Targeting DNA Secondary Structures Using Chemically Modified Oligonucleotides

A thesis presented in partial fulfilment
of the requirements for the degree of

Doctor of Philosophy
in Chemistry

Massey University,
Palmerston North, New Zealand

Yongdong Su

2021

For the memory of Len Blackwell

Abstract

Chemical modifications bring in additional features to oligonucleotides (ONs), including enhanced stability against nucleases, increased binding affinity towards DNA or RNA, improved cellular uptake, etc. This Thesis describes several strategies and chemical modifications used for targeting DNA duplexes and G-quadruplexes.

We introduced a pyrene analogue, (*R*)-1-*O*-[2-(1-pyrenylethynyl)phenylmethyl]-glycerol, called *ortho*-TINA (twisted intercalating nucleic acid) monomer into a native duplex DNA. The affinity of *ortho*-TINA modified strands was low to each other, whereas the affinity of *ortho*-TINA sequence towards complementary DNA was increased. This property of *ortho*-TINA duplex was applied for targeting native duplexes in a sequence-specific manner using a process called dual duplex invasion (DDI). The speed of DDI is increased with the increased number of *ortho*-TINA pairs present in the duplex, as well as with the rise of temperature from 4 to 37 °C. However, DDI against duplexes longer than the probe is compromised. To improve the kinetics of DDI, we designed and synthesised DNA probes with zwitterionic moieties, 4-(trimethylammonium)butylsulfonyl phosphoramidate groups (N+), in which the negatively charged phosphate is neutralised by the positively charged quaternary amine. We assume that several N+ moieties in the DNA probe should reduce the electrostatic repulsion between the probe and the target duplex, and in this way, enhance DDI. However, no improvement of kinetics was achieved using N+ modifications in the probe alone and in combination with *ortho*-TINA monomers. Application of ONs bearing N+ modifications was explored further in parallel DNA triplexes and G-quadruplex. The initial stage of assembly of N+TG$_4$T proceeded faster in the presence of Na^+ than K^+ ions, which contrasted the trend observed for unmodified sequences, and this process was independent of the ionic strength in solution.

We also evaluated several other phosphate modifications alongside for a comparison with our N+ modified DNA. Finally, several directions of future work are proposed based on the results obtained in the present Thesis.

Acknowledgements

I would like to thank my supervisor, Associate Professor Vyacheslav Filichev for providing me with the opportunity to do a PhD in his laboratory. Many thanks for your instruction, help and assistance throughout this journey. I would also like to thank my co-supervisors, Dr Richard Winkworth and Dr Tracy Hale for their support in plant genetics and cellular and molecular biology. Thanks also Associate Professor Len Blackwell for his ideas and instructions in the construction of lateral flow biosensors.

Special thanks to Dr Patrick J. B. Edwards and David Lun for helping me with so many times of NMR and mass-spectrometry measurements during the five years of my PhD. Many thanks to Dr Maitsetseg Bayarjargal for her help in the cell uptake assay.

I wish to thank my former colleague Dr Maksim V. Kvach for his valuable discussions and suggestions. Thanks also my group members, Harikrishnan Mohana Kurup and Bruce Chilton for their accompany. I am also grateful for the assistance of departmental staff over the years I have been at Massey University.

Finally, Massey University is gratefully acknowledged for providing us facilities and an environment for doing researches.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

A:	Adenine
ABE:	Adenine base editors
ACN:	Acetonitrile
AID:	Activation-induced cytidine deaminase
BCNs:	Branched, charge neutralising sleeves
bisLNA:	Bis-locked nucleic acid
BNA:	2′,4′-Bridged nucleic acid
bps:	Base pairs
C:	Cytosine
Cas:	CRISPR-associated
CD:	Circular dichroism
CPG:	Controlled pore glass
C-rich:	Cytosine rich
CRISPR:	Clustered regularly interspace short palindromic repeats
CRISPRi:	CRISPR interference
crRNA:	CRISPR RNA
CSDPR:	Circular strand displacement reaction
dCas9:	Dead Cas9
DCM:	Dichloromethane
DDI:	Dual duplex invasion
DMT:	4,4’-Dimethoxytrityl
DNA:	Deoxyribonucleic acid
DSB:	Double stranded break
dsDNA:	Double-stranded DNA
ESI-MS:	Electrospray ionisation mass spectrometry
ETT:	5-Ethylthio-1*H*-tetrazole
G:	Guanine
G4:	G-quadruplex
GNA:	Glycol nucleic acid
G-rich:	Guanine rich
gRNA:	Guide RNA
HDR:	Homology directed repair
HGPS:	Hutchinson-Gilford Progeria Syndrome

HMBC:	Heteronuclear multiple bond coherence
HMQC:	Heteronuclear multiple quantum coherence
HP:	2-Hydroxypyrrole
HP1α:	Heterochromatin Protein 1α
HPLC:	High-performance liquid chromatography
ICL:	Interstrand crosslinking
IE-HPLC:	Ion-exchange HPLC
Im:	Imidazole
INA:	Intercalating nucleic acid
indels:	Insertions or deletions
LHCP:	Left-handed circularly polarised light
LNA:	Locked nucleic acid
MALDI:	Matrix-assisted laser desorption ionisation
MGBs:	Minor groove binders
MQ water:	Milli-Q water
MW:	Molecular weight
NHEJ:	Non-homologous end joining
NMR:	Nuclear magnetic resonance spectroscopy
NOESY:	Nuclear Overhauser effect spectroscopy
ON:	Oligonucleotide
OTRs:	Oligonucleotide-templated reactions
PAGE:	Polyacrylamide gel electrophoresis
PAM:	Protospacer adjacent motif
PCR:	Polymerase chain reaction
PNA:	Peptide nucleic acid
Py:	Pyrrole
RHCP:	Right-handed circularly polarized light
RNA:	Ribonucleic acid
RNAi:	RNA interference
RNP:	Ribonucleoprotein
RP-HPLC:	Reverse-phase HPLC
RT:	Room temperature
RVDs:	Repeated variable di-residues
SaRNA:	Sulfonamide RNA

SE-HPLC:	Size-exclusion HPLC
SI:	Structure index
SNPs:	Single-nucleotide polymorphisms
SpyCas9:	*Streptococcus pyogenes* Cas9
ssDNA:	Single-stranded DNA
ST:	Salmon testes DNA
T:	Thymine
TALENs:	Transcription-activator like effector nucleases
TALEs:	Transcription-activator like effectors
TD:	Thermodynamic product
TERRA:	Telomeric repeat containing RNA
TFO:	Triplex forming oligonucleotide
THF:	Tetrahydrofuran
TINA:	Twisted intercalating nucleic acid
TOCSY:	Total correlation spectroscopy
TOF:	Time of flight
tracrRNA:	Trans-activating CRISPR RNA
U:	Uracil
ZFNs:	Zinc finger nucleases

Chapter 1. Introduction

1.1. Nucleic acids

Nucleic acids are biopolymers consisting of long chains of units called nucleotides, each containing a phosphate group and a nucleoside (Fig. 1.1). The nucleoside is composed of a nucleobase and a pentose sugar ring which is 2′-D-deoxyribofuranose for DNA and D-ribofuranose for RNA. The four bases found in DNA are bicyclic purines adenine (A) and guanine (G), and monocyclic pyrimidines cytosine (C) and thymine (T). A fifth base called uracil (U) was found in RNA as a replacement of thymine in DNA (Fig. 1.1A).[1] Nucleosides are joined by phosphate groups through phosphodiester bonds, forming one oligonucleotide strand.[2]

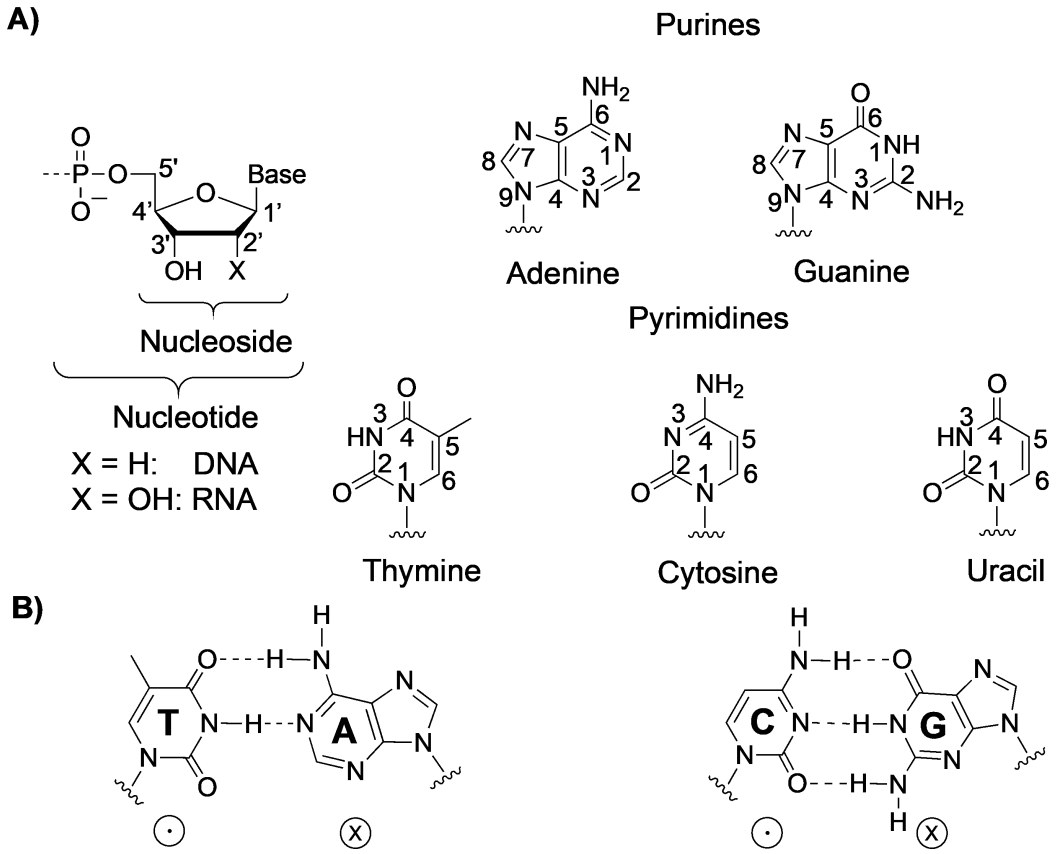


Figure 1.1. A) Nucleotides in DNA and RNA; B) Watson/Crick base-pairing in DNA. The relative orientation of phosphodiester backbones is indicated by the symbols, "⊙" for 5′- to 3′-end, and "⊗" for 3′- to 5′- end.

The 1′ - position of deoxyribose sugar is an anomeric centre (Fig. 1.2A). If a substituent attached to the 1′ - carbon lies on the opposite side of the sugar's 5′-hydroxyl group, it is known as the α-anomer. If the substituent is on the same face as the sugar's 5′-OH, it is the β-anomer. All the nucleotides in DNA or RNA adopt the β-configuration. L-

enantiomer of DNA (L-DNA) were explored for clinical purposes because of its high resistance to enzymatic digestion (Fig. 1.2A).[3] The bond joining the 1′-carbon of the deoxyribose sugar to the heterocyclic base is the glycosidic bond. Rotation about this bond leads to either *anti*- or *syn*- conformation (Fig. 1.2B), in which *anti*-conformation is generally favoured for nucleotides in DNA structures.

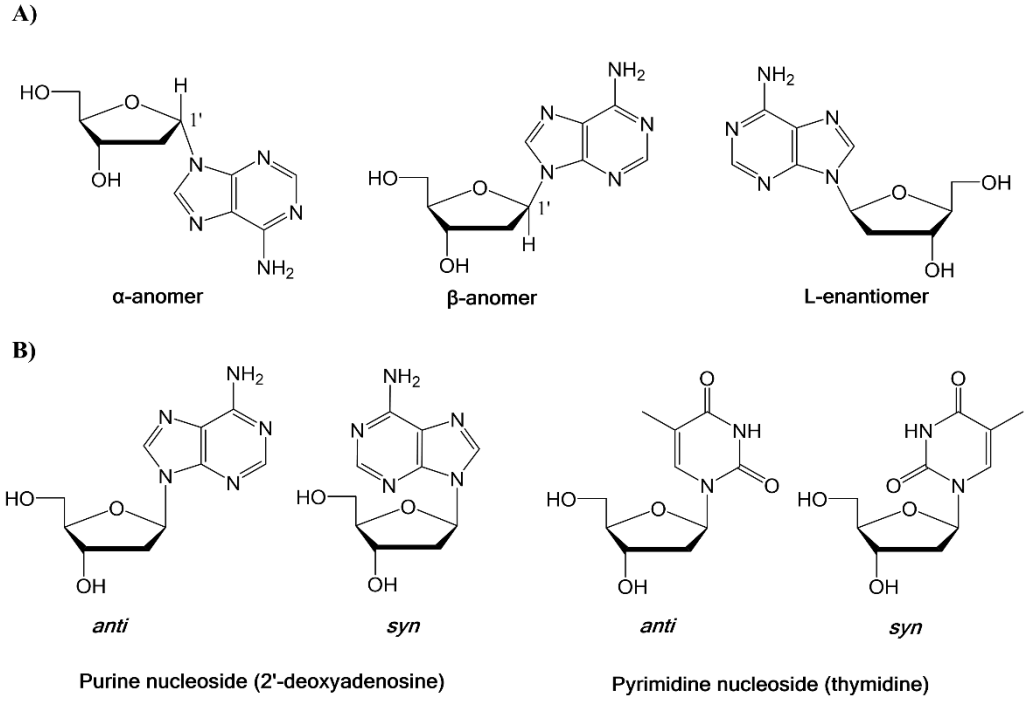


Figure 1.2. Sugar configuration and glycosidic bond conformation of a nucleoside.

Two complementary DNA strands can form a duplex through Watson/Crick hydrogen bonding (Fig. 1.1B).[1] As a result of π-π stacking, hydrophobic effect, and electrostatic repulsion between negatively charged phosphates, DNA duplex forms a helical structure with two grooves, the major groove (22 Å wide) and the minor groove (12 Å wide).[4] The two DNA strands are antiparallel to each other, with the 5′-end of one strand next to the 3′-end of the other (Fig. 1.3).

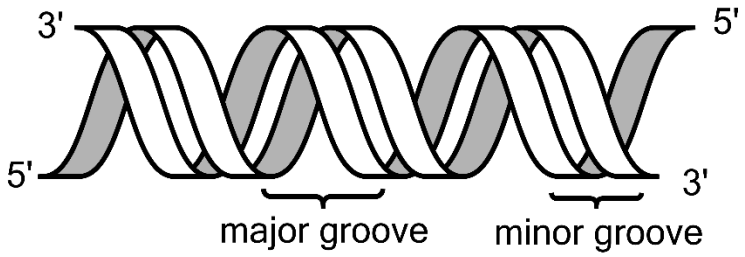


Figure 1.3. Directionality of the nucleic acid chain in antiparallel DNA duplex.

Although the DNA duplex is the most widely known form of DNA, several other high-order structures such as triplexes, G-quadruplexes and i-motifs are found in nature. Besides these naturally occurring structures, many artificial nucleic acids have been synthesised, including L-DNA, peptide nucleic acid (PNA),[5] locked nucleic acid (LNA)[6] and glycol nucleic acid (GNA).[7] Most artificial nucleic acids substitute the sugar ring for another moiety, and others can carry modifications on the heterocyclic base or the phosphate group.

1.1.1. DNA secondary structures

DNA was initially believed to exist only as a duplex, in which two oligonucleotide (ON) strands orientated in an anti-parallel pattern, forming right-handed double-helix, also known as B-DNA (Fig. 1.4, Table 1.1). The primary driving force is the exclusion of water, whereas Van der Waals contacts between the aromatic surfaces of the bases play a minor role. The twist between adjacent bps is 36°, and the space between them is 3.4 Å. The length of A/T base pairs (bps) and G/C bps are identical, which results in a constant width of 23.7 Å in the helix. Besides B-DNA, there are many other duplex variants exist, including A-DNA and Z-DNA. A-DNA is mainly found in dehydrated DNA samples, in which a lack of stabilising water causes changes in the twist between bp (33.6° *versus* 36° in B-DNA).[8] As a result, a more compact helix with ~11 bp per turn is formed (10 bp for B-DNA and 12 bp for Z-DNA). In contrast to A-DNA and B-DNA, which form a right-handed helix, Z-DNA forms a left-handed helix where the bps are nearly perpendicular to the phosphate backbone.[9] Z-DNA is majorly formed by chemically synthesised DNA strands d(CG)n that possessing the alternating purine and pyrimidine nucleotides. Z-DNA has a dinucleotide repeat unit that makes the backbone not smooth but appears as a ‘zigzag’ pattern. The negatively charged phosphates in Z-DNA are closer to each other than those in B-DNA (8 Å *versus* 11.7 Å).[10] The biological functions of Z-DNA and protein binding to Z-DNA are not thoroughly studied. It is believed that Z-DNA can be involved in the regulation of gene expression in cells.[11]

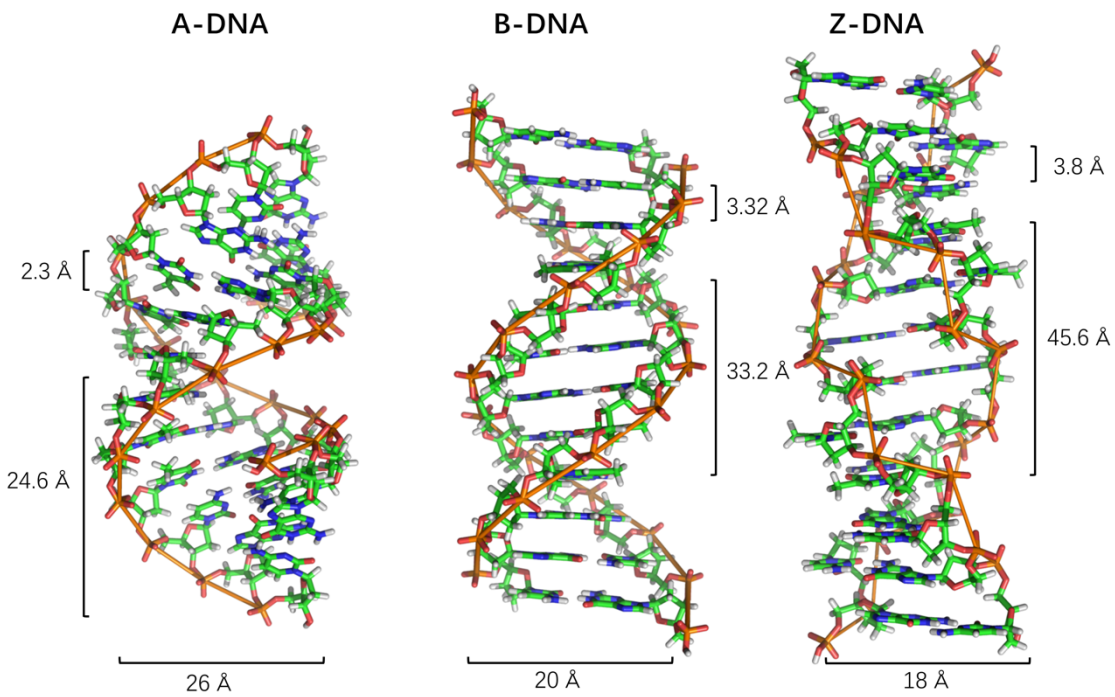


Figure 1.4. A-, B-, and Z- DNA.

Table 1.1. Structural parameters of A-, B- and Z- DNA.[10]

	A-DNA	**B-DNA**	**Z-DNA**
Pattern	RNA like	DNA like	Zigzag
Helix sense	Right-handed	Right-handed	Left-handed
Repeating unit	1 bp	1 bp	2 bp (alternating G-C sequences)
Rotation/bp	33.6°	35.9°	60°/2
Mean bp/turn	10.7	10.0	12.0
Inclination of bp to axis	+19°	-1.2°	-9°
Rise/bp along axis	2.3 Å	3.32 Å	3.8 Å
Pitch/turn of helix	24.6 Å	33.2 Å	45.6 Å
Mean propeller twist	+18°	+16°	0°
Glycosyl angle	anti	anti	C: anti, G: syn
Sugar pucker	C3′-endo	C2′-endo	C: C2′-endo, G: C2′-exo
Diameter	26 Å	20 Å	18 Å
Major groove	Narrow and very deep	Wide and quite deep	flat
Minor groove	Very broad and shallow	Narrow and quite deep	Very narrow and deep

1.1.2. Non-canonical DNA secondary structures

DNA does not exist exclusively as a duplex but can also form many other types of secondary structures. Watson/Crick hydrogen bonds between base pairs of A/T and G/C allow for the formation of helical antiparallel DNA duplexes. The H-bond donors and acceptors in the major groove of DNA duplex allow a third strand to bind, forming Hoogsteen or reverse Hoogsteen base pairing, resulting in the formation of a DNA triplex.[12] For a sequence that is rich in one particular nucleotide such as G or C, quadruplex structures such as G-quadruplex and i-motif structures, respectively, can be formed through Hoogsteen bonding. Investigations of these non-canonical secondary structures brought new insights into the use of synthetic nucleotides in various applications in material and life sciences.

1.1.2.1. DNA triplex

A DNA triplex is formed when a third single-stranded ON named triplex-forming oligonucleotide (TFO)[13] binds in a sequence-specific manner to the major groove of the double-stranded DNA (dsDNA) (Fig. 1.5).[14] Triplexes are formed when the pyrimidine bases of the TFO strand form Hoogsteen or reversed Hoogsteen hydrogen bonds[15] with the purine bases that have already been involved in Watson–Crick base pairing. DNA triplexes are generally classified as either parallel or antiparallel based on the orientation of the TFO relative to the homopurine strand of the dsDNA.

In forming the parallel triple-helix structure, a homopyrimidine TFO binds to dsDNA through Hoogsteen base pairing, in which the cytosine bases in the TFO must be protonated at the N3 atom (Fig. 1.5B). As a result, the formation of parallel triplexes at neutral pH is compromised.[16] Most studies on DNA triplexes have focused on the improvement of the formation of pH-sensitive parallel triplexes.[16]

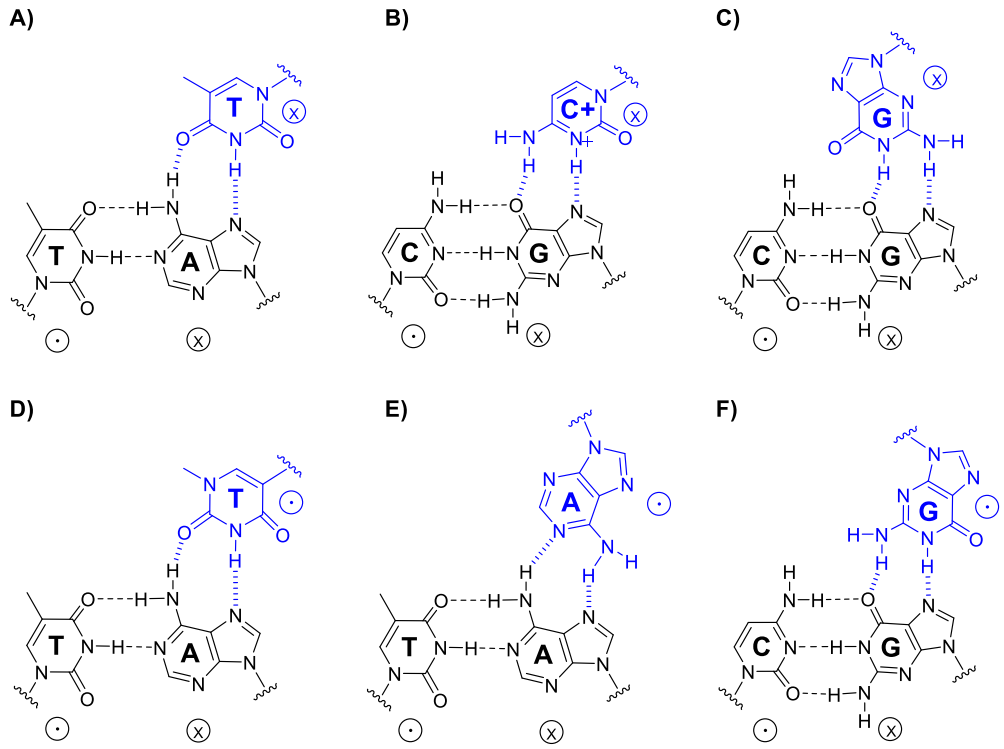


Figure 1.5. Hydrogen bonding patterns in DNA triplexes. A, B and C occur in parallel DNA triplexes. D, E and F occur in antiparallel DNA triplexes. Watson/Crick hydrogen bonding is presented in black dash line and Hoogsteen hydrogen bonding is presented in blue hash line. The relative orientation of phosphodiester backbones is indicated by the symbols, "⊙" for 5′- to 3′- end, and "⊗" for 3′- to 5′- end.

In contrast to parallel triplexes, antiparallel triplexes feature reversed Hoogsteen base pairings and do not require acidic pH (Fig. 1.5D, E and F). However, G-rich TFOs tend to form highly stable structures like G-quadruplexes rather than triplex. This sometimes can lead to controversial results whether the biological effects of G-rich TFOs are due to the formation of a G-quadruplex or because of the triplex formation.

1.1.2.2. G-quadruplex

Guanine-rich (G-rich) DNA or RNA sequences can form a highly stable secondary structure named a G-quadruplex (G4), in which four guanine bases self-assemble through the formation of eight Hoogsteen hydrogen bonds, forming a G-tetrad (Fig. 1.6A). Two or more G-tetrads can stack on top of each other, forming a G-quadruplex. The assembly of G-tetrads brings guanine O6 oxygens together, which creates electrostatic repulsion between O6. This repulsion can be neutralised by cations, typically Na^+ or K^+, which reside in the central cavity of the G-quadruplexes, stabilising the overall structure. For unmodified G4s, the thermal stability and association rate are dependent on the cations, following the order of $Li^+ < Na^+ < K^+$.[16]

G4s can form intermolecularly or intramolecularly and can display a wide variety of topologies depending on strand direction as well as loop size and composition of a sequence.[17] The G4 topology is classified as either parallel or antiparallel topology based on strand directionality changes. In parallel G4s, all of the G-tracts are aligned in the same direction, 5′ – 3′. When at least one of the four strands is oriented antiparallel to all others, a G-quadruplex is classified as antiparallel (Fig. 1.6). One typical difference between parallel and antiparallel structures is the conformation of 2′-deoxyguanosine, which is *anti*- in parallel G4s and *syn*- in antiparallel G4s. RNA G4s predominantly adopt parallel topology due to the preferred *anti*-conformation of guanosine.[18]

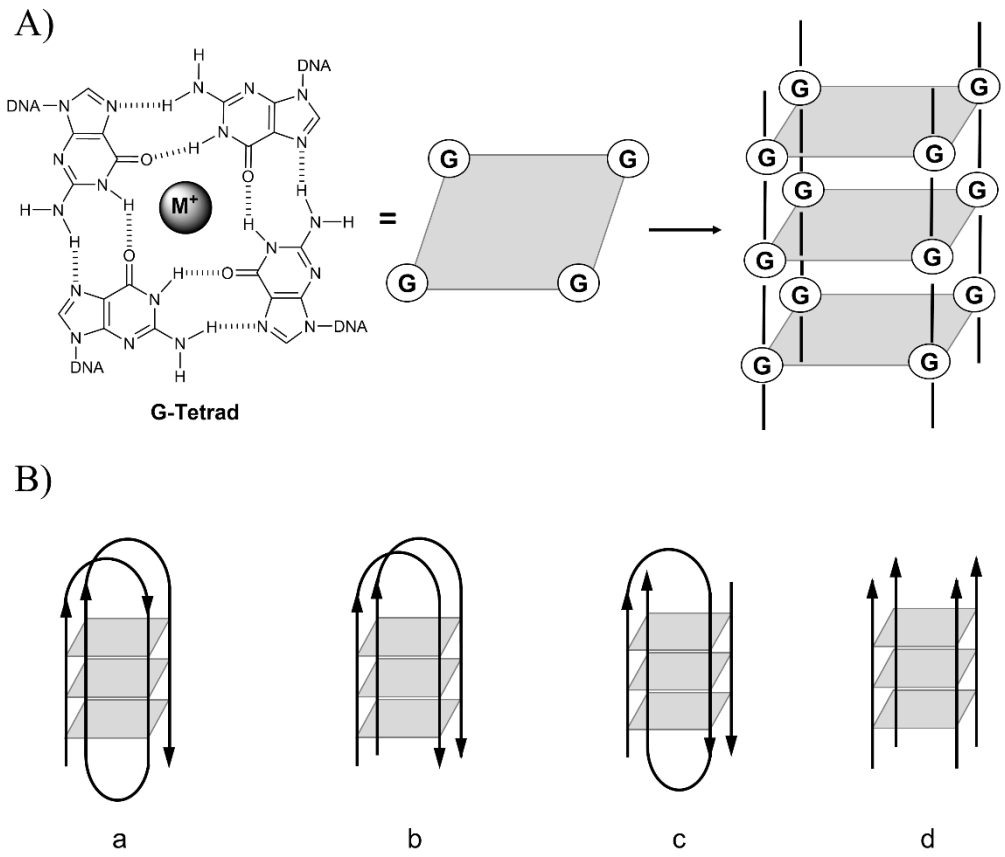


Figure 1.6. Illustration of G4 structures. A) Hydrogen bond formation in a G-tetrad, $M^+ = Na^+$ or K^+; B) G4s form intramolecular (a), bimolecular (b and c), or tetramolecular (d) structures. Depending on the direction of the strands or parts of a strand that form the tetrads, structures may be described as antiparallel (a, b, and c) or parallel (d).

The loops connecting the G-tracts of G4s are characterised as three types: propeller loop, which links the bottom G-tetrad with the top G-tetrad in adjacent parallel strands (Fig. 1.7a, (i)); lateral loop joins adjacent G-tetrad (Fig. 1.7b, (ii)); and diagonal loop that joins opposite G-strands and the directionality of adjacent strands must alternate between

parallel and antiparallel (Fig. 1.7b, (iii)). Variation of these loops can affect the stability and biological property of G4s. More importantly, it plays an important role in interacting with proteins. Increasing the loop length between G-tracts may result in decreased thermal stability of the G4s. However, this reduction in thermal stability may be accompanied by a topological change in the complex.[19]

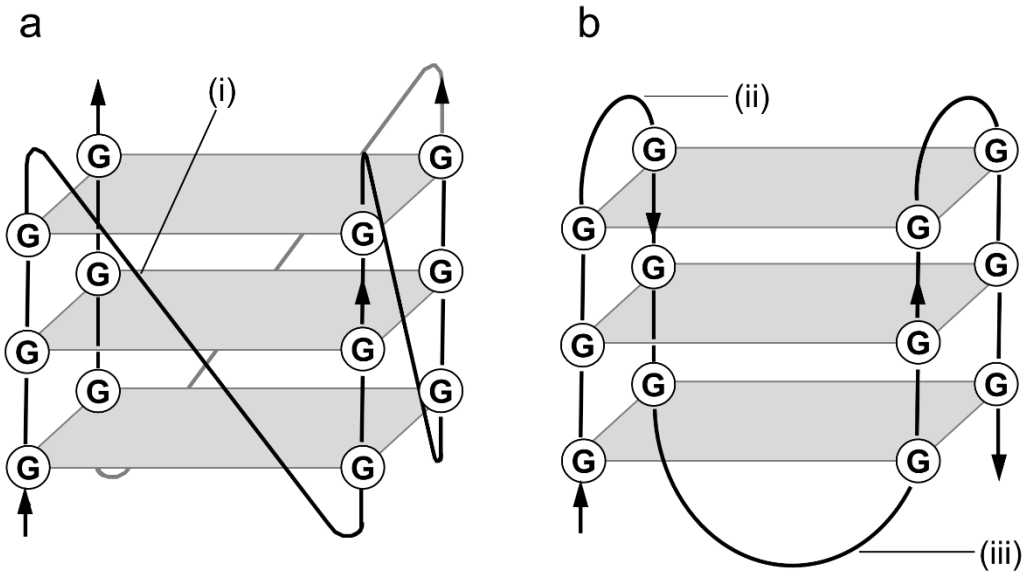


Figure 1.7. Topologies and types of loops in monomolecular G4s: (a) parallel and (b) antiparallel G4s. Loops: (i) Propeller loop; (ii) Lateral loop; and (iii) Diagonal loop.

G4s have been shown to have potential biological relevance within the human genome. A telomeres sequence with $(G_3TTA)_n$ repeat at the end of chromosomes is known to fold into several G4 topologies under various conditions. Telomeric DNA has been linked to genome stability, ageing and cancer.[20] G4s may also form in promoter regions of some oncogenes and play a role in regulating gene expression during transcription.[21],[22] Moreover, some proteins exhibit specific affinity for G4, such as heterochromatin protein 1α (HP1α), which shows preferential binding to parallel rather than antiparallel DNA and RNA G4s.[23] The study of G4 structures and their properties is of great value in designing anti-cancer related drugs.

1.1.2.3. i-Motif

G-rich DNA sequences can fold into four-stranded G-quadruplexes in the presence of metal ions, while the complementary cytosine-rich (C-rich) sequences can fold into i-motifs at low pH as a result of hydrogen bonding between a pair of cytosines sharing a

proton (C-C+), which then leads to the formation of an i-motif quadruplex structure (Fig. 1.8). Although poly-cytosine nucleotides were first studied in the early 1960s, it was not until 1993 that the i-motif was first structurally characterised at acidic to near-neutral pH.[24]

i-Motifs are four-stranded structures formed by C-rich DNA. Similar to G4s, i-motifs can also be formed intramolecularly (Fig. 1.8a) and intermolecularly (Fig. 1.8b and c). The requirement of protonated cytosines in the formation of the i-motif leads to a strong pH sensitivity, which has been used in the design of biosensors, nanomachines, and molecular switches.[25]

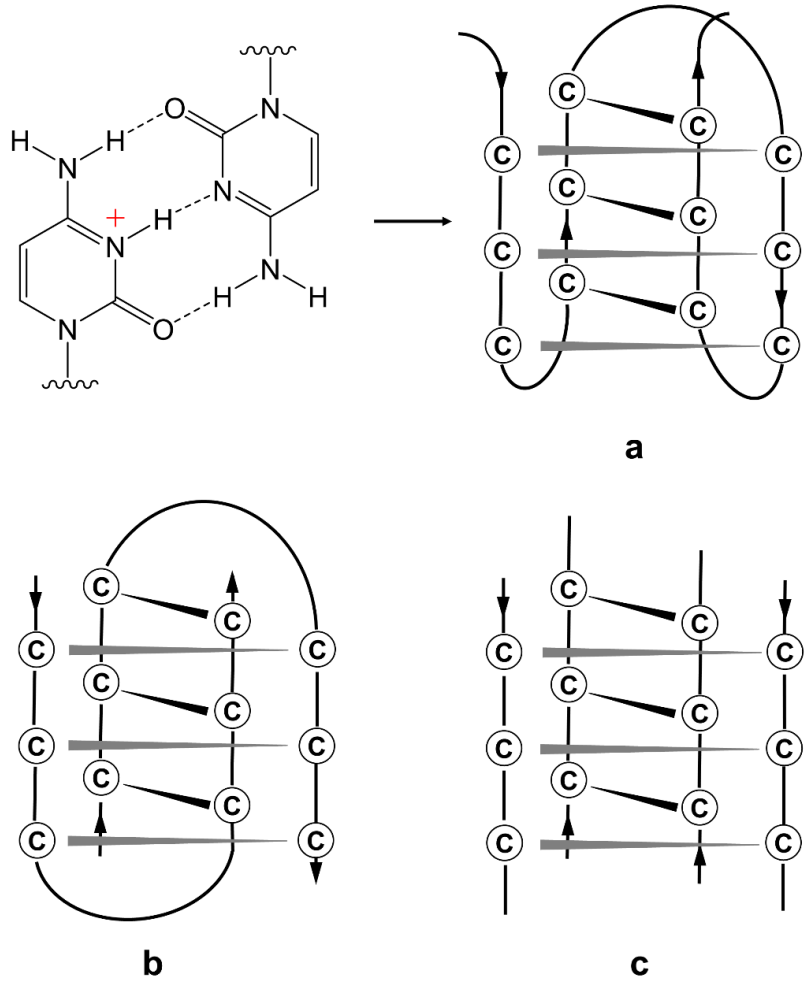


Figure 1.8. Schematic i-motif structure showing hydrogen bonding in cytosine-protonated cytosine (C·C+) base pair and several i-motif topologies (a, b and c). i-Motif can form intramolecularly (a) and intermolecularly including bimolecular (b) and tetra-molecular (c) i-motifs.

Little attention has been paid to the C-rich strand due to the reason that i-motifs are traditionally believed to be less stable at neutral pH and therefore, the biological relevance of these structures is a matter of debate. However, i-motifs have been recently observed in the nuclei of human cells and shown to play a role in cell reproduction.[26] C-rich DNA regions are commonly found in gene regulation portions of the genome, such as telomeric DNA, which is capable of forming i-motif structures. These G-rich (on one strand) and C-rich (on the complementary strand) sequences, which are found in the telomeres of most eukaryotic organisms, are ideal for the formation of both G4s and i-motifs. It is believed that these four-stranded structures could be involved in a variety of biological processes, such as telomere maintenance, regulation of oncogene expression, and DNA repair.

1.1.3. Strategies for targeting genomic DNA

1.1.3.1. Genome editing

One promising approach for targeting genomic DNA is genome editing, also called genome engineering or gene editing, which relies on molecules that cut or open the target duplex. Genome editing uses various technologies to induce permanent changes in the genomic DNA sequence of a cell or an organism. A common form of genome editing relies on DNA double-stranded break (DSB) repair mechanisms. There are two major pathways to repair DSB: non-homologous end joining (NHEJ) and homology-directed repair (HDR).[27] NHEJ uses enzymes to directly join the DNA ends from DSB, while HDR uses a homologous sequence as a template for regeneration of missing DNA sequences at the breakpoint (Fig. 1.9), and it is more accurate. The desired change can be inserted at the site of the DSB by using HDR.

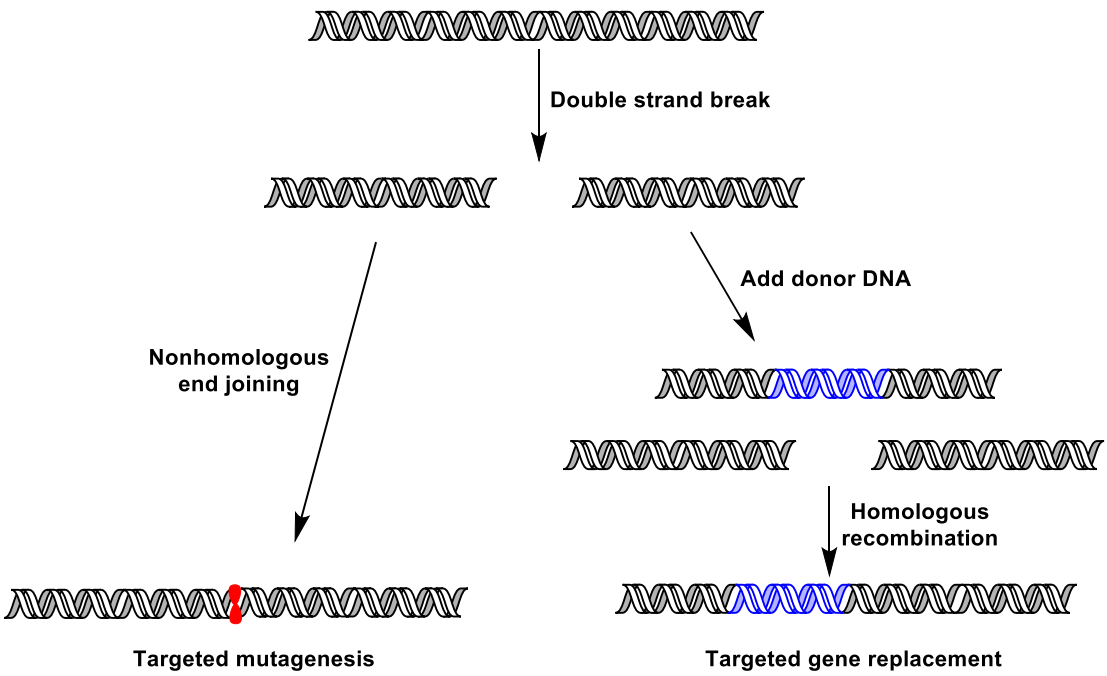


Figure 1.9. Different DNA repair pathways used to repair DSB.

The key to genome editing is specifically creating a DSB at a point within the genome. Commonly used restriction enzymes can cut DNA effectively but generally recognise and cut at multiple sites. Three distinct classes of nucleases have been discovered and bioengineered to create site-specific DSB, the meganucleases, Zinc finger nucleases (ZFNs), and transcription-activator like effector nucleases (TALENs). Meganucleases have the benefit of causing less toxicity in cells than methods such as ZFNs. Yet, the construction of sequence-specific enzymes for all possible sequences is more costly and time-consuming than ZFNs and TALENs. Recently, the clustered regularly interspaced short palindromic repeats (CRISPR/Cas9) system was more efficient and accurate in genome editing.

1.1.3.1.1. Meganucleases

Meganucleases are enzymes in the endonuclease family capable of recognising and cutting large DNA sequences (from 14 to 40 base pairs).[28] Meganucleases are found commonly in microbial species. Because of the unique property of recognising long DNA sequences, meganucleases show high sequence specificity.[29] However, it is practically impossible to find the specific meganuclease required for a chosen DNA sequence. To overcome this challenge, various attempts including mutagenesis,[30] high throughput screening methods,[31] and rationally designed meganuclease that alters the DNA interacting amino acids of the meganuclease to increase sequence specificity has been

made.[32] A more recent approach uses computer models to predict, as accurately as possible, the activity and specificity of the modified meganucleases.[33] A large protein bank has been created containing tens of thousands of units, which can be combined to obtain meganucleases that can recognise the target site, thus providing research and developmental tools for fundamental analysis, health, agriculture and industry research.[34]

1.1.3.1.2. Zinc finger nucleases

ZFNs are useful and powerful reagents for genome editing. Unlike meganucleases, ZFNs fuse a non-specific DNA cutting enzyme with specific DNA sequence recognising peptides. ZFNs have been used to direct DSBs at a specific genomic point and repair these breaks by normal cellular repair pathways. It has been reported that engineered ZFNs can induce targeted knockout mutations in cultured human cells and other model organisms such as zebrafish.[35]

The design of ZFNs is based on naturally occurring structural motifs, the Cys2His2 zinc-finger motif, which is well characterised and has distinct metal (Zn^{2+}) and DNA binding properties. The Zn^{2+}, found in 8% of all human proteins, is vital in organising their three-dimensional structures. The C-terminal part of each finger is responsible for the specific recognition of the DNA sequence. The recognised sequences are as short as three bps. Proteins specifically recognising sequences around 20 base pairs can be obtained by combining 6 to 8 zinc fingers. Therefore, it is possible to fuse a constructed protein with an endonuclease that can induce a DSB and use these engineered ZFNs as genome editing tools.[36]

There are several approaches for designing ZFNs for specific sequences. The most commonly used one involves combining zinc-finger units with known specificities. Various techniques have been developed using bacteria, yeast, or mammal cells to identify the best specificity and best cell tolerance combinations.[37]

1.1.3.1.3. TALENs

Similar to ZFNs, TALENs are specific DNA-binding proteins composed of the transcription activator-like effectors (TALEs) fused to the DNA cutting domain of a nuclease, which can bind to practically any desired DNA sequence. TALEs consist of repeated domains, and each domain contains a highly conserved sequence of 34 amino acids that can recognise a single DNA nucleotide within the target site. Each repeat is conserved, except for the so-called repeat variable di-residues (RVDs) at amino acid

positions 12 and 13, determining the DNA sequence that the TALE will bind. The one-to-one correspondence between the TALE repeats and the DNA sequence makes the recognition simple and straightforward.

Along with ZFNs, TALEN is a prominent tool in the genome-editing field. The significant difference between these engineered nucleases is in the DNA recognition peptide: ZFNs rely on Cys2-His2 zinc fingers, whereas TALEN constructs TALEs. Cys2-His2 Zinc fingers typically happen in repeats that are three bps apart. Each finger is entirely independent, and the binding capacity is impacted by its neighbour. TALEs, on the other hand, are found in repeats with a one-to-one recognition ratio between the amino acids and the recognised nucleotide pairs. In comparison with ZFNs, TALENs have higher binding specificity towards DNA, lower off-target effects, and easier construction of DNA-binding domains.[29]

1.1.3.1.4. CRISPR-Cas9

Early methods that use ZFNs or TALENs are expensive, slow, and challenging to design and perform compared to newer techniques based on the CRISPR-Cas system. CRISPR is a family of DNA sequences found in the genomes of prokaryotic organisms such as bacteria and archaea. These sequences are found in approximately 50% sequenced bacterial genomes, and nearly 90% sequenced archaea.[38]

CRISPR are derived from DNA fragments of bacteriophages that had previously infected the prokaryote. These sequences play a vital role in the antiviral (i.e., anti-phage) defense system of prokaryotes as they are used to detect and destroy DNA from similar bacteriophages during subsequent infections.

Cas (CRISPR-associated) enzymes are a part of a bacterial immune system that incorporates short, viral DNA sequences into the bacterial genome. The process is not yet entirely understood due to its complexity. The incorporated viral DNA sequences can be transcribed into guide RNA (gRNA) when needed. When the same kind of virus tries to infect the bacterium again, the CRISPR system can cut the invading viral DNA by using the gRNA and Cas enzyme. The combination of gRNA and Cas enzyme to cleave the target DNA is adopted as a CRISPR-Cas system for genome editing in laboratories. It was recognised by a Nobel Prize in Chemistry in 2020, being awarded to Emmanuelle Charpentier and Jennifer Doudna.

The CRISPR-Cas systems commonly utilise the Cas9 enzyme. The CRISPR-Cas editing system has a wide range of applications, including fundamental biological research, development of biotechnology products, and more importantly, treatment of diseases.

The CRISPR-Cas9 system starts with the annealing of CRISPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA) together, forming a complete gRNA. Next, gRNA binds to the Cas9 enzyme, forming ribonucleoprotein (RNP). The RNP is then delivered into the cells, and the gRNA directs RNP to the DNA target inside the cells through base pairing with a protospacer (target). Adjacent to the protospacer is a short sequence that is only a few bases long, called protospacer adjacent motif (PAM). PAM is essential for the target DNA to be recognised by the Cas9 enzyme.[38] Finally, the DNA target is cut precisely by the Cas9 enzyme (Fig. 1.10).

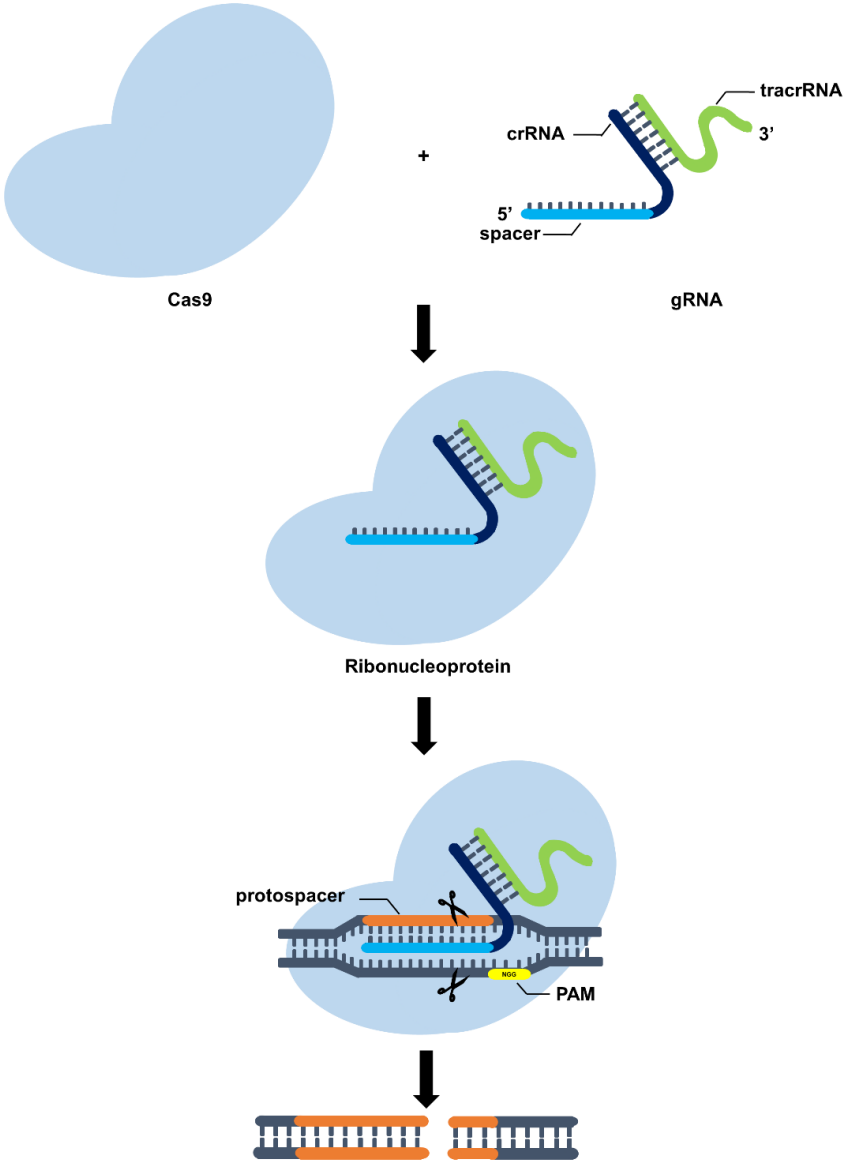


Figure 1.10. Schematic illustration of gene editing using CRISPR-Cas9 system.

CRISPR/Cas9 genome editing technique is remarkably reliable and flexible for many applications. This system offers to scaling far beyond the capacities of ZFNs and TALEN based methods. It has been proven to be more efficient for site-directed genome editing in a wide variety of biological systems. However, besides its 'off-target' effects, ethical concerns are involved when using CRISPR genome editing. In 2015, Chinese scientists attempted to alter the DNA of non-viable human embryos using CRISPR to correct a mutation that causes beta-thalassemia, a lethal heritable disorder. The experiments successfully changed only some of the intended genes but had off-target effects on other genes. These results indicated that CRISPR is not ready for clinical application in reproductive medicine.

1.1.3.1.5. CAS9-constructs

Among various Cas enzymes found in bacteria, Cas9 and Cas12a are the most commonly used in genome editing. The Cas9 enzyme is a non-specific endonuclease derived from *Streptococcus pyogenes* (*S. pyogenes*). The Cas12a enzyme (also named as Cpf1) is often derived from *Acidaminococcus* or *Lachnospiraceae*. Cas9 and Cas12a can target and cleave DNA complementary to gRNA in the presence of divalent metal ions like Mg^{2+}.[39] However, it has been reported that some other Cas enzymes, such as *Francisella tularensis novicida* (Fno) Cas12a, FnoCas9, and *Streptococcus pyogenes* Cas9 (SpyCas9) can cleave DNA without a gRNA in the presence of Mn^{2+} ions.[39] These observations indicated that alongside the Cas9/Cas12a ortholog used, the identities and levels of intracellular metal ions could influence the result of genome editing.

The use of CRISPR-Cas9 as a therapeutic reagent is impeded by the off-target effects. Although rationally designed *S. pyogenes* Cas9 (SpCas9) variants display higher specificities than the wild-type SpCas9 protein, the efficiency of these attenuated Cas9 variants in human cells are often inadequate.[40] Other Cas9 constructs with higher specificities, such as Sniper-Cas9, obtained from Escherichia coli **(***E. coli*), have been designed. Sniper-Cas9 shows reduced off-target effects and works well in a preassembled RNP to allow genome editing in human cells.[40]

Soon after the discovery of CRISPR genome editing, scientists produced a catalytically inactive enzyme named dCas9 (dead Cas9).[41] dCas9 cannot cut DNA but can still be targeted to a specific genomic site. After binding, dCas9 can inactivate gene transcription at the target site by preventing the binding of the cellular transcription machinery to the gene. The relevant protein is not produced without transcription, and the gene is

effectively silenced until the cell naturally eliminates the dCas9 enzyme. This approach was named CRISPR interference (CRISPRi) as CRISPR is used to temporarily silence a gene rather than cutting the DNA. CRISPRi is similar to RNA interference (RNAi), the technology for studying gene function that does not involve altering the DNA but more efficient and easier to design.[42]

1.1.3.1.6. Programmable deaminases

The above-mentioned genome-editing technologies introduce DSB at a target site as the first step of gene correction. These approaches are generally inefficient and may induce small insertions or deletions (indels) at a target site. A new tehcnique called 'base editing' was invented for genome editing, allowing direct, irreversible conversion of one target DNA base into another in a programmable manner without creating DSB and indels.[43] This new approach utilises programmable deaminases, which comprise a DNA-binding module and a cytidine deaminase that converts cytosine (C) to uracil (U) (G or A, to a lesser extent) within a window of approximately five nucleotides at a target site. Programmable deaminases can correct point mutations that cause genetic diseases in human cells, animals, and plants.[43] The direct conversion of one DNA base to another at a programmable target site without DSBs could increase the gene correction efficiency. Four different classes of programmable deaminases have been reported: (1) base editors that fuse dCas9 or D10A nCas9 and rAPOBEC1, a cytidine deaminase from rat;[44] (2) target-AID composed of dCas9 or nCas9 and PmCDA1, an activation-induced cytidine deaminase (AID) ortholog from sea lamprey,[45] or human AID;[46] (3) CRISPR-X composed of dCas9 and sgRNAs linked to MS2 RNA hairpins to recruit a hyperactive AID variant fused to MS2-binding protein;[47] (4) zinc-finger proteins or TALEs fused to a cytidine deaminase.[48]

It has been shown recently that adenine base editors (ABEs) can convert A•T base pairs to G•C base pairs directly and correct the pathogenic Hutchinson-Gilford Progeria Syndrome (HGPS) mutation in cultured fibroblasts that are derived from children with progeria.[49] In patient-derived cells, base editing using ABE efficiently corrected the pathogenic allele (87–91% correction), subsequently decreasing the progerin protein level and rescued nuclear morphology abnormalities. A very low degree of off-target editing was detected in patient-derived fibroblasts.[49] These results indicate that *in vivo* base editing is the potential and promising treatment for HGPS and other genetic diseases.

Despite broad interest in base editing, genome-wide target specificities of these enzymes remain unclear due to a lack of appropriate methods. It has been speculated that Cas9 nickase-induced mutations at cytosines in the non-template site might arise from their exposure to cellular cytosine deaminases.[50] The mutation of the non-template site might induce genomic toxicity at that site.[51]

1.1.3.2. Antigene strategy

Meganuclesases, ZFNs, and TALENs are modular enzymes used to recognise and cut DNA sequences, allowing the designed sequence to participate in natural DNA repair mechanisms or homologous recombination.[52-55] However, existing strategies for targeting dsDNA rely on engineered proteins with very narrow specificity. [56, 57] The recently discovered CRISPR-Cas9 system can be engineered to target almost any DNA sequence but depends on the initial unwinding of the DNA helix by a large exogenous Cas9 protein.[58]

Another approach using chemically modified ONs to specifically bind to native DNA can induce genomic changes and/or interfere with gene expression, the called antigene strategy.[59] ONs are perfectly suitable for this purpose because of the unique recognition through Watson/Crick base pairing. Moreover, in comparison with proteins, ONs are much smaller in size. In antigene strategy (Fig. 1.11), ONs (TFO) can bind to genomic dsDNA, forming a DNA triplex and thus prevent transcription of the target DNA into mRNA (Fig. 1.11a). Selective modification of the genome, such as site-specific mutagenesis or homologous recombination through triplex delivered donor DNA via DNA repair, could also be achieved (Fig. 1.11b). Further applications arise, including cleavage, site-specific cross-linking or alkylation, which can be accomplished using a triplex-targeted chemical modification (Fig. 1.11c).[60]

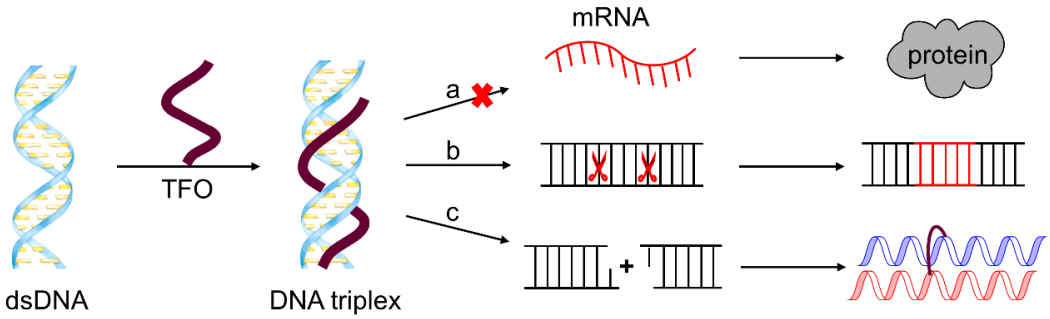


Figure 1.11. TFO based antigene strategy and corresponding applications. a) Transcription inactivation; b) Directed mutagenesis through homologous recombination; c) DNA cleavage or crosslinking at selected site.

There are several approaches using chemically modified ONs to target dsDNA (Fig. 1.12). Triplex formation (Fig. 1.12, strategy a) relies on a TFO binding to the major groove of dsDNA.[61-63] However, such binding is preliminary limited to homopurine regions and depending on the type of triplexes formed has additional restrictions. Triplex invasion (Fig. 1.12, strategy b) uses homopyrimidine TFO analogues to target homopurine regions, forming extremely stable triplex and P-loop structures.[64, 65] Another strategy is strand invasion (Fig. 1.12, strategy c), in which a single-stranded DNA (ssDNA) bind to supercoiled DNA by hybridising to one strand of the duplex through Watson/Crick base-pairing while displacing the other.[66] Because strand invasion is relatively slow and inefficient, dual duplex invasion (DDI), also known as double strand invasion or double duplex invasion (Fig. 1.12, strategy d) is more attractive but challenging in the design.[67] A pseudo-complementary duplex has to be obtained: the affinity of invading strands must be low to each other, but the affinity to unmodified strands has to be high, exceeding the stability of the native duplex.

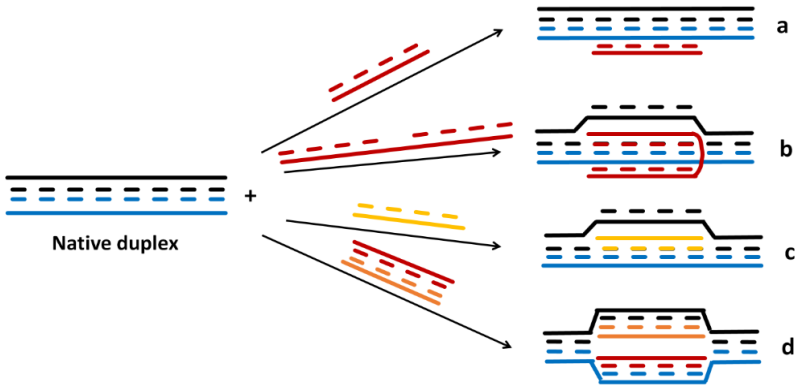


Figure 1.12. dsDNA targeting strategies by modified nucleic acids. a) Triplex formation by TFO; b) Triplex invasion; c) Strand invasion: ssDNA invades the dsDNA target forming Watson/Crick hydrogen bonds, leaving the non-complementary strand as a loop; d) Double duplex invasion: modified duplex invades into dsDNA, occupying both target strands via Watson/Crick base pairing.

1.1.3.3. Minor groove binders

Proteins can bind to dsDNA in a site-specific manner by distinguishing between the major and minor grooves' geometry and making close contact with individual bases. Many studies have been focused on the development of small non-natural ligands binding to the minor groove of DNA, mimicking a DNA–protein interaction. Groove binding usually does not induce huge structural/conformational changes in the DNA duplex. This mode of binding may be considered similar to a standard lock and key recognition.[68]

Minor groove binders (MGBs) are usually isohelical, crescent-shaped molecules, which are compatible with the shape of the minor groove of the DNA helix. MGBs have found application in treating many diseases, including cancer, parasitic, bacterial and viral infections.[69] The first two MGBs reported were distamycin A and netropsin (Fig. 1.13). These natural molecules are concave-shaped and characterised by repeating N-methylpyrrole units with one or more positively charged nitrogen atoms at the end, referred to as 'shape-selective' binders.[70] They selectively interact with A·T rich regions (at least four A·T bps) in the minor groove through hydrogen bonding interaction (between the bps and the amides) and electrostatic stabilising interactions (between the protonated amines and negatively charged phosphates).[71-74] These molecules were reported as inhibitors of Werner and Bloom syndrome helicases and dual topoisomerase I/II inhibitors.[75, 76]

Distamycin A itself had no anticancer effects, but it has been the base of interesting anticancer derivatives.[77] Netropsin showed antiviral activity as well as being active against Gram-positive and Gram-negative bacteria (Fig. 1.13).[78] There are three moieties identified in the structure of distamycin: (i) 'head' (formamide), (ii) 'polyamide core' (or DNA binding region, the pyrroles connected by peptide bonds) and (iii) 'tail' (cationic amide) (Fig. 1.13). Various modifications have been introduced to these three moieties to improve the anti-tumour activity of distamycin A. Tallimustine (Fig. 1.13), as a derivative of distamycin A, with the 'head' substituted by a nitrogen mustard and one of the pyrroles of the 'polyamide core' substituted by a benzoyl group, has shown excellent anti-tumour activity in preclinical tests. However, it also shows strong myelotoxicity.[79] Novel cinnamic analogues such as PNU 157911 and PNU 160666 (Fig. 1.13) have been reported showing improved activity and less myelotoxicity.[80]

A series of α-halogenoacrylamido derivatives of distamycin has been designed and synthesised, which displayed improved activity.[81] For instance, brostallicin (Fig. 1.13) showed potent and broad anti-tumour activity accompanied by strongly reduced myelotoxicity.[82, 83] In brostallincin, a guanidinium cation has been introduced in the 'tail' and an α-bromo-acrylamide group has been used as the 'head'.

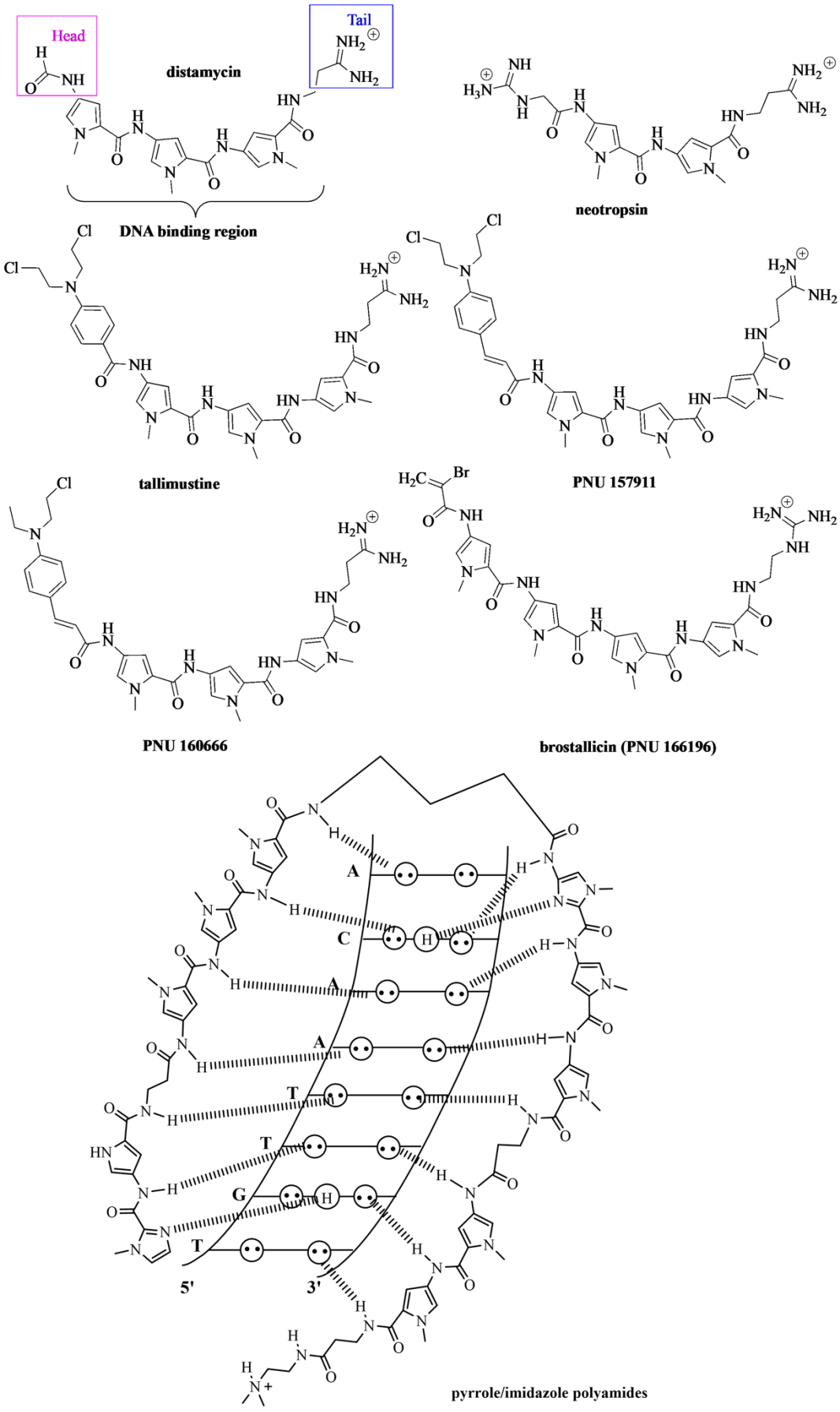


Figure 1.13. Structures of naturally occurring and synthetic hybrid minor groove binders.

Despite various distamycin and netropsin analogues being developed, these analogues have limited efficacy in recognising long DNA duplexes. Dervan et al.[84, 85] and other groups[86] developed a series of polyamides composed of pyrrole (Py) and imidazole (Im) rings (pyrrole/imidazole polyamides, Fig. 1.13) to achieve better sequence specificity. It has been reported that this 'hairpin' structure of pyrrole/imidazole polyamides bind to DNA in the minor groove with high affinity and specificity. By carefully designing and selective placement of an aliphatic *β*-alanine (*β*) residue paired side-by-side with either a Py or Im amino acid, discrimination of G·C/C·G from A·T/T·A or vice versa can be achieved.[87] Moreover, by introducing another aromatic amino acid, 3-hydroxypyrrole (Hp), HP–Im–Py polyamides can distinguish all four Watson/Crick bps in the minor groove of DNA.[87] These polyamides show the successful design of synthetic MGBs that can bind to specific DNA sequences and compete with protein–DNA binding interactions in the minor grooves of DNA.

1.1.4. Molecules targeting dsDNA

Several artificial nucleic acids, such as peptide nucleic acids (PNAs)[88] and locked nucleic acids (LNAs),[89] have been designed and synthesised to improve the triplex affinity of TFOs. PNAs, also named polyamide nucleic acids, were firstly invented by Nielsen *et al.*,[5] to mimic DNA or RNA structure with a neutral backbone connected by peptide bonds (Fig. 1.14, PNA). It has various purine or pyrimidine bases linked to the backbone, therefore can hybridise with the complementary DNA or RNA strand via Watson/Crick base-pairing. PNAs can form PNA/DNA duplex with higher thermal stability than DNA/DNA duplex because of their neutral backbone. Synthetic PNAs have been used in molecular biology procedures, diagnostic assays, and antisense therapies.[90, 91]

Besides modifications on the backbone or nucleobases such as PNAs, modifications in the sugar moieties have also been well developed. LNA, also known as 2′,4′-bridged nucleic acid (BNA), as an example of RNA modifications, was independently synthesised in 1997[92, 93] and has been applied in chemistry, biochemistry and biotechnology successfully (Fig. 1.14, LNA).[94, 95]

PNA, LNA and their derivatives (Fig. 1.14) have been widely used in triplex formation[62, 89, 96] and triplex invasion.[97-100] PNAs can also invade into dsDNA through strand invasion with higher efficiency than using modified DNA or RNA because of the uncharged backbone.[101] For dual duplex invasion (DDI), PNAs have been modified

using pseudo-complementary nucleobases 2,6-diaminopurine and 2-thiouracil as derivatives of A and T, respectively.[67] Although PNA and modified PNAs have excellent chemical stability and high binding affinity towards complementary DNA and RNA, several drawbacks such as the tendency to aggregate, low solubility in water, and requirement of high salt concentrations in DDI necessitate further functionalisation to achieve targeting of DNA under physiological ionic strength.[102-105] One possible approach to overcome this challenge is to introduce modifications in PNA that can perform interstrand cross-linking (ICL), such as furan moieties (Fig. 1.14, Furan-PNA monomer). Furan-modified PNA probes allow targeting of both ssDNA as well as dsDNA sequences through subsequent crosslinking.[106] Another promising approach is to introduce a positive charge onto the PNA backbone. It has been reported that positively charged PNA binds more strongly to complementary DNA and RNA than negatively charged PNA at low salt concentrations (0 – 100 mM Na^+), whereas the trend is reversed at medium to high salt concentrations (250 – 1000 mM Na^+).[107] Incorporation of a positively charged monomer into PNA, *N*-(2-aminoethyl)-D-lysine (Fig. 1.14, chiral PNA monomer), destabilised the PNA/PNA duplex, whereas the corresponding PNA/DNA duplex was stabilised. This modification triggered the DDI of chemically modified PNA duplexes into dsDNA with great specificity.[102] One can also incorporate cyclic derivatives in the peptide backbone to increase the affinity and sequence selectivity of PNA. It has been reported that PNA with *N*-(2-aminoethyl) glycine backbone (Fig. 1.14, aegPNA monomer) is potentially capable of binding dsDNA through triplex formation, triplex invasion, strand invasion as well as DDI. However, using aegPNA to target dsDNA failed in many cases due to the decreased binding affinity.[108] Another PNA analogue with cyclic structures as part of the backbone, the acpcPNA (Fig. 1.14, acpcPNA monomer) exhibits significant features. In contrast to conventional PNA systems that require three or more equivalents of PNA, only 1.5 equivalents acpcPNA are sufficient to achieve efficient DDI.[109] Another way of improving DDI efficiency is to use energetically favourable γ-modified PNAs (Fig. 1.13, γ-PNA monomer), which contain a stereogenic centre at the γ-position of each monomer. The stereogenic centre forces the single-stranded oligomer into a pre-organised right-handed helix, thus reducing the entropic penalty during duplex formation. In combination with the elimination of the electrostatic (enthalpic) penalty because of the neutral PNA backbone, γ-PNA shows a very high sequence selectivity. It has been used to directly and rapidly detect bacterial and fungal pathogens from blood without culturing.[110]

LNAs were used for targeting dsDNA through triplex formation and triplex invasion. However, the use of LNAs in DDI is problematic due to the increased thermal stability of the LNA/LNA duplex.[93] Bis-locked nucleic acids (bisLNAs) can invade into dsDNA, forming clamp-constructs, in which Hoogsteen-arm was formed first, followed by Watson–Crick arm invasion. BisLNA invasion is more efficient than other LNA analogues, and the efficiency can be further improved by substituting LNA with 2′-glycylamino-LNA or 2′-(3-piperazino-propanoyl)-amino-LNA, which provides a positive charge (Fig. 1.14).[111]

One promising way of destabilising the duplex of the probe is the insertion of a pair of aromatic chromophores, usually pyrenes, in the helix structure. If two pyrene units are positioned opposite each other, the accommodation of these insertions in the duplex leads to a local unwinding of the double helix and destabilising the DNA duplex, creating an energetic hotspot. Various pyrene-functionalised LNA and 2′-*O*-modified RNA analogues were designed, synthesised, and evaluated for strand invasion and DDI.[112] Incorporation of a single pyrene-functionalised 2′-amino-α-L-LNA monomer (Fig. 1.14) into ON backbone led to excellent strand invasion property as well as for DDI.[113-115] ONs incorporated with +1 interstrand zippers of 2′-*O*-(pyren-1-yl)methyl-RNA monomers (Fig. 1.14) has also been reported as invasion probes for recognition of dsDNA.[116] However, the recognition efficiency of these probes was influenced by the composition of the energetic hotspots, in which C and/or U monomers displayed favourable dsDNA-binding affinities, whereas hotspots constructed using one or two G monomers (or two A monomers) are the least activated constructs for dsDNA recognition.[116]

Another type of neutral artificial nucleic acid, morpholino oligonucleotide (PMO, Fig. 1.14),[117-119] is built upon a backbone of morpholine rings connected by phosphorodiamidate linkages. PMO binds to complementary DNA or RNA sequences through Watson/Crick base pairing with higher affinity because of its uncharged backbone. It possesses potential in the application of targeting dsDNA. PMOs exhibit higher affinity for their target nucleic acid sequences, greater resistance to enzymatic degradation and extremely low toxicity. In contrast to DNA-based modified backbones, RNA-PMO hybrids are not substrates for RNase H, and thus the mRNA is not degraded. These properties were used to develop exon skipping PMOs that found their application in clinics with FDA-approved eteplirsen and golodirsen.[120, 121] However, despite their

nonionic nature, PMOs still have insufficient delivery into tissues, especially in the heart.[122]

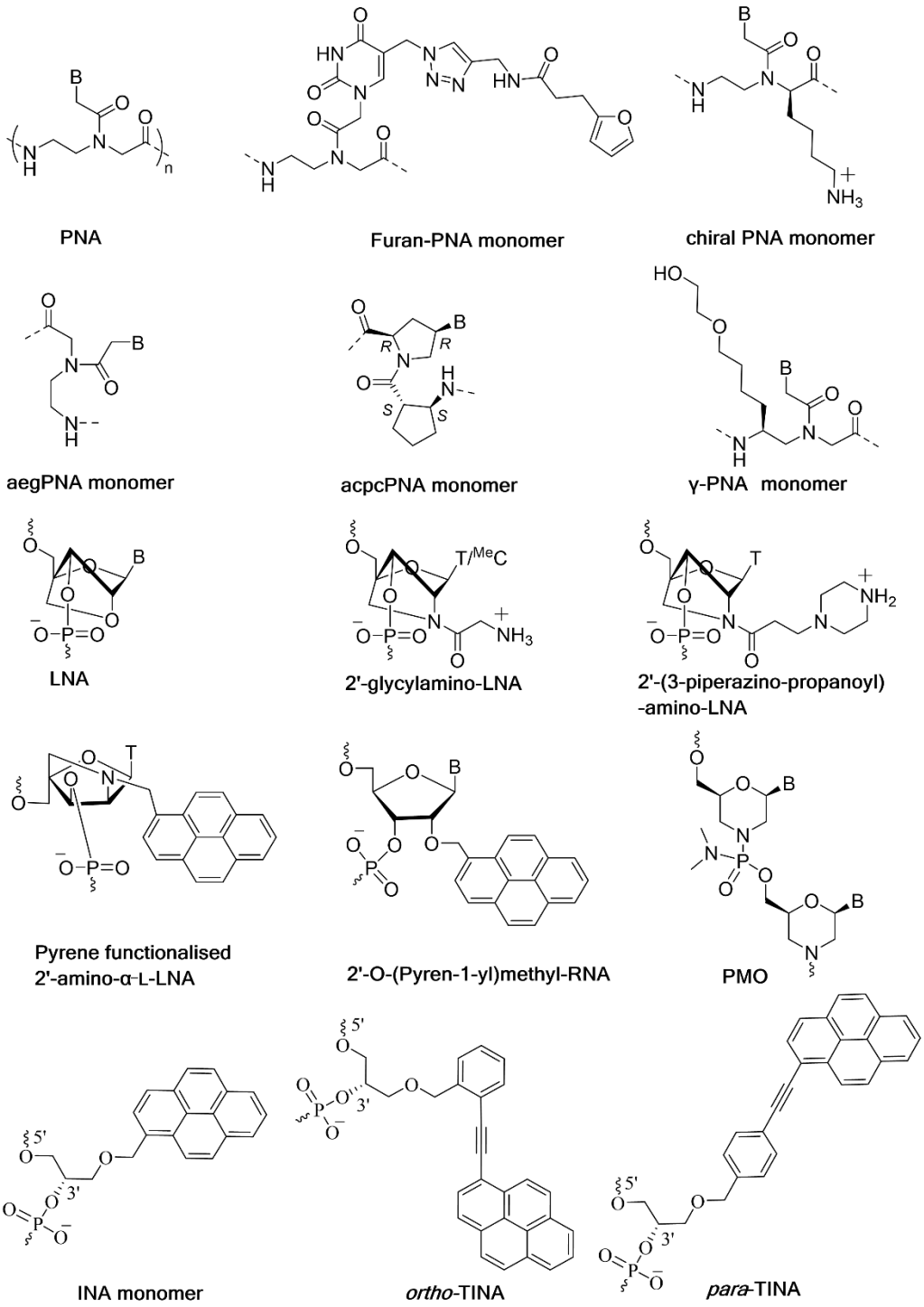


Figure 1.14. Chemically modified ON analogues and molecules for targeting dsDNA. B, heterocyclic base, A: adenin-9-yl, C: cytosin-1-yl, G: guanin-9-yl, T: thymin-1-yl, and U: uracil-1-yl.

