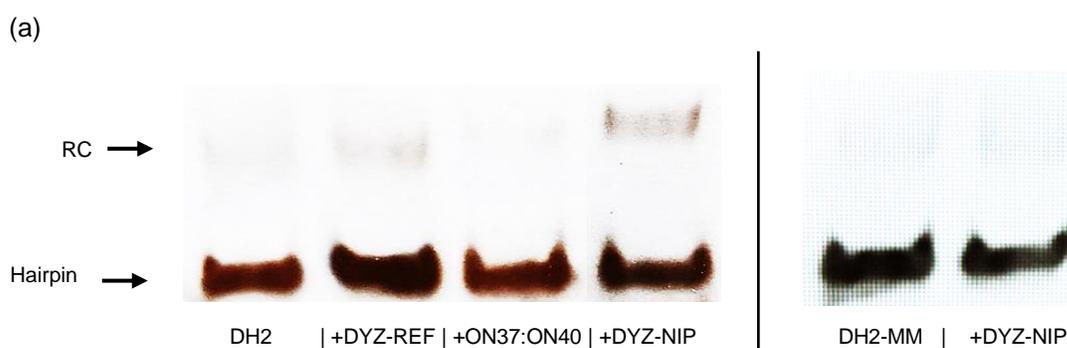
Figure 4.6. Representative gel electrophoretograms from recognition experiments in which complementary target **DH2** or non-complementary target **DH2-MM** were incubated with a 50-fold molar excess of different DYZ-targeting probes at 37 °C for 17 h. See Table 4.4 and Fig 4.41 for sequences of **DH2** and **DH2-MM** and structures of recognition complexes formed. **DYZ**-REF = 5'-Cy3-T**U**A**U**ATGCTG**U**TCTC:3'-AA**U**A**U**ACGACA**A**GAG-Cy3. Experimental conditions are as described in Fig. 4.2.

The binding specificity of **DYZ-NIP** was evaluated using DNA hairpin **DH2-MM**, which was designed akin to the previously discussed **MM3**, i.e., comprising a region that differs in sequence at three positions vis-à-vis **DYZ-NIP**. Importantly, **DYZ-NIP** displays excellent discrimination of **DH2-MM**, as evidenced by the absence of the recognition complex band (Fig. 4.6).

Next, nd-FISH experiments were carried out in which fixed interphase nuclei from a male bovine kidney cell line were incubated with **DYZ-NIP**. Gratifyingly, **DYZ-NIP** recognizes the chromosomal DNA targets as evidenced by the formation of single localized Cy3-signals in ~40% of the nuclei, whereas the conventional Invader probe **DYZ-REF** did not result in signal formation (Fig. 4.7). As reported in our earlier study, the corresponding toehold version of **DYZ-NIP**, i.e., **ON23**:**ON26** also results in efficient recognition of the *DYZ-1* target (~40 % recognition of target nuclei).

Control experiments were performed in which the Y-chromosome-targeting **DYZ-NIP** was incubated with nuclei from a female bovine endothelial cell line, which lacks the *DYZ-1* region. Importantly, no signals were produced in the context of nd-FISH experiments (Fig. 4.43), strongly suggesting that nicked Invader probes target mixed-sequence chromosomal DNA under non-denaturing conditions with excellent binding specificity.

(a)

(b)
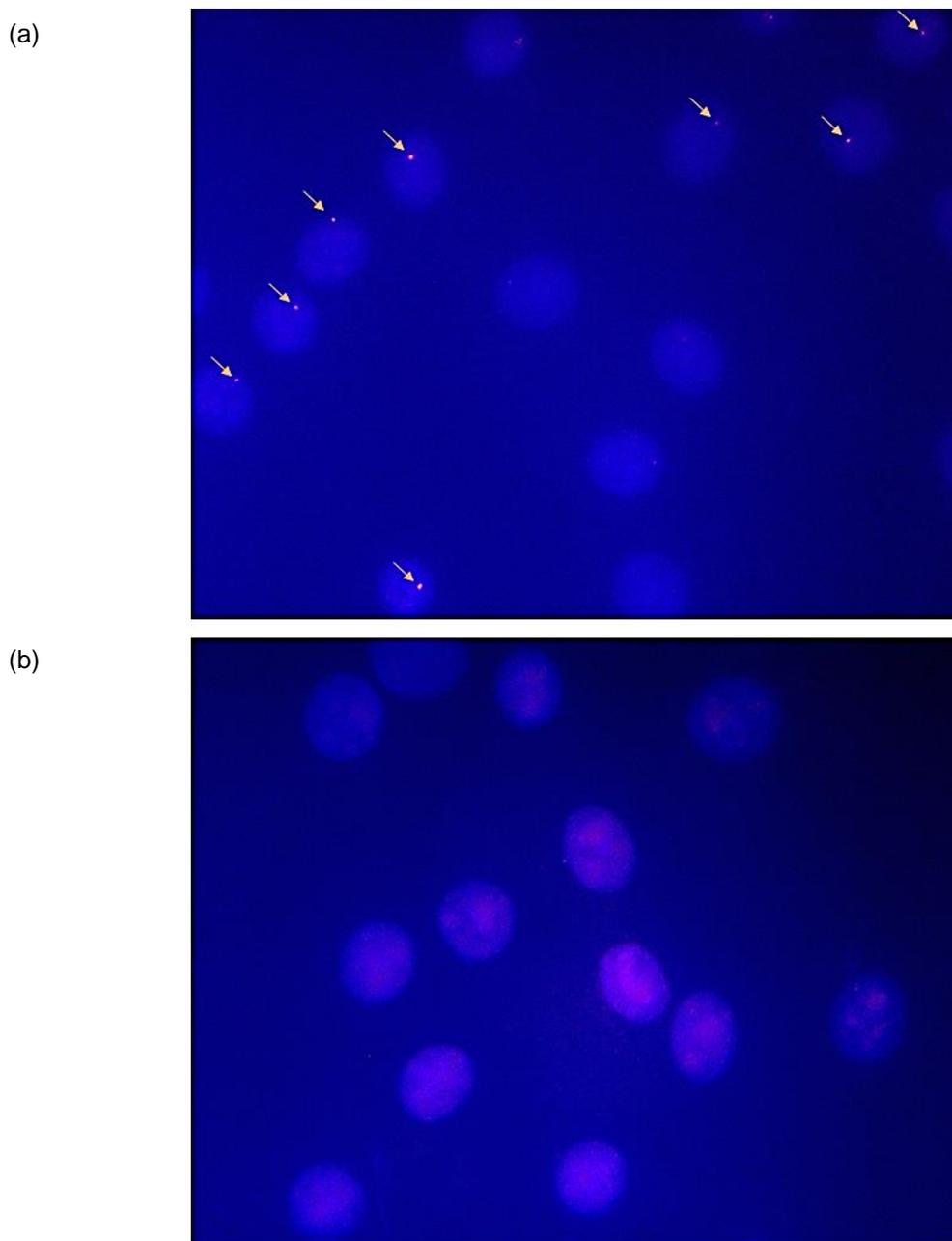
Figure 4.7. Images from nd-FISH experiments using (a) nicked Invader probe **DYZ-NIP** (1 ng per 200μl of PCR buffer, 3 h, 37.5 °C), or (b) conventional Invader probe **DYZ-REF** (6 ng, 3 h, 37.5 °C). Fixed isolated interphase nuclei from male bovine kidney cells were incubated with probes in a Tris buffer (20 mM Tris-Cl, 100 mM KCl, pH 8.0) and counterstained with DAPI. Images were obtained by overlaying images from Cy3 (red) and DAPI (blue) channels and adjusting the exposure. Nuclei were viewed at 60X magnification using a Nikon Eclipse Ti-S

inverted microscope. The amount of **DYZ-NIP** was chosen based on optimization studies (Fig. 4.42).

4.3 Conclusion

Herein we used nicked Invader probes with central overlap dsDNA Invader region ranging from 1-13 nts and auxiliary/toehold regions ranging from 6-12 nts long. From the thermal denaturation studies, we observe that probes with 9 or 13 nts (**NIP1** and **NIP2**, respectively) central overlapping region hold the main strands to form a stable duplex in the central region of the construct (T_m ~ 37-49 °C), while probes containing shorter overlapping regions 1-nt (**NIP4**) or 5-nt (**NIP3**) have a low T_m ($T_m < 15$ °C), and thus the central dsDNA region holding the nicked construct is too unstable. Because of this instability, these probes lose the advantage gained by nicked construct. Moreover, in the recognition experiments, nicked probes with stable (and longer) overlapping regions (**NIP1**, **NIP2**) displayed enhanced recognition, as suggested by their lower C_{30} values, than the probes with short central overlapping regions (**NIP3**, **NIP4**). Similarly, Analogous toehold probes with shorter central dsDNA overlap (**ON9**:**ON12**, **ON13**:**ON16**) region used herein are inefficient in recognition of complementary target strands in a hairpin dsDNA target while probes with longer toeholds displayed efficient recognition of the dsDNA target (**ON1**:**ON4**/**ON5**:**ON8**). Nicked Invaders displayed excellent mismatch discrimination against mismatched targets **MM1**, **MM2**, and **MM3**, which proves that they are very specific to recognize their targets. Besides, Cy3 labeled **DYZ2** was used to target a region in the Y-chromosome of male bovine cells, where single localized Cy3 signals were observed. When female cell lines were targeted, no signal was observed, hence nicked probe is target-specific. In summary, Nicked Invader probes can recognize the target regions in chromosomal DNA and displayed enhanced target recognition compared to corresponding probes with blunt-ended and toehold architect as evident by greater recognition of hairpin targets, as well as exhibiting excellent mismatch discrimination against singly mismatched non-complementary targets. Nicked Invader probes appear to be a viable method for chromosomal DNA recognition, paving the way for the development of novel tools for molecular biology, genomic engineering, and nanotechnology applications.

4.4 Acknowledgments

4.5 Experimental section

Synthesis and purification of ONs. Modified ONs were synthesized on an automated DNA synthesizer (0.2 µmol scale) using a long chain alkyl amine-controlled pore glass (LCAA-CPG) solid support with a pore size of 500 Å. The corresponding protected phosphoramidites of monomer **U̲**, **A̲**, and, **C̲** were prepared as previously described[15,16] and incorporated into ONs via hand-couplings (0.05 M in acetonitrile, ~50-fold molar excess) using 0.01 M 4,5-dicyanoimidazole as the activator (15 min) and 0.02 M iodine in THF/H_2O/pyridine for extended oxidation (45 s). 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). Subsequent treatment with 32% ammonia (55 °C, 17 h) ensured deprotection and cleavage from the solid support of the crude DMTr-protected ONs, which were purified via ion-pair reverse phase HPLC (XTerra MS C18 column: 0.05 M triethylammonium acetate and acetonitrile gradient) followed by detritylation (80% aq. acetic acid, 20 min) and precipitation (NaOAc, $NaClO_4$, acetone, -18 °C, 16 h). The purities of the synthesized ONs were verified using analytical HPLC (>85% purity, see Figs. 4.26 – 4.28), while identity was verified either by MALDI-MS (using 2,4,6-trihydroxy acetophenone as a matrix) or LC-ESI-MS analysis (Waters/Acquity C18 column; triethylammonium formate and acetonitrile gradient) recorded on a Quadrupole Time-of-Flight (Q-TOF) mass spectrometer (Table 4.3 and Figs. 4.8 - 4.25).

Thermal denaturation experiments. The concentrations of ONs were estimated using the following extinction coefficients (OD_{260}/µmol): G (12.01), A (15.20), T/U (8.40), C (7.05), Cy3 (4.93), and pyrene (22.4).[23] T_ms of duplexes (0.5 µM final concentration of each strand) were determined using a Cary 100 UV/Vis spectrophotometer equipped with a 12-cell Peltier temperature controller and determined 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 having a path-length of 1.0 cm and annealed by heating to 92 °C

(2 min) followed by cooling to the starting temperature of the experiment. A temperature range from either 5 °C or 20 °C to at least 15 °C above the duplex T_m was used, with T_ms determined as the average of at least two experiments within ±1.0 °C (except for DYZ-REF, where only one experiment was conducted). A temperature ramp of 1 °C/min was used in all experiments.

Electrophoretic mobility shift assay. Unmodified DNA strands were obtained from commercial sources and used without further purification. Target strands were DIG-labelled using the 2nd generation DIG Gel Shift Kit (Roche Applied Bioscience). Briefly, 11-digoxigenin-ddUTP was incorporated at the 3′-end of the strand (100 pmol) using a recombinant DNA terminal transferase. The reaction mixture was quenched through the addition of EDTA (50 mM), and then diluted to 100 nM in 2X HEPES buffer and used without further processing. The recognition experiments were conducted essentially as previously reported.[5] Thus, Invader probes (concentration as specified) were annealed (90 °C for 2-3 min, followed by cooling to room temperature) and subsequently incubated with separately pre-annealed DIG-labelled DNA (50 nM final concentration in 1X 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 ± 2 °C for 17 h.

Following incubation, loading dye (6X) was added and the mixtures were then loaded onto 12 % (DNA hairpin targets) non-denaturing TBE-PAGE gels (45 mM tris-borate, 1 mM EDTA; acrylamide:bisacrylamide (19:1)). Mixtures were resolved via electrophoresis, which was performed using constant voltage (70 V) at ~4 °C. Bands were blotted onto positively charged nylon membranes (100 V, 30 min, ~4 °C) and cross-linked through exposure to UV light (254 nm, 5 × 15 W bulbs, 3 min). The membranes were incubated with anti-digoxigenin alkaline phosphatase F_{ab} fragments as recommended by the manufacturer and transferred to a hybridization jacket. Membranes were incubated with the chemiluminescence substrate (CSPD) for 10 min at 37 °C, and chemiluminescence was captured on X-ray films. Digital images of developed X-ray films were obtained using a BioRad ChemiDoc™ MP Imaging system and used for densitometric quantification of the bands. The percentage of dsDNA recognition was calculated as the intensity ratio between the recognition band relative to the total lane. An average of at least three independent experiments is reported along with standard deviations (±). Electrophoretograms may be composite images 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,[24] was used to fit the following equation to the data points: $y = C + A (1 - e^{-kt})$ where C, A and k are constants. The resulting equation was used to calculate C_{30} values by setting $y = 30$ and solving for t.

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) 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 suspension of nuclei. 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.

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 which allowed the suspension 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%.

Fluorescence in situ hybridization experiments. An aliquot of labelling buffer (~200 µL of a solution containing 1-30 ng of Cy3-labeled probes in 1x PCR buffer (10 mM Tris, 50 mM KCl, pH 8.0)) was placed on each slide. Preliminary assay optimization studies revealed that a probe amount of ~1 ng per 200 µL of labeling buffer resulted in the best signal-to-background for nicked Invader probes under non-denaturing conditions (Fig. 4.42). Slides with labelling 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 to capture many nuclei in one image. Images of fluorescently labelled nuclei were captured using a 14-bit CoolSNAP HQ2 cooled CCD camera and processed with the NISElements BR 4.20 software. The percentage of nuclei presenting representative signals (i.e., signal coverage) was estimated by evaluating >50 nuclei per Invader probe.

4.6. Supplementary data

Table 4.3. MS data of ONs used in this study.[a]

ON	Sequence	Calculated m/z (M+H)$^+$	Observed m/z (M±H)$^±$
1[b]	5'- UGC ACA GGT AUA UAT AGG C -3'	6731.0	6731.5
2	5'- CG CAU A -3'	2223.0	2223.5
3	3'- ACG TGU -5'	2254.0	2254.5
4[b]	3'- CCA TAU AUA TCC GGC GTA U -5'	6642.0	6642.5
5	5'- UGC ACA GGT AUA UAT AG -3'	6112.5	6113.0
6	5'- GCC GCA UA -3'	2841.0	2841.5
7	3'- AC GTG UCC -5'	2832.0	2833.5
8	3'- ATA UAU ATC CGG CGT AU -5'	6063.5	6064.5
9	5'- UG CAC AGG TAU AUA T-3'	5470.0	5470.0
10	5'- AG GCC GCA UA -3'	3483.5	3484.0
11	5'- AC GTG UCC AT-3'	3449.5	3449.5
12	3'- A UAU ATC CGG CGT AU -5'	5446.0	5447.0
13	5'- UG CAC AGG TAU AU-3'	4852.5	4853.0
14	5'- A TAG GCC GCA UA-3'	4101.0	4101.5
15	3'- AC GTG UCC ATA U-5'	4283.0	4283.0
16	5'- AU ATC CGG CGT AU -3'	4612.5	4613.0
17[b]	5'-GG TAU AUA TAG GCC GCA UA- 3'	6731.0	6732.0
18[b]	3'-ACG TGU CCA TAU AUA TCC G-5'	6265.5	6266.0
19[b]	5'- GGT AUA UAT AGG C -3'	4446.0	4447.5
20[b]	3'- CCA TAU AUA TCC G -5'	4326.0	4327.0
21[c]	5'-Cy3-TUA UAT GCT GUT CTC-3'	5680.0	5678.0
22[c]	3'- AAU AUA CGA CAA GAG-Cy3-5'	5789.0	5787.0
23[*]	5'- Cy3-T GUG TUA TAU GCU GTT C	6570.0	6570.0
24[*]	5'- TC AGC CCT-3'	2806.0	2806.0
25[*]	3'- A CAC AAU A	2833.0	2833.0
26[*]	3'- TA CGA CAA GAG UCG GGA-Cy3-5'	6683.0	6681.5