Besides pyrene-functionalised LNA analogues, insertion of a pair of pyrenes opposite each other in the helix can also lead to DDI potential. Using nuclear magnetic resonance spectroscopy (NMR), one can see that two (*R*)-1-*O*-(pyren-1-ylmethyl)glycerol moieties

(intercalating nucleic acid, Fig. 1.14, INA monomer) are nicely accommodated in the helix with two pyrenes stacking on top of each other.[123] Such a duplex is less thermally stable than the unmodified dsDNA, and the corresponding INA/DNA duplexes in which the INA monomer is inserted as a bulge are stabilised.[124] This property of INA was used in the design of easily denaturing nucleic acids that are suitable for application in DDI.[125, 126] Another type of pyrene analogue, twisted intercalating nucleic acid (TINA) monomers (Fig. 1.14, *ortho*-TINA and *para*-TINA monomers), has been reported to change hybridisation properties of ONs, providing DNA probes with an improved binding affinity towards natural dsDNA. The insertion of *para*-TINA monomer as a bulge in the middle of ON leads to the enhanced thermal stability of Hoogsteen-type duplexes and triplexes. In contrast, Watson/Crick type DNA and RNA duplexes are destabilised.[127-131] On the other hand, the use of *ortho*-TINA monomer leads to the enhanced thermal stability of antiparallel DNA duplexes as well as Hoogsteen-type duplexes and triplexes, although the stabilising effect on triplexes is less pronounced as for *para*-TINA.[132]

This Thesis will initially focus on evaluating the ability of *ortho*-TINA modified duplexes as invading probes for targeting native dsDNA in a sequence-specific manner. This approach is particularly attractive as *ortho*-TINA modified ONs are already commercially available, and no additional chemical modifications are required for obtaining effective invasion probes.

1.1.5. Molecules targeting G4 DNA/RNA

Ligands or molecules targeting G-quadruplex are usually composed of a core of planar aromatic rings capable of interacting with G-tetrads through π-π stacking, which may increase stabilisation or induce the change of topology of G4 structures.[133] Unlike molecules targeting duplexes with modifications on the sugar, nucleobase or a phosphate, ligands for targeting G4 structures are usually incorporated at one or both terminus of the sequence. It has been reported that large aromatic units such as perylene PDI and porphyrin TCPP (Fig. 1.15) can convert a bimolecular, antiparallel G4 formed by d($G_4T_4G_4$), into a parallel, presumably tetramolecular complex.[134] It has also been shown that a human telomeric sequence d($G_3T_2AG_3$) conjugated with dibromoperylene derivative BrPDI (Fig. 1.15) at the 5′-end can induce the formation of a parallel dimeric G4 with both 5′-ends located at the same side of the structure with increased thermal stability. Moreover, the kinetic products of annealing (antiparallel species) were reported

to be converted into thermodynamically stable parallel G4s after incubation for two days at RT.[135]

Another type of well-studied ligand are porphyrin-based ligands. The porphyrin core is close to that of the G-tetrads in G4s, allowing for π-π interactions with G-tetrads. Porphyrin-based ligands can bind to G4s in three modes, either binding to the top and bottom of G4s (capping), binding between strands of G4s or intercalation between G-tetrads within a G4. All of the three binding modes are capable of stabilising the G4 structure. Among which the capping of porphyrin ligands to the end of G4s is of particular interest as it can stabilise G4 structures at the ends of chromosomal DNA.[136]

TMPyP4 (Fig. 1.15) is a widely used cationic porphyrin ligand that can stabilise G4 structures. The four positively charged side arm substituents can bind to negatively charged phosphate backbones of loops of G4s through electrostatic attraction, further stabilising the complex.[137]

A pentacationic manganese (III) porphyrin (Fig. 1.15) inhibits the activity of telomerase, an enzyme responsible for the elongation of the telomeres through the proposed stabilisation of G4s. The growth of cancer cells can be restricted by inhibiting the elongation of telomeres.[138]

Another well-designed and characterised ligand, cyclic helicene M1 (Fig. 1.15), can selectively bind to the telomeric multimeric G4s rather than to the telomeric monomeric G4s because of the orientation of the left-handed helicene and the shape of the structure. Molecular modelling studies demonstrated that the cyclic helicene M1 bound to the cleft between two adjacent G4 units, thus inhibiting the activity of telomerase. Moreover, it was proposed that the dihedral angle of the helical frameworks was a prominent factor in the design of ligands recognising the telomeric multimeric G4s.[139]

Based on this assumption, a hybrid oxazole-triazole ligand DR4-47 (Fig. 1.15) was designed and synthesised. It was reported to bind to the multimeric telomeric G4s, forming a parallel folded structure.[140] MD simulations suggested that the specific and effective binding of DR4-47 to telomeric multimeric G4s results from two factors, the π-π stacking between DR4-47 that is sandwiched between adjacent G4 units and the terminal G-tetrads, and the extra groove interaction by electrostatic attraction.

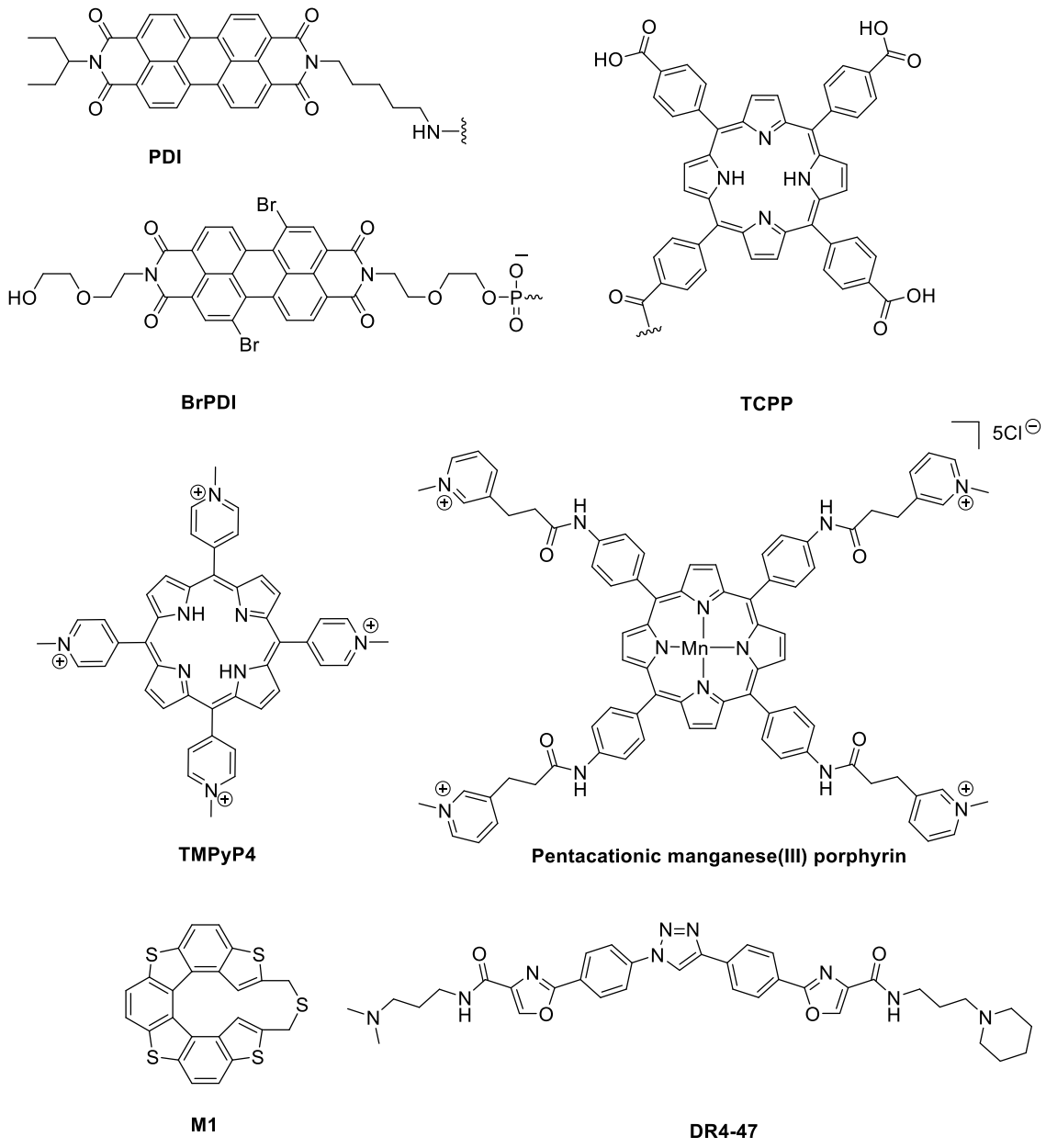


Figure 1.15. Structures of ligands that bind to G4s.

1.1.6. Modified ONs for invasion into G4 DNA/RNA

Compared with studies of ligands or small molecules that bind to G4s, research about modified ONs binding or invading into native G4 DNA/RNA is less reported. PNA and PNA analogues have been used for targeting G4s through various pathways.[141] G-rich PNA can target the G-rich DNA duplex in the promoter region, such as the *c-myc* oncogenic promoter sequence, via forming a hybrid PNA/DNA G4 as well as forming PNA/DNA duplex with the complementary DNA strand through Watson/Crick base pairing at the same time. Through this G4 invasion process, PNA incorporated with a L-

phosphoserine group attached at the N-termini can induce a sequence-specific DNA break in the target sequence of human genome.[142]

It has also been reported that G-rich PNA with acridone conjugate can target the human telomeric DNA sequence, forming a hybrid PNA/DNA G4 that mimics the biologically relevant (3+1) DNA dimeric telomeric quadruplex. The PNA/DNA bimolecular G4 showed increased stability at the low salt concentration compared to the corresponding dimeric G4 DNA. However, this PNA-acridone conjugate cannot invade a dimeric G4 by displacing of one of the DNA strands.[143] γ-PNA, alongside its DDI potential, was also reported as a probe for invading stable RNA G4s and inhibiting translation in cell lysate.[144]

Pyrene moieties have been applied in G4 structures. As shown in Fig. 1.16, a carefully designed light-switching pyrene probe can be used to detect G4s in living cells. The fluorescence of the pyrene probe switches from monomer to excimer emission when forming a G4 structure. This light-switching pyrene probe has been used to detect Telomeric repeat-containing RNA (TERRA), a non-coding RNA in mammalian cells. It has been found that human TERRA RNA forms a parallel G4 structure in living cells, providing the *in vivo* evidence for the presence of the G4 in human TERRA RNA. Moreover, the colocalised image showed that TERRA RNA G4s localised to telomere DNA at cell nuclei. Although the detailed structure and function of the TERRA RNA at human chromosome ends are not fully understood, this RNA G4 structure may be a key component of the telomere machinery.[145]

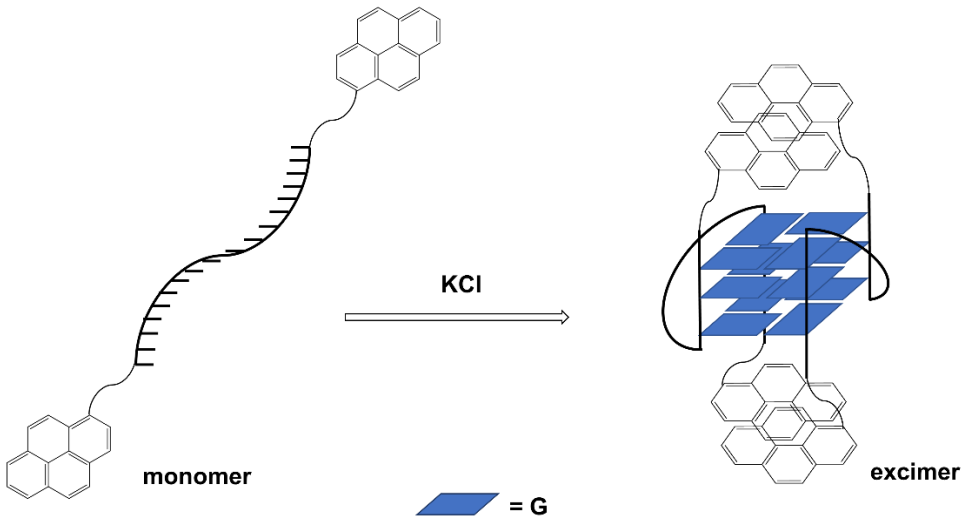


Figure 1.16. Use of the pyrene probe to detect TERRA RNA G4 structure. The pyrene molecule has monomer emission near 400 nm. G4 formation brings the pyrene molecules close to each other, forming pyrene excimers. Green light (~480 nm) is emitted after photoexcitation.

Pyrene moieties have also been attached to the nucleobase of the nucleotide next to the G-tetrad (Fig. 1.17, Pyrene-conjugated G-tract). This pyrene conjugated ON has been used to recover the oxidised G4 by replacing the oxidised guanine tract, forming a stable intermolecular G4 with the other intact guanine tracts. This recombination process induced by G4 invasion is a promising approach for therapeutic applications targeting genetic diseases.[146]

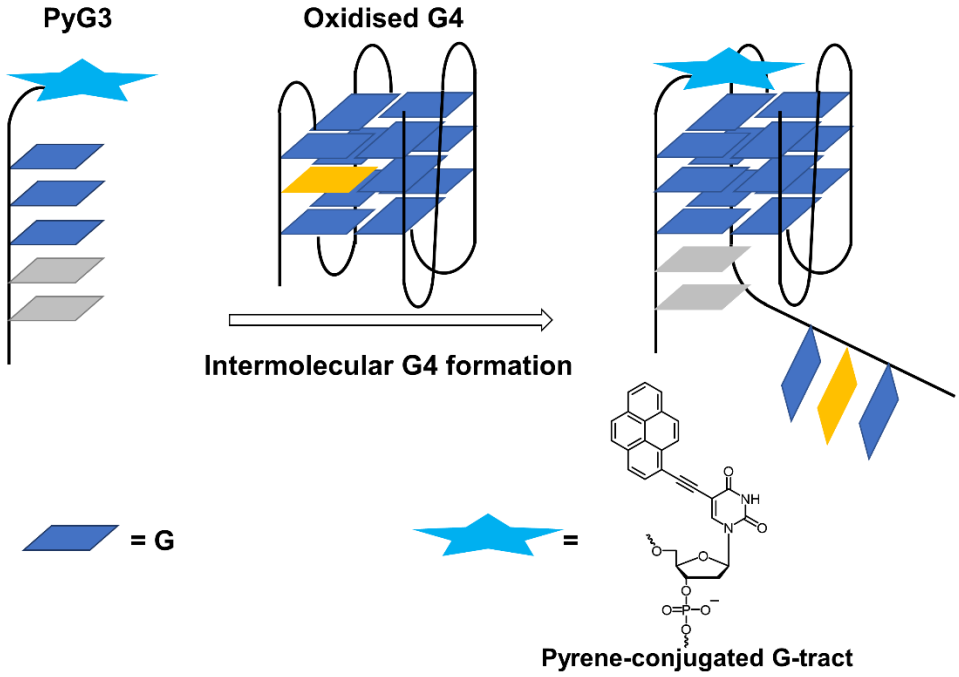


Figure 1.17. The schematic of intermolecular G4 formation with PyG3. The pyrene moiety of PyG3 is shown as a cyan star, and the oxidised guanine is highlighted in yellow.

1.2. Research aim

Research in this Thesis is aimed to develop novel probes for targeting native DNA duplexes or G4s using chemically modified ONs. As discussed above, DDI is more attractive than forming a DNA triplex or stand invasion because of higher efficiency and specificity. We will focus on the designing of probes suitable for targeting native DNA duplexes via DDI.

INA monomers were successfully applied in the design of easily denaturing nucleic acids suitable for DDI.[125, 126] It has also been reported that DNA duplex possessing two *ortho*-TINA monomers opposite each other was less thermally stable than the native duplex.[132] Considering the fact that *ortho*-TINA monomer is a mimic of INA monomer (Fig. 1.14), we believe that DNA sequences with *ortho*-TINA monomer can be used for DDI: the two

complementary *ortho*-TINA modified strands should have a lower binding affinity with each other but higher binding affinity towards the native duplex.

As G-quadruplex structures present at elevated levels in cancer cells than in normal cells,[147] we will also focus on designing chemically modified ONs that can invade into native G4s, and in this way, may be used to stop proliferation of cancer cells.

1.3. Thesis objectives

This Thesis discusses the development of chemically modified ONs for targeting native DNA duplexes or G4s.

In Chapter 2, we designed *ortho*-TINA monomer-based probes for sequence-specific targeting of native DNA duplex through DDI. Various research methods, including UV/Vis denaturation experiments, fluorescence spectrometry, and 20% native gel electrophoresis, were applied to evaluate properties of *ortho*-TINA modified duplexes, and the viability of DDI was proved. One problem observed during the investigation of *ortho*-TINA probes in DDI was the difficulty of invading into long duplexes. Several attempts were made, but the kinetics of invasion was not improved.

Chapter 3 describes a positively charged group (N+ modification) that was introduced into the phosphate of the DNA backbone through Staudinger reaction, forming a zwitterionic phosphate moiety. The N+ modification was incorporated into *ortho*-TINA probes, resulting in further increased thermal stability of corresponding invasion products. However, the improvement of the kinetics of DDI using this zwitterionic N+ modification was marginal.

Chapter 4 focuses on evaluating ONs possessing different numbers of N+ modifications in the formation of parallel DNA triplexes and antiparallel DNA and RNA duplexes. Resistance of modified ONs towards enzymatic digestion as well as their cell uptake was investigated.

In Chapter 5, we designed and synthesised chemically modified G-rich ONs bearing two types of charge-neutral modifications and investigated the properties of tetramolecular G4s formed by charge-neutral G-rich ONs. Various assessments, including NMR, electrospray ionisation mass spectrometry (ESI-MS), size-exclusion (SE-) high-performance liquid chromatography (HPLC) and native gel electrophoresis were conducted for the confirmation of assembly of charge-neutral G4 structures. The changes in thermal stability, association rate as well as dependence on the ionic strength of charge-

neutral G4 species were evaluated and compared with both negatively charged modified and unmodified G4s.

Research design and results in this Thesis might be useful in the field of DNA nanotechnology, antigene/antisense therapy, etc. Direction of future work and possible applications of chemically modified ONs are discussed in Chapter 6, and experimental procedures are included in Chapter 7.

1.4. Overview of methods used for the synthesis and evaluation of DNA assemblies

1.4.1. Description of automated DNA synthesis

ONs are synthesised using a MerMaid-4 automated DNA synthesiser (BioAutomation Corp.). Automated DNA synthesis allows the rapid production of ONs up to 200 bases long. Synthesis of ONs starts from 3′- to 5′- end through the stepwise addition of nucleoside phosphoramidites (or H-phosphonates) to each other in a programmable manner. Stepwise assembly of DNA/RNA is performed on a controlled pore glass (CPG) that allows for removing excess reagents by filtration and eliminates the requirement of purification between base additions.

As shown in Fig. 1.18, the synthetic cycle begins with a 3'-hydroxyl nucleoside fixed onto the CPG support by a base sensitive linker. This nucleoside contains a 4,4′-dimethoxytrityl (DMT) protecting group at 5′-OH (step 1, deprotection), which can be cleaved with dichloroacetic acid in DCM, providing a free 5'-hydroxyl for the subsequent coupling reaction. An activated trivalent phosphoramidite derivative is then added, reacting with the free 5'-hydroxyl on the primary nucleoside to create a phosphite bond (step 2, activation and coupling). Activation occurs with the simultaneous addition of 5-ethylthio-1*H*-tetrazole (ETT) that protonates the 3′-*O*-phosphoramidite and promotes coupling with the 5′-OH of the nucleoside. Self-condensation of the activated nucleoside is prevented by the protection of the 5′-OH with DMT. Any chains that do not undergo coupling are capped with acetic anhydride, 2,6-lutidine, and 1-methylimidazole (step 3, capping). The phosphite linker is then oxidised to phosphotriester by iodine/pyridine (step 4, oxidation), and the cycle is continued until the desired full length ON is obtained (step 5). When the synthetic cycle is finished, the ON is cleaved from the CPG support and the cleavage of the β-cyanoethyl protecting groups is carried out with concentrated aqueous ammonia (~ 28%) at RT for 2 h followed by incubation at 55 °C for 12 h to cleave the protecting groups on the nucleobases (step 6). One should notice that if the temperature

is raised before β-cyanoethyl groups are removed, there is a risk of cleavage of the phosphodiester backbone.

The desired ON is further purified from any short oligonucleotide sequences using various techniques, such as reversed-phase (RP-) or ion-exchange (IE-) HPLC or denaturing gel electrophoresis.

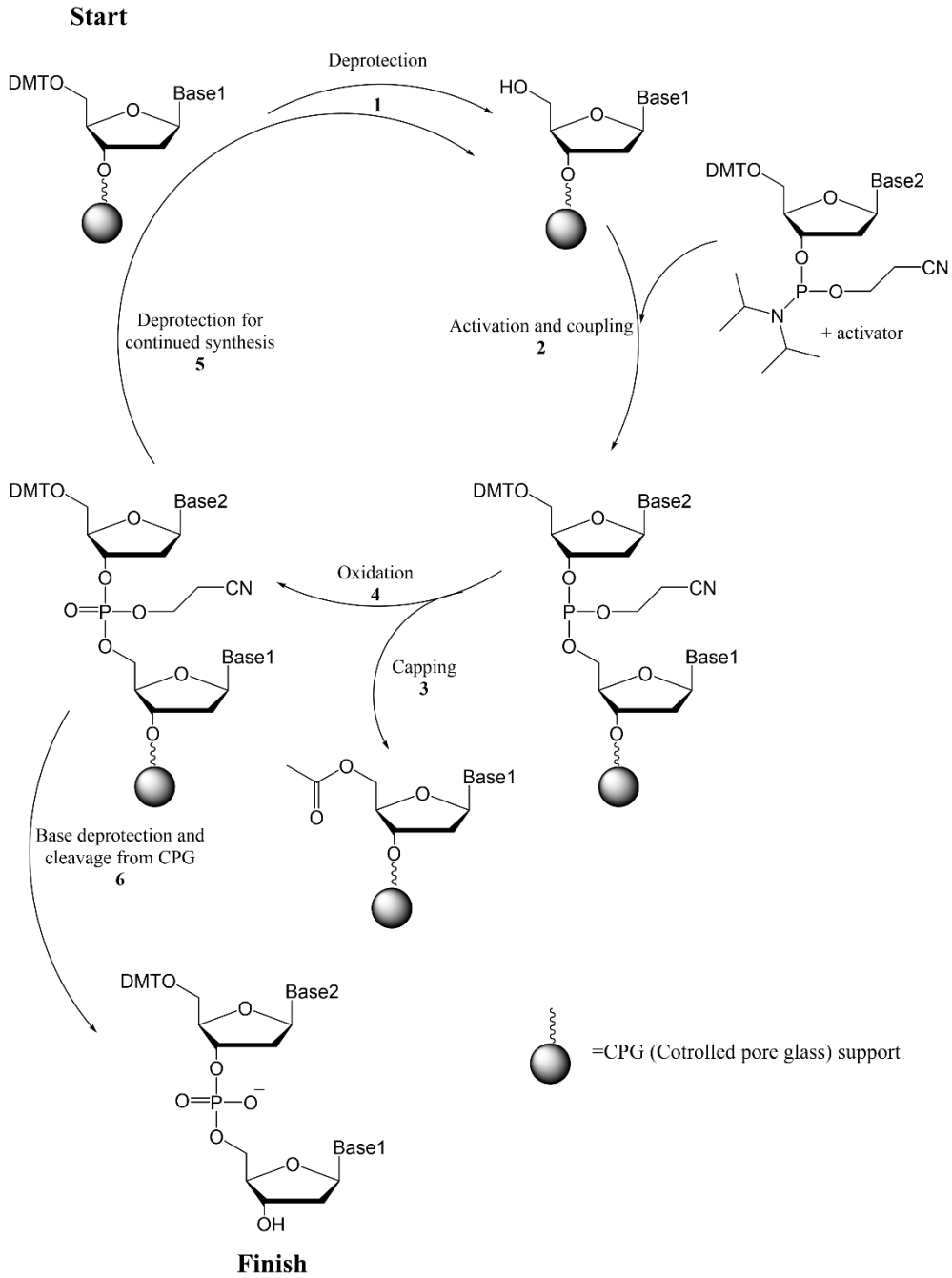


Figure 1.18. Synthetic steps in automated DNA synthesis using DNA phosphoramidites.

Table 1.2. Reagents used for automated ON synthesis (5 μmol synthesis scale)

		Reagents/amount/Supplier	Solvent/Volume (mL) /Supplier	Inject volume
Deprotection		Dichloroacetic acid/12 mL/Koch-Light Laboratories, Ltd.	DCM/388 mL/ Fisher Chemical	3 x 500 μL (240 s)
Activation		5-Ethylthio-1*H*-tetrazole (ETT)/23.4 g/Azide Chemical CO., Ltd. (China)	ACN/400 mL/Fisher Chemical	2 x 275 μL (240 s)
Capping	**A**	Acetic Anhydride/ 50 mL/ Asia Pacific Speciality Chemicals (APS, China) 2,6-Lutidine/50 mL/Sigma Aldrich	THF/400 mL/ ScharLan (China)	750 μL (80 s)
	B	1-Methylimidazole/70 mL/ACROS Organics	THF/380 mL/ ScharLan (China)	
Oxidation		Iodine/2 g/DIFCO Laboratories (U. K.) Pyridine/40 mL/CABLO ERBA reagents (S.A.S.) H_2O/8 mL	THF/352 mL/ ScharLan (China)	3 x 500 μL (240 s)

Abbreviations: DCM: dichloromethane; ACN: acetonitrile; THF: tetrahydrofuran.

1.4.2. Methods for studying DNA secondary structures and invasion process

For the investigation and characterisation of ONs, we have used several techniques, including UV-Vis spectroscopy, circular dichroism (CD) spectroscopy, polyacrylamide gel electrophoresis (PAGE), NMR, and ESI-MS. Various chromatographic techniques were also employed depending on the chemical modification incorporated in ONs and the properties evaluated.

1.4.2.1.UV-Vis Spectroscopy

Nucleotides exhibit strong UV absorption with λ_{max} values close to 260 nm.[148] Since neither the pentose nor the phosphate in the nucleotides shows any significant UV absorption above 230 nm,[148] the UV absorption of nucleotides is mainly contributed by the UV absorption profiles of nucleobases. When the nucleotides are in close proximity to the neighbouring bases, for example, in the duplex structure, there is a remarkable decrease in the intensity of UV absorption (by up to 40%) due to the π-π electron interactions between stacked bases that affect the transition dipoles of the bases (Fig. 1.19).[148] This phenomenon is known as hypochromicity, which is reversible when the duplex is dissociated. There are two major applications of this phenomenon, the determination of temperature-dependent changes between structured and ssDNA (using UV-Vis spectroscopy) as well as the monitoring of changes in the asymmetric environment of the bases (using CD spectroscopy).

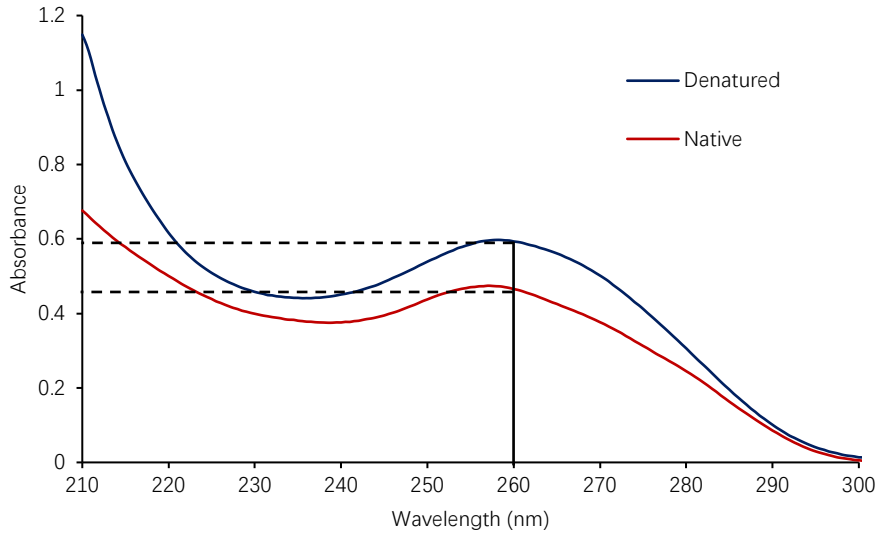


Figure 1.19. UV-Vis spectra showing the difference between dsDNA (20 °C) and denatured ssDNA (90 °C).

UV-Vis spectroscopy is mainly used to determine the concentration of DNA in solution and to evaluate the thermal stability of DNA assemblies. The concentration of a DNA sample is determined using Beer's law based on the extinction coefficient (ε) of an ON and the UV absorbance at a certain wavelength (normally at 260 nm).[149] The thermal stability of ON assemblies is determined by monitoring changes in absorption at a certain wavelength upon a temperature ramp. The cooperative assembly of DNA generally leads to a change in absorption that follows a sigmoidal curve (Fig. 1.20, melting curve). The point of inflection (the maximum in the first derivative) of the melting curve represents the temperature at which 50 % of the duplex is denatured. This temperature is defined as T_m value, which presents the thermal stability of a particular assembly. Dissociation and association must be reversible for the T_m to be considered an accurate representation of thermodynamic stability. Hysteresis between melting and annealing profile is undesirable as the apparent T_m becomes a function of the rate of heating/cooling. The hysteresis can be reduced by lowering the rate of temperature ramp for DNA complexes. However, it is often unavoidable when studying bimolecular and tetramolecular G4s because these complexes form so slowly that a sufficiently low ramp is unfeasible. In this situation, the thermal stability of G4 is presented as a $T_{1/2}$ value with the heating rate specified (Fig. 1.20).[150]

There are mainly three ways to determine the T_m value of DNA secondary structures. As shown in Fig. 1.21, the most common one is to use the maximum of the first derivative of the melting curve (T_m1). Alternatively, one can draw baselines at the top and the bottom of the melting curve, and a median line is then placed between these lines. The temperature at which the median line intersects the melting curve is the T_m value (T_m2). In order to analyse thermodynamic parameters, the melting curve can be converted into a fraction folded (Θ) *versus* temperature representation followed by adequate upper and lower baselines chosen:

$$\Theta_T = (L0_T - C_T) / (L0_T - L1_T) \qquad \text{(Equation 1)}$$

$L0_T$ and $L1_T$ correspond to the baseline values of the unfolded and folded species, respectively. Θ should be a number between 0 and 1: Θ = 0 for $T \gg T_m$, Θ = 1 for $T \ll T_m$. T_m corresponded to the temperature where Θ = 0.5 (T_m3).[151] The procedures of calculating thermodynamic parameters will be discussed in Chapter 2. There is only a marginal difference between the T_m values obtained using these three methods.[151]

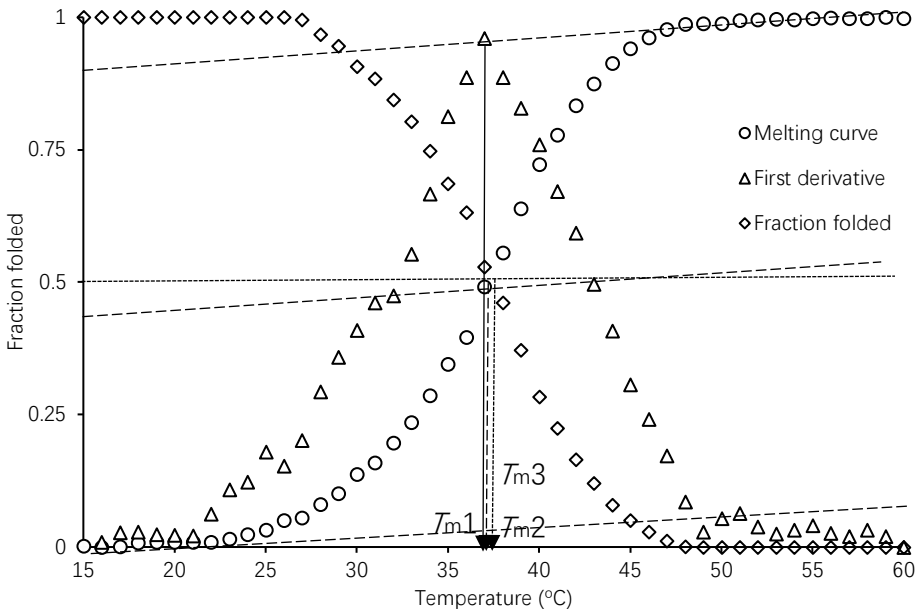


Figure 1.20. Determination of the T_m value of a DNA duplex. Using either the maximum of the first derivative (triangles, T_m1), the median method (circles, T_m2) or the fraction folded of the melting curve (T_m3) give approximately equal T_m value.

1.4.2.2. Circular Dichroism (CD) Spectroscopy

Circular dichroism (CD) spectroscopy is performed based on the property that chiral entities interact with left-handed circularly polarised light (LHCP) and right-handed circularly polarised light (RHCP) to a different extent.[152] CD data is presented as the difference in absorption between LHCP and RHCP. The observed difference in absorption is reported in degrees of ellipticity or as molar ellipticity, and the latter has been corrected for concentration at 1 cm pathlength.

CD is useful for investigating the topology of protein and nucleic acid secondary structures. Isolated purine and pyrimidine bases are planar and non-chiral, therefore exhibit no CD signal. When incorporated into nucleotides, the glycosylic bond from the C-1' atom of the sugar to either the N-9 of purines or N-1 of pyrimidines creates a chiral centre and hence can absorb LHCP and RHCP to a different extent. Depending on the relative orientation of the transition dipoles, the difference in absorption of LHCP and RHCP gives rise to either positive or negative peaks in the CD spectrum, which are called Cotton effects (Fig. 1.21A). The observed signal for complex structures like DNA and protein is the sum of the Cotton effects, which gives rise to bisignate curves as shown in Fig. 1.21B.

The CD signal of nucleic acids increases with the length of a sequence due to the co-operativity from the nearest neighbour interactions between contiguous bases. This effect is a result of both sequence effects arising from nearest-neighbour interactions and the

overall secondary structure. CD spectra for a B-DNA duplex, ssDNA and G4s show that each secondary structure exhibits its own signature spectrum (Table 1.3). For example, parallel G4s can be characterised by a positive maximum at 260 nm and a negative minimum at 240 nm, whereas antiparallel G4s are identified by a positive maximum at 295 nm and a negative minimum at 265 nm.[153]

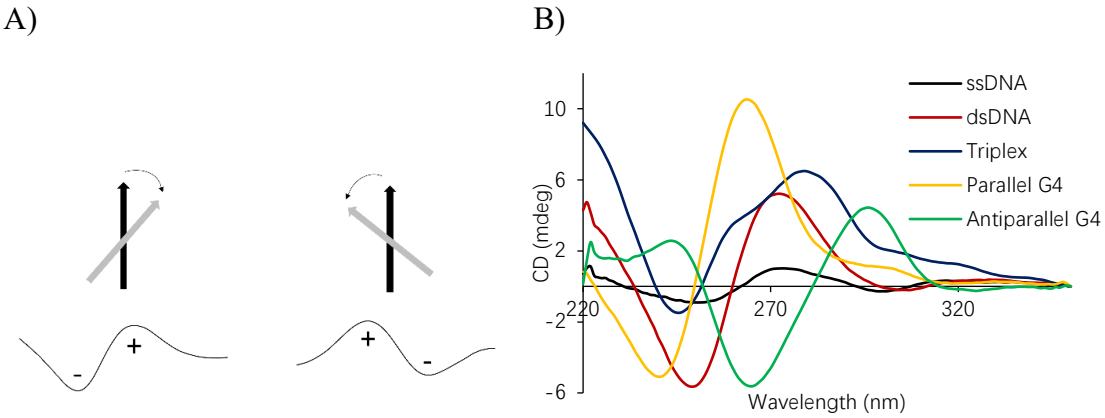


Figure 1.21. Bisignate curves and CD spectra for DNA secondary structures. A) left: A positive bisignate curve resulting from exciton coupling of dipoles that are in clockwise orientation; right: A negative bisignate curve resulting from exciton coupling of dipoles that are in anticlockwise orientation. B) CD spectra of ssDNA, duplex B-DNA triplex, parallel G4 and antiparallel G4. DNA triplexes can be characterised using CD, but their characteristic peaks vary significantly depending on the sequence.

Table 1.3. Characteristic CD peaks of various DNA secondary structures.

Topology	Characteristic peaks (nm)	Glycosidic Bond Conformation
B-DNA	+260, -280, -245[a]	Anti
Z-DNA	+260, -290, -205	G: Syn, C: Anti
i-Motif	+290, -260	Anti
Parallel G4	+260, -240	Anti
Antiparallel G4	+290, -260	Syn

[a]The + and – values indicate the phase of the peak. Duplex DNAs (B-DNA and Z-DNA) have multiple characteristic CD peaks based on sequence composition.

1.4.2.3. Gel Electrophoresis

Gel electrophoresis is a useful tool for the purification of synthesised ONs and the characterisation of nucleic acid secondary structures. In gel electrophoresis, DNA migrates through a cross-linked polymer gel towards the positive electrode in the presence of a current, and are separated according to their size, charge, and shape of the DNA assembly. Shorter ONs with a higher charge/mass ratio migrate faster through the gel. PAGE is commonly used for short ONs up to ~200 nucleotides in length,[154] and depending on the information that is desired, PAGE can be performed in either denaturing or non-denaturing conditions.

Denaturing PAGE is applied mainly for assessing the purity of ONs. Samples are incubated in the presence of a chaotropic agent like urea (7 or 8 M) at high temperature to disrupt high-order structures. Same chaotropic agents are also incorporated into the gel to prevent the assembly of high-order structures during the experiment. Samples migrate as unstructured, single-stranded ONs and a pure product appears as a single band. Denaturing PAGE can also be used for purifying ONs, and in this case, a thicker gel is usually applied because a larger quantity of samples is loaded. However, in cases such as G-quadruplexes, the chaotropic agent employed in denaturing PAGE is not enough to disrupt the assembly of high-order structures.

Non-denaturing (native) PAGE is used to evaluate properties of assembled structures such as duplexes, triplexes or G-quadruplexes. Native PAGE uses mild conditions to ensure that samples are in equilibrium and migrating on the gel as fully assembled structures. The same concentration of cations (Na^+ or K^+) as in the samples is included in the gel and the buffer. However, increased ionic strength will lead to further heating of the gel and subsequently destabilise DNA assemblies. Lower voltages in comparison with denaturing gel are used to reduce the temperature of the gel and denaturation of the samples during the experiment. The migration rate of a sample can provide information on the size and molecularity. For example, two duplexes will show different mobility on native PAGE because of different molecular weights.

The results of gel electrophoresis are usually visualised by immersing the gel in a solution of dye, typically Stains-all (Sigma-Aldrich). Gels are exposed to Stains-all for approximately five min, followed by destaining for approximately thirty min in water to remove the dye bound to the polyacrylamide gel. DNA samples are stained purple and can be visualised using a standard camera.

1.4.2.4. NMR Spectroscopy

NMR is a useful technique in studying nucleic acids, especially for short nucleic acids up to 28 - 30 nucleotides.[155] NMR measurements of ONs allow for quick recognition of bases forming the complex. However, full characterisation of a structure requires the assignment of the whole sugar-phosphate backbone. A typical 1- (frequency) dimensional (1D) ^{1}H spectrum for a DNA/RNA hybrid shows signals for over 300 different protons, which is impractical to be readily distinguished and assigned a chemical shift. A variety of experiments have been developed by introducing a second frequency dimension to alleviate the overcrowding problem. For example, in 2D (^{1}H- ^{1}H) total correlation spectroscopy (TOCSY) spectrum, cross peaks for coupled protons can be observed. Moreover, cross peaks are observed not only for nuclei which are directly coupled, but also between nuclei that are connected by a chain of couplings. In 2D (^{1}H- ^{1}H) Nuclear Overhauser effect spectroscopy (NOESY) spectrum, the obtained cross peaks connect resonances from nuclei that are spatially close to each other. The 2D (^{1}H- ^{13}C) Heteronuclear multiple quantum coherence (HMQC) detects correlations between two different types of nuclei (^{1}H and ^{13}C), which are separated by one bond. Heteronuclear multiple bond coherence (HMBC) is similar to HMQC, but it detects heteronuclear correlations over longer ranges of about 2–4 bonds. By combining these 2D experiments, different nucleotide residues may be identified. However, complete resolution of an assembled DNA secondary structure requires a huge amount of work due to the flexibility of sequences as well as the modifications incorporated.

NMR is particularly useful when studying G4 structures. For a reference sequence TGGGGT (TG_4T), there are sixteen separate guanine nucleobases in the tetramolecular G4 structure. It will be very difficult to solve the entire structure as each of the four guanines in a particular strand experiences a slightly different environment and will provide a slightly different set of signals in NMR. In this situation, we will focus mainly on the Guanine imino (H1) protons region between 9 – 12 ppm that shows protons involved in the H-bonding of the G-tetrad. For G4s NMR, assembled samples are dissolved in an aqueous buffer supplemented with cations (Na^+ or K^+) and 10 % D_2O. Guanine imino protons in unstructured ssDNA disappear in the NMR spectrum due to the exchange with deuterium. However, since TG_4T in Na^+ or K^+ solutions forms a parallel, symmetrical, tetramolecular structure with all dGs in an *anti*-conformation, four sharp peaks between 9 – 12 ppm are observed in the NMR spectrum, indicating the formation

of a four-layered, symmetrical G4 (Fig. 1.22). These imino signals persist at relatively high temperatures, indicating the formation of a stable G4, especially in K^+ solution.

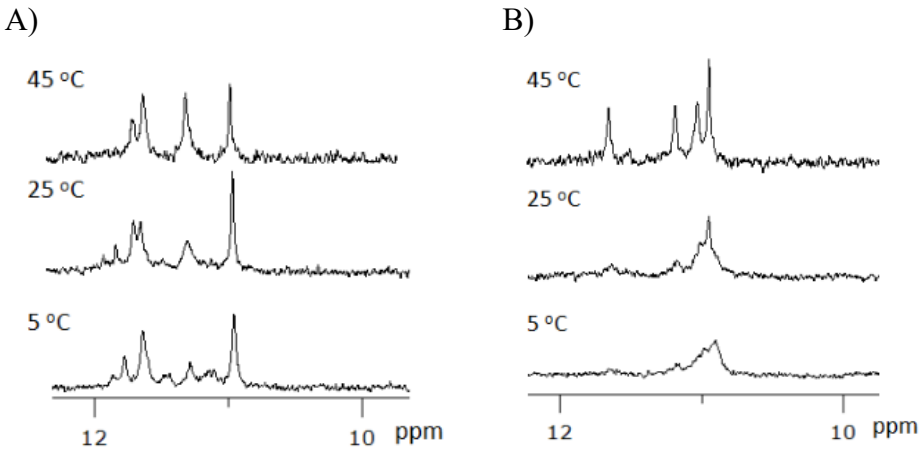


Figure 1.22. ^{1}H NMR spectra in the imino region of $[dTG_4T]_4$ at indicated temperatures. 10 mM Li-cacodylate buffer containing 100 mM: A) NaCl and B) KCl.

One noticeable drawback of NMR is the requirement of relatively high ON concentrations (minimum 200 μM) as well as sample volumes (~250 μL).[156]

1.4.2.5. Mass Spectrometry

Mass spectrometry is considered one of the most reliable methods to detect nucleic acids. It can provide accurate molecular masses of single-stranded nucleic acid as well as its assemblies. Matrix-assisted laser desorption ionisation (MALDI) and electrospray ionisation mass spectrometry (ESI-MS) are the two most useful methods in nucleic acid research.

In MALDI analysis, the nucleic acid sample is embedded in a crystalline matrix of a small molecule (e.g. 3-hydroxypicolinic acid). This target is excited by an ultraviolet laser beam in a high vacuum, resulting in intact molecules of the ON sample. The molecules are then desorbed into the gas phase and ionised by the UV radiation to give singly charged ions. The resulting molecular ions are analysed by mass using Time-of-Flight (TOF) mass spectrometry, and therefore can avoid any need for prior separation of a mixture.

MALDI is suitable for measuring the mass of individual ONs rather than observing nucleic acid assemblies due to the use of a laser, limiting MALDI's utility to easily ionise short ONs. Also, the need for dry samples co-crystallised with a matrix reduces a great deal of control needed to ensure the proper assembly of DNA complexes. MALDI is mostly suitable for the confirmation of products after DNA synthesis.

In ESI-MS, samples are ionised directly from the micro-droplets of a solution. Samples are aspirated and passed through a fine, charged nozzle producing charged droplets, which undergo a number of desolvating processes that ultimately produce individual ions. As a result, high mass nucleic acid molecules can be observed as multiply charged ions at relatively low mass to charge (m/z) ratios.

In comparison with MALDI, ESI-MS is slower and is more sensitive to salt in the sample. It is more suitable for studying species with mass up to several thousand Daltons. Moreover, ESI-MS is especially useful when studying the assembled structures such as G4s since the binding interactions are maintained during the transit from solution to micro-droplets in the gas phase.[157],[158]

Chapter 2. *ortho*-Twisted Intercalating Nucleic Acid (*ortho*-TINA) in specific recognition of mixed-sequence double-stranded DNA via dual-duplex invasion

2.1. Introduction

Using chemically modified ONs to sequence-specifically target native dsDNA is of great interest in the field of therapeutics. In comparison with triplex formation and triplex invasion that rely on TFOs to form Hoogsteen base pairing, utilising modified ONs to invade into dsDNA through Watson/Crick base pairing is more attractive. This approach was called dual duplex invasion (DDI, Fig. 2.1B) or double duplex invasion.[67] An ideal invading probe for DDI requires the two chemically modified invading strands having a better binding affinity toward native strands than with each other.

A large number of molecules have been designed, synthesised and evaluated for the targeting of dsDNA trough DDI. One major issue in designing the invading probe is that an efficient binding affinity of modified ONs toward ssDNA usually results in even higher binding affinity towards the complementary invading strands, as seen for PNAs and LNAs.[126] INA monomers (Fig. 2.1A) were used in the design of invasion probes for targeting dsDNA through DDI.[125, 126] Same principle was also applied in several pyren-1-yl modified nucleotides in which the zipping arrangement of these modifications forced pyrenes in the same region of the helix leading to the destabilised duplexes.[112, 113, 115, 159-165] However, a multi-step synthesis of individual modified nucleotides is usually required. It has been reported that the DNA duplex possessing two *ortho*-TINA monomers (Fig. 2.1A) opposite each other destabilised native duplex.[132] We proposed that dsDNA probes with *ortho*-TINA monomers could serve as promising tools for invading into native DNA duplexes because the *ortho*-TINA monomer mimics the INA structure. The challenge of targeting and sequence-specific recognition of long relaxed dsDNA under physiological conditions using short invaders has not been met.

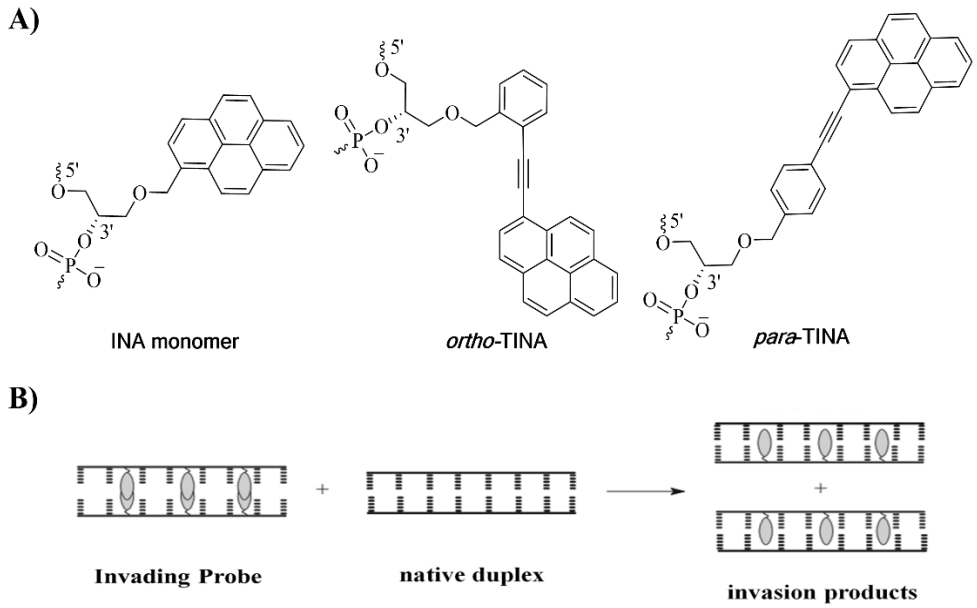


Figure 2.1. Pyrene analogues A) and illustration of DDI B).

2.2. Hypothesis

We hypothesised that DNA sequences incorporated with *ortho*-TINA monomer could be used for DDI: the two complementary *ortho*-TINA modified strands would have lower binding affinity with each other but higher binding affinity towards the native strands. And this process should be sequence-specific. We proposed that the speed of DDI should be in line with the number of *ortho*-TINA pairs present in the probe: the more *ortho*-TINA pairs present, the faster the DDI.

2.3. Methodology and methods used for evaluation of DDI

Based on the hypothesis mentioned above, we designed two sets of 16-mer DNA duplexes that differed in G/C content for evaluating the potential of *ortho*-TINA containing ONs in DDI: duplexes in Set 1 contained a single G/C bp, whereas in Set 2, each duplex had six (Table 2.1). Higher G/C content in Set 2 duplexes results in higher thermal stability of duplexes. For DDI to occur, the affinity of invading strands must be low to each other, but the affinity to unmodified strands has to be high, exceeding the stability of the native duplex. In other words, the invasion products are thermodynamically preferable. As shown in Figure 2.2, to validate the feasibility of DDI, we initially performed thermal stability analysis of native duplexes, probes and invasion products and obtained thermodynamic parameters. Then we applied fluorescence spectroscopy and 20% native PAGE to monitor and visualise the process of DDI both indirectly and directly.

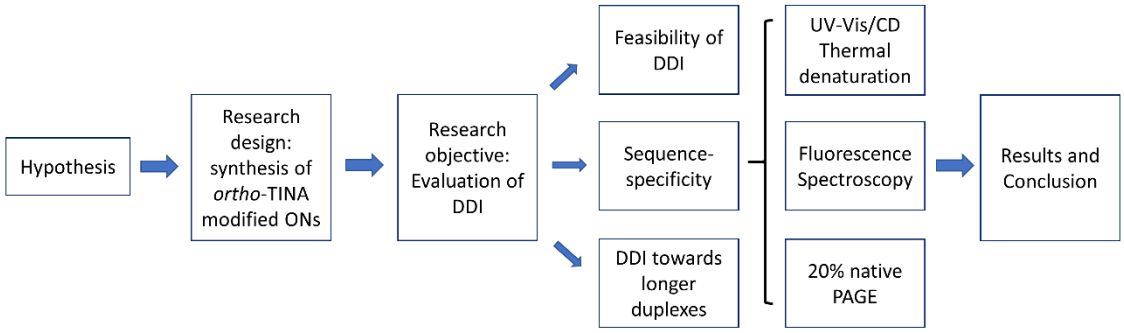


Figure 2.2. Flowchart of methodology and methods used for evaluation of DDI.

Thermal denaturation and renaturation experiments of duplexes studied were performed using UV-Vis and CD spectroscopy. Based on the melting profiles of duplex structures, we obtained thermodynamic parameters (ΔH, $T\Delta S$ and ΔG) for each duplex. Gibbs free energy was then used to predict the occurrence of DDI after the corresponding probes and a native duplex are mixed. Here, we used Gibbs free energy of recognition at 298 K, ΔG^{298}_{rec}, defined as:

$$\Delta G^{298}_{rec} = \Delta G^{298}\,(\text{invasion products}) - \Delta G^{298}\,(\text{starting duplexes}) \qquad \text{(Equation 1)}$$

The obtained ΔG^{298}_{rec} values for the invasion process should be preferably negative for the spontaneous occurrence of DDI under the experimental conditions stated.

We then verified the results using two techniques, fluorescence spectroscopy and 20% native PAGE. Fluorescence spectroscopy provided indirect evidence of DDI as a result for DDI is reflected by a change in fluorescence signals of TINA monomers. On the other hand, 20% native PAGE provided direct evidence of DDI because of the direct visualisation of invasion products. Kinetics of DDI using probes possessing different numbers of *ortho*-TINA monomers was monitored using fluorescence spectroscopy. The influence of a number of G/C pairs present in the duplex on DDI was also the subject of evaluation.

2.3.1. Research design: synthesis of *ortho*-TINA modified ONs

Unmodified as well as *ortho*-TINA-modified sequences were purchased from TriLink Biotechnologies (USA). Native duplexes **C1** (ON1/ON2, 5′-TTAATAATATGATTTA, 5′-TAAATCATATTATTAA) and **C2** (ON9/ON10, 5′-TAGTTGGTAGTGCTTT, 5′-AAAGCACTACCAACTA) were selected as controls for Set 1 and Set 2, respectively (Table 2.1). In each set, invading probes were designed possessing two, three and four *ortho*-TINA pairs (i.e., duplexes **P1** – **P3** for Set 1 and **P4** – **P6** for Set 2). In *ortho*-TINA

containing duplexes, monomers were separated by at least three unmodified nucleotides (Table 2.1).[127, 130]

One should mention that unless otherwise stated, the abbreviation for the sequences and duplexes only stands for sequences and duplexes studied in the current Chapter.

Table 2.1. Abbreviation and sequence for native duplexes (**C**), probes (**P**), and corresponding invasion products (**IP**) in Set 1 and Set 2.

	Probes	ON	Sequences	Invasion products	
Set 1	**C1**	ON1 ON2	5′-TTAATAATATGATTTA 3′-AATTATTATACTAAAT		
	P1	ON3 ON4	5′-TTA**X**ATAATATGAT**X**TTA 3′-AAT**X**TATTATACTA**X**AAT	**IP1**	5′-TTAATAATATGATTTA 3′-AAT**X**TATTATACTA**X**AAT
				IP2	5′-TTA**X**ATAATATGAT**X**TTA 3′-AATTATTATACTAAAT
	P2	ON5 ON6	5′-TTA**X**ATAAT**X**ATGAT**X**TTA 3′-AAT**X**TATTA**X**TACTA**X**AAT	**IP3**	5′-TTA**X**ATAATATGAT**X**TTA 3′-AATTATTATACTAAAT
				IP4	5′-TTAATAATATGATTTA 3′-AAT**X**TATTA**X**TACTA**X**AAT
	P3	ON7 ON8	5′-TTA**X**ATA**X**ATAT**X**GAT**X**TTA 3′-AAT**X**TAT**X**TATA**X**CTA**X**AAT	**IP5**	5′-TTAATAATATGATTTA 3′-AAT**X**TAT**X**TATA**X**CTA**X**AAT
				IP6	5′-TTA**X**ATA**X**ATAT**X**GAT**X**TTA 3′-AATTATTATACTAAAT
Set 2	**C2**	ON9 ON10	5′-TAGTTGGTAGTGCTTT 3′-ATCAACCATCACGAAA		
	P4	ON11 ON12	5′-TAG**X**TTGGTAGTG**X**CTTT 3′-ATC**X**AACCATCAC**X**GAAA	**IP7**	5′-TAGTTGGTAGTGCTTT 3′-ATC**X**AACCATCAC**X**GAAA
				IP8	5′-TAG**X**TTGGTAGTG**X**CTTT 3′-ATCAACCATCACGAAA
	P5	ON13 ON14	5′-TAG**X**TTGG**X**TAGTG**X**CTTT 3′-ATC**X**AACC**X**ATCAC**X**GAAA	**IP9**	5′-TAGTTGGTAGTGCTTT 3′-ATC**X**AACC**X**ATCAC**X**GAAA
				IP10	5′-TAG**X**TTGG**X**TAGTG**X**CTTT 3’-ATCAACCATCACGAAA
	P6	ON15 ON16	5′-TAG**X**TTG**X**GTA**X**GTGC**X**TTT 3′-ATC**X**AAC**X**CAT**X**CACG**X**AAA	**IP11**	5′-TAGTTGGTAGTGCTTT 3′-C**X**AAC**X**CAT**X**CACG**X**AAA
				IP12	5′-AG**X**TTG**X**GTA**X**GTGC**X**TTT 3′-ATCAACCATCACGAAA

X denotes *ortho*-TINA monomer; **C** stands for the target native duplex; **P** stands for a probe; and **IP** stands for an invasion product.

To ensure sequence specificity of our *ortho*-TINA probes, we tested invasion against totally (**MM1**, ON17/ON18, 5′-TAAATATTTTTCTTC and 3′-ATTTATAAAAAAGAAG) and partially mis-matched duplexes (**MM2**, ON19/ON20, 5′-TTTTAAATATTTTTCTTCAAT and 3′-AAAATTTATAAAAAGAAGTTA, underlined nucleotides are found in the native duplex). Moreover, since there are many more non-targeted sequences (un-complementary sequences) in the cell than the target duplex, it is necessary to investigate the sequence specificity of *ortho*-TINA modified probes mimicking physiological conditions. We conducted competitive experiments using salmon testes DNA (100 × concentration) as non-targeted sequences and monitored DDI using *ortho*-TINA modified probes.

Based on the previous study about easily denaturing nucleic acids formed by INA, we knew that there was a general problem that invader probes have difficulties invading into long duplexes.[126] We investigated whether our *ortho*-TINA probes can invade into long DNA duplexes, using two unmodified complementary DNA duplexes, **LC1**, (ON21/ON22, 5′-TTTTTAATAATATGATTTAATA and 3′-AAAAATTATTATACTAAATTAT; 3 bps longer at each end) and **LC2** (ON23/ON24, 5′-CTTTTTTTAATAATATGATTTAATAAAA and 3′-GAAAAAAATTATTATACTAAATTATTTT; 6 bps longer at each end).

2.3.2. Research methods used for evaluation of DDI

2.3.2.1. UV-Vis/CD denaturation and renaturation experiments

Thermal stability (T_m) of DNA duplexes was initially evaluated using UV-Vis denaturation and renaturation experiment. The T_m value was determined as the maximum of the first derivative of the melting curve.[151] However, occasionally some duplexes with chemical modifications (such as *ortho*-TINA monomers) did not provide a clear transition upon thermal denaturation due to the interactions between aromatic moieties or between modifications and nucleobases. Alternatively, we performed thermal denaturation experiments using CD upon heating from 5 to 90 °C with given equilibration time (e.g., every 5 °C for 5 min, Fig. 2.3A). CD signal change at a certain wavelength (e.g., 249 nm) was then extracted and plotted against temperature (Fig. 2.3B). As described in the first chapter, T_m corresponded to the crossing point between the obtained melting curve and the median (Fig. 2.3B).[151]

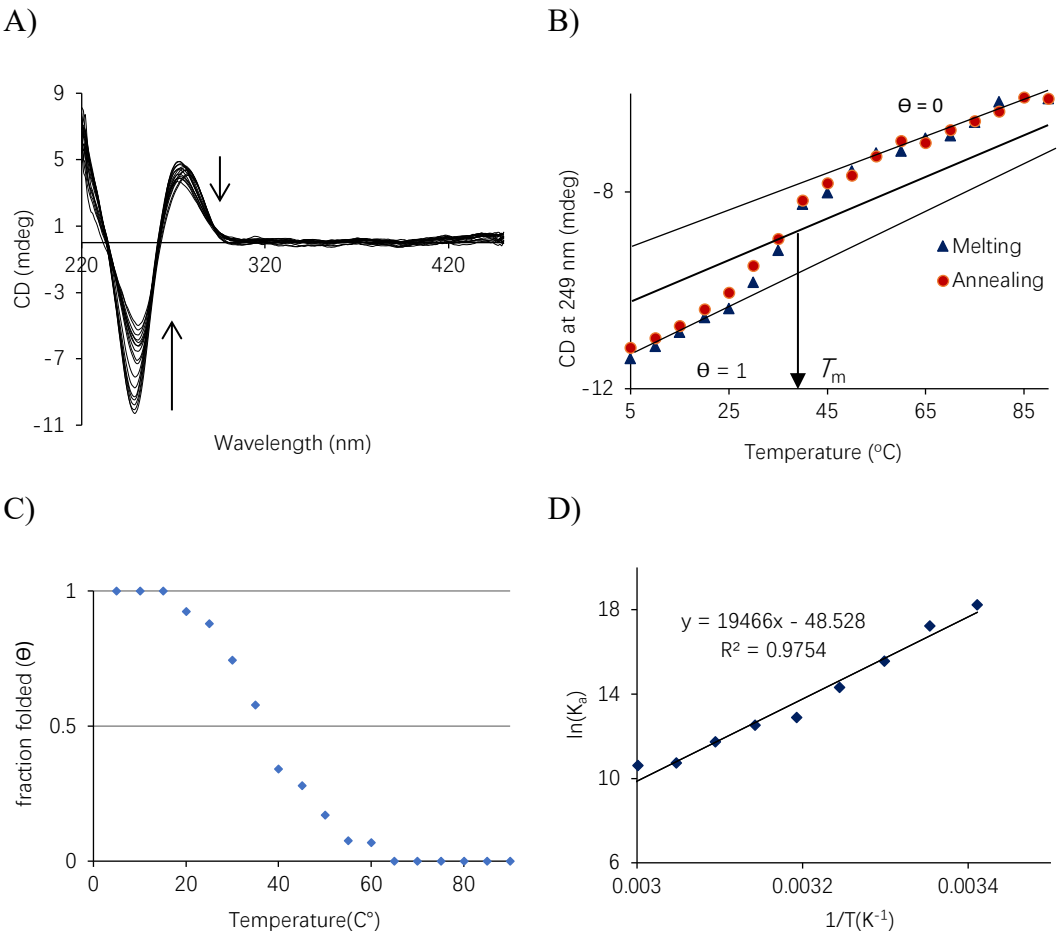


Figure 2.3. Determination of T_m and thermodynamic parameters of DNA duplex **C1**. A) CD melting spectra of a duplex formed by ON1 and ON2 (2 μM strand concentration), recorded every 5 °C from 220 nm to 450 nm. Signal changes at 249 nm and 273 nm were observed during melting; B) The melting and annealing profile of signal change at 249 nm, the lower baseline corresponds to the associated form (Θ = 1), the upper baseline corresponds to the dissociated form (Θ = 0); T_m was determined as 37 °C; Conditions: 10 mM Na-phosphate buffer (140 mM NaCl, 0.1 mM Na_2-EDTA, pH 7.0 at 25 °C); C) Fraction folded as a function of temperature. D) van't Hoff plot used for determination of ΔH^0 and ΔS^0. Θ values are extracted from the melting curve shown in Fig. 2.3B (melting of a bimolecular structure). Not all points are plotted in this figure, only Θ values significantly higher than 0 and lower than 1 are used.

The thermodynamic parameters of the duplex can be obtained by analysing the melting profiles ignoring the changes in DNA and salt concentrations induced by solution evaporation, the change of pH during heating, and assuming that there is no change in the heat capacity ($\Delta Cp = 0$).[151] Ellipticity for melting profiles *versus* temperature plot (Fig. 2.3B) was converted into a fraction folded (Θ) *versus* temperature representation (Fig. 2.3C).

By definition, the free Gibbs enthalpy may be written as:

$$\Delta G^0 = -RT \ln(K_a) = \Delta H^0 - T \times \Delta S^0 \quad \text{(Equation 2)}$$

Where R = 8.3145 J/(K·mol), T is the temperature in Kelvin, ΔH^0 is the standard enthalpy of the reaction, and ΔS^0 is the standard entropy.

Equation 2 can be deduced as:

$$\ln(K_a) = -\Delta H^0 / R \times (1/T) + \Delta S^0 / R \quad \text{(Equation 3)}$$

Therefore, the following step required a van't Hoff plot of the natural logarithm of the affinity constant ($\ln(K_a)$) as a function of the reciprocal of the temperature (1/T in K^{-1}).[166]

For bimolecular equilibrium A + B ⇌ C:

$$K_a = [C] / ([A] [B])$$

$$= (C_c \times \Theta) / ([C_A \times (1 - \Theta)] \times [C_B \times (1 - \Theta)]) \quad \text{(Equation 4)}$$

When A and B is present at the same initial strand concentration C_0

$$K_a = \Theta / (C_0 \times (1 - \Theta)^2) \quad \text{(Equation 5)}$$

where C_0 is the initial strand concentration and Θ is Θ_T at each temperature. It should be noted that the analysis should be restricted between the temperature range for which $0.03 < \Theta < 0.97$ as it is relatively difficult to evaluate the affinity constant when almost all or almost none of the molecules are associated.[167]

Following the calculations described above, Fig. 2.3C can be converted into Fig. 2.3D. The so-called van't Hoff relation ($\ln(K_a)$ *versus* 1/T) should give a straight line, with a slope of $-\Delta H^0/R$ and Y-axis intercept of $\Delta S^0/R$. ΔG^0 can be calculated using Equation 2 based on the obtained ΔH^0 and ΔS^0 values. The same procedure was used for the determination of T_m and thermodynamic parameters for all complexes studied using melting profiles (Tables 2.2 and 2.3).

2.3.2.2. Fluorescence spectroscopy

As derivatives of phenylethynylpyren-1-yl, *ortho*-TINA monomers are fluorescent and when associated with DNA exhibit structured monomer peaks at 405 and 421 nm upon excitation at 373 nm.[132, 168, 169] It has also been shown that a fluorescent excited dimer (excimer) centred at 510 nm can be detected when two *ortho*-TINA monomers were placed opposite each other in the duplex.[132] In contrast, the intensity of the excimers was reduced for invasion products, whereas the monomeric fluorescence was increased. The excimers and monomers signal change of invading probes after the addition of native duplex provided evidence for the DDI process. Moreover, these signal changes can be

used for monitoring the process of DDI in real-time and kinetic parameters can be obtained.

2.3.2.3. 20% native PAGE

We performed 20% native PAGE analysis to provide direct evidence of DDI. 20% Native PAGE can be used to evaluate the properties of assembled structures. The migration rate of different samples varies depending on the size and molecularity of the sample. The incorporation of *ortho*-TINA monomers onto dsDNA resulted in slower migration of invading probes relative to the native duplex on the gel. The band corresponding to invasion products sits in between because of strand exchange. Based on the difference in the mobility of invasion products and starting materials, the process of DDI can be visualised.

2.4. Chapter summary

We demonstrated in this chapter that *ortho*-TINA modified duplexes could invade into native dsDNA in a sequence-specific manner. As expected, the speed of invasion was dependent on the number of *ortho*-TINA pairs incorporated in the duplex. However, the speed of invasion was also limited by the increased G/C content in the duplex. This approach is particularly attractive as *ortho*-TINA modified ONs are already commercially available, and no additional chemical modifications are required for obtaining effective invasion probes. We also demonstrated several strategies attempted for improving DDI kinetics towards long duplexes, although a remarkable improvement was not accomplished.

2.5. Results

2.5.1. Evaluation of thermal stability of duplexes possessing *ortho*-TINA monomers

For duplexes in Set 1, the inclusion of *ortho*-TINA monomers resulted in marginal changes in thermodynamic and thermal stabilities relative to the unmodified duplex (Table 2.2). In contrast, for duplexes formed between TINA-modified strands and native strands, both T_m and ΔG_{298} values were close to or higher than those for the native duplex. Relative to the native duplex, thermal stabilisation due to TINA-modified strand insertion varied from +3 °C to +7 °C (Table 2.2). In Set 2, duplexes containing paired *ortho*-TINA monomers resulted in 4-6 °C reductions in thermal stability relative to the native duplex (Table 2.3). However, thermal stabilities for duplexes formed between TINA-modified strands and native strands were 1-5 °C higher than those for the native duplex.

It is interesting to note that the introduction of *ortho*-TINA pairs in the duplex resulted in a less favoured ΔH in comparison with native duplexes for both sets, especially for **P1** and **P4** possessing two *ortho*-TINA pairs. With increasing numbers of *ortho*-TINA pairs, the ΔH became more favourable, similar to that of the native duplex. In contrast, invading probes exhibited a favourable entropy cost $T\Delta S$. For duplexes in Set 1 with only one G/C pair, the corresponding invasion products showed ΔH and $T\Delta S$ values that were similar to the starting materials, whereas for duplexes in Set 2 with six G/C pairs, ΔH values for invasion products were much more preferable than both native and invading duplexes (Tables 2.2 and 2.3).

To quantitatively assess the thermodynamics of the dsDNA-targeting potential of *ortho*-TINA modified ONs, the available free energy of recognition was estimated for isosequential DNA at RT (ΔG^{298}_{rec} values in Tables 2.2 and 2.3). Based on these ΔG^{298}_{rec} values, which ranged between -9 and -29 kJ/mol, we predict that the exchange of strands between native duplexes and *ortho*-TINA-modified duplexes should occur at temperatures well below the T_m for all duplexes studied.

Table 2.2. Hybridisation data for Set 1 and their free energy, ΔG^{298}_{rec} for recognition of dsDNA target C1 at 298K.

Entry	Name	T_m (°C)	ΔH^0 (kJ/mol)	$T_{298}\Delta S^0$ (kJ/mol)	ΔG_{298} (kJ/mol)	ΔG^{298}_{rec} (kJ/mol)
1	**C1**	37	-240 (±20)	-200 (±20)	-40 (±30)	
2	**P1**	37	-186 (±8)	-143 (±8)	-43(±11)	
3	**P2**	38	-197 (±12)	-153 (±11)	-44 (±16)	
4	**P3**	38	-234 (±14)	-188 (±13)	-46 (±19)	
5	**IP1**	41	-220 (±13)	-174 (±12)	-46 (±18)	-9
6	**IP2**	40	-190 (±8)	-144 (±7)	-46 (±11)	
7	**IP3**	44	-227 (±16)	-179 (±15)	-50 (±20)	-15
8	**IP4**	44	-196 (±9)	-147 (±9)	-49 (±13)	
9	**IP5**	44	-220 (±20)	-170 (±18)	-50 (±30)	-11
10	**IP6**	44	-236 (±14)	-189 (±13)	-47 (±19)	

T_m (°C) values (±1.5 °C) were determined by baseline fitting of melting curve (Abs_{249} from CD *versus* T) against increasing temperature (1.0 °C min^{-1}) on equimolar mixtures (2.0 μM in each strand) of ONs in 10 mM Na-phosphate, 140 mM NaCl, 0.1 mM Na_2-EDTA, pH 7.0. Thermodynamic parameters are calculated as described above at 298 K, errors were calculated as described in reference.[170]

Table 2.3. Hybridisation data for Set 2 and their free energy, ΔG^{298}_{rec} for recognition of dsDNA target **C2** at 298 K.