[a] MALDI-MS used for all ONs except **23**-**26**[*] (LC-ESI-MS). For MALDI-MS, the m/z is for the (M+H)$^+$ ion, while it is for the (M-H)$^-$ ion with LC-ESI-MS.
[b] Data for **1**, **4** and **17**-**20** previously reported in reference 17.[27]
[c] Data for **21** and **22** previously reported in reference 22.[22]
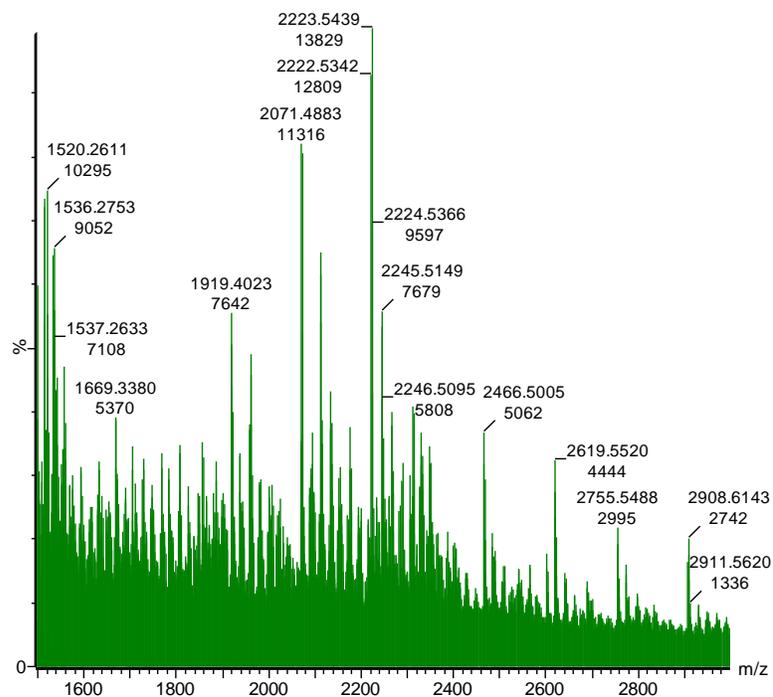
Figure 4.8. MALDI-MS spectrum of **ON2**.
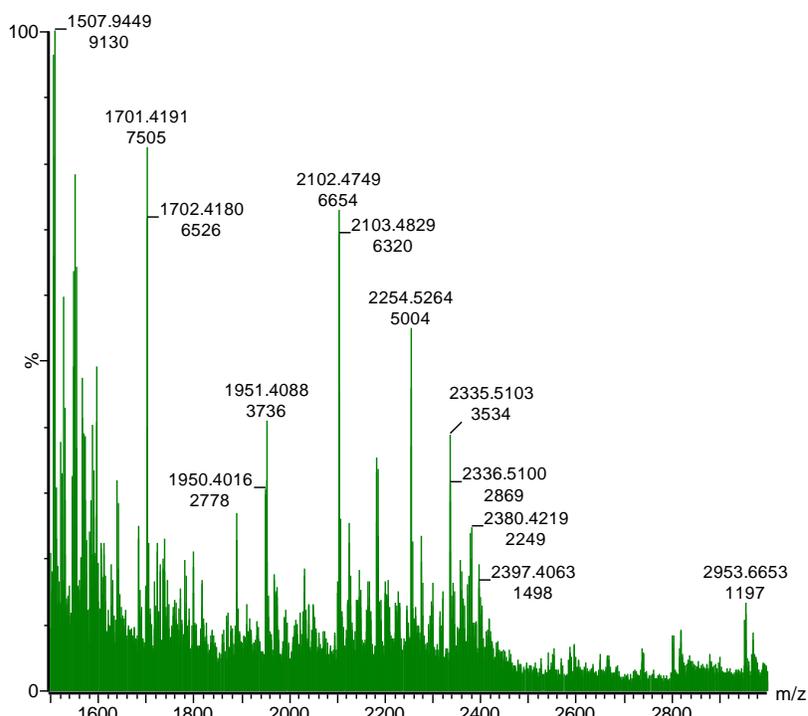
Figure 4.9. MALDI-MS spectrum of **ON3**.
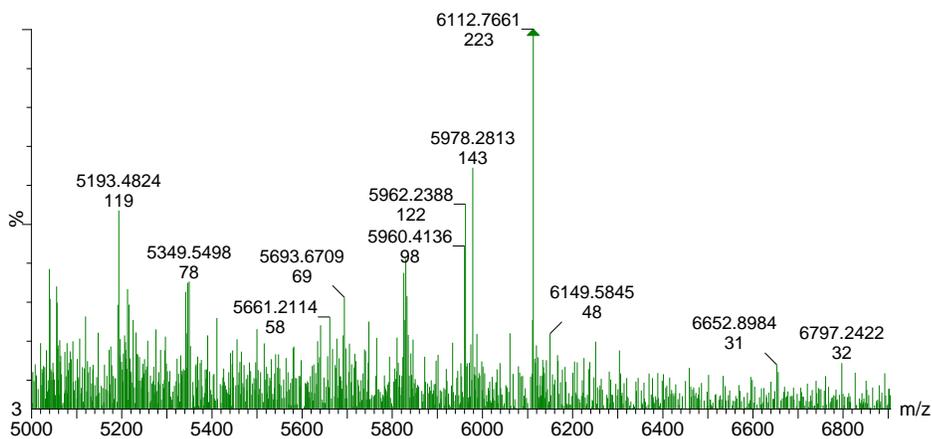
Figure 4.10. MALDI-MS spectrum of **ON5**.
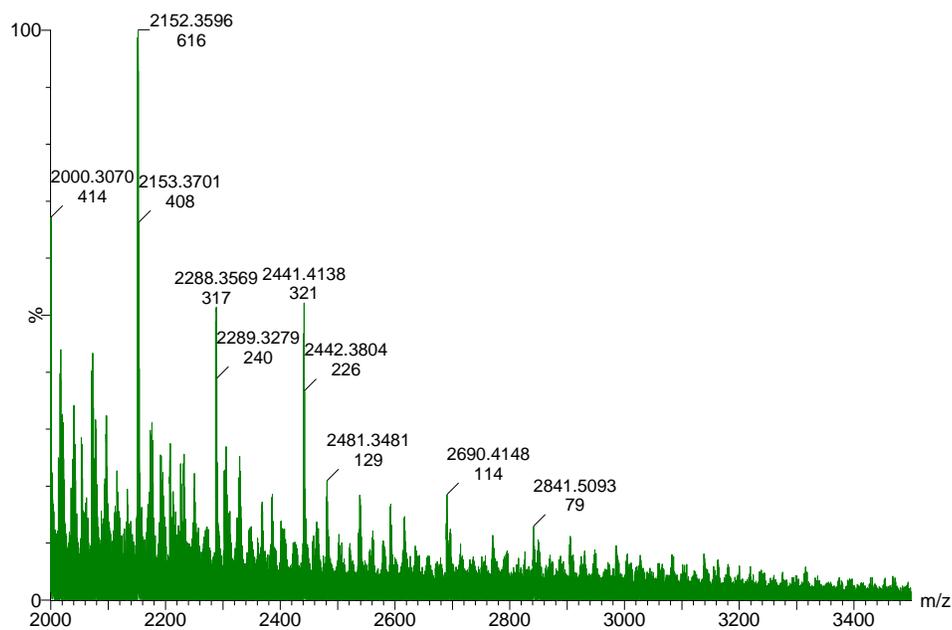
Figure 4.11. MALDI-MS spectrum of **ON6**.
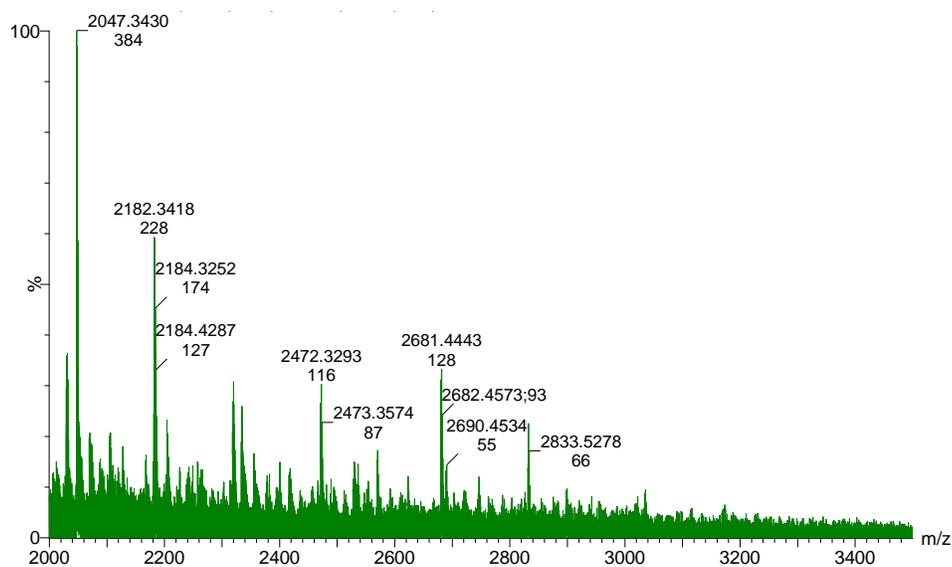
Figure 4.12. MALDI-MS spectrum of **ON7**.
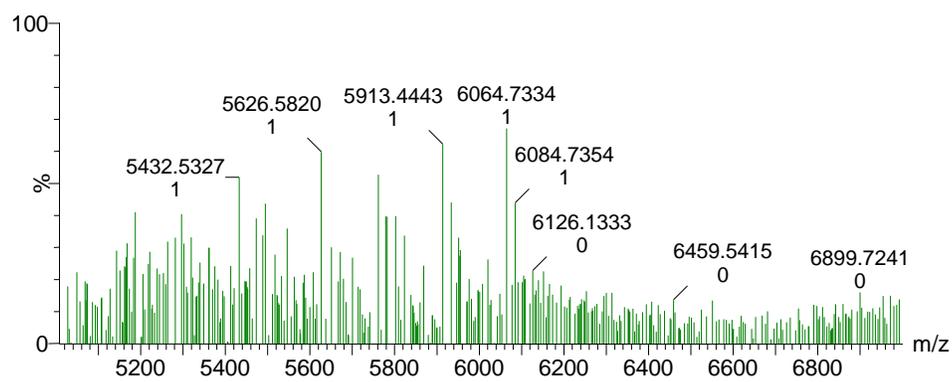
Figure 4.13. MALDI-MS spectrum of **ON8**.
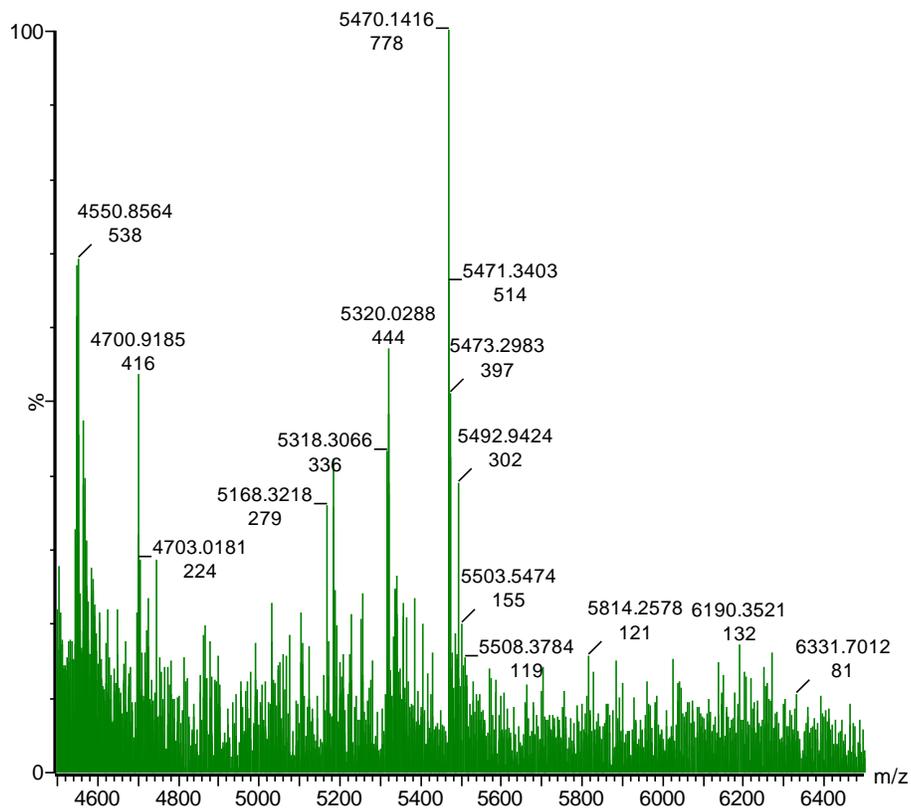
Figure 4.14. MALDI-MS spectrum of **ON9**.
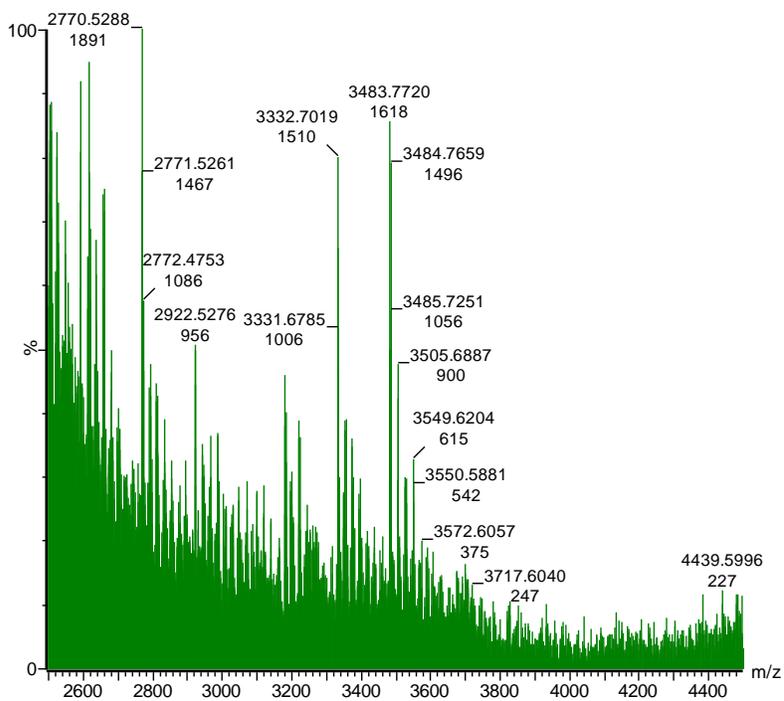
Figure 4.15. MALDI-MS spectrum of **ON10**.
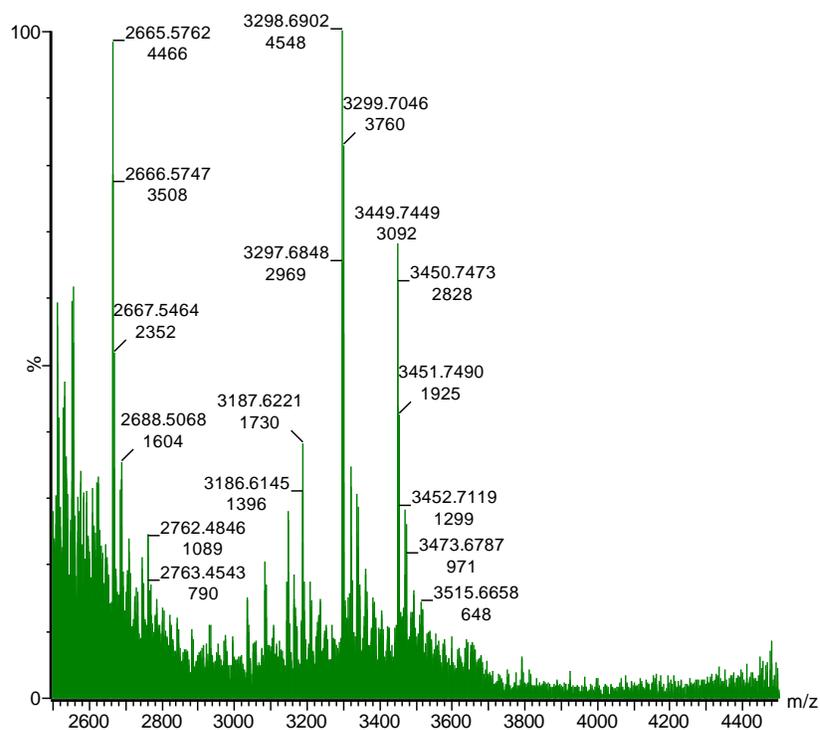
Figure 4.16. MALDI-MS spectrum of **ON11**.
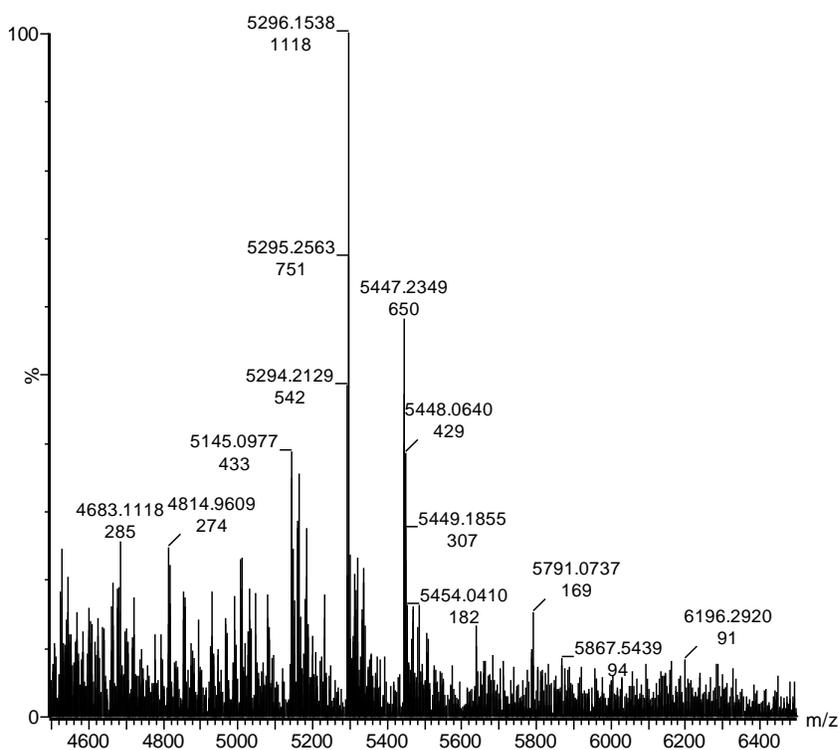
Figure 4.17. MALDI-MS spectrum of **ON12**.
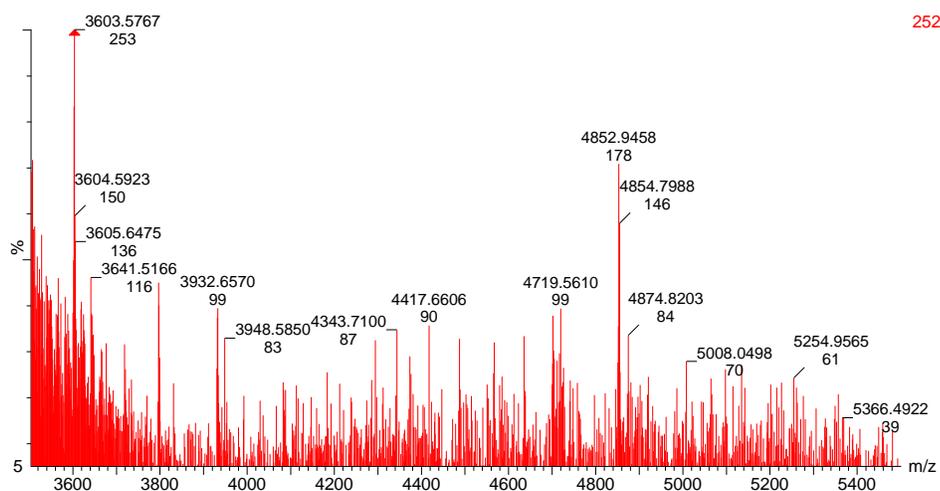
Figure 4.18. MALDI-MS spectrum of **ON13**.
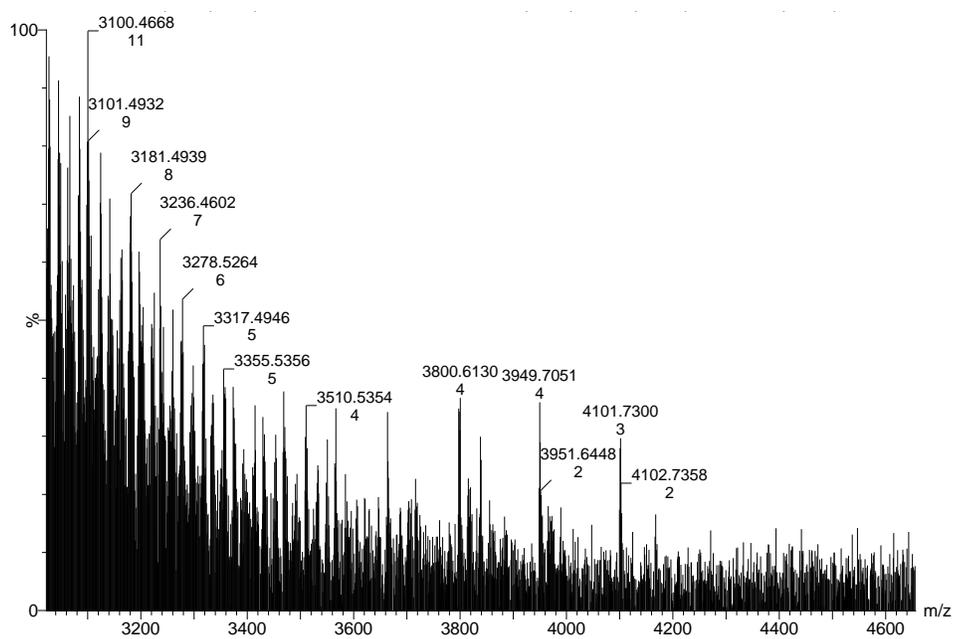
Figure 4.19. MALDI-MS spectrum of **ON14**.
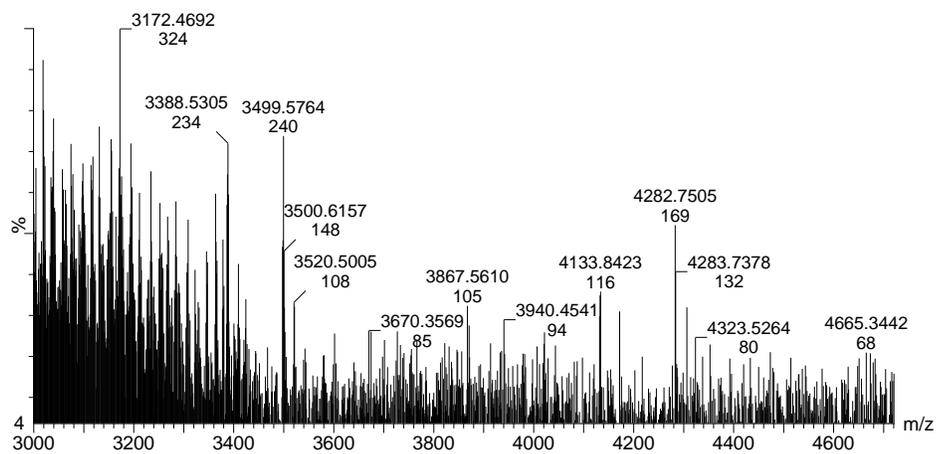
Figure 4.20. MALDI-MS spectrum of **ON15**.
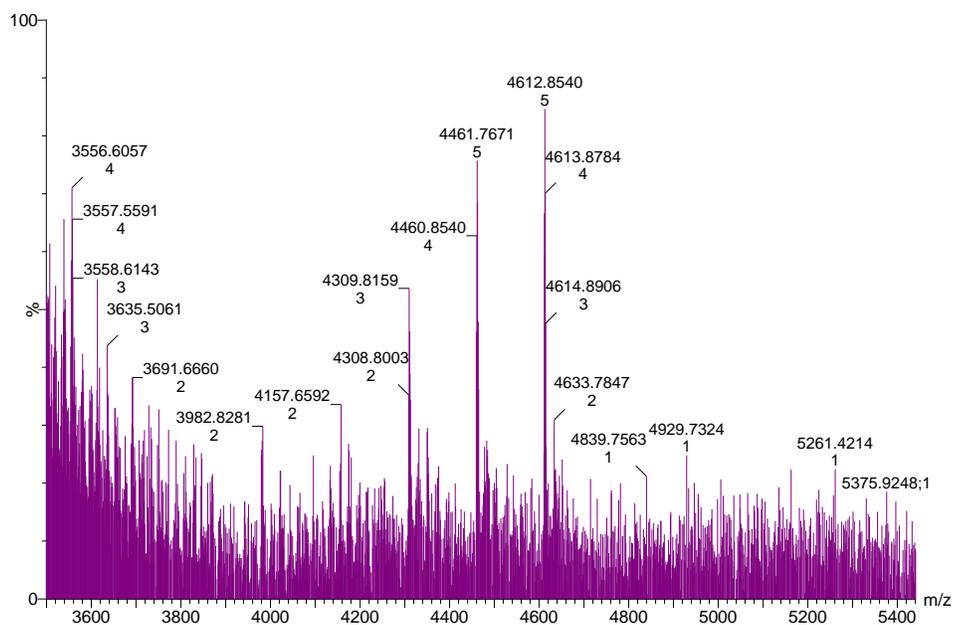
Figure 4.21. MALDI-MS spectrum of **ON16**.
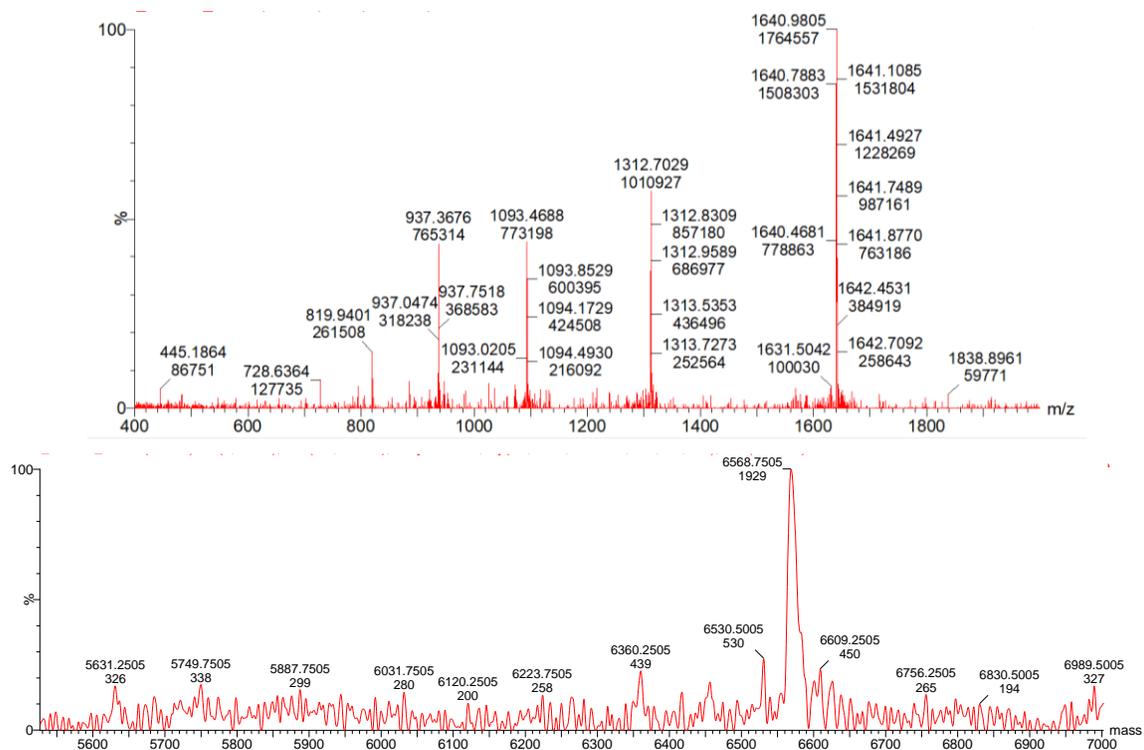
Figure 4.22. Unprocessed (upper panel) and deconvoluted (lower panel) ESI-MS spectrum of **ON23**.
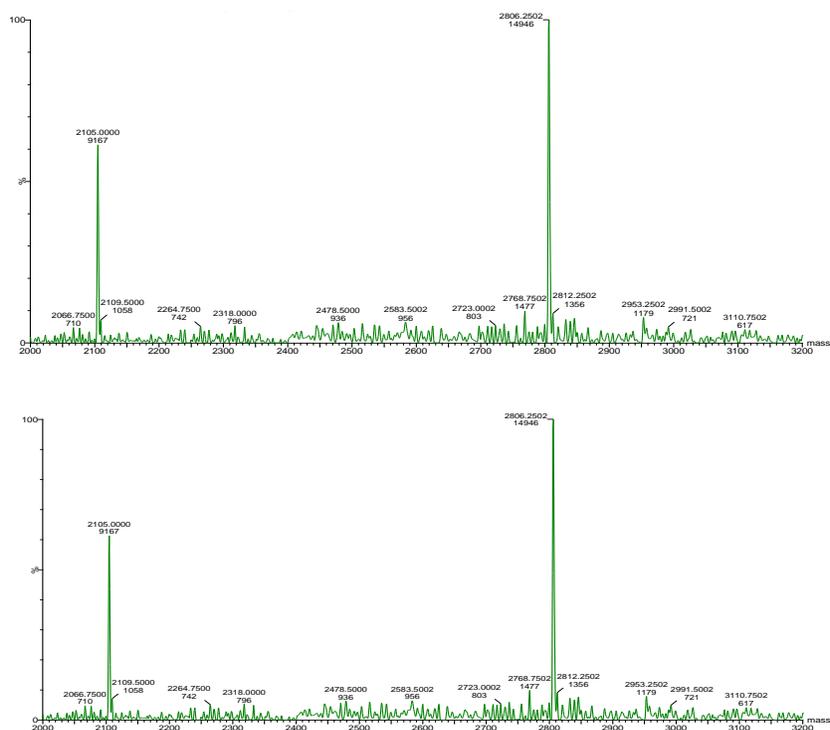
Figure 4.23. Unprocessed (upper panel) and deconvoluted (lower panel) ESI-MS spectrum of **ON24**
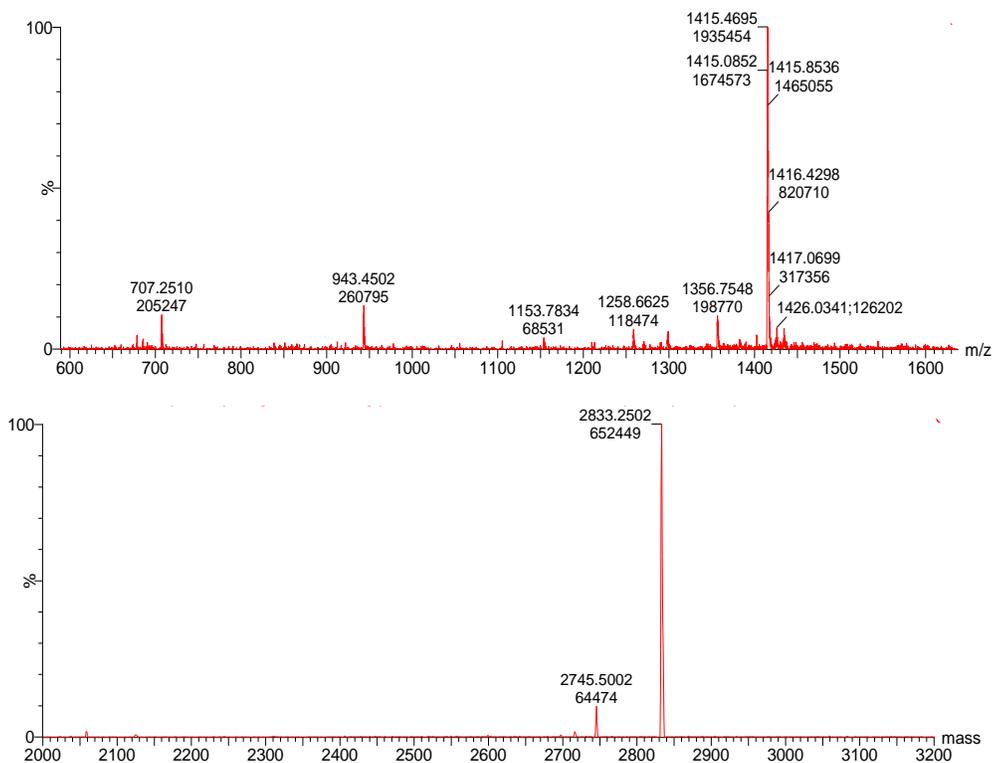
Figure 4.24. Unprocessed (upper panel) and deconvoluted (lower panel) ESI-MS spectrum of **ON25**.
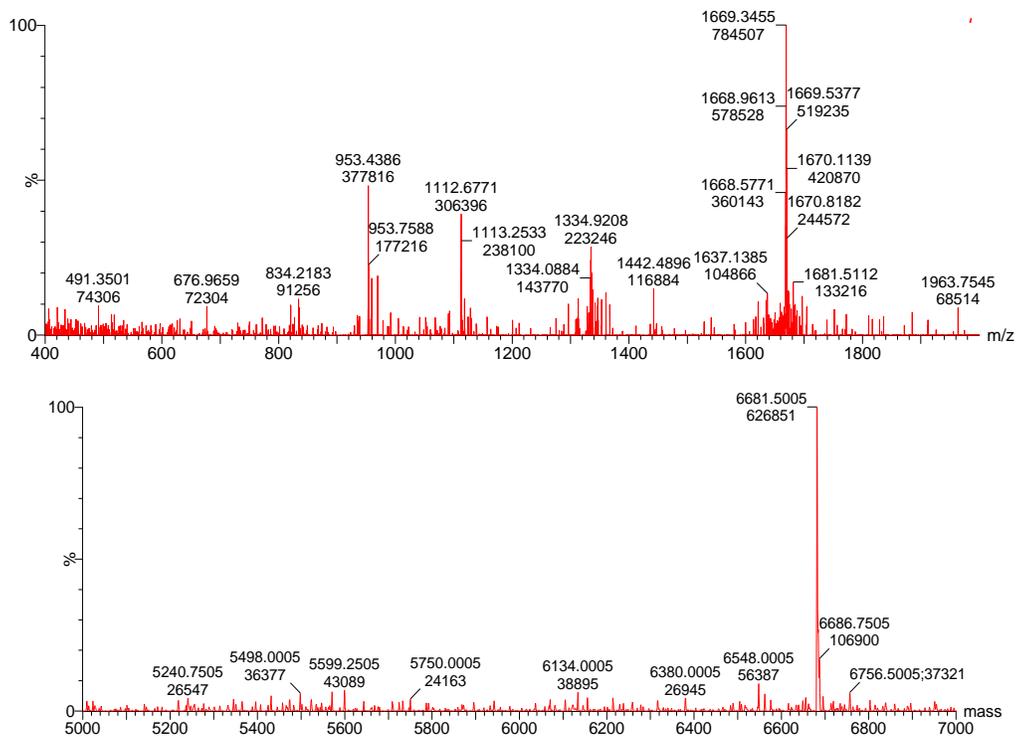
Figure 4.25. Unprocessed (upper panel) and deconvoluted (lower panel) ESI-MS spectrum of **ON26**.

ON2

ON3

ON5

ON6

ON7

ON8

ON9

ON10
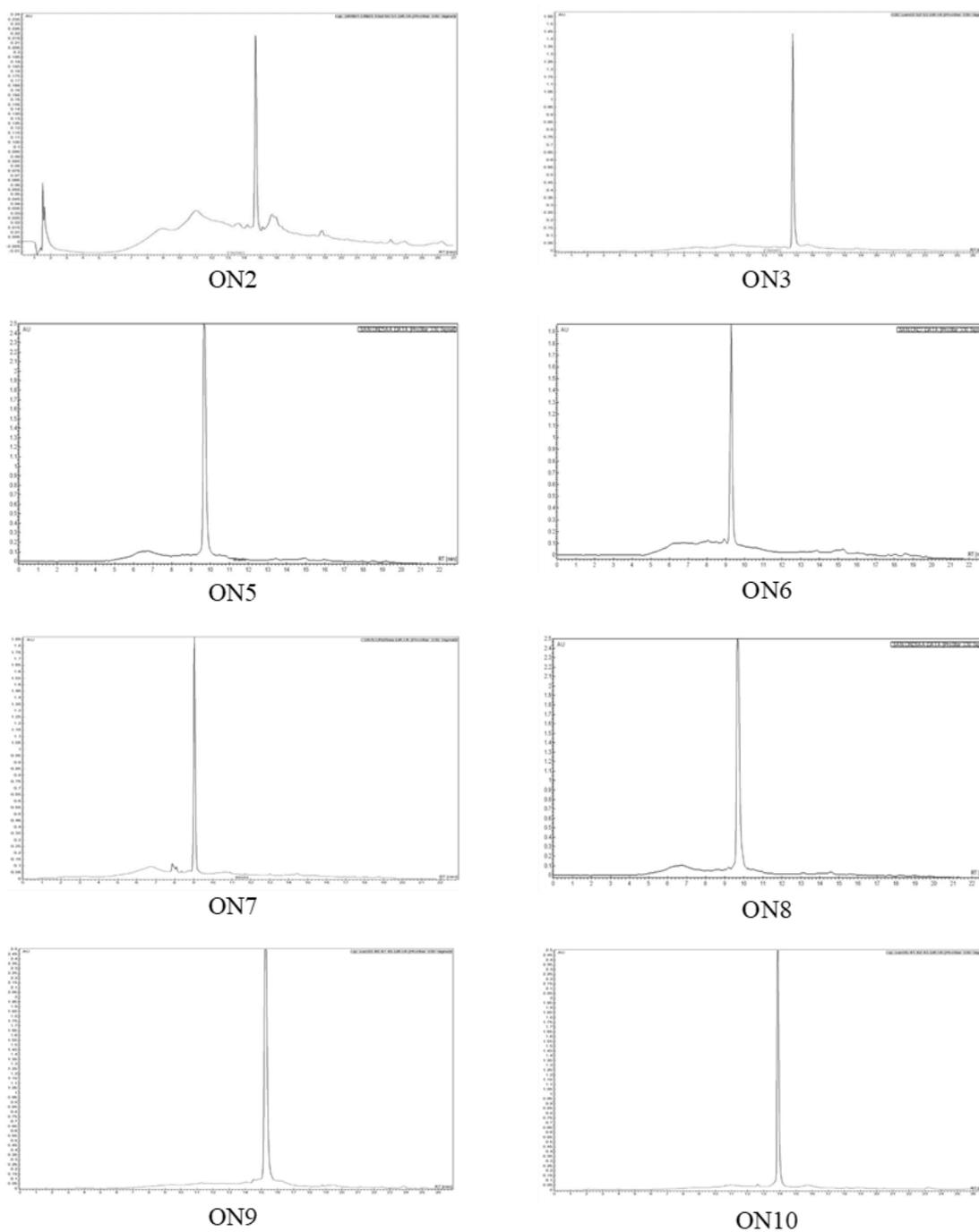
Figure 4.26. HPLC traces of **ONs** used in this study.