Entry	Name	T_m (°C)	ΔH^0 (kJ/mol)	$T_{298}\Delta S^0$ (kJ/mol)	ΔG_{298} (kJ/mol)	ΔG^{298}_{rec} (kJ/mol)
1	**C2**	55	-200 (±10)	-150 (±9)	-50 (±13)	
2	**P4**	51	-166 (±9)	-120 (±9)	-46 (±13)	
3	**P5**	49	-194 (±9)	-144 (±8)	-50 (±12)	
4	**P6**	50	-199 (±6)	-151 (±5)	-48 (±8)	
5	**IP7**	58	-262 (±12)	-202 (±11)	-60 (±16)	-29
6	**IP8**	57	-330 (±20)	-269 (±20)	-65 (±28)	
7	**IP9**	57	-254 (±13)	-197 (±12)	-57 (±18)	-19
8	**IP10**	56	-315 (±16)	-252 (±15)	-62 (±22)	
9	**IP11**	60	-229 (±7)	-172 (±7)	-57 (±10)	-20
10	**IP12**	58	-287 (±11)	-226 (±10)	-61 (±15)	

Conditions as specified in Table 2.2.

2.5.2. Evaluation of DDI by *ortho*-TINA probes using fluorescent spectrometry

As described above, fluorescence spectrometry was applied to evaluate the DDI process of *ortho*-TINA probes. We focused on the signal change of fluorescent excited dimer (excimer) centred at 510 nm and the monomer peaks at 405 nm upon excitation at 373 nm. Here, we observed that the intensity of the excimer band relative to that of the monomer was increased with the number of *ortho*-TINA pairs in the duplex. Specifically, I_{Fl}(510 nm)/I_{Fl}(405 nm) = 0.46 for **P1** (two monomer pairs), 0.60 for **P2** (three monomer pairs) and 2.67 for **P3** (four monomer pairs) (Fig. 2.4A, C, and E, respectively). A similar trend, but a more pronounced intensity of excimer bands relative to monomeric fluorescence bands, was observed for Set 2 (Fig. A.2 in Appendix A). In contrast, the intensity of the excimers was reduced for invasion products, whereas the monomeric

fluorescence was increased (**IP1** and **IP2** *versus* **P1**, Fig. 2.4A). This intensity shift was most noticeable for duplexes with four TINA modifications (Fig. 2.4E).

We used the described changes in the fluorescence spectra to monitor the exchange of strands at 25 °C (Figures 2.4B, D and F).[126] In these experiments, the native and *ortho*-TINA modified duplexes were mixed at a 1:1 ratio. Over the reaction time of 30 min, the intensity of excimer bands decreased, whereas that of monomeric bands intensified. This behaviour corresponds to the DDI illustrated in Fig. 2.1. The kinetic spectra of the reactions were obtained by monitoring the intensity of signals at 405 nm for Set 1 and 510 nm for Set 2 over time (Fig. 2.5), and kinetic parameters were calculated (Table 2.4). Strand exchange in Set 1 was significantly faster than for Set 2 at 25 °C. The speed of invasion was correlated with the number of *ortho*-TINA pairs in the duplex, which was more noticeable for Set 2 (see $t_{50\%}$ and $t_{75\%}$ values in Table 2.4). We assume that a more thermodynamically stable native duplex in Set 2 with six G/C pairs is a greater obstacle for invasion than the native duplex in Set 1. This is despite the fact that the driving force for the strand exchange (ΔG^{298}_{rec}) is better for Set 2.

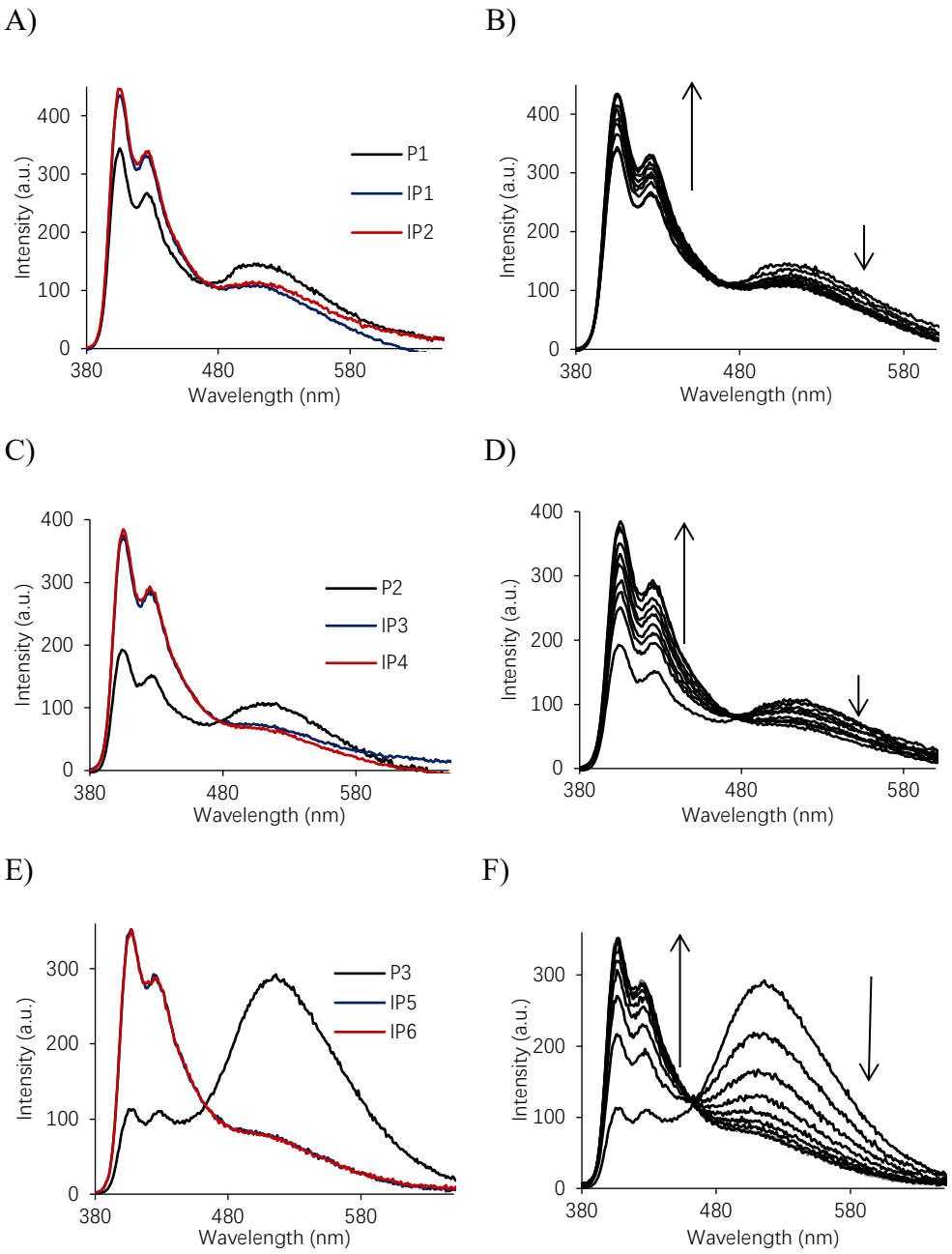


Figure 2.4. Steady state fluorescence emission spectra of duplexes with *ortho*-TINA pairs and corresponding invasion products (A: **P1** and **IP1** and **IP2**, C: **P2** and **IP3** and **IP4**, E: **P3** and **IP5** and **IP6**); Fluorescence kinetic spectra upon addition of pre-annealed native duplex **C1** to pre-annealed duplexes with TINA-TINA pairs (B: **P1**, D: **P2**, F: **P3**) over 30 min. All spectra were recorded using FuoroMax-4, each spectrum was recorded every 4 min; all samples were at 2.0 μM strand concentration at 25 °C in 10 mM Na-phosphate buffer (140 mM NaCl, 0.1 mM Na_2-EDTA, pH 7.0). For convenience the values in the y-axis (fluorescence intensity) were divided by 10^3. Arrows indicate the direction of signal change of fluorescent monomer and excimer signals over time.

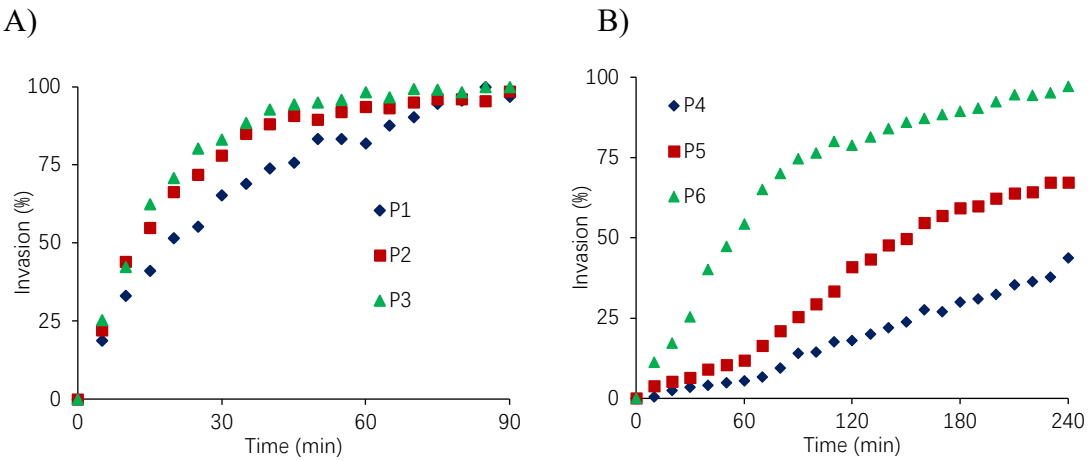


Figure 2.5. Representative time course of A) invasion for **P1**, **P2**, and **P3** upon addition of pre-formed 1.0 eq of native duplex **C1**; B) invasion for **P4**, **P5**, and **P6** upon addition of pre-formed 1.0 eq of native duplex **C2**. Spectra were recorded using fluorescence spectrometer (see experimental section for conditions) at 25 °C in 10 mM Na-phosphate buffer (140 mM NaCl, 0.1 mM Na_2-EDTA, pH 7.0) without sample shaking. Concentration of each strand is 1.0 µM.

Table 2.4. Kinetic parameters ($t_{50\%}$ and $t_{75\%}$ [min, ± 0.5 min]) for recognition of control duplexes by DNA probes with *ortho*-TINA pairs.

Set	Probes	$t_{50\%}$ (min)	$t_{75\%}$ (min)
Set 1	**P1**	21	40
	P2	13	25
	P3	11	22
Set 2	**P4**	>240	-[a]
	P5	150	-
	P6	56	95

[a] not reached

Circular dichroism (CD) spectra of duplexes containing *ortho*-TINA pairs had characteristics of antiparallel B-type duplexes with positive ellipticity at 275 nm and negative ellipticity at 250 nm.[152] The formation of several *ortho*-TINA pairs in the duplex results in a pronounced positive CD exciton couplet (+/− pattern, Fig. 2.6A) that indicates a clockwise orientation for *ortho*-TINA transition dipoles.[171] The intensity of the exciton couplet is correlated with the number of *ortho*-TINA pairs in the duplex for both sets. In contrast, only a negative Cotton effect was observed in CD spectra for invasion products (Fig. 2.7D) with peaks at 365 and 385 nm. Upon addition of the pre-formed target duplex **C1** to the complex showing the highest exciton couplet (**P3**), we

observed a significant increase in the intensity of bands at 250 and 275 nm (Fig. 2.6C), whereas the positive signal at 415 nm was decreasing as a result of disruption of *ortho*-TINA pairs and formation of invasion products (Fig. 2.6D). This correlates with the disappearance of *ortho*-TINA excimer in fluorescence spectra during strand exchange described above. However, the difference in intensity of signals in the TINA absorbance region in CD spectra for starting materials and products is not convenient for monitoring the strand exchange using CD (see Fig. A.1 in Appendix A for **P1** and **P2**).

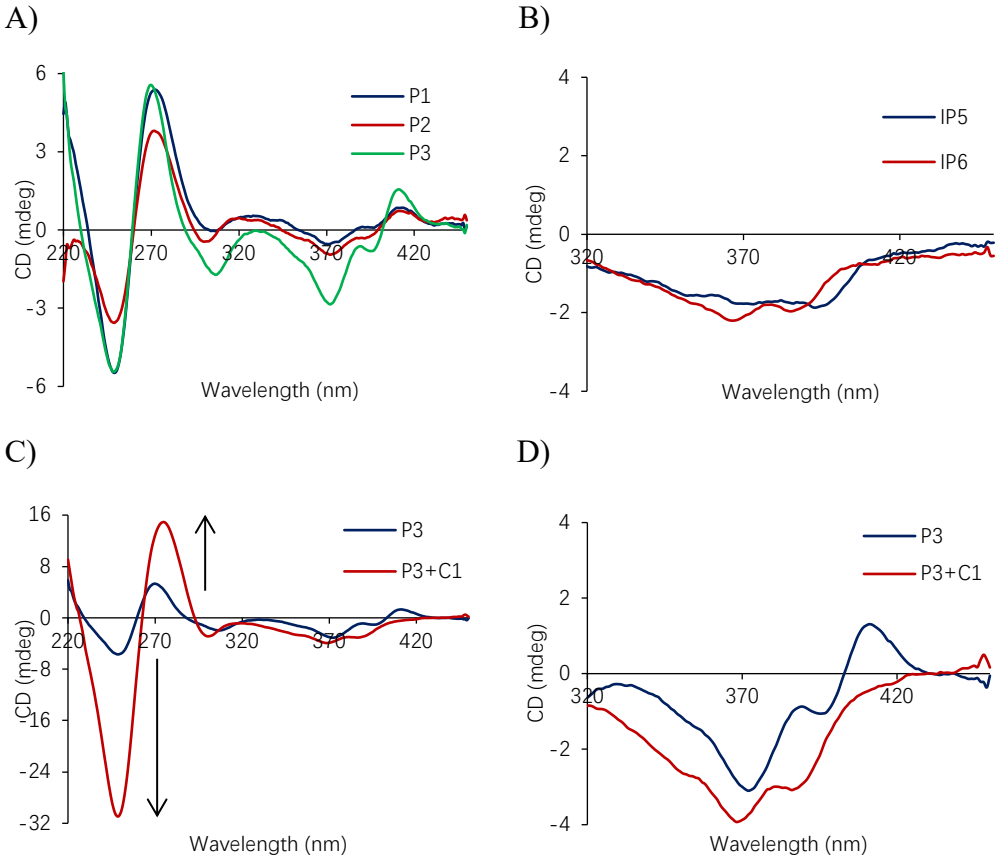


Figure 2.6. CD spectra for **P1**, **P2**, and **P3**. A) CD spectra of DNA duplexes with two, three and four *ortho*-TINA-pairs, 2.0 µM strand concentration at 25 °C in 10 mM Na-phosphate buffer (140 mM NaCl, 0.1 mM Na_2-EDTA, pH 7.0); B) CD spectra of invasion products **IP5** and **IP6** in the TINA absorbance region (320 to 450 nm); C) CD spectra of *ortho*-TINA duplex **P3** before and after addition of pre-annealed dsDNA target **C1** (all strands 2.0 µM); D) CD spectra of C) in the TINA absorbance region (320 to 450 nm). Arrows indicate the direction of CD signal change.

To directly monitor dsDNA-recognition by *ortho*-TINA probes and validate the fluorescence-based assay, we performed experiments using 20% non-denaturing PAGE (Fig. 2.7 and Figures A.4 and A.5 in Appendix A).

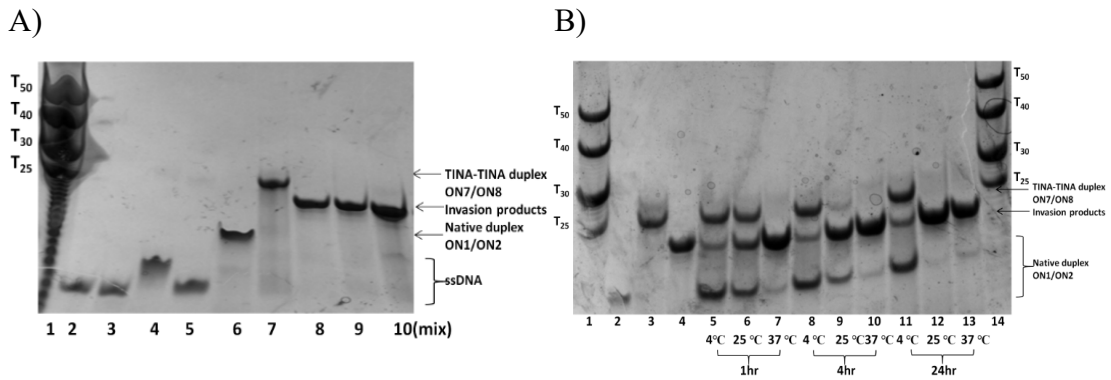


Figure 2.7. 20% Native PAGE showing strand exchange. A) 20% Native PAGE showing difference in mobility of unmodified duplex **C1** and duplexes with four *ortho*-TINA pairs **P3** in 1 × TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM Na_2-EDTA, pH 8.0) supplemented with 150 mM NaCl at 50 μM strand concentrations. Lanes are 1: ladder 2: **ON1**; 3: **ON2**; 4: **ON7**; 5: **ON8**; 6: **C1**; 7: **P3**; 8: **IP5**; 9: **IP6**; 5: mixture of **ON1**, **ON2**, **ON3**, **ON4**, heated at 90 °C for 5 min and then slowly cooled down to RT; B) 20% Native PAGE showing invasion of **P3** into **C1** in 1 × TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM Na_2-EDTA, pH 8.0) supplemented with 150 mM NaCl at 50 μM strand concentrations at different temperatures and times. Lanes are 1: ladder; 2: **C1**; 3: **P3**; 4: **IP5**; 5: **P3** + **C1**, 1 hr, 4 °C; 6: **P3** + **C1**, 1 hr, 25 °C; 7: **P3** + **C1**, 1 hr, 37 °C; 8: **P3** + **C1**, 4 h, 4 °C; 9: **P3** + **C1**, 4 h, 25 °C; 10: **P3** + **C1**, 4 h, 37 °C; 11: **P3** + **C1**, 24 h, 4 °C; 12: **P3** + **C1**, 24 h, 25 °C; 13: **P3** + **C1**, 24 h, 37 °C; 14: ladder.

As shown in Fig. 2.7A, the duplex with four *ortho*-TINA pairs (**P3**, lane 7, 13503.4 Da) had lower migration on the native gel than the invasion products (**IP5**, lane 8, 1160.5 Da and **IP6**, lane 9, 11629.5 Da) and the native duplex (**C1**, lane 6, 9755.6 Da). This was a result of the different molecular weights of these duplexes. Such difference in mobility on the gel was also seen for duplexes with three *ortho*-TINA pairs, but it becomes less pronounced for duplexes with two *ortho*-TINA pairs for both sets (Fig. A.4 for Set 1 and Fig. A.6 for Set 2 in Appendix A). To evaluate if the invasion products are thermodynamically more stable than the starting materials, all four strands were mixed in a buffer (**ON1**, **ON2**, **ON7** and **ON8**), heated at 90 °C and then slowly cooled down. Native gel shows that this procedure led to the formation of a band (lane 10) that had the same mobility as the expected invasion products (lanes 8 and 9, Fig. 2.7A). Incubation of equimolar quantities of *ortho*-TINA duplex **P3** and dsDNA target **C1** at 25 °C (Fig. 2.7B) resulted in the complete disappearance of bands corresponding to the starting materials after 24 h (lane 12 *versus* lanes 9 & 6, Fig. 2.7B). One should mention that the concentration of duplexes on the gel (50 μM) is significantly higher than concentrations used in the fluorescence experiments (1 and 2 μM), which inevitably results in increased thermal stability of all duplexes and may cause a delay in strand exchange. The invasion process was significantly faster at 37 °C than at 25 and 4 °C with the full completion of

the reaction after 1 hr. In contrast, reactions performed at 4 °C were not completed even after 24 h for Set 1. The slow kinetics of strand exchange observed for Set 2 in the fluorescence experiments was also seen on the native PAGE. For example, the strand exchange reactions for the duplex with two *ortho*-TINA pairs were not complete even after 24 h at 37 °C (Fig. A.6 in Appendix A for **P4**).

2.5.3. Evaluation of sequence specificity of duplexes possessing *ortho*-TINA monomers

In order to check the specific recognition properties of our *ortho*-TINA probes, we tested invasion against two unmodified DNA duplexes, one was mis-matched unmodified DNA duplex **MM1**, the other duplex **MM2** had partial complementarity with an unmodified DNA duplex. Our *ortho*-TINA probes proved to have specific recognition properties as no change in fluorescence emission spectra was observed upon addition of both mis-matched duplexes (Fig. 2.8).

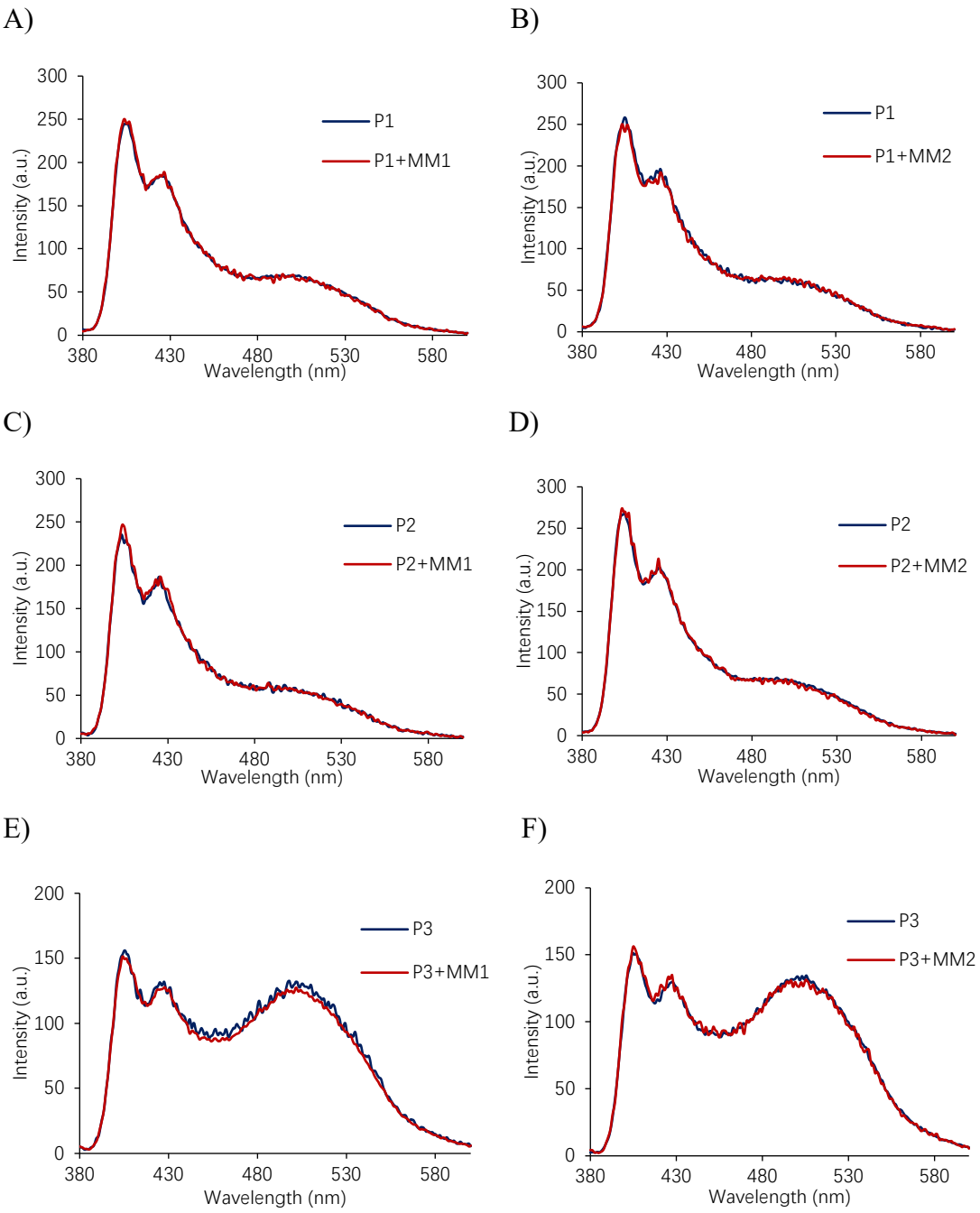


Figure 2.8. Fluorescence emission spectra of *ortho*-TINA duplexes **P1** - **P3** in the absence and in the presence of mis-matched duplexes **MM1** and **MM2**, respectively. Fluorescence emission spectra of *ortho*-TINA duplexes (A: **P1**, C: **P2**, E: **P3**) before and after addition of pre-annealed mis-matched duplex **MM1**; Fluorescence emission spectra of *ortho*-TINA-DNA duplexes (B: **P1**, D: **P2**, F: **P3**) before and after addition of pre-annealed partially complementary duplex **MM2**. All spectra were recorded using Perkin-Elmer LS55 after 24 h of incubation at RT; all strands are at 1.0 µM strand concentration at 25 °C in 10 mM Na-phosphate buffer (140 mM NaCl, 0.1 mM Na_2-EDTA, pH 7.0).

Sequence-specific targeting properties of Set 2 were evaluated using 20% native PAGE. No invasion was observed as upper and lower bands corresponding to starting materials were separated clearly after incubation at 37 °C for 24 h (Figures A.9, A.10 and A.11 in Appendix A).

2.5.4. Monitoring DDI in competitive environment using excess of salmon testes DNA

We then performed competitive experiments using salmon testes DNA (ST, 100 × concentration) to monitor invasion of *ortho*-TINA probes into target native duplex mimicking natural conditions.

As a positive control, a fluorescence emission spectrum of *ortho*-TINA duplex (1 μM in 10 mM Na-phosphate buffer (140 mM NaCl, 0.1 mM Na_2-EDTA, pH 7.0)) was recorded (take **P2** as an example). Then 100 × ST was added, and after shaking was incubated at RT for 24 h. After recording the spectrum, 1 eq of pre-annealed native duplex was added to the solution, and the fluorescence emission spectrum was recorded every 15 min over 2 h. Finally, the solution was heated at 90 °C for 5 min, cooled down to RT slowly to get the spectrum of thermodynamic products (TD) (Fig. 2.9A). We expected to see no changes in fluorescence signals of *ortho*-TINA duplex **P2** before and after adding of ST, and the kinetics of invasion against **C1** should not be influenced by the presence of ST.

As a negative control, 1 eq of pre-annealed native duplex was added to the solution of 1 μM *ortho*-TINA duplex in 10 mM Na-phosphate buffer, and fluorescence emission spectrum was recorded every 15 min over 2 h. Then 100 × of ST was added, and after mixing was incubated at RT for 24 h. After recording the fluorescence spectrum, the solution was heated at 90 °C for 5 min, cooled down to RT slowly to get the spectrum of TD (Fig. 2.9B). We expected to see the fluorescence signals of invasion products to be the same before and after adding ST.

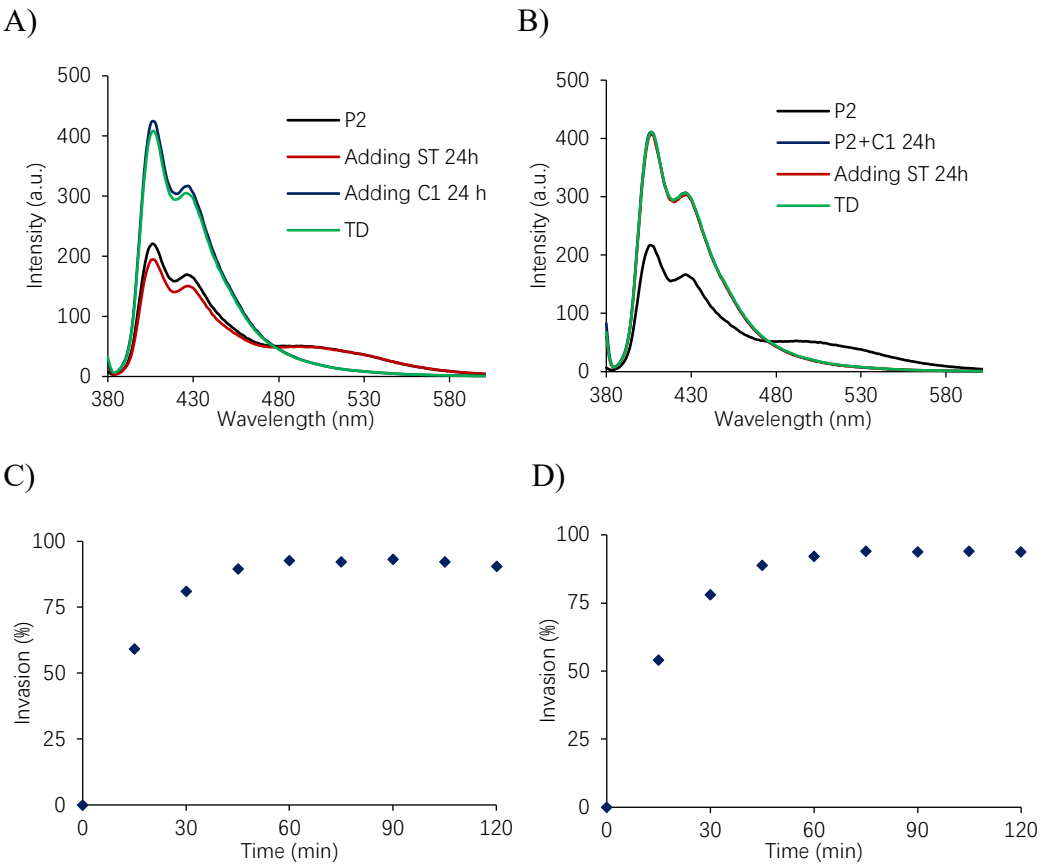


Figure 2.9. Competitive experiments for **P2** in the presence of excess of salmon testes DNA. A) Fluorescence emission spectra of **P2** (alone), **P2** + **ST** (100 ×), **P2** + **ST** + **C1**, and TD; B) Fluorescence emission spectra of **P2** (alone), **P2** + **C1, P2** + **C1**+ **ST** (100 ×), and TD; C) Time course of invasion upon addition of **C1** to **P2** showed in A; D) Time course of invasion upon addition of **C1** to **P2** showed in B. All spectra were recorded using Perkin-Elmer LS55 at 25 °C. Invading probes and target native duplexes are at 1.0 μM strand concentration at 25 °C in 10 mM Na-phosphate buffer (140 mM NaCl, 0.1 mM Na_2-EDTA, pH 7.0).

As shown in Fig. 2.9A, the addition of ST (100 ×) did not change the fluorescence spectra of **P2** over 24 h incubation, which means that *ortho*-TINA modified duplexes maintained their sequence specificity in the condition mimicking high concentration of DNA in cells. Furthermore, in the presence of excess non-targeted sequences (ST, 100 ×), the recognition of the native duplex occurred as expected, and the speed of the invasion was not influenced by the presence of non-complementary duplex. In the negative control, no signal change was observed after the invasion process was finished, and ST-DNA was added and incubated for 24 h (Fig. 2.9B). The time-course of invasion progress upon addition of **C1** for both positive (Fig. 2.9C) and negative controls (Fig. 2.9D) provided similar $t_{50\%}$ and $t_{75\%}$ values as reported in Table 2.4.

The same experiments were conducted for duplexes in Set 2, and the time course was plotted monitoring the signal change at 510 nm, giving similar $t_{50\%}$ and $t_{75\%}$ values as given in Table 2.4.

2.5.5. Invasion of *ortho*-TINA probes against long DNA duplexes

In order to check whether our *ortho*-TINA-modified probes can invade into long DNA duplexes, we tested invasion against two long native duplexes, **LC1** and **LC2**, with three and six bps longer at each end than **C1**, respectively.

Duplexes possessing two (Fig. 2.10A) and three *ortho*-TINA pairs (Fig. 2.10C) showed good invasion properties into six bps longer duplexes. However, for duplexes possessing four *ortho*-TINA pairs, invasion into long duplex is more difficult, the invasion took more than 3 h to occur (Fig. 2.10E). For a duplex with 12 bps longer, DDI was problematic. Only **P2** with three *ortho*-TINA pairs showed marginal changes in the fluorescence signal (Fig. 2.10D) and no changes in fluorescence spectra were detected for **P1** and **P3** probes mixed with **LC2** after 24 h incubation at RT (Fig. 2.10B and F).

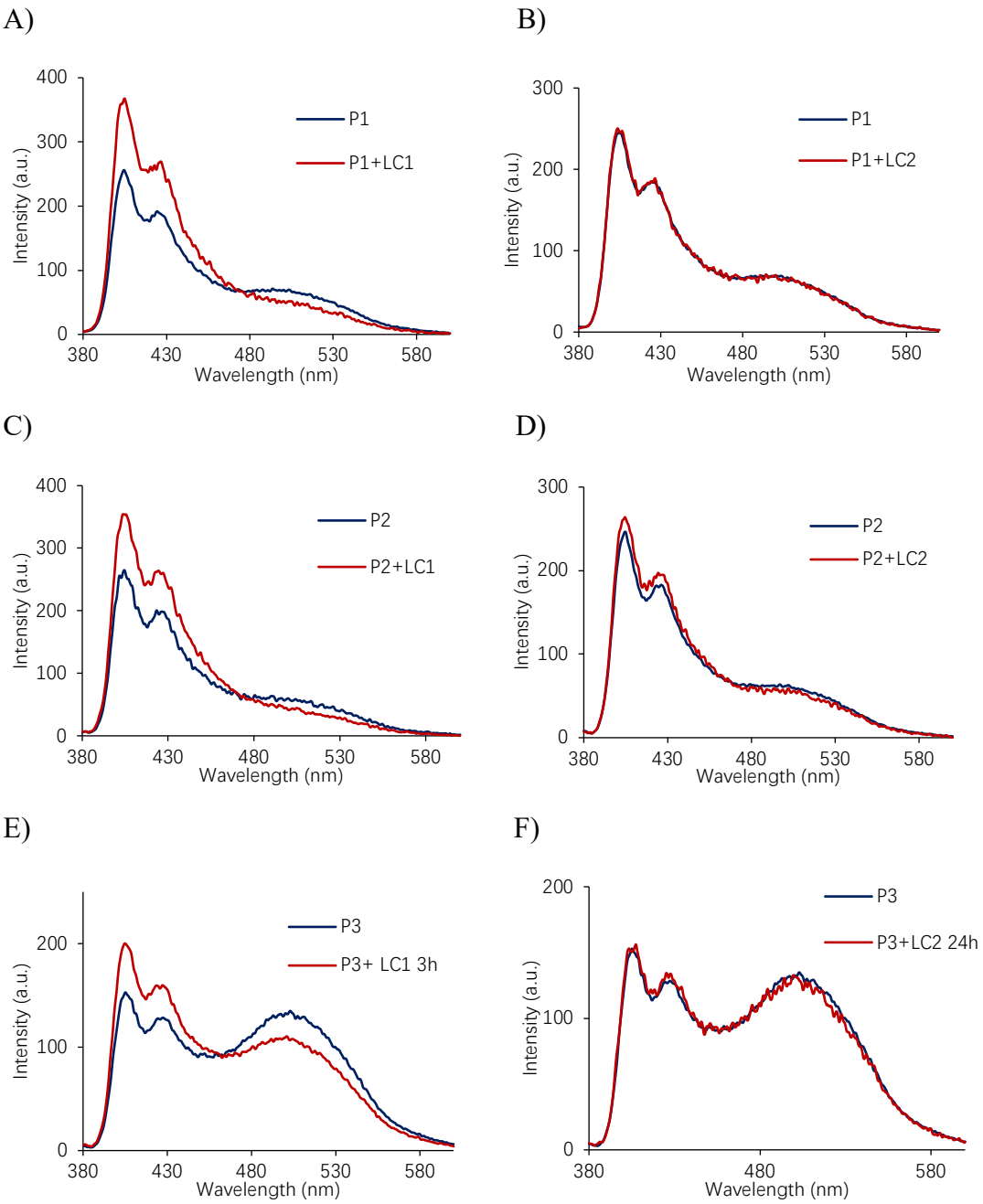


Figure 2.10. Fluorescence emission spectra of *ortho*-TINAs probes (A and B for **P1**; C and D for **P2**; E and F for **P3**) before and after addition of pre-annealed long duplexes **LC1** (A, C and E) and **LC2** (B, D and F). All spectra were recorded using Perkin-Elmer LS55 at 25 °C; all samples were at 1.0 μM strand concentration at 25 °C in 10 mM Na-phosphate buffer (140 mM NaCl, 0.1 mM Na_2-EDTA, pH 7.0). A), B), C), and D) were recorded after 120 min of incubation at 25 °C, whereas E) and F) was recorded after incubation at 25 °C for 3 h and 24 h, respectively.

2.5.6. Attempts for improving DDI efficiency against long duplexes

Invasion into duplexes longer than the probes is still a challenge for our *ortho*-TINA probes. In order to make it suitable for invasion into longer duplexes, we designed *ortho*-TINA modified duplexes with tails at the 5′- or 3′- end, as shown in Table 2.5 and 2.6, respectively. The idea was that tails could hybridise with the complementary region of a long duplex, and therefore serve as a starting point and at the same time help opening up the long duplex and initiate DDI. The new design should be able to fulfil the requirements of improving binding affinity towards long duplexes as well as accelerating the speed of invasion.

Table 2.5. Abbreviation and sequence for *ortho*-TINA probes with tails at the 5′- end.

Tail	Probe	Sequences	Invasion products towards LC1	
Tails at the 5′- end	P7	5′-**X**TTTTTA**X**ATAAT**X**ATGAT**X**TTA 3′-AAT**X**TATTA**X**TACTA**X**AATTAT**X**	IP13	5′- TTTTTAATAATATGATTTAATA 3′- AAT**X**TATTA**X**TACTA**X**AATTAT**X**
			IP14	5′-**X**TTTTTA**X**ATAAT**X**ATGAT**X**TTA 3′-AAAAATTATTATACTAAATTAT
	P8	5′-**X**TTTTTA**X**ATAATATGAT**X**TTA 3′-AAT**X**TATTATACTA**X**AATTAT**X**	IP15	5′- TTTTTAATAATATGATTTAATA 3′- AAT**X**TATTATACTA**X**AATTAT**X**
			IP16	5′-**X**TTTTTA**X**ATAATATGAT**X**TTA 3′- AAAAATTATTATACTAAATTAT
	P9	5′-TTTTTA**X**ATAAT**X**ATGAT**X**TTA 3′-AAT**X**TATTA**X**TACTA**X**AATTAT	IP17	5′- TTTTTAATAATATGATTTAATA 3′- AAT**X**TATTA**X**TACTA**X**AATTAT
			IP18	5′-TTTTTA**X**ATAAT**X**ATGAT**X**TTA 3′-AAAAATTATTATACTAAATTAT
	P10	5′-TTTTTA**X**ATAATATGAT**X**TTA 3′-AAT**X**TATTATACTA**X**AATTAT	IP19	5′- TTTTTAATAATATGATTTAATA 3′- AAT**X**TATTATACTA**X**AATTAT
			IP20	5′-TTTTTA**X**ATAATATGAT**X**TTA 3′-AAAAATTATTATACTAAATTAT

X denotes *ortho*-TINA monomer.

Table 2.6. Abbreviation and sequence for *ortho*-TINA probes with tails at the 3′- end.

Tail	Probe	Sequences	Invasion products towards LC1	
Tails at the 3′-end	**P11**	5′-TTA**X**ATAAT**X**ATGAT**X**TTA-2′-OMe(AUA) 3′-2′-OMe(AAA)AAT**X**TATTA**X**TACTA**X**AAT	**IP21**	5′- TTTTTAATAATATGATTTAATA 3′-2′-OMe(AAA)AAT**X**TATTA**X**TACTA**X**AAT
			IP22	5′- TTA**X**ATAAT**X**ATGAT**X**TTA-2′-OMe(AUA) 3′- AAAAATTATTATACTAAATTAT
	P12	5′-TTA**X**ATAAT**X**ATGAT**X**TTAATA 3′-AAAAAT**X**TATTA**X**TACTA**X**AAT	**IP23**	5′- TTTTTAATAATATGATTTAATA 3′- AAT**X**TATTA**X**TACTA**X**AATTAT**X**
			IP24	5′-**X**TTTTTA**X**ATAAT**X**ATGAT**X**TTA 3′- AAAAATTATTATACTAAATTAT

X denotes *ortho*-TINA monomer.

Taking duplexes with tails at the 5′-end (**P7** – **P10**, Table 2.5) as an example. The invasion potential of *ortho*-TINA modified duplexes with tails towards long duplex **LC1** was evaluated by thermal denaturation experiments (Table A.1 in Appendix A). The available free energy of recognition for isosequential DNA at room temperature (ΔG^{298}_{rec}) was calculated based on thermodynamic parameters obtained in the UV-Vis melting experiments.

Relative to the starting materials, T_m values of corresponding invasion products were at least higher than one of the starting duplexes. Based on the ΔG^{298}_{rec} values, which ranged between -5 and -15 kJ/mol, we can predict that the exchange of strands between duplexes with tails at the 5′-end and duplex **LC1** should occur at the experimental temperature (25 °C), well below the T_m for all duplexes studied.

ortho-TINA modified duplexes with tails also exhibited structured monomer peaks at 405 and 421 nm and an excimer peak at 510 nm upon excitation at 373 nm in the fluorescence spectra.[132, 168, 169] We can still follow these changes in the fluorescence spectra to monitor the exchange of strands at 25 °C (Fig. 2.11 and 2.12).[126] In these experiments, the target duplexes (same length: **LC1**, Fig. 2.11; longer duplex: **LC2**, Fig. 2.12) were added to the *ortho*-TINA modified duplexes with tails at the 5′-end at a 1: 1 ratio. During the reaction (4 h) for **LC1** (Fig. 2.11), the intensity of the excimer bands was reduced, whereas the intensity of monomeric bands was getting stronger. However, the change in the excimer band was not as significant as for duplexes in Set 1. For invasion towards longer duplex **LC2**, during the reaction (6 h), marginal changes at 510 nm were observed for all four duplexes with tails, which contradicted our assumption that tails at the 5′-end may improve DDI.

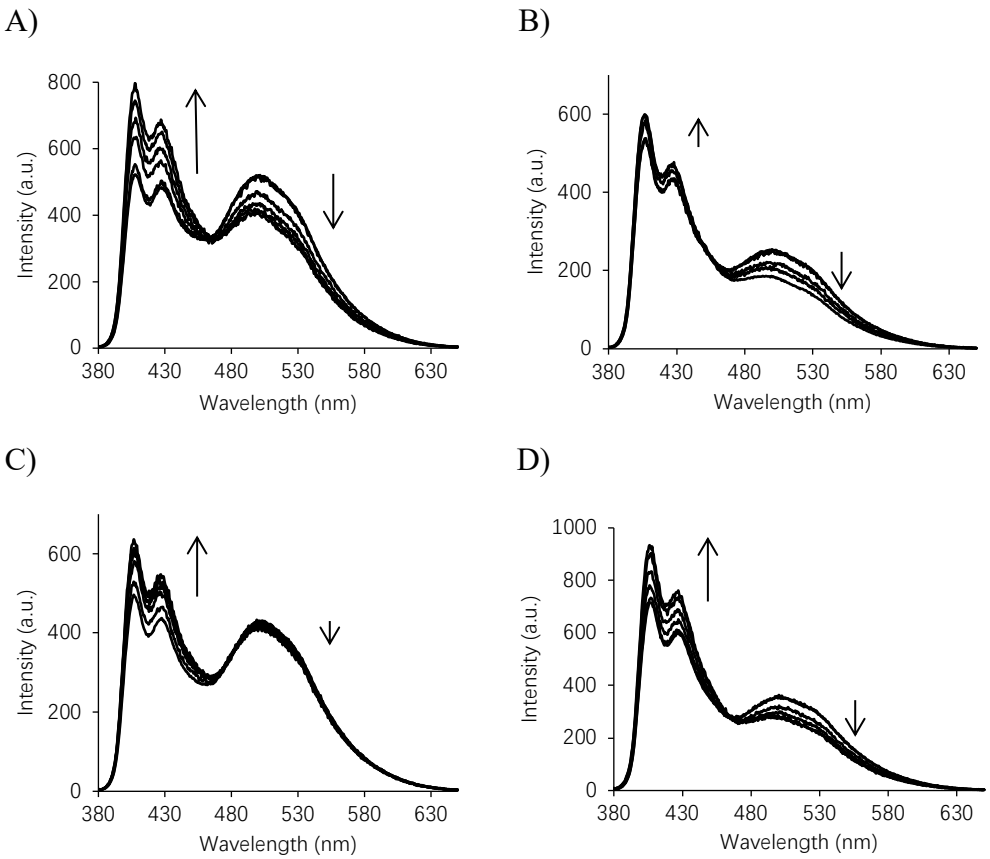


Figure 2.11. Fluorescence emission spectra upon addition of pre-annealed **LC1** to pre-annealed *ortho*-TINA modified duplexes with tails at the 5′-end (A: **P7**; B: **P8**; C: **P9**; D: **P10**) over 4 h. All spectra were recorded using FluoroMax-4 every 15 min at 37 °C; all samples were at 2.0 μM strand concentration at 25 °C in 10 mM Na-phosphate buffer (140 mM NaCl, 0.1 mM Na_2-EDTA, pH 7.0). For convenience the values in the y-axis (fluorescence intensity) were divided by 10^2. Direction of arrows indicates the signal change from 0 to 4 h.

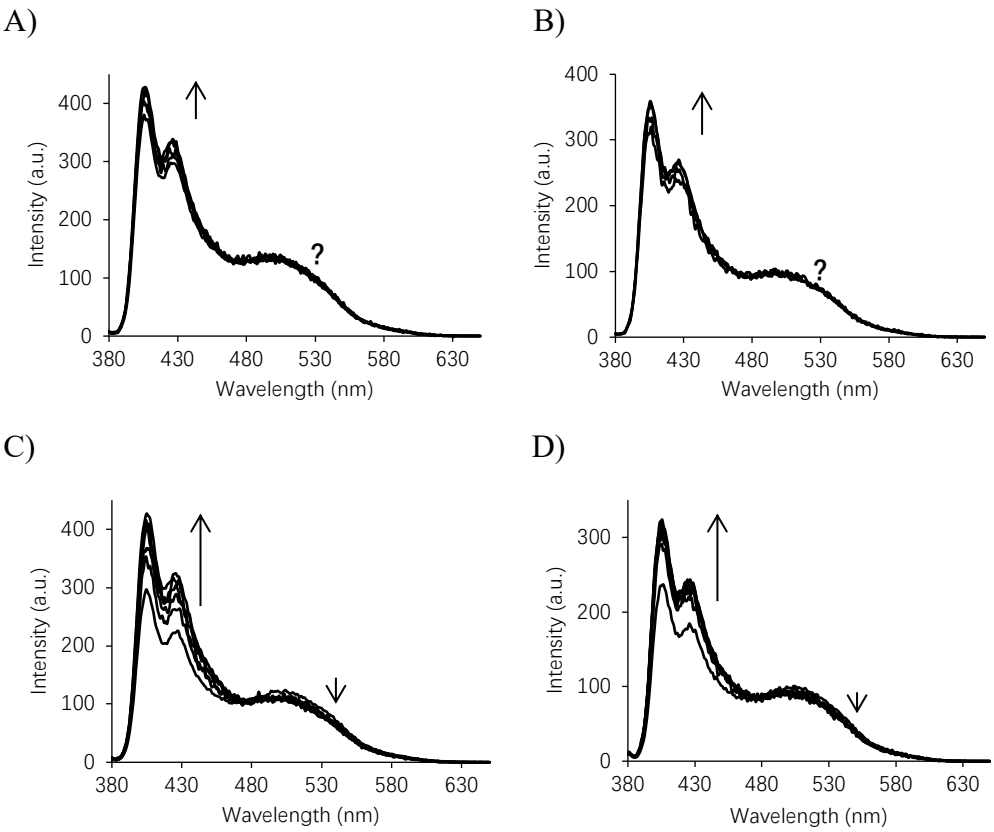


Figure 2.12. Fluorescence emission spectra upon addition of pre-annealed **LC2** to pre-annealed *ortho*-TINA modified duplexes with tails at the 5′-end (A: **P7**; B: **P8**; C: **P9**; D: **P10**) over 6 h. All spectra were recorded using FluoroMax-4 every 30 min at 37 °C; all samples were at 2.0 μM strand concentration at 25 °C in 10 mM Na-phosphate buffer (140 mM NaCl, 0.1 mM Na_2-EDTA, pH 7.0). For convenience the values in the y-axis (fluorescence intensity) were divided by 10^2. Direction of arrows indicates the signal change from 0 to 6 h.

In terms of *ortho*-TINA modified duplexes with a tail at the 3′-end, which is **P11** with 2′-OMe group and **P12** with three nucleotides (Table 2.6), fluorescence emission kinetic spectra were recorded upon addition of native duplex **C1**, duplex with the same length (**LC1**), and longer duplex **LC2**, at 1: 1 ratio (Fig. 2.13 for **P11** and Fig. 2.14 for **P12**). During reaction (4 h), we observed that the reducing intensity of excimer bands and the increasing intensity of monomeric bands were more pronounced in comparison with that of duplexes with tails at the 5′-end. However, during the recording of spectra upon addition of mis-matched duplex **MM1** (Fig. 2.13D and Fig. 2.14D) and partially complementary duplex **MM2** (Fig. 2.13E and Fig. 2.14E) at 1: 1 ratio, a noticeable signal decay at 510 nm was observed, which means that tails at the 3′-end lead to the loss of sequence-specificity.

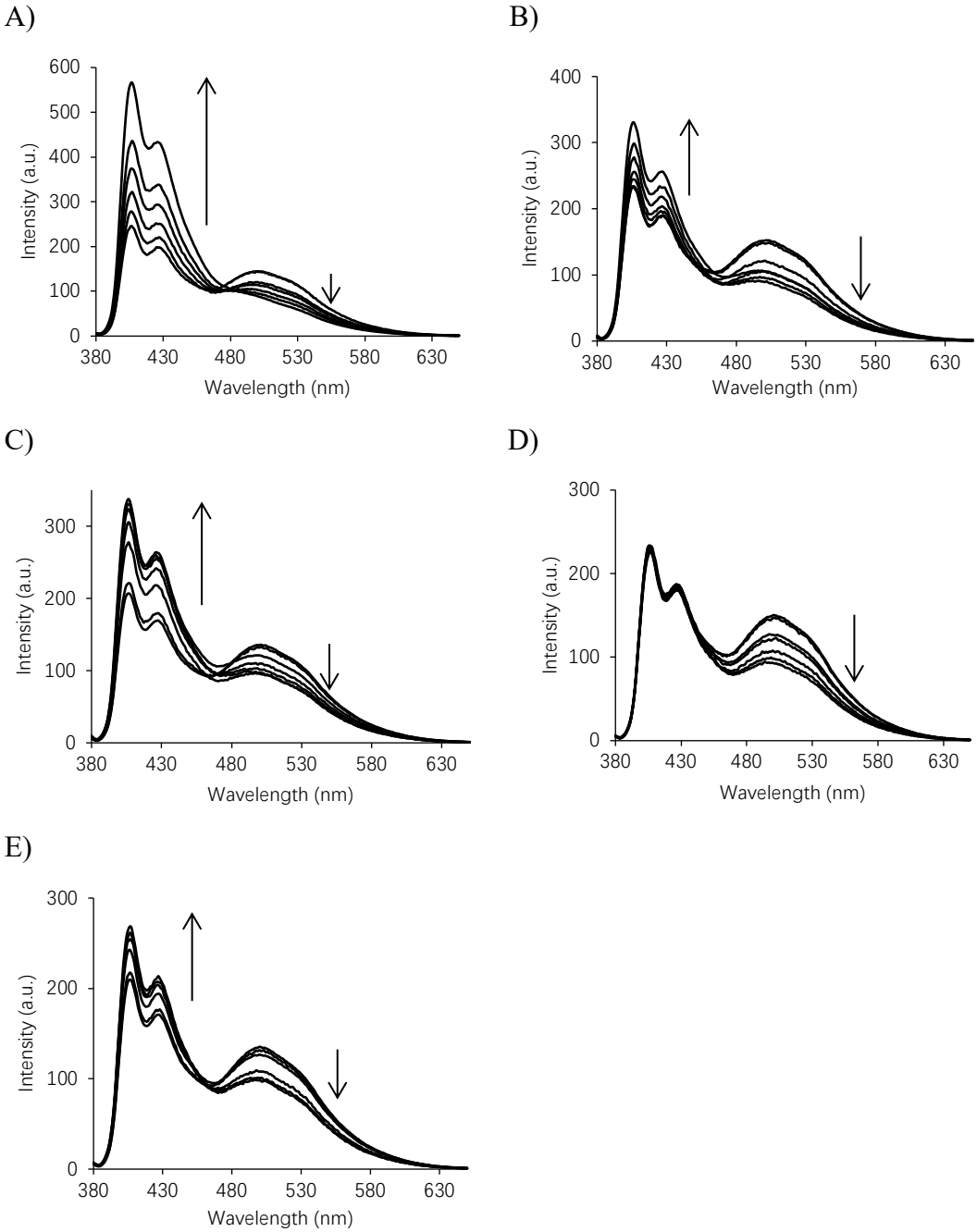


Figure 2.13. Fluorescence emission spectra upon addition of pre-annealed target duplexes (A: **C1**; B: **LC1**; C: **LC2**; D: **MM1**; E: **MM2**) to pre-annealed *ortho*-TINA modified duplexes with 2′-OMe group at the 3′-end **P11** over 4 h. All spectra were recorded using FluoroMax-4 every 20 min at 37 °C; all samples were at 2.0 μM strand concentration at 25 °C in 10 mM Na-phosphate buffer (140 mM NaCl, 0.1 mM Na_2-EDTA, pH 7.0). For convenience the values in the y-axis (fluorescence intensity) were divided by 10^2. Direction of arrows indicates the signal change from 0 to 4 h.

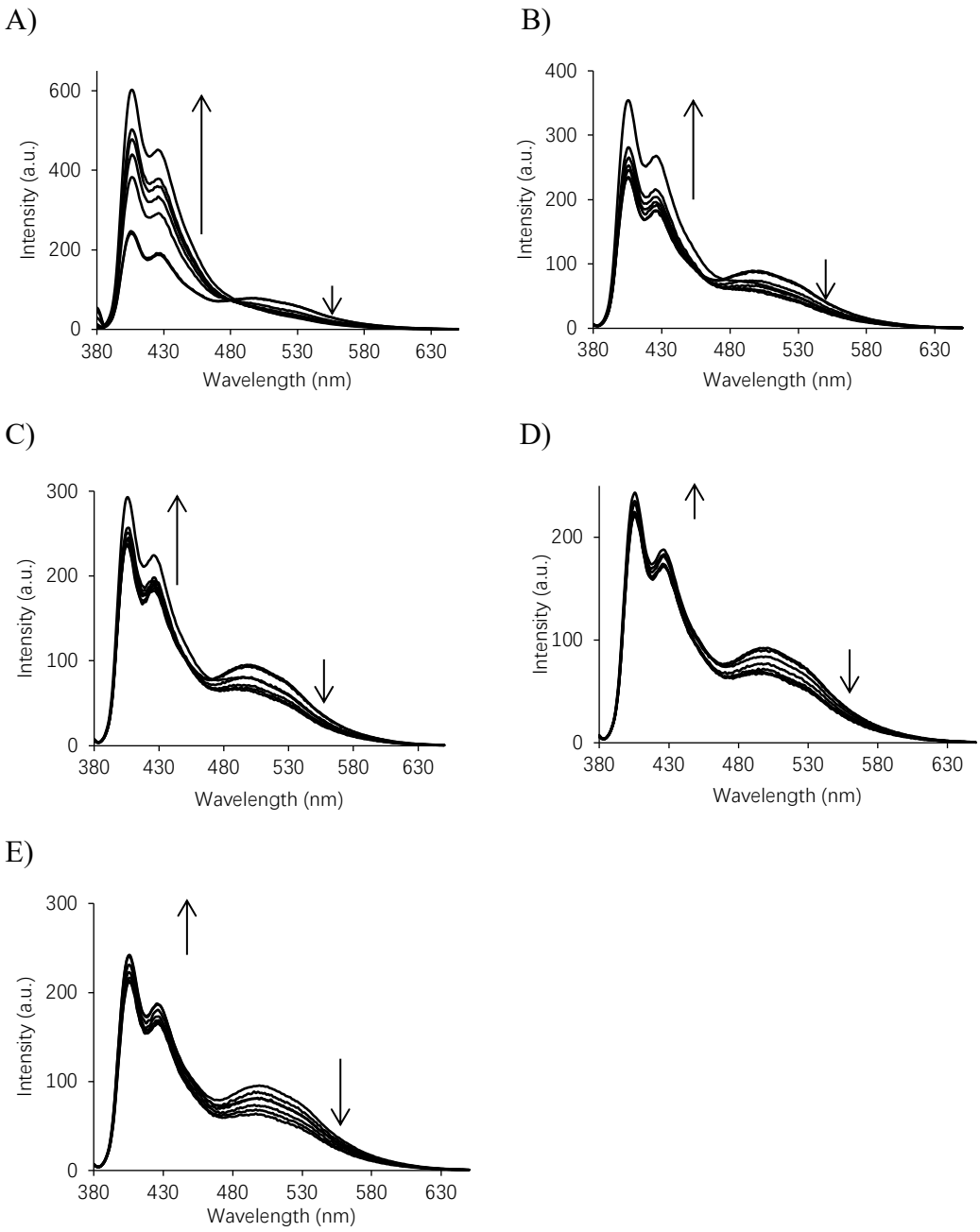


Figure 2.14. Fluorescence emission spectra upon addition of pre-annealed target duplexes (A: **C1**; B: **LC1**; C: **LC2**; D: **MM1**; E: **MM2**) to pre-annealed *ortho*-TINA modified duplexes with tails at the 3′-end **P12** over 4 h. All spectra were recorded using FluoroMax-4 every 30 min at 37 °C; all samples were at 2.0 μM strand concentration at 25 °C in 10 mM Na-phosphate buffer (140 mM NaCl, 0.1 mM Na_2-EDTA, pH 7.0). For convenience the values in the y-axis (fluorescence intensity) were divided by 10^2. Direction of arrows indicates the signal change from 0 to 6 h.

2.6. Discussion

DDI is an attractive mechanism for the recognition of dsDNA because the probes are involved in Watson/Crick base-pairing with both target strands. This results in an energy gain during complex formation compared to both strand invasion and triplex formation. Moreover, in contrast to DNA triplexes, DDI does not require the presence of divalent cations, such as Mg^{2+}, and has no sequence restrictions. This means that, theoretically, probes may be designed for any DNA target. In the context of DDI, PNAs containing pseudo-complementary base-pairs (2,6-diaminopurine and 2-thiouracil)[67] do have desirable affinity characteristics. However, these PNA probes require salt conditions outside the normal physiological range.[67] Over the last decade, several examples of pyren-1-yl modified nucleotides have been used for DDI,[112, 113, 115, 159-165] but the requirement for multi-step synthesis hinders their widespread application. It has been shown previously that INA-modified oligonucleotides can be used for DDI at RT and close to physiological conditions.[123, 125, 126] *para*-TINA monomers have also been developed for DNA triplexes.[127, 128] When an ethynylpyren-1-yl moiety is placed in the *ortho*-position of the benzene ring, it (*ortho*-TINA) mimicks INA in the structure of DNA and thus can be used for targeting DNA via Watson/Crick hybridisation.[132] This property improves the performance of DNA primers in the polymerase chain reaction (PCR)[172] and has led to the commercialisation of *ortho*-TINA monomers. An important observation is that both *para*- and *ortho*-TINA modified ONs have considerably reduced affinity towards complementary RNAs (ΔT_m= -6 – -10 °C per modification),[132] meaning that recognition of dsDNA can occur in the presence of RNA.[131]

The use of *ortho*-TINA monomers in DDI has not previously been evaluated. By placing two – four *ortho*-TINA pairs separated by at least three native nucleotide pairs, we obtained DNA helices with several features useful for DDI. Their thermal stabilities were close to or lower than those for native duplexes. A strong exciton couplet for *ortho*-TINA pairs was observed in CD spectra, accompanied by a strong excimer in fluorescence emission spectra. The increased intensity of the excimer and the exciton couplet bands correlated with the increasing number of *ortho*-TINA pairs formed. Contrary to that, duplexes formed by *ortho*-TINA-modified ONs and unmodified ONs exhibited higher thermal stabilities exceeding the T_m values of native duplexes. No exciton couplets and low excimer intensities were observed for these duplexes. The ability of duplexes with *ortho*-TINA pairs to participate in DDI was evaluated by fluorescence spectroscopy (indirectly) and 20% native PAGE (directly). Both methods showed that the reaction is

faster for Set 1 than for Set 2 despite the fact that Set 2 had preferable Gibbs free energy (ΔG^{298}_{rec} values were -29 to -19 kJ/mol for Set 2 *versus* -15 to -9 kJ/mol for Set 1). This phenomenon has also been reported earlier,[115] indicating that thermodynamics does not determine the kinetics in DDI and other factors such as the presence of G/C pairs that may influence the kinetics. The speed of DDI increased in the order of *ortho*-TINA pairs present in the starting duplex and was faster at higher experimental temperatures. It is important that no invasion was observed for mis-matched and partially complementary targets.

Even though the insertion of *ortho*-TINA monomers in duplexes showed excellent DDI potential, it did not overcome the difficulty that most invasion probes have, to invade into long native duplexes. A strategy that put a tail at the 5′- or 3′-end is theoretically able to help open the long duplex and accelerate invasion. Tails were designed with and without *ortho*-TINA monomers, and 2′-OMe nucleotides at the 3′-end were also evaluated. However, fluorescence spectroscopy results showed that tails at the 5′-end, with and without *ortho*-TINA monomers, did not show any benefit in invading into long duplexes in comparison with probes without a tail. It was, in contrast, very slow with minimal invasion occurring; tails at the 3′-end, both 2′-OMe nucleotides and DNA nucleotides, were able to open up the target long duplex to some extent, but a loss in specificity was observed at the same time. Based on these results, other modifications on tails at the 3′-end, which can help to open longer duplexes and maintain sequence-specificity is worth trying.

2.7. Conclusion

In this Chapter, we have demonstrated the use of *ortho*-TINA modified ONs in sequence-specific recognition of mixed-sequence dsDNA targets. Commercial availability, compatibility with a range of DNA modifications and the ease of design of *ortho*-TINA-DNAs for DDI are attractive features that can appeal to people working in the field of dsDNA targeting. This method can be used to improve the detection of pathogenic DNA, sequence-selective treatment of dsDNA for mutagenesis or recombination of genes and can also be applied in drug design against genetic diseases.

Chapter 3. Synthesis and evaluation of invasion probes modified with zwitterionic phosphate moiety (N+) and *ortho*-TINA monomers

3.1. Introduction

ONs (DNA or RNA) possessing modifications on the phosphate backbone have gained a lot of attention in recent years because such modifications can improve the nuclease resistance of ONs to a greater extent in comparison with native ONs. Charge-neutral phosphate modification is of particular interest as it can also enhance the binding affinity of ONs towards complementary DNA/RNA/dsDNA as well as their cell uptake. PNA (Fig. 3.1) is one successful example that has improved binding towards complementary DNA and RNA strands because of the lack of a negative charge.

Methylphosphonate linkage (PMe, Fig. 3.1), as the first charge-neutral phosphate mimic, was designed and introduced into the DNA backbone to improve stability of ONs against enzymatic digestion as well as binding affinity towards ssDNA and dsDNA.[173] However, poor aqueous solubility,[174] reduced binding affinity towards complementary RNA[175] and the destabilising effect on the thermal stability of G4s[176] hinders its applications. Phosphate methylated linkage (POMe, Fig. 3.1), in contrast, show improved specificity with marginal destabilisation towards complementary DNA sequences.[177] An internucleotide tetramethyl phosphoryl guanidine group (Tmg, Fig. 3.1) was introduced into ONs and led to slight changes in binding affinity towards complementary DNA and RNA sequences.[178] Introduction of a single sulfonamide RNA (SaRNA, Fig. 3) dimer into ON backbone led to decreased duplex thermal stability with complementary DNA and RNA, but the level of destabilisation was lower for ON/RNA duplexes than that for ON/DNA duplexes.[179] Branched, charge-neutralising sleeves (BCNSs, Fig. 3.1) were incorporated into the ON backbone,[180] forming self-neutralising ONs with good aqueous solubility, enhanced nuclease stability, low cytotoxicity and increased thermal stability in the context of 2′-OMe-RNA duplexes.

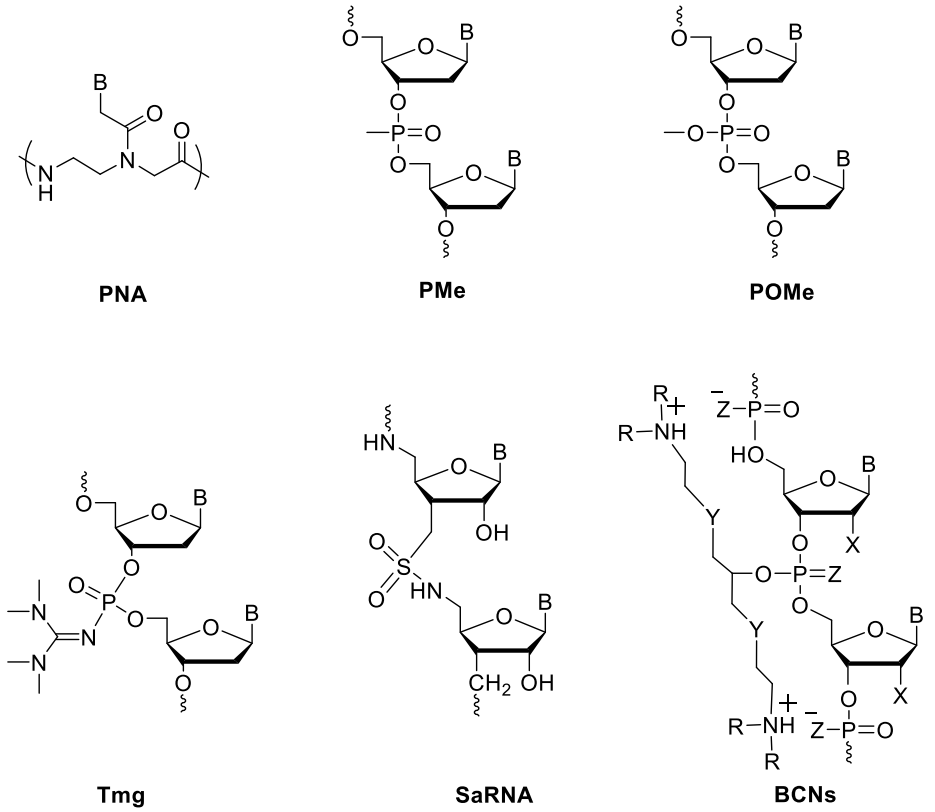


Figure 3.1. Structures of chemically modified ON analogues. SaRNA stands for sulfonamide RNA; BCNs stands for branched charge-neutralising sleeves, X is H or OMe, Z is O or S, Y is O or CH_2, and R is H or CH_3. B, heterocyclic base.

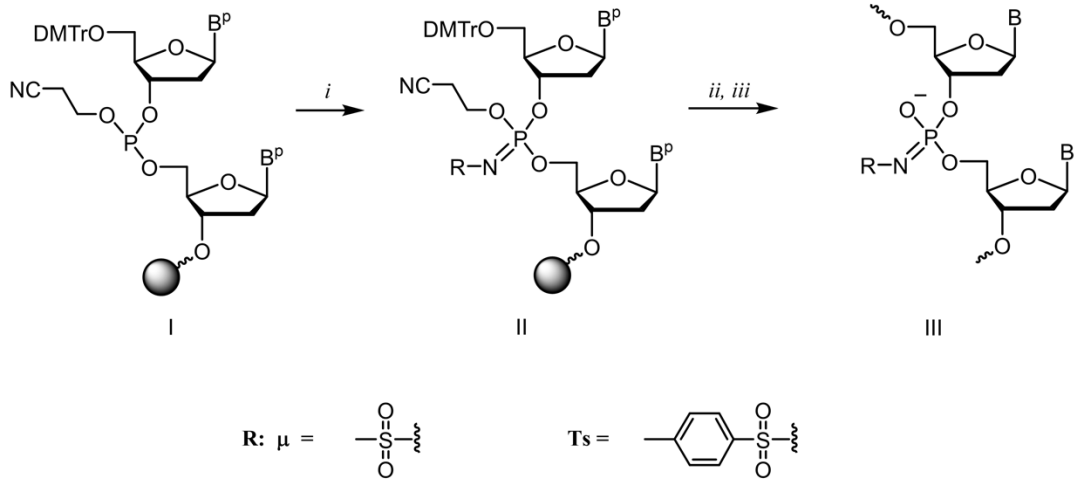
3.2. Hypothesis

Considering the fact that chemically modified ONs bearing several neutral phosphate mimics could hybridise with complementary ssDNA or RNA with higher affinity in comparison with native ONs, we hypothesised that the introduction of several charge-neutral modifications into the backbone of *ortho*-TINA probe might be able to improve the kinetics of DDI as the invasion products can be further stabilised due to the reduced electrostatic repulsion between negatively charged phosphates.

Phosphate modifications can be introduced into the DNA backbone using the Staudinger reaction between sulfonyl azide and 3',5'-dinucleoside β-cyanoethyl phosphites (Scheme 3.1, I), forming *N*-modified iminophosphoranes (Scheme 3.1, II).[181] Sulfonyl azides have advantages over aryl or alkyl azides because of facile hydrolysis of the corresponding aryl or alkyl phosphoramidates under acidic conditions, whereas sulfonyl phosphoramidates are stable during the oligonucleotide synthesis, particularly under acidic detritylation conditions.[181] Moreover, the use of sulfonyl azides is compatible with standard automated solid phase DNA or RNA synthesis as the modification can be introduced as a replacement of a standard iodine/pyridine oxidation step.

It has been reported that *N*-(methanesulfonyl)phosphoramidate (mesyl, or μ-modification, scheme 3.1)[182] and *N*-tosyl phosphoramidate (Ts- modification, scheme 3.1) were introduced into the DNA backbone using the Staudinger reaction during automated DNA synthesis. The introduction of single μ- or Ts- modification into DNA only resulted in marginally destabilised duplexes with complementary DNA and RNA.[181, 182]

The introduction of every single μ- or Ts- modification creates a chiral centre at the phosphorus atom resulting in a mixture of 2^n diastereomers, where n is the number of modified phosphate groups. The RP-HPLC purification only occasionally results in the separation of individual diastereomers (usually for ONs with a single modification), which supports the insignificant difference in the lipophilicity of the diastereomeric ONs.



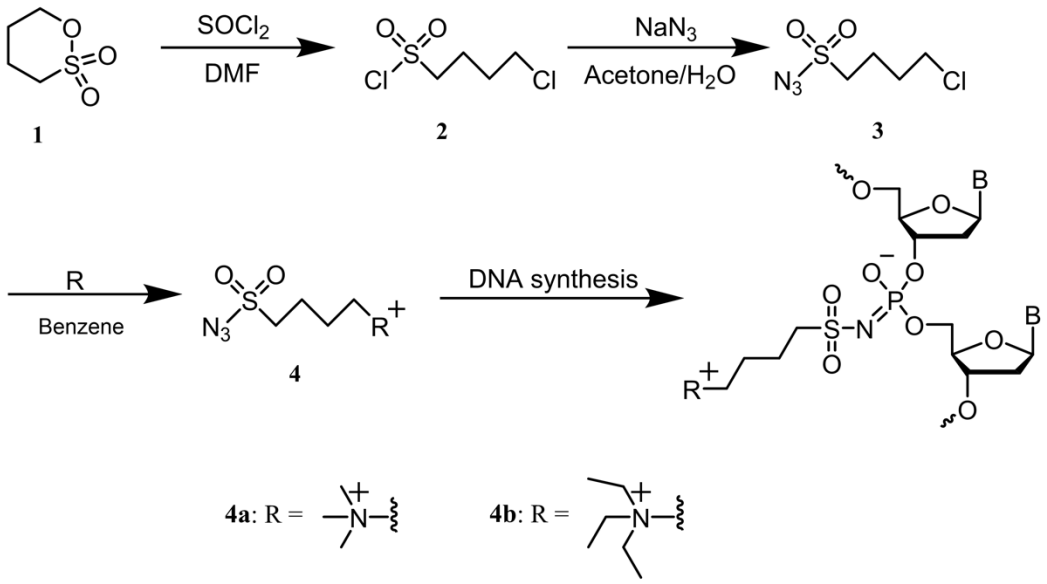
Scheme 3.1. Synthesis of ONs with μ- and Ts- modifications using Staudinger reaction during solid-phase synthesis. (*i*) 0.5 M μ-N_3 or Ts-N_3 in MeCN, 37 °C, 30 min; (*ii*) DNA synthesis; (*iii*) conc. aq. NH_3, 55°C, 12 h; B^p/B, protected/deprotected heterocyclic base.