ON11

ON12

ON13

ON14

ON15

ON16
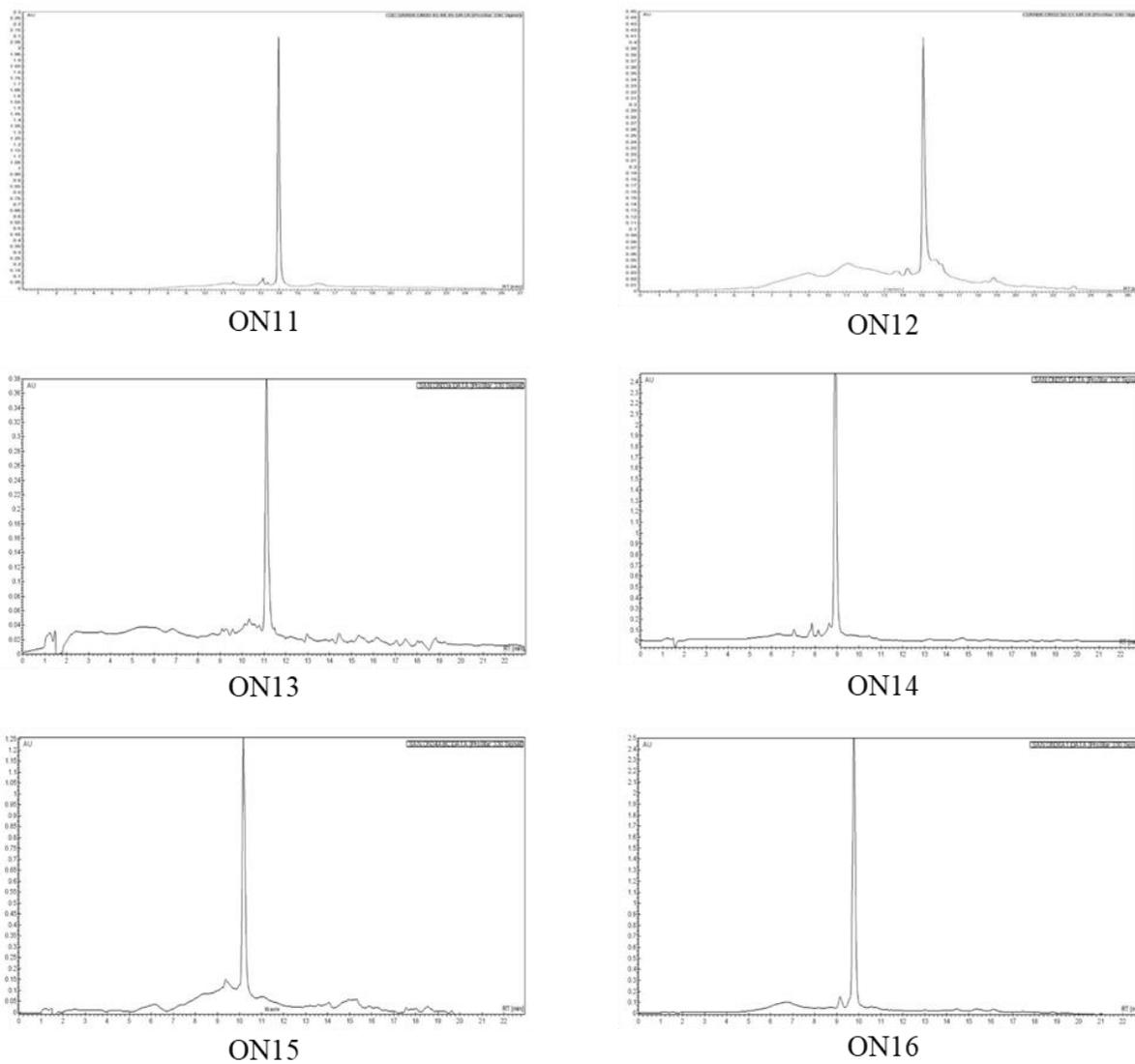
Figure 4.27. HPLC traces of **ONs** used in this study.
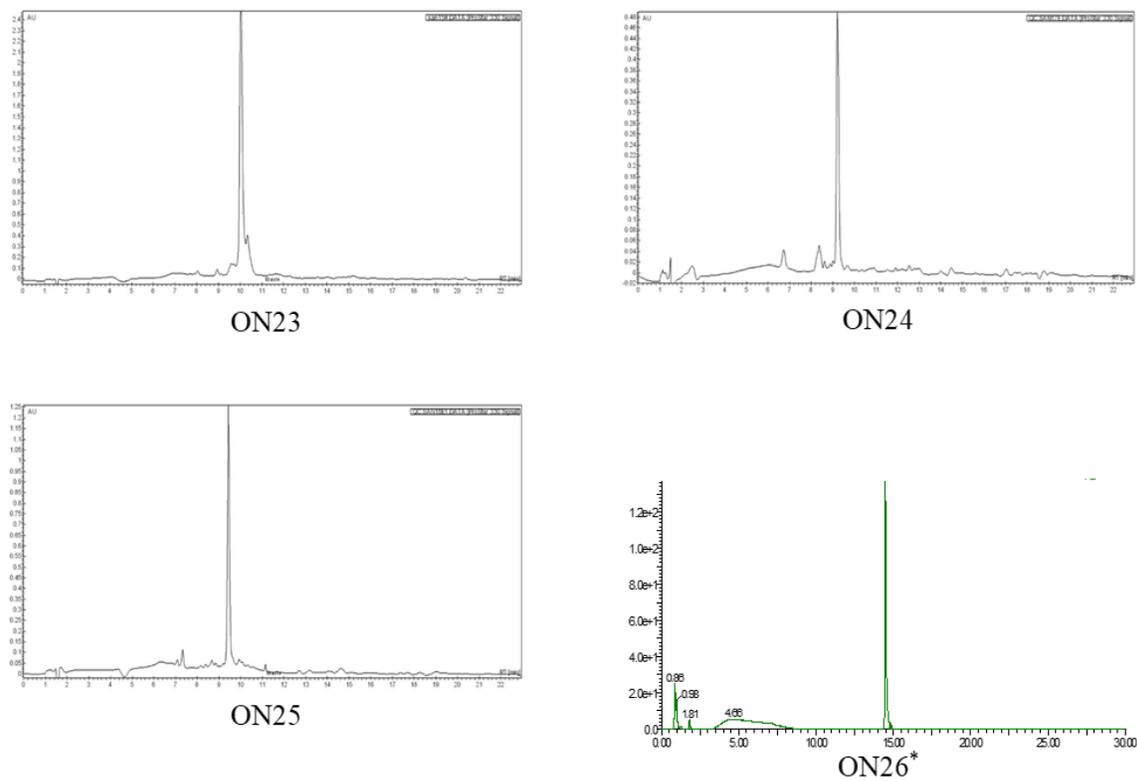
Figure 4.28. HPLC traces of **ONs** used in this study. *Chromatogram for **ON26** obtained from LC-MS.
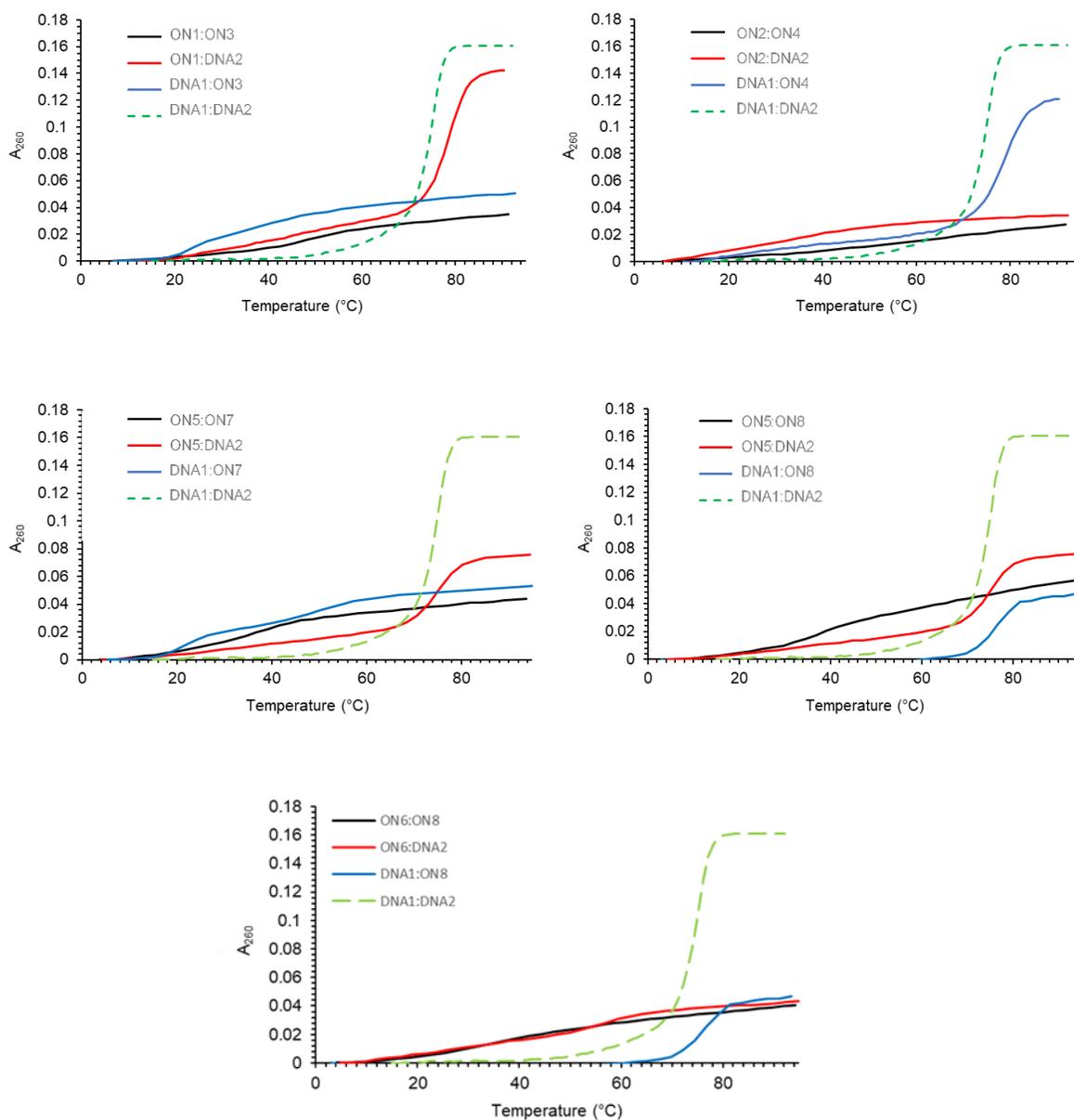
Figure 4.29. Representative thermal denaturation curves of double-stranded probes, duplexes between individual probe strands and 33-mer ssDNA targets harboring complementary regions, and unmodified reference duplex **DNA1**:**DNA2**. For experimental conditions, see Table 4.1.
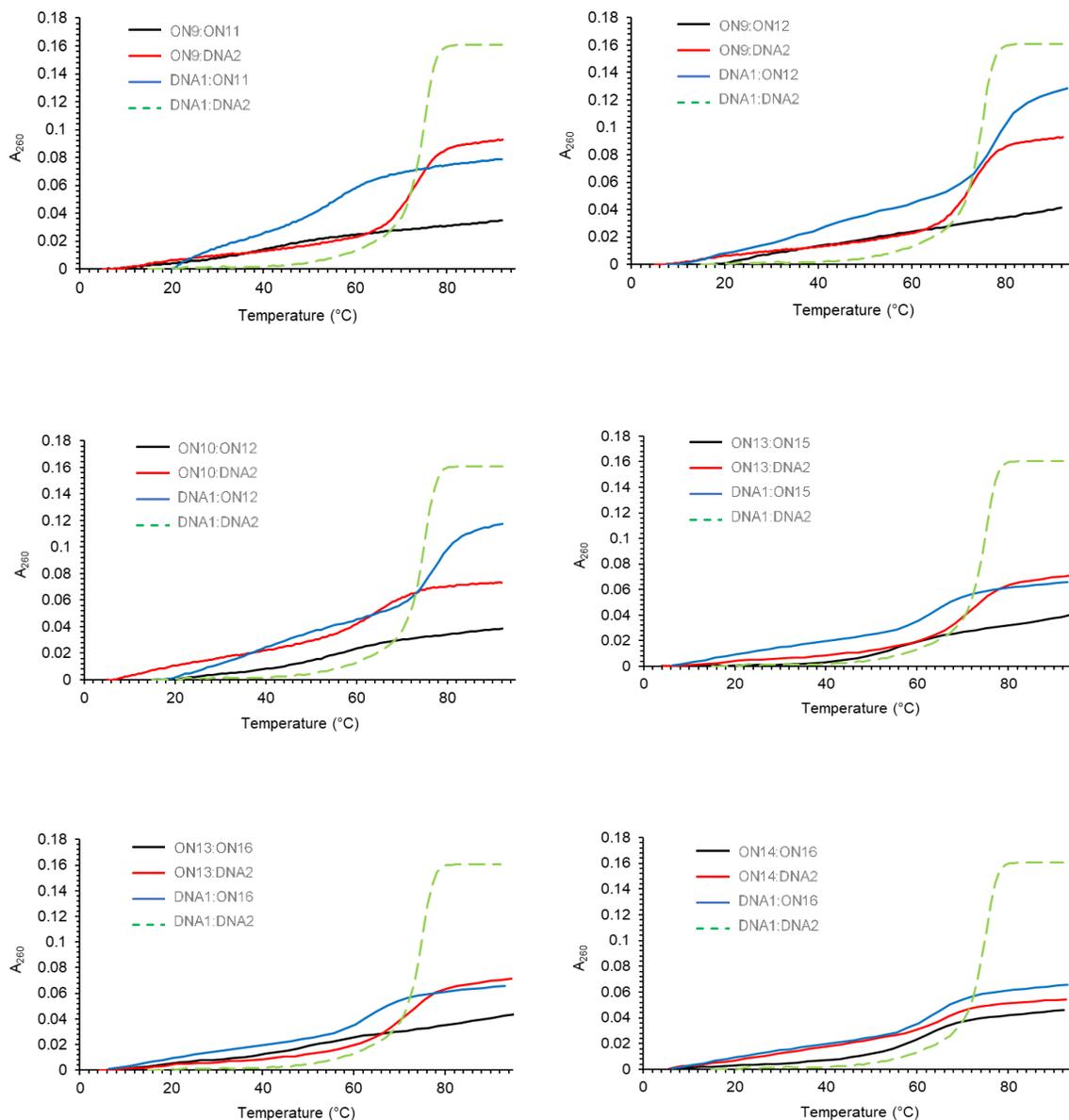
Figure 4.30. Representative thermal denaturation curves of double-stranded probes, duplexes between individual probe strands and 33-mer ssDNA targets harboring complementary regions, and unmodified reference duplex **DNA1:DNA2**. For experimental conditions, see Table 4.1.
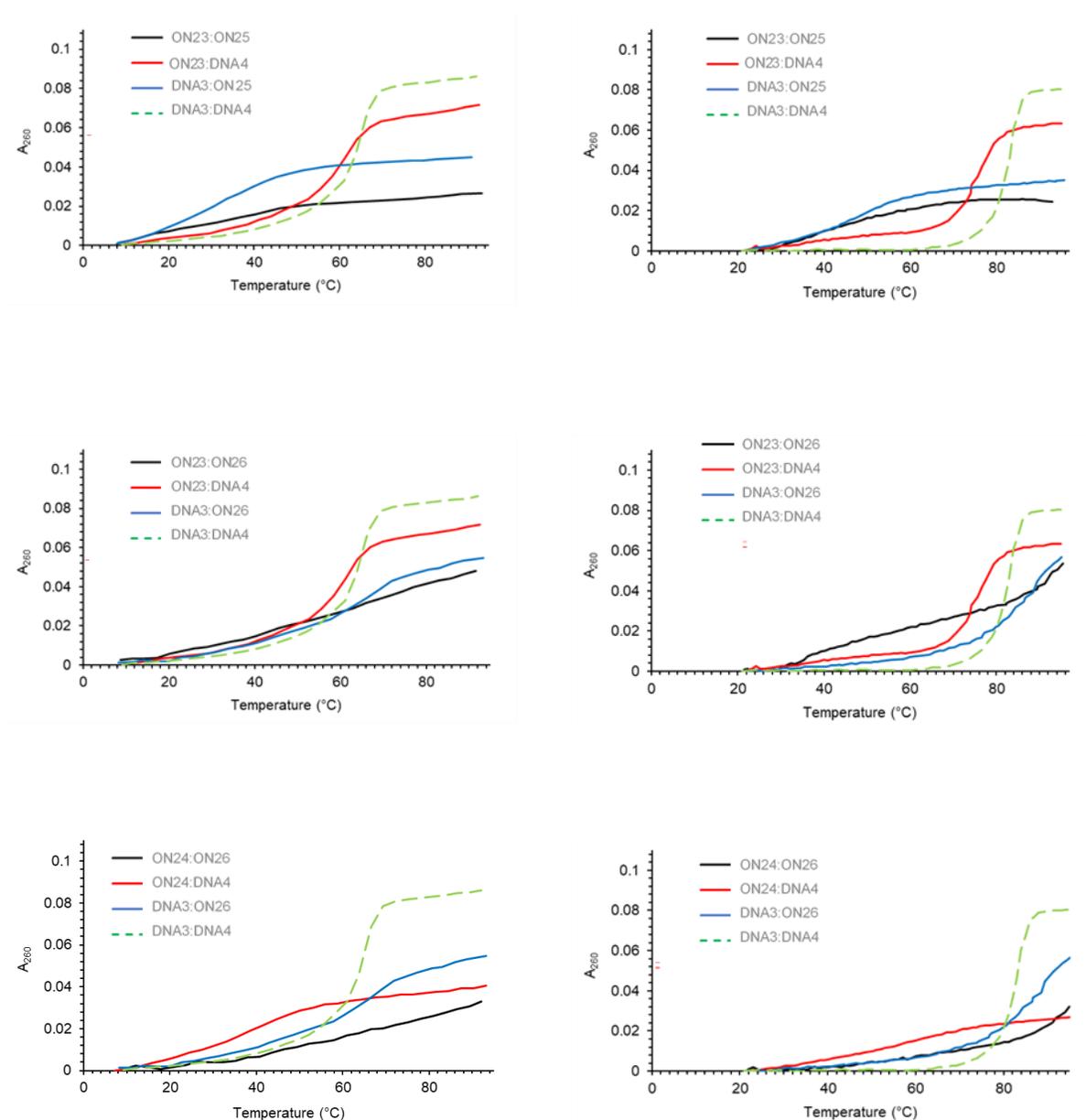
Figure 4.31. Representative thermal denaturation curves of double-stranded probes (used in FISH experiments), duplexes between individual probe strands and 33-mer ssDNA targets harboring complementary regions, and unmodified reference duplex **DNA3**:**DNA4** recorded in medium (left) or high salt buffer (right). For experimental conditions, see Table 4.1.
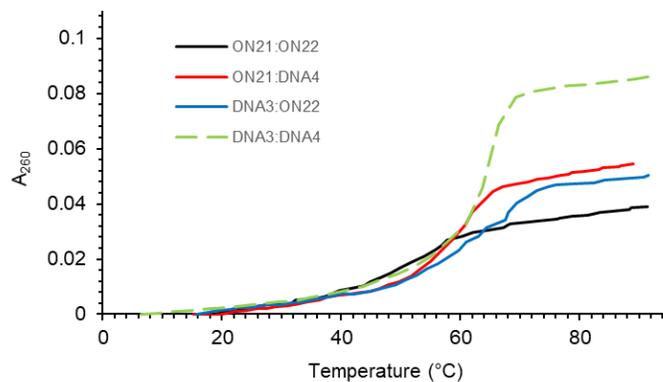
Figure 4.32. Representative thermal denaturation curves of double-stranded probes **DYZ-REF** (used in FISH experiments), duplexes between individual probe strands and 33-mer ssDNA targets harboring complementary regions, and unmodified reference duplex **DNA3**:**DNA4** recorded in medium salt buffer. For experimental conditions, see Table 4.1.

Supplemental discussion regarding **ON1:ON3** *and* **ON1:ON4**. While evaluating the stability of the various duplex segments of **NIP1**, we observed a transition for **ON1**:**ON3** at a temperature of ~45 °C, which, based on our prior experience with these type of probes, seemed unlikely to stem from melting of a 6-bp duplex with two energetic hotspots; the single-stranded overhangs were expected only to have a minor influence on the T_m value.[17] Indeed, when recording the thermal denaturation profile for **ON1** in absence of **ON3**, a transition at a similar temperature was observed leading us to suggest that the ~45 °C transition likely is due to denaturation of a secondary structure that only entails **ON1** (Fig. 4.33). No other transitions were observed for **ON1**:**ON3**, which is why we list a T_m value of <15 °C for that duplex.

ON1:**ON4**, in turn, displays a transition at 49 °C, which is close to the T_m value observed for **ON1**-only (Fig. 4.33). To determine if the 49 °C transition is due to duplex formation between **ON1** and **ON4** rather than denaturation of a secondary structure entailing only **ON1**, we additionally recorded denaturation curves for **ON1:ON4** at a 10-fold higher concentration, i.e., 5 µM. A sharply increased T_m value (~57 °C, Fig. 4.33) was observed, consistent with bimolecular duplex formation between **ON1** and **ON4**. Additional support for duplex formation between **ON1** and **ON4** was obtained from EMSA experiments conducted at non-denaturing conditions as evidenced by the formation of slower moving bands when **ON1** and **ON4** were mixed (Fig. 4.34).