3.3. Methodology and methods used for evaluation of N+ modified *ortho*-TINA probes in DDI

Based on the hypothesis proposed above, we planned to introduce a zwitterionic sulfonyl phosphoramidate group into DNA via the Staudinger reaction. Here, the negatively charged phosphoramidate is neutralised by the positively charged quaternary ammonium group, providing a zwitterionic phosphate mimic (Scheme 3.2).

For that purpose, we designed cationic sulfonyl azides, 4-(azidosulfonyl)-*N,N,N*-trimethylbutan-1-aminium iodide (Scheme 3.2, compound **4a**) or 4-(azidosulfonyl)-*N,N,N*-triethylbutan-1-aminium iodide (Scheme 3.2, compound **4b**) which can be obtained in three synthetic steps. As shown in scheme 3.2, the quaternary amine

(compounds **4 a** and **b**) can be obtained from the reaction between 4-chlorobutane-1-sulfonyl azide (compound **3**) and triethylamine (Et_3N) or trimethylamine (Me_3N). Compound **3** can be prepared by substitution of 4-chlorobutane-1-sulfonyl chloride (compound **2**) with NaN_3 using the standard protocol with aqueous acetone as a solvent. Compound **2** can be synthesised from commercially available 1,4-butane sultone (1,2-oxathiane 2,2-dioxide).[183]



Scheme 3.2. Retrosynthetic analysis for the preparation of a quaternary amine (**4**).

In this Chapter, we evaluated *ortho*-TINA probes bearing zwitterionic phosphate modifications in the context of DDI. As described in the previous Chapter and illustrated in Figure 3.2, we firstly performed UV-Vis/CD thermal denaturation and renaturation experiments for the confirmation and evaluation of DDI, followed by fluorescence spectroscopy for monitoring the kinetics of DDI and 20% native gel electrophoresis for direct visualisation of DDI.

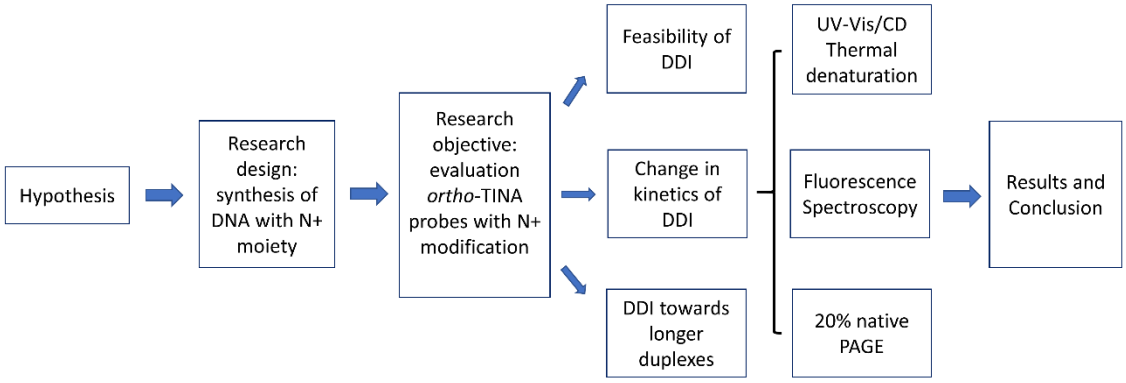


Figure 3.2. Flowchart of methodology and methods used for evaluation of DDI.

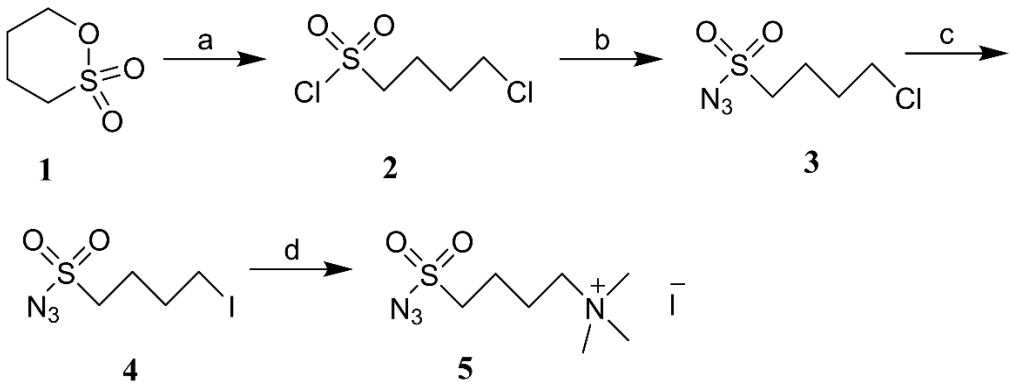
3.4. Chapter summary

We demonstrated in this Chapter that a cationic sulfonyl azide was successfully synthesised and introduced into the DNA backbone. Purification of ONs with two types of modifications (N+ and *ortho*-TINA monomers) was challenging but can be achieved using ion-exchange (IE-) HPLC. In contrast to our hypothesis, the incorporation of N+ modification into *ortho*-TINA probes brought marginal improvement in the kinetics of DDI. Invasion against long duplexes was still an obstacle.

3.5. Results

3.5.1. Synthesis of 4-(azidosulfonyl)-*N,N,N*-trimethylbutan-1-aminium iodide (5)

The quaternary amine salt was prepared as shown in Scheme 3.3. We decided to use Me_3N instead of Et_3N to make the quaternary amine salt as it was less bulky. The 4-(azidosulfonyl)-*N,N,N*-trimethylbutan-1-aminium iodide (**5**) was obtained in four steps rather than three steps as mentioned earlier in Scheme 3.2.



Scheme 3.3. Preparation of 4-(azidosulfonyl)-*N,N,N*-trimethylbutan-1-aminium iodide (5). Reagents and conditions: (a) $SOCl_2$ in DMF, 70 °C, 9 days; (b) NaN_3, Bu_4NHSO_4, $H_2O/CHCl_3$ (1: 1), 0 °C, 2 h; (c) NaI, acetone, reflux, 20 h; (d) Me_3N, diethyl ether, 3 days, RT.

According to the published protocol, ring-opening of 1,4-butane sultone was carried out using excess $SOCl_2$ (5 eq).[183] Compound **2** (68% yield) was used in the next step without further purification. Conversion of sulfonyl chloride to sulfonyl azide was initially performed using aqueous acetone, giving compound **3** with only 34% yield. Such a low yield was a result of the hydrolysis of sulfonyl chloride. Alternatively, a phase-transfer reaction was performed with two different reagents, tetra-*n*-butylammonium hydrogen sulfate ($Bu_4N^+HSO_4^-$) and tetraethylammonium bromide ($Et_4N^+Br^-$). Use of ($Bu_4N^+HSO_4^-$) resulted in a higher yield than using ($Et_4N^+Br^-$): 80% *versus* 58%, respectively. Alkyl chloride **3** was converted to a more reactive iodide **4** by a Finkelstein reaction[184] due to the fact that the Menshutkin reaction[185] between compound **3** and Me_3N (saturated in benzene, 22.3%, *v/v*) did not occur even after one week. The final step of preparation of a quaternary amine from iodoalkane **4** was conducted using a saturated solution of Me_3N in benzene or diethyl ether (10.7%, *v/v*), respectively. The use of diethyl ether led to a faster reaction and purer compound **5** than using benzene. The whole synthesis starting from 1,4-butane sultone was monitored by ^1H NMR and IR (see Fig. B.1 to B.13 in Appendix B) and carried out without any silica gel column chromatography.

3.5.2. Synthesis, purification and chemical stability evaluation of N+ modified ON

As a model study, we synthesised ON possessing one N+ modification at the 5′-end (5′-C_{N+}CCCTTTCTTTTTT, **5N′+ON2**). N+ modified ONs were synthesised using the standard β-cyanoethyl phosphoramidite approach on an automated DNA synthesiser, substituting the Staudinger reaction of the 3',5'-dinucleoside β-cyanoethyl phosphite triester (I) with sulfonyl azide (compound **5)** for standard aqueous iodine oxidation.[181] In order to get a higher coupling efficiency, DMF was used as a solvent instead of acetonitrile as the solubility of azide **5** in DMF was higher than in acetonitrile (700 *versus* 77 mM, respectively). Moreover, extended time (30-45 min) and elevated temperature (37 °C) for the Staudinger reaction of azide **5** were applied to achieve higher conversion. After synthesis, the ONs were cleaved from the solid support and deprotected using 28% NH_4OH. The obtained N+ modified ONs were initially purified using reverse-phase (RP-) HPLC. However, separation of ONs with varying number of modifications was not ideal as the changes in the retention time in RP-HPLC were only marginal. Alternatively, ion-exchange (IE-) HPLC was used for purifying these ONs. As a result of substitution of each phosphate with N+ modification, a shorter retention time (τ) in IE-HPLC was observed ($\Delta\tau$ = -2.5 min/modification, Table B.1 in Appendix B). Composition of the

ONs was confirmed by electrospray ionisation-mass spectrometry (ESI-MS) in the negative mode (Fig. 3.3A, see also Table B.1 in Appendix B).

The solubility of the N+ONs was not influenced by the introduction of modifications, as the purified, desalted, and lyophilised ONs were fully dissolved in 50 µL H_2O. The chemical stability of **5N′+ON2** was evaluated at various pHs (5.5, 7.0, and 8.5) upon incubation in 10 mM Na-phosphate buffer (140 mM NaCl, 0.1 mM Na_2-EDTA) at 50 °C for 24 h. No degradation was observed according to IE-HPLC analysis (Fig. 3.3B), which ensures that N+ONs will be chemically stable during the evaluation of thermal stability of complexes formed with complementary DNA.

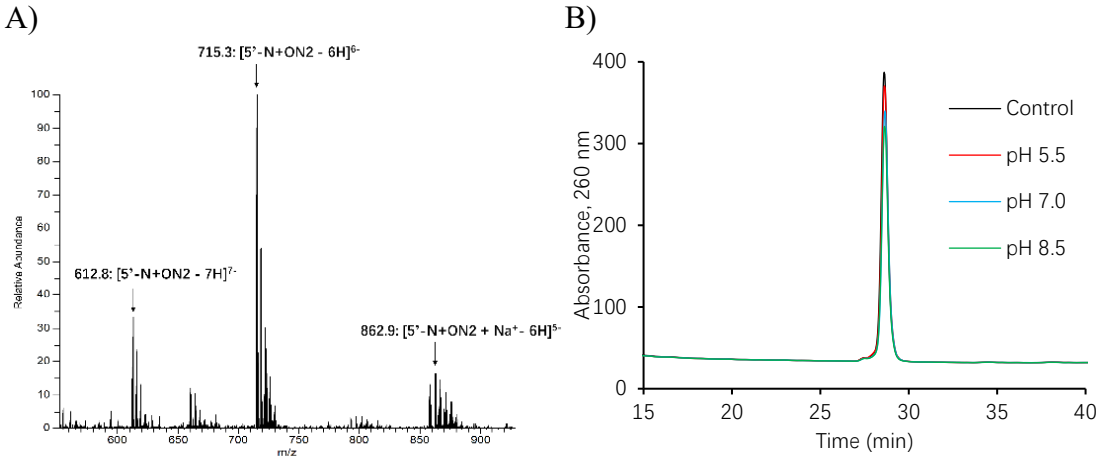


Figure 3.3. Composition and chemical stability evaluation of **5N′+ON2**. A) ESI-MS result for **5N′+ON2**; Calculated molecular weight of [**5N′+ON2**- 7H]$^{3-}$ is 612.8, observed: 612.8. B) IE-HPLC profiles of samples of **5′-N+ON2** after incubation at 55 ºC in Na-phosphate buffer (pH 5.5, 7.0 and 8.5, respectively) for 24 h. Control is a sample of **5′-N+ON2** in Na-phosphate buffer (pH 7.0) without incubation at 55 ºC.

We then synthesised modified native duplexes for Set 1 with N+ modifications (Table 3.1) for DDI.

Table 3.1. Abbreviation and sequence of N+ modified duplexes and corresponding theoretical invasion products against native duplex **C1** of Set 1

Duplex		**Sequences**	**Theoretical invasion products**	
C1		5′-TTAATAATATGATTTA 3′-AATTATTATACTAAAT		
2N+P 1	2N+ON 1 2N+ON 2	5′- T_{N+}TAATAATATGATTT$_{N+}$ A 3′- A_{N+}ATTATTATACTAAA$_{N+}$ T	**2N+IP 1**	5′- T_{N+}TAATAATATGATTT$_{N+}$ A 3′-AATTATTATACTAAAT
			2N+IP 2	5′-TTAATAATATGATTTA 3′- A_{N+}ATTATTATACTAAA$_{N+}$ T
3N+P 2	3N+ON 1 3N+ON 2	5′- T_{N+}TAATAAT$_{N+}$ATGATTT$_{N+}$A 3′- A_{N+}ATTATTA$_{N+}$TACTAAA$_{N+}$T	**3N+IP 3**	5′- T_{N+}TAATAAT$_{N+}$ATGATTT$_{N+}$A 3′-AATTATTATACTAAAT
			3N+IP 4	5′-TTAATAATATGATTTA 3′- A_{N+}ATTATTA$_{N+}$TACTAAA$_{N+}$

3.5.3. Evaluation of probes bearing N+ modification in DDI

N+ modified sequences of Set 1 were synthesised without incorporation of *ortho*-TINA monomers, and DDI for N+ modified probes was evaluated.

Table 3.2. Hybridisation data for Set 1 and their free energy for recognition, ΔG^{298}_{rec} of dsDNA target **C1** (ON1/ON2 in Chapter 2) at 298 K.

Entry	Duplex	T_m (°C)	ΔH (kJ/mol)	$T\Delta S$ (kJ/mol)	ΔG_{298} (kJ/mol)	ΔG^{298}_{rec} (kJ/mol)
1	**C1**	36	-338 (±7)	-290 (±7)	-48 (±10)	
2	**2N+P1**	36	-326 (±7)	-280 (±7)	-46 (±10)	
3	**3N+P2**	39	-317 (±13)	-266 (±13)	-50 (±18)	
4	**2N+IP1**	37	-380 (±20)	-325 (±19)	-50 (±30)	-6
5	**2N+IP2**	36	-332 (±17)	-283 (±16)	-50 (±20)	
6	**3N+IP3**	38	-335 (±16)	-284 (±16)	-50 (±20)	-5
7	**3N+IP4**	38	-350 (±20)	-297 (±19)	-50 (±30)	

T_m (°C) values (±1.5 °C) were determined by baseline fitting of melting curve (Abs_{260} from UV-Vis *versus* T) against increasing temperature (1.0 °C min^{-1}) on equimolar mixtures (1.0 μM in each strand) of ONs in 10 mM Na-phosphate buffer (140 mM NaCl, 0.1 mM Na_2-EDTA, pH 7.0). Calculation of thermodynamic parameters is described in Chapter 2.

As we can see from Table 3.2, the introduction of two N+ modifications in the duplex led to a marginal change in the thermal stability of the invading probe (entry 2) as well as invasion products (entries 4 and 5). Introduction of three N+ modifications led to slightly stabilised invasion products (entries 6 and 7), whereas the invading probe was further stabilised (entry 3) due to reduced electrostatic repulsion. Similar to *ortho*-TINA monomers, incorporation of N+ modifications also resulted in less favoured ΔH and more favourable $T\Delta S$ of the invading probe.

Duplexes with two and three N+ modifications showed slightly favourable ΔG^{298}_{rec} values (-6 and -5 kJ/mol, respectively) for DDI. This encouraged us to conduct the evaluation of DDI using 20% native PAGE. However, there was a minimal amount of invasion product detected even for the thermodynamic product of reaction (line 6, Fig. 3.4). No invasion was detected between N+ modified probes and the native duplex **C1** (line 7, Fig. 3.4). These results indicate that N+ modification does not contribute to DDI, and *ortho*-TINA monomers might improve DDI for N+ containing probes.

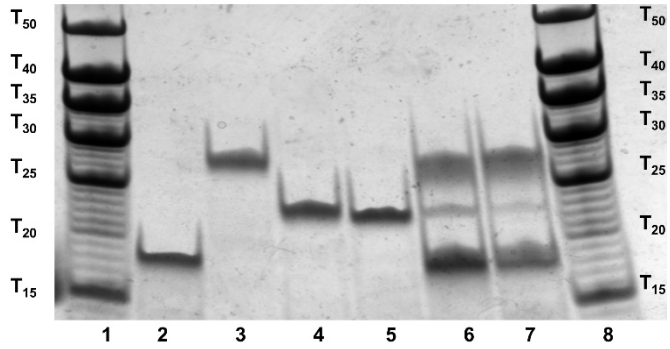


Figure 3.4. 20% Native PAGE showing difference in mobility of unmodified duplex **C1** and duplex **3N+P2**. Lanes are 1: ladder; 2: **C1**; 3: **3N+P2**; 4: **3N+IP3**; 5: **3N+IP4**; 6: mixture of **C1** and **3N+P2**, heated at 90 °C for 5 min and then slowly cooled down to RT; 7: **C1** + **3N+P2**, 24 h, 4 °C; 8: ladder. Conditions: 50 μM strand concentrations in 1 × TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM Na_2-EDTA) supplemented with 100 mM NaCl.

We then synthesised DNA probes containing both *ortho*-TINA monomers and N+ modifications. N+ modifications were incorporated in the duplexes opposite each other to reduce the thermal stability of the invading probe (Table 3.3 for duplexes in Set 1 and Table 3.4 for duplexes in Set 2). N+ modified *ortho*-TINA duplexes were tested for DDI. *ortho*-TINA phosphoramidite ((*S*)-1-*O*-(4,4′-dimethoxytriphenylmethyloxy)-3-*O*-[2-(pyren-1-ylethynyl)benzyloxy]propan-2-yl 2-cyanoethyl diisopropylphosphoramidite) were synthesised following the reported protocol (Fig. B.14 – B.18 in Appendix B).[132] *ortho*-TINA monomers were introduced through automated DNA synthesis using an increased deprotection time (100 s) as well as coupling time (2 min) for a 75 mM solution of the *ortho*-TINA phosphoramidite in a 1: 1 mixture of dry MeCN/CH_2Cl_2.[132]

Table 3.3. Abbreviation and sequence of N+ modified *ortho*-TINA duplexes and corresponding invasion products against native duplex **C1** of Set 1

Duplex		**Sequences**	**Invasion products**	
C1		5′-TTAATAATATGATTTA 3′-AATTATTATACTAAAT		
4N+P3	ON7 4N+ON8	5′-TTA**X**ATA**X**ATAT**X**GAT**X**TTA 3′-A_{N+}AT**X**TA_{N+}T**X**TATA**X**C$_{N+}$TA**X**AA_{N+}T	**IP6**	5′- TTA**X**ATA**X**ATAT**X**GAT**X**TTA 3′-AATTATTATACTAAAT
			4N+IP5	5′-TTAATAATATGATTTA 3′-A_{N+}AT**X**TA_{N+}T**X**TATA**X**C$_{N+}$TA**X**AA_{N+}T
3N+P4	ON7 3N+ON8	5′-TTA**X**ATA**X**ATAT**X**GAT**X**TTA 3′-A_{N+}AT**X**TAT**X**TATA**X**C$_{N+}$TA**X**AA_{N+}T	**IP6**	5′- TTA**X**ATA**X**ATAT**X**GAT**X**TTA 3′-AATTATTATACTAAAT
			3N+IP6	5′-TTAATAATATGATTTA 3′-A_{N+}AT**X**TAT**X**TATA**X**C$_{N+}$TA**X**AA_{N+}T

X denotes *ortho*-TINA monomer; **IP6** was the same as listed in Chapter 2.

Table 3.4. Abbreviation and sequences of N+ modified *ortho*-TINA duplexes and corresponding invasion products against native duplex **C2** of Set 2

Duplex		**Sequences**	**Invasion products**	
C2		5′-TAGTTGGTAGTGCTTT 3′-ATCAACCATCACGAAA		
4N+P5	4N+ON15	5′-T_{N+}AG**X**TT_{N+}G**X**GTA**X**GT_{N+}GC**X**TT_{N+}T	**4N+IP7**	5′- T_{N+}AG**X**TT_{N+}G**X**GTA**X**GT_{N+}GC**X**TT_{N+}T 3′-ATCAACCATCACGAAA
	4N+ON16	3′-A_{N+}TC**X**AA_{N+}C**X**CAT**X**CA_{N+}CG**X**AA_{N+}A	**4N+IP8**	5′-TAGTTGGTAGTGCTTT 3′-A_{N+}TC**X**AA_{N+}C**X**CAT**X**CA_{N+}CG**X**AA_{N+}A
3N+P6	3N+ON15	5′-T_{N+}AG**X**TTG**X**GTA**X**GT_{N+}GC**X**TT_{N+}T	**3N+IP9**	5′-T_{N+}AG**X**TTG**X**GTA**X**GT_{N+}GC**X**TT_{N+}T 3′-ATCAACCATCACGAAA
	ON16	3′-ATC**X**AAC**X**CAT**X**CACG**X**AAA	**IP11**	5′-TAGTTGGTAGTGCTTT 3′-ATC**X**AAC**X**CAT**X**CACG**X**AAA

X denotes *ortho*-TINA monomer; **IP11** was the same as listed in Chapter 2.

Table 3.5. Hybridisation data for N+ modified *ortho*-TINA probes in Set 1 and their free energy of recognition, ΔG^{298}_{rec} of dsDNA target **C1** (ON1/ON2) at 298 K.

Duplex	T_m (°C)	ΔH (kJ/mol)	TΔS (kJ/mol)	ΔG_{298} (kJ/mol)	ΔG^{298}_{rec} (kJ/mol)
C1	37	-230 (±15)	-191 (±14)	-40 (±20)	
4N+P3	45	-175 (±11)	-132 (±11)	-43 (±16)	
3N+P4	40	-170 (±10)	-121 (±9)	-49 (±13)	
IP6	44	-236 (±14)	-189 (±13)	-47 (±19)	-24
4N+IP5	45	-355 (±29)	-300 (±30)	-59 (±40)	
IP6	44	-236 (±14)	-189 (±13)	-47 (±19)	-22
3N+IP6	47	-398 (±34)	-340 (±30)	-60 (±50)	

Conditions as specified in Table 3.2.

As shown in Table 3.5, the introduction of N+ modification into one strand of *ortho*-TINA duplex led to the formation of much more stable invading probes (+8 °C for **4N+P3** and +3 °C for **3N+P4**) than the native duplex **C1**. In contrast, the stabilisation of invasion products was not as great, with ΔT_m of only +1 – +3 °C than the duplex without N+ modifications (**IP6**). As described in the last Chapter, invasion products showed more favourable ΔH, and this gain in ΔH was further enlarged by the introduction of N+ modifications, alongside the larger entropy cost. As a result, the changes in ΔG_{298} were more noticeable in comparison with duplexes without N+ modifications (-29 to -22 kJ/mol *versus* -11 kJ/mol for **P3**). This allows us to conduct 20% native gel electrophoresis to evaluate DDI using N+ modified *ortho*-TINA probes against native duplex **C1** (Fig. 3.5).

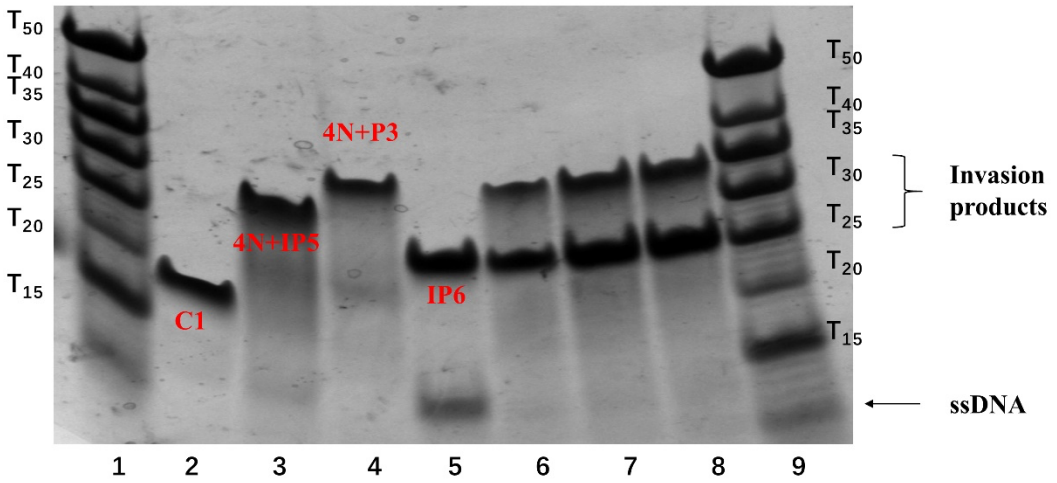


Figure 3.5. 20% Native PAGE showing strand exchange between pre-annealed **4N+P3** and **C1**. Lanes are 1: ladder; 2: **C1**; 3: **4N+IP5**; 4: **4N+P3**; 5: **IP6**; 6: mixture of **C1** and **4N+P3**, heated at 90 °C for 5 min and then slowly cooled down to RT; 7: mixture of mixture of **C1** and **4N+P3**, incubated overnight at 25 °C; 8: mixture of mixture of **C1** and **4N+P3**, incubated overnight at 37 °C; 9: ladder. Conditions: 50 μM strand concentrations in 1 × TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM Na_2-EDTA) supplemented with 100 mM NaCl.

Similar to what has been observed previously, native gel showed that incubation of equimolar quantities of invading probe **4N+P3** and dsDNA target **C1** at 25 °C and 37°C resulted in complete disappearance of bands corresponding to the starting materials after 24 h (lanes 7 and 8 *versus* lanes 2 and 4, Fig. 3.5), showing that the DDI was efficient for duplexes with less G/C content.

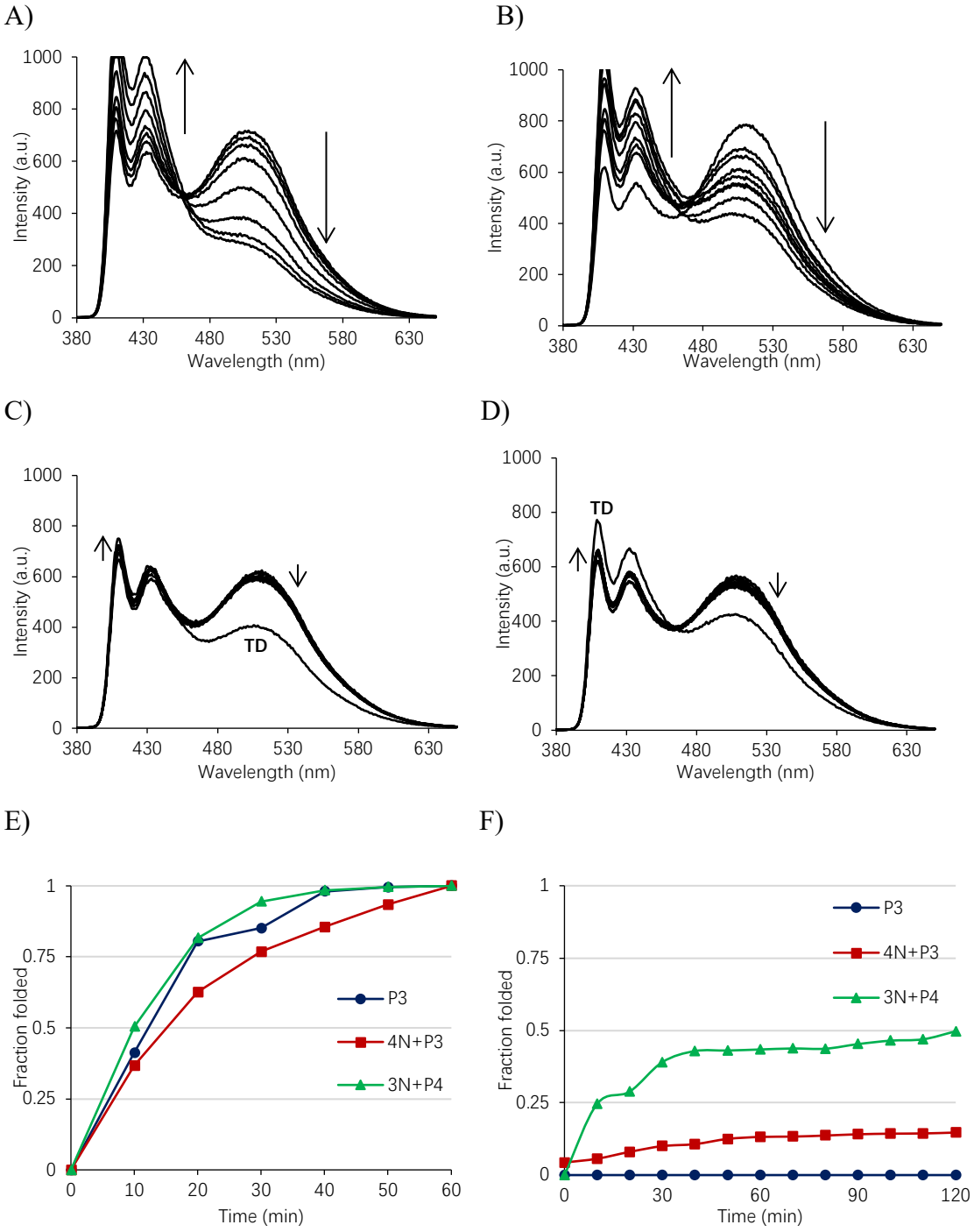


Figure 3.6. Fluorescence emission spectra upon addition of A) pre-annealed **C1** to pre-annealed invading duplex **4N+P3**; B) pre-annealed **C1** to pre-annealed invading duplex **3N+P4**; C) pre-annealed **LC1** to pre-annealed invading duplex **4N+P3**; D) pre-annealed **LC1** to pre-annealed invading duplex **3N+P4**. Time course of invasion of **4N+P3** and **3N+P4** into native duplex **C1** E), and into long duplex **LC1** F). Spectra were recorded every 4 min over 1 h (for **C1**) and 2 h (for **LC1**) using Perkin-Elmer LS55 at 25 °C in 10 mM Na-phosphate buffer (140 mM NaCl, 10 mM Na_2-EDTA, pH 7.0), without sample shaking. Concentration of each strand is 1.0 μM. TD stands for Thermodynamic products. Direction of arrows indicates the signal changes over time.

We then monitored the DDI process using fluorescence spectroscopy. The kinetics of DDI against native duplex **C1** and longer duplex **LC1** is presented in Fig 3.6 and Table 3.6. We can see that for the target duplex of the same length, using N+ modifications in one of the strands of *ortho*-TINA duplexes (**3N+P4**) slightly increased the kinetics of DDI ($t_{50\%}$ = 10 min *versus* 13 min for **P3**, Fig. 3.6E). When targeting a longer duplex, the control probe showed no invasion after 120 min, whereas the probes with the same number of N+ modifications and *ortho*-TINA monomers (four each, **4N+P3**) exhibited marginal decay in the exited dimer region, indicating the initiation of DDI (Fig. 3.6C). In contrast, **3N+P4** with one less N+ modification showed great improvement of DDI kinetics against the long duplex, **LC1**, with $t_{50\%}$ = 120 min (Fig. 3.6F and Table 3.6).

Table 3.6. Kinetic parameters ($t_{50\%}$ and $t_{75\%}$ [min, ± 0.5 min]) for recognition of C1 and LC1 by probes possessing N+ and *ortho*-TINA modifications.

Target duplex	Probes	$t_{50\%}$ (min)	$t_{75\%}$ (min)
C1	**P3**	13	22
	4N+P3	15	30
	3N+P4	10	20
LC1	**P3**	-[a]	-
	4N+P3	-	-
	3N+P4	120	-

[a] not reached

Table 3.7. Hybridisation data for N+ modified *ortho*-TINA probes in Set 2 and their free energy of recognition, ΔG^{298}_{rec} of dsDNA target **C2** (ON9/ON10 in Chapter 2) at 298 K.

Duplex	T_m **(°C)**	ΔH **(kJ/mol)**	$T\Delta S$ **(kJ/mol)**	ΔG_{298} **(kJ/mol)**	ΔG^{298}_{rec} **(kJ/mol)**
C2	55	-199 (±12)	-147 (±11)	-52 (±16)	
4N+P5	57	-182 (±7)	-127 (±6)	-54 (±9)	
3N+P6	54	-209 (±14)	-153 (±12)	-56 (±18)	
4N+IP7	64	-329 (±11)	-257 (±10)	-72 (±15)	-40
4N+IP8	63	-350 (±30)	-276 (±20)	-74 (±40)	
3N+IP9	65	-209 (±8)	-156 (±8)	-53 (±11)	-5
IP11	60	-229 (±7)	-172 (±7)	-57 (±10)	

Conditions were the same as Table 3.5.

The same evaluation was also performed for duplexes designed for Set 2. As shown in Table 3.7, the introduction of N+ modifications into both strands of invading probes caused a dramatic decrease of the ΔG^{298}_{rec} value (-40 kJ/mol), whereas the introduction of N+ into one strand of the duplex with six G/C bps led to less favourable ΔG^{298}_{rec} in comparison with **P6** (-5 kJ/mol *versus* -29 kJ/mol). We then performed an evaluation of the probe with the most promising ΔG^{298}_{rec} value, **4N+P5** against native duplex **C2** in native gel electrophoresis (Fig. 3.7). Surprisingly, even after incubation of the probe **4N+P5** and native duplex **C2** overnight at 37 °C, the invasion product was barely detectable on the native gel (lane 6, Fig. 3.7).

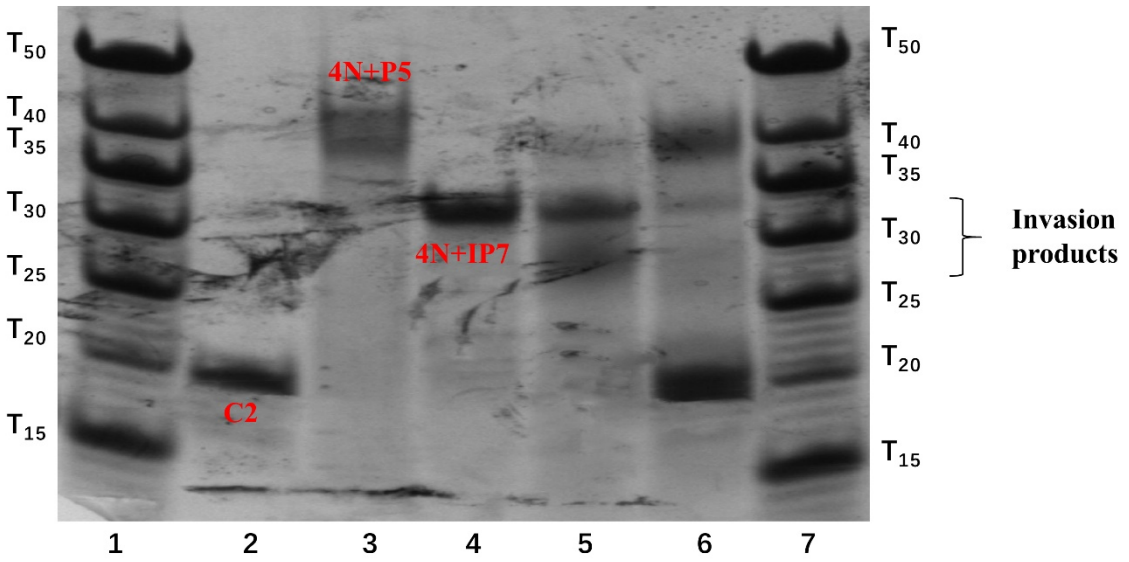


Figure 3.7. 20% Native PAGE showing strand exchange between pre-annealed **4N+P5** and **C2**. Lanes are 1: ladder; 2: **C2**; 3: **4N+P5**; 4: **4N+IP7**; 5: mixture of **C2** and **4N+P5**, heated at 90 °C for 5 min and then slowly cooled down to RT; 6: mixture of mixture of **C1** and **4N+P3**, incubated overnight at 37 °C; 8: ladder. Conditions: 50 μM strand concentrations in 1 × TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM Na_2-EDTA) supplemented with 100 mM NaCl.

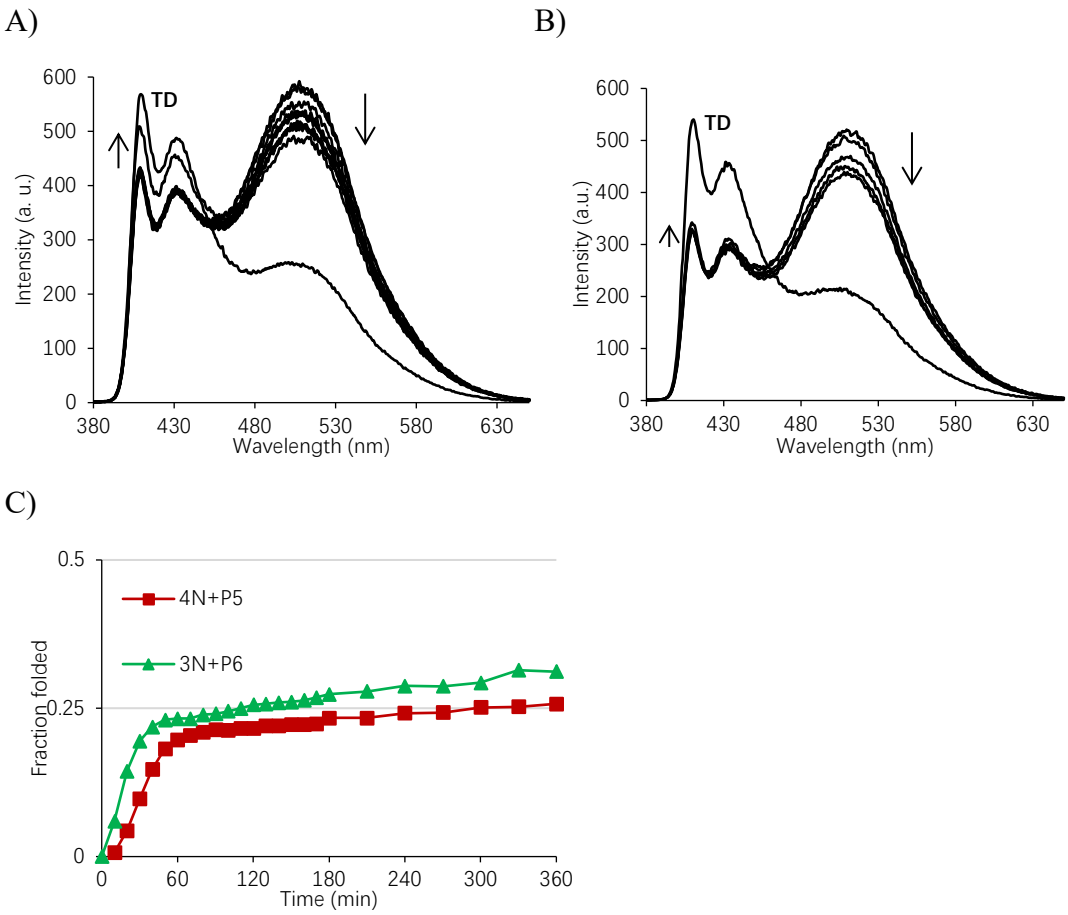


Figure 3.8. Fluorescence emission spectra upon addition of pre-annealed native duplex **C2** to pre-annealed duplexes A) **4N+P5** and B) **3N+P6** over 6 h. C) Time course of A) and B). Spectra were recorded every 30 min using Perkin-Elmer LS55 at 37 °C, without sample shaking. Concentration of each strand is 1.0 μM at 25 °C in 10 mM Na-phosphate buffer (140 mM NaCl, 10 mM Na_2-EDTA pH 7.0). TD stands for Thermodynamic products. Direction of arrows indicates the changes in signal changes from 0 to 6 h.

Monitoring of DDI using fluorescence spectroscopy for duplexes in Set 2 showed that incorporation of four N+ modification pairs into TINA-TINA duplexes played no role in improving the kinetics of invasion. DDI is probably hindered by the increased thermal stability of one of the invasion probes (57 °C for **4N+P5** *versus* 50 °C for **P6**). Only 25% of the invasion was accomplished using N+ modified probes over 6 h (Fig. 3.8C), whereas $t_{50\%}$ = 1 hr for **P6** with no N+ modifications.

3.6. Discussion

Chemical modification provides an excellent and efficient way of designing probes for targeting dsDNA based on the nucleic acid scaffold. In order to be compatible with standard solid-phase DNA or RNA synthesis, the introduction of SaRNA monomers into the RNA backbone involved a 14-step-preparation of phosphoramidite SaRNA-TT dinucleotide.[179] Similarly, incorporation of BCNS groups into the DNA backbone also required the synthesis of thymidine and 2'-OMe-uridine phosphoramidites bearing BCNS groups with nine and ten synthetic steps, respectively. [180] In comparison, the synthesis of the N+ monomer requires only four steps of synthesis starting from commercially available 1,4-butane sultone. Our synthesis does not require any silica gel purification. Incorporation of N+ modification into DNA phosphate is performed during automated DNA synthesis. From a synthetic perspective, it is advantageous that N+ modification can be introduced into any position in the sequence, which is not the case for SaRNA and BCNs nucleic acid analogues.

Unlike PNAs and some BCN groups, N+ modified ONs demonstrated excellent chemical stability and solubility in buffer solutions. Incorporation of N+ modifications onto TINA-TINA duplexes led to some changes in the property of the duplexes, such as increased thermal stability and preferable enthalpy of formation.

For A/T rich duplexes, the incorporation of N+ modification into one strand of the probe slightly improved the kinetics of DDI towards the native duplex with the same length, and it initiated the DDI process against long duplex, which was not the case for *ortho*-TINA probes without N+ modification. The duplex with four *ortho*-TINA pairs having three N+ modifications in one strand led to improved invasion kinetics towards longer duplex, with $t_{50\%}$ around 120 min.

For the duplexes with high G/C content, no obvious improvement of invasion kinetics was observed. In contrast, incorporation of N+ modifications in both strands resulted in even slower kinetics, despite the fact that preferable ΔG values were obtained.

3.7. Conclusion

We described in this Chapter the design and synthesis of a chemically modified ON with zwitterionic N+ group as a phosphate mimic. ONs possessing N+ modifications showed good aqueous solubility and chemical stability, which allowed us to conduct further assessments. We observed changes in the thermal stability of duplexes induced by incorporation of N+ modification into the duplex alone and in combination with *ortho*-TINA monomers.

The synthesis of N+ group and its introduction into DNA was efficient, yet the overall yield of N+ modified *ortho*-TINA probes decreased with the increased number of modifications used due to the insufficient coupling efficiency at each modification.

As we expected, the presence of N+ modifications on the internucleotidic phosphates enhanced the binding affinity of modified ONs towards complementary ssDNA because of the reduced electrostatic repulsion. However, N+ modification did not improve the kinetics of DDI. The application of this type of zwitterionic N+ modification will be further evaluated.

Next, we decided to further investigate the properties of N+ containing nucleic acid assemblies.

Chapter 4. DNA with zwitterionic and negatively charged phosphate modifications: Formation of DNA triplexes, duplexes and cell uptake studies

Content in this Chapter was published in Beilstein J. Org. Chem. (2021), *17*, 749–761.

Su, Y., Bayarjargal, M., Hale, T. K., & Filichev, V. V. DNA with zwitterionic and negatively charged phosphate modifications: Formation of DNA triplexes, duplexes and cell uptake studies.

In this article, I synthesised N+ and Ts- modified ONs, evaluated properties of parallel triplexes, antiparallel duplexes with DNA and RNA and performed an enzymatic digestion assay. Cell uptake studies were conducted by Dr M. Bayarjargal and Dr Tracy K. Hale (Both School of Fundamental Sciences, Massey University, Palmerston North).

4.1. Introduction

In the previous Chapter, we described the synthesis of ONs with phosphates replaced by 4-(trimethylammonio)butylsulfonyl phosphoramidate group (N+ modification, Fig. 4.1) that was designed to obtain a formally charge-neutral zwitterionic N+ON sequence. In this Chapter, the properties of DNA triplexes and duplexes having one to four N+ modifications were evaluated. For comparison, we also synthesised ONs modified with tosyl sulfonyl phosphoramidate group (Ts- modification, Fig. 4.1) that results in negatively charged phosphate mimic using tosyl azide in the same protocol for the Staudinger reaction during automated DNA synthesis, as described in the previous Chapter.[181]

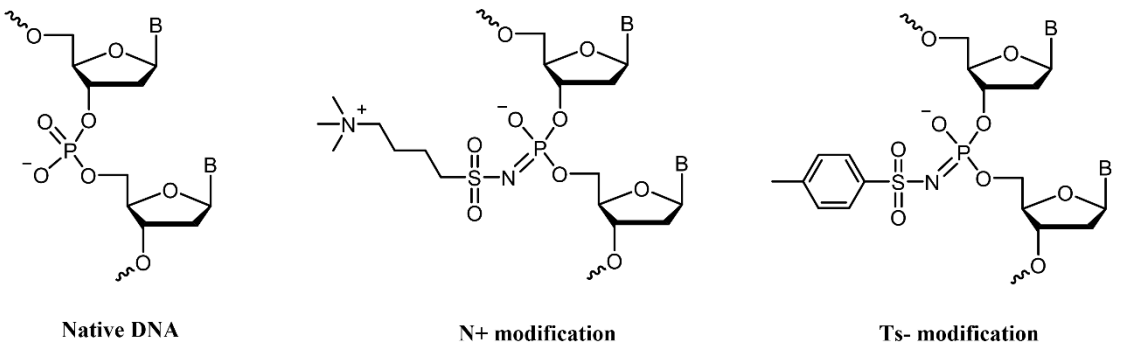


Figure 4.1. Native and chemically modified phosphate linkages in DNA. The N+ group is zwitterionic and the Ts- modification is negatively charged.

4.2. Hypothesis

We hypothesised that the N+ONs should hybridise with complementary ssDNA or RNA with higher affinity compared to native ONs because of less charge repulsion between negatively charged phosphates. We proposed that the thermal stability of duplexes bearing zwitterionic N+ modifications should be less dependent on the salt concentration. Moreover, zwitterionic phosphate modifications could lead to increased enzymatic resistance as well as cell uptake of ONs.

4.3. Methodology and methods used for evaluation of G4 formation

Based on the assumption proposed above, we synthesised 14-mer ONs with one, two, three, and four N+ and Ts- modifications introduced in various positions in the sequence for evaluation of parallel DNA triplexes and duplexes with DNA and RNA counterparts (Table 4.1). Exonuclease stability and cell uptake of N+ and Ts- modified ONs were also the subject of evaluation in this Chapter.

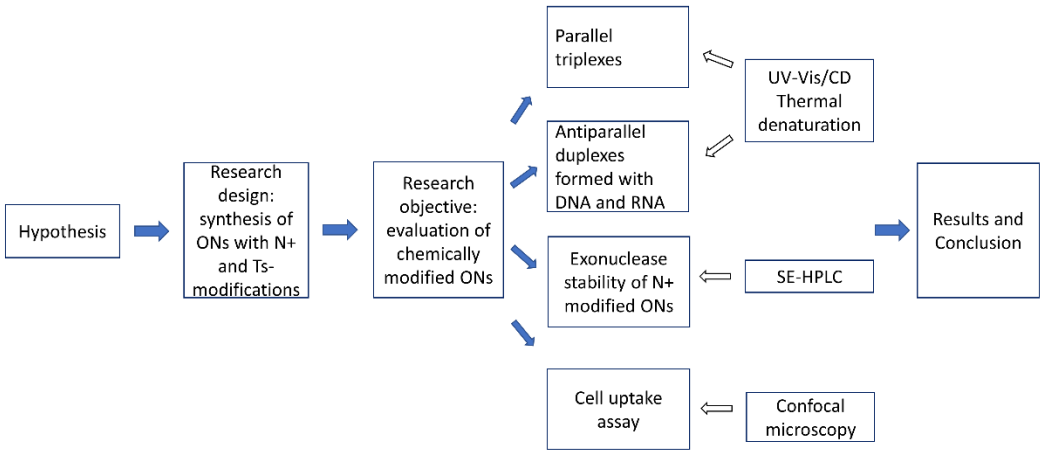


Figure 4.2. Flowchart of methodology and methods used for evaluation of chemically modified ONs.

As shown in Figure 4.2, UV-Vis/CD thermal denaturation and renaturation experiments were also involved in the determination of thermal stability as well as for the calculation of thermodynamic parameters of triplexes and duplexes. The resistance of chemically modified ONs towards enzymatic digestion by snake venom phosphodiesterase was evaluated using SE-HPLC, which separates intact oligos from digested products. Cell permeability assays were performed in mouse fibroblast cells using fluorescein labelled ONs. After fixation of cells, evaluation of the distribution of the ONs inside the cells was conducted using confocal microscopy.

Table 4.1. Abbreviation and sequence of the N+ and Ts- modified ONs.

Name	Sequence
ON1	5'-CCCCTTTCTTTTTT
5′-N+ON2	5'-C$_{N+}$CCCTTTCTTTTTT
m-N+ON3	5'-CCCCTTT$_{N+}$CTTTTTT
3′-N+ON4	5'-CCCCTTTCTTTTT$_{N+}$T
2N+ON5	5'-C$_{N+}$CCCTTTCTTTTT$_{N+}$T
3N+ON6	5'-C$_{N+}$CCCTTT$_{N+}$CTTTTT$_{N+}$T
4N+ON7	5'-C$_{N+}$CCCT$_{N+}$TTCT$_{N+}$TTTT$_{N+}$T
4N+{FAM}	5'-C$_{N+}$CCCT$_{N+}$TTCT$_{N+}$TTTT$_{N+}$T{FAM}
5′-Ts-ON8	5'-C$_{Ts}$CCCTTTCTTTTT
m-Ts-ON9	5'-CCCCTTT$_{Ts}$CTTTTTT
3′-Ts-ON10	5'-CCCCTTTCTTTTT$_{Ts}$T
2Ts-ON11	5'-C$_{Ts}$CCCTTTCTTTTT$_{Ts}$T
3Ts-ON12	5'-C$_{Ts}$CCCTTT$_{Ts}$CTTTTT$_{Ts}$T
4Ts-ON13	5'-C$_{Ts}$CCCT$_{Ts}$TTCT$_{Ts}$TTTT$_{Ts}$T
4Ts-{FAM}	5'-C$_{Ts}$CCCT$_{Ts}$TTCT$_{Ts}$TTTT$_{Ts}$T{FAM}
m-N+ON14	5'-CCCCTTTCTTT$_{N+}$TTT
m-N+ON15	5'-CCCCTTT C$_{N+}$TTTTT
m-N+ON16	5'-CCCC$_{N+}$TTTCTTTTT
m-N+ON17	5'-CC$_{N+}$CCTTTCTTTTT
3N+ON18	5'-C$_{N+}$CCCTT$_{N+}$TCTTTTT$_{N+}$T

The prefix 5′- or 3′- N+ or Ts- means the first phosphate at 5′- or 3′-end was modified; m- N+ or Ts- means modification was incorporated in the middle of the sequence; 2N+ – 4N+ or 2Ts- – 4Ts- means 2– 4 modifications were distributed evenly in the sequence.

4.4. Chapter summary

We demonstrated in this chapter that N+ and Ts- modified ONs formed thermally stabilised parallel DNA triplexes and stabilised antiparallel duplexes with complementary RNA, regardless of the number of modifications incorporated. The incorporation of N+ modifications led to the formation of DNA duplexes with thermal stability that was less dependent on the ionic strength of the solution. Enhanced resistance against enzymatic digestion and improved cell uptake of N+ and Ts- ONs were observed.

4.5. Results

4.5.1. Synthesis and purification of modified ONs

4-(Azidosulfonyl)-*N,N,N*-trimethylbutan-1-aminium iodide was synthesised as described in the previous Chapter. Tosyl azide (*p*-toluenesulfonyl azide, TsN_3) was synthesised following the previously reported protocol.[181] One should note that TsN_3 should be handled with caution as it has the shock sensitivity of tetryl (*N*-methyl-*N*-2,4,6-tetranitroaniline) and the explosiveness of TNT[186]. N+ and Ts- modified ONs were synthesised and purified using IE-HPLC following the description in the previous Chapter. Introduction of N+ modifications resulted in a shorter retention time (τ) in IE-HPLC ($\Delta\tau$ = -3 min/modification, Table C.1 in Appendix C), whereas incorporation of Ts- modifications led to increased retention time ($\Delta\tau$ = +1.5 – 2 min/modification, Table C.1 in Appendix C) of Ts-ONs in comparison with the native sequence as a result of increased hydrophobicity. The composition of the ONs was confirmed by ESI-MS in the negative mode (Table C.1 and Figures C.2 – C.5 in Appendix C).

4.5.2. Thermal stability evaluation

Thermal stability of antiparallel ON/RNA and ON/DNA duplexes, and parallel DNA triplexes were evaluated using UV-Vis thermal denaturation experiments, and the results are summarised in Table 4.2.

Table 4.2. T_m (°C, ± 0.5 °C) data for triplex and duplex melting, taken from UV-Vis melting curves (λ = 260 nm).

Entry		Antiparallel duplex			Triplex[c]	
		RNA[a]	DNA[b]			
		pH 7.0	pH 5.0	pH 7.0	pH 5.0[d]	pH 6.0
1	**ON1**	46	48	50	45	28
2	**5′-N+ON2**	53 (+ 7.0)	44 (- 4.0)	51 (+ 1.0)	40 (- 5.0)	25 (- 3.0)
3	**m-N+ON3**	47 (+ 1.0)	43 (- 5.0)	48 (- 2.0)	55 (+ 10.0)	28
4	**3′-N+ON4**	58 (+ 12.0)	46 (- 2.0)	52 (+ 2.0)	56 (+ 11.0)	29 (+ 1.0)
5	**2N+ON5**	53 (+ 7.0)	44 (- 4.0)	52 (+ 2.0)	56 (+ 11.0)	28
6	**3N+ON6**	41 (- 5.0)	45 (- 3.0)	51 (+ 1.0)	48 (+ 3.0)	< 15
7	**4N+ON7**	55 (+ 9.0)	48 (0.0)	51 (+ 1.0)	48 (+ 3.0)	28
8	**5′-Ts-ON8**	54 (+ 8.0)	39 (- 9.0)	46 (- 4.0)	51 (+ 6.0)	24 (- 4.0)
9	**m-Ts-ON9**	44 (- 2.0)	31 (- 17.0)	37 (- 13.0)	51 (+ 6.0)	< 15
10	**3′-Ts-ON10**	57 (+ 11.0)	44 (- 4.0)	49 (- 1.0)	54 (+ 9.0)	27 (- 1.0)
11	**2Ts-ON11**	56 (+ 10.0)	43 (- 5.0)	51 (+ 1.0)	53 (+ 8.0)	25 (- 2.0)
12	**3Ts-ON12**	38 (- 8.0)	40 (- 8.0)	45 (- 5.0)	49 (+ 4.0)	< 15
13	**4Ts-ON13**	53 (+ 7.0)	39 (- 9.0)	44 (- 6.0)	47 (+ 2.0)	20 (- 8.0)
14	**m-N+ON14**	54 (+ 8.0)	-[f]	51 (+ 1.0)	-	-
15	**m-N+ON15**	56 (+ 10.0)	-	51 (+ 1.0)	-	-
16	**m-N+ON16**	54 (+ 8.0)	-	55 (+ 5.0)	-	-
17	**m-N+ON17**	56 (+ 10.0)	-	52 (+ 2.0)	-	-
18	**3N+ON18**	54 (+ 8.0)	-	51 (+ 1.0)	-	-

[a] RNA sequence for antiparallel duplex formation is **ON19**: 3′-rGGGGAAAGAAAAAA; C = 1.0 μM of each strand in 20 mM sodium cacodylate buffer (100 mM NaCl, 10 mM $MgCl_2$, pH 7.0). T_m values for ON/RNA duplexes were confirmed by CD melting experiments (Figures C.7 and C.8, Table C.2 in Appendix C).

[b] DNA sequence for antiparallel duplex formation is **ON20**: 3′-GGGGAAAGAAAAAA; C = 1.0 μM of each strand in 20 mM sodium cacodylate buffer (100 mM NaCl, 10 mM $MgCl_2$, pH 5.0 and 7.0).

[c] C = 1.5 μM of **ON1– 13** and 1.0 μM of each strand of dsDNA (**D1**:3'-CTGCCCCTTTCTTTTTT/ 5′-GACGGGGAAGAAAAAA) in 20 mM sodium cacodylate bufer (100 mM NaCl, 10 mM $MgCl_2$, pH 5.0, 6.0 and 7.0); duplex T_m = 56.5 °C (pH 5.0), 58.5 °C (pH 6.0), and 57.0 °C (pH 7.0). Triplex formation was confirmed by SE-HPLC in sodium cacodylate buffer (pH 5.0 and pH 6.0, Fig. C.15 in Appendix C), no triplex formation was observed at pH 7.0.

[d] T_m for triplex melting was determined by subtraction of duplex melting curve from overlaid melting curve (Fig. C.6 in Appendix C).

[f] not performed

Sequences possessing various numbers of N+ and Ts- modifications were studied initially in an antiparallel duplex formed with complementary RNA and compared with corresponding antiparallel DNA duplexes at pH 7.0. Except for ONs with modification in the middle of the sequence (entries 6, 9 and 12 in Table 4.2), considerably stabilised ON/RNA duplexes were obtained for both N+ONs (ΔT_m = +1 – +12 °C, entries 2 – 7, Table 4.2) and Ts-ONs (ΔT_m = +7 – +11 °C, entries 8– 13, Table 4.2). The highest thermal stabilisation against RNA induced by a single modification at the 3′-end was observed (ΔT_m = +12 °C and +11 °C for N+ and Ts- modifications, respectively). Corresponding antiparallel DNA duplexes were less thermally stabilised (ΔT_m = -1 – +2 °C). The same trend was seen for ONs having two modifications at both 5′- and 3′-ends. ON/RNA duplexes were more stabilised (ΔT_m = +7 °C for **2N+ON5** and ΔT_m = +10 °C for **2Ts-ON11**) than corresponding ON/DNA duplexes (ΔT_m = +2 °C for **2N+ON5** and ΔT_m = +1 °C for **2Ts-ON11**). Thermal stability of ON/RNA duplexes was not further improved with the increased number of N+ or Ts- modifications in ONs.

Both N+ and Ts- modifications destabilised Watson/Crick type DNA duplex at pH 5.0. The destabilisation was more noticeable for Ts- than N+ modification (ΔT_m = -17 – -4 °C for Ts- and -5 – 0 °C for N+ modifications, respectively). For antiparallel ON/DNA duplexes formed at pH 7.0, N+ modifications led to higher T_m values than corresponding Ts- ONs. Incorporation of three and four Ts- modifications resulted in further decreased T_m values, whereas corresponding N+ONs did not interrupt duplex thermal stability.

These results indicated that both N+ and Ts- modifications could be used for targeting RNA in cells, because their use in ONs led to higher ΔT_m values for ON/RNA than for ON/DNA duplexes at pH 7.0.

The same sequences (**ON1** – **13**) were studied in a pH-dependent Hoogsteen-type parallel triplex[187] towards the duplex **D1**. As shown in Table 4.2, all parallel triplexes studied at pH 5.0 were more thermally stable than at pH 6.0, and no triplex formation was observed at pH 7.0, which is consistent with the trend for parallel triplexes based on CT-TFOs.[16] Some fluctuations were observed for Hoogsteen-type triplexes formed by N+ONs. The N+ modification at 5′- end led to destabilised triplex at both pH 5.0 and 6.0 (ΔT_m = -5 °C and -3°C, respectively, Table 4.2). All other N+ONs formed more thermally stable triplexes with **D1** at pH 5.0, whereas marginal changes were observed at pH 6.0 except for **3N+ON6** with three modifications that did not form a triplex. Introduction of Ts- moieties resulted in stabilised parallel triplexes at pH 5.0 (ΔT_m = +2 – +9 °C, Table 4.2),

whereas same triplexes at pH 6.0 were destabilised (ΔT_m = -1 – -8 °C). For Ts-ONs possessing modification right in the middle of the sequence (**m-Ts-ON9** and **3Ts-ON12**), no triplex was formed at pH 6.0 at RT.

These results showed that Hoogsteen-type triplexes with a single modification at the 3′-end are more thermally stable than that at the 5′-end. Increased number of N+ or Ts-modifications in TFOs showed no benefit for T_m of parallel triplexes.

The fact that ONs with a single modification in the middle of the sequence (in TC motif) formed less thermally stable duplexes than native ON suggests a position-dependent influence of N+ and Ts- moieties on T_m values. We synthesised another set of N+ONs, with single (**ON14** – **17**, entries 14 –17, Table 4.2) and triple modifications (**ON18**, entry 18, Table 4.2) that had no modification right in the middle of the sequence. The thermal stability of antiparallel duplexes formed with complementary RNA and DNA at pH 7.0 was evaluated (Table 4.2). Results showed that **ON14** – **ON17** stabilised both RNA (ΔT_m = +8 – +10 °C) and DNA (ΔT_m = +1 – +5 °C) duplexes. It is interesting to see that sequences having N+ modification in the CT motif (**m-N+ON15** and **m-N+ON16**) did not cause destabilisation of antiparallel duplexes as for N+ modification in the TC motif (**m-N+ON3** and **3N+ON6**). One possible explanation for this position-dependent influence might be a result of a propeller twist[188] in the TC dinucleotides that interfere with N+ and Ts- moieties that cause destabilisation of DNA and RNA duplexes.

Table 4.3. T_m [°C] data for antiparallel duplex melting at different NaCl concentrations, taken from UV-Vis melting curves (λ = 260 nm)[a]

Type of complex	Antiparallel Duplex	NaCl (mM)	T_m (°C) [b]	ΔH [c] (kJ/mol)	TΔS (kJ/mol)	ΔG_{298} (kJ/mol)
ON/DNA	ON1/ON20	**25**	19	-430 (±20)	-400 (±20)	-30 (±30)
		50	37	-400 (±9)	-350 (±8)	-50 (±12)
		100	50	-368 (±8)	-305 (±7)	-63 (±10)
	4N+ON7/ON20	**25**	41 (+ 22.0)[d]	-388 (±9)	-334 (±8)	-54 (±12)
		50	44 (+ 7.0)	-322 (±15)	-266 (±14)	-56 (±20)
		100	51 (+ 1.0)	-320 (±8)	-260 (±7)	-60 (±10)
	4Ts-ON13/ON20	**25**	29 (+ 10.0)	-382 (±14)	-340 (±14)	-42 (±19)
		50	34 (- 3.0)	-372 (±12)	-326 (±11)	-46 (±16)
		100	45 (- 5.0)	-354 (±7)	-297 (±7)	-57 (±10)
ON/RNA	ON1/ON19	**25**	44	-390 (±10)	-332 (±9)	-58 (±20)
		100	46	-306 (±17)	-248 (±17)	-58 (±24)
	4N+ON7/ON19	**25**	45 (+ 1.0)	-419 (±7)	-359 (±6)	-60 (±9)
		100	55 (+ 9.0)	-329 (±13)	-264 (±12)	-66 (±17)
	4Ts-ON13/ON19	**25**	38 (- 6.0)	-420 (±20)	-370 (±20)	-50 (±30)
		100	53 (+ 7.0)	-407 (±13)	-337 (±12)	-70 (±17)

[a] 1μM of each strand in 20 mM sodium cacodylate buffer (pH 7.0, supplemented with 25 mM, 50 mM or 100 mM NaCl, respectively).

[b] T_m values are reported with ± 0.5 °C uncertainties as determined from at least two repeated experiments.

[c] Thermodynamic parameters are calculated as described in Chapter 2.

[d] Values in parentheses are ΔT_m values calculated as T_m (sample) - T_m (unmodified duplex) at the same salt concentration.

Next, we evaluated the binding affinity of N+ and Ts- modified ONs toward complementary ssDNA and RNA at various ionic strengths (25, 50, and 100 mM NaCl, Table 4.3). It has been stated that the thermal stability of DNA duplexes decreases at a reduced salt concentration due to increased electrostatic repulsion between negatively charged phosphates.[189] Native DNA duplex **ON1**/**ON20** showed a decrease in T_m values from 50 to 37 to 19 °C when the NaCl concentration was reduced from 100 to 50 and to 25 mM, respectively (Table 4.3). Similar behaviour was observed for Ts-ONs as the backbone of Ts-ONs is still entirely negatively charged. In contrast, N+ONs showed resistance to changes of salt concentration: T_m decreased with the reduction of NaCl concentration but not as significant as for negatively charged ONs. The T_m value for **4N+ON7/ON20** at 25 mM NaCl was 20 °C higher than that for **ON1**/**ON20** and 12 °C higher than for **4Ts-ON13/ON20**. However, such behaviour was not as pronounced for duplexes formed with complementary RNA. The T_m value of the control ON/RNA duplex **ON1/ON19** decreased by 2 °C with the salt concentration decreased from 100 mM to 25 mM, whereas the T_m value of **4N+ON7**/**ON19** decreased by 10 °C, but it was still more thermally stable than the control duplex. In contrast, duplex formed by Ts-ON and RNA (**4Ts-ON13**/**ON19)** was destabilised (ΔT_m = -6 °C) at the lowest salt concentration tested. We analysed the melting profiles and obtained thermodynamic parameters of duplexes at different salt concentrations. As the thermal stability of DNA duplex increased with the increment of salt concentration, we expected a favourable ΔH of duplex formation at higher concentration of NaCl. However, in all cases studied, ΔH is less favourable at 100 mM than at 25 mM NaCl. Similar conclusions have been reported for DNA modified with methyl phosphotriester linkage (POMe) using results obtained from isothermal titration calorimetry (ITC) measurements.[177]

For the native DNA duplex **ON1/ON20**, the more favourable ΔH at 25 mM NaCl was compensated by an even higher entropy penalty (larger reduction), which led to a loss in ΔG ($\Delta\Delta G_{298}$ = 33 kJ/mol), therefore resulted in lower T_m value at low salt concentration. For the native RNA duplex **ON1/ON19**, $\Delta\Delta H$ between 25 mM and 100 mM NaCl is -84 kJ/mol, and the corresponding entopic factor $\Delta(T\Delta S)$ is -84 kJ/mol. As a result, $\Delta\Delta G_{298}$ is not changed, as reflected in T_m values that a decrease by 2 °C was observed when the salt concentration reduced from 100 mM to 25 mM.

When N+ or Ts- modification was introduced, ΔH for ON/DNA duplexes is less favourable at the same salt concentration than that for the native duplex, whereas $T\Delta S$ is more favourable. According to Kuo. et. al.,[177] the increase of salt concentration stabilises

the DNA duplexes mainly by reducing the entropy costs of duplex formation rather than by lowering the strand charge repulsion, and the reduction of entropy costs of duplex formation is in the endothermic release of DNA-hydrating water molecules. As the introduction of N+ modification led to less charged ON backbone, the entropy cost for N+ON/DNA is less than that for the native duplex ($\Delta(T\Delta S)$ = -74 kJ/mol for N+ON/DNA and -95 kJ/mol for the native DNA duplex, respectively). For Ts- modified ONs, the increased hydrophobicity induced by the introduction of tosyl groups led to fewer water molecules involved in the formation of H-bonds and polar interactions with dsDNA, hence the reduction of entropy costs is mainly in reducing strand charge repulsion ($\Delta(T\Delta S)$ = -43 kJ/mol).

For duplexes formed by N+ and Ts- ONs with complementary RNA, both ΔH and $T\Delta S$ values were more negative at the same salt concentration than that for native ON/RNA duplex, which is opposite for duplexes formed with DNA. This led to more significant enthalpy-entropy compensation at the medium salt concentration (100 mM). When the salt concentration decreased from 100 mM to 25 mM, despite the reduction of entropy costs for Ts-ON/RNA duplex is minimal ($\Delta(T\Delta S)$ = -33 kJ/mol), the loss in ΔG is larger than that for native and N+ON/RNA duplexes ($\Delta(\Delta G_{298})$ =0, 6, and 20 kJ/mol for native ON/RNA, N+ON/RNA, and Ts-ON/RNA, respectively). This resulted in the destabilised Ts-ON/RNA duplex at low salt concentration.

4.5.3. Evaluation of nuclease stability of N+ and Ts- modified ONs

Nuclease stability of modified ONs was evaluated using snake venom phosphodiesterase (phosphodiesterase I, Sigma) and compared with degradation of unmodified sequence **ON1**. Snake venom phosphodiesterase is commonly used for the evaluation of the nuclease stability of modified ONs in the field of research. As an isolated enzyme, it is more powerful in digesting ONs than the action of nucleases *in vivo*. The advantage of using it is that if the modified ONs is stable in the snake venom phosphodiesterase, it will also be stable in biological media.

Under the conditions used in this experiment, **ON1** was completely degraded within 30 min at RT (Fig. 4.3E, see also Fig. C.17 in Appendix C). Both N+ and Ts- modified ONs showed increased nuclease stability when modifications were present at 3′-end and /or in the middle of the sequence. A single N+ or Ts- modification at the 5′-end did not provide protection of ONs against phosphodiesterase. The extent of resistance to the nuclease of modified ONs increased with the increasing number of modifications. N+ONs showed

higher resistance to nuclease degradation than Ts-ONs with the same number of modifications. For example, 92.0 ± 1.8 % of **4N+ON7** remained intact, whereas only 54.0 ± 2.7 % of intact **4TS-ON13** was found after 2 h of enzymatic digestion (Fig. 4.3).

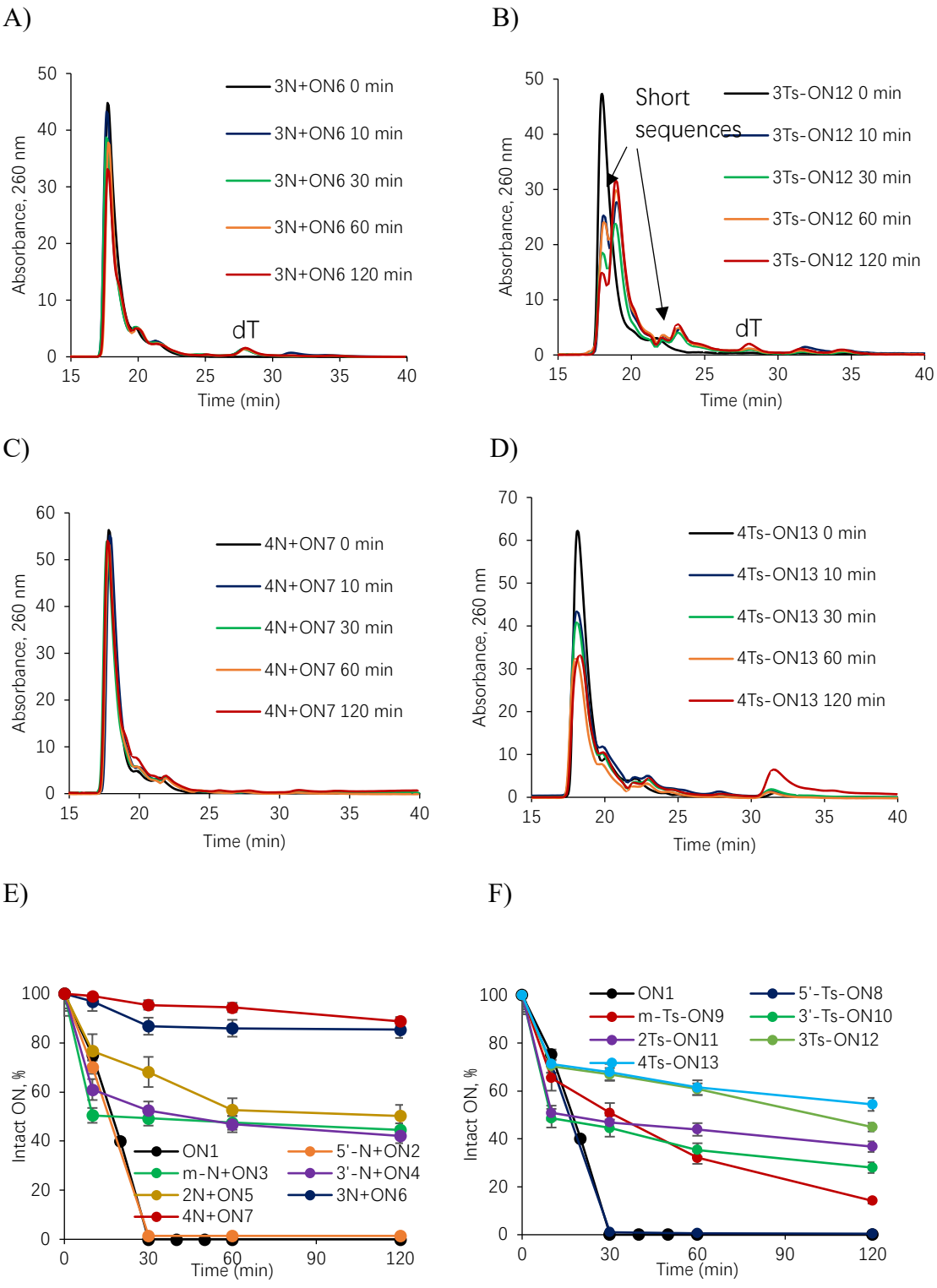


Figure 4.3. Percentage of intact ONs in 120 min. A) N+ONs; B) Ts-ONs. Percentage of intact ONs was determined by the ratio of full-length ONs at individual time in comparison with the sample at 0 min.

4.5.4. Cell uptake assay

Unmodified ONs are unable to penetrate cell membranes as the negatively charged phosphate backbone is repelled by the negatively charged cellular membranes. A transfection reagent is usually required for ONs to cross the cell membrane. We decided to evaluate the cellular uptake of our N+ modified ONs without using any transfection reagent as the negative charge was reduced by the quaternary amine. Hence cell uptake should be improved.

Two modified ONs were synthesised possessing four N+ or Ts- modifications (Table 4.1, 4N+{FAM} and 4Ts-{FAM}, respectively) and a fluorescent label (6-FAM) at the 3′-end. The ONs were incubated with asynchronously growing NIH3T3 mouse fibroblasts for 12 h, washed with Dulbecco's phosphate buffered saline containing Ca^{2+} and Mg^{2+} cations (DPBS, Thermofisher Scientific), cells were fixed in 4% paraformaldehyde, and washed again with DPBS before proceeding to fluorescent confocal microscopy.

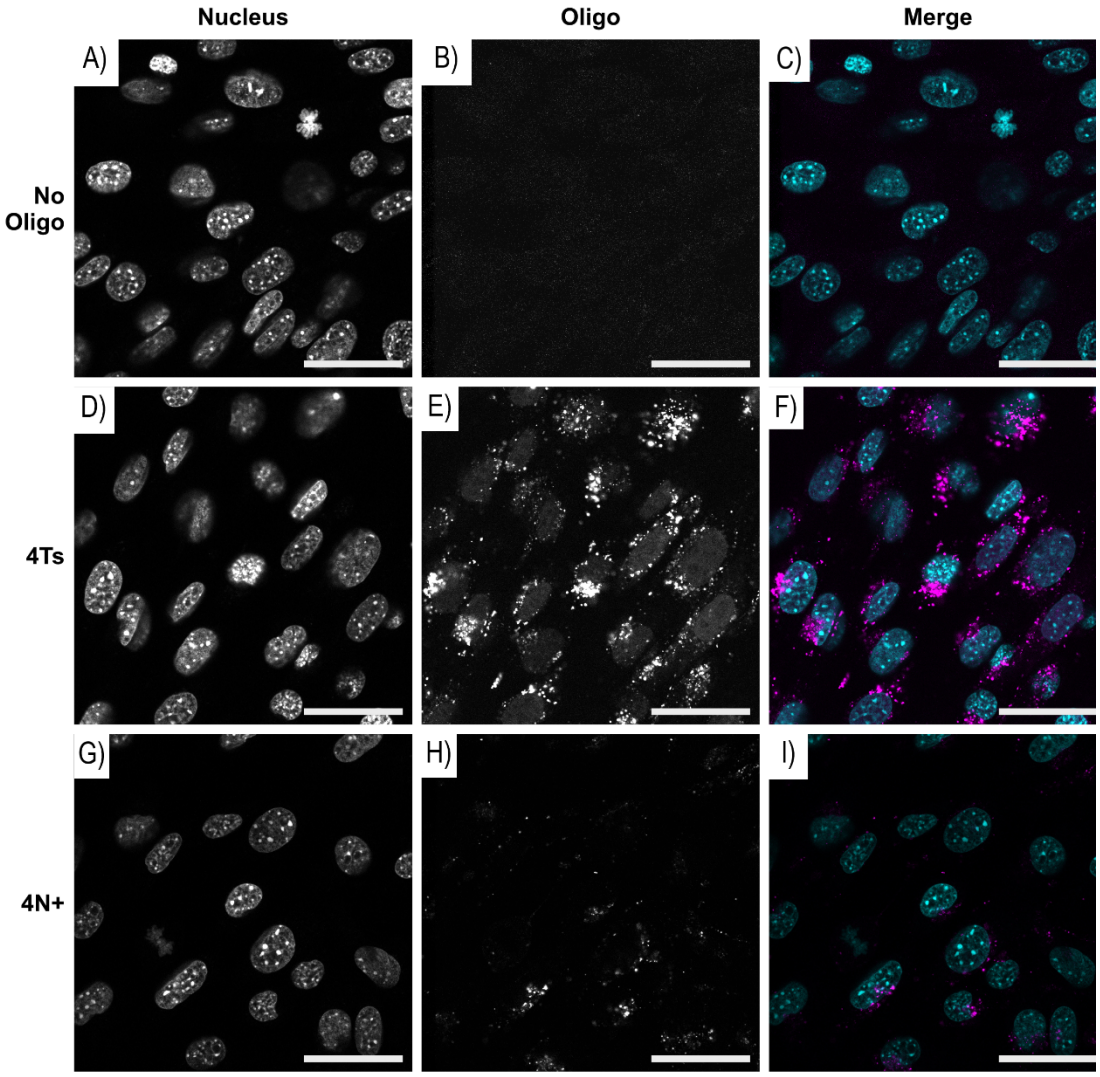


Figure 4.4. Representative images of mouse NIH 3T3 fibroblasts incubated with either (A) No Oligo or 20 μM of (B) 4Ts-, and (C) 4N+ FAM labelled ONs for 12 h. Nuclear DNA was stained with Hoechst 3342 and imaged with scanning confocal microscope. Individual panels are shown for each section: Nucleus/Hoechst 3342 (left column), Oligo/FAM (middle column), and Merge where pseudo-coloured panels are overlaid (right column), nucleus is presented in blue and ONs are in magenta. Scale bar: 40 μm.

For visualisation of cellular uptake, confocal sections that dissect the nucleus were collected to ensure that the FAM-ON imaged were at the edge of the nucleus and not on the cell surface. The optical section was collected after excitation with 405 nm (for DAPI) and 495 nm (for FAM) lasers, respectively (Fig. 4.4).

Figure 4.4 shows that FAM-labelled Ts- and N+ modified DNAs are concentrated in vesicles (bright spots in Fig. E, and H in the middle column) that accumulate around the nuclear membrane. The Ts-modified oligo has also diffused into the nucleus, as indicated by the light grey ovals (Fig. 4.4E) that overlap with the nuclear DNA stained with DAPI (Fig. 4.4D). This contrasts with the lack of colocalisation of the FAM signal with the

nuclear DNA in the negative control (No Oligo) and for the N+ modified oligos (Fig. 4.4B and H, respectively). Figure 4.5 shows that the cellular membrane stained with CellBrite Fix 640 (yellow) also confirmed that the Ts- ON foci are present within the cytoplasm.

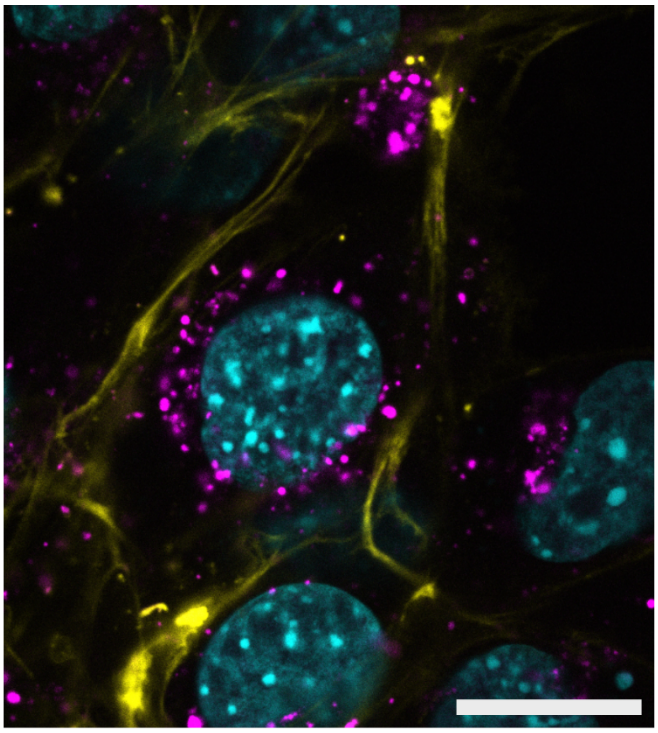


Figure 4.5. Representative confocal microscopy section showing the FAM vesicles inside the cell. Mouse NIH 3T3 fibroblasts were incubated with 20 μM of the FAM labelled 4Ts ON for 12 hours and then stained with Hoechst 3342 and CellBrite Fix 640 to identify the nuclear DNA and the cell membrane respectively. Overlaid pseudo-coloured panels of a section are shown: DNA of the nucleus (blue), 4Ts-FAM (magenta), cell membrane (yellow). Scale bar: 20 μm.

4.6. Discussion

Chemically modified nucleic acid analogues are widely applied in antigene/antisense therapies. For effective inhibition or alteration of transcription or translation, chemically modified ONs should be capable of penetrating the cell membrane, be resistant to nuclease degradation, have no toxicity, and more importantly, bind to the target DNA or RNA sequence-specifically with high affinity.[190] The electrostatic repulsion between negatively charged phosphates is one factor that determines the thermal stability of nucleic acid secondary structures. Neutral or positively charged oligonucleotide analogues should bind more tightly with complementary DNA or RNA. Several studies have introduced positively charged groups to the base,[191, 192] sugar[193, 194] or backbone,[195-197] which led to stable duplexes and triplexes.[198] The introduction of SaRNA monomers into the RNA backbone as a replacement of the phosphodiester backbone led to charge-neutral sulfonamide antisense oligonucleotides (SaASOs), which resulted in lower destabilisation for DNA-RNA duplex than that for DNA-DNA duplex.[179] In contrast, incorporating BCNSs groups into the DNA backbone led to self-neutralising oligonucleotides that did not induce a change of T_m values against complementary DNA sequences,[180] regardless of the number of BCNSs incorporated. This was in line with our results that an increasing number of N+ or Ts-modifications showed no benefit in the T_m of the antiparallel duplex formed by ON, with RNA or DNA. We conducted thermal denaturation experiments, nuclease stability and cell uptake assays and compared our N+ONs with Ts-ONs. The presence of N+ modification enhanced the stability of parallel triplexes at pH 5.0. Ts- modification also stabilised parallel triplexes at pH 5.0, but the extent of stabilisation was decreased with the increasing number of Ts-modifications incorporated. Apart from ONs with N+ or Ts- modification right in the middle of the sequence, both types of modified ONs hybridised with complementary RNA with higher thermal stability than DNA. This slightly contrasts with reported results that ONs with Ts- groups destabilise duplexes with complementary RNA (ΔT_m = -1.6 – -1.2 °C/modification).[181]

The thermal stability of duplexes formed by N+ONs with complementary DNA sequence was less dependent on the salt concentration, which was predicted for zwitterionic nucleotides that can bind to natural DNA at low ionic strength well as or better than natural DNA does with itself.[199] Similar results have been reported for ONs with BCNS groups: the T_m of 2′-OMe duplexes was increased with the increasing number of BCNSs at low ionic strength (25 mM HEPES buffer, pH 7.3). When binding to complementary

RNA, such behavior was not as noticeable. It has been reported that T_m values for ON/RNA duplexes are less sensitive to changes in ionic strength in comparison with ON/DNA duplexes.[200] Ts-modification stabilised duplexes with RNA at high salt concentrations but destabilised RNA duplex at low salt concentration. The polyelectrolyte ion condensation theory can explain how N+ modification can stabilise duplex formation: for natural DNA, the double-helical form has a higher charge density compared to the single-stranded form. Upon denaturation, a portion of the counterions bound to DNA is lost to the bulk solvent due to the reduction in charge density. When one strand is zwitterionic in the DNA duplex, the charge density between duplex and single-stranded states is balanced, and only a fraction of the counterions should be lost during denaturation. As a result, the thermal stability of zwitterionic N+DNA duplexes was less dependent on the ionic strength.[199],[201] This is in line with the lower entropy costs for dsDNA with less charge density.

Native DNA and RNA sequences are highly susceptible to nuclease degradation within the cellular environment. The phosphate modifications can reduce the possibility of enzymic degradation. Even introducing a single N+ and Ts- modification at 3′-end but not 5′- end led to resistance of modified ONs to enzymatic digestion by snake venom phosphodiesterase I, which could be helpful in the cellular applications of N+ and Ts-modified ONs. Modified ONs with the same number of N+ modifications showed higher nuclease resistance than ONs with Ts- modifications.

In the cell uptake study, ONs were delivered without any transfection reagent. Cells were stained and fixed. Visualisation of the optical confocal cross-section through the nucleus can ensure that the fluorescent labelled ONs have entered the cell, not just on the cell surface. In contrast to our design and other neutral ONs such as BCNSs with enhanced cellular uptake, 4N+{FAM} showed only a slight degree of cell uptake after 12 h of incubation. In contrast, ON with four Ts- modifications displayed a very high level of cellular uptake and 4Ts-{FAM} was also found inside the nucleus of the cells.

4.7. Conclusion

We described in this Chapter that chemically modified ONs with zwitterionic N+ group as a phosphate mimic were synthesised. The formation and properties of antiparallel duplexes with RNA and DNA, and in a parallel triplex were evaluated.

The presence of N+ and Ts- modifications on the internucleotidic phosphates enhanced the binding affinity of modified ONs towards complementary RNA. Increased enzymatic stability of N+ and Ts-ONs was demonstrated by nuclease resistance assay. ONs with the same number of N+ modifications showed higher nuclease resistance than Ts-ONs. Ts-ONs can enter the cell and be found in the nucleus, whereas N+ONs showed a limited amount of celluar uptake without using any transfection reagents. These improvements made N+ or Ts- modified ONs promising candidates for gene expression modulation therapeutics or research tools. This study warrants further investigation and development of various phosphate modifications.

Chapter 5. The importance of phosphates for DNA G4 formation: Evaluation of charge-neutral and zwitterionic G-rich ONs

Content in this Chapter was summarised from the following two articles published in peer-review journals:

1) Su, Y., Fujii, H., Burakova, E. A., Chelobanov, B. P., Fujii, M., Stetsenko, D. A., & Filichev, V. V. (2019). Neutral and Negatively Charged Phosphate Modifications Altering Thermal Stability, Kinetics of Formation and Monovalent Ion Dependence of DNA G-Quadruplexes. *Chemistry - An Asian Journal*, *14*(8), 1212-1220.

   In this article, I conducted several experiments, including CD, native ESI-MS, and SE-HPLC, to confirm the G4 formation of chemically modified G-rich ONs, which were synthesised in the laboratory of Dr Stetsenko (Novosibirsk State University, Novosibirsk, Russia). I also performed CD denaturation and association experiments to evaluate the thermal stability and kinetic properties of G4s formed by chemically modified G-rich ONs.

2) Su, Y., Edwards, P. J. B., Stetsenko, D. A., & Filichev, V. V. (2020). The Importance of Phosphates for DNA G-Quadruplex Formation: Evaluation of Zwitterionic G-Rich Oligodeoxynucleotides. *ChemBioChem*, *21*(17), 2455-2466.

   In this article, I synthesised and purified the zwitterionic N+TG$_4$T sequence and evaluated the formation, thermal stability, kinetics, and enzymatic resistance of G4s formed by N+TG$_4$T. I performed all analyses of the derived data, compared it with results from the previous article. I prepared the first draft of the article.

The work reported in this Chapter was also presented at the 2nd International symposium on 'Functional nucleic acids: From laboratory to targeted molecular therapy', Perth, Australia (22 – 23 November, 2018): *Neutral and Negatively Charged Phosphates alter thermal stability and association of G-quadruplexes* as a selected oral presentation.

5.1. Introduction

G-rich DNA or RNA strands can self-assemble into G4s through G-tetrad formation.[202, 203] The G4 structure is stabilised by cations, typically Na^+ or K^+, which reside in the central cavity of the structure forming coordination bonds to the guanine O6 oxygen atoms (Fig. 5.1).[204, 205] Moreover, the presence of the salt is also required to minimise Coulombic repulsion between negatively charged phosphates in the DNA backbone. In general, the degree of stabilisation as well as the association rate of G4s are dependent on the radii of monovalent cations and follow a preference order of $Li^+ < Na^+ < K^+$.[150]

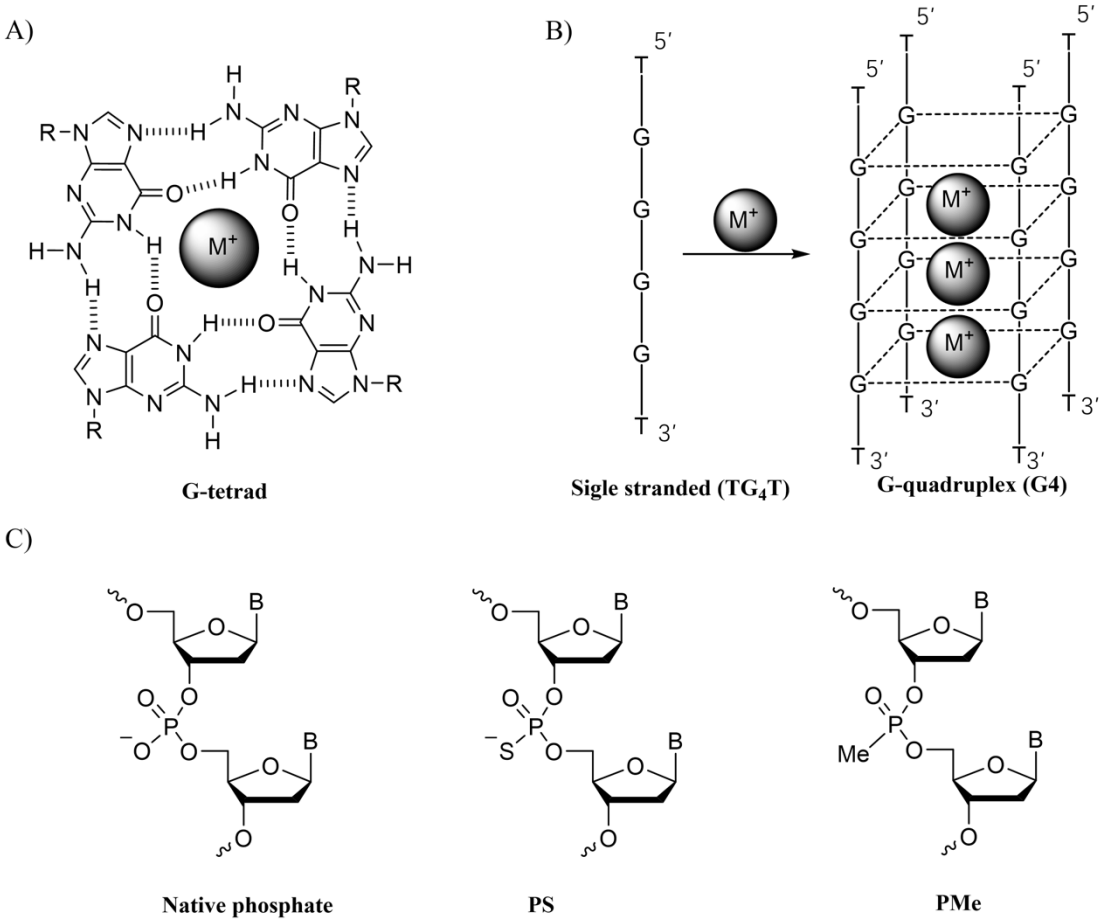


Figure 5.1. Illustration of a G-tetrad and a parallel G4. A) Hydrogen bond formation in a G-tetrad, M^+: Na^+ or K^+; B) tetramolecular parallel G-quadruplex $[TG_4T]_4$ formed by four $d(TG_4T)$ sequences; C) native and chemically modified phosphates previously evaluated in G4s. B, heterocyclic base.

Chemically modified G-rich ONs, including modifications in the nucleobases, sugar residue, and the loop region of G-rich ONs, were widely explored to obtain G4s with the desired structure or topology.[16, 206-208] However, the effect of the chemical alteration of phosphates in the the formation of G4s has not been investigated in detail. Commercially

available phosphorothioates (PS, Fig. 5.1C)[209-211] and methyl phosphonates (PMe, Fig. 5.1C)[209] resulted in formation of G4s with lower thermal stability, but the overall change in properties was relatively insignificant.[209] G-rich PNAs, including their conformationally constrained variants,[212-218] as an example of the replacement of the entire sugar-phosphate backbone, have also been reported to invade native DNA[212, 219, 220] and RNA G4s,[221] which was applied for targeting G4s in living cells.[142] These results suggested that G-rich PNA and PNA analogues can serve as a new method for designing efficient cancer-related sequence-specific cleaving reagents. The ability to invade into native G4s of PNAs has encouraged us to design charge-neutral G-rich ONs that are similar to PNAs and develop novel probes for targeting G-rich regions in genomic DNA.

The *Tetrahymena* telomeric repeat sequence d(TGGGGT) or (TG_4T) forms a tetramolecular parallel G-quadruplex structure (Fig. 5.1B) with all guanines adopting the *anti*-configuration of the glycosidic bond.[150, 222] It is well-characterised and can be used as a reference for evaluating chemically modified G-rich ONs. First, the association of the TG_4T sequences follows a fourth-order reaction. Second, the stabilisation and association rate of $[TG_4T]_4$ follows an order of $Li^+ < Na^+ < K^+$. Last but not least, the dissociation process of $[TG_4T]_4$ is independent of the salt concentration, whereas their association is dependent on ionic strength: the higher the ionic strength, the faster the complex formation.[150]

In this Chapter, three types of chemical modifications have been introduced into ONs as a replacement of all the native phosphates in the sequence: *N*-methanesulfonyl (mesyl) phosphoramidate group (μ-modification, Fig. 5.2B), which is negatively charged;[181] a phosphoryl guanidine group (PG-modification, Fig. 5.2C), which led to a charge-neutral DNA modification,[178, 223] and the 4-(trimethylammonio) butylsulfonyl phosphoramidate (N+, Fig. 4.2D) which resulted in a zwitterionic phosphate mimic.

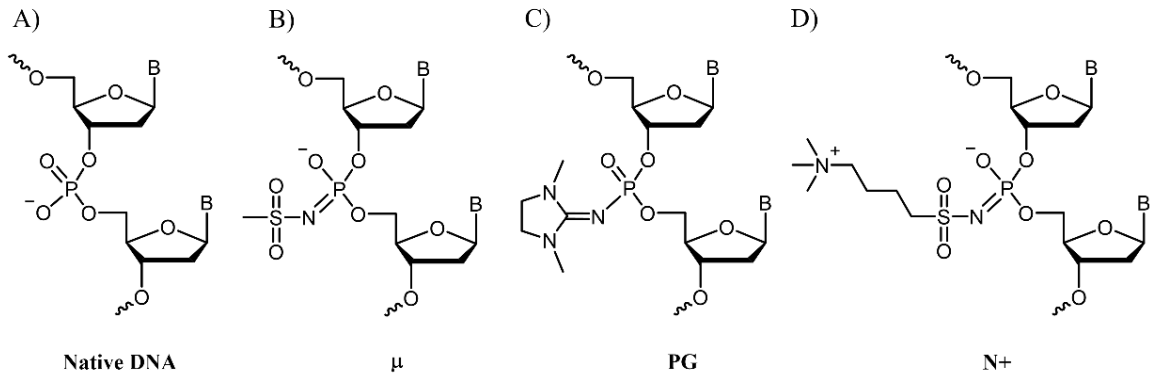


Figure 5.2. Native and chemically modified phosphate linkages in DNA. (A) Native phosphate; B) Mesyl phosphoramidate (μ-modification); C) Phosphoryl guanidine (PG-modification); and D) 4-(Trimethylammonio)butylsulfonyl phosphoramidate linkage (N+). The μ-modification is negatively charged similarly to the native phosphate linkage, whereas the PG group is uncharged and the N+ group is zwitterionic.

5.2. Hypothesis

We hypothesised that the introduction of the μ-modification should not disrupt G4s significantly and can serve as a reference for our studies together with unmodified G4s. G-rich sequences bearing charge-neutral or zwitterionic phosphate mimics (PG and N+ modified TG_4T) should form G4s faster, and this process should be independent of ionic strength due to the lack of charge in the DNA sequence.

5.3. Methodology and methods used for evaluation of G4 formation

Based on the hypothesis made, d(TG_4T) fully modified with the μ-modification (Fig. 5.2B) or PG groups (Fig. 5.2C) were synthesised by Dr Stetsenko in Novosibirsk, Russia, as previously reported,[181, 223] and the synthetic protocol is described in Appendix D. The zwitterionic N+TG_4T (Fig. 5.2D) sequence was synthesised following the protocol described in Chapters 2 and 3. μ-TG_4T was purified by denaturing gel, and PG-TG_4T was purified by RP-HPLC, followed by desalting. Purification of N+TG_4T was accomplished using IE-HPLC using 0.01 M NaOH (pH 12.0) as a mobile phase to disrupt G4 formation during purification. The composition of the ONs was analysed by ESI-MS in the negative ion detection mode (Fig. D.3 in Appendix D).

As illustrated in Fig. 5.3, the formation of G4s was initially evaluated using CD spectroscopy and ESI-MS analysis. However, the ESI-MS analysis for neutral species is problematic due to the lack of charge in the DNA sequence. Alternatively, ^{1}H NMR and SE-HPLC analysis were applied for the confirmation of G4 formation by charge-neutral species.

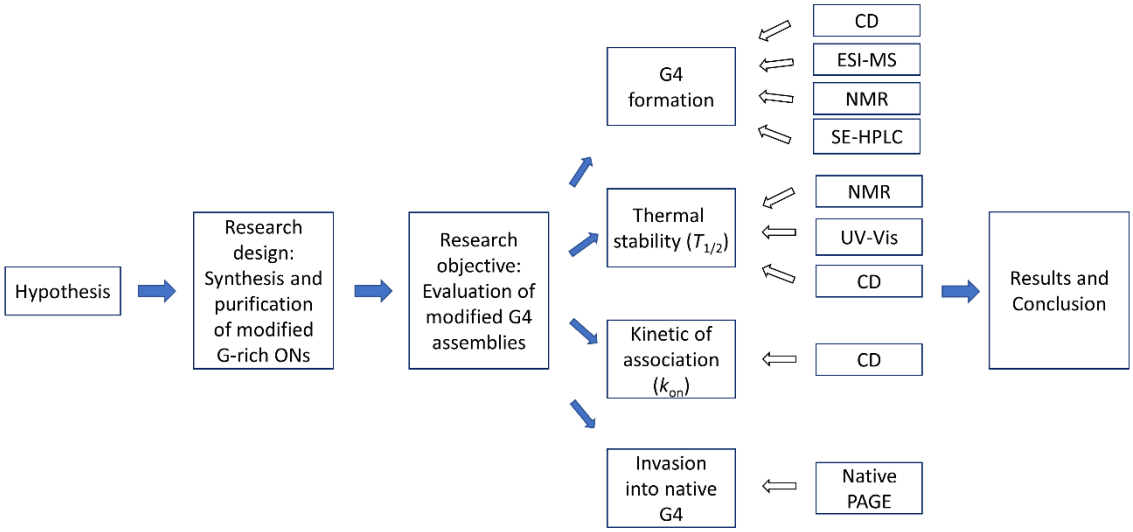


Figure 5.3. Flowchart of methodology and methods used for evaluation of chemically modified G-rich ONs in the formation of G4 assemblies.

The thermal stability of G4s is one of the key factors in evaluating chemically modified DNA secondary structures. The thermal stability of a G4 is presented as $T_{1/2}$ value, the temperature at which half of the sample is folded, and half is unfolded. Due to the hysteresis between melting and annealing profiles of intermolecular G4s, the $T_{1/2}$ value is usually dependent on the heating/cooling rate. The analysis of thermal stability of chemically modified G4s was performed using CD denaturation and renaturation experiment using the same heating/colling rate.

Another key point for assessing the property of modified G4s is the association rate (k_{on}) of G4 formation. A faster association rate is always favoured. Association rate (k_{on}) could be obtained by recording a signal change at a certain wavelength (normally 262 nm for tetramolecular parallel G4s) from denatured to a structured sample using CD over a period of time.

The last objective of this research is to evaluate whether charge-neutral PG or zwitterionic N+ sequences can invade into the pre-formed native G4s, as has been shown for PNA and PNA analogues.[212, 219, 220] We performed 20% native PAGE for the evaluation of G4 invasion since different native species have different mobility on the gel. The neutral species will not enter the gel due to the lack of charge. If invasion occurs, one or several retarded bands should be observed in the mixture of modified and native G4s in comparison with the unmodified G4s, which confirms the formation of hetero complexes with a reduced net charge. Neutral G4s on the gel might not be visible by standard staining of DNA using Stains All, which can complicate the analysis. We also applied SE-HPLC

as an alternative to native PAGE for the investigation of invasion into G4. The differences in retention time of modified and native G4s *versus* invasion products can provide sufficient evidence for the occurrence of invasion.

5.3.1. Research methods used for evaluation of modified G4 assemblies

CD spectroscopy is widely used for distinguishing G4 topologies.[153] There are three types of G4 topologies that can be easily evaluated using CD spectroscopy, parallel-stranded (Type I), mixed or hybrid (both parallel and antiparallel G4s, Type II) and antiparallel-stranded (Type III) topologies. As shown in Figure 5.4, an intense CD band at approximately 264 nm and a negative band at approximately 245 nm are observed for type I topology (Fig. 5.4A). In contrast, type III antiparallel topology shows positive ellipticity at 290 nm and negative ellipticity at 264 nm (Fig. 5.4C). Type II shows a mixed topology with positive bands at 264 and 290 nm and a negative band at 240 nm (Fig. 5.4B).[153]

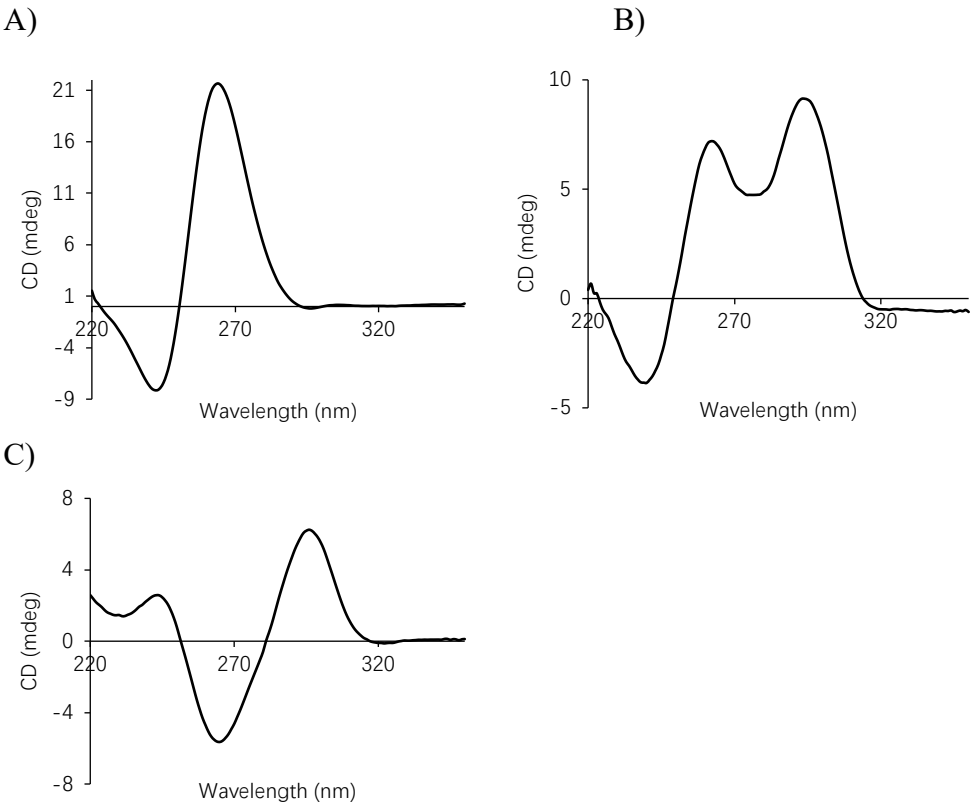


Figure 5.4. CD spectra for a set of standard structurally characterised G4 topologies. A) CD spectrum for parallel stranded G4 of type I; B) CD spectrum for type II G4; C) CD spectrum for antiparallel stranded type III G4.

In comparison with CD and NMR, SE-HPLC uses a lower amount of sample. Moreover, SE-HPLC does not depend on the m/z ratio of the structures as ESI-MS does. Therefore, it can be used for the structural analysis of neutral species.

The structure index (SI) of a given peak is an arbitrary score, calculated using the following equation:[224]

$$SI = V_e/V_0 \times \log_{10}(MW_{structure}) \quad \text{(Equation 1)}$$

Where V_e/V_0 is the relative elution volume, V_e is the elution volume, and V_0 is the dead volume of a SE column. $\log_{10}(MW_{structure})$ is the logarithm of the molecular weight (MW) of the whole structure. SE-HPLC chromatograms from recording the absorbance at 260 nm were normalised to [0, 1] and used to plot a $\log_{10}(MW_{structure})$ against V_e/V_0.

The experimental set up of SE-HPLC started with calibration of a SE column using ssDNA d(T_n) (n = 5, 10, 15, 20, 25, 30, 35, 40, 50) in Na^+ and K^+ solutions, respectively. The relative elution volume V_e/V_0 decreased with increasing of ON's length, ranging from 1.6 to 1.3. Dead volume V_0 was determined using blue dextran (average MW = 2000 kDa) as 2.01 mL, with a retention time of 13.4 min in Na^+ and K^+ solutions. The plot of the decimal logarithm of the MW $\log_{10}(MW)$ against V_e/V_0 gives a linear slope, with R^2 = 0.9995 in both Na^+ and K^+ (Fig. 5.5).

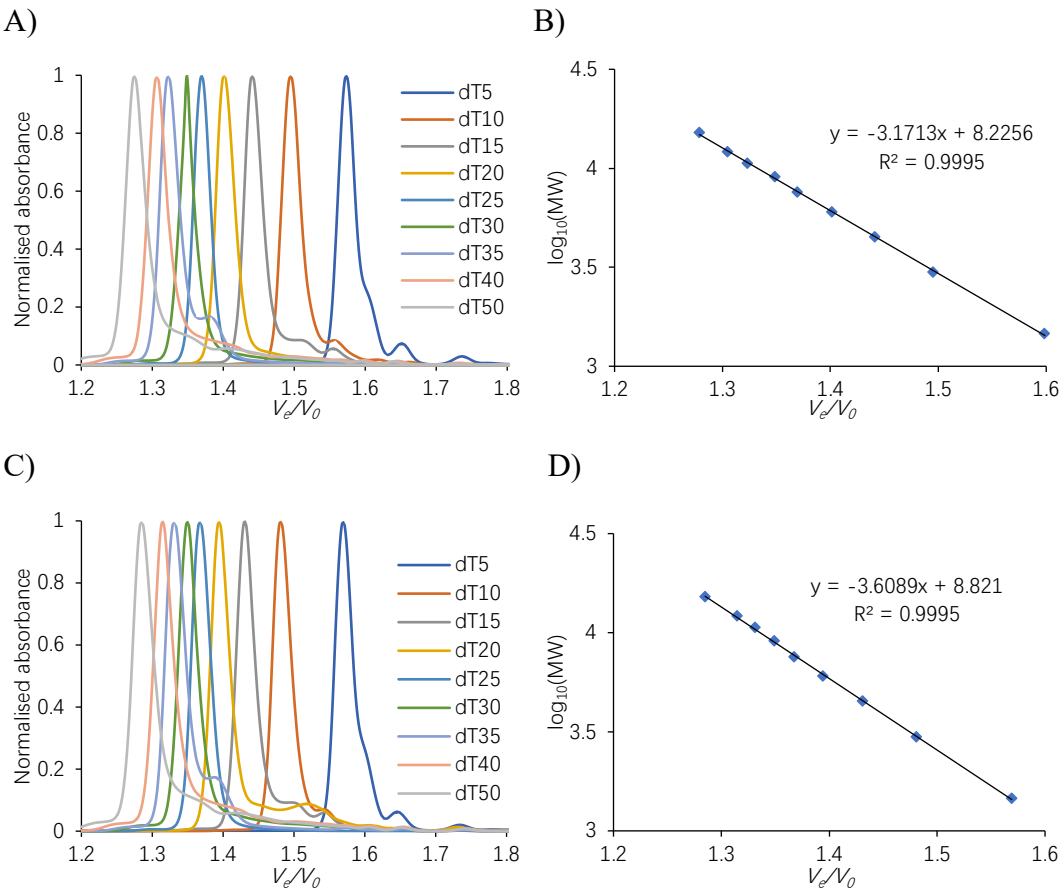


Figure 5.5. Normalised chromatograms of unstructured polypyrimidine d(T_n) oligonucleotides in: A) Na^+, C) K^+. Plot of the molecular weight decimal logarithm against relative elution volume for d(T_n) in: B) Na^+, D) K^+.

Then the experiment was carried out by calibrating the SE-column using natural G4s, d(TG_3T)/d(G_4T)/d(TG_4T)/d(TG_5T)/d(TG_6T). The relative linear alignment of $\log_{10}$(MW) against V_e/V_0 can be used for predicting the formation of G4 structures by chemically modified ONs (Fig. 5.6).

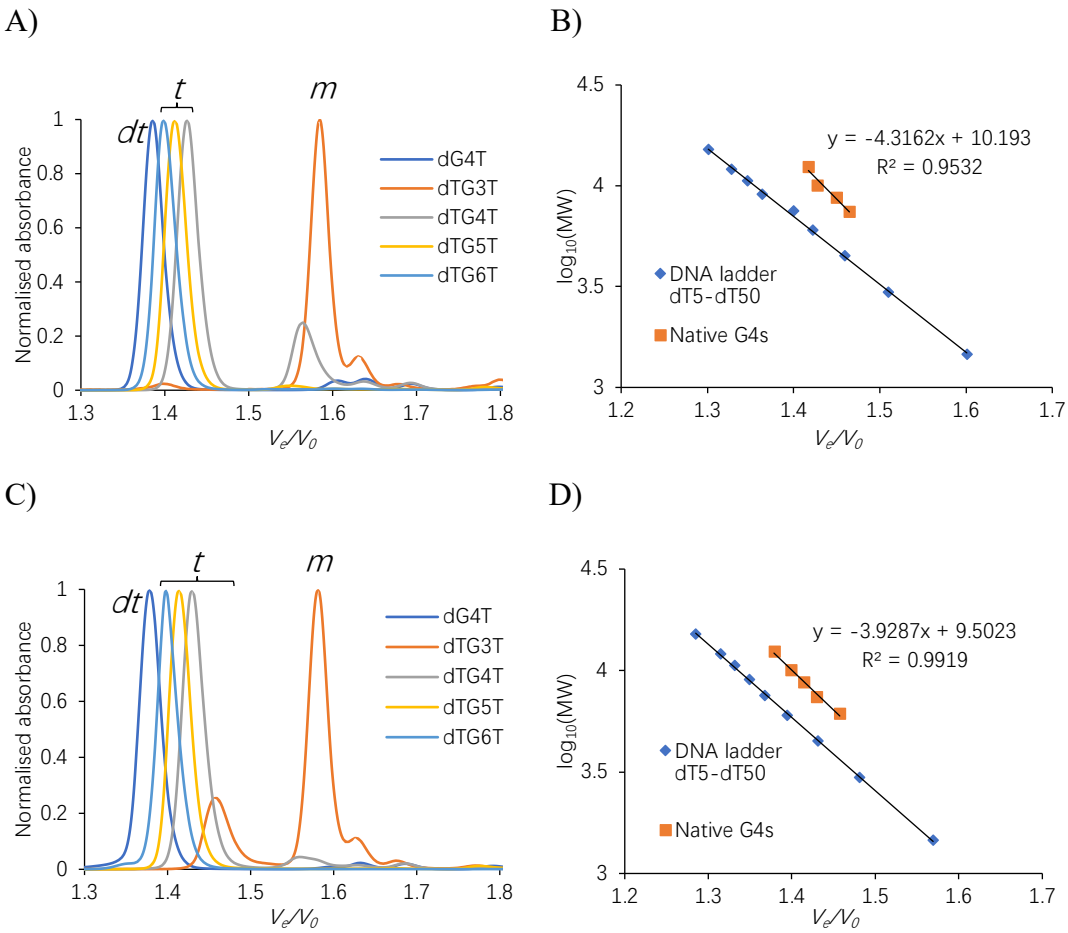


Figure 5.6. Normalised chromatograms of unmodified G4s in: A) Na^+, C) K^+. Plot of the MW decimal logarithm against relative elution volume for G4-DNAs in: B) Na^+, D) K^+. Monomer species are indicated with a *m* and tetramers with a *t*, *dt* stands for dimer of tetramers.

One should mention that $[dG_4T]_4$ formed a dimer of G4[150] and therefore had the fastest retention time, which corresponds to the double molecular weight, i.e. $[dG_4T]_8$. For dTG_3T, the formation of G4s was very slow in Na^+, its chromatogram showed a ssDNA peak. dTG_4T also showed a ssDNA peak of low intensity in Na^+ (Fig. 5.6A). In K^+, $[dTG_3T]_4$ has partially formed, and the formation of $[dTG_4T]_4$ was completed (Fig. 5.6C).

$T_{1/2}$ value of G4s can be obtained using either UV-Vis or CD denaturation and renaturation experiments. By recording spectra using CD at a given heating rate (for example, every 2.5 °C with equilibration for 2.5 min at each temperature) from 5 to 90 °C, a series of spectra at different temperatures can be obtained (Fig. 5.7A). Signal change at 262 nm can be plotted against temperature to give melting and annealing curves. (Fig. 5.7B).

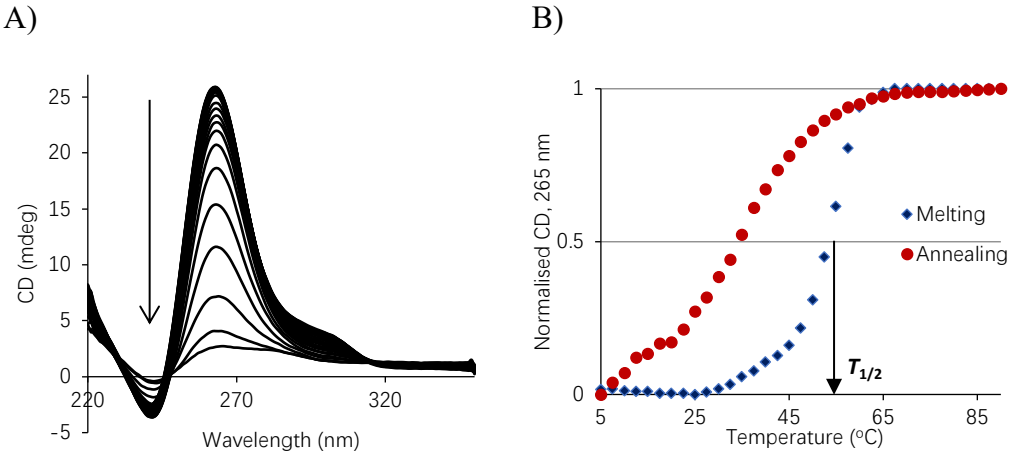


Figure 5.7. A) CD spectra of $[dTG_4T]_4$ (100 μM strand concentration in Na^+), recorded every 2.5 °C from 5 to 90 °C; B) CD signal change at 265 nm *versus* temperature. Conditions: 10 mM Li-cacodylate buffer (pH 7.2, 100 mM NaCl). Arrow in A) indicates the signal change at 265 nm from 5 to 90 °C.

An association curve of the change in absorbance *versus* time (s) could be obtained by recording signal change at 262 nm from denatured to a structured sample using CD at initial strand concentrations, solution conditions, and temperatures indicated. This association curve can be fit to the data using the following equation to obtain association rate (k_{on}):[150]

$$y = M_2 + (M_1 - M_2) \cdot [1 + (C_0)^{n-1} \cdot (n-1) \cdot k_{on} \cdot x]^{(1/(1-n))} \quad \text{(Equation 2)}$$

Where M_1 refers to the absorbance of the sample when all strands exist in unfolded, single-stranded form, M_2 represents the absorbance of the sample when equilibrium is reached between folded and unfolded forms. C_0 represents the initial strand concentration, n represents the reaction order, which is 4 for tetramolecular complexes, and x represents the time elapsed in seconds.

Half-association times ($t_{½}$) were calculated using following formula:[150]

$$t_{½} = (0.5^{(1-n)} - 1) / (C_0^{(n-1)} (n-1) k_{on}) \quad \text{(Equation 3)}$$

in which n and k_{on} parameters obtained from the fit of experimental data, C_0 is the initial strand concentration (100 μM, unless otherwise stated).

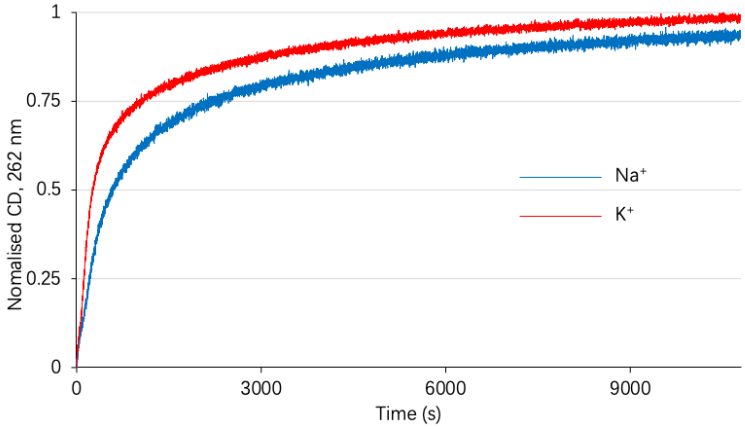


Figure 5.8. Association curves of 100 μM native TG_4T at 5 ºC in 10 mM Li-cacodylate buffer, pH 7.2, supplemented with 100 mM Na^+ or K^+. Start and end point was determined as fully denatured and fully assembled structure.

Take native TG_4T as an example, by fitting the association curve, as shown in Figure 5.8 to Equation 2, the association rate (k_{on}) and n values can be obtained, $k_{on}= 2.97 \times 10^7$ $M^{-3}\cdot s^{-1}$ for Na^+, $k_{on}= 1.1 \times 10^9$ $M^{-3}\cdot s^{-1}$ for K^+, with n = 3.5 and 3.6, respectively. The association half-times ($t_{½}$) at various strand concentrations were then calculated from the above-mentioned Equation 3 using obtained k_{on} and n values (Table 5.1), providing that n and k_{on} do not change at various strand concentrations.

Table 5.1. Calculated association half-times $t_{½}^{[a]}$ for native $[TG_4T]_4$ G-quadruplex at various strand concentrations in 100 mM NaCl or KCl.

Cation	Na^+					K^+				
Stand concentration	1 μM	10 μM	20 μM	50 μM	100 μM	1 μM	10 μM	20 μM	50 μM	100 μM
$t_{1/2}$	2 years	3 days	10 ± 3 h	60 ± 19 min	600 ± 200 s	60 days	5 ± 1 h	48 ± 15 min	4 ± 1 min	40 ± 10 s

5.4. Chapter summary

This chapter focuses on the changes in G4 properties as a result of the introduction of neutral (PG and N+) and negatively charged (μ- modification) into d(TG_4T). Thermal stability, association, and ability of invading into native G4s were evaluated using CD, UV/Vis, ^{1}H NMR, ESI-MS, SE-HPLC and gel electrophoresis analyses. Results in this Chapter highlight the importance of phosphates for the G4 formation and show that further investigations of G-rich strands possessing various phosphate mimics are essential.

5.5. Results

5.5.1. Formation of G4s by chemically modified TG_4T

The formation and topology of native and modified G4s were initially evaluated using CD spectroscopy. As shown in Figure 5.9, CD spectra of complexes formed by all sequences studied at 20 °C in both NaCl and KCl solutions had characteristic features of Type I CD signature with a positive ellipticity at 264 nm and negative ellipticity at 245 nm, which is typical for unmodified tetramolecular, parallel G4s (Fig. 5.9A, B).[225] The positive peak at ~ 295 nm (Fig. 5.9 C and D for μ-TG_4T, E and F for PG-TG_4T, and G and H for N+TG_4T) indicates the presence of dG residues in the *syn* glycosidic conformation,[226, 227] which is also observed in native tetramolecular G4s.[228]

One should note that the introduction of every single modification (μ-, PG-, and N+) creates a chiral centre at the phosphorus atom resulting in a mixture of 2^n diastereomers, where n is the number of modified phosphate groups in the sequence. Therefore, tetramolecular G4s formed by chemically modified TG_4T sequence consists of 128 individual complexes ($2^5 \times 4 = 128$).

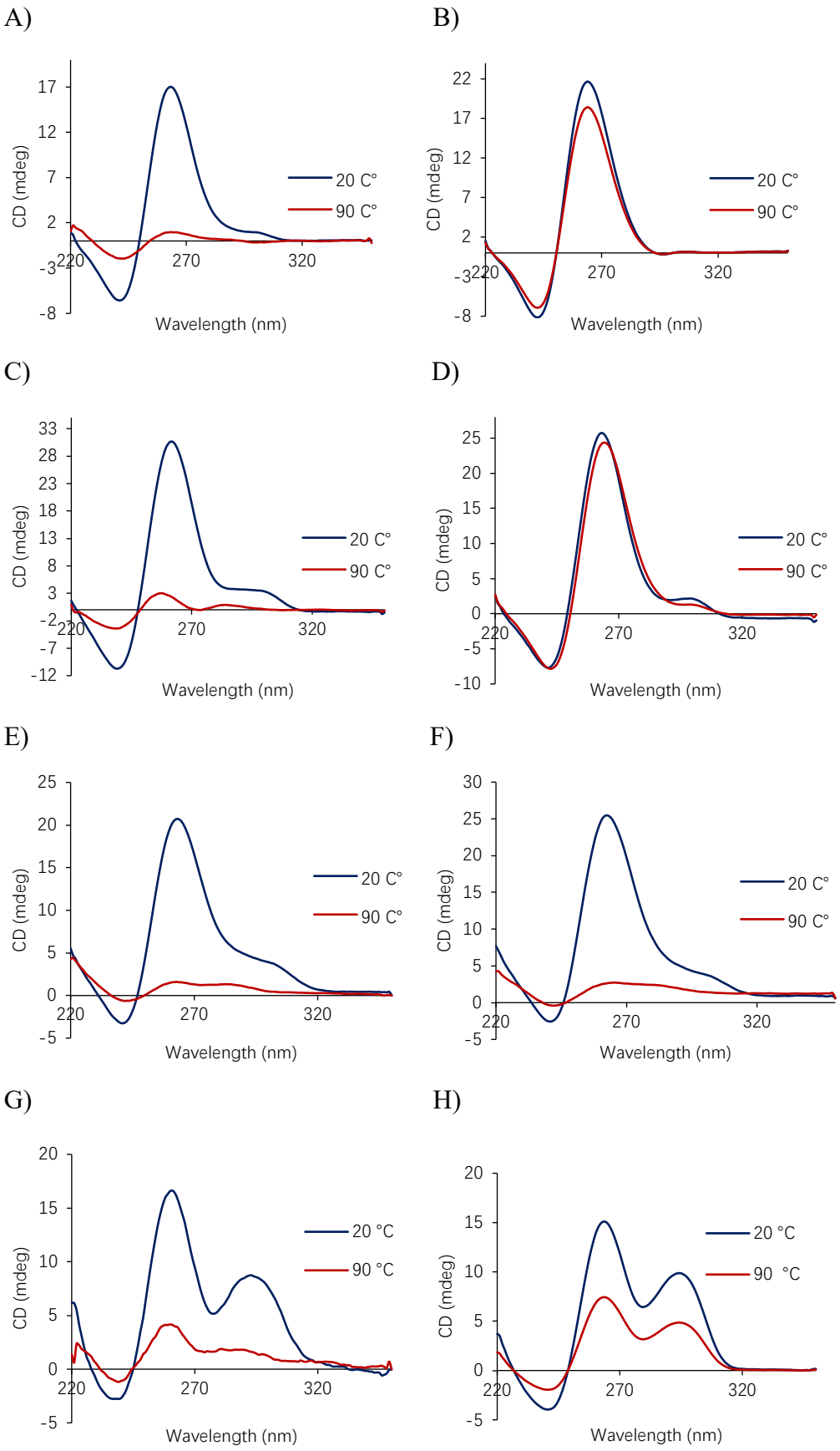


Figure 5.9. CD spectra at 20 and 90 °C for the G-quadruplexes formed by native TG_4T (A and B), μ-TG_4T (C and D), PG-TG_4T (E and F) and N+TG_4T (G and H) in 10 mM Li-cacodylate buffer (pH 7.2) supplemented with 100 mM NaCl (left) or KCl (right).

We then evaluated G4 formation using native ESI-MS. As shown in Fig. 5.10A and B, native dTG$_4$T and μ-TG$_4$T complexes displayed several m/z peaks, attributed to salt adducts of tetramolecular G4s with various charges. The analysis of the G4 stoichiometry for neutral species is problematic due to the lack of charge. The ESI-MS spectra of the PG-TG$_4$T sequence showed the dimers in the 2- charge state accompanied by dimeric species with two or more sodium adducts (Fig. 5.10C). Similarly, ESI-MS spectra of the complex formed by N+TG$_4$T showed a major peak for ssDNA, a peak of small intensity for ssDNA with two sodium adducts and a minor peak of a dimer in the 3- charge state (Fig. 5.10D).

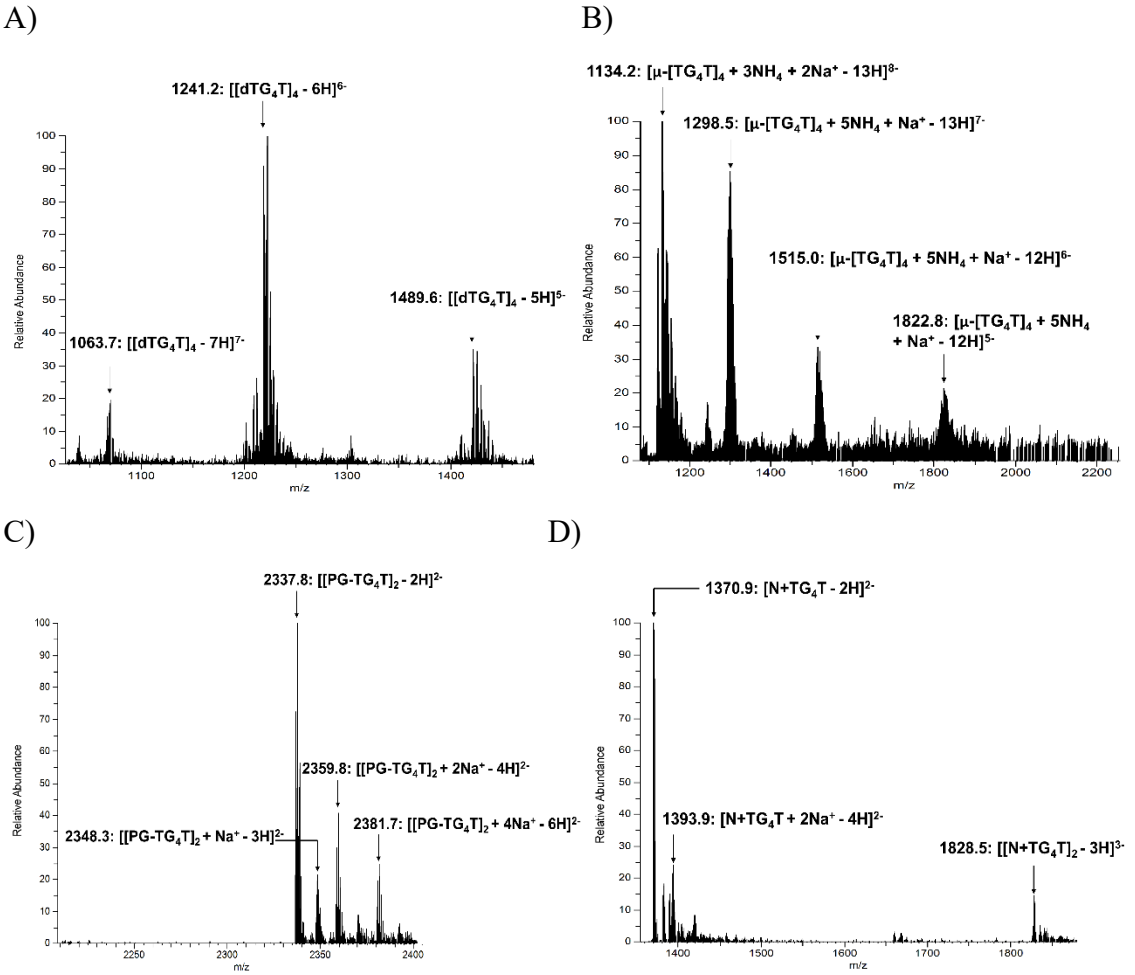


Figure 5.10. ESI-MS spectra: A) native dTG$_4$T complexes; B) μ-TG$_4$T complexes; C) PG-TG$_4$T complexes; and D) N+TG$_4$T complexes in 150 mM ammonium acetate buffer, 20% MeOH, pH 7.0.

To further confirm G4 formation, we recorded ^{1}H NMR spectra of unmodified as well as modified G4s (Fig. 5.11). Similar to native TG_4T (Fig. 5.11A), μ-TG_4T showed four signals in the region from 10.8 to 12 ppm (Fig. 5.11B), indicating that these ONs form highly symmetric complexes containing four G-tetrads. These imino signals also persist at relatively high temperatures indicating the formation of stable G4s. ^{1}H NMR spectra of PG-TG_4T exhibit imino signals at around 11 ppm that were broad and unresolved due to the existence of a mixture of diastereomers and low thermal stability (Fig. 5.11C). Similarly, N+TG_4T also showed several unresolved imino proton signals in the region from 10.5 to 12 ppm (Fig. 5.11D), indicating the existence of a mixture of G4 complexes, which were also stable at relatively high temperatures.

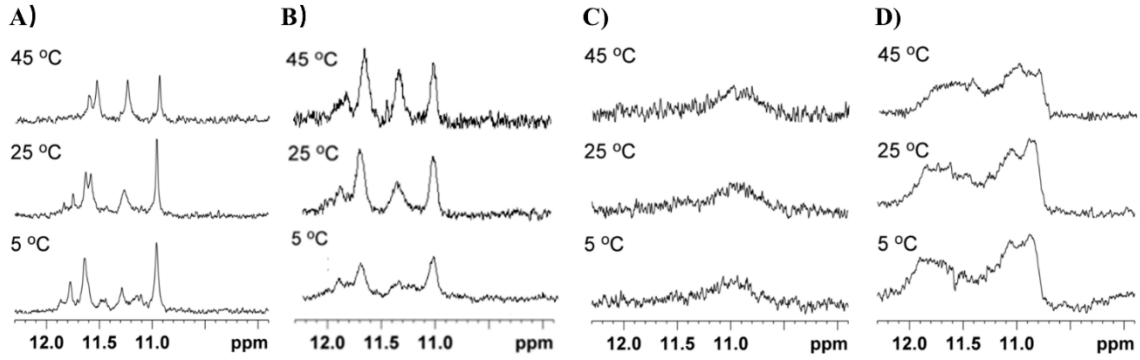


Figure 5.11. ^{1}H NMR spectra in the imino region of complexes formed by TG_4T (A), μ-TG_4T (B), PG-TG_4T (C) and N+TG_4T (D) at indicated temperatures in 10 mM Li-cacodylate buffer containing 100 mM NaCl and 10% D_2O. Sample concentration: 0.3 - 0.7 mM.

We then performed SE-HPLC analysis of G4 formation of native and modified TG_4T sequences to gain further insight into the stoichiometry of complexes. SE-HPLC confirmed the formation of G4s for both native and μ-DNA sequences: Predictive plot of the $\log_{10}$(MW) against the relative elution volume (V_e/V_0)[229] showed that G4s formed by natural strands and μ-DNAs are on the same line (Fig. 5.12B). That confirms that both negatively charged native and μ-DNAs form similar G4s. SE-HPLC profile of PG-TG_4T showed the existence of several entities in solution with significantly larger relative elution volume (V_e/V_0 = 1.6 – 2.0, Fig 5.12A) in comparison with negatively charged complexes reported earlier[229] (V_e/V_0 = 1.3 – 1.5 for tetrameric species and 1.5 – 1.6 for monomers). The range in V_e/V_0 values and SE-profiles for PG-TG_4T indicates the possible co-existence of monomers, dimers, trimers, and tetramers in solution (Fig. 5.12A, green line). The predictive plot for PG-TG_4T species shows linear trends for multi-stranded complexes, which at least confirms the formation of multi-stranded entities in solution

(molecular weights for monomers, dimers, trimers, and tetramers were used in Fig. 5.12C). We isolated individual peaks for PG-TG$_4$T by SE-HPLC and then re-analysed them under identical conditions. Multiple peaks similar to the composition of the original mixtures were observed. This indicates a dynamic equilibrium of different complexes in solution, of which the dimers are the most stable because they can be detected by ESI-MS.

A complex formed by N+TG$_4$T showed a single peak with the same retention time as for native TG$_4$T tetramer (18.5 min, Fig. 5.12A) and a small peak corresponding to the TG$_4$T monomer (19.7 min, Fig. 5.12A), indicating that N+TG$_4$T sequence formed tetramolecular structure ([N+TG$_4$T]$_4$) similar to TG$_4$T.

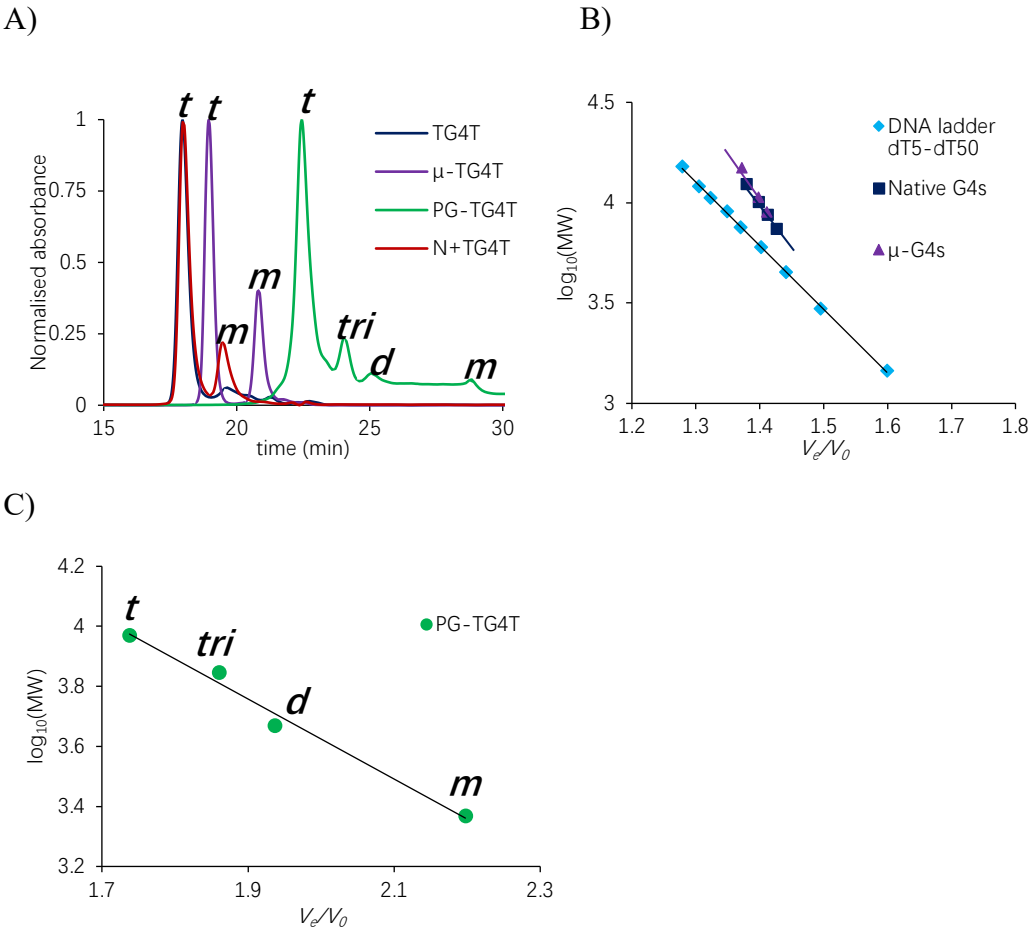


Figure 5.12. SE-HPLC profiles for native and modified G4s. A) Normalised SE-HPLC elution profiles of complexes formed by TG$_4$T, µ-TG$_4$T, PG-TG$_4$T and N+TG$_4$T sequences in 50 mM Tris·HCl buffer, 100 mM Na^+, pH 7.5, strand concentration 100 µM. Monomers are indicated with *m*, dimer with *d*, trimer with *tri*, and tetramers with *t*. B) Predictive plots of the $\log_{10}$(MW) against the relative elution volume (V_e/V_0) for native and µ- G4s. C) Predictive plots of the $\log_{10}$(MW) against the relative elution volume (V_e/V_0) for multi-stranded complexes of PG-TG$_4$T in 100 mM NaCl (molecular weights for monomers, dimers, trimers, and tetramers were used).

5.5.2. Thermal stability and kinetic of formation of chemically modified G4s

CD spectra recorded at 90 °C showed that both native and modified TG_4T complexes denatured fully in NaCl and partially in KCl at elevated temperatures (Fig. 5.8). This allowed us to conduct thermal denaturation experiments and obtain non-equilibrium melting temperatures ($T_{1/2}$, Table 5.2).

Table 5.2. ON sequences, apparent melting temperatures ($T_{½}$), association rate constants (k_{on}), order of reaction (n) and association times ($t_{½}$) in 100 mM NaCl or KCl.

Sequence, 5'-3'	Na^+			K^+		
	$T_{½}$, ºC *[a]*	k_{on} *[b]*	$t_{½}$ *[c]*, s	$T_{½}$, ºC	k_{on}	$t_{½}$, s
TG_4T	56	2.9×10^7 $n = 3.5$	600 ± 200	>90	1.1×10^9 $n = 3.6$	40 ± 10
μ-TG_4T	53	7.5×10^6 $n = 3.3$	260 ± 80	>90	20 $n = 1.9$	190 ± 60
PG-TG_4T	48	300 n = 2.1	180 ± 60	84	1.8×10^{11} n = 4.3	270 ± 80
N+TG_4T	64	0.016 n = 1.1 *[d]*	150 ± 46	>90	0.116 n = 1.4	333 ± 105

[a] (Non-equilibrium) melting temperature of the preformed complexes, in ºC (± 1.5 °C), determined from thermal denaturing profiles recorded with a temperature gradient of 0.24 ºC/min.
[b] Association rate constant at 5 ºC, pH 7.2, with 100 mM Na^+ or K^+. In situations when n ~ 4, k_{on} is given in $M^{-3}\cdot s^{-1}$, when $n \sim 2$, k_{on} is expressed in $M^{-1}\cdot s^{-1}$, when $n \sim 1$, k_{on} is expressed in s^{-1}. In all cases, k_{on} and n are reported with ± 30% and ± 10% uncertainties as determined from several experiments, respectively.
[c] Time required for 50% complex formation ($t_{½}$) calculated from Equation 3 using k_{on} and n parameters obtained from the fit of experimental data and initial strand concentration of 100 μM. Uncertainties in $t_{½}$ was calculated using error-propagation method using uncertainties in k_{on} and n mentioned in [b].
[d] n value for N+TG_4T in Na^+ was determined by fitting the initial curve within 1800 s for a better fit, see Fig D.5 in Appendix D.

According to Table 5.2, both μ- and PG-modified G4s were less thermally stable than the natural ones, possibly due to additional substituents attached to the phosphorus atoms. However, G4 formed by neutral PG-DNAs was significantly less thermally stable than the complex formed by a negatively charged μ-TG_4T, possibly due to steric bulkiness of the 1,3-dimethylimidazolidine-2-imino modification and potential clash between the PG-groups belonging to the opposite stereoisomers in the adjacent strands. N+TG_4T was stabilised considerably in Na^+-containing buffer in comparison with native TG_4T, μ-TG_4T, and PG-TG_4T by $\Delta T_{1/2}$ of + 8 °C, + 11 °C, and + 16 °C, respectively. The use of K^+ in the buffer stabilised modified G4s considerably compared to Na^+, which is in good agreement with previous observations for unmodified G4s.[150]

Next, we evaluated the kinetics of G4 formation by recording the intensity of a CD signal at 262 nm every 0.25 s at 5 °C over 3 h. It was surprising to observe that in contrast to negatively charged sequences (Fig. 5.13A and B), neutral PG-TG_4T assembled faster in Na^+ than in K^+ containing solutions (Fig. 5.13C). Assembly of N+TG_4T was initially faster in Na^+ than K^+ ions, which was in line with PG-TG_4T but contrasted the trend observed for negatively charged sequences. However, N+TG_4T in solution with Na^+ showed a slower association than in solution with K^+ after complex formation exceeded 75% (Fig. 5.13D). As we mentioned earlier, every N+ modification creates a chiral centre at the phosphorus atom resulting in a mixture of diastereomers. Therefore, the general trend of association is the average of assembly of 128 complexes ($2^5 \times 4 = 128$). The overall assembly of neutral ONs was faster in the K^+ solution, which is in line with the general trend for unmodified TG_4T, but some ONs assemble faster in Na^+. This led to the initial association speed in Na^+ solution being faster than in K^+ solution.

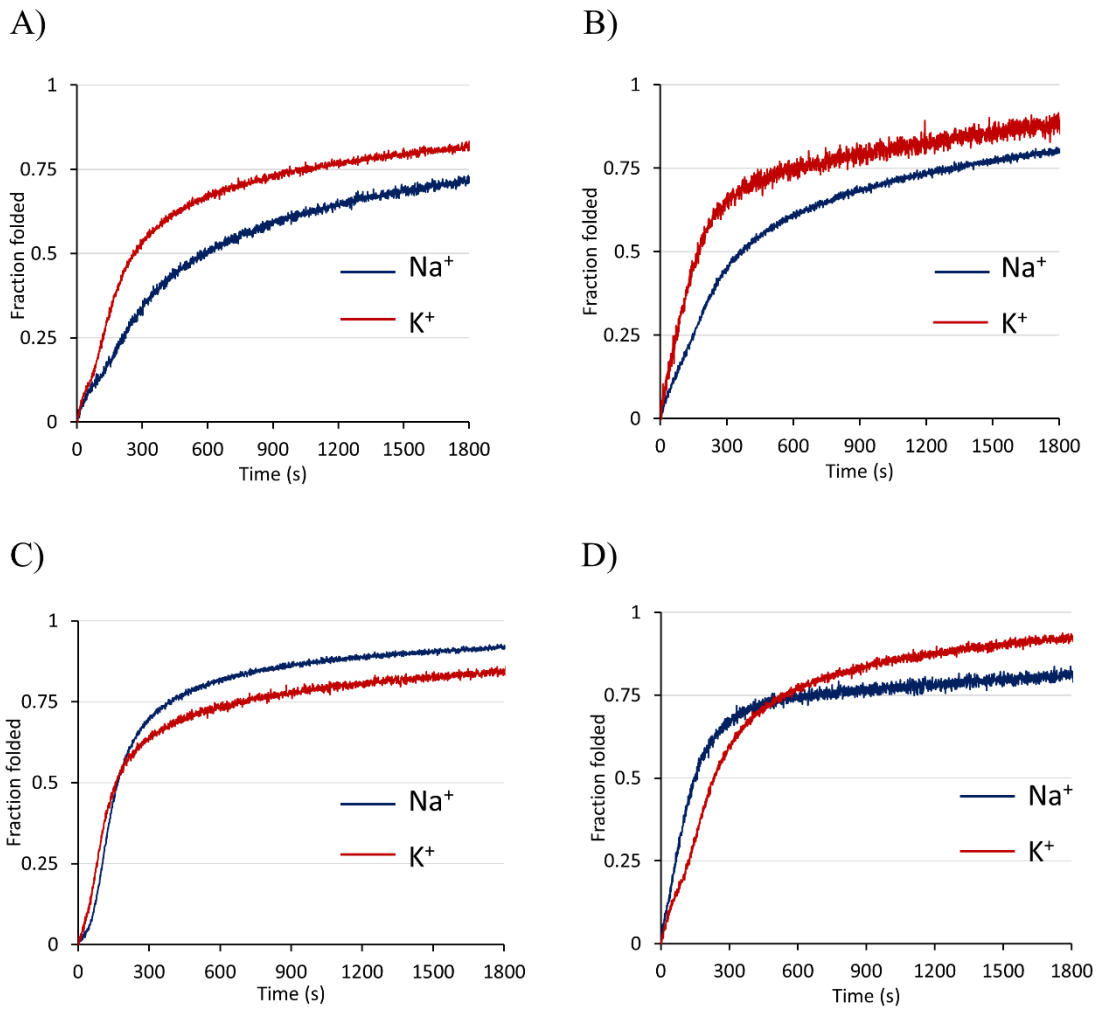


Figure 5.13. Association of A) native TG_4T; B) µ-TG_4T; C) PG-TG_4T; and D) N+TG_4T at 100 µM strand concentration at 5 ºC in 10 mM Li-cacodylate buffer, pH 7.2, supplemented with 100 mM Na^+ or K^+.

The association rate k_{on} and the order of reaction (n) were obtained by fitting the association curve to the above-mentioned Equation 2 in this Chapter. The order of reaction (n) in Equation 2 was set as unknown for a better fit. However, for N+TG_4T in Na^+, n was determined by fitting the association curve during 1200 s of assembly that gave the best fit (see Fig. D.5A, C and D in Appendix D).

Despite the fact that the complex formed by N+TG_4T showed a single peak for the tetramer on the SE-HPLC profile (Fig. 5.12), the obtained order of the reaction (n) for N+TG_4T in both Na^+ and K^+ solutions was in a range of 1.1 – 1.4, rather than around 4, which is usually observed for unmodified tetramolecular G4s. Such a drastic difference in the order of reaction can be a result of a reduction of electrostatic repulsion for N+TG_4T. Native TG_4T assembles very slow due to the electrostatic repulsion between negatively

charged phosphates. Usually, the association of unmodified TG_4T follows a two-step model, where single strands and dimers are in a rapid pre-equilibrium that favours single strands, and tetramer formation from the dimer is a rate-limiting step.[150, 230] As a result, a fourth-order reaction with n between 3 and 4 is usually detected. By reducing the electrostatic repulsion, the assembly of four zwitterionic N^+TG_4T strands is much faster than that of unmodified ones, as shown in Fig. 5.13D. This means that the tetramer formation from the dimer or directly from the single strand is fast, leading to the observed order of the reaction in the 1.1 – 1.4 range. In fact, dimer intermediates are detectable in ESI-MS (Fig. 5.10D).