(a)   (b)
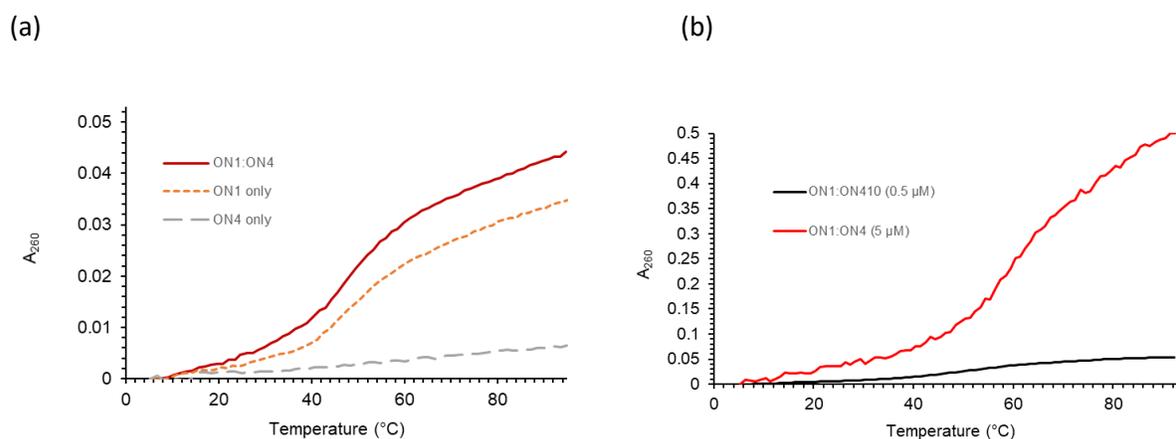
Figure 4.33. (a) Representative thermal denaturation curves of single-stranded probes and double-stranded Invader probes in medium salt buffer. All experiments were performed using

0.5 µM concentration of each strand. (b) Concentration dependent thermal denaturation curve of **ON1** and **ON4**. For experimental conditions, see Table 4.1.
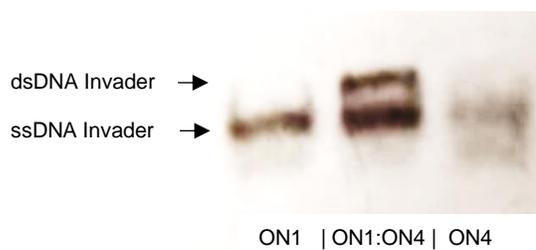
Figure 4.34. Representative gel electrophoretograms from non-denaturing PAGE runs aiming to determine if **ON1** forms a duplex with **ON4** in HEPES buffer. Experimental conditions are as outlined in Fig. 4.3.

Supplemental discussion regarding secondary structure of DNA1. While evaluating the Tm curves, we noticed that the probe-target duplexes, **DNA1:ON5** and **DNA3:ON9** displayed two transitions, which is likely because of some secondary structure formed by 33-mer strand DNA1. While analyzing the target strand using oligoanalyzer tool (IDT), we noticed that **DNA1** forms a hairpin that melts at ~25 °C. Experimentallly, **DNA1** displayed a transition at <30 °C (see Fig. 4.35 (a), and (b)), which is in agreement with the observation using oligoanalyzer tool. Thus, the lower transition observed in the probe-target duplexes are due to the melting of this secondary structure formed by **DNA1,** and the actual transition temperature of the duplex is the higher T_m.

Meanwhile, at the same temperature range, we observed a transition in probe-traget duplex, **DNA1:ON3** at ~ 25 °C (Fig. 4.35 (c)). We estimated a differential thermal denaturation curve diff (**DNA1:ON3** - **DNA1**) and found that **DNA:ON3** forms a labile duplex ($T_m > 25$ °C).
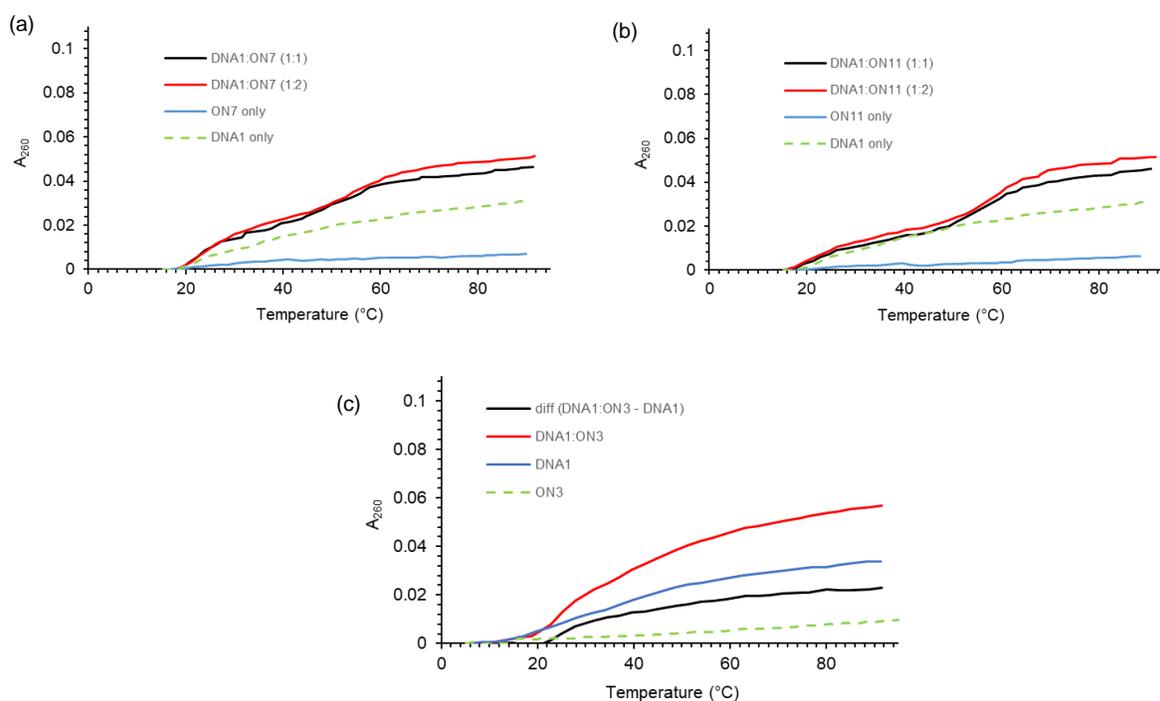
Figure 4.35. Representative thermal denaturation curves of single-stranded target **DNA1** and probe-target duplexes in medium salt buffer formed by (a) ON5, (b) ON9, and (c) ON3. For experimental conditions, see Table 4.1. In (a) and (b), 1:1 represents the equimolar concentration of both strands while 1:2 represents 2-fold molar concentration was used of the latter.

Table 4.4. Sequences of DNA hairpin targets used in this study.

DNA hairpin target	Sequence
DH1	5'- AAG CTG CAC AGG TAT ATA TAG GCC GCA TAT GCA 3'- TTC GAC GTG TCC ATA TAT ATC CGG CGT ATA CGT T_{10}
MM1	5'- AAG CTG CAC AGG TAT **T**TA TAG GCC GCA TAT GCA 3'- TTC GAC GTG TCC ATA **A**AT ATC CGG CGATATA CGT T_{10}
MM2	5'- AAG CTG **G**AC AGG TAT ATA TAG GCC GC**T** TAT GCA 3'- TTC GAC **C**TG TCC ATA TAT ATC CGG CG**A** ATA CGT T_{10}
MM3	5'- AAG CTG **G**AC AGG TAT **T**TA TAG GCC GC**T** TAT GCA 3'- TTC GAC **C**TG TCC ATA **A**AT ATC CGG CG**A** ATA CGT T_{10}
DH2	5'- ACT GTG TGT TAT ATG CTG TTC TCA GCC CTA CTG 3'- TGA CAC ACA ATA TAC GAC AAG AGT CGG GAT GAC T_{10}
DH2-MM	5'- ACT GTG TG**A** TAT ATG **G**TG TTC TCA G**G**C CTA CTG 3'- TGA CAC AC**T** ATA TAC **C**AC AAG AGT C**C**G GAT GAC T_{10}

The yellow highlights indicate the position of mismatched base-pairs relative to the probes.

Table 4.5. Quantification of **DH1**-recognition when using a 50-fold molar excess of different Invader probes.[a]

Probe	Construct	Sequence	Rec. (%)
NIP1	6-13-6 nicked Invader	5'- UGC ACA GGT AUA UAT AGG CCG CAU A -3' 3'- ACG TGU CCA TAU AUA TCC GGC GTA U -5'	~30%
NIP2	8-9-8 nicked Invader	5'- UGC ACA GGT AUA UAT AGG CCG CAU A -3' 3'- ACG TGU CCA TAU AUA TCC GGC GTA U -5'	~40%
NIP3	10-5-10 nicked Invader	5'- UGC ACA GGT AUA UAT AGG CCG CAU A -3' 3'- ACG TGU CCA TAU AUA TCC GGC GTA U -5'	~25%
NIP4	12-1-12 nicked Invader	5'- UGC ACA GGT AUA UAT AGG CCG CAU A -3' 3'- ACG TGU CCA TAU AUA TCC GGC GTA U -5'	~15%
ON1:**ON4**	6-13-6 toe. Invader	5'- UGC ACA GGT AUA UAT AGG C -3' 3'- CCA TAU AUA TCC GGC GTA U -5'	~20%
ON5:**ON8**	8-9-8 toe. Invader	5'- UGC ACA GGT AUA UAT AG -3' 3'- A TAU AUA TCC GGC GTA U -5'	<15%
ON9:**ON12**	10-5-10 toe. Invader	5'- UGC ACA GGT AUA UAT -3' 3'- AU AUA TCC GGC GTA U -5'	<5%
ON13:**ON36**	12-1-12 toe. Invader	5'- UGC ACA GGT AUA U -3' 3'- AUA TCC GGC GTA U -5'	<5%
ON19:**ON20**	13-mer conv. Invader	5'- GGT AUA UAT AGG C -3' 3'-CCA TAU AUA TCC G-5'	<5%
ON17:**ON4**	19-mer conv. Invader	5'- GGT AUA UAT AGG CCG CAU A -3' 3'- CCA TAU AUA TCC GGC GTA U -5'	<15%
ON1:**ON18**	19-mer conv. Invader	5'- UGC ACA GGT AUA UAT AGG C -3' 3'- ACG TGU CCA TAU AUA TCC G -5'	<15%

[a]Experiments were performed in triplicate. Corresponding electrophoretograms shown in Fig. 4.3. Rec. = Recognition
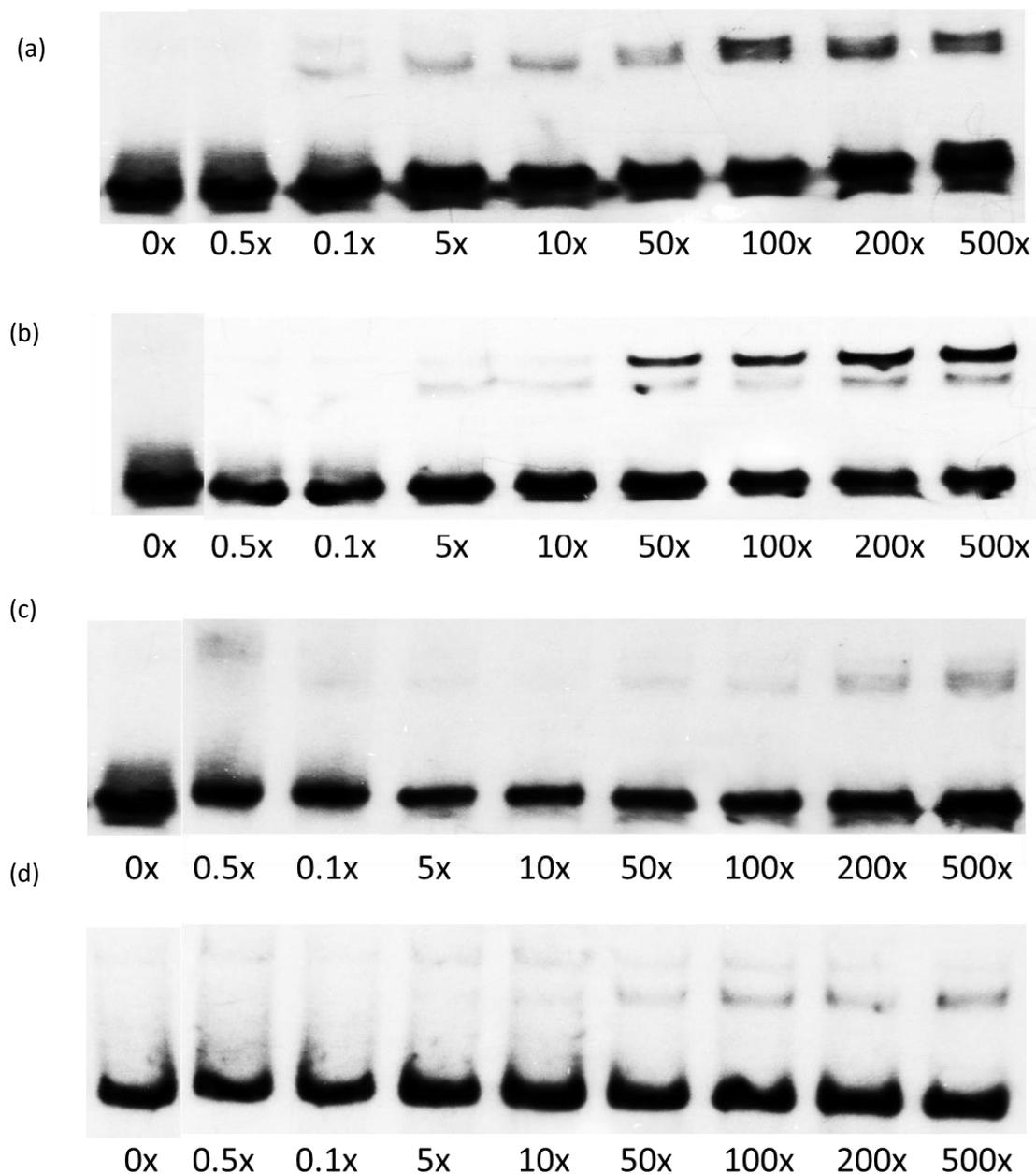
Figure 4.36. Representative gel electrophoretograms from dose-response experiments in which DNA hairpin **DH1** (50 nmol) was incubated with a variable molar excess of different nicked Invader probes and the corresponding toehold probes: (a) **NIP1** (**ON1/2:ON3/4**), (b) **NIP3** (**ON9/10:ON11/12**), (c) **ON1:ON4** and (d) **ON9:ON12** at 37 °C. Experimental conditions are as described in Fig. 4.3.
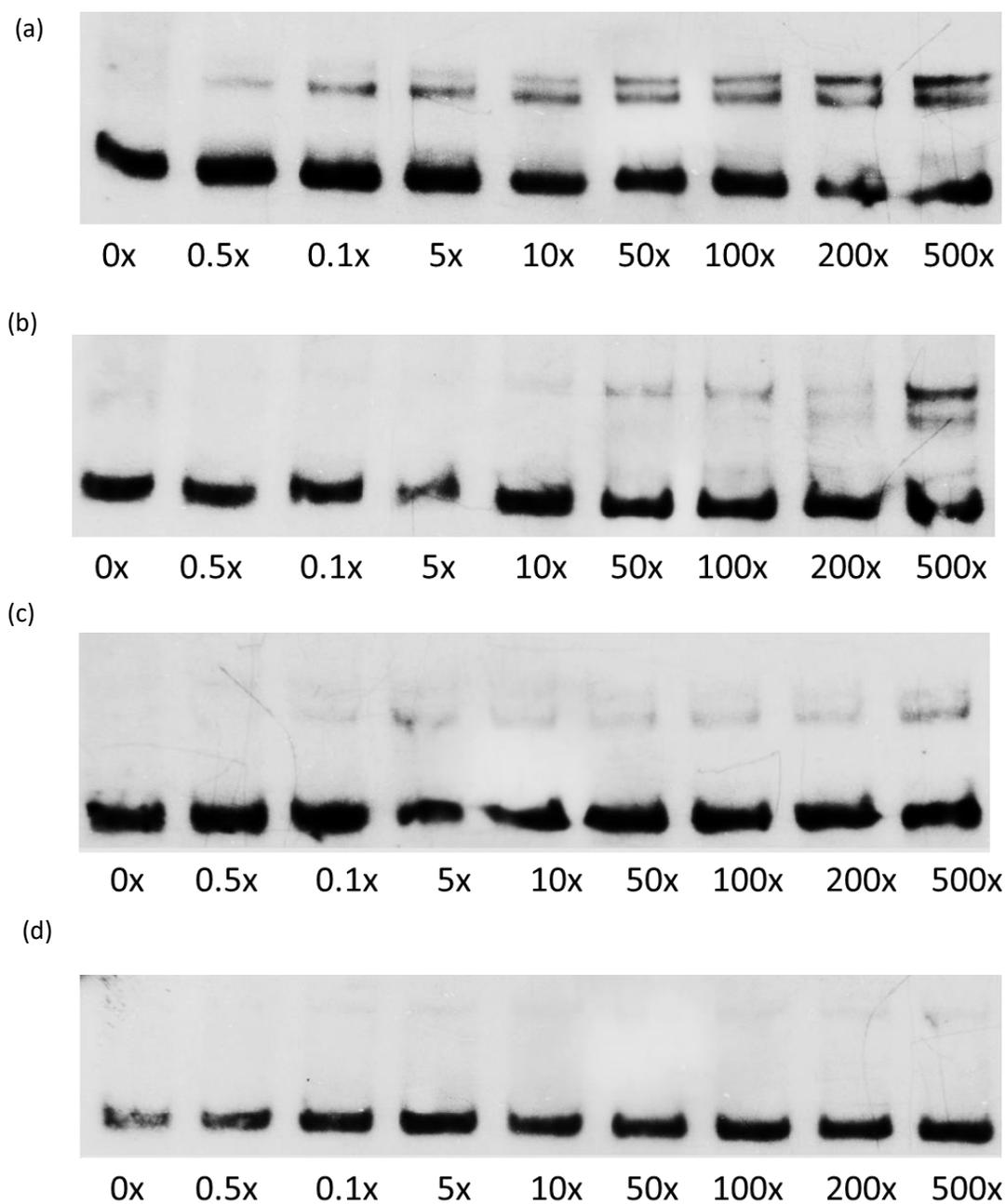
Figure 4.37. Representative gel electrophoretograms from dose-response experiments in which DNA hairpin **DH1** (50 nmol) was incubated with a variable molar excess of different nicked Invader probes and the corresponding toehold probes: (a) **NIP2** (**ON5+6**:**ON7+8**), (b) **NIP3** (**ON13+14**:**ON15+16**) (c) **ON5**:**ON8** and, (d) **ON13**:**ON36** at 37 °C. Experimental conditions are as described in Fig. 4.3.
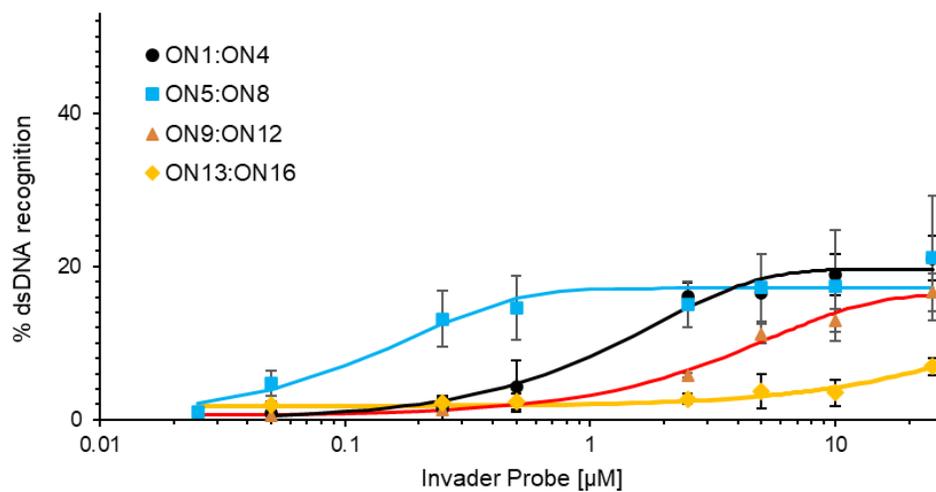
Figure 4.38. Dose-response profiles for recognition of **DH1** by toehold Invader probes. Curves are constructed based on the gel electrophoretograms shown in Figs. 4.36 and 4.37. Experimental conditions are as described in Fig. 4.3.
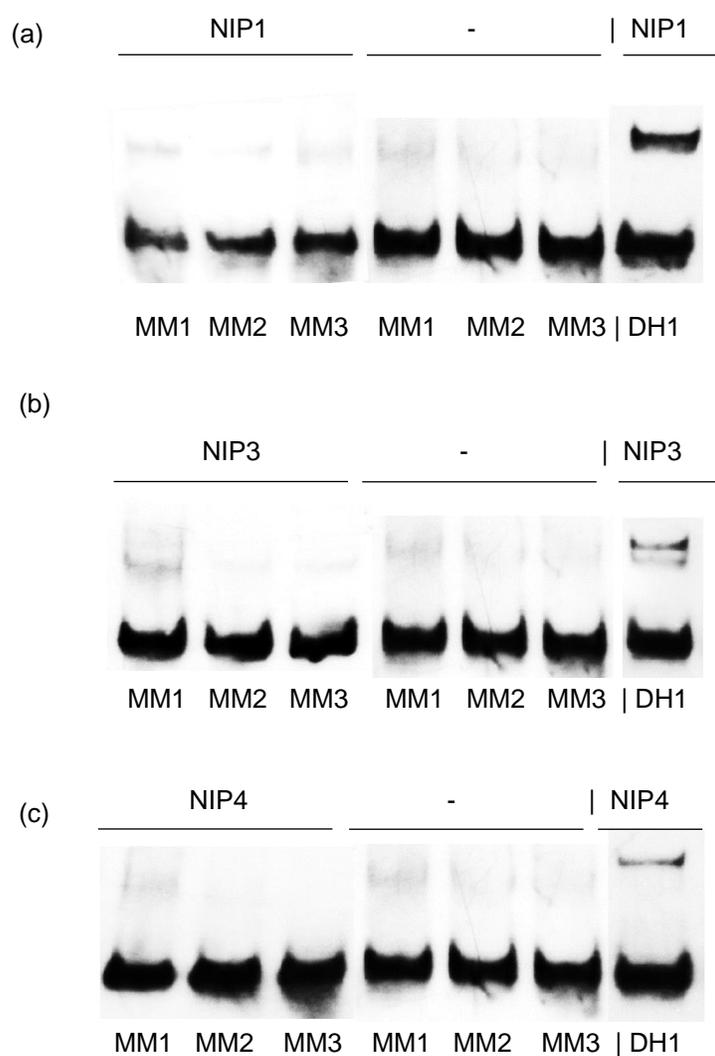
Figure 4.39. Binding specificities of nicked Invader probes. Representative electrophoretograms from experiments in which a 50-fold molar excess of pre-annealed (a) **NIP1**, (b) **NIP3** or (c) **NIP4** was incubated with pre-annealed non-complementary **MM1**-**MM3** targets in HEPES buffer at 37 °C for 17 h as described in Fig. 4.5. For sequences of **MM1**-**MM3**, see Table 4.4. For an illustration of the mismatched recognition complexes that would ensue upon recognition, see Fig. 4.40.
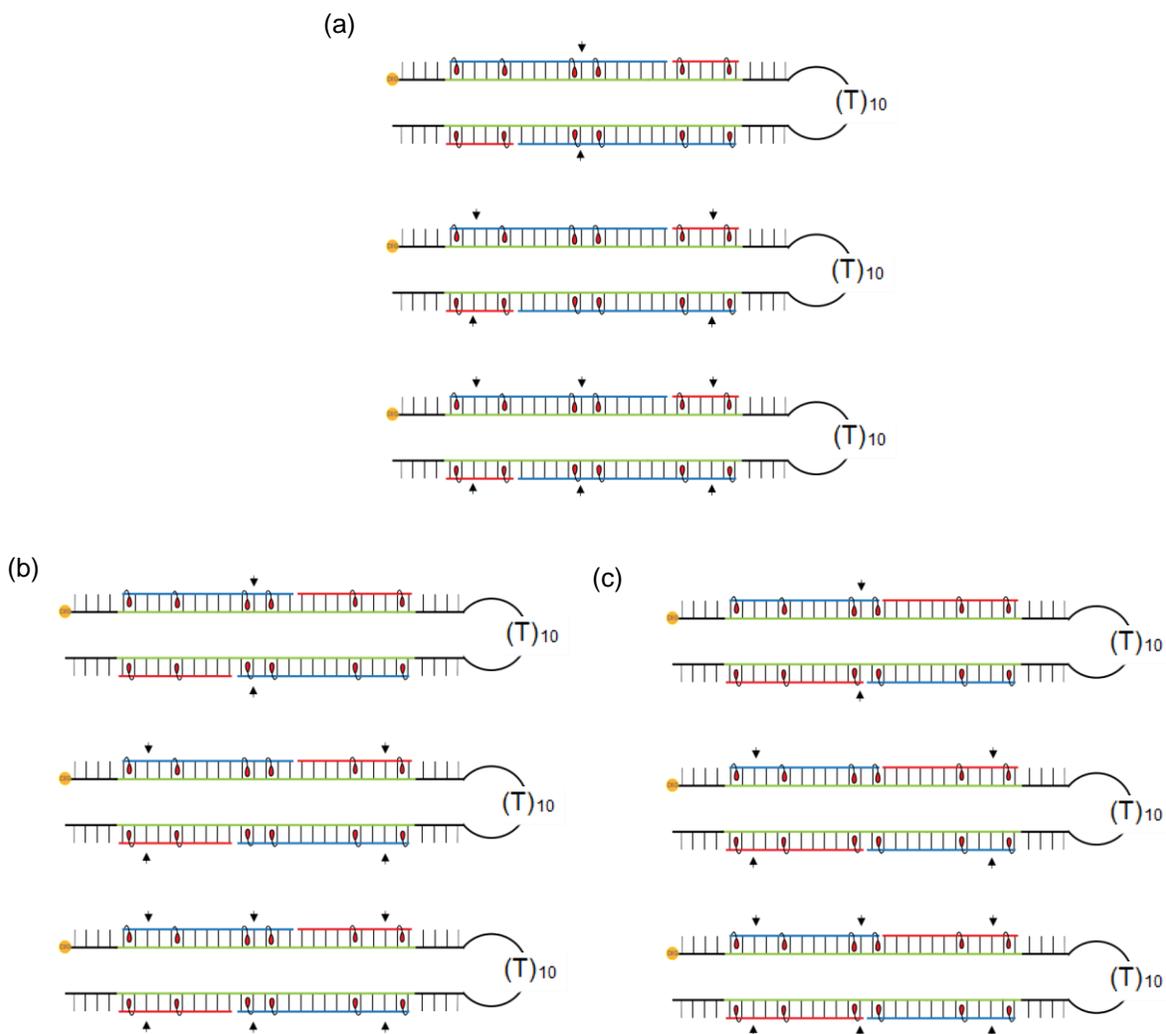
Figure 4.40. Illustration of the mismatched recognition complexes that would ensue upon recognition of **MM1**-**MM3** by (a) **NIP1**, (b) **NIP3**, and (c) **NIP4**; arrows indicate position of mismatched base-pairs. The arc connecting the two strand of hairpin is $(T)_{10}$ linker. For sequences of **MM1**-**MM3**, see Table 4.4.