Table 5.3. Calculated association half-times $t_{½}$[a] for TG_4T quadruplexes at various strand concentrations in 100 mM NaCl or KCl.

DNA sequence, 5'-3'	Na^+				K^+			
	1 µM	10 µM	20 µM	50 µM	1 µM	10 µM	20 µM	50 µM
TG_4T (calculated)	2 years	3 days	10 ± 3 h	60 ± 19 min	60 days	5 ± 1 h	48 ± 15 min	4 ± 1 min
µ-TG_4T (calculated)	100 days	13 h	90 ± 12 min	17 ± 3 min	3 h	30 min	18 ± 5 min	5 ± 2 min
PG-TG_4T (calculated)	10 h	39 ± 12 min	18 ± 6 min	6 ± 2 min	35 years	6 days	15 ± 5 h	43 ± 14 min
N^+TG_4T (calculated)	273 ± 86 s	202 ± 64 s	185 ± 60 s	164 ± 52 s	38 ± 12 min	15 ± 5 min	11 ± 3 min	7 ± 3 min
N^+TG_4T (experimental)[b]	nd[c]	550 s	298 s	182 s	nd	20 min	15 min	10 min

[a] Calculated from above-mentioned Equation 2 for $t_{1/2}$ using k_{on} and n values from Table 4.1.

[b] Experimental $t_{1/2}$ values were reported with ± 10% uncertainties as determined as an average value from at least three repeated experiments with ± 15% uncertainties, see Fig. D.6 in Appendix D for association curves at different strand concentrations for N^+TG_4T.

[c] G4 formation at 1 µM strand concentration cannot be detected due to low absorption of DNA samples.

Since the experimental order of reaction varies dramatically (from $n = 3.5$ for native TG_4T to $n = 2$ or 1 for charge-neutral PG-TG_4T and zwitterionic N+TG_4T sequences, respectively), a direct comparison of rate constants for these three sequences is problematic. For a better comparison, we calculated $t_{½}$ values for the association of each sequence at various strand concentrations (1, 10, 20, and 50 μM, Table 5.3),[150] assuming that the order of reaction n does not deviate significantly at different strand concentrations. In Na^+ solutions, both PG- and N+ TG_4T strands form G4s faster in dilute solutions than native TG_4T. However, in K+ solutions, assembly of PG-TG_4T sequence was significantly slower than the formation of all complexes studied.

As the obtained order of reaction (n) of N+TG_4T was 1.1 – 1.4, the association of zwitterionic N+TG_4T sequence should be less dependent on the strand concentration: when concentrations changed dramatically (100 μM – 1 μM), the calculated $t_{1/2}$ value for N+TG_4T in Na^+ solutions did not change as much as for the native, μ- and PG sequences. Similar but less pronounced property was also seen for K^+-containing solutions: even at 1 μM strand concentration, N+TG_4T sequence can form G4s relatively fast ($t_{½}$ = 38 min compared to $t_{½}$ = 60 days and 35 years for native TG_4T and PG-TG_4T, respectively, Table 5.3). The observed $t_{1/2}$ values for N+TG_4T in K^+ showed a good agreement with calculated values, whereas the observed $t_{1/2}$ values for G4s formed in Na^+ were slightly higher than the calculated ones due to the noisy data observed for low DNA concentrations (10 and 20 μM, Fig. D.6 in Appendix D). However, as for all tetramolecular G4s, an increase in strand concentration led to faster G4 association for N+TG_4T, a result of that the order of reaction is higher than 1.0.

Increasing the ionic strength played no role in the thermal dissociation of both modified and unmodified complexes ($\Delta T_{½}$ < 2 ºC, Fig. 5.14A).[150] However, as described in the past,[150] varying the concentration of Na^+ had a dramatic effect on the association constant k_{on} of negatively charged species (Fig. 5.14B): the higher the Na^+ concentration, the higher the k_{on}. In contrast, the formation of G4s by charge-neutral PG- and zwitterionic N+ DNAs was independent of Na^+ concentration (Fig. 5.14C and D), which was expected for the sequences lacking the negative charge.

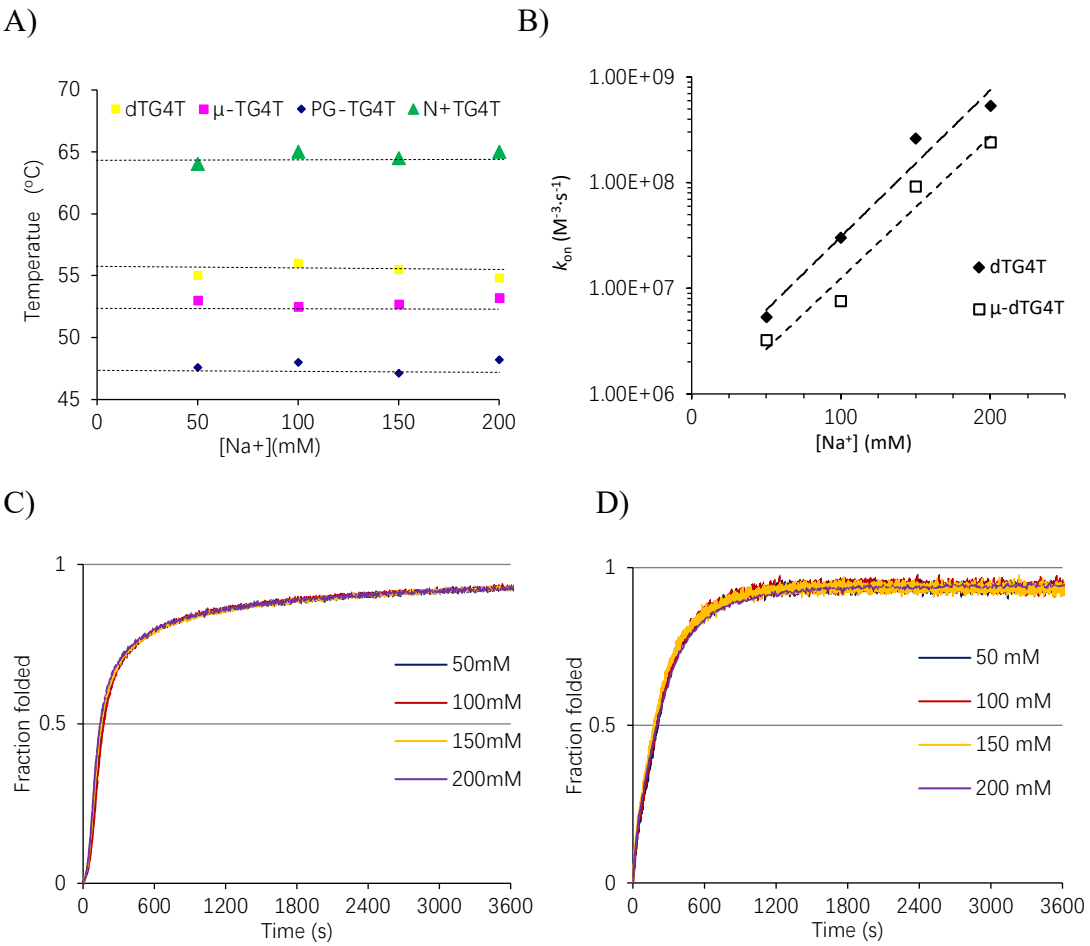


Figure 5.14. Dependence of the $T_{1/2}$ values and k_{on} values of native and modified G4s at various Na^+ concentrations (50 mM to 200 mM). A) Dependence of the $T_{1/2}$ value (obtained upon heating at a 0.24 ºC/min) of the pre-formed modified and natural TG_4T species in 10 mM Li-cacodylate buffer (pH 7.2) with various Na^+ concentration. B) Dependence of k_{on} values as a function of Na^+ concentration for negatively charged native and µ-TG_4T strands. C) Association of neutral PG-TG_4T at various Na^+ concentrations in 10 mM Li-cacodylate buffer, pH 7.2, at 5 ºC. D) Dependence of association at 5 ºC of N+TG_4T in 10 mM Li-cacodylate buffer, pH 7.2, supplemented with various Na^+ concentrations. All samples were evaluated at 100 µM per strand concentration.

Unmodified G-rich sequences are unable to form G4s in Li^+ solutions. Since the assembly of N+TG_4T sequence is less dependent on the ionic conditions, we assume that N+TG_4T can form G4s in Li^+ solutions. As shown in Figure 5.15A, after incubation in Li^+ solution for four days, we observed for the first-time complex formation even at 10 µM strand concentration. The positive ellipticity at ~ 255 nm observed for single-stranded DNA shifted to 264 nm, clearly indicating the formation of G4 (Fig. 5.15A). In water, the peak with a positive ellipticity at ~ 255 nm remained unchanged even after one week (Fig. 5.15B). Native TG_4T is unable to form G4 in Li^+ solutions, especially at such a low strand concentration.[231] Thermal denaturation experiment (Fig. 5.15C) showed that $T_{1/2}$ of $[N+TG_4T]_4$ in Li^+ solution was 56 °C (Fig. 5.15D), which was lower than $T_{1/2}$ observed

for $[N+TG_4T]_4$ in Na^+ solution (65 °C). In this regard, $N+TG_4T$ complex follows the thermal stability pattern established for the native sequences: $Li^+ < Na^+ < K^+$.

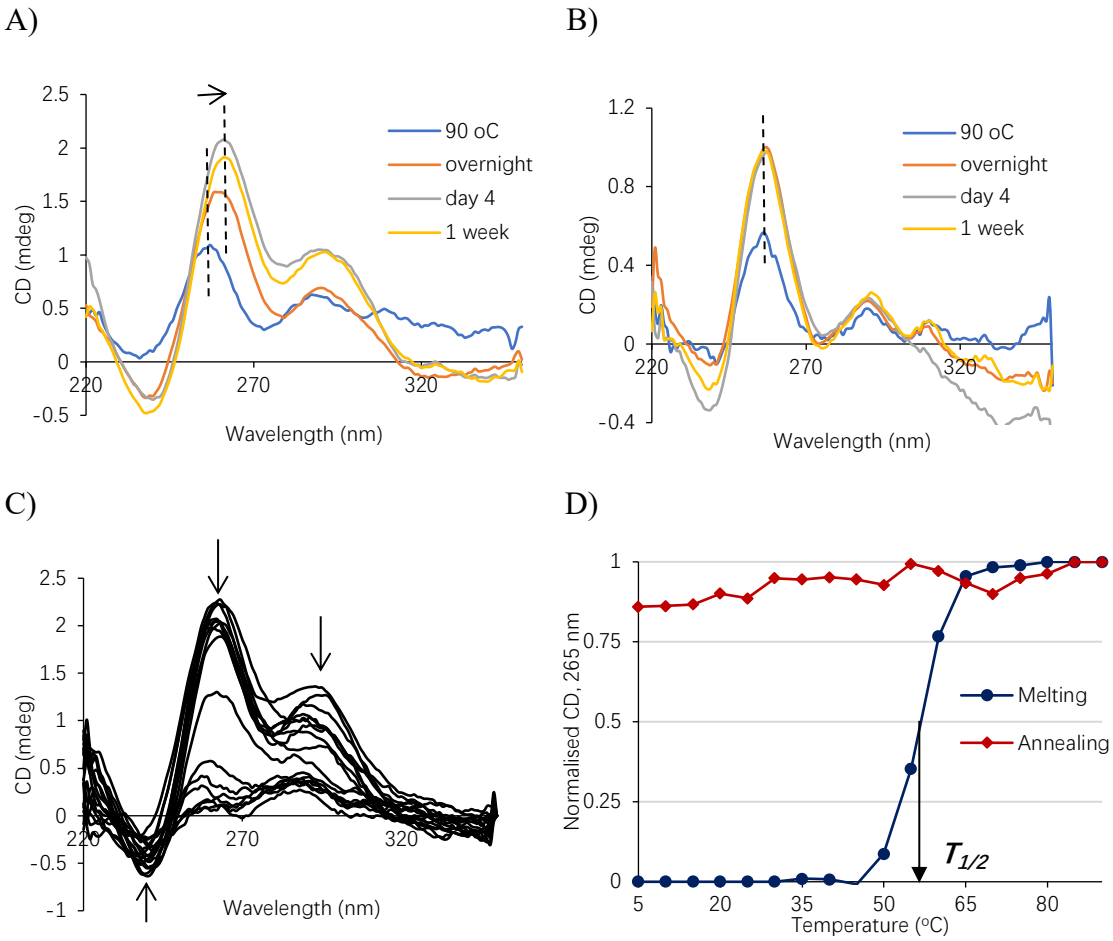


Figure 5.15. CD spectra of 10 μM $N+TG_4T$ in A) 10 mM Li-cacodylate buffer (pH 7.2) supplemented with 100 mM LiCl and B) in H_2O. C) CD melting spectra of $[N+TG_4T]_4$ (10 μM strand concentration in Li^+), recorded every 5 °C/5 min from 220 nm to 350 nm; arrow indicates direction of changes in the peak intensity from low to high temperatures. D) The melting and annealing profiles of signal changing at 265 nm, $T_{1/2}$ was determined as 56 °C.

5.5.3. Evaluation of invasion of PG- and N+TG$_4$T sequences into the native [TG$_4$T]$_4$

Next, we assessed if the neutral sequences could form mixed G4s with native DNA strands. Equimolar mixtures of pre-formed complexes at 100 µM concentration were incubated for 16 h at 5 ºC and analysed by SE-HPLC (Fig. 5.16A). A slight delay in the retention times of native G4s and appearance of a new peak for TG$_4$T samples suggest formation of mixed G4s. Peaks at the relative elution volume of ~1.4 were isolated and analysed using ESI-MS, which showed the formation of several complexes with various ratios of PG-modified and native TG$_4$T strands (Fig. 5.16B). These observations indicate that neutral PG-TG$_4$T can invade the already formed G4s. However, in-depth characterisation of the thermodynamics and kinetics of such invasion is beyond the scope of this study.

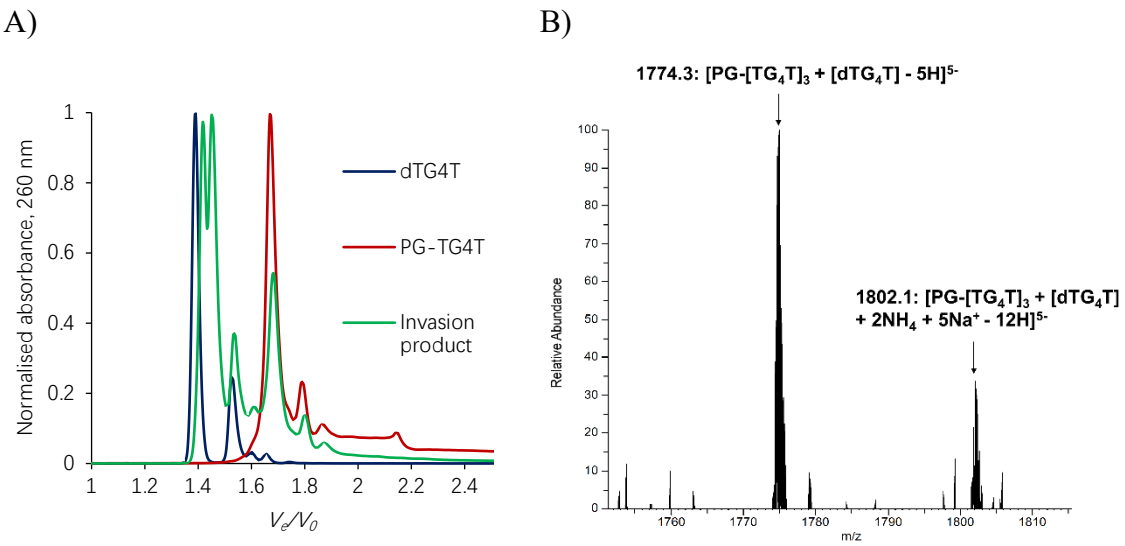


Figure 5.16. A) Normalised SE-HPLC elution profiles showing the relative elution volume of native and PG-TG$_4$T oligos before and after mixing in the presence of 100 mM NaCl. B) ESI-MS spectra of corresponding invasion products.

We assumed that N+TG$_4$T might exhibit similar properties as PG-TG$_4$T. We prepared fluorescein labelled TG$_4$T (TG$_4$T{FAM}) and analysed the formation of mixed G4s using 20% native PAGE. To form thermodynamic products of the reaction, we mixed N+TG$_4$T and TG$_4$T{FAM} samples in TBE buffer (100 mM Na^+, pH 8.0), heated it at 90 ºC for 10 min, cooled down, and incubated at 5 ºC overnight. Two retarded bands were observed for thermodynamic products (Fig. 5.17A, lane 3) in comparison with the [TG$_4$T{FAM}]$_4$ (Fig. 5.17A, lane 2), which confirmed the formation of heterocomplexes with a reduced net charge. Pre-formed [TG$_4$T{FAM}]$_4$ and [N+TG$_4$T]$_4$ samples at different molar ratios

(1: 1, 1: 2, and 1: 5, respectively, Fig. 5.17A, lanes 4-6) were incubated at 5 ºC for one week, but no retarded bands were observed. Only after three weeks of incubation, two new retarded bands were observed in native PAGE (Fig. 5.17B, lane 4). The band in lane 3 (Fig. 5.17B) corresponding to thermodynamic product shows that even at low strand concentration (10 μM), the thermodynamic product was able to form after three weeks, indicating that although products formed in lane 3 (both A and B in Fig. 5.17) were thermodynamically preferable, kinetically the process is slow. Slow invasion of N+TG_4T into native $[TG_4T]_4$ is most likely due to the increased thermal stability of $[N+TG_4T]_4$.

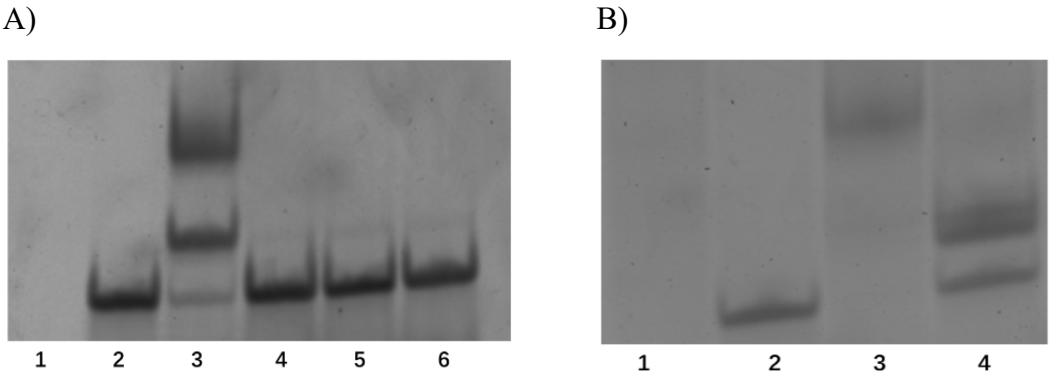


Figure 5.17. 20% native PAGE showing difference in mobility of G4s in 1 × TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM Na_2-EDTA, pH 8.0) supplemented with 100 mM NaCl at 100 μM strand concentrations. A) Lanes are: 1: $[N+TG_4T]_4$; 2: $[TG_4T\{FAM\}]_4$; 3: Mixture of pre-formed $TG_4T\{FAM\}$ and $N+TG_4T$ complexes (1: 1), heated at 90 °C for 10 min, slowly cooled down to RT; 4: Mixture of pre-formed $TG_4T\{FAM\}$ and $N+TG_4T$ complexes (1: 1); 5: Mixture of pre-formed $TG_4T\{FAM\}$ and $N+TG_4T$ complexes (1: 2); 6: Mixture of pre-formed $TG_4T\{FAM\}$ and $N+TG_4T$ complexes (1: 5). B) Lanes are: 1: $N+TG_4T$; 2: $[TG_4T\{FAM\}]_4$; 3: Mixture of $N+TG_4T$ and pre-formed $[TG_4T\{FAM\}]_4$ (1: 1), heated at 90 °C for 10 min, slowly cooled down to RT; 4: Mixture of $N+TG_4T$ and pre-formed $[TG_4T\{FAM\}]_4$ (1: 1).

5.6. Discussion

Various chemical modifications have been introduced into G-rich sequences to affect the assembly, stability, and topology of tetramolecular G4s, leading to certain changes in the pattern of G4 formation. Nucleobase modifications[232-234] can affect the conformation of a glycosidic bond (*anti*- or *syn*-), which is very useful for the NMR investigation of G4 assemblies. Parallel G4s with sugar modifications such as 2′-OMe RNA[176] or LNA[235] were more thermally stable than unmodified ones and showed increased anti-HIV-1 activity.[236] Chemical modification of the internucleosidic phosphate linkage barely change the overall trend of properties but usually had a negative impact on the thermodynamic stability of complexes. Fully modified phosphorothioate $d(T_2G_2T_2)$

sequence destabilised G4s but kept the fourth-order of reaction for the assembly.[237] Methyl phosphonates,[209] historically the first charge-neutral phosphate mimic (as PNA has no sugar-phosphate backbone), showed strong thermal destabilisation and so slow kinetics that in some cases, no formation of tetramolecular G4 could be observed.[150] Introduction of mesyl phosphoramidate groups (μ-modification) into internucleotidic position destabilised parallel tetramolecular G4s, but the preference order for monovalent cations for both association and thermal stability were the same as for the native sequences. Our results on the thermal stability of G4s formed by μ- or PG-modified DNA strands are in line with previous observations for methyl phosphonate[209] and PNA[217] analogues, when strong destabilisation was observed for tetramolecular assemblies. In contrast, the use of zwitterionic N+ modifications stabilised the G4s considerably ($\Delta T_{1/2}$ = + 8 °C and + 16 °C in comparison with native $[TG_4T]_4$ and $[PG\text{-}TG_4T]_4$, respectively). Stabilisation of $[N{+}TG_4T]_4$ and is $[PG\text{-}TG_4T]_4$ dependent on the radii of monovalent cations and follows the established pattern of $Li^+ < Na^+ < K^+$, and their dissociation process is independent of ionic strength.

Generally, the degree of stabilisation as well as the association rate of G4s are dependent on the radii of monovalent cations and follow a preference order of $Li^+ < Na^+ < K^+$.[150] It is expected that G4 formation by neutral DNA strands should be independent of ionic strength but should follow the same preference order for monovalent cations for both association and thermal stability. Such behaviour was observed for G4s formed by PNA.[217] In our case, the dependence of association on monovalent cations for PG-modified strands is reversed, i.e. it is faster in the presence of Na^+ than in the presence of K^+. However, the zwitterionic TG_4T sequence with N+ modification exhibited different properties. The assembly of $N{+}TG_4T$ is faster in Na^+ ions at the beginning of the association process, whereas it becomes faster in K^+ ions after association exceeds 80%, indicating that $N{+}TG_4T$ in Na^+ solutions experience a multi-step association process. The melting curves for $[N{+}TG_4T]_4$ derived from CD melting and annealing experiments at 260 and 295 nm showed that non-equilibrium melting temperature values ($T_{½}$) were different ($T_{½}$ = 65 °C at 262 nm and $T_{½}$ = 60 °C at 295 nm, see Fig. D.8 in Appendix D). Also, the annealing process at 262 nm was faster than that at 295 nm, which also indicate that $N{+}TG_4T$ exists as several G4 structures in solution. The association of PG- and $N{+}TG_4T$ is independent of salt concentration.

Variation in strand concentration had a marginal effect on $N{+}TG_4T$ G4 assembly, whereas formation of other $[TG_4T]_4$ complexes studied is highly dependent on strand

concentration: the higher the DNA concentration, the faster G4 formation. This property is reflected in the order of reaction being in the range of 1.1 – 1.4 for N+TG$_4$T and ~ 2 – 4 for other species.

Tetramolecular G4s were believed to be unable to fold in Li^+-containing solutions.[231] However, N+TG$_4$T can form G4 in Li^+ containing buffer even at 10 μM strand concentration, which has never been reported for native G4s. There are two possibilities how N+ modification can stabilise G4 formation leading to the assembly of TG$_4$T at low strand concentrations and in Li^+ containing buffer. The first and the most plausible explanation is that the quaternary nitrogen of the N+ linker can interact with the oxygen atom bearing a negative charge of the neighbouring N+ modification and in this way stabilise the assembly. However, this interaction will depend on the stereochemistry of neighbouring phosphates leading to the formation of complexes at a different speed. This agrees with our kinetic experiments. The second explanation relies on NH_4^+-induced stabilisation of G4 assemblies by minimising the Coulombic repulsion between oxygen atoms of carbonyls in the core of G4.[238] However, according to molecular modelling, the length of N+ modification between the phosphate and the quaternary amine is 8.11 Å (Fig. D.7B in Appendix D), which is too short to reach the centre of the G-tetrad, because the distance between the phosphate of the DNA backbone and the guanine O6 in the centre of G-tetrad is 9.14 Å in [TG$_4$T]$_4$ crystal structure (pdb: 2O4F, Fig. D.7A in Appendix D).[239] In addition, we observed the formation of G4 by N+TG$_4$T in Li^+ containing buffer but not in water which supports the first explanation as the presence of Li^+ was still required to stabilise G-tetrads.

According to SE-HPLC analysis in Na^+ buffer, PG-ONs existed as complexes of various molecularities in solution, which complicated the analysis of G4 formation. N+TG$_4$T, on the other hand, showed one single peak in SE-HPLC similarly to the native TG$_4$T tetramer, confirming the formation of tetramolecular G4. Invasion of PG-G4s into preformed native G4s was observed using both SE-HPLC and 20 % native PAGE. However, the evaluation of the mixture of [N+TG$_4$T]$_4$ and [TG$_4$T{FAM}]$_4$ by 20% native PAGE showed that although the mixed G4s were formed, the invasion was extremely slow, probably due to the high thermal stability of [N+TG$_4$T]$_4$.

5.7. Conclusion

To conclude, we evaluated for the first time the properties of chemically modified G-rich DNA sequences bearing either negatively charged mesyl phosphoramidate (μ) or charge-neutral phosphoryl guanidine (PG) or zwitterionic 4-(trimethylammonio)butylsulfonyl phosphoramidate (N+) groups instead of native phosphates. Various biophysical methods confirmed the formation of G4-assemblies of lower (μ- and PG-TG_4T) and higher (N+TG_4T) thermal stabilities relative to the unmodified $[TG_4T]_4$. G-rich μ-TG_4T retaining the negative charge behaved similarly to the native G4s, whereas ONs with a different type of charge-neutral modifications such as uncharged PG and zwitterionic N+ resulted in considerably different properties during G4 formation. This study highlights the importance of phosphates for the G4 formation and necessitates further investigation of G-rich strands possessing various phosphate mimics. We observed that neutral PG-modified DNA strands had a faster association in Na^+ than in K^+, which was independent of ionic strength. PG-ONs could also invade the pre-assembled native G4s forming mixed-G4s of variable stoichiometry. We also demonstrated that a fully phosphate modified N+ G-rich sequence does not easily invade into the native G4 complex suggesting the possible creation of phosphate modified nucleic acids that are orthogonal to the native DNA and RNA. Increased stability of N+TG_4T G4s and the fast association that is independent of ionic strength are useful properties for life and material science applications. It opens new avenues for designing novel chemical probes targeting enzymes and proteins interacting with genomic G-rich sequences.

Chapter 6. Summary and future directions

6.1 Summary

This Thesis aims to design novel probes for targeting native DNA assemblies using chemically modified nucleic acid analogues. Chapter 2 described the application of *ortho*-TINA modified duplexes as invading probes for targeting native double-stranded DNA (dsDNA) through dual duplex invasion (DDI). Chapter 3 dealt with the improvement of the kinetics of DDI using a zwitterionic N+ phosphate moiety. Chapter 4 further investigated the changes in the property induced by N+ phosphate modification in the context of a parallel DNA triplex and antiparallel duplexes with complementary DNA and RNA. Finally, three types of phosphate modifications were used for the evaluation of G-quadruplex (G4) assembly, which is detailed in Chapter 5.

In Chapter 2, we reported the use of *ortho*-TINA monomers in the duplex that fulfilled the requirement of DDI. Native duplex can be targeted by *ortho*-TINA probes in a sequence-specific manner. However, the slow kinetics of DDI against long duplexes, especially for probes with high G/C content, possibly limits its further applications. Several strategies, including the introduction of tails at each end of invasion probes, were attempted, but the improvement of kinetics has not been achieved. The introduction of tails at the 3′-end resulted in a loss in sequence-specificity.

The difficulty in invading into long duplexes has been a common issue for dsDNA invading probes. We designed and synthesised sulfonate azide carrying a quaternary amine (N+) and used it to modify phosphates in the DNA backbone, which was described in detail in Chapter 3. The negative charge on the phosphate was neutralised by the positive charge in the quaternary amine, forming a zwitterionic N+ phosphate moiety. The introduction of N+ modification into the invading probe did not initiate DDI, which highlighted the importance of *ortho*-TINA monomers in the DDI process. Various combinations of N+ and *ortho*-TINA monomers in the duplex were designed and evaluated, yet the improvement of kinetics was marginal. Only in one case, which was a combination of three N+ modifications and four *ortho*-TINA monomers in one strand and four *ortho*-TINA monomers in the complementary strand, achieved a faster invasion against long native dsDNA.

In Chapter 4, we conducted a series of evaluations of N+ modified ONs in the formation of DNA parallel triplexes, antiparallel duplexes with complementary DNA and RNA. For comparison, a *p*-tosyl (Ts-) group was introduced into the phosphate, which did not change the overall charge of the sequence. Incorporation of both N+ and Ts-

modifications led to the formation of thermally stabilised parallel triplexes, regardless of the number of modifications incorporated. For both N+ and Ts- modified ONs, antiparallel duplexes with complementary RNA were stabilised (except for ONs with modification right in the middle of the sequence) to a greater extent than with complementary DNA. Introduction of N+ and Ts- modifications enhanced resistance of ONs towards snake venom phosphodiesterase. Cell uptake studies showed that without any transfection reagents used, Ts-ONs diffused into the nucleus of the cells, whereas N+ONs were concentrated in vesicles in the cytosol.

In Chapter 5, we evaluated the properties of G4 assemblies formed by chemically modified G-rich sequences, in which all phosphates in TG_4T sequence were substituted with either mesyl (μ)- modifications that retained negative charge, or phosphoryl guanidine group (PG-modifications) that was charge-neutral, or N+ modification that resulted in a zwitterionic phosphate mimic. PG- and N+ modifications changed tetramolecular parallel G4 assembly dramatically, which highlighted the importance of negatively charged phosphates in the formation of G4s.

This Thesis provides new insights for the targeting of native DNA assemblies using chemically modified nucleic acid scaffolds. Future research should focus on the applications of chemically modified ONs bearing the same or similar modifications.

6.2. Future directions

6.2.1. Incorporation of minor groove binders into *ortho*-TINA probe

Despite the fact that the invasion products of DDI are thermodynamically preferable (ΔG_{298} is negative), their formation through DDI is kinetically slow. Our attempt to improve the kinetics of DDI using tails at the 5′- or 3′-end was not successful. An alternative solution might be introducing a minor groove binder (MGB) into the *ortho*-TINA probe, and as a result, the DDI process will be initiated by the binding of MGB to the minor groove of the target native duplex. The intermediate state of DDI might be changed in this way, and therefore the kinetics of DDI might be improved.

6.2.2. Design of new modifications with a positive charge

Improving the kinetics of DDI using neutral or positively charged probes is attractive. It has been reported that the neutral DNA backbone can help increase enzymatic resistance as well as cell uptake.[180] PNA with positive charges destabilised PNA/PNA duplexes, whereas the thermal stability of PNA/DNA duplexes was increased, leading to increased

kinetics of DDI.[102, 107] Despite the fact that our N+ modification did not lead to faster kinetics of DDI, the introduction of other positively charged moieties into the DNA phosphate backbone could serve as a promising solution for improving the kinetics of DDI using *ortho*-TINA probe. As mentioned previously, a bulkier N+ modification could be obtained using triethylamine instead of trimethylamine (I and III in Fig. 6.1). This might lead to a less thermally stable invading probe. Also, instead of 1,4-butane sultone, one can use 1,3-propyl sultone (1,2-oxathiolane 2,2-dioxide) to obtain a shorter linkage between the quaternary amine and the phosphate (II and III in Fig. 6.1). By introducing two positive charges (IV in Fig. 6.1) into phosphate, we can obtain a positively charged phosphate mimic, which will increase the thermal stability of the invasion products to a large extent, which may drive the reaction (DDI) to occur.

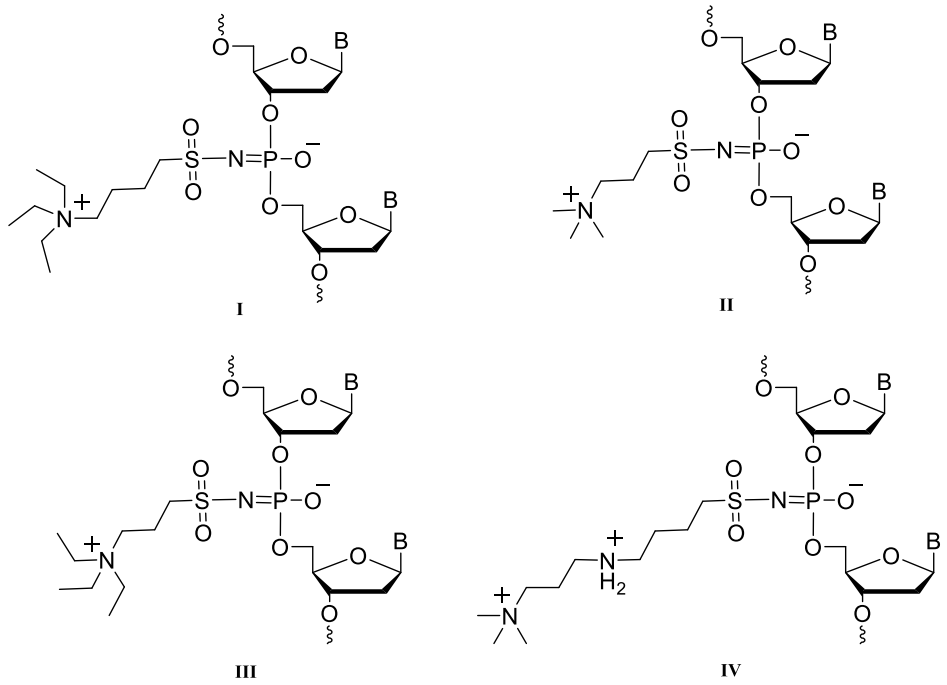


Figure 6.1. Possible phosphate modifications bearing one or more positive charges.

6.2.3. Other applications of N+ modification

The incorporation of N+ modification into the DNA backbone was not able to improve the kinetics of DDI. However, it is worth trying other applications of this N+ modification because it has a number of advantages, including the ease of preparation, the flexibility of introduction into the sequences, as well as increased enzymatic resistance and higher binding affinity towards complementary RNA. N+ modified RNA or 2′-*O*Me RNA might be suitable for applications in antisense technology.[240] Other phosphate modifications, such as Ts- or μ- modifications,[182, 183] can also be evaluated in various applications.

6.2.4. Combination of *ortho*-TINA monomer with phosphate modifications

It has been reported that INA monomers were successfully applied in sequence-unrestricted targeting of telomeric DNA.[241] It is reasonable to predict that DNA sequences with an *ortho*-TINA monomer that mimics the properties of an INA monomer might also be used to target telomeric DNA *in vivo*. Mesyl phosphoramidate was introduced into AOs as an alternative to phosphorothioates with significantly higher nuclease resistance, lower toxicity and good cell permeability.[242]

Based on the above-mentioned results, probes with the combination of *ortho*-TINA monomers and μ- modified phosphates should be able to penetrate cell membranes and perform excellent sequence-specific targeting of telomeric DNA.

6.2.5. Biosensor based on DDI

The development and application of a lateral flow biosensor (LFB) to detect a specific DNA sequence have gained significant interest recently as it minimises the expensive cost of personnel and instrumentation for PCR,[243, 244] loop-mediated isothermal amplification (LAMP),[245] and DNA microarrays.[246] We assume that an LFB for specific, sensitive, and visual detection of pathogenic DNA based on DDI might be feasible. Super-paramagnetic particles (SPMP) can be conjugated to *ortho*-TINA probes as a signal indication. SPMP has no background signal, it exhibits only diamagnetism, a linear magnetic behaviour, which does not disturb nonlinear magnetisation measurement.[247] Moreover, SPMP is small enough to penetrate extracellular and intracellular barriers with biological safety, and therefore could be used *in vivo*.[247, 248] Finally, SPMP have low detection limits, which can be detected at the pM range. It can produce a steady magnetic signal that can be entirely captured by a device such as a magnetic reader, thus improving the sensitivity of the LFB and allowing the labelled analyte to be quantitatively detected.[249]

We assume that the sequence specificity of DDI provides us with the possibility to avoid DNA amplification steps so that this LFB can be used by untrained personnel (e.g., a farmer in the field).

As shown in Figure 6.2, the LFB could be constructed based on a nitrocellulose membrane. The sample pad and conjugation pad are attached at the bottom side of the membrane, whereas the absorption pad is attached at the top. *ortho*-TINA probe (ssDNA in Set 1) conjugated with SPMP is immobilised on the conjugation pad. On the test line

(TL), streptavidin-biotinylated dsDNA (*ortho*-TINA duplex in Set 2) is immobilised. Streptavidin-biotinylated ssDNA complementary to the ssDNA used in the conjugation pad (Set 1) is immobilised on the control line (CL) to capture excess amounts of SPMP-DNA conjugate. When the sample is applied on the sample pad and submerged in the developing solution, the sample in the solution will move towards the absorption pad due to the capillary force. If the sample contains a target pathogenic DNA, strand invasion of Set 1 duplex should occur initially at the conjugation pad, forming SPMP labelled sample. The sample with SPMP will keep moving forward to the test line, where the DDI with Set 2 duplex should occur, and part of the SPMP labelled sample is captured at the TL. An accumulation of SPMP at TL will generate a magnetic signal that can be detected by a magnetic reader. The excess amount of SPMP in a labelled sample is moving to the CL and captured by the ssDNA on CL through hybridisation, giving a second magnetic signal. Only when the magnetic signal is detected on both TL and CL lines, the result is 'positive/yes'; otherwise the result is 'negative/no'.

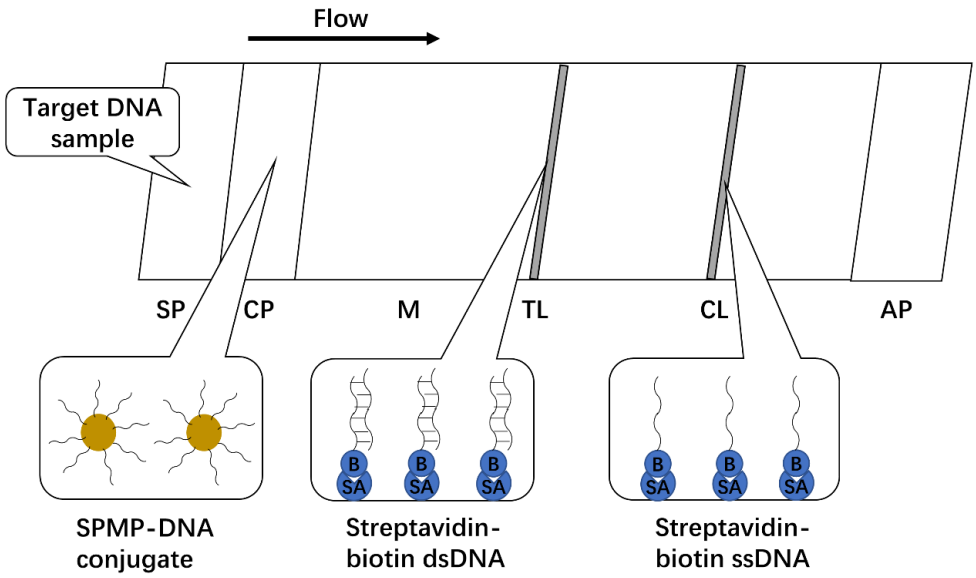


Figure 6.2. Design of the portable device for detection of pathogenic DNA based on LF assay. Arrow shows the direction of flow. SP: sample pad, CP: conjugation pad, M: membrane, TL: test line, CL: control line, AP: absorption pad, SA: streptavidin, B: biotin.

6.2.6. Neutral G-rich sequences (G4s) for labelling or fixing G4s *in vivo*.

Neutral G4s that are formed by charge neutral G-rich sequences have several attractive features. For example, PG-G4s that invade into the pre-formed native G4s and N+ modifications can stabilise native G4s. Moreover, assembly of both PG- and N+G4s are less dependent on ionic strength, and more stable against endo- and exo-nucleases. Based on the results we obtained, we can conclude that the modifications proposed in Figure 6.1 could also bring dramatic changes to G4 formation. These properties make neutral G4s promising tools for labelling or fixing G4 structures *in vivo*.

Chapter 7. Appendix

Appendix A for Chapter 2

1. UV/Vis Spectroscopy

UV-Vis spectra were recorded using Cary 100 Bio UV-Vis spectrometer using quartz cuvettes with 1 cm pathlength (1 mm for G4 study) and a 2 × 6 multicell block with a Peltier temperature controller.

2. Circular Dichroism (CD) Spectroscopy

CD spectra were recorded using Chirascan CD spectrophotometer (150 W Xe arc) from Applied Photophysics with a Quantum Northwest TC125 temperature controller. CD spectra (average of at least 3 scans) were recorded between 220 and 450 nm with 1 nm intervals, 120 nm/min scan rate and 1 cm pathlength followed by subtraction of a background spectrum (buffer only). CD spectra of ONs were recorded at 2 μM strand concentration in 10 mM Na-phosphate buffer (140 mM NaCl, 0.1 mM Na_2-EDTA, pH 7.0). CD spectra of G-quadruplexes were recorded at 100 μM strand concentration in 10 mM Li-cacodylate buffer (100 mM NaCl or KCl, pH 7.0).

3. Fluorescence Spectroscopy

Fluorescence spectroscopy studies were performed using Fluorescence Spectrometer LS55 from PerkinElmer, or FluoroMax-4 Spectrofluorometer from HORIBA Scientific. Fluorescence emission spectra were recorded for all TINA-containing duplexes between 380 and 600 nm upon excitation at 373 nm. To make sure that the most intense fluorescent signal is below limitation (1000 a.u. for LS55 and 2,000,000 counts for FluoroMax-4), slit widths were set differently (emission slit = 10 nm, excitation slit = 0 nm for LS55 and emission slit = 3 nm, excitation slit = 3 nm for FluoroMax-4, respectively). Time course spectra for Set 1 were recorded using LS55. Fluorescence intensity at 405 nm was recorded every five min after addition of 1.0 eq of pre-formed native duplex **C1** to *ortho*-TINA-duplexes (**P1**, **P2** and **P3**) over 90 min, without sample shaking. Data for Set 2 was recorded using FluoroMax-4. Fluorescence intensity at 520 nm was recorded every ten min after addition of 1.0 eq of pre-formed native duplex **C2** to *ortho*-TINA-duplexes (**P4**, **P5** and **P6**) over 240 min, without sample shaking. Extracted data was then converted into percentage of invasion based on fluorescence spectra before (0 %) and 24 h after

invasion (100 %). Buffer was the same as used for UV-Vis and CD measurements, concentration of ONs is described in Tables and Figures.

4. Denaturing and native PAGE

Purity of the synthesised ONs was assessed using 20% denaturing PAGE. Gels were prepared in 1 × TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM Na_2-EDTA, pH 8.0) in 7 M urea with 0.5 mm thickness, 17.5 × 14.5 cm^2 (19: 1 acrylamide/bis-acrylamide ratio). Samples were incubated at high temperature (typically 90 °C) in 7 M urea (7.5 µL) for 5 min to disrupt high-order assemblies. 1 × TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM Na_2-EDTA, pH 8.0) was used as a running buffer.

For non-denaturing PAGE, gels were prepared in 1 × TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM Na_2-EDTA, pH 8.0) supplemented with 150 mM NaCl (the same buffer was used as a running buffer) with 0.5 mm thickness, 17.5 × 14.5 cm^2 (19: 1 acrylamide/bis-acrylamide ratio). ONs were prepared at 50 µM strand concentration in a running buffer.

All gel electrophoresis was performed at 15 W per gel at 4 °C to avoid overheating. After the electrophoresis, gels were stained with 5% Stains-All® solution in water/formamide (1: 1) for 5 – 10 min and then destained in H_2O until complete fading of the dye from the gel background.

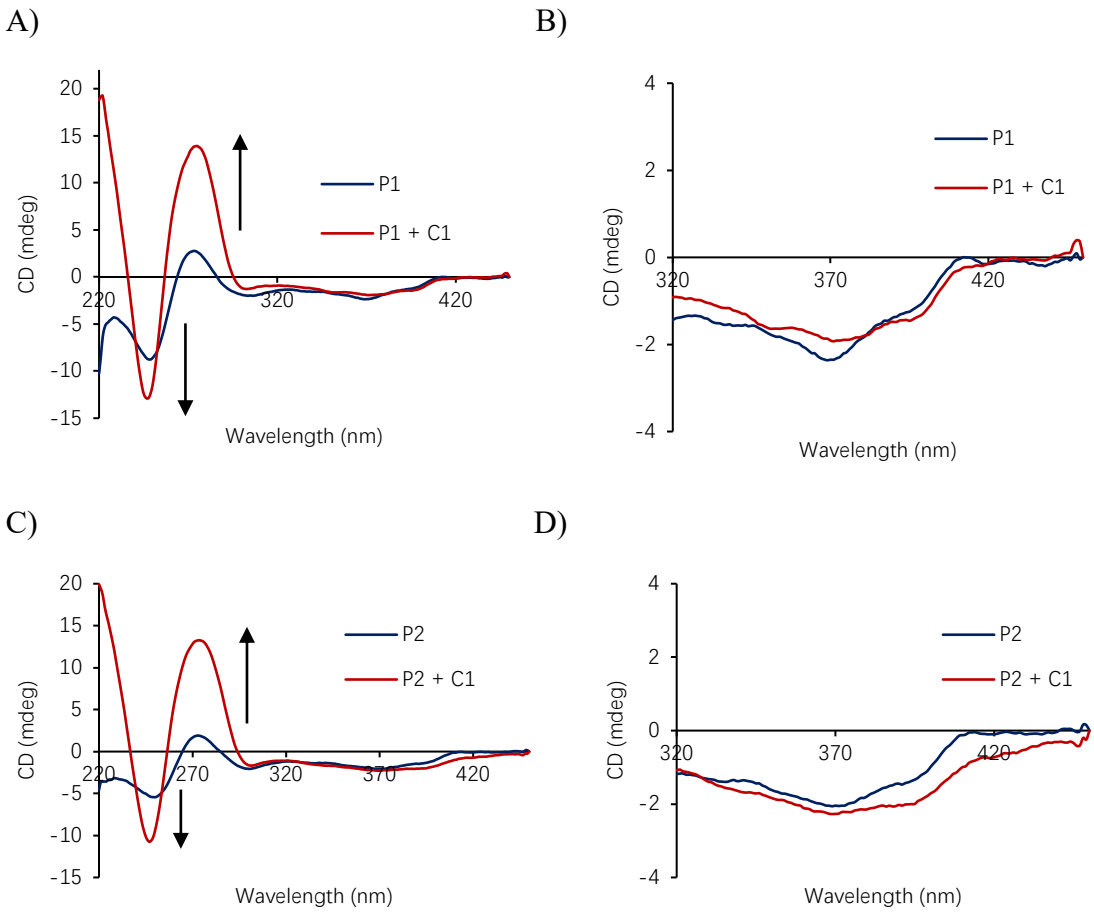


Figure A.1. CD spectra of *ortho*-TINA duplexes (A: **P1**, and C: **P2**) before and after addition of the pre-annealed DNA target **C1** (all strands 2.0 µM); B) CD spectra of A) in the TINA absorbance region (320 to 450 nm); D) CD spectra of C) in the TINA absorbance region. Arrows indicate the direction of CD signal change.

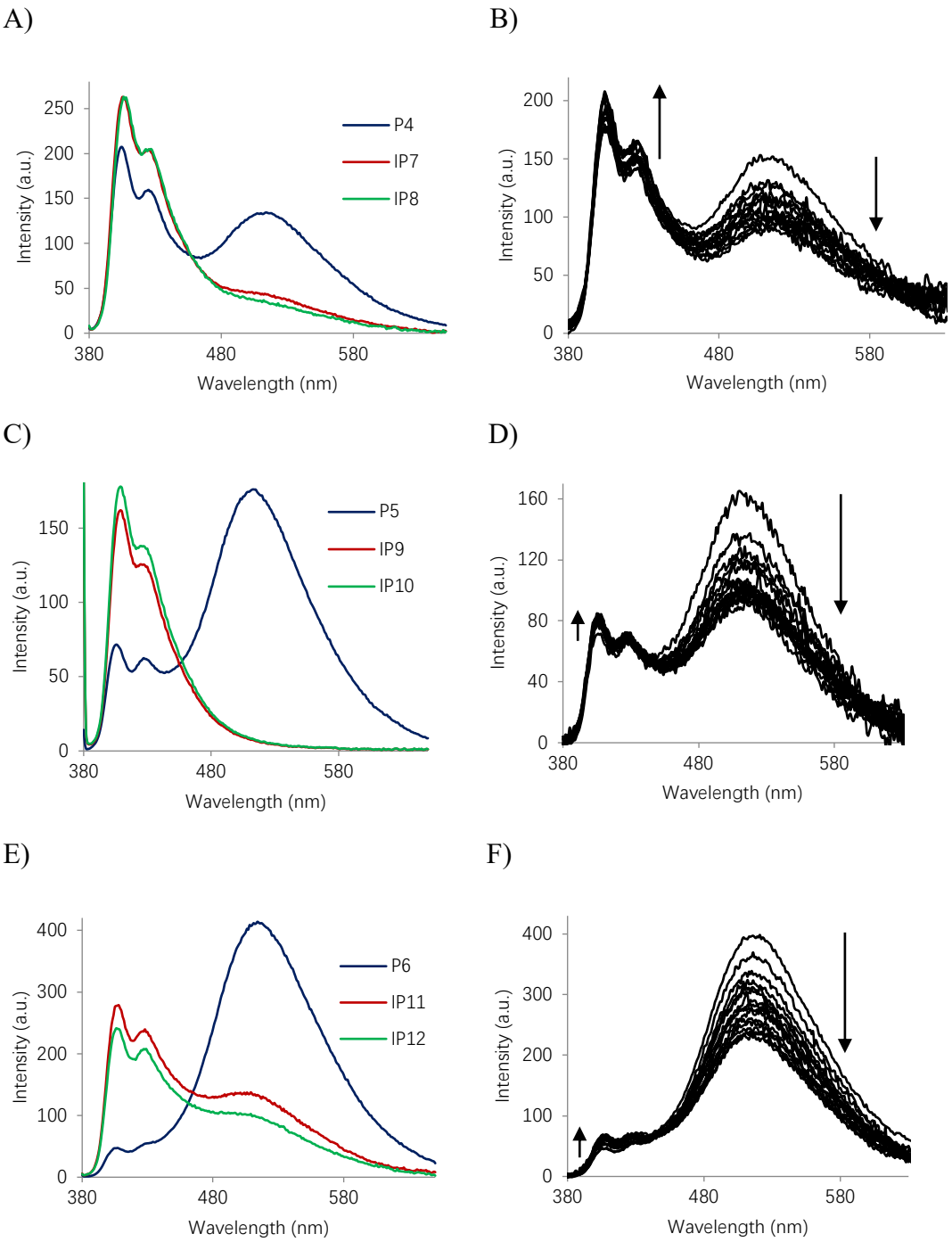


Figure A.2. Steady-state fluorescence emission spectra of duplexes with *ortho*-TINA pairs and corresponding invasion products in Set 2. A: **P4** and **IP7** and **IP8**, C: **P5** and **IP9** and **IP10**, E: **P6** and **IP11** and **IP12**; Fluorescence spectra upon addition of the pre-annealed **C2** to the pre-annealed duplexes with *ortho*-TINA pairs (B: **P4**, D: **P5**, F: **P6**). All spectra were recorded using FluoroMax-4 over 2 h at 25 °C, each spectrum was recorded every 10 min; all samples were at 2.0 µM strand concentration in 10 mM Na-phosphate buffer (140 mM NaCl, 0.1 mM Na_2-EDTA, pH 7.0). For convenience the values in the y-axis (fluorescence intensity) were divided by 10^2. Arrows indicate the direction of signal change of fluorescent monomer and excimer signals from 0 to 2 h.

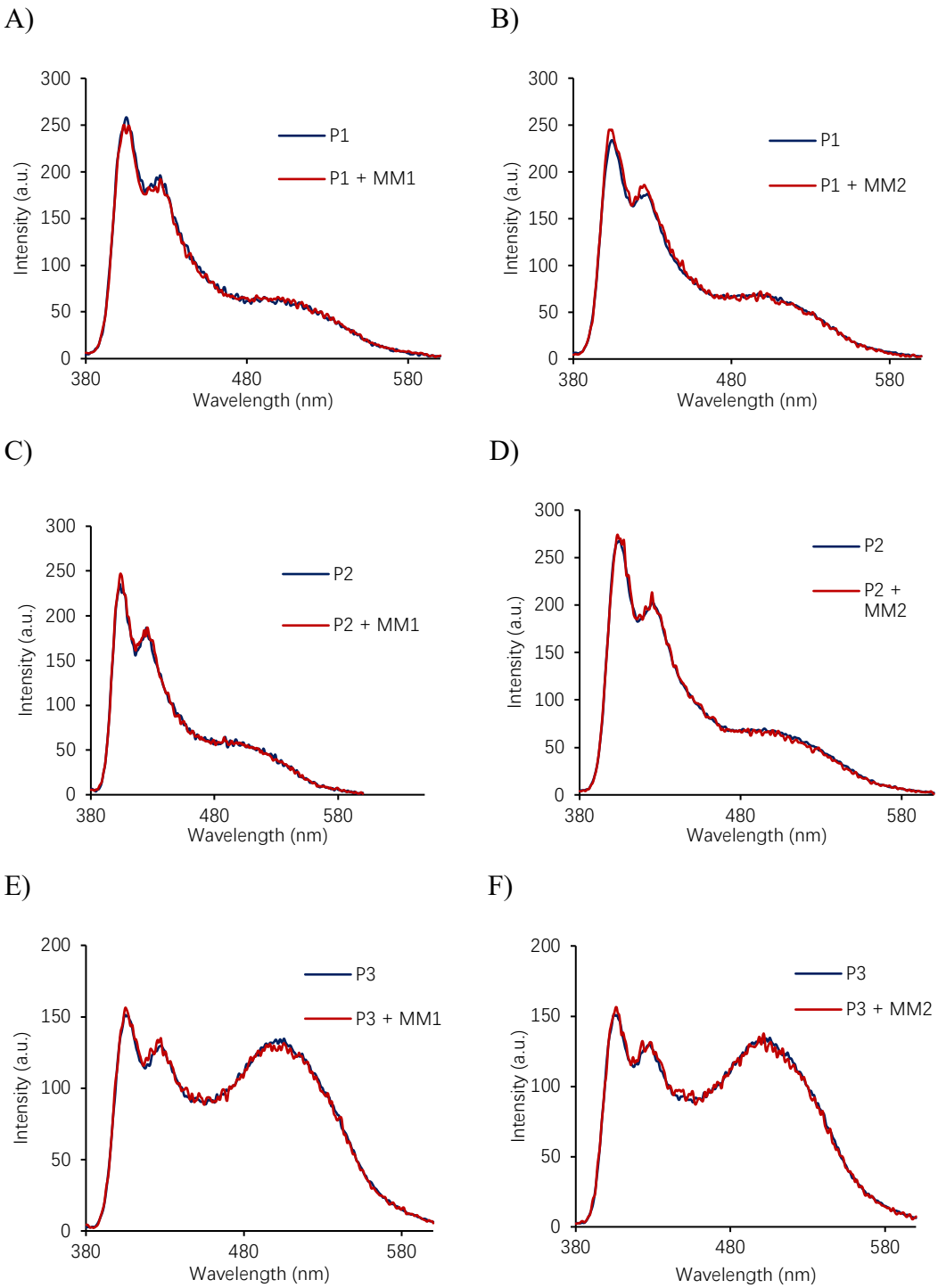


Figure A.3. Fluorescence emission spectra of *ortho*-TINA duplexes (A: **P1**, C: **P2**, E: **P3**) before and after addition of pre-annealed mis-matched duplex **MM1** (all strands 2.0 μM) over 40 min; Fluorescence emission spectra of *ortho*-TINA duplexes (B: **P1**, D: **P2**, F: **P3**) before and after addition of pre-annealed partially complementary duplex **MM2** (all strands 2.0 μM) over 40 min. All spectra were recorded using LS55 fluoremeter; all strands were 1.0 μM in 10 mM Na-phosphate buffer (140 mM NaCl, 0.1 mM Na_2-EDTA, pH 7.0).

As shown in Fig. A.3, for both mis-matched **MM1** and partially complementary **MM2** duplexes, no change in signal intensity was observed during 40 min of incubation, which indicated that invasion only happens between *ortho*-TINA probes and complementary native duplexes.

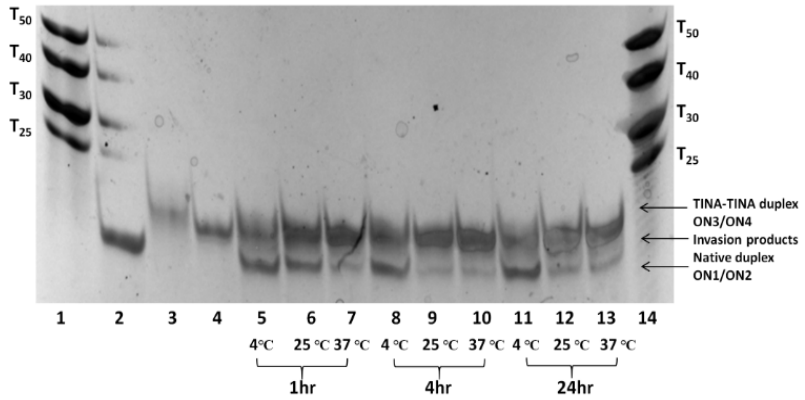


Figure A.4. 20% Native PAGE showing invasion of **P1** into **C1** in 1 × TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM Na_2-EDTA, pH 8.0) supplemented with 150 mM NaCl at 50 μM strand concentrations at different temperatures and times. Lanes are 1: ladder; 2: **C1**; 3: **P1**; 4: **IP1**; 5: **P1** + **C1**, 1 hr, 4 °C; 6: **P1** + **C1**, 1 hr, 25 °C; 7: **P1** + **C1**, 1 hr, 37 °C; 8: **P1** + **C1**, 4 h, 4 °C; 9: **P1** + **C1**, 4 h, 25 °C; 10: **P1** + **C1**, 4 h, 37 °C; 11: **P1** + **C1**, 24 h, 4 °C; 12: **P1** + **C1**, 24 h, 25 °C; 13: **P1** + **C1**, 24 h, 37 °C; 14: ladder.

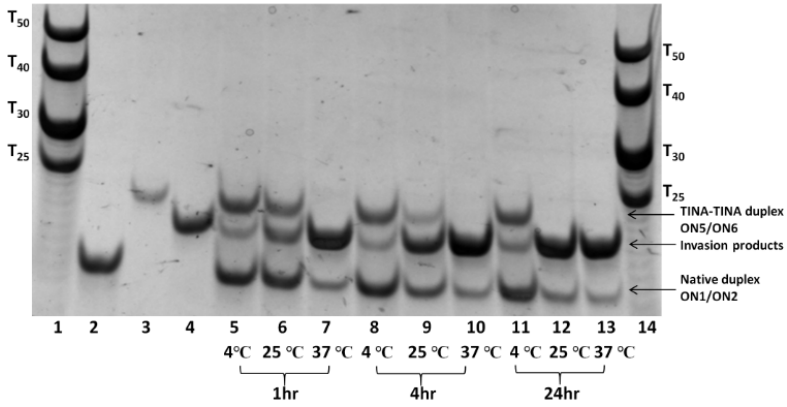


Figure A.5. 20% Native PAGE showing invasion of **P2** into **C1** in 1 × TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM Na_2-EDTA, pH 8.0) supplemented with 150 mM NaCl at 50 μM strand concentrations at different temperatures and times. Lanes are 1: ladder; 2: **C1**; 3: **P2**; 4: **IP3**; 5: **P2** + **C1**, 1 hr, 4 °C; 6: **P2** + **C1**, 1 hr, 25 °C; 7: **P2** + **C1**, 1 hr, 37 °C; 8: **P2** + **C1**, 4 h, 4 °C; 9: **P2** + **C1**, 4 h, 25 °C; 10: **P2** + **C1**, 4 h, 37 °C; 11: **P2** + **C1**, 24 h, 4 °C; 12: **P2** + **C1**, 24 h, 25 °C; 13: **P2** + **C1**, 24 h, 37 °C; 14: ladder.

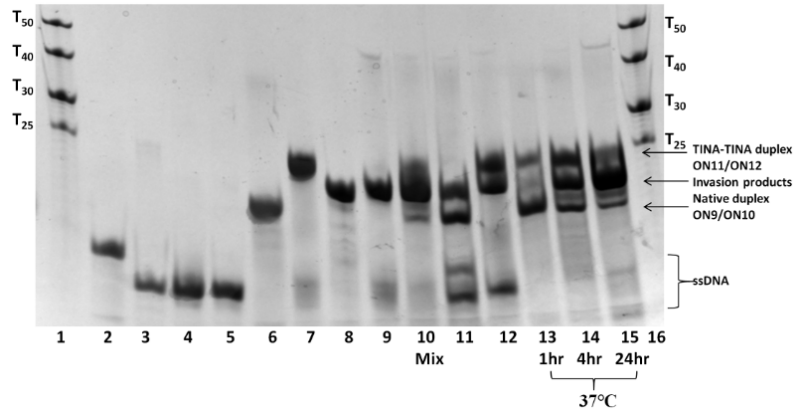


Figure A.6. 20% Native PAGE showing invasion of **P4** into **C2** in 1 × TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM Na_2-EDTA, pH 8.0) supplemented with 150 mM NaCl at 50 µM strand concentrations. Lanes are 1: ladder; 2: ON9; 3: ON10; 4: ON11; 5: ON12; 6: **C2**; 7: **P4**; 8: **IP7**; 9: **IP8**; 10: Mixture of ON9, ON10, ON11, ON12, heated at 90 °C for 5 min and then slowly cooled down to RT; 11: **C2** + ON11, 24 h, 37 °C; 12: **P4** + ON9, 24 h, 37 °C; 13: **P4** + **C2**, 1 hr, 37 °C; 14: **P4** + **C2**, 4 h, 37 °C; 15: **P4** + **C2**, 24 h, 37 °C; 16: ladder.

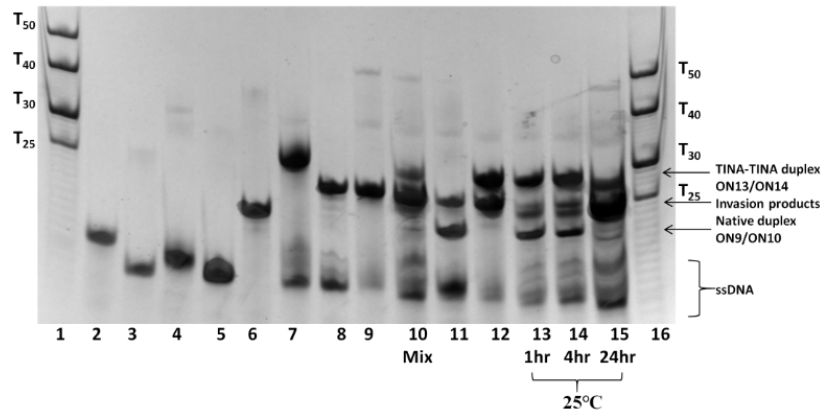


Figure A.7. 20% Native PAGE showing invasion of **P5** into **C2** in 1 × TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM Na_2-EDTA, pH 8.0) supplemented with 150 mM NaCl at 50 µM strand concentrations. Lanes are 1: ladder; 2: ON9; 3: ON10; 4: ON13; 5: ON14; 6: **C2**; 7: **P5**; 8: **IP9**; 9: **IP10**; 10: Mixture of ON9, ON10, ON13, ON14, heated at 90 °C for 5 min and then slowly cooled down to RT; 11: **C2** + ON13, 24 h, 25 °C; 12: **P5** + ON9, 24 h, 25 °C; 13: **P5** + **C2**, 1 hr, 25 °C; 14: **P5** + **C2**, 4 h, 25 °C; 15: **P5** + **C2**, 24 h, 25 °C; 16: ladder.

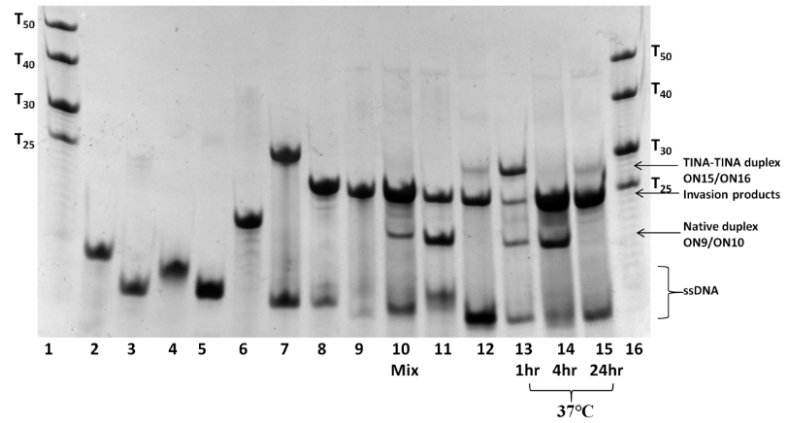


Figure A.8. 20% Native PAGE showing invasion of **P6** into **C2** in 1 × TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM Na_2-EDTA, pH 8.0) supplemented with 150 mM NaCl at 50 μM strand concentrations. Lanes are 1: ladder; 2: ON9; 3: ON10; 4: ON15; 5: ON16; 6: **C2**; 7: **P6**; 8: **IP11**; 9: **IP12**; 10: Mixture of ON9, ON10, ON15, ON16, heated at 90 °C for 5 min and then slowly cooled down to RT; 11: **C2** + ON15, 24 h, 37 °C; 12: **P6** + ON10, 24 h, 37 °C; 13: **P6** + **C2**, 1 hr, 37 °C; 14: **P6** + **C2**, 4 h, 37 °C; 15: **P6** + **C2**, 24 h, 37 °C; 16: ladder.

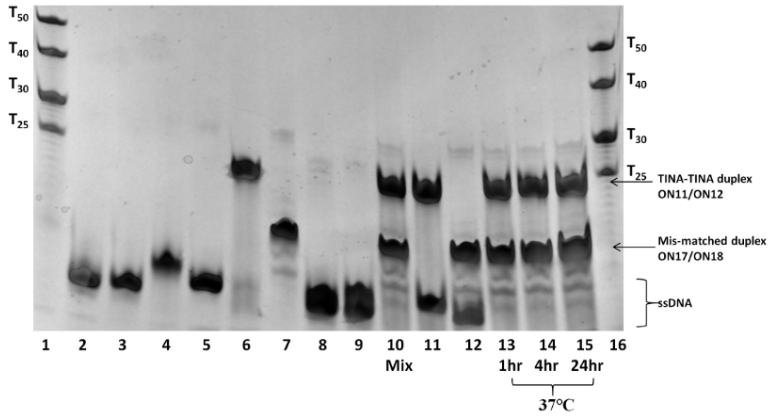


Figure A.9. 20% Native PAGE showing no invasion between **P4** and mis-matched duplex **MM1** in 1 × TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM Na_2-EDTA, pH 8.0) supplemented with 150 mM NaCl at 50 μM strand concentrations. Lanes are 1: ladder; 2: ON11; 3: ON12; 4: ON17; 5: ON18; 6: **P4**; 7: **MM1**; 8: ON11 + ON18; 9: ON17 + ON12; 10: Mixture of ON11, ON12, ON17, ON18, heated at 90 °C for 5 min and then cooled down to RT; 11: **P4** + ON18; 12: **MM1** + **ON12**; 13: **P4** + **MM1**, 1 hr, 37 °C; 14: **P4** + **MM1**, 4 h, 37 °C; 15: **P4** + **MM1**, 24 h, 37 °C; 16: ladder.

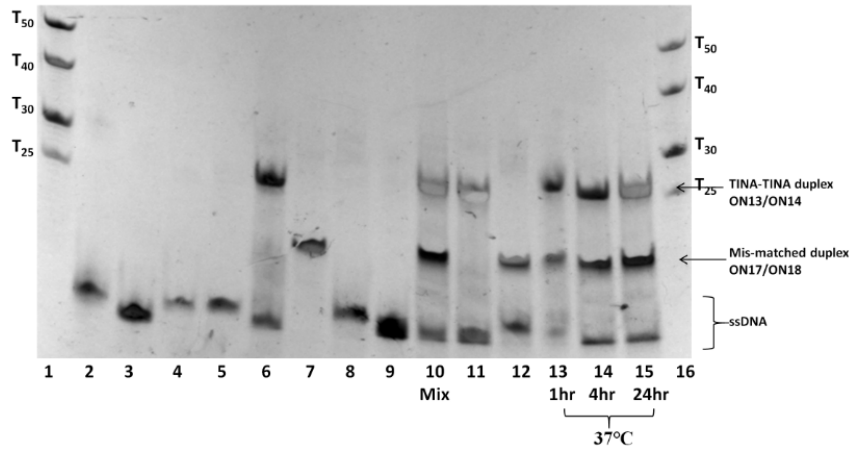


Figure A.10. 20% Native PAGE showing no invasion between **P5** and mis-matched duplex **MM1** in 1 × TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM Na_2-EDTA, pH 8.0) supplemented with 150 mM NaCl at 50 μM strand concentrations. Lanes are 1: ladder; 2: ON13; 3: ON14; 4: ON17; 5: ON18; 6: **P5** (excess of ON14); 7: **MM1**; 8: ON13 + ON18; 9: ON17 + ON14; 10: Mixture of ON13, ON14, ON17, ON18, heated at 90 °C for 5 min and then slowly cooled down to RT; 11: **P5** + ON18; 12: **MM1** + ON14; 13: **P5** + **MM1**, 1 hr, 37 °C; 14: **P5** + **MM1**, 4 h, 37 °C; 15: **P5** + **MM1**, 24 h, 37 °C; 16: ladder.

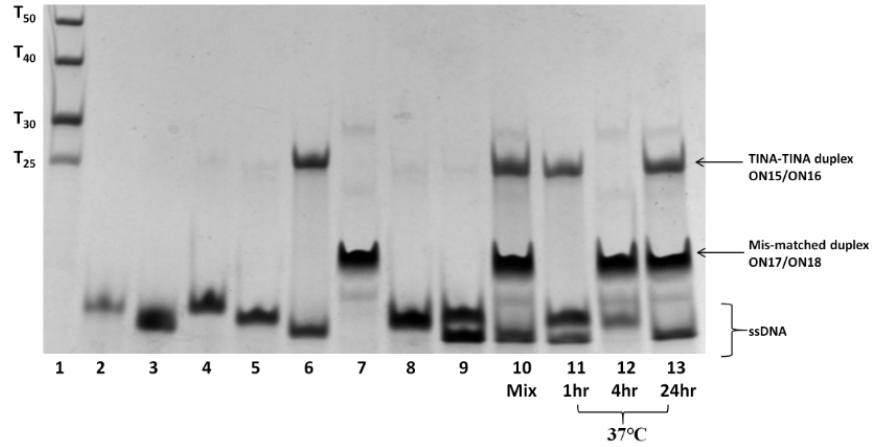


Figure A.11. 20% Native PAGE showing no invasion between **P6** and mis-matched duplex **MM1** in TBE buffer (pH 7.0) supplemented with 150 mM NaCl at 50 μM strand concentrations. Lanes are 1: ladder; 2: ON15; 3: ON16; 4: ON17; 5: ON18; 6: **P6** (excess of ON16); 7: **MM1**; 8: ON15 + ON18; 9: ON17 + ON16; 10: Mixture of ON15, ON16, ON17, ON18, heated at 90 °C for 5 min and then slowly cooled down to RT; 11: **P6** + ON17, 24 h, 37 °C; 12: **MM1** + ON15, 24 h, 37 °C; 13: **P6** + **MM1**, 24 h, 37 °C.

As shown in Fig. A.9, A.10, and A.11, for *ortho*-TINA duplexes in Set 2, no invasion occurrs when using a mis-matched duplex, as no interaction between *ortho*-TINA duplexes and mis-matched duplexes was observed after incubation for 24 h at 37 °C.

Table A.1. Hybridisation data for duplexes with tails at 5′-end and their free energy for recognition (ΔG^{298}_{rec}) of dsDNA target **LC1** at 298K.