Table 4.6. Sequences of probes used in FISH experiments, T_ms of probe duplexes and duplexes with DNA targets, and thermal advantages (*TA*s) of probes at high salt conditions.[a]

ONs	Sequence	T_m (°C)						
		5'-main probe vs 3'-aux probe	5'-main probe vs 3'-main probe	5'-aux probe vs 3'-main probe	5'-main probe vs DNA4	5'-aux probe vs DNA4	3'-aux probe vs DNA3	3'-main probe Vs DNA3
23/24	5'-Cy3-TG**U**GT**U**ATA**U**GC**U**GTTCT**C**AGCC**C**T-3'	37.0	36.0	nt	75.0	58.0	43.0	>76.0
25/26	3'-ACA**C**AA**U**ATA**C**GA**C**AAGAG**U**CGGG**A**-Cy3-5'							

[a] The unmodified **DNA3**:**DNA4** duplex (T_m = 82.0 °C) is the model dsDNA duplex, where **DNA3** = 5'- ACTG TG TGT TAT A TGC TGT TC TCA GCC CT ACTG and **DNA4** = 5'-CAGT AG GGC TGA GA ACA GCA T ATA ACA CA CAGT. Thermal denaturation curves were recorded as described in Table 4.1 with the exception that in a high salt phosphate buffer was used ([Na$^+$] = 710 mM, [Cl$^-$] = 710 mM, pH 7.0 (NaH$_2$PO$_4$/Na$_2$HPO$_4$), [EDTA] = 0.2 mM), with each ON present at 0.5 μM concentration. nt = no sigmoidal transition observed, indicating that if a duplex is formed it is highly distorted.
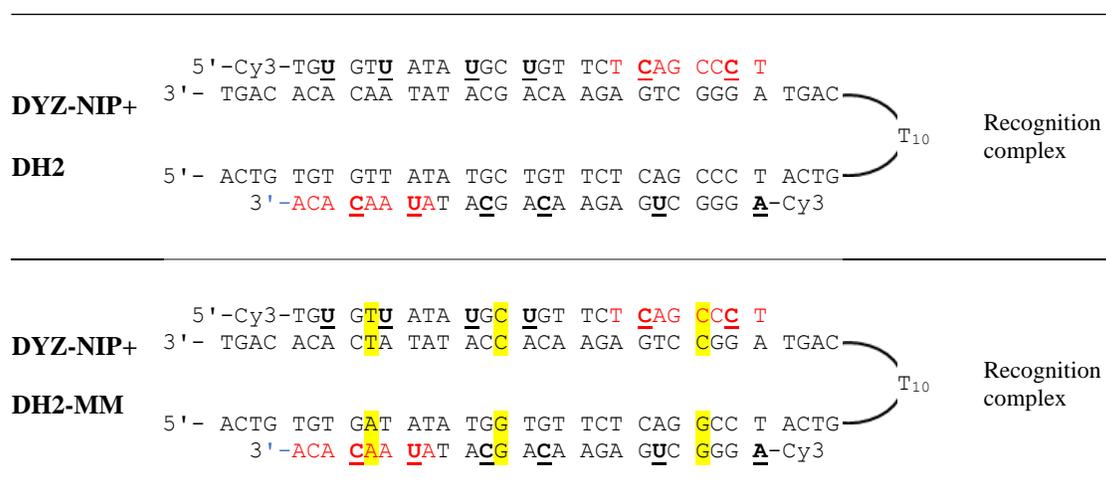
Figure 4.41. Recognition complex and mismatched recognition complex that would ensue upon recognition of complementary target **DH2,** and non-complementary target **DH2-MM**, with yellow highlighted indicating the position of mismatched base-pairs (relative to chromosomal region targeting probe **DYZ-NIP**). T$_{10}$ loop of hairpin is represented by extended arcs.
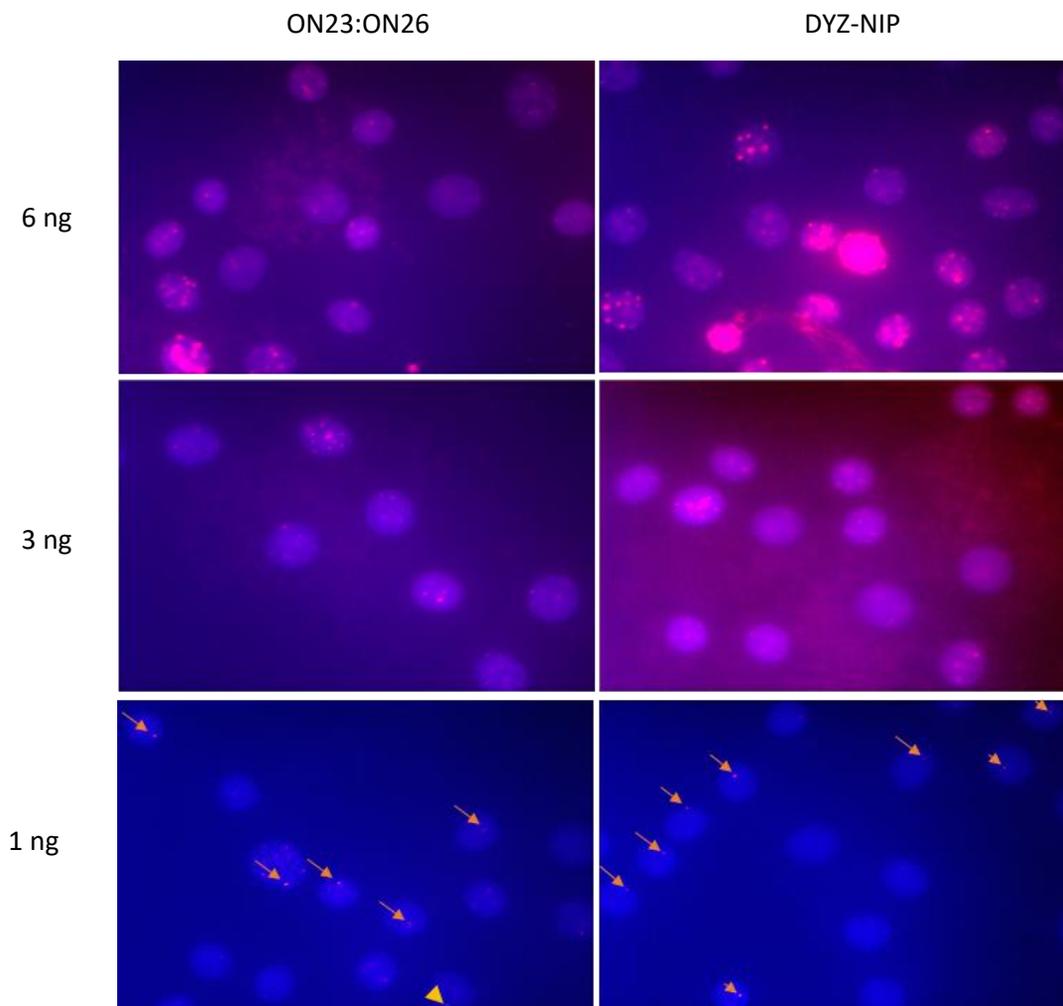
ON23:ON26 DYZ-NIP

Figure 4.42. Images from nd-FISH experiments in which different concentrations of toehold Invader probe **ON23:ON26** or nicked Invader probe **DYZ-NIP** (1-6 ng per 200 ul of PCR buffer) were incubated with fixed isolated nuclei from male bovine kidney cells for 3 h at 37.5 °C in a Tris buffer (20 mM Tris-Cl, 100 mM KCl, pH 8.0) and counterstained with DAPI. Experimental are as specified in Fig. 4.7. High background is observed when using **DYZ-NIP** and, **ON23:ON26** at 3 or 6 ng (per 200 ul of PCR buffer), whereas 1 ng per 200 ul of PCR buffer yielded centrally localized Cy3 signals. One sentence comment what was best for toehold probe.
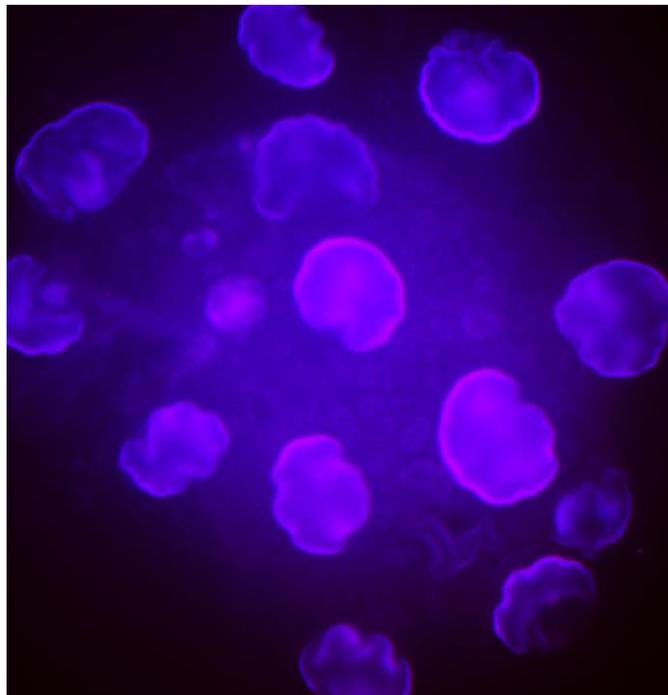
Figure 4.43. Images from nd-FISH experiments in which nicked Invader probe **DYZ-NIP was** incubated with isolated fixed nuclei from a female bovine endothelial cell line (15 ng per 200 µl of PCR buffer, 3 h, 37.5 °C). Experimental conditions and image analysis was carried out as described in Fig. 4.7.

4.7 REFERENCES AND NOTES.

1 Adhikari, S. P., Emehiser, R. G., Karmakar, S. & Hrdlicka, P. J. Recognition of mixed-sequence DNA targets using spermine-modified Invader probes. *Org. Biomol. Chem.* **17**, 8795–8799 (2019).

2 Duca, M., Vekhoff, P., Oussedik, K., Halby, L. & Arimondo, P. B. The triple helix: 50 years later, the outcome. *Nucleic Acids Res.* **36**, 5123–5138 (2008).

3 Dervan, P. B. & Edelson, B. S. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr. Opin. Struct. Biol.* **13**, 284–299 (2003).

4 Ricciardi, A. S., Quijano, E., Putman, R., Saltzman, W. M. & Glazer, P. M. Peptide Nucleic Acids as a Tool for Site-Specific Gene Editing. *Mol. Basel Switz.* **23**, (2018).

5 Guenther, D. C. *et al.* Invader probes: harnessing the energy of intercalation to facilitate recognition of chromosomal DNA for diagnostic applications. *Chem. Sci.* **6**, 5006–5015 (2015).

6 Rapireddy, S., He, G., Roy, S., Armitage, B. A. & Ly, D. H. Strand Invasion of Mixed-Sequence B-DNA by Acridine-Linked, γ-Peptide Nucleic Acid (γ-PNA). *J. Am. Chem. Soc.* **129**, 15596–15600 (2007).

7 Hsu, P. D., Lander, E. S. & Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157**, 1262–1278 (2014).

8 Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **168**, 20–36 (2017).

9 Hrdlicka, P. J., Kumar, T. S. & Wengel, J. Targeting of mixed sequence double-stranded DNA using pyrene-functionalized 2′-amino-α-L-LNA. *Chem. Commun.* 4279–4281 (2005) doi:10.1039/B506986F.

10 Crothers, D. M. Calculation of binding isotherms for heterogeneous polymers. *Biopolymers* **6**, 575–584 (1968).

11 Tsai, C., Jain, S. C. & Sobell, H. M. Visualization of drug-nucleic acid interactions at atomic resolution: I. Structure of an ethidium/dinucleoside monophosphate crystalline complex, ethidium:5-Iodouridylyl (3′–5′) adenosine. *J. Mol. Biol.* **114**, 301–315 (1977).

12 Sau, S. P. *et al.* Identification and characterization of second-generation invader locked nucleic acids (LNAs) for mixed-sequence recognition of double-stranded DNA. *J. Org. Chem.* **78**, 9560–9570 (2013).

13 Karmakar, S., Madsen, A. S., Guenther, D. C., Gibbons, B. C. & Hrdlicka, P. J. Recognition of double-stranded DNA using energetically activated duplexes with interstrand zippers of 1-, 2- or 4-pyrenyl-functionalized O2′-alkylated RNA monomers. *Org. Biomol. Chem.* **12**, 7758–7773 (2014).

14 Hu, P. *et al.* Cooperative Toehold: A Mechanism To Activate DNA Strand Displacement and Construct Biosensors. *Anal. Chem.* **90**, 9751–9760 (2018).