Entry	Name	T_m /°C	ΔH^0 /kJ/mol	$T_{298}\Delta S^0$ / kJ/mol	ΔG_{298} / kJ/mol	ΔG^{298}_{rec} / kJ/mol
1	**LC1**	46	-661	-582	-79	
2	**P7**	47	-237	-187	-50	
3	**P8**	45	-336	-280	-56	
4	**P9**	51	-217	-165	-52	
5	**P10**	45	-228	-180	-48	
6	**IP13**	51	-489	-416	-73	-13
7	**IP14**	49	-469	-400	-69	
8	**IP15**	49	-419	-353	-66	-5
9	**IP16**	47	-586	-512	-74	
10	**IP17**	50	-433	-366	-67	-10
11	**IP18**	47	-543	-469	-74	
12	**IP19**	49	-428	-360	-68	-12
13	**IP20**	47	-554	-483	-71	

Appendix B for Chapter 3

NMR spectra were recorded on a 500 MHz Bruker Avance spectrometer equipped with a cryoprobe and processed with Topspin software (v. 2.1.8). The internal standard used in ^{1}H NMR spectra was $CDCl_3$ (δ: 7.26); in ^{13}C NMR the standard was $CDCl_3$ (δ: 77.0). IR spectra were recorded using Thermo Scientific Nicolet iS5 FT-IR spectrometer with Thermo Scientific iD7 ATR accessory and Thermo Scientific OMNIC Spectra software. ESI-MS spectra were recorded using Thermo Scientific Q-Exactive Focus Hybrid Quadrupole-Orbitrap Mass Spectrometer. Samples (4 μM, 5 μL) were injected via Dionex Ultimate 3000 HPLC system running at 0.1 mL/min CH_3OH. CD spectra were recorded using a Chirascan CD spectrophotometer (150 W Xe arc) from Applied Photophysics with a Quantum Northwest TC125 temperature controller.

1. Synthesis of N+ azide (compound 5 in Scheme 3.2)

4-Chlorobutane-1-sulfonyl chloride (2)

To a mixture of 1,2-oxathiane 2,2-dioxide (30 g, 220.3 mmol) and $SOCl_2$ (131.0 g, 80 ml, 1.1 mol), DMF (2.3 g, 2.4 mL, 30.8 mmol) was added dropwise and the mixture was stirred under nitrogen at 70 °C for 3 days. A second portion of $SOCl_2$ (0.5 eq) was added to the reaction mixture and stirred at 70 °C for 3 days. A third portion of $SOCl_2$ (0.25 eq) was added to the reaction mixture and stirred at 70 °C for 3 days. Reaction was monitored by ^{1}H NMR. After starting material was consumed, the reaction mixture was cooled to room temperature, concentrated *in vacuo* and diluted with toluene (30 mL). The mixture was co-evaporated with toluene several times and the residue was dried *in vacuo* overnight providing compound **2** as a yellow oil (28.6 g, 67.8%).

^{1}H NMR: 500 MHz, $CDCl_3$: δ, ppm 2.01 - 2.07 (m, 2H, *H*-3), 2.25 - 2.31 (m, 2H, *H*-2), 3.60 (t, J = 6.15 Hz, 2H, *H*-4), 3.72 (t, J = 7.55 Hz, 2H, CH_2, *H*-1).

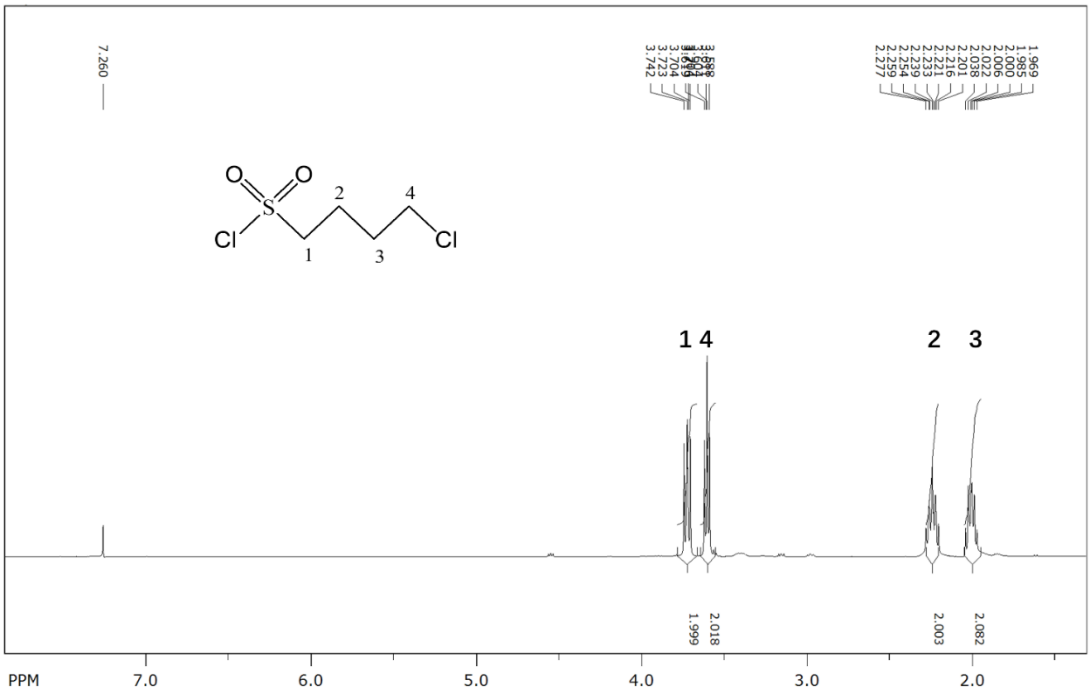


Figure B.1. ^{1}H NMR spectrum of 4-chlorobutane-1-sulfonyl chloride (**2**) in $CDCl_3$.

^{13}C NMR: 125.8 MHz, $CDCl_3$: δ = 21.97 (*C*-2), 30.07 (*C*-3), 43.50 (*CH₂Cl*), 64.53 (*C*-1).

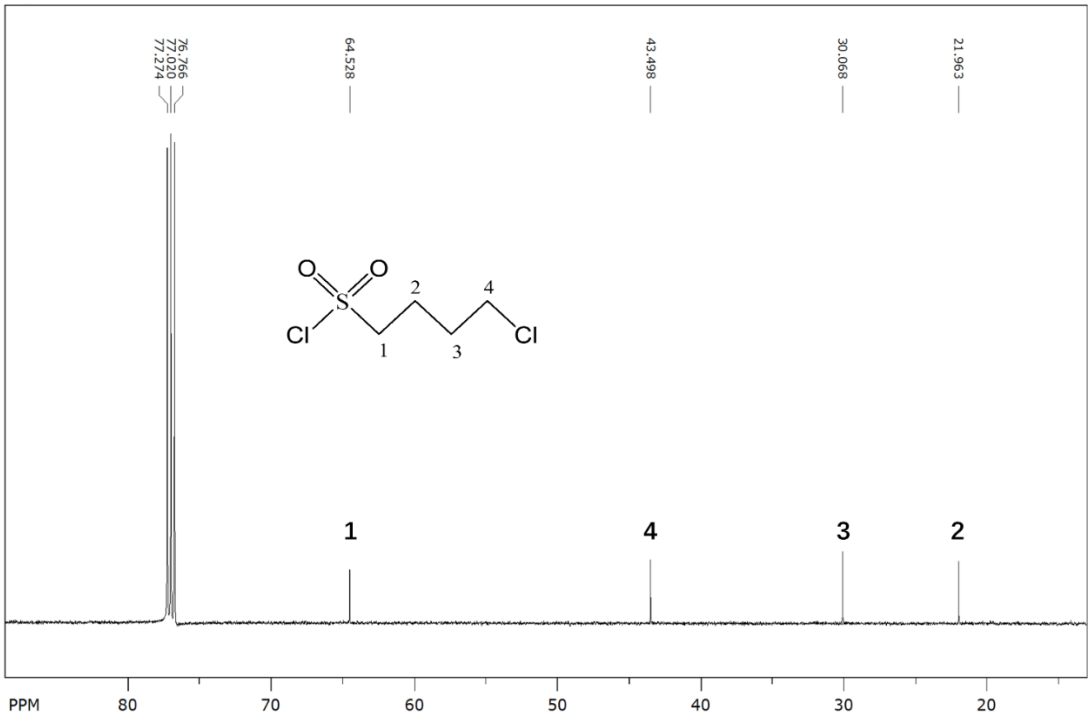


Figure B.2. ^{13}C NMR spectrum of 4-chlorobutane-1-sulfonyl chloride (**2**) in $CDCl_3$.

IR (ATR), cm^{-1}: ν (CH_2) 2963 - 2877; N (RSO_2-Cl) 1703; N (RSO_2) 1367.

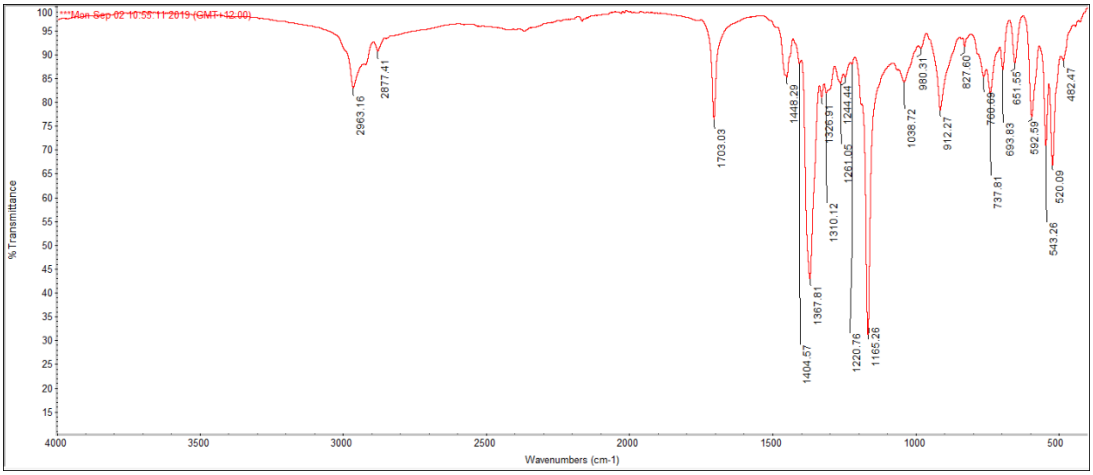


Figure B.3. IR spectrum of 4-chlorobutane-1-sulfonyl chloride (**2**).

4-Chlorobutane-1-sulfonyl azide (3)

NaN_3 (11.7 g, 179.3 mmol) and $Bu_4N^+HSO_4^-$ (60.9 g, 179.3 mmol) were dissolved in water (25 mL). Compound **2** (28.6 g, 149.4 mmol) dissolved in $CHCl_3$ (25 mL) was added to the aqueous phase with vigorous stirring at 0 °C (ice bath). Reaction was monitored using ^{1}H NMR and was finished within 2 h. Organic layer was separated and washed with sat. $NaHCO_3$ (3 × 25 mL), dried over $MgSO_4$, filtered, and evaporated *in vacuo*, yielding azide **3** as an orange oil (29.2 g, 80%).

^{1}H NMR: 500 MHz, $CDCl_3$: δ, ppm 1.99 - 2.05 (m, 2H, *H*-3), 2.10 - 2.16 (m, 2H, *H*-2), 3.40 (t, J = 8.45 Hz, 2H, *H*-1), 3.61 (t, J = 6.15 Hz, 2H, *H*-4).

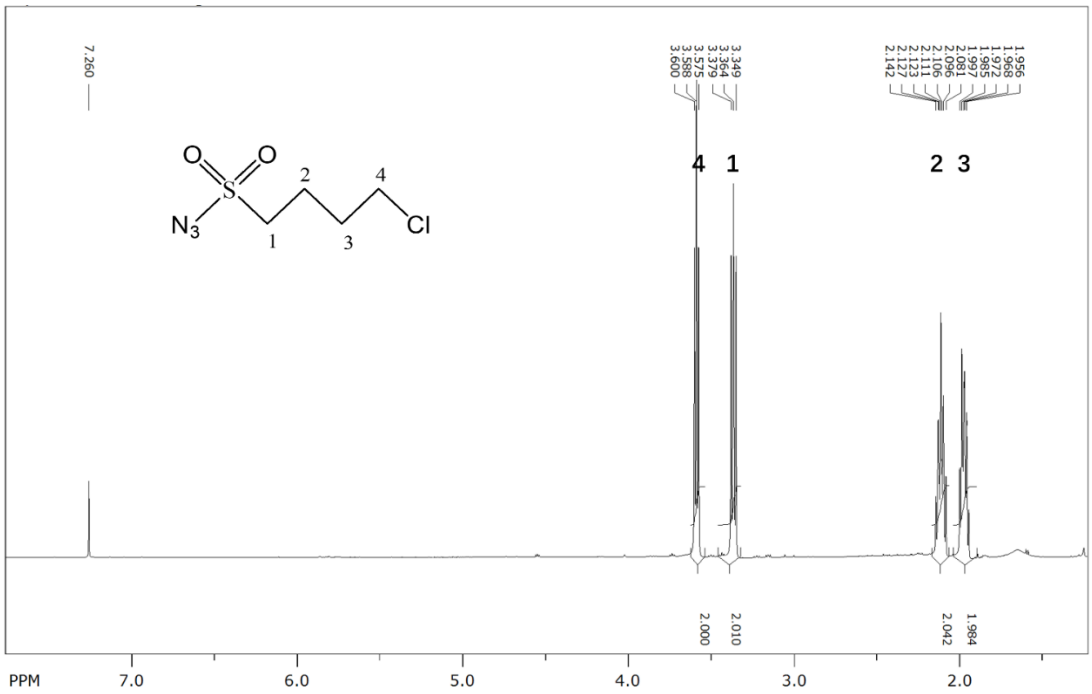


Figure B.4. ^{1}H NMR spectrum of 4-chlorobutane-1-sulfonyl azide (**3**) in $CDCl_3$.

^{13}C NMR: 125.8 MHz, $CDCl_3$: δ, ppm 20.98 (*C*-2), 30.38 (*C*-3), 43.60 (*CH₂Cl*), 55.03 (*C*-1).

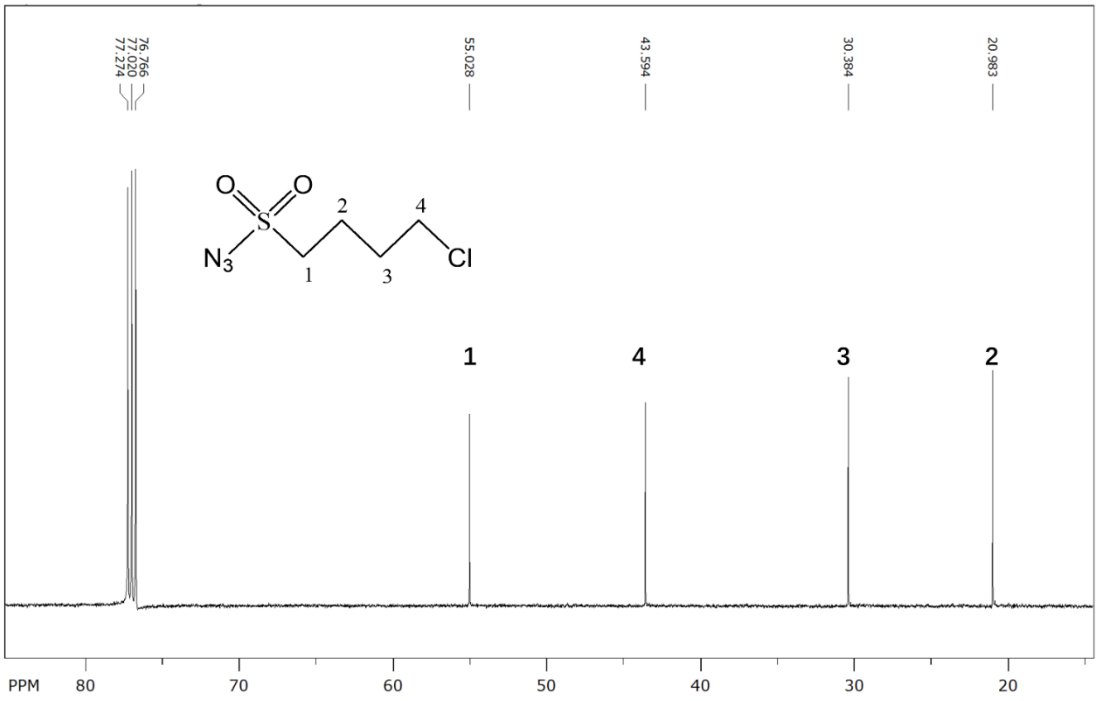


Figure B.5. ^{13}C NMR spectrum of 4-chlorobutane-1-sulfonyl azide (**3**) in $CDCl_3$.

IR (ATR), cm^{-1}: ν (CH_2) 2963 - 2877; ν (N_3) 2138; ν (RSO_2) 1360.

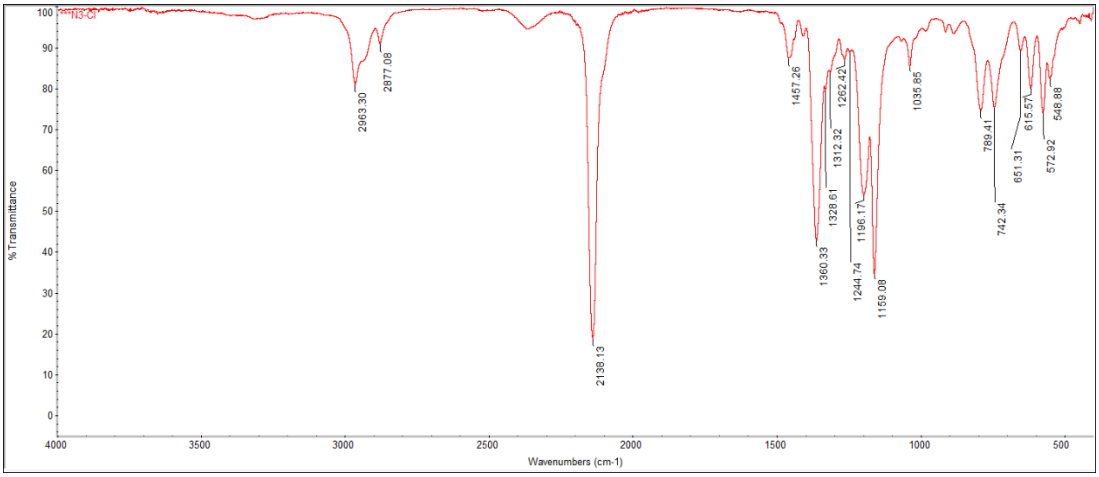


Figure B.6. IR spectrum of 4-chlorobutane-1-sulfonyl azide (**3**).

4-Iodobutane-1-sulfonyl azide (4)

Compound **3** (46.9 g, 237.5 mmol) and NaI (35.6 g, 237.9 mmol) were dissolved in acetone (100 mL), stirred and refluxed for 30 h. Reaction was monitored using ^{1}H NMR. After starting material was consumed, reaction mixture was filtered, concentrated *in vacuo* and then diluted with CH_2Cl_2 (25 mL). Solution was washed with H_2O (3 × 40 mL), dried over $MgSO_4$, filtered, evaporated *in vacuo*, yielding 58.1 g (84.3 %) of iodide **4** as a dark orange oil.

^{1}H NMR: 500 MHz, $CDCl_3$: δ, ppm 2.00 - 2.06 (m, 2H, *H*-3), 2.07 - 2.13 (m, 2H, *H*-2), 3.24 (t, J = 6.55 Hz, 2H, *H*-4), 3.37 (t, J = 7.1 Hz, 2H, *H*-1).

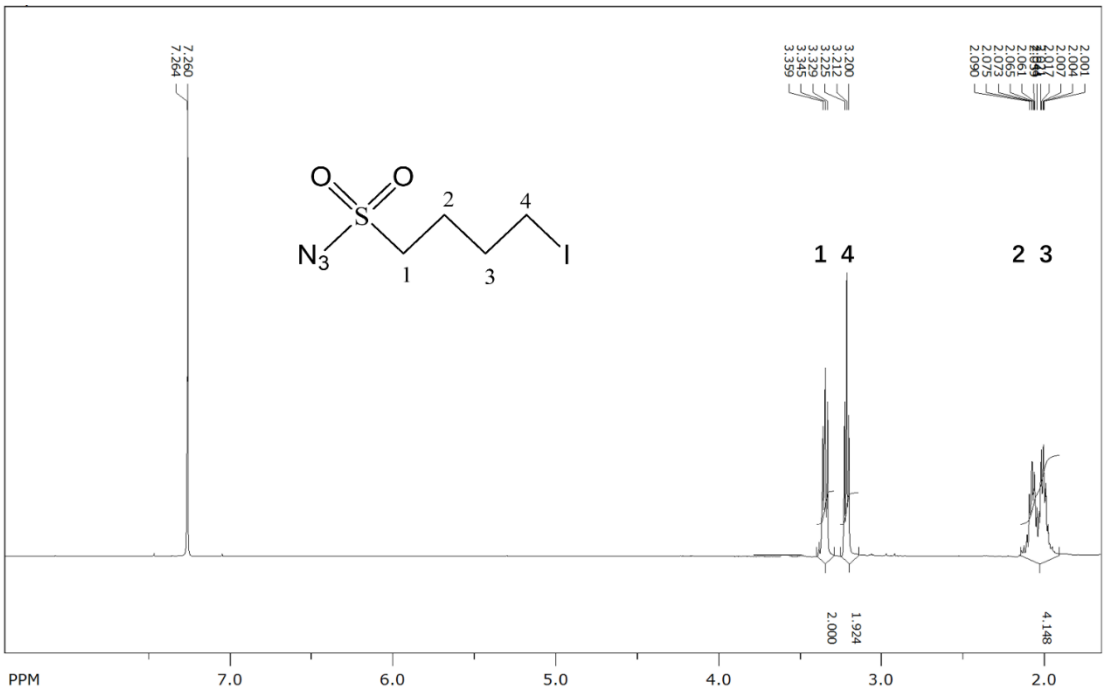


Figure B.7. ^{1}H NMR spectrum of 4-iodobutane-1-sulfonyl azide (**4**) in $CDCl_3$.

^{13}C NMR: 125.8 MHz, $CDCl_3$: δ, ppm 4.04 (*C*H_2*I*), 24.45 (*C*-2), 31.12 (*C*-3), 54.67 (*C*-1).

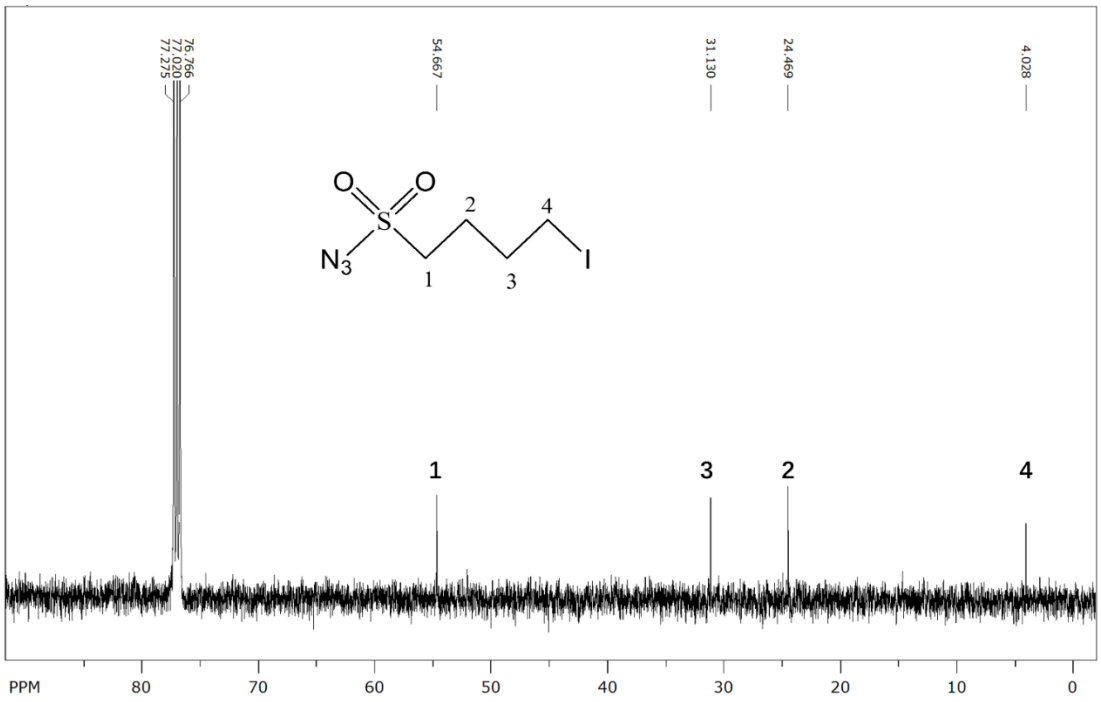


Figure B.8. ^{13}C NMR spectrum of 4-iodobutane-1-sulfonyl azide (**4**) in $CDCl_3$.

IR (ATR), cm^{-1}: ν (CH_2) 2955 - 2871; ν (RSO_2-I) 2370; ν (N_3) 2149; ν (RSO_2) 1350.

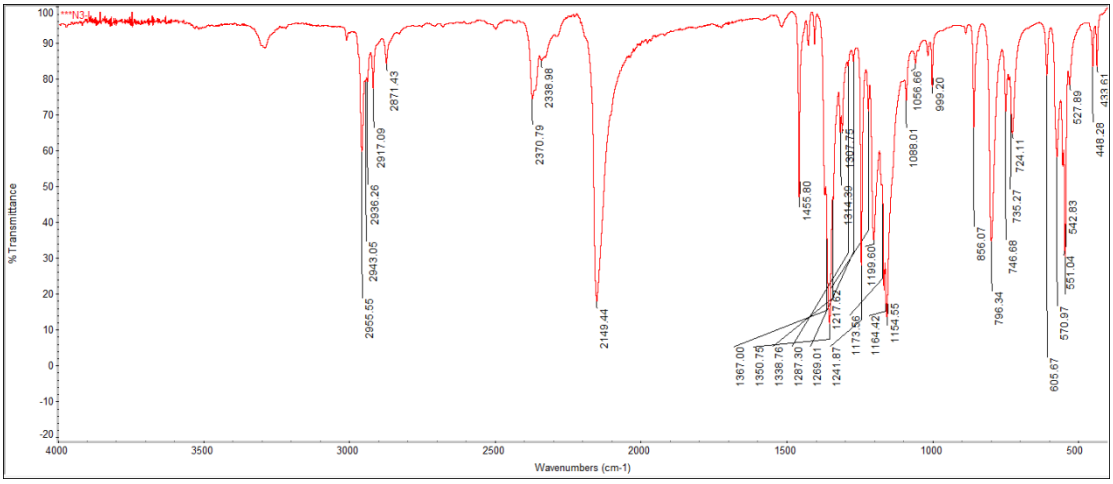


Figure B.9. IR spectrum of 4-iodobutane-1-sulfonyl azide (**4**).

4-(Azidosulfonyl)-*N,N,N*-trimethylbutan-1-aminium iodide (5)

Compound **4** (10 g, 34.5 mmol) was dissolved in diethyl ether (50 ml), and Me_3N (saturated solution in diethyl ether, 51.7 mmol, 43.3 ml) was added. Reaction mixture was stirred vigorously at room temperature for 3 days, then transferred into four 50 mL centrifuge tubes (Cellstar) and centrifuged (2900 × g, 5000 rpm, Fiberlite F15-8×50cy rotor from Thermo Fischer Scientific, 5 min) using Heraeus multifuge X1R centrifuge (Thermo Scientific). Supernatant was discarded, the pellet was washed with diethyl ether (3 × 25 mL) followed by CH_2Cl_2 (3 × 25 mL), and dried *in vacuo*, providing 8.8 g (67.3 %) of compound **5** as a white powder.

^{1}H NMR: 500 MHz, D_2O: δ, ppm 1.95 - 1.99 (m, 4H, *H*-3 + *H*-2), 3.11 (s, 9H, *H*-5,6,7), 3.35 (t, *J* = 9.15 Hz, 2H, *H*-4), 3.70 (t, *J* = 7.05 Hz, 2H, *H*-1).

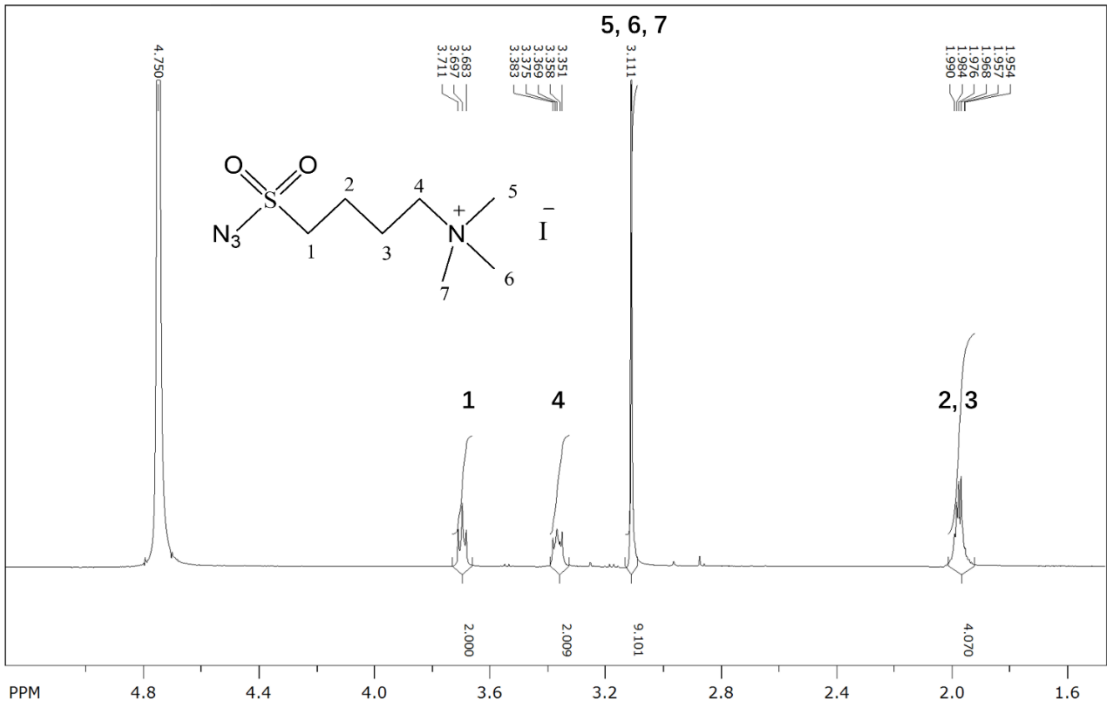


Figure B.10. ^{1}H NMR spectrum for 4-(azidosulfonyl)-*N,N,N*-trimethylbutan-1-aminium iodide (**5**) in D_2O.

^{13}C NMR: 125.8 MHz, D_2O: δ, ppm 19.82 (*C*-2), 20.81 (*C*-3), 52.89, 52.91, 52.94 (*C*-5,6,7), 54.13 (*C*-1), 65.48 (*CH₂N*).

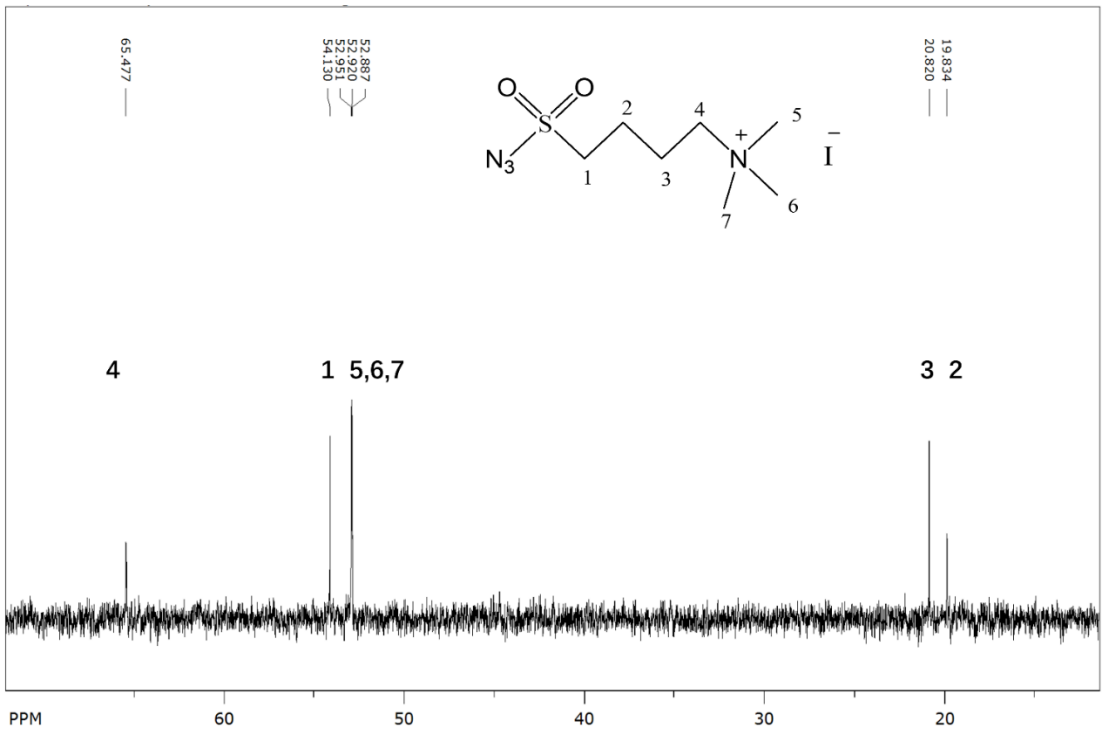


Figure B.11. ^{13}C NMR spectrum of 4-(azidosulfonyl)-*N,N,N*-trimethylbutan-1-aminium (**5**) in D_2O.

IR (ATR), cm^{-1}: ν (CH_3) 3018 - 2996; ν (CH_2) 2968 - 2885; ν (N+) 2386; ν (N_3) 2145; N (RSO_2)1341.

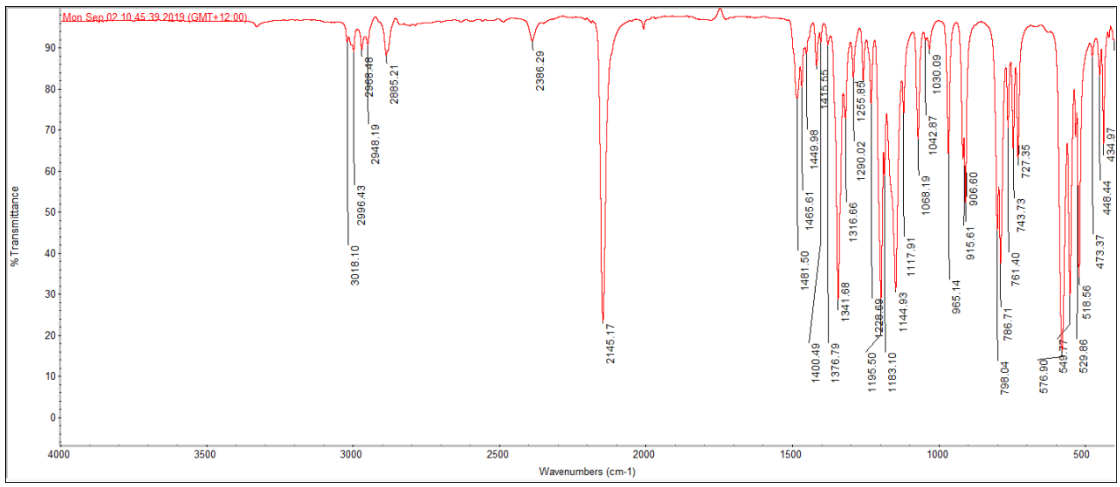


Figure B.12. IR spectrum of 4-(azidosulfonyl)-*N,N,N*-trimethylbutan-1-aminium (**5**).

ESI-MS calcd. for $C_7H_{17}N_4O_2S^+$ $[M]^+$: m/z 221.1067; found: m/z 221.1068.

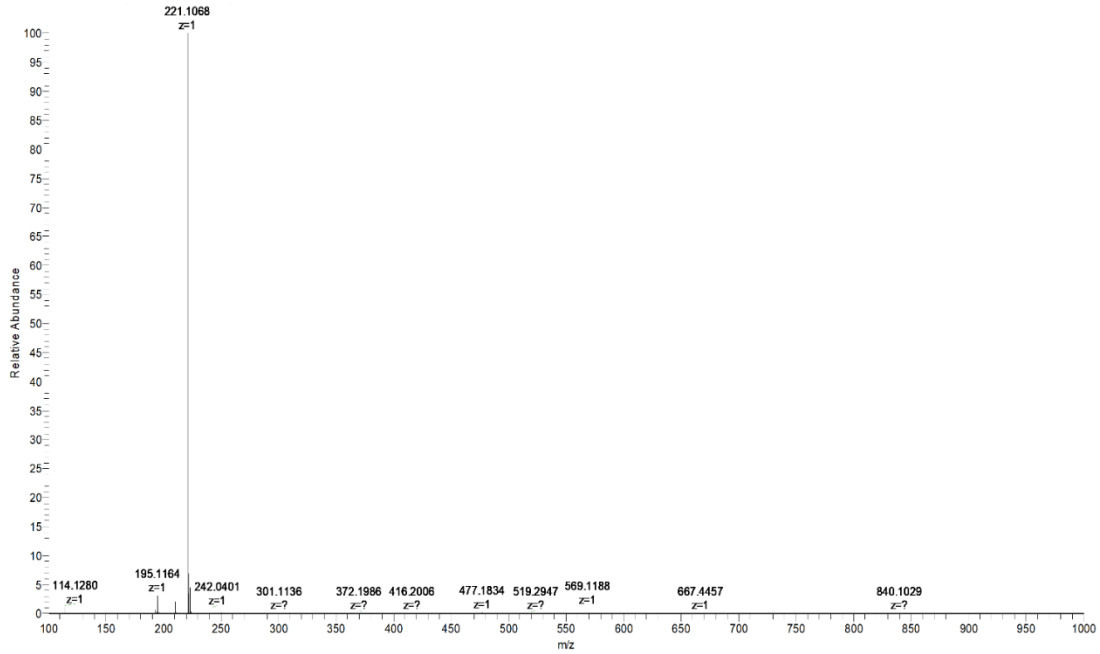
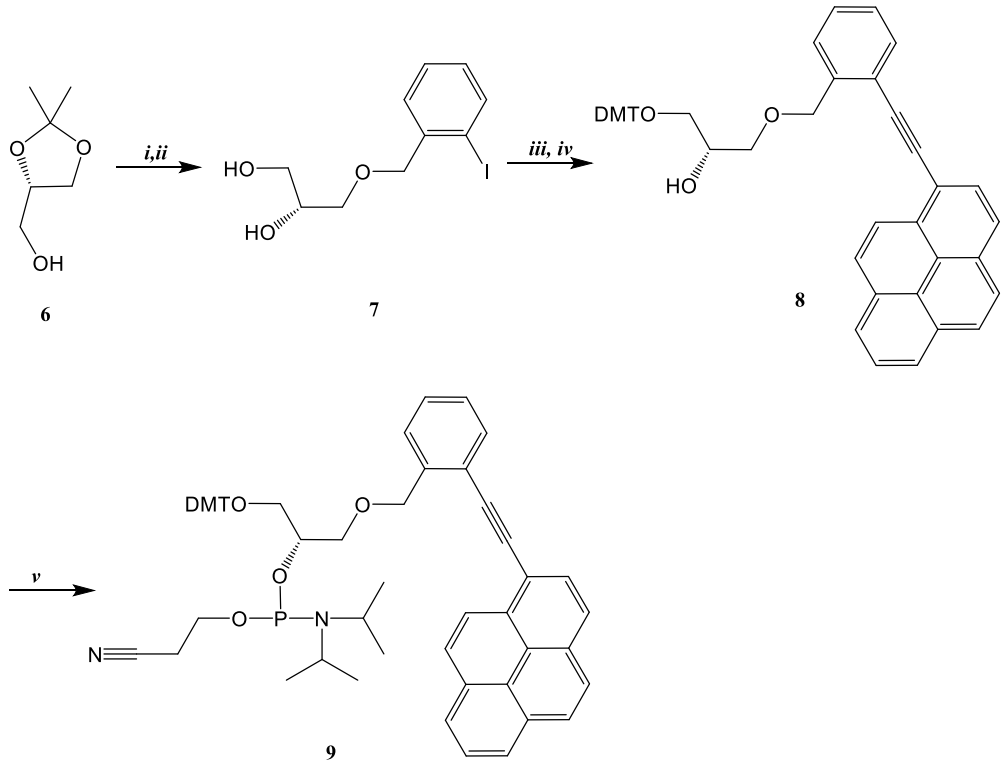


Figure B.13. ESI-MS spectrum (positive mode) of 4-(azidosulfonyl)-*N,N,N*-trimethylbutan-1-aminium (**5**) in MeCN.

2. Synthesis of *ortho*-TINA phosphoramidite



Scheme B.1. Synthesis of *ortho*-TINA phosphoramidite. Reagents and conditions: *i*) 2-iodobenzylbromide, KOH, toluene, reflux, 7 h; *ii*) 80% aq. CF_3COOH, RT, 4 h; *iii*) 1-Ethynylpyrene, Et_3N, $Pd(PPh_3)_4$, CuI, DMF, RT under argon, 26 h; *iv*) DMT-Cl, pyridine, RT, 2 h; v) $NC(CH_2)_2OP(N^iPr_2)_2$, diisopropylamonium tetrazolide, CH_2Cl_2, RT, 1.5 h.

The protocol of synthesis of *ortho*-TINA phosphoramidite was adopted from reported procedure[132] and modified as follow:

Synthesis of DMT-protected *ortho*-TINA precursor: (*S*)-1-*O*-(4,4′-dimethoxytriphenylmethyloxy)-3-*O*-[2-(pyren-1-ylethynyl)benzyloxy]propan-2-ol (8)

(*S*)-(+)2,2-Dimethyl-1,3-dioxolane-4-methanol (**1**, 500 mg, 3.8 mmol) and 2-iodobenzylbromide (1.1 g, 3.7 mmol) were refluxed under Dean–Stark conditions in toluene (250 mL) in the presence of KOH (1.94 g, 34.6 mmol) for 7 h. The reaction mixture was allowed to cool to RT and H_2O (50 mL) was added. After separation of the phases, the water layer was washed with toluene (2 × 25 mL). The combined organic layers were washed with H_2O (30 mL) and concentrated in vacuo. The residue was treated with 80% aq. CF_3CO_2H (25 mL) for 4 h at RT. The solvent was removed *in vacuo* and the residue was co-evaporated twice with toluene/EtOH (30 mL, 5:1, v/v) and then with dry pyridine (20 mL). The residue was dried under diminished pressure to afford (*R*)-3-

(2-iodobenzyloxy) propane-1,2-diol (compound **7**, 80%, 943 mg) as yellowish oil that was used in the next step without further purification.

Compound **7** (943 mg; 3.06 mmol) was dissolved in DMF (7 mL) and triethylamine (7 mL). Argon was bubbled through the mixture for at least 30 min. CuI (21.5 mg; 0.153 mmol) and $Pd(PPh_3)_4$ (176.8 mg; 0.153 mmol) were added and the mixture was bubbled with argon for another 5 min. 1-Ethynylpyrene (1.23 g; 5.45 mmol) was added and the reaction mixture was intensively bubbled with argon for 10 min, followed by stirring at RT in the dark under argon atmosphere. The reaction was finished after 26 h according to silica TLC (5% MeOH/$CHCl_3$). The reaction mixture was filtered and CH_2Cl_2 (30 mL) was added to the mixture. The mixture was washed with 0.3 M Na_2-EDTA (3 × 15 mL), H_2O (2 × 20 mL), dried over $MgSO_4$, filtered, concentrated *in vacuo* to afford a dark brown oil. This oil (995 mg, 2.5 mmol) was dissolved in anhydrous pyridine (25 mL) and 4,4'-dimethoxytrityl chloride (914 mg, 2.7 mmol) was then added under argon. After 2 h, ethanol (2 mL) followed by EtOAc (150 mL) were added and the mixture was extracted with sat. $NaHCO_3$ (3 × 40 mL). The water phase was extracted with EtOAc (50 mL). The combined organic layers were dried over Na_2SO_4, filtered, and evaporated under diminished pressure. The residue was coevaporated twice with toluene/EtOH (25 mL, 1: 1, v/v). The residue was dried, dissolved in toluene, and purified using silica gel flash chromatography with EtOAc (0– 20%, v/v) in toluene to afford compound **8** (63%, 1.1 g) as a yellow foam.

^{1}H NMR ($CDCl_3$) δ = 3.25 (m, 2H, CHOHC*H*$_2$OCH$_2$), 3.69 (s, 6H, 2 × OCH_3), 3.77 (m, 2H, CH_2ODMT), 4.00–4.16 (m, 1H, C*H*OH), 4.98 (s, 2H, CH_2Ar), 6.71–6.75 (m, 4H, DMT), 7.17–7.29 (m, 7H, Ar), 7.36–7.45 (m, 4H, Ar), 7.48–7.50 (m, 1H, Ar), 7.66–7.80 (m, 1H, Ar), 7.94–8.30 (m, 8H, Ar), 8.65 (d, *J* = 9.1 Hz, 1H, Ar).

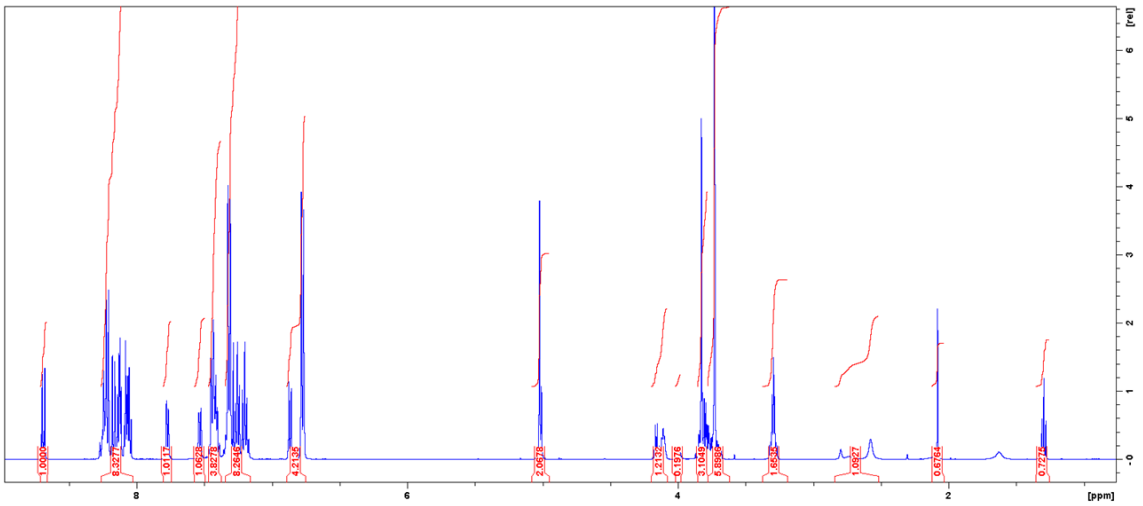


Figure B.14. ^{1}H NMR spectrum of (*S*)-1-*O*-(4,4′-dimethoxytriphenylmethyloxy)-3-*O*-[2-(pyren-1-ylethynyl)benzyloxy]propan-2-ol (**8**) in $CDCl_3$.

^{13}C NMR ($CDCl_3$) δ = 55.11 (2 × OCH_3), 64.41 (CH_2ODMT), 70.02 (CH(OH)*C*H_2OCH_2), 71.76 (OCH_2-phenyl), 72.03 (CHOH), 86.08 (CPh_3), 92.72, 93.22 (C≡C), 113.06, 117.65, 122.28, 124.33, 124.51, 124.58, 125.45, 125.63, 125.72, 126.26, 126.70, 127.24, 127.61, 127.76, 128.01, 128.11, 128.24, 128.50, 128.59, 129.65, 130.02, 131.06, 131.25, 131.37, 131.87, 132.35, 135.99, 139.71, 144.84, 158.41 (Ar).

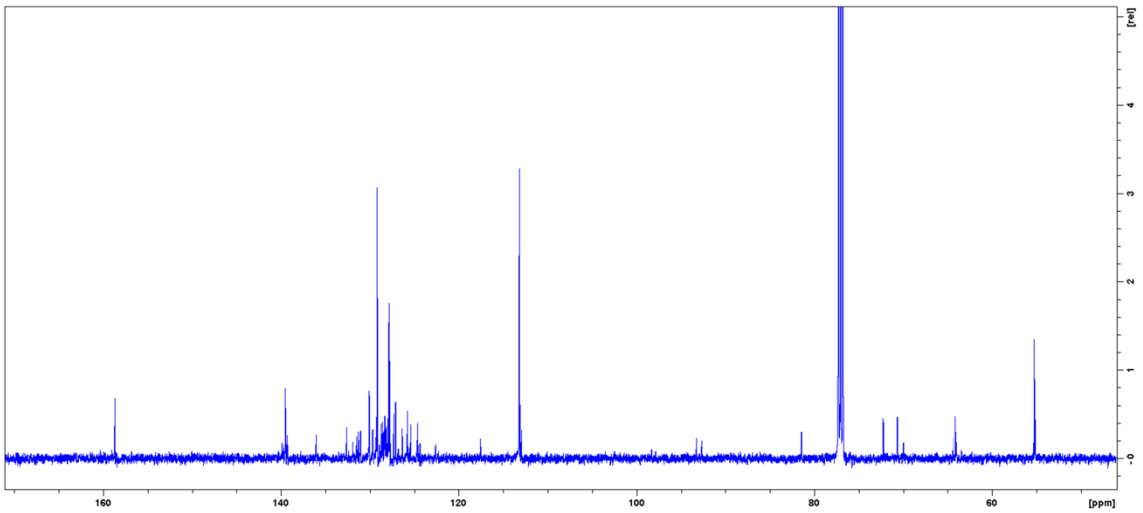


Figure B.15. ^{13}C NMR spectrum of (*S*)-1-*O*-(4,4′-dimethoxytriphenylmethyloxy)-3-*O*-[2-(pyren-1-ylethynyl)benzyloxy]propan-2-ol (**8**) in $CDCl_3$.

ESI-MS: calcd. for $C_{49}H_{40}O_5$ $[M^+]^+$: m/z 708.2876; found: m/z 708.28.

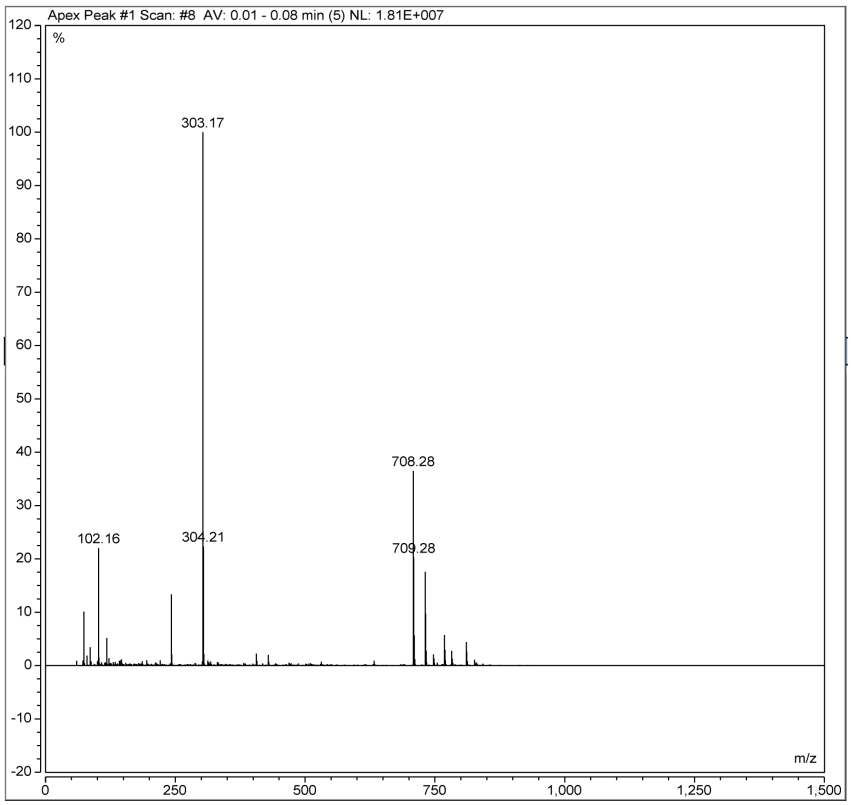


Figure B.16. ESI-MS spectrum (positive mode) of (*S*)-1-*O*-(4,4′-dimethoxytriphenylmethyloxy)-3-*O*-[2-(pyren-1-ylethynyl)benzyloxy]propan-2-ol (**8**) in MeCN.

Sythesis of *ortho*-TINA phosphoramidite:

(*S*)-1-*O*-(4,4′-dimethoxytriphenylmethyloxy)-3-*O*-[2-(pyren-1-ylethynyl)benzyloxy]propan-2-yl 2-cyanoethyl diisopropylphosphoramidite (9)

Compound **8** (500 mg; 0.7 mmol) was dissolved in dry CH_2Cl_2 (freshly distilled, 50 mL). *N,N*-Diisopropylammonium tetrazolide (181 mg; 1.06 mmol) and 2-cyanoethyl *N,N,N′,N′*-tetraisopropylphosphoane (340 µL; 1.12 mmol) were added to the mixture under an argon atmosphere. The solvent was concentrated *in vacuo* to half volume (25 mL) and stirred at RT for 1.5 h until analytical TLC showed no more starting material (ethyl acetate: toluene: triethylamine, 100: 200: 1, v/v/v). The reaction was taken up in CH_2Cl_2 (25 mL) and washed with saturated $NaHCO_3$ (3 × 50 mL), followed by brine (2 × 50 mL) and then dried over anhydrous Na_2SO_4, filtered, and evaporated under reduced pressure. The residue was dissolved in toluene and purified using silica gel flash chromatography with acetone (0 – 5%, v/v, 1% Et_3N in acetone) in toluene to afford compound **4** in 65% yield (419 mg) as a yellow foam.

^{31}P NMR ($CDCl_3$) δ ppm = 149.44, 149.23.

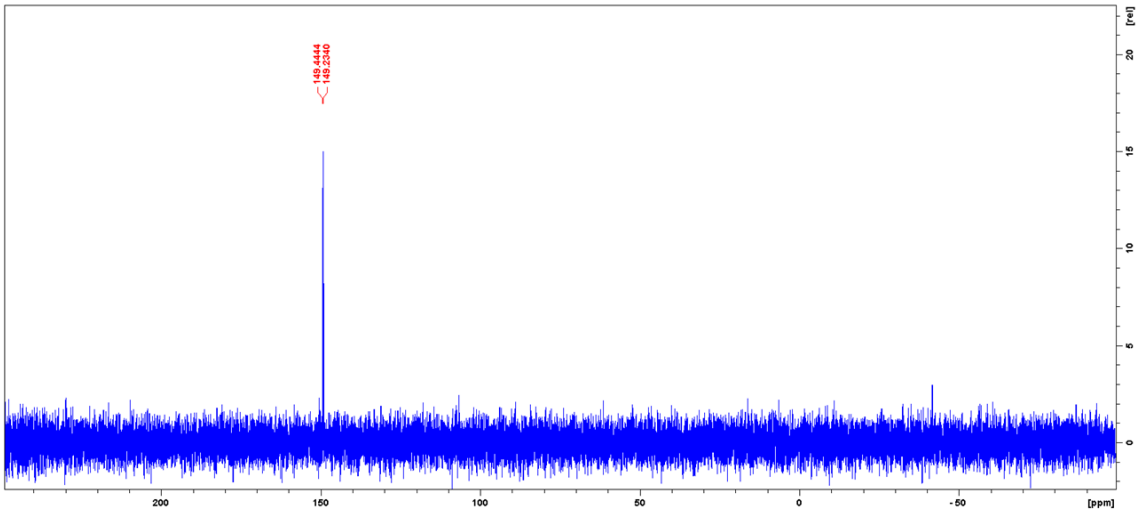


Figure B.17. ^{31}P NMR spectrum of (*S*)-1-*O*-(4,4′-dimethoxytriphenylmethyloxy)-3-*O*-[2-(pyren-1-ylethynyl)benzyloxy]propan-2-yl 2-cyanoethyl diisopropylphosphoramidite (**9**) in $CDCl_3$.

ESI-MS: calcd for $C_{58}H_{57}N_2O_6P$ $[M+]^+$: m/z 909.3954; found: m/z 909.4015.

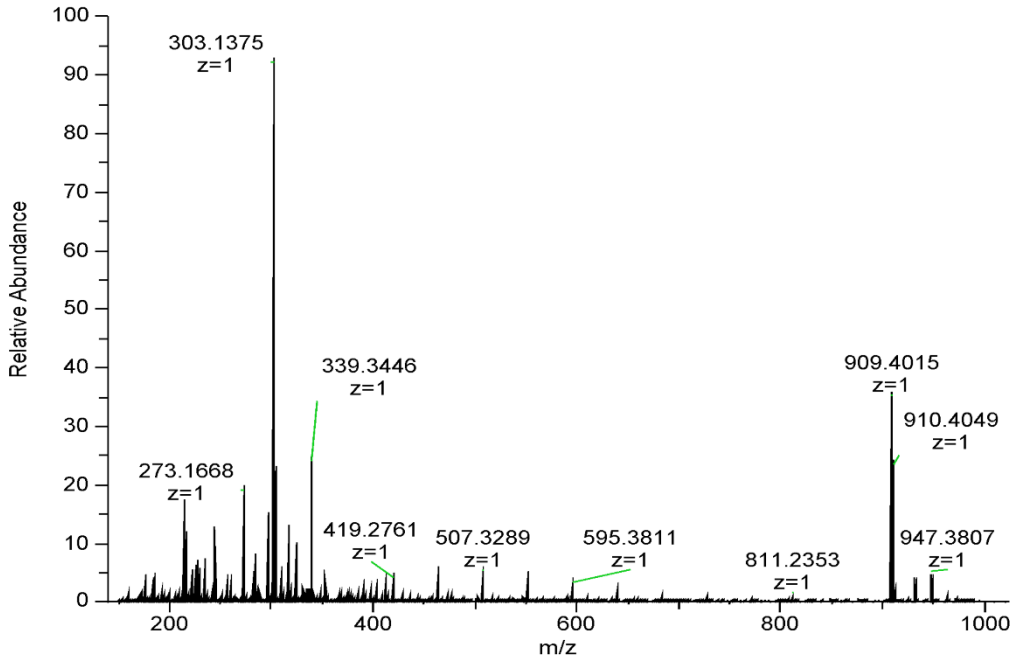


Figure B.18. ESI-MS spectrum (positive mode) of (*S*)-1-*O*-(4,4′-dimethoxytriphenylmethyloxy)-3-*O*-[2-(pyren-1-ylethynyl)benzyloxy]propan-2-yl 2-cyanoethyl diisopropylphosphoramidite (**9**) in MeCN.

3. ON synthesis and purification

Unmodified ONs were purchased from Integrated DNA Technologies (IDT, Singapore). *ortho*-TINA- and N+ modified ONs were synthesised using a Mermaid-4 automated DNA synthesiser (BioAutomation Corp.) using 5-ethylthio-1*H*-tetrazole (ETT) as an activator. Oxidation and deprotection times were 60 s (repeat 3 times) and coupling time 90 s for 5 µmol synthesis scale. An increased deprotection time (100 s) and coupling time (2 min) for 0.075 M solutions of the *ortho*-TINA phosphoramidites in 1:1 mixture of dry MeCN/CH_2Cl_2 were applied.[250] For introduction of N+ moiety, the automatic synthesis was paused after coupling step, the column was taken out from DNA synthesiser and connected to a micro tube pump (Fig. B.19), 4-(azidosulfonyl)-*N,N,N*-trimethylbutan-1-aminium iodide (**5**) in DMF (saturated and degassed, 2 mL) was pumped through the column for 30 min at 37 °C. The column was placed back in the DNA synthesiser to continue the synthesis. The resulting ON was cleaved from the solid support, and deprotected with concentrated aqueous ammonia (~ 28%) at RT for 2 h, followed by 55 °C for 12 h.

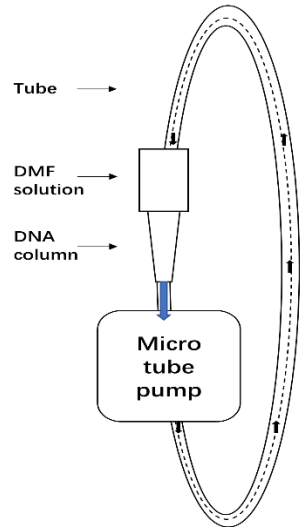


Figure B.19. Micro tube pump for pumping solution through the column containing DNA-CPG for introduction of N+ moiety into DNA.

Purification of ONs was accomplished by HPLC using IE-column (TSKgel Super Q-5PW). Buffer A (20 mM Tris-HCl, 1 mM EDTA, pH 9.0), buffer B (20 mM Tris-HCl, 1 mM EDTA, 1 M NaCl, pH 9.0). Gradients: 3.7 min 100% A, convex curve gradient to 30% B in 11.1 min, linear gradient to 50% B in 18.5 min, concave gradient to 100% B in 7.4 min, keep 100% B for 7.4 min and then 100% A in 7.3 min. Collected individual UV-absorbing fractions (λ = 260 nm) were desalted using NAP-25 column.

Table B.1. Abbreviation of the ONs, their retention time on ion-exchange column and their composition.

	Retention time (min)	**Calculated MW**	**Observed MW** [a]
ON1[b]	28.51	4890.8681	-
ON2	28.23	4859.8736	-
2N+ON1	23.01	5245.0790	5245.7203
2N+ON2	22.82	5214.0845	5214.4638
3N+ON1	20.31	5422.1845	5422.3754
3N+ON2	19.33	5391.1899	5390.4537
ON8	40.02	6736.2000	-
4N+ON8	30.79	7440.6190	7440.6447
3N+ON8	33.90	7267.5163	7267.6531
ON15	39.57	6807.0000	-
ON16	39.84	6701.2000	-
4N+ON15	26.31	7511.4190	7510.6088
4N+ON16	27.05	7405.6190	7405.6439
3N+ON15	28.43	7338.3163	7337.4367

[a] Based on ESI-MS in the negative mode; [b] Obtained from Integrated DNA Technologies; ESI-MS spectra are provided in Figures B.20 – B.22.

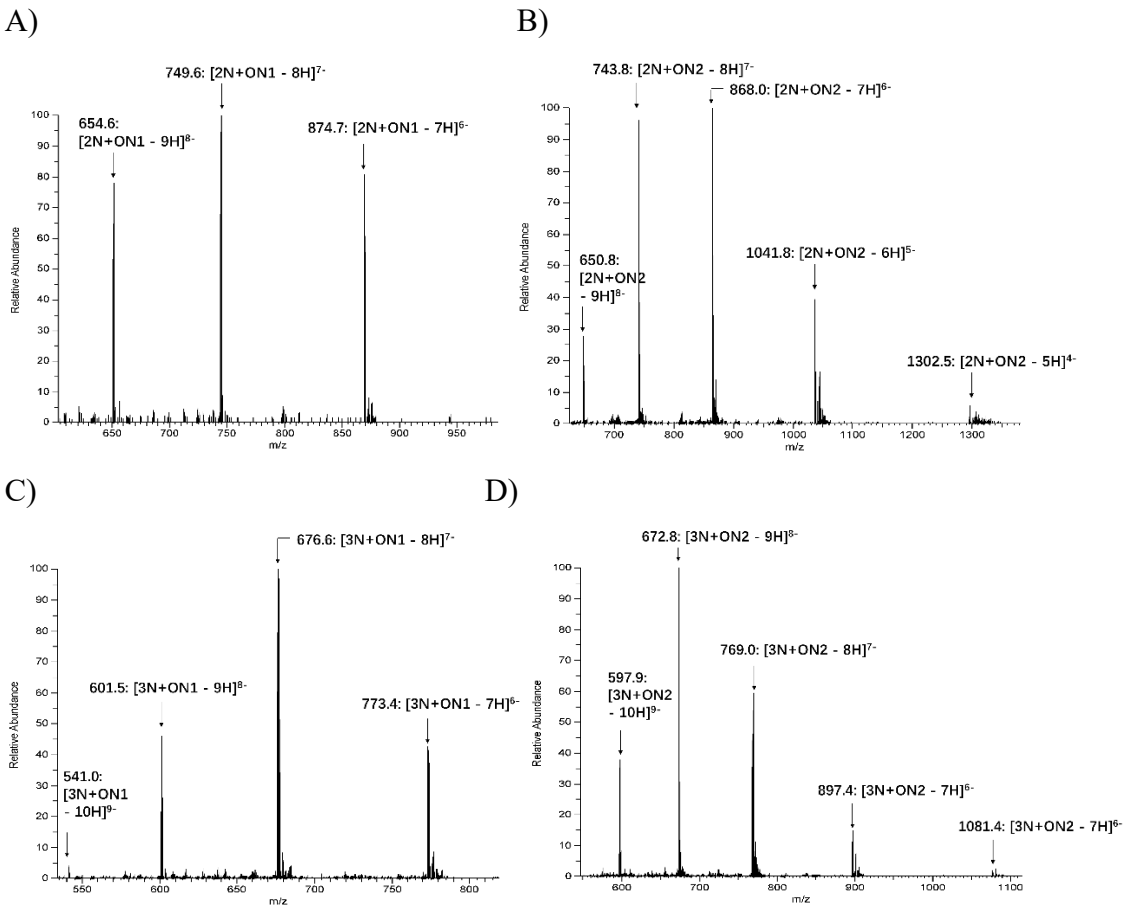


Figure B.20. ESI-MS results for N+ modified ONs in Set 1. Calculated and found molecular weights of each modified ONs are listed in Table B.1.

A) B)

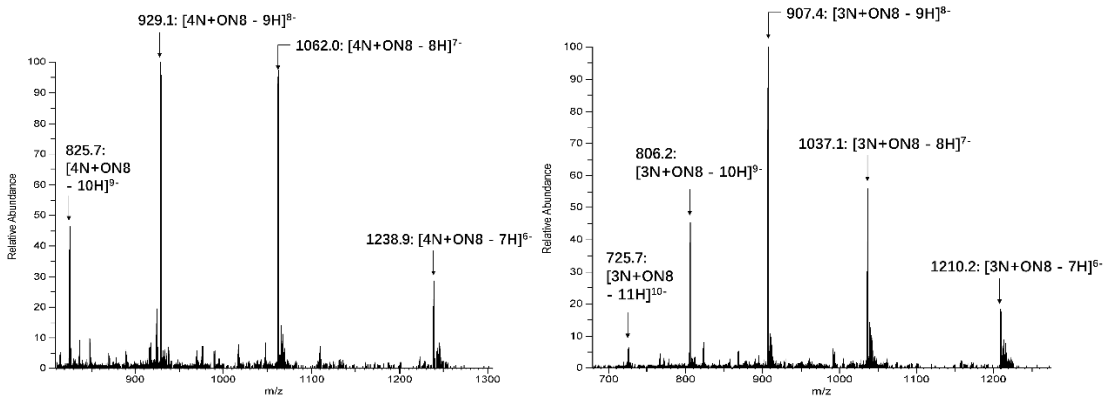
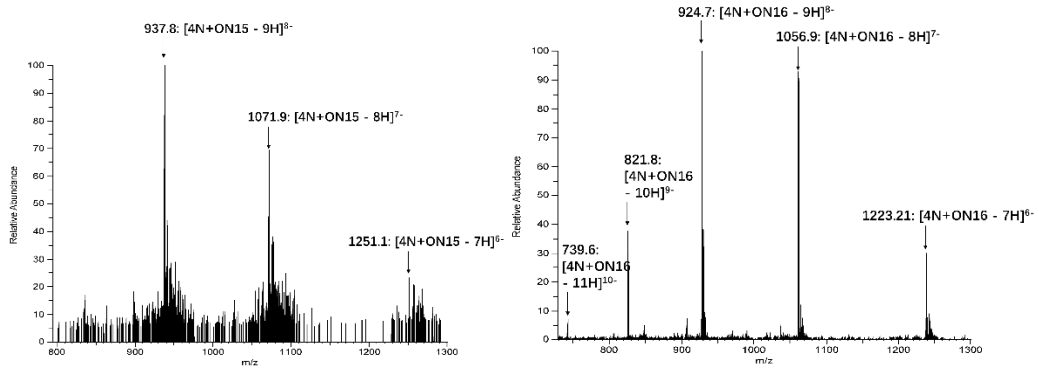


Figure B.21. ESI-MS results for N+ modified *ortho*-TINA sequences in Set 1. Calculated and found molecular weights of each modified ONs are listed in Table B.1.

A) B)



C)

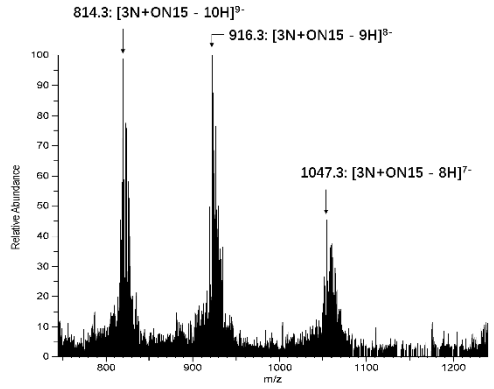


Figure B.22. ESI-MS results for N+ modified *ortho*-TINA sequences in Set 2. Calculated and found molecular weights of each modified ONs are listed in Table B.1.

Appendix C for Chapter 4

1. Synthesis and purification of chemically modified ONs

4-(Azidosulfonyl)-*N,N,N*-trimethylbutan-1-aminium iodide was synthesised as described in Chapter 2. Tosyl azide (*p*-toluenesulfonyl azide, TsN_3) were synthesised as previously reported. [181]

Unmodified ONs were purchased from Integrated DNA Technologies (IDT, Singapore). N+ and Ts- modified ONs were synthesised and purified following the previously described protocol.

The composition of each fraction was confirmed by ESI-MS (Table C. 1 and Figures C.2 – C.5). The purity of chemically modified ONs was confirmed using 20% denaturing gel (Figure C.1).

Table 1. Abbreviation of the ONs synthesised, their sequence, retention time on ion-exchange column,[a] composition, and isolated yield.

	Retention time (min)	Calculated MW	Observed MW [b]	Isolated yield, %
ON1	31.53	4121.7	-	
5′-N+ON2	27.76	4296.7958	4297.7588 [d]	9 [e]
m-N+ON3	27.43	4296.7958	4297.7455 [d]	11 [e]
3′-N+ON4	27.74	4296.7958	4297.7466 [d]	20 [f]
2N+ON5	23.42	4473.9012	4473.8248 [d]	8 [e]
3N+ON6	20.77	4649.0067	4650.1412 [d]	10 [e]
4N+ON7	17.09	4826.0965	4826.0516 [d]	23 [f]
4N+{FAM}	- [g]	5396.4012	5396.1460 [h]	20
5′-Ts-ON8	32.80	4272.7151	4274.6664 [d]	4 [e]
m-Ts-ON9	32.75	4272.7151	4273.4728 [d]	5 [e]
3′-Ts-ON10	32.81	4272.7151	4273.4694 [h]	8 [e]
2Ts-ON11	35.67	4425.7398	4427.7401[h]	6 [e]
3Ts-ON12	36.76	4578.7646	4578.6076 [i]	30 [f]
4Ts-ON13	40.07	4731.7893	4731.7620 [j]	39 [f]
4Ts-{FAM}	- [g]	5301.2398	5302.8380 [h]	26
m-N+ON14	27.50	4296.7958	4296.7520 [d]	19 [f]
m-N+ON15	27.43	4296.7958	4296.7300 [d]	16 [f]
m-N+ON16	27.45	4296.7958	4296.7440 [d]	17 [f]
m-N+ON17	27.62	4296.7958	4296.7350 [d]	21 [f]
3N+ON18	20.44	4649.0067	4649.937 [d]	17 [f]

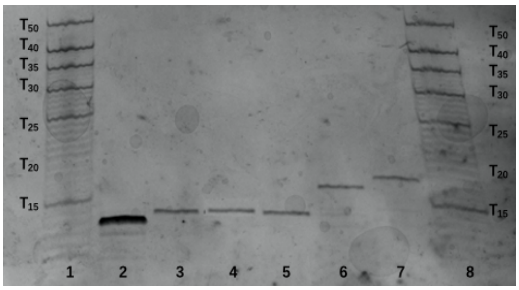
[a] IE-HPLC was performed on IE-column (TSKgel Super Q-5PW) using a gradient of NaCl concentration (0 – 0.5 M) in 20 mM Tris-HCl, 1 mM EDTA, pH 9.0 over 30 min. [b] Based on ESI-MS in the negative mode; [c] Obtained from Integrated DNA Technologies; [d] Calculated for $[M -6H]^{6-}$; [h] for $[M -7H]^{5-}$; [i] for $[M +K^{+} -9H]^{8-}$; [j] for $[M -4H]^{4-}$. ESI-MS spectra are provided in Figures C.2 – C.5.

[e] Synthesised in 1 µmol scale using previously reported procedure with transferring solid support from a column into a vial for reactin with sulfonyl azide for 30 min at room temperature. Afterwards, the solid support was transferred back to the column to continue DNA synthesis[181]. A part of solid support was lost during the transfer and wash, especially for multiple modifications, which is the major reason for the low yield.

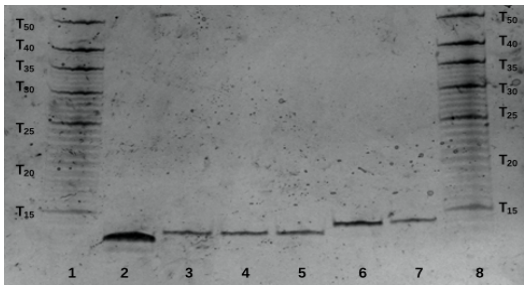
[f] Synthesised in 1 µmol scale following a modified procedure using a micro tube pump for delivery of a solution with sulfonyl azide into the column with CPG-support at 37 °C.

[g] Synthesised in 3 – 4 µmol scale, purified by 20% denaturing PAGE (7M urea), followed by extraction from the gel and desalting.