15 Karmakar, S. *et al.* High-affinity DNA targeting using readily accessible mimics of N2'-functionalized 2'-amino-α-L-LNA. *J. Org. Chem.* **76**, 7119–7131 (2011).

16 Karmakar, S., Guenther, D. C. & Hrdlicka*, P. J. Recognition of mixed-sequence DNA duplexes: Design guidelines for Invaders based on 2′-*O*-(pyren-1-yl)methyl-RNA monomers. *J. Org. Chem.* **78**, 12040–12048 (2013).

17 Unpublished data. PhD dissertation work, S.P. Adhikari -Optimizing Invader probe architect for sequence-unrestricted recognition of dsDNA targets.

18 Demidov, V. V. & Frank-Kamenetskii, M. D. Two sides of the coin: affinity and specificity of nucleic acid interactions. *Trends Biochem. Sci.* **29**, 62–71 (2004).

19 Demidov, V. V. *et al.* Kinetics and mechanism of the DNA double helix invasion by pseudocomplementary peptide nucleic acids. *Proc. Natl. Acad. Sci.* **99**, 5953–5958 (2002).

20 Chen, Y.-J., Groves, B., Muscat, R. A. & Seelig, G. DNA nanotechnology from the test tube to the cell. *Nat. Nanotechnol.* **10**, 748–760 (2015).

21 Perret, J., Shia, Y.-C., Fries, R., Vassart, G. & Georges, M. A polymorphic satellite sequence maps to the pericentric region of the bovine Y chromosome. *Genomics* **6**, 482–490 (1990).

22 Unpublished data, C. Shepard, Thesis - Invader-mediated targeting of chromosomal DNA.

23 Dioubankova, N. N. *et al.* Pyrenemethyl ara-Uridine-2′-carbamate: A Strong Interstrand Excimer in the Major Groove of a DNA Duplex. *ChemBioChem* **4**, 841–847 (2003).

24 Brown, A. M. A step-by-step guide to non-linear regression analysis of experimental data using a Microsoft Excel spreadsheet. *Comput. Methods Programs Biomed.* **65**, 191–200 (2001).

CHAPTER 5: LNA-modified toehold probes for efficient mixed-sequence recognition of dsDNA-targets at physiological conditions.

Shiva P. Adhikari,[a] Raymond Emehiser,[a] Michaela Brown, [a] and, Patrick J. Hrdlicka.[a]

[a] Department of Chemistry, University of Idaho, Moscow, ID-83844, USA

Abstract

In a recent study, toehold invader probes were shown to result in more efficient recognition of complementary double-stranded DNA target regions than blunt-ended Invader probes, thus highlighting the importance of toeholds for dsDNA recognition. In the present study, we set out utilize the toehold strategy with commercially available nucleotide monomers. Thus, we set out to explore dsDNA recognition properties of LNA-modified toehold probes. We compared the recognition efficiency of LNA-modified toehold probes relative to Zorro-LNA. Lastly, we introduced the concept of tiled LNA, which are two single-stranded non-overlapping LNA strands that target adjacent sites of complementary strands in dsDNA targets. Toehold-modified LNA probes and tiled LNA probes display efficient recognition of dsDNA targets at physiological conditions.

5.1. Introduction

Oligonucleotide (ON)-based probes capable of sequence-specific targeting of double-stranded DNA (dsDNA) at physiological conditions offer a wide range of diagnostic and therapeutic applications. Established DNA-targeting strategies such as minor-groove binding polyamides only recognize short double-stranded (ds) DNA regions (<8 bp),[1] because their shape complementarity is lost with longer probes, whereas triplex-forming oligonucleotides (TFOs) and peptide nucleic acid (PNAs) require long polypurine stretch.[2,3] Further, triplex-forming TFOs/PNAs require low pH for efficient binding to dsDNA targets.[2,3] Later, dsDNA-invading probes were developed (eg. chiral γPNA,[4] pseudocomplementary PNA (pcPNA),[5,] etc.) for dsDNA-targeting at physiological conditions. However, chiral γPNA are challenging to synthesize, while pcPNA only display satisfactory recognition of AT-rich sequences. Further studies to overcome these limitations are undergoing. Recently, Invader probes were introduced by the Hrdlicka laboratory, which efficiently recognize dsDNA at physiological conditions.[6,7]

These Invader probes are short DNA duplexes containing one or more +1 interstrand zipper arrangements of intercalator-functionalized nucleotides like 2′-O-(pyren-1-yl)methyl-RNA.[6,7] Later, overhangs or toeholds were introduced at the 5'-ends of Invader probes allowing for recognition of longer target regions.[8] Toehold Invader probes display enhanced recognition of dsDNA-targets vis-à-vis conventional blunt-ended Invader probes, and they were shown to recognize a chromosomal DNA target region, which was refractory to detection by conventional Invader probes. From this study, we realized that toehold facilitates dsDNA-recognition and thus we set out to explore the properties of LNA-modified toehold probes utilizing commercially available LNA-monomers, for recognition of dsDNA-targets. Here we compare the recognition efficiency of LNA-modified toehold probes vis-à-vis a previously introduced DNA-targeting approach, i.e., Zorro-LNA.[9] Zorro-LNAs are LNA-based constructs consisting of two highly LNA-modified strands that form a short non-targeting overlapping , thus yielding two long single-stranded overhangs.[10] We also introduce the concept of tiled LNA probes for dsDNA-recognition. The study presented here is very preliminary and further experiments and evaluation are needed for a full characterization of these LNA-based probes.

5.2. Results and discussion:

Design of LNA-modified toehold probes.

A series of LNA-modified ONs were previously prepared by a former Ph.D.-student in the Hrdlicka laboratory (Table 5.1 and Figure 5.1).[11] These ONs were purified using HPLC and characterized using MALDI-MS previously (See Table 1 footnote).[11] Access to these ONs allowed for assembly of different LNA-based probes: i) LNA-modified toehold probes, ii) Zorro-LNA, and iii) tiled LNA probes. LNA-toehold probes are DNA duplexes with single-stranded 5'-overhangs that are 50% LNA-modified. LNA-toehold probes **LNA3:LNA4**, **LNA5:LNA6**, and **LNA7:LNA8** – comprised of individual strands that are 16 nucleotides (nts) - have an overlapping double-stranded DNA region of 4, 6, and 8 base pairs (bps), and single-stranded overhangs that are 16, 12, and 10 nts, respectively (Figure 5.1). Additionally, corresponding LNA-modified toehold probes featuring LNA monomers in the overlap regions (**LNA3L:LNA4L** - **LNA7L:LNA8L**) were constructed to explore the effect of LNA-induced stabilization in the overlapping region. Zorro-LNA construct **LNA9:LNA10** consists of two LNA-modified strands that target adjacent regions on opposing strands of complementary

dsDNA. The two strands are connected via a 7-bp non-targeting duplex. The equivalent LNA tiling probe **LNA1**+**LNA2** targets the same region but differs from the Zorro LNA inasmuch the strands are not connected (Figure 5.1). Additionally, to compare different features in the designs e.g., the impact overall probe length, toehold length, overlap length, other LNA-based probes were prepared (Fig.5.1). All of the probes were designed to recognize complementary mixed-sequence regions embedded within a 32-mer model dsDNA target **DNA1**:**DNA2**. The thermal denaturation properties of the LNA-modified probes have been previously reported (see Table 1).[11] The LNA-modified strands form very duplexes with DNA harboring complementary regions (T_m ~ 70.5 °C - 83.0 °C). The stability of the overlaps of the LNA-modified toehold probes ranges from moderate to high (T_m ~ 39.5 °C - 72.0 °C).[11]
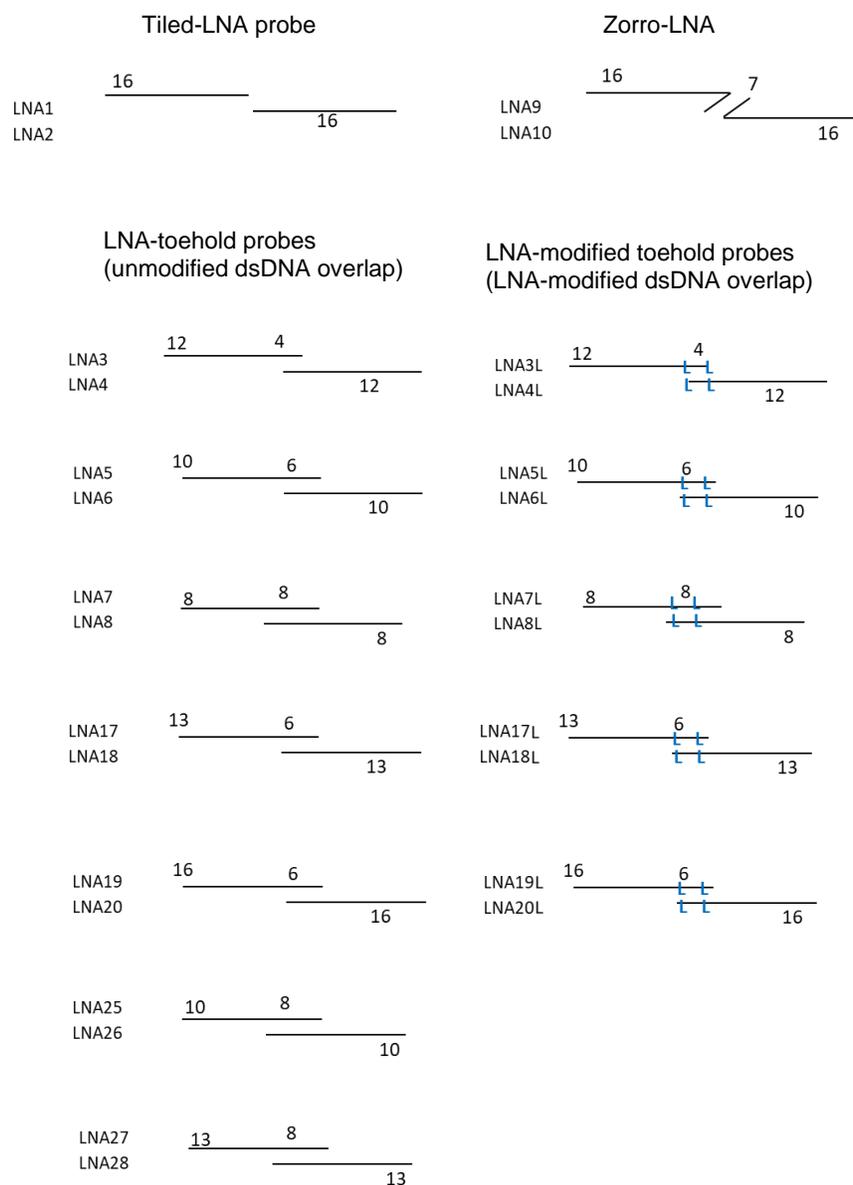
Figure 5.1. Illustration of different LNA-based probe designs. See Table 5.1 for sequences.

Characterization of dsDNA-recognition by LNA-modified probes. The dsDNA-targeting properties of the LNA-modified probes were characterized using an electrophoretic mobility shift assay in which a doubly 3'-digoxigenin (DIG)-labeled version of **DNA1**:**DNA2** was used as a model dsDNA target (Fig. 5.2). Recognition of **DNA1**:**DNA2** by the LNA-modified probes results in the formation of two duplexes between individual probe strands and the corresponding 33-mer DNA strands harboring the complementary region, i.e., two probe-target duplexes, (PTDs) which have greater mobility than **DNA1**:**DNA2** (Fig. 5.2).
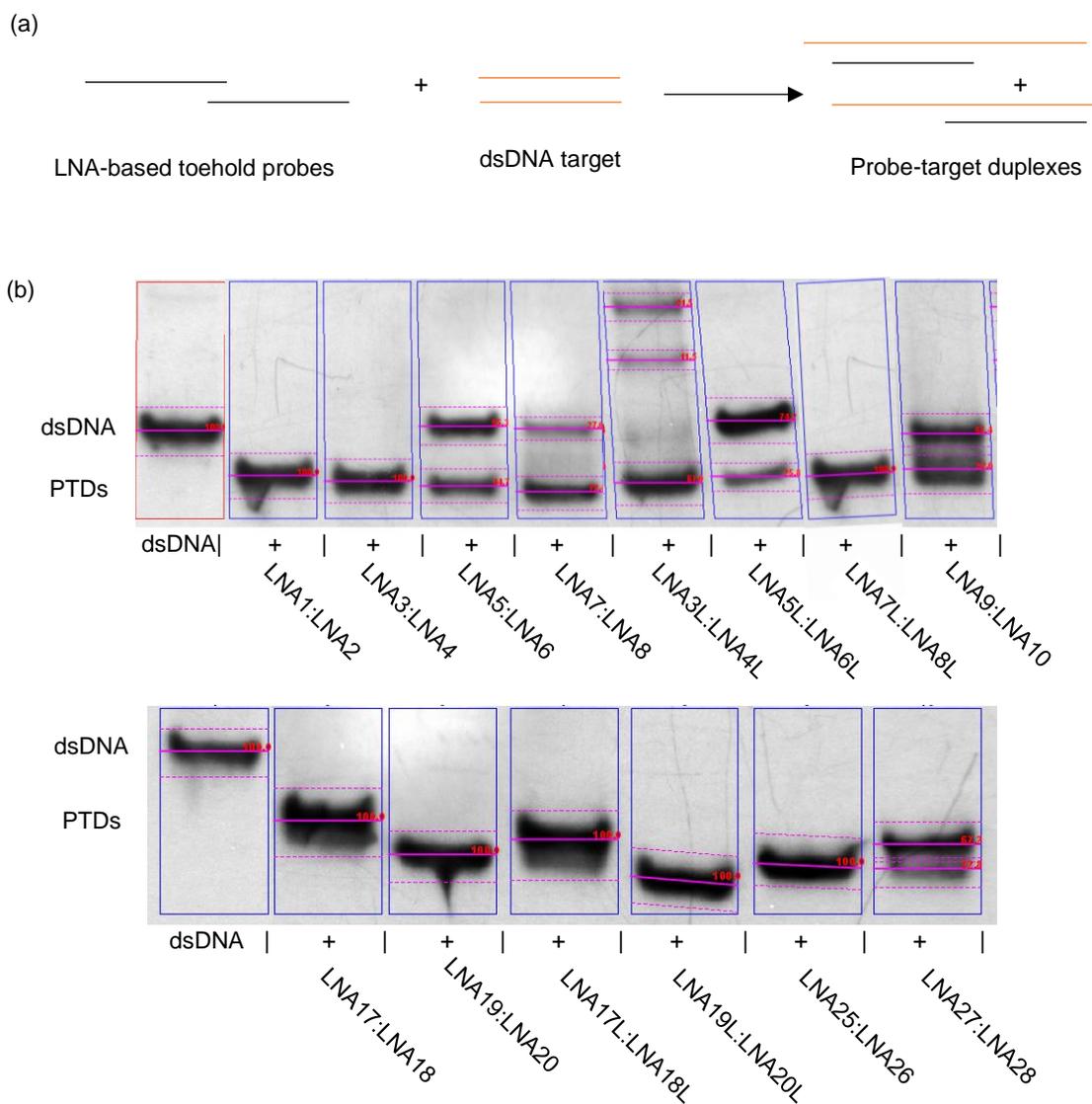
Figure 5.2. (a) Illustration of assay used to evaluate toehold-LNA-mediated recognition of model dsDNA targets. (b) Representative gel electrophoretograms from recognition experiments in which **DNA1**:**DNA2** (5'-GAA GAT CAG TTG GGT GCA CGA GTG GGT TAC AT-DIG:3'-DIG-CTT CTA GTC AAC CCA CGT GCT CAC CCA ATG TA) was incubated with a 100-fold molar excess of different probes. PTDs = probe-target duplexes. Only one replicate was performed given the preliminary nature of this study. Conditions: pre-annealed doubly 3'-DIG-labeled **DNA1**:**DNA2** (50 nM) was incubated with a 100-fold molar excess of the specified pre-annealed probe 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 17 h. Mixtures were resolved on 16% nd-PAGE gels.

The tiled LNA probe combination **LNA1+LNA2** and Zorro-LNA **LNA9:LNA10**, have identical total probe length (32 nt). The tiled LNA probe combination **LNA1+LNA2** results in complete recognition of **DNA1**:**DNA2**, whereas only ~50% recognition is observed with Zorro LNA **LNA9:LNA10**. This suggests that linking the two LNA strands of Zorro LNAs actually might be detrimental. Similarly, toehold probe **LNA17**:**LNA18**, which also targets a 32 bp region, also results in complete recognition of **DNA1**:**DNA2**. Thus Zorro-LNA construct appears to be unfavorable relative to tiledLNA probes or LNA-modified toehold constructs..

All the strands entailed in making LNA-modified toehold probes **LNA3**:**LNA4-LNA7**:**LNA8** are 16 nt long. The toehold probes differ in their architecture (i.e., toehold and overlap length) and – accordingly the length of the region that they are targeting. Thus, **LNA3**:**LNA4** (12-4-12) which has toeholds that are 12 nt long, an overlap that is 4 bps long, and targets a 28 bp region, results in complete recognition of **DNA1**:**DNA2**. When the overlapping region is extended – and total number of target base-pairs is reduced – the recognition efficiency decreases (compare recognition efficiency of **LNA3**:**LNA4**, **LNA5**:**LNA6** ~ 55% and **LNA7**:**LNA8** ~ 70%). This trend might be due to the higher stability of the overlapping region or a less favorable driving force since fewer base-pairs are targeted. All of the LNA-modified toehold probes with unmodified overlapping regions display more efficient recognition of **DNA1**:**DNA2** than Zorro LNA **LNA9**:**LNA10** despite targeting shorter regions. It is likely that toehold designs are more favorable than zorro-LNA design.

When LNA monomers are introduced in the overlapping dsDNA regions of LNA-toehold probes recognition is less efficient with **LNA3L**:**LNA4L**, and **LNA5L**:**LNA6L** but not with **LNA7L**:**LNA8L** for which complete recognition is observed. The loss of recognition for **LNA3L**:**LNA4L** and **LNA5L**:**LNA6L** might be due to higher stability of the overlapping regions arising from the LNA modifications, leading to hampered dissociation. However, further study is needed to fully understand the properties of **LNA7L**:**LNA8L** which completely recognizes the target despite having an 8-mer LNA-modified overlapping region. When LNA-modifications were introduced in overlaps of **LNA17**:**LNA18** and **LNA19**:**LNA20** to form

LNA17L:**LNA18L** and **LNA19L**:**LNA20L**, the effects of LNA-introduction in the overlap are inconclusive, as all four probes results in complete recognition of **DNA1:DNA2**.

LNA1+LNA2 and **LNA17**:**LNA18** both target 32 bp regions and the lengths of the single-stranded regions are identical (16 nt per strand), but differ in the presence/absence of an overlap (0 and 6 bp, respectively). Since, both probes result in complete recognition of **DNA1**:**DNA2**, the impact of overlap is unclear.

LNA5:**LNA6** and **LNA25**:**LNA26** have toeholds of identical length (10 nt), but different overlap lengths (6 and 8 bps, respectively). **LNA25**:**LNA26** (10-8-10) appears to result in favorable recognition (~100%) despite having a more extensive overlap- which renders the result counterintuitive. The more efficient dsDNA-recognition displayed by **LNA25**:**LNA26** might instead be the consequence of targeting a longer region relative to **LNA5**:**LNA6** (28 vs 26) relative to probes.

LNA-toehold probes **LNA3**:**LNA4** and **LNA25**:**LNA26**, target regions of identical length (28 bps) but have different toehold designs (12-4-12 and 10-8-10 respectively). Since both designs result in complete recognition of target, the effect of the different toehold designs is not clear from this experiment.

LNA-toehold probes **LNA5**:**LNA6** has same overlap as **LNA17**:**LNA18** or **LNA19**:**LNA20** (6 bps) but different overhang lengths (10, 13, and 16 nts respectively) and, accordingly, target regions of different lengths (26, 32, and 38 bps long). **LNA5**:**LN6** (featuring a 10-6-10 design) is less effective (~55 %) relative to **LNA17**:**LNA18** (13-6-13) or **LNA19**:**LNA20** (16-6-16) both of which result in complete recognition, which could be due to their longer toeholds or since longer regions are targeted. Comparing three probes, longer toehold length – or targeting longer regions – is beneficial. Similarly, LNA-modified toehold probes **LNA7**:**LNA8**, **LNA25**:**LNA26**, **LNA27**:**LNA28** have overlaps of equaly length (8 bps), but different toehold lengths and accordingly target regions of different lengths (24, 28 and 34 bps, respectively) . **LNA7:LNA8** (8-8-8) is less effective (~70 % recognition) than **LNA25:LNA26** (10-8-10) or **LNA27:LNA28** (13-8-13) (later two ~ complete recognition), which again might be due to the longer toeholds or because longer regions are targeted.

LNA1:**LNA2** (16-0-16) and **LNA19**:**LNA20** (16-6-16) have identical toehold lengths (16 nt), but differ in the length of the overlap (0 and 6 bps, respectively); accordingly, the length of the targeted region differs (32 and 38 bps, respectively). The effect of overlap length is unclear since both probes display complete recognition.

Similarly, **LNA5**:**LNA6** and **LNA25**:**LNA26** (10-8-10 and 10-6-10) also have identical toehold lengths, but the length of the overlap – and, thus, length of the targeted region – differs. **LNA25**:**LNA26** results in more efficient recognition than **LNA5**:**LNA6** which is somewhat counterintuitive given the more extensive overlap which should render it more difficult for probes to dissociate, Instead, it seems likely that the better dsDNA-recognition of **LNA25**:**LNA26** is due to the fact that a longer region is targeted.

LNA17:**LNA18** and **LNA27**:**LNA28** have toeholds of identical length (13 nt), but different overlaps (6 and 8 bps, respectively) and accordingly target regions of different lengths (25 and 31 bps, respectively). The effect of overlap length (and total probe length) is uncertain as both probes completely recognize the target (faster moving band observed for both cases), although the results are not completely clear for **LNA 27**:**LNA28** as two bands formed.

Collectively, these observations indicate that LNA-modified toehold probes and tiled-LNA probes offer advantages for recognition of mixed-sequence dsDNA target regions vis-à-vis Zorro LNA. Moreover, some of the results suggest that the overlapping regions must be kept short and labile for dsDNA-recognition to be. Lowering the probe concentration might help us see the effect, thus further exploration is required to evaluate the properties of these probes.

Discussion on T_m values of LNA3:LNA4 and LNA5:LNA6

We noticed that LNA-modified toehold probe **LNA3:LNA4** has an unreasonably high T_m (~ 59.5 °C) given the 4 bp double-stranded overlap. Similarly, **LNA5**:**LNA6** has a 6-bp overlap dsDNA region and a T_m ~ 59.0 °C. First, we evaluated T_ms of single strands (**LNA3/LNA4/LNA5/LNA6**) to investigate if there is any secondary structure formation or self-hybridization. No sigmoidal transitions were observed suggesting that secondary structures or self-hybridization is not happening. (Fig. 5.3).
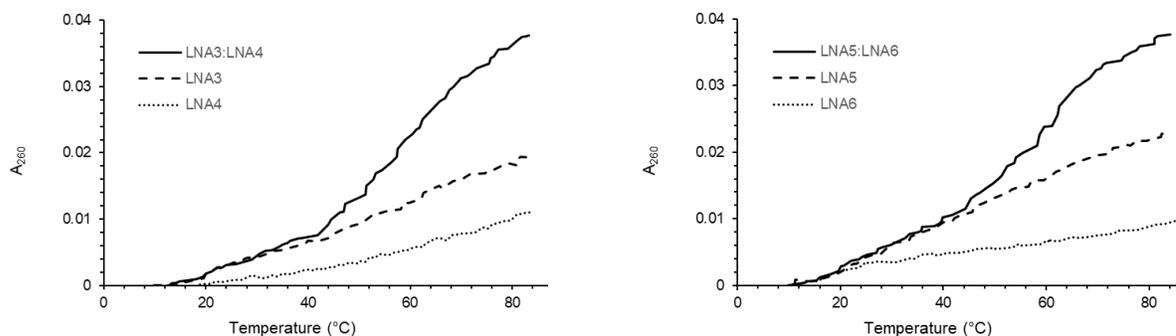
Figure 5.3. Thermal denaturation curves of LNA-toehold probes **LNA3:LNA4, LNA5:LNA6** and the individual strands thereof.

Next, we performed analysis of the probe design by using oligoanalyzer tool (IDT),[12] and we found an alternative **LNA3**:**LNA4** heterodimer structure (Fig. 5.4) more stable ($\Delta G \sim$ -9.43 kcal/mol) than the design we anticipated ($\Delta G \sim$ -7.05 kcal/mol). Additionally, the **LNA3**:**LNA4** heterodimer structure is further stabilized by the presence of four LNA-monomer in the dsDNA region of this structure. Similarly, when analyzing **LNA5**:**LNA6**, we found an alternative structure that is as stable as the structure we anticipated. Further, this alternative **LNA5**:**LNA6** heterodimer structure consists of four LNA-monomers in dsDNA region, which further stabilizes the duplex (Fig. 5.4). Presumably, the high T_m is the result of the alternative structures stabilized by LNA-monomers.

(a) LNA3:LNA4

Delta G: -9.43 kcal/mole **Base Pairs:** 5

```
5'  AGATCAGTTGGGTGCA
      :   :   |||||
3'  ACGTGCTCACCCAATG
```

Alternative structure predicted by oligoanalyzer tool

Delta G: -7.05 kcal/mole **Base Pairs:** 4

```
5'  AGATCAGTTGGGTGCA
                ||||
3'              ACGTGCTCACCCAATG
```
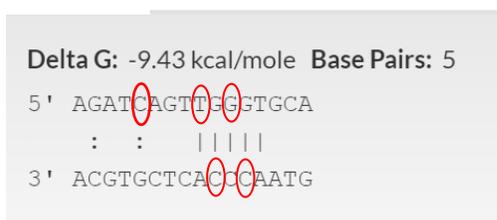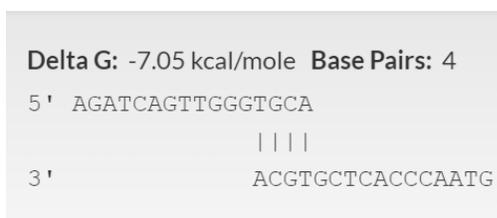
Anticipated structure

(b) LNA5:LNA6

Delta G: -9.73 kcal/mole **Base Pairs:** 6

```
5'  GATCAGTTGGGTGCAC
                ||||||
3'              CACGTGCTCACCCAAT
```

Anticipated structure

Delta G: -9.43 kcal/mole **Base Pairs:** 5

```
5'    GATCAGTTGGGTGCAC
        :   :   |||||
3'  CACGTGCTCACCCAAT
```
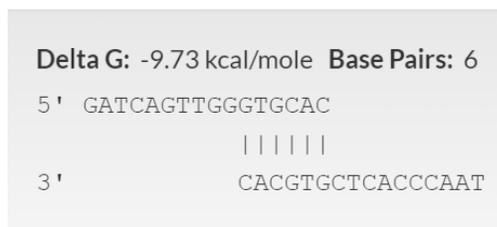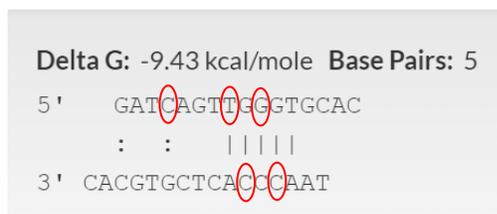
Alternative structure predicted by oligoanalyzer tool

Figure 5.4. Secondary structure prediction using oligoanalyzer tool (IDT).[12] The red circle represents LNA-monomers (Only the LNA-monomers that form base-pairs with corresponding ON are shown, for detail sequence see Table 5.1). The Delta G values are calculated by considering unmodified monomers using oligoanlayzer tool.

5.3. Conclusion and future directions

LNA-based dsDNA targeting probes (tiled LNA probes and some LNA-modified toehold probes) display enhanced dsDNA recognition relative to equivalent Zorro LNA. The properties and length optimization of the dsDNA overlap region in LNA-toehold probes needs to be carried out more rigorously to fully understand the trends. Cy3-labeled LNA-based probes can

be designed to envision recognition of target regions within chromosomal DNA. The successful recognition would provide a potential application in molecular biology and diagnostics.

5.4. Supplementary data

Table 5.1. Sequences of LNA probes.[a,b]

ONs	Sequences
LNA1	5'-gAa GaT cAg TtG gGt G
LNA2	5'-aT gTa AcC cAc TcG tG
LNA3	5'-a GaT cAg TtG gGT GCA
LNA4	5'-gTa AcC cAc TcG TGC A
LNA5	5'-GaT cAg TtG gGT GCA C
LNA6	5'-Ta AcC cAc TcG TGC AC
LNA7	5'-aT cAg TtG GGT GCA CG
LNA8	5'-a AcC cAc TCG TGC ACC
LNA9	5'-gAa GaT aAg TtG gGt GTC tAa Ct
LNA10	5'-aT gTa AcC cAc TcG tG AG tTA GA
LNA3L	5'-a GaT cAg TtG gGa GCt
LNA4L	5'-gTa AcC cAc TcG tGC a
LNA5L	5'-GaT cAg TtG gGt GCa C
LNA6L	5'-Ta AcC cAc TcG tGC aC
LNA7L	5'-aT cAg TtG GGt GCa CG
LNA8L	5'-a AcC cAc TCG tGC aCC
LNA17	5'-gAa GaT cAg TtG gGT GCA C
LNA18	5'-aT gTa AcC cAc TcG TGC AC
LNA19	5'-GcT gAa GaT cAg TtG gGT GCA C
LNA20	5'-Tc GaT gTa AcC cAc TcG TGC AC
LNA17L	5'-gAa GaT cAg TtG gGt GCa C
LNA18L	5'-aT gTa AcC cAc TcG tGC aC
LNA19L	5'-GcT gAa GaT cAg TtG gGt GCa C
LNA20L	5'-Tc GaT gTa AcC cAc TcG tGC aC
LNA25	5'-aGa TcA gTt GGG TGC ACG
LNA26	5'-gTa AcC cAc TCG TGC ACC
LNA27	5'-TgA aGa TcA gTt GGG TGC ACG
LNA28	5'-GaT gTa AcC cAc TCG TGC ACC

[a] LNA monomers shown in lower case letters.

[b] See reference 11 for MALDI-MS, HPLC-chromatogram, and T_m data.

Protocol - Thermal denaturation experiments. The concentrations of ONs were estimated using the following extinction coefficients (OD$_{260}$/µmol): G (12.01), A (15.20), T/U (8.40), C (7.05).[13] T_ms of duplexes (1 µM final concentration of each strand) were determined using a Cary 100 UV/Vis spectrophotometer equipped with a 12-cell Peltier temperature controller and determined 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$_2$HPO$_4$ and 5 mM Na$_2$HPO$_4$). Strands were mixed in quartz optical cells having 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. A temperature range from 10 °C to at least 15 °C above the duplex T_m was used, with T_ms determined as a single experiment. A temperature ramp of 1 °C/min was used in all experiments.

Protocol - electrophoretic mobility shift assay. Unmodified DNA strands were obtained from commercial sources and used without further purification. Target strands were DIG-labelled using the 2nd generation DIG Gel Shift Kit (Roche Applied Bioscience). Briefly, 11-digoxigenin-ddUTP was incorporated at the 3′-end of the strand (100 pmol) using a recombinant DNA terminal transferase. The reaction mixture was quenched through the addition of EDTA (50 mM), and then diluted to 100 nM in 2X HEPES buffer and used without further processing. The recognition experiments were conducted essentially as previously reported.[6] Thus, LNA-probes (concentration as specified) were annealed (90 °C for 2-3 min, followed by cooling to room temperature) and subsequently incubated with separately pre-annealed DIG-labelled DNA (50 nM final concentration in 1X 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 ± 2 °C for 17 h.

Following incubation, loading dye (6X) was added and the mixtures were then loaded onto 16 % non-denaturing TBE-PAGE gels (45 mM tris-borate, 1 mM EDTA; acrylamide:bisacrylamide (19:1)). Mixtures were resolved via electrophoresis, which was performed using constant voltage (70 V) at ~4 °C. Bands were blotted onto positively charged nylon membranes (100 V, 30 min, ~4 °C) and cross-linked through exposure to UV light (254 nm, 5 × 15 W bulbs, 3 min). The membranes were incubated with anti-digoxigenin alkaline phosphatase F$_{ab}$ fragments as recommended by the manufacturer and transferred to a

hybridization jacket. Membranes were incubated with the chemiluminescence substrate (CSPD) for 10 min at 37 °C, and chemiluminescence was captured on X-ray films. Digital images of developed X-ray films were obtained using a BioRad ChemiDoc™ MP Imaging system and used for densitometric quantification of the bands. The percentage of dsDNA recognition was calculated as the intensity ratio between the recognition band relative to the total lane. An average of at least three independent experiments is reported along with standard deviations (±). Electrophoretograms may be composite images from different runs.

5.5. References and Notes.

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

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

3 K. Kaihatsu, B. A. Janowski and D. R. Corey, *Chem. Biol.*, 2004, **11**, 749–758.

4 B. Raman, Q. Elias, A. M. Nicole, L. Yanfeng, C. B. Dinesh, L.-G. Francesco, J. F. Rachel, M. S. William, H. L. Danith and M. G. Peter, *Curr. Gene Ther.*, 2014, **14**, 331–342.

5 N. Shigi, Y. Mizuno, H. Kunifuda, K. Matsumura and M. Komiyama, *Bull. Chem. Soc. Jpn.*, 2019, **92**, 330–335.

6 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.

7 P. J. Hrdlicka, T. S. Kumar and J. Wengel, *Chem. Commun.*, 2005, 4279–4281.

8 Unpublished results; S. P. Adhikari, P. Vukelich, D. C. Guenther, S. Karmakar and P. J. Hrdlicka (Recognition of double-stranded DNA using toehold Invader probes)

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

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

11 R. G. Emehiser (Ph.D. dissertation; Exploring Double-Stranded DNA-Targeting Strategies).

12 Oligo Analyzer, https://www.idtdna.com/calc/analyzer, (accessed 27 July 2021).

13 N. N. Dioubankova, A. D. Malakhov, D. A. Stetsenko, M. J. Gait, P. E. Volynsky, R. G. Efremov and V. A. Korshun, *Chembiochem Eur. J. Chem. Biol.*, 2003, **4**, 841–847.

Chapter 6: Conclusion and Future Perspective

DNA-targeting methods have the potential to treat diseases at their source because all diseases, in principle, are of genetic origin. By resolving the underlying restrictions that need to be improved in order to ensure improved recognition of dsDNA targets with excellent mismatch specificity, the platform for the advancement of DNA targeting tools will be ensured. In this regard, we optimized Invader probe architecture for DNA-targeting by using 2'-O-(pyren-1-yl)methyl RNA monomer and adding destabilizing or affinity-enhancing features to the probes. The dissertation work focuses on designing and developing alternative Invader probe architectures for enhanced sequence-unrestricted recognition of dsDNA targets with excellent mismatch specificity under physiological conditions. The intercalator arrangement in Invader probes destabilizes the duplex and energetically activates it, causing each strand to have a very high affinity for complementary DNA sequences, as recognition leads to very favorable stacking interactions between intercalators flanking base pairs. However, highly GC-rich and/or longer targets pose challenges since corresponding probes will be high-melting, thus interfering with the probe dissociation which is needed for this approach to be effective.

The spermine-bulged Invader project involves inserting non-nucleotidic spermine bulges into Invader probes, which results in more labile probe duplexes aided by the hotspots and bulges for further probe destabilization. At low incubation temperatures, Invader probes having one or two spermine bulges denature more quickly than conventional Invader probes, resulting in more efficient, yet highly specific recognition of mixed-sequence dsDNA targets. Thus, spermine-modified Invader probes are expected to be especially useful for recognizing high-melting DNA targets, such as extended and/or highly GC-rich regions, which are currently difficult to recognize because the corresponding conventional Invader probes are also very high-melting. Similarly, Invader probes modified with C_9-bulges are also useful for recognizing high-melting DNA targets, as they displayed greater destabilization than spermine-bulges. Because the addition of positively charged moieties like spermine is known to aid the delivery of oligonucleotides to the cell, these constructs could be further evaluated to target chromosomal DNA in the absence of transfection agents.[1] Further research into this area will reveal the potential for combining these two approaches in biological applications.

The toehold Invader probes were designed to recognize longer dsDNA targets by trimming the probes to shorten the double-stranded region of the probe to expose single-stranded overhangs. The single-stranded overhangs or toeholds can be additionally modified with affinity-enhancing modifications to further increase cDNA affinity, thereby increasing the overall thermodynamic driving force for dsDNA-recognition. The enhanced recognition efficiency of toehold Invaders against complementary targets, as well as chromosomal DNA, implies that toehold Invaders might be used in therapeutic and diagnostic applications on biological targets.

Nicked Invader probes in which nicks are introduced in longer probes to facilitate dissociation, recognizes target regions within chromosomal DNA and displayed improved target recognition over blunt-ended and toehold architecture probes and excellent mismatch discrimination against singly mismatched non-complementary targets. Nicked Invader probes, look to be a feasible approach for chromosomal DNA recognition, paving the way for the development of novel tools for molecular biology, genomic engineering, and nanotechnology applications. In a head-to-head comparison of toehold Invaders vs. Nicked Invaders, both were effective for recognition of *DYZ-1* target in Y-chromosome (nd-FISH assay). An extensive detailed study involving different target regions in chromosomal DNA and different lengths of probe architect might provide a more complete data set for evaluation of nicked probes vs toehold probes; however, it is clear from the hairpin assay that nicked Invaders display enhanced recognition compared to corresponding toehold probes and are also more efficient than blunt-ended Invaders.

To summarize, nicked Invader probes are the most promising for future applications due to their potential for improved DNA recognition and excellent mismatch specificity. Moreover, Nicked Invader probes provide a basis for subsequent probe design to construct concatemers, allowing for recognition of even longer target regions. Concatemers[2] are supramolecular polymer structures formed by the self-assembly of short DNA segments through wc interactions, and a very lengthy probe design targeting a larger area in the plasmid/chromosome may be imagined. However, when utilizing such lengthy probe designs, target specificity should be tackled in a biological context, and therefore the study concerning specificity should be addressed. Concatemers have also been found to help physiologically active oligonucleotide derivatives penetrate eukaryotic cells.[2] It will be feasible to employ nicked Invader probes for cellular

applications by combining improved probe designs with recent improvements in cellular delivery. To identify the entire extent of improved nicked Invader probes for biological applications, further research is required. Still promising, but potentially more challenging to use are Invader-modified toehold Invader probes, which display enhanced target recognition potential but suffer from poor specificity. Gratifyingly, LNA-modified toehold Invader probes are very promising due to the large improvement seen in DNA recognition over conventional Invader probes and improved mismatched specificity relative to Invader-modified toehold Invader probes. Also, LNA monomers are now commercially available thus, LNA probes are easy to obtain.

Overall, this dissertation work advances DNA targeting knowledge and will aid in the selection of the best probe architecture to improve DNA targeting for future applications in molecular biology, biotechnology, and biomedicine.

References:

1 K. T. Gagnon, J. K. Watts, H. M. Pendergraff, C. Montaillier, D. Thai, P. Potier and D. R. Corey, *J. Am. Chem. Soc.*, 2011, **133**, 8404–8407.

2 O. N. Simonova, A. V. Vladimirova, M. A. Zenkova and V. V. Vlassov, *Biochim. Biophys. Acta BBA - Biomembr.*, 2006, **1758**, 413–418.