A)



B)



C)

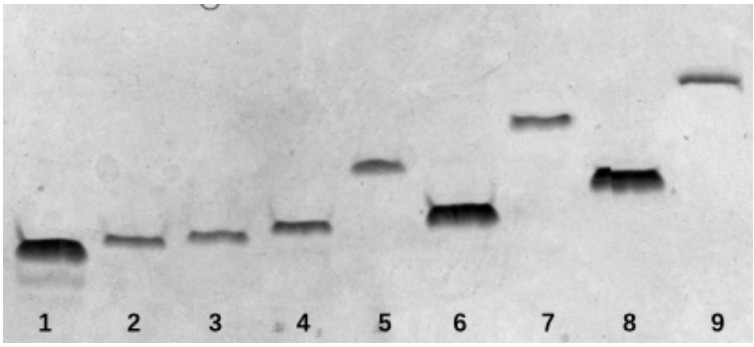


Figure C.1. 20% Denaturing gel (7M urea) of chemically modified ONs. A) Lanes are 1: ladder; 2: ON1; 3: **5′-N+ON2**; 4: **m-N+ON3**; 5: **3′-N+ON4**; 6: **2N+ON5**; 7: **3N+ON6**; 8: ladder. B) Lanes are 1: ladder; 2: **ON1**; 3: **5′-Ts-ON8**; 4: **m-Ts-ON9**; 5: **3′-Ts-ON10**; 6: **2Ts-ON11**; 7: **3Ts-ON12**; 8: ladder. C) Lanes are 1: **ON1**; 2: **5′-Ts-ON8**; 3: **5′-N+ON2**; 4: **2Ts-ON11**; 5: **2N+ON5**; 6: **3Ts-ON12**; 7: **3N+ON6**; 8: **4Ts-ON13**; 9: **4N+ON7**. Ladder is a mixture of polythimidylates as indicated on the gel picture.

2. ESI-MS analysis of modified ONs

ESI-MS spectra were recorded using Thermo Scientific Q-Exactive Focus Hybrid Quadrupole-Orbitrap Mass Spectrometer. Modified ONs were prepared into 4 μM strand concentration with water: methanol 4: 1 (HPLC grade from Fisher Scientific). Samples (5 μL) were injected via Dionex Ultimate 3000 HPLC system running at 0.1 mL/min CH_3OH.

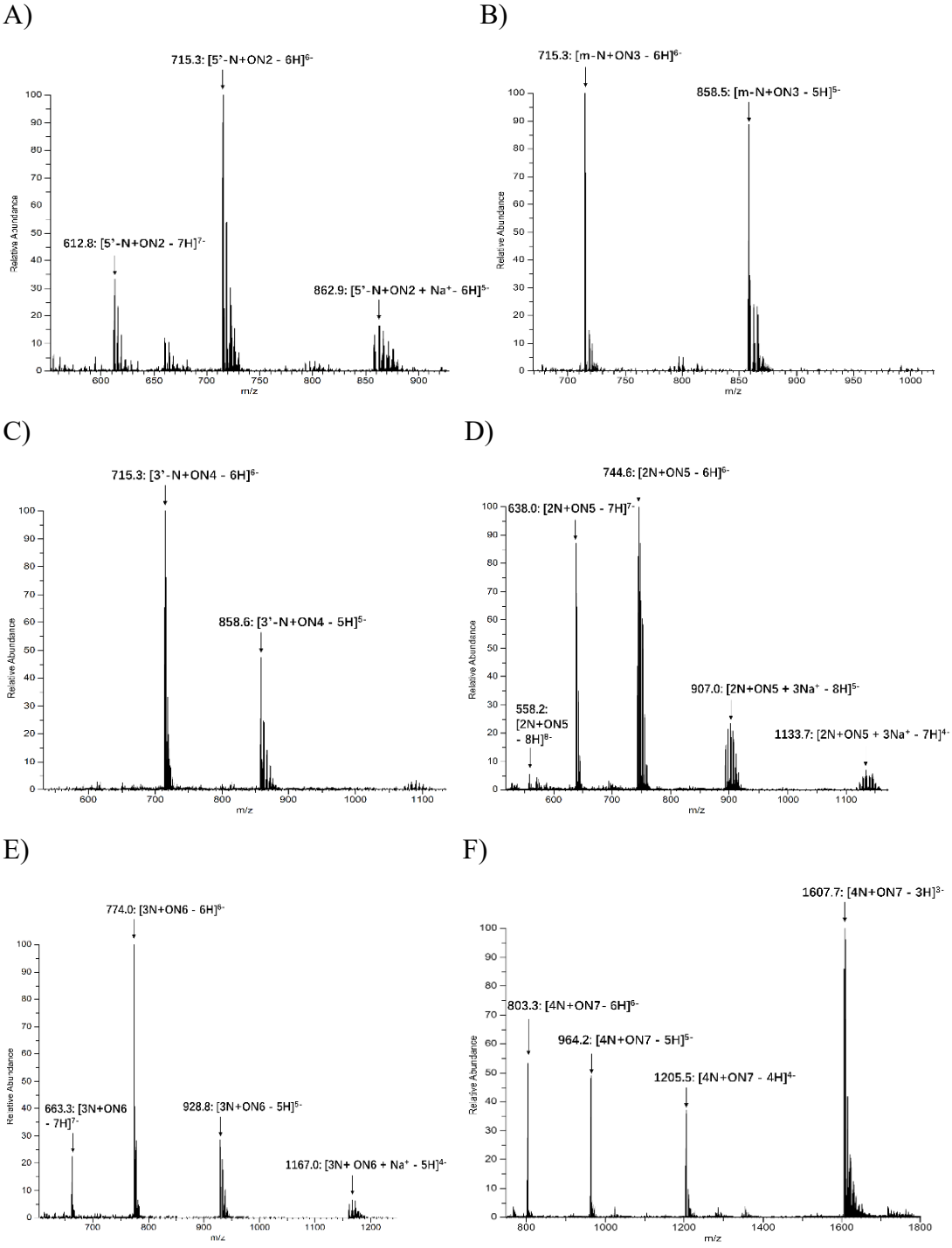


Figure C.2. ESI-MS results for **ON2** – **ON7**. Calculated and found molecular weights of each modified ONs are listed in Table C.1.

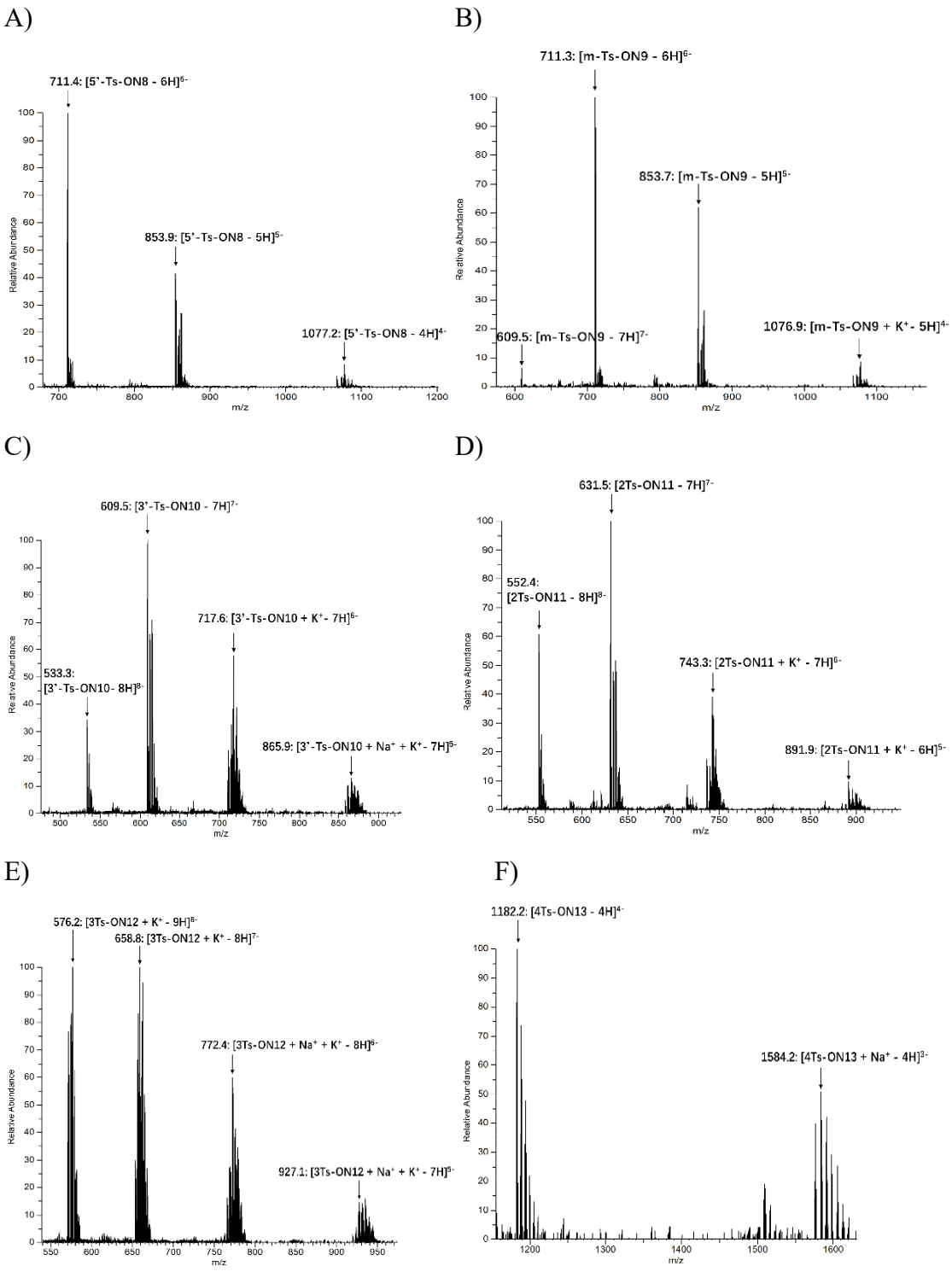


Figure C.3. ESI-MS results for **ON8** – **ON13**. Calculated and found molecular weights of each modified ONs are listed in Table C.1.

A) B)

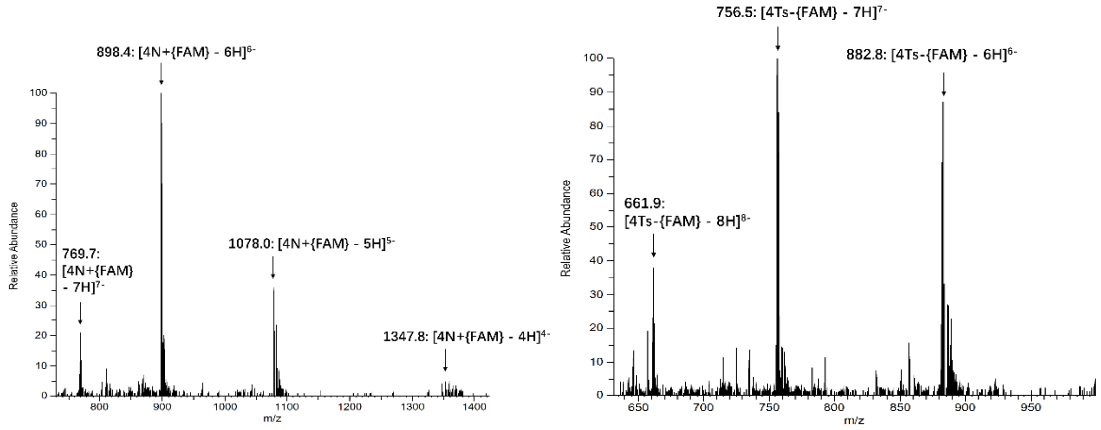


Figure C.4. ESI-MS results for fluorescently labelled ONs. Calculated and found molecular weights of 4N+{FAM} and 4Ts-{FAM} ONs are listed in Table C.1.

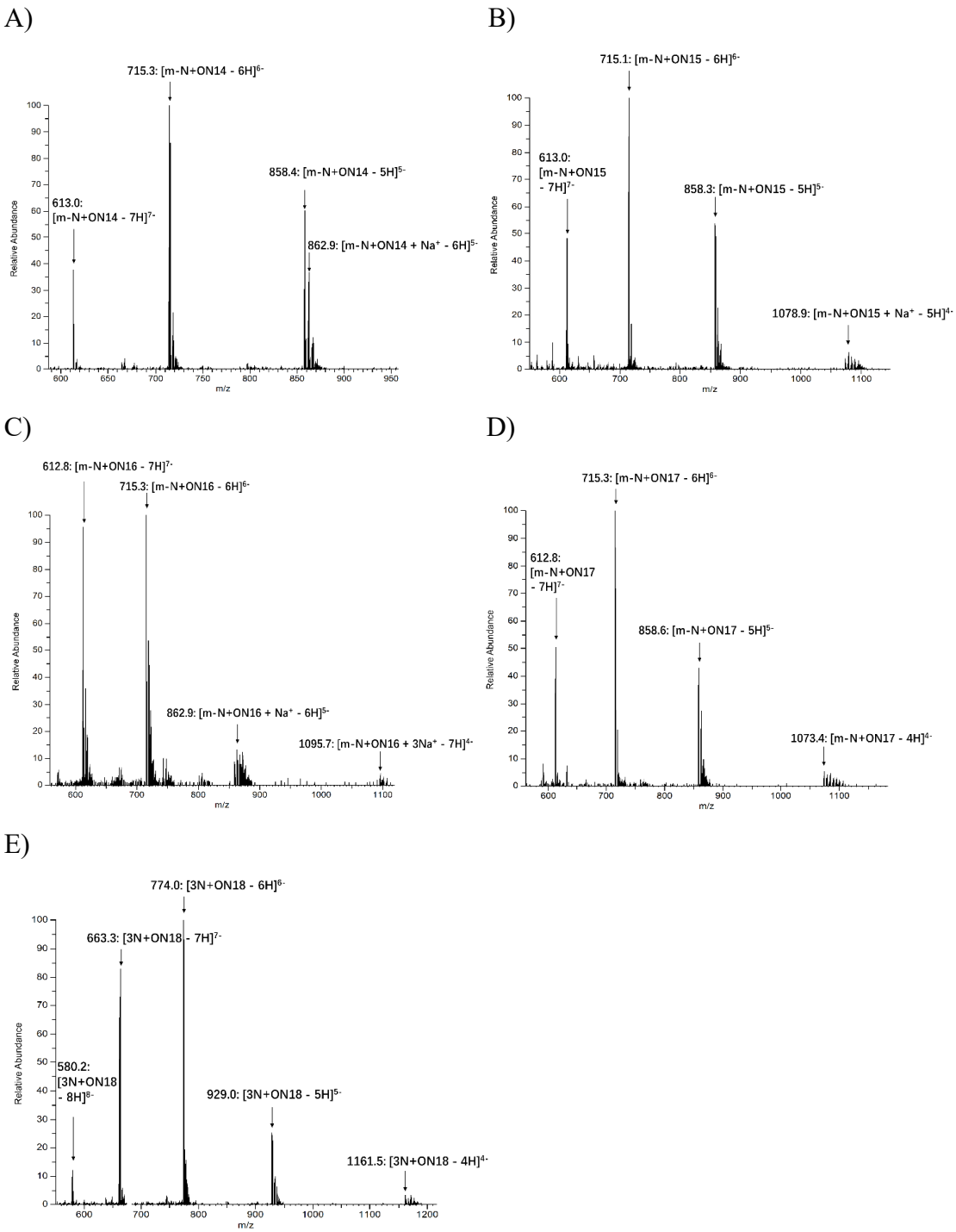


Figure C.5. ESI-MS results for **ON14** – **ON18**. Calculated and found molecular weights of each modified ONs are listed in Table C.1.

3. Determination of melting temperature of duplexes and triplexes using UV-Vis spectrometry

Parallel and antiparallel duplexes were formed by mixing two strands (each at a concentration of 1.0 μM) in the corresponding buffer solution as listed in Table 1. Each solution was heated to 80 °C for 5 min and cooled to RT. Triplexes were formed by first mixing the two strands of the Watson/Crick duplex, each at a concentration of 1.0 μM in the corresponding buffer. The solution was heated to 80 °C for 5 min and cooled to RT, and the third strand (TFO) was added and then kept at 15 °C for at least 30 min.

Melting temperature measurements were performed on a Cary-100Bio UV-Vis spectrometer using quartz cuvettes with 10 mm pathlength and a 2 × 6 multicell block with a Peltier temperature controller. The melting temperature (T_m, °C) for antiparallel duplexes was determined as the maximum of the first derivative plots of the melting curves obtained by measuring absorbance at 260 nm against increasing temperature (0.5 °C/min).

For parallel triplex at pH 5.0 and pH 6.0, a melting curve with two transition states (Fig. C.6C, blue line) was obtained due to the triplex and duplex melting curves were overlaid. A melting triplex profile was obtained by subtracting the melting curve of duplex **D1** (Fig. C.6C, black line) from the triplex melting curve. T_m value for a parallel triplex was determined as a cross point of the obtained triplex melting curve (Fig. C.6C, red line) with its median of the upper and lower baselines (Fig. C.6C). All melting temperatures are within the uncertainty of ± 0.5 °C as determined by repetitive experiments.

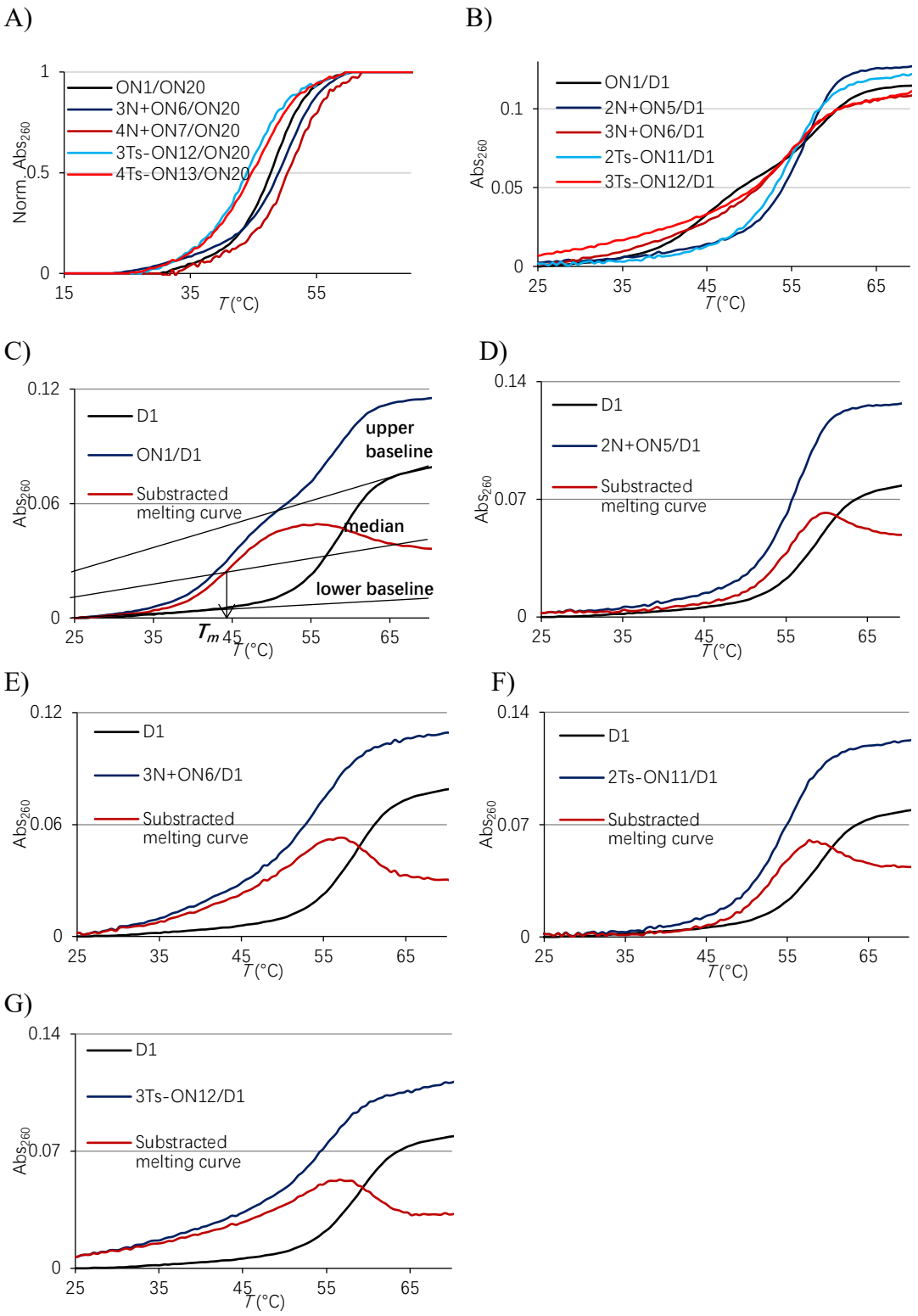


Figure C.6. Representative UV-vis melting curves recorded at 260 nm against increasing temperature for A) antiparallel DNA duplexes at pH 7.0; and B) parallel triplexes at pH 5.0; C) A triplex melting profile (red solid line) obtained after subtraction of a duplex melting curve (**D1**, black line) from the triplex melting curve (**ON1/D1**, blue line). D) – F) Subtracted menting curve for parallel triplexes showed in B). Triplex formation of **3Ts-ON12/D1** in pH 5.0 was confirmed by SE-HPLC (Fig. C.15)

3.1. Determination of melting temperature using CD measurements

To confirm the T_m values of antiparallel duplexes formed by ONs/RNA obtained from UV-Vis denaturation experiment, CD denaturation experiments were conducted using the same buffer and DNA concentrations as for UV-Vis measurements. CD denaturation and renaturation experiments were performed by recording spectra every 2.5 °C with equilibration for 2.5 min at each temperature from 10 to 75 °C (Fig. C.7A). Signal change at 265 nm was extracted and plotted against temperature to give T_m value (Fig. C.7B). The comparison of T_m values obtained from UV-Vis and CD melting experiments is shown in Table C.2.

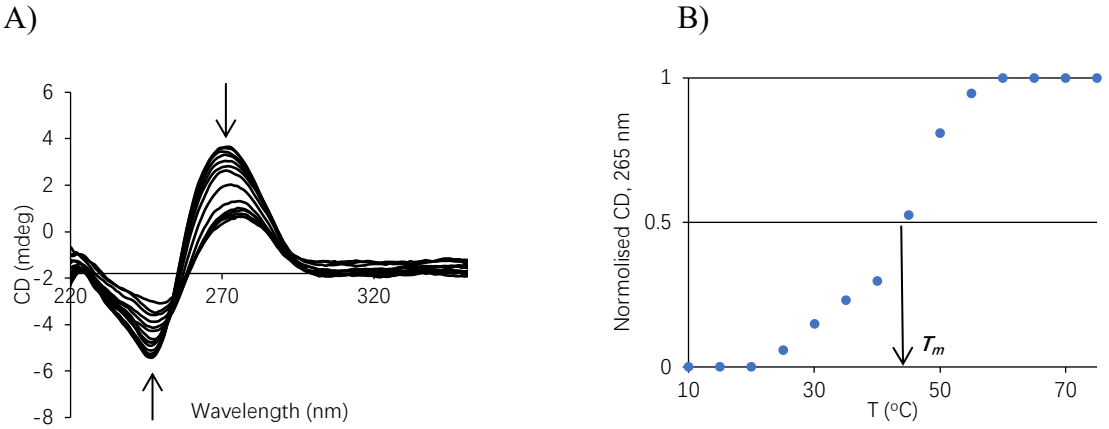


Figure C.7. A) CD melting profile of DNA/RNA duplex **ON1**/**ON16** from 10 – 75 °C; The arrow indicates direction of changes in the peak increasing from low to high temperatures. B) Plotted melting curve of extracted data at 265 nm against temperature.

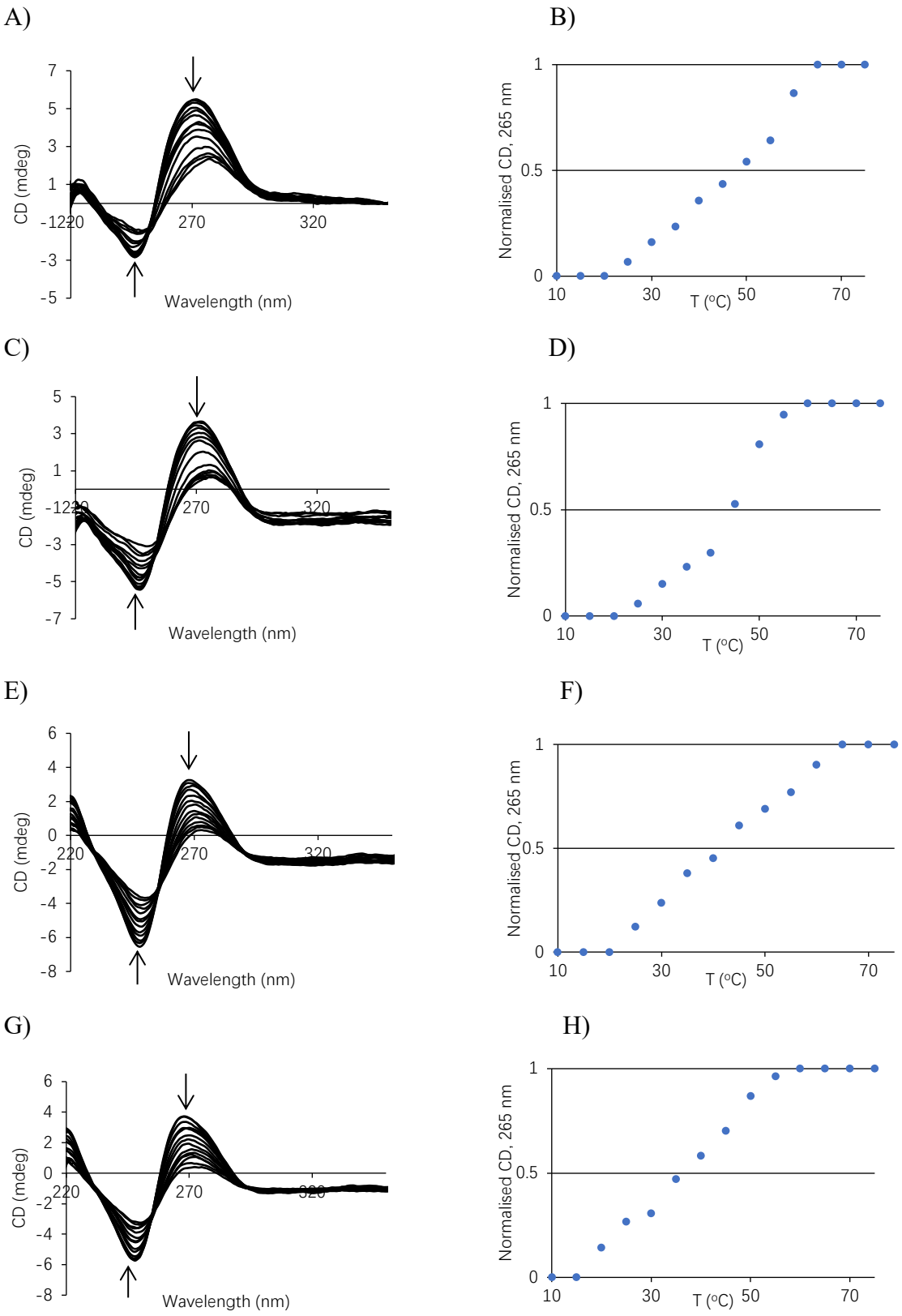


Figure C.8. Left: CD melting profiles of DNA/RNA duplexes, A) **m-N+ON3/ON19**; C) **3N+ON6/ON19**; E) **m-Ts-ON9/ON19**; G) **3Ts-ON12/ON19** from 10 – 75 °C. The arrow indicates direction of changes in the peak increasing from low to high temperatures. Right: Plotted melting curve of extracted data at 265 nm against temperature, B) **m-N+ON3/ON19**; D) **3N+ON6/ON19**; F) **m-Ts-ON9/ON19**; H) **3Ts-ON12/ON19**.

Table C.2. Comparison of T_m (°C) values for ON/RNA duplexes obtained by UV-Vis and CD denaturation experiments.

	ON19: 3'-GCGGAAAGAAAAAA	
	UV-Vis melting	CD-melting
ON1: 5'-CCCCTTTCTTTTTT	46	47
m-N+ON3: 5'-CCCCTTT$_{N+}$CTTTTTT	47	48
3N+ON6: 5'-C $_{N+}$CCCTTT $_{N+}$CTTTTT $_{N+}$T	41	41
m-Ts-ON9: 5'-CCCCTTT$_{Ts}$CTTTTTT	44	44
3Ts-ON12: 5'-C$_{Ts}$CCCTTT$_{Ts}$CTTTTT$_{Ts}$T	38	37

The determination of thermodynamic parameters for all complexes studied in Table 4.3 in the main text of Chapter 4 was following the calculations described in Chapter 2. Fig. C.9A was converted into Fig. C.9B. The van't Hoff relation ($\ln(K_a)$ *versus* 1/T) should give a straight line, with a slope of $-\Delta H^0/R$ and Y-axis intercept of $\Delta S^0/R$.

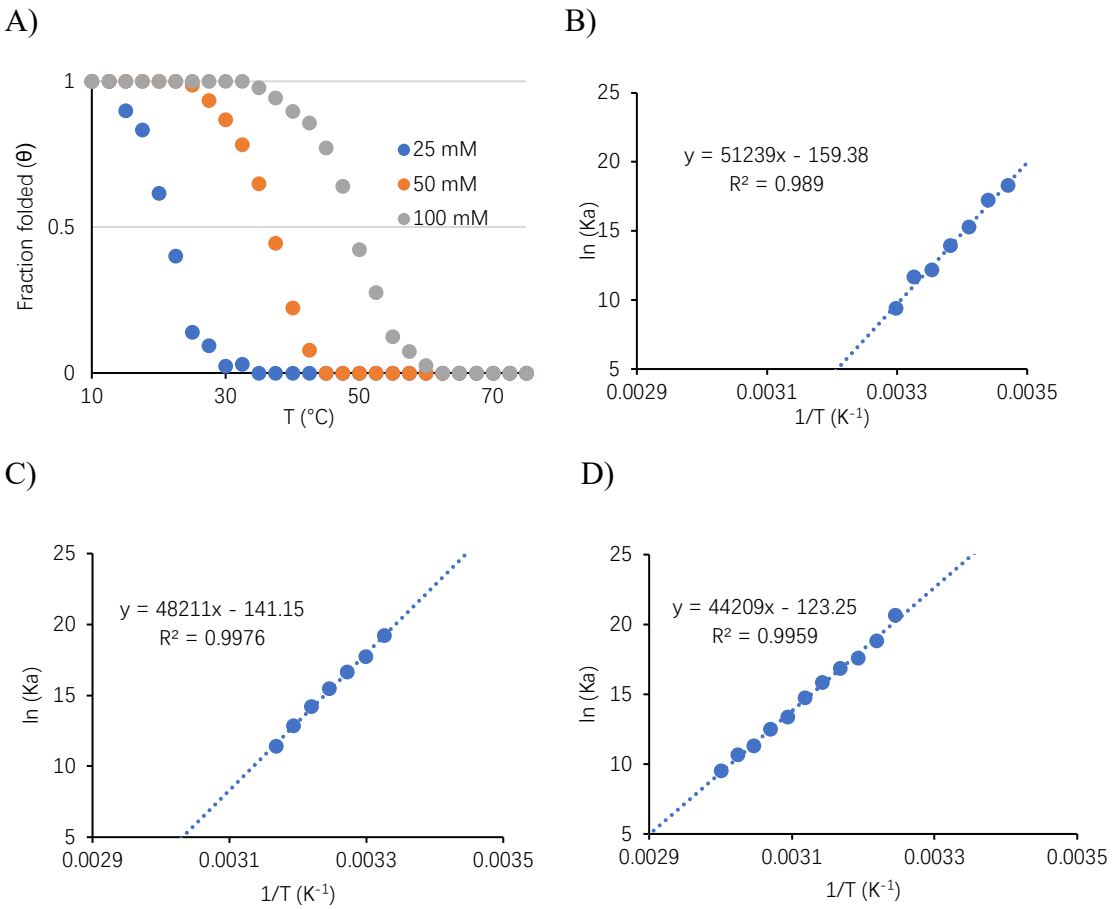


Figure C.9. Determination of thermodynamic parameters. A) Fraction folded as a function of temperature for control **ON1/ON20** duplex. B), C) and D) van't Hoff plot used for determination of ΔH and ΔS at 25, 50 and 100 mM NaCl concentration used respectively. Θ values are extracted from the melting curve obtained by UV-Vis denaturation experiments. Not all Θ points are plotted in this figure, only Θ values significantly higher than 0 and lower than 1 are used.

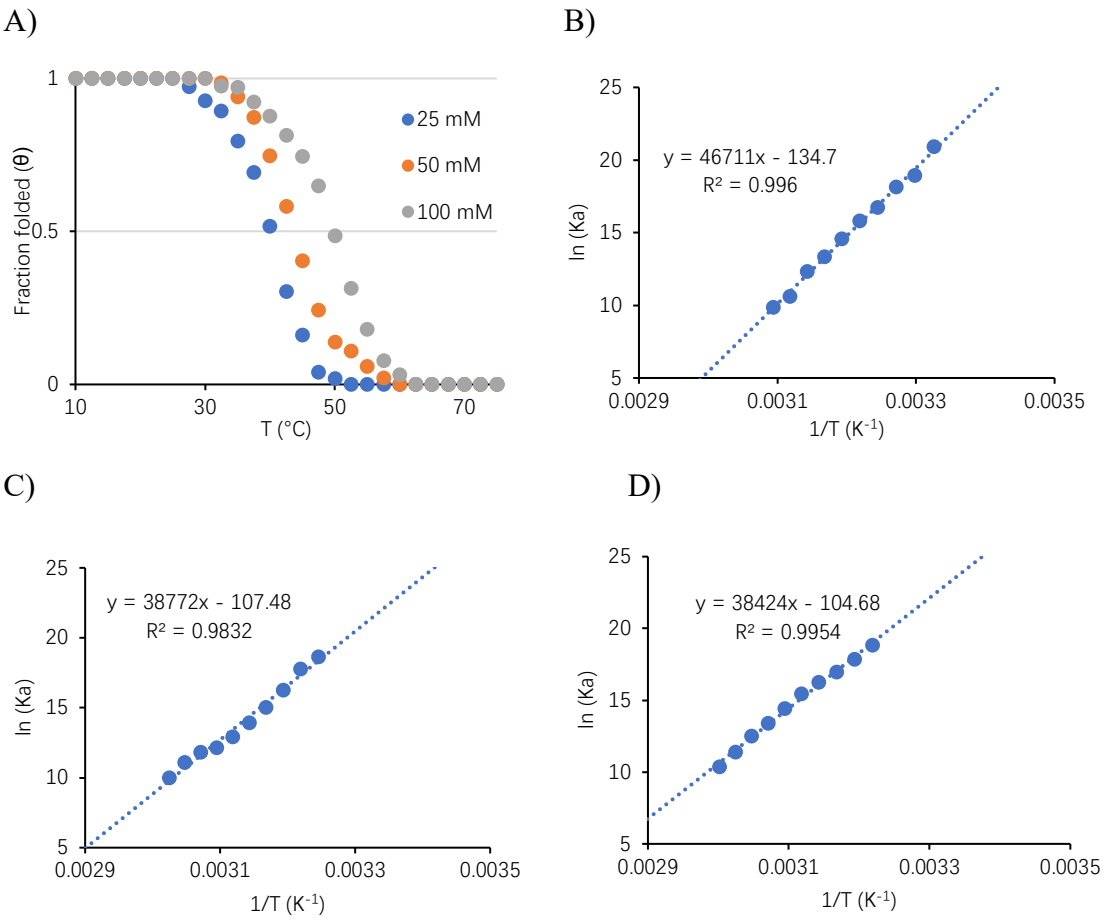


Figure C.10. Determination of thermodynamic parameters. A) Fraction folded as a function of temperature for **4N+ON7/ON20** duplex. B), C) and D) van't Hoff plot used for determination of ΔH and ΔS at 25, 50 and 100 mM NaCl concentration used respectively. Θ values are extracted from the melting curve obtained by UV-Vis denaturation experiments. Not all Θ points are plotted in this figure, only Θ values significantly higher than 0 and lower than 1 are used.

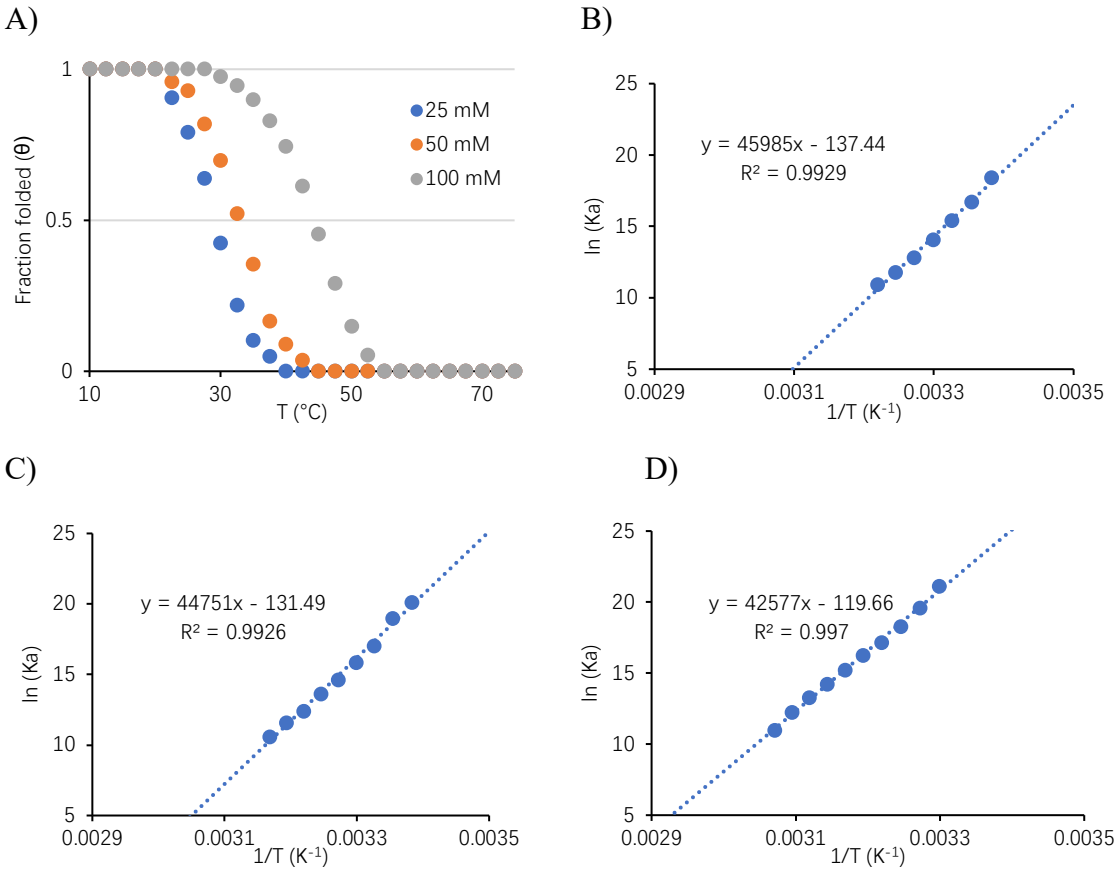


Figure C.11. Determination of thermodynamic parameters. A) Fraction folded as a function of temperature for **4Ts-ON13/ON20** duplex. B), C) and D) van't Hoff plot used for determination of ΔH and ΔS at 25, 50 and 100 mM NaCl concentration used, respectively. Θ values are extracted from the melting curve obtained by UV-Vis denaturation experiments. Not all Θ points are plotted in this figure, only Θ values significantly higher than 0 and lower than 1 are used.

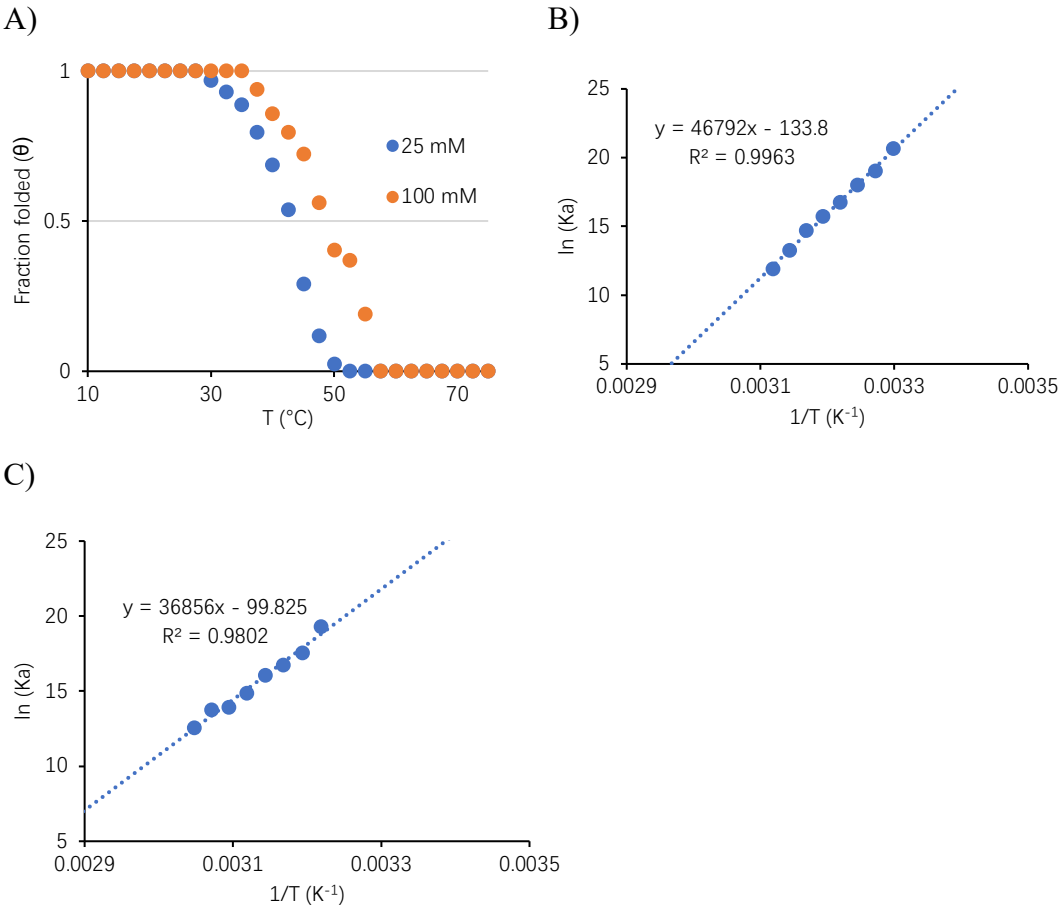


Figure C.12. Determination of thermodynamic parameters. A) Fraction folded as a function of temperature for control ON/RNA duplex: **ON1/ON19**. B), C) and D) van't Hoff plot used for determination of ΔH and $\Delta \mathbf{S}$ at 25, 50 and 100 mM NaCl concentration used, respectively. Θ values are extracted from the melting curve obtained by UV-Vis denaturation experiments. Not all Θ points are plotted in this figure, only Θ values significantly higher than 0 and lower than 1 are used.

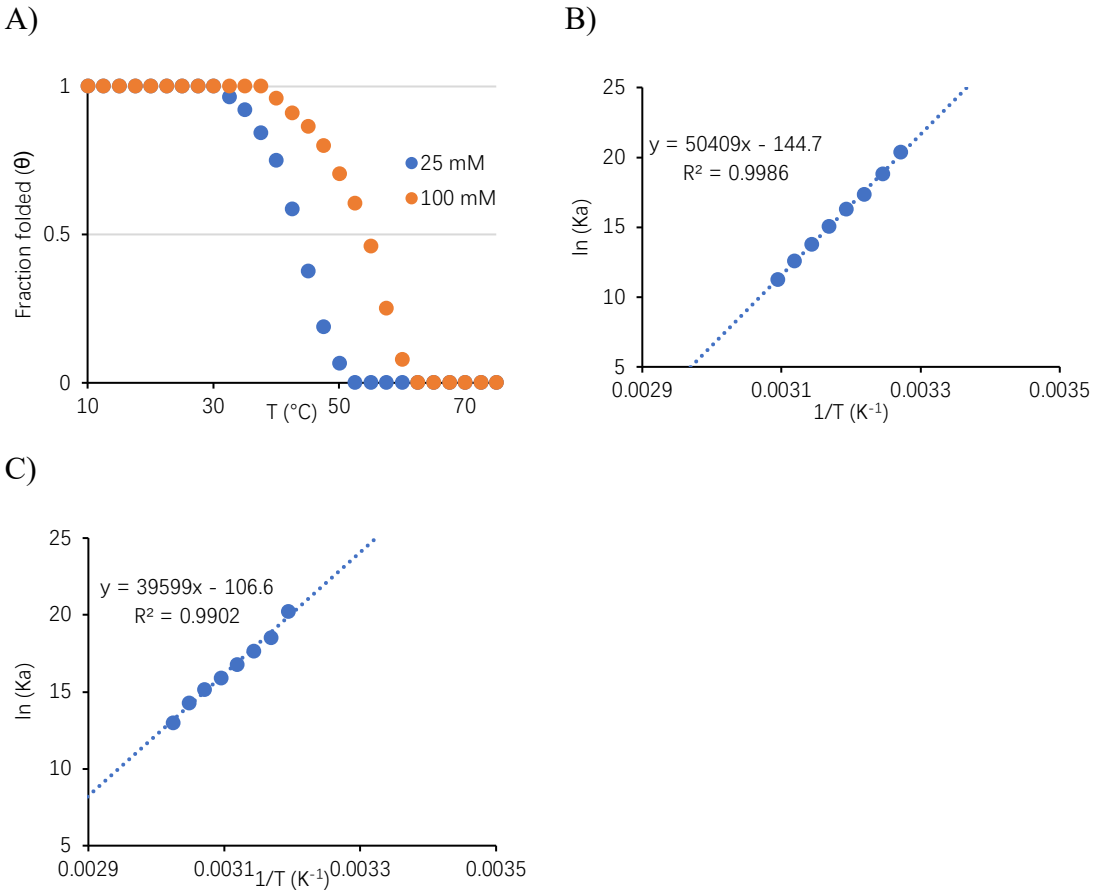


Figure C.13. Determination of thermodynamic parameters. A) Fraction folded as a function of temperature for **4N+ON7/ON19** duplex. B), C) and D) van't Hoff plot used for determination of ΔH and ΔS at 25, 50 and 100 mM NaCl concentration used, respectively. Θ values are extracted from the melting curve obtained by UV-Vis denaturation experiments. Not all Θ points are plotted in this figure, only Θ values significantly higher than 0 and lower than 1 are used.

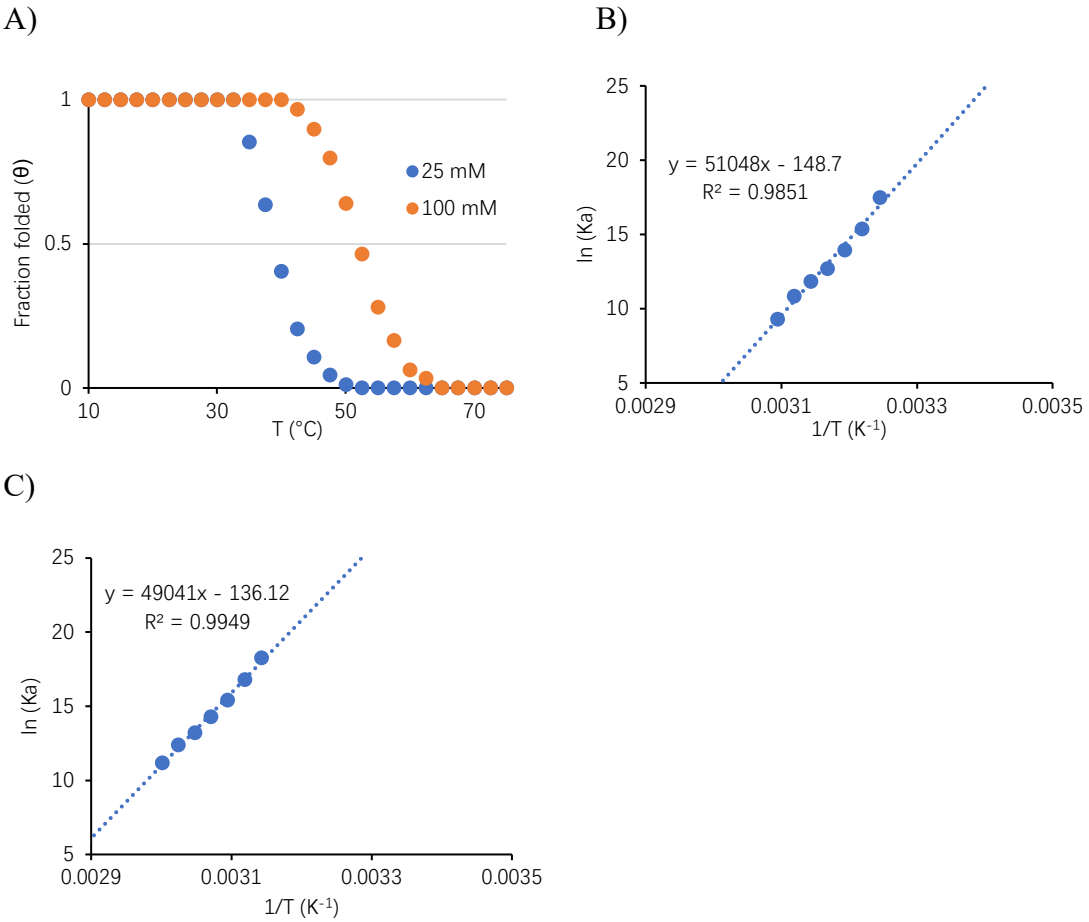


Figure C.14. Determination of thermodynamic parameters. A) Fraction folded as a function of temperature for **4Ts-ON13/ON19** duplex. B), C) and D) van't Hoff plot used for determination of ΔH and ΔS at 25, 50 and 100 mM NaCl concentration used respectively. Θ values are extracted from the melting curve obtained by UV-Vis denaturation experiments. Not all Θ points are plotted in this figure, only Θ values significantly higher than 0 and lower than 1 are used.

4. Evaluation of triplex formation using size-exclusion HPLC

To confirm triplex formation, we performed size-exclusion (SE) HPLC evaluation of several samples at pH 50 and 6.0 at RT. Preformed complex samples (1 µM, 100 µL) were analysed on an Ultimate 3000 HPLC system, equipped with an autosampler, a diode array detector detecting absorbance at 260 nm and a Thermo Acclaim SEC-300 column (4.6 × 300 mm; 5-µm hydrophilic polymethacrylate resin spherical particles, 300 Å pore size). 10 mM Na-cacodylate buffer (pH 5.0 and pH 6.0, respectively) supplemented with 100 mM NaCl and 10 mM $MgCl_2$ was used as a mobile phase.

A) B)

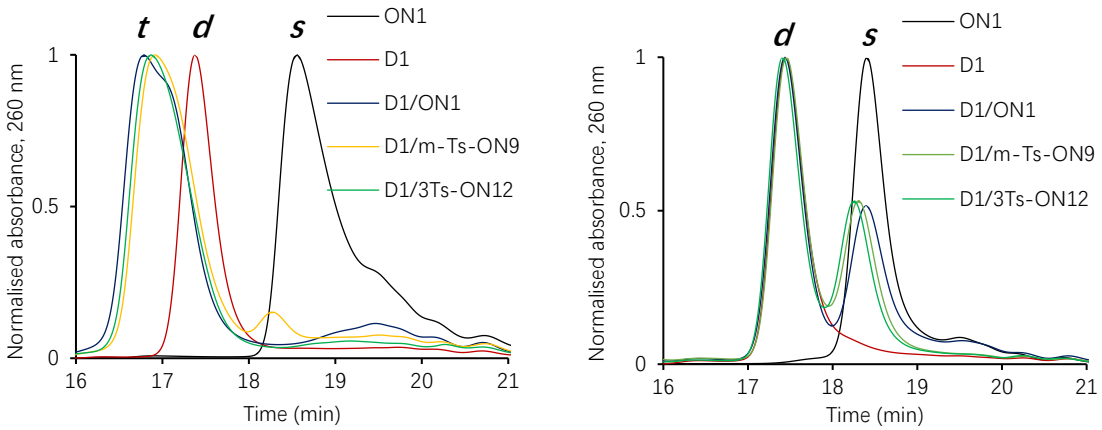


Figure C.15. A) SE-HPLC profiles confirming formation of a triplex at pH 5.0; B) SE-HPLC spectra confirming no triplex formation at pH 6.0. Triplexes are indicated with a *t*, duplexes with a *d*, and *s* stands for single-stranded DNA.

As shown in Fig. C.15B, samples of **D1/ON1**, **D1/m-Ts-ON9** and **D1/3Ts-ON12** at pH 6.0 showed no triplex formation at RT. Only peaks corresponding to the duplex **D1** (17.4 min) and single-stranded DNAs (ssDNAs, ~ 18.5 min) were observed. ssDNAs have similar retention time to **ON1** (18.3 min) in SE-HPLC. In contrast, the same samples at pH 5.0 showed new peaks with faster retention time at 16.8 min, corresponding to DNA triplexes.

5. Enzymatic digestion of ONs by phosphodiesterase I

Enzymatic digestion experiments were carried out under the following conditions: 7.5 µM ON concentration, 0.16 units/L of snake venom phosphodiesterase (Sigma) in 5 mM Tris-HCl buffer (10 mM $MgCl_2$, pH 8.0) in a final volume of 60 µL. Samples were incubated at 37 °C and 10 µL samples were collected at 0, 10, 30, 60 and 120 min, respectively, heated at 90 °C for 5 min to deactivate the enzyme. Aliquots of the samples were analysed by SE-HPLC (same buffer as above mentioned, 0.15 mL/min) to evaluate the amount of intact ON remaining. The percentage of intact ON in each sample was calculated and plotted against the digestion time to obtain a degradation curve (Fig. 4.3, main text in Chapter 4). Percentage of intact ON was determined by the ratio of full-length ONs at each time evaluated to the sample at 0 min.

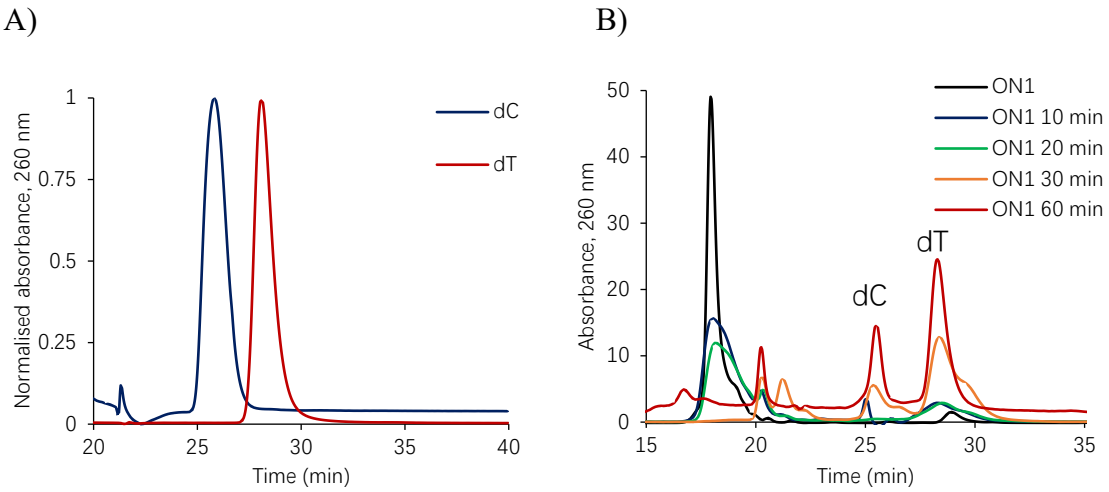


Figure C.17. SE-HPLC profiles of A) 2'-deoxycytidine (dC) and thymidine (dT); B) Unmodified sequence **ON1** after incubation in 5 mM Tris-HCl buffer (10 mM $MgCl_2$, pH 8.0) containing 0.16 units/mL of snake venom phosphodiesterase at 0, 10, 30, and 60 min, respectively.

6. Cell culture and uptake assay

The mouse fibroblast cell line NIH 3T3 (ATCC) was maintained in DMEM (Gibco, Thermofisher Scientific) with 1% penicillin/streptomycin (Gibco) and 10% calf serum (Thermofisher Scientific) at 37 °C with 5% CO_2 in a humidified atmosphere.

Ts- and N+ modified FAM labelled ONs were diluted in Opti-MEM medium (Gibco) and then added to a final concentration of 20 μM to NIH 3T3 cells growing asynchronously on fibronectin coated coverslips. After 12 h, cells were washed with Dulbecco's phosphate buffered saline with Ca^{2+} and Mg^{2+} (DPBS, Thermofisher Scientific), fixed in 4% paraformaldehyde, washed again with DPBS before staining with 1 μg/mL Hoechst 3342. The coverslips were then rinsed in DPBS and mounted in Slowfade Diamond Antifade Mountant (Invitrogen, Thermofisher Scientific) on a microscope slide. For membrane staining, cells were stained with CellBrite Fix 640 (Biotium) for 15 min at RT before fixation and then processed as above.

Slides were imaged using a Leica SP5 DM6000B Scanning Confocal Microscope equipped with 63x/1.40 objective lens. Probes were excited with 405 nm, 495 nm and 640 nm excitation lasers, running LAS X software (Leica) and digitally processed for presentation with Affinity Designer v1.6.1 (Serif Ltd).

Appendix D for Chapter 5

1. Synthesis and purification of μ- and PG- modified ONs

Mesyl- (μ-) and phosphoryl guanidine (PG-) modified ONs were synthesised on an automated DNA/RNA synthesiser ASM-800 (Biosset, Russia) on a 0.4 μmol scale using 5'-dimethoxytrityl (DMTr) G(ib) and T 2-cyanoethyl-*N*,*N*-diisopropyl 3'-phosphoramidites and a Controlled Pore Glass (CPG) 1000Å support loaded with T (25-35 μmol/g) (Sigma-Aldrich, USA) in the DMTr-OFF mode using manufacturer's standard protocols with the following modifications. The μ-ONs and PG-ONs were obtained by replacing the iodine oxidation step using either Mesyl azide or 2-azido-1,3-dimethylimidazolinium hexafluorophosphate (TCI, Japan) in the automated solid-phase synthesis protocol by Staudinger reaction as described in Appendix B.

After the end of the synthesis, ONs were cleaved from solid support and deprotected by treatment with AMA reagent (25% aq. ammonia – 40% aq. methylamine 1:1 v/v) at 55 °C for 15 min. After the removal of volatiles *in vacuo*, the oligonucleotides were dissolved in water (μ-oligonucleotides) or 50% acetonitrile (PG oligonucleotides) and subjected to HPLC analysis and purification as follows.

Analytical RP-HPLC was performed on an Agilent 1220 system (Agilent Technologies, USA) with UV detection at 260 nm using a ZORBAX Eclipse XDB-C18 5 μm 4.6 × 150 mm column (Agilent Technologies, USA) eluted with a 0–60% gradient of acetonitrile (eluent B) in 20 mM TEAA, pH 7.0 (eluent B) for 30 min, flow rate 1 ml/min. Semi-preparative isolation of PG oligonucleotides was carried out on a Waters 600E chromatograph (Waters Corp., USA) with UV detection at 190, 260 and 280 nm using a ZORBAX Eclipse PrepHT XDB-C18 column (7 μm, 21.2 × 150 mm, Agilent Technologies, USA) and the same gradient of eluent B in eluent A at flow rate of 21 ml/min. Fractions containing the product were collected and lyophilized. Oligonucleotide concentration was calculated by measuring Abs_{260} of the solutions on a NanoDrop 2000c UV-Vis spectrophotometer (ThermoFisher Scientific, USA).

Electrophoretic analysis and purification to homogeneity of the μ-ONs were performed in 20% polyacrylamide gel (acrylamide – *N*,*N*'-methylene-bis-acrylamide 30: 1) using 1× TBE buffer (90 mM tris-borate, 2 mM Na_2-EDTA, pH 8.3) containing 8 M urea. Oligonucleotide solutions containing 8 M urea, 0.05% Xylene Cyanol FF and 0.05% Bromophenol Blue were loaded on the gel. The bands in the analytical gel were visualised by staining with Stains-All after electrophoresis. The bands in the preparative gel were visualised by UV shadowing, excised and the μ-oligonucleotides were eluted by 0.3 M

$LiClO_4$, desalted on NAP-25 column eluted by water, lyophilised, redissolved in water, passed through NAP-25 column again and lyophilised.

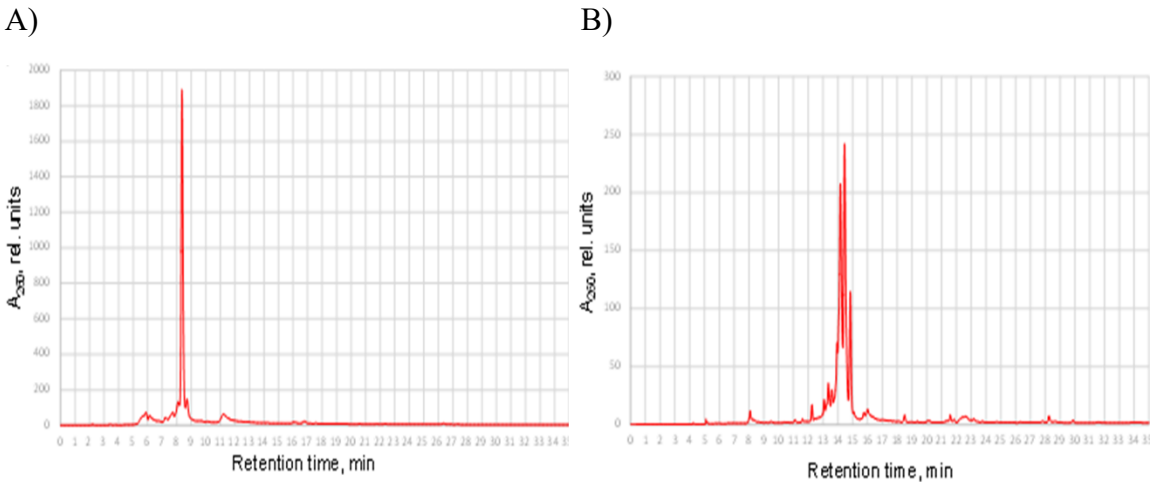


Figure D.1. Analytical RP-HPLC elution profiles of the purified oligonucleotides. A) µ-TG_4T and B) PG-TG_4T. Multiple peaks correspond to diastereomers.

2. Synthesis and purification of N+TG_4T

N+TG_4T sequence was synthesised as described in Appendix B.

Purification of N+ TG_4T sequence was accomplished using IE-HPLC on a TSKgel Super Q-5PW column. Buffer A (0.01 M NaOH, pH 12.0), buffer B (0.01M NaOH, 1 mM EDTA, 2M NaCl, pH 12.0). Gradients: 3.7 min 100% A, convex curve gradient to 30% B in 11.1 min, linear gradient to 50% B in 18.5 min, concave gradient to 100% B in 7.4 min, keep 100% B for 7.4 min and then 100% A in 7.3 min. Collected fractions (Fig. D.2) were desalted using NAP-25 column. The composition of each fraction was analysed by ESI-MS.

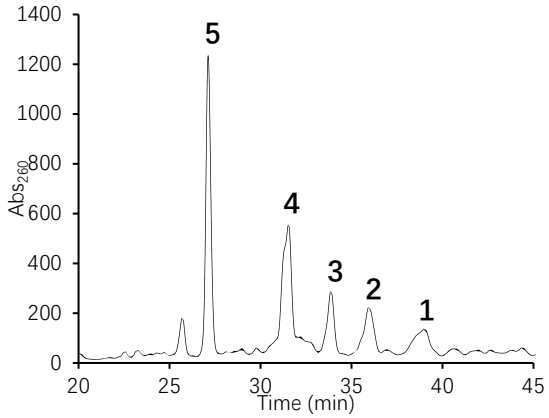


Figure D.2. Preparative IE-HPLC elution profiles of the crude N+TG_4T, numbers above the peaks indicate number of N+ modifications in the sequence.

3. ESI-MS of single-stranded ONs

Collected fractions form IE-HPLC were prepared at 4 μM strand concentration in water: methanol (4: 1, HPLC grade from Fisher Scientific). ESI-MS spectra were obtained using Thermo Scientific Q-Exactive Focus Hybrid Quadrupole-Orbitrap Mass Spectrometer. Samples (5 μL) were injected via Dionex Ultimate 3000 HPLC system running at 0.1 mL/min in CH_3OH.

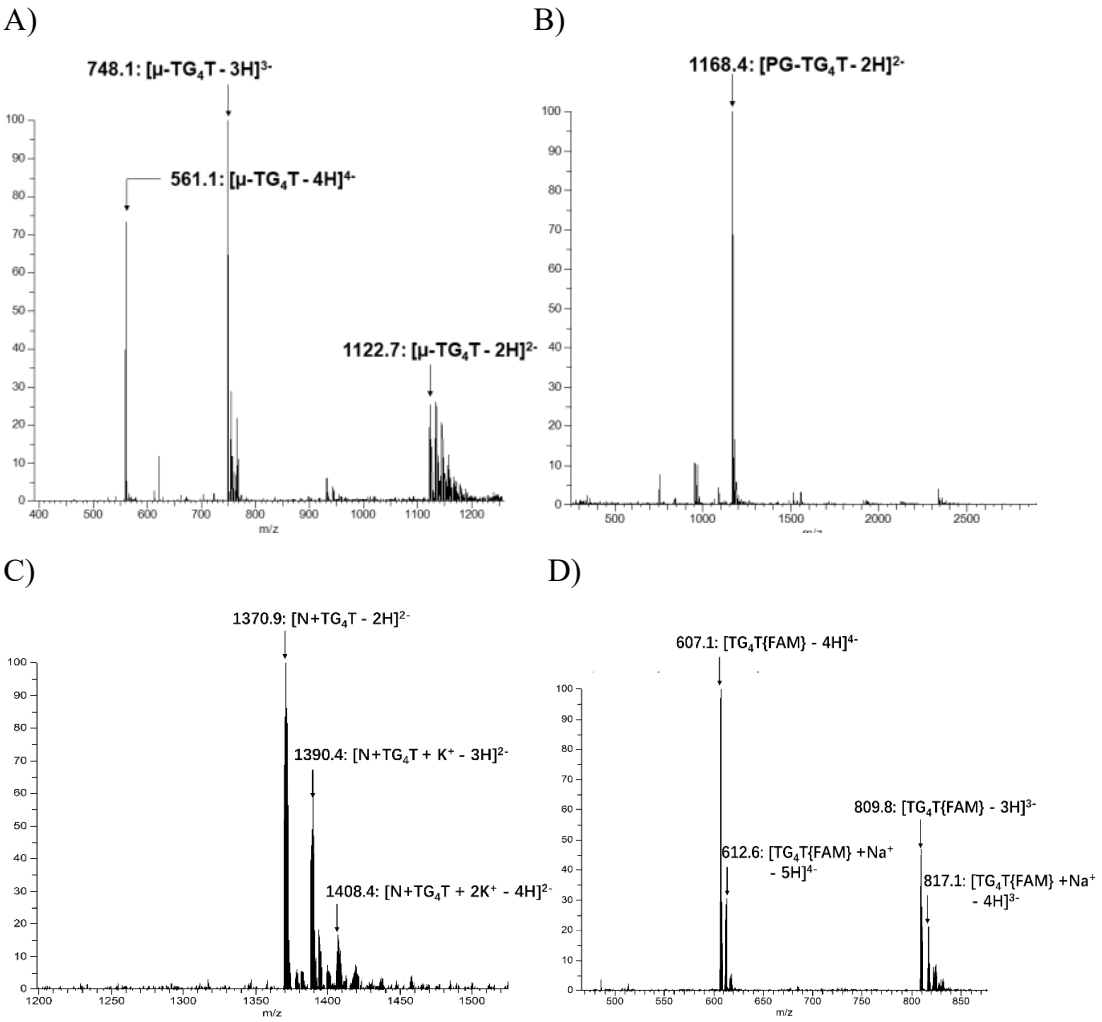


Figure D.3. ESI-MS spectra of purified single-stranded ONs: A) μ-TG_4T, MW = 2248.8, calcd for [μ-TG_4T-3H]$^{3-}$: 748.6, found: 748.1; B) PG-TG_4T, MW = 2337.8, calcd for [PG-TG_4T-2H]$^{2-}$: 1167.9, found: 1168.4; C) N+TG_4T, MW = 2742.8, calcd for [N+TG_4T-2H]$^{2-}$: 1370.4, found: 1370.9; D) TG_4T{FAM}, MW = 2432.7, calcd for [TG_4T{FAM}-4H]$^{4-}$: 607.2, found: 607.1.

3.1. ESI-MS under native conditions

Modified and natural G-rich sequences were dissolved in 150 mM NH_4OAc (pH 7.0) at 300 μM strand concentration, preheated for 5 min and allowed to cool down to RT. Samples were left at 5 ºC for three to four days to ensure G4 formation. Prior ESI-MS analysis, samples were diluted to 30 μM strand concentration with water: methanol 4: 1 (HPLC grade from Fisher Scientific). Results are presented in Fig. 5.10 in Chapter 5.

4. Determination of G4 association rate (k_{on})

G4-association was monitored at initial strand concentrations, solution conditions, and temperatures indicated (in Chapter 5). Samples were incubated at 90 – 100 °C for 30 min to dissociate any secondary structures. As soon as the samples were denatured, changes in CD signal at 262 nm were then recorded every 0.25 s for 3 h at 5 °C.

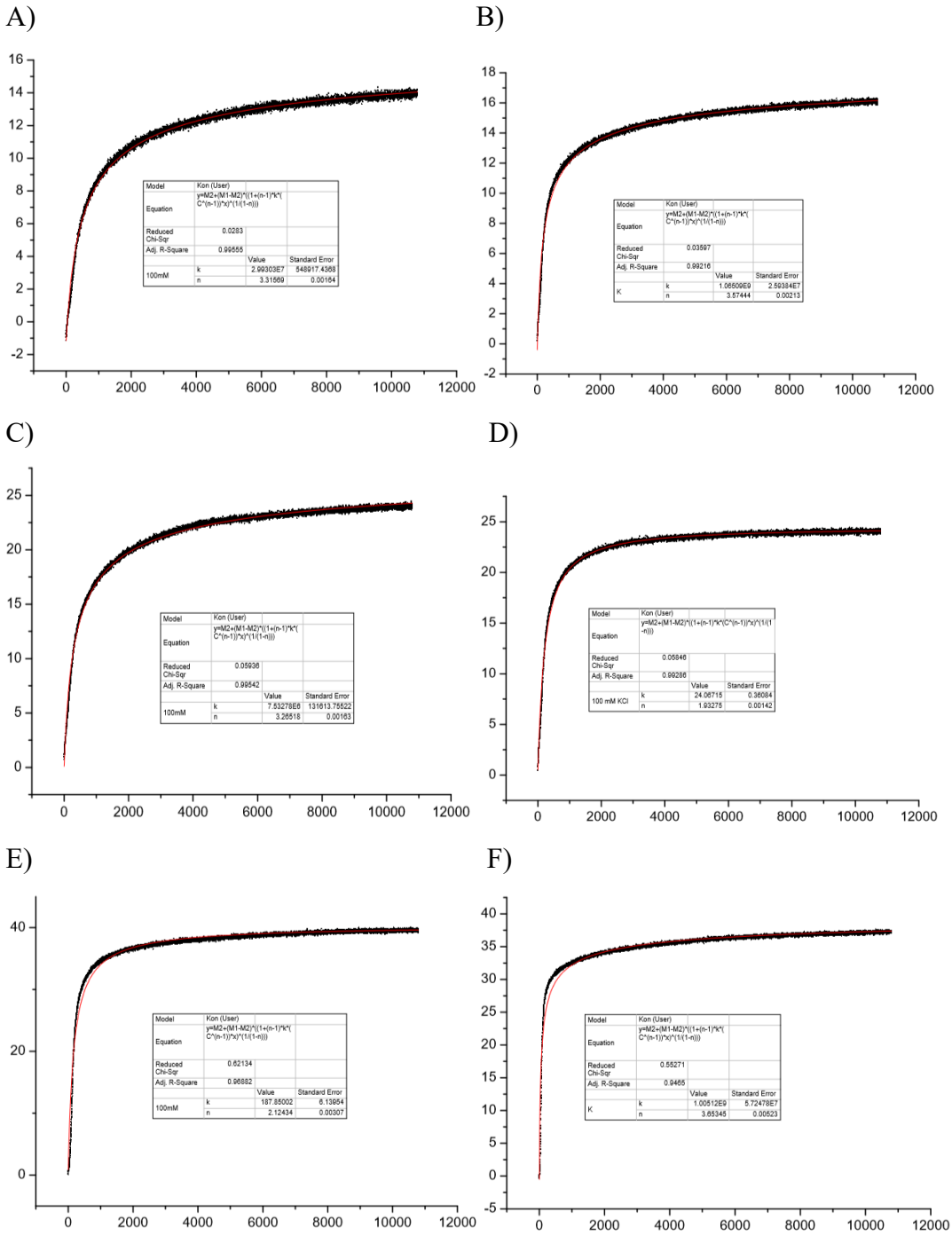


Figure D.4. Isothermal G4 association at 5 °C (black line) and mathematical fits (Equation 1, red line) for complexes formed by A) and B) d(TG_4T); C) and D) μ-TG_4T; E) and F) PG-TG_4T in NaCl (left) and KCl (right).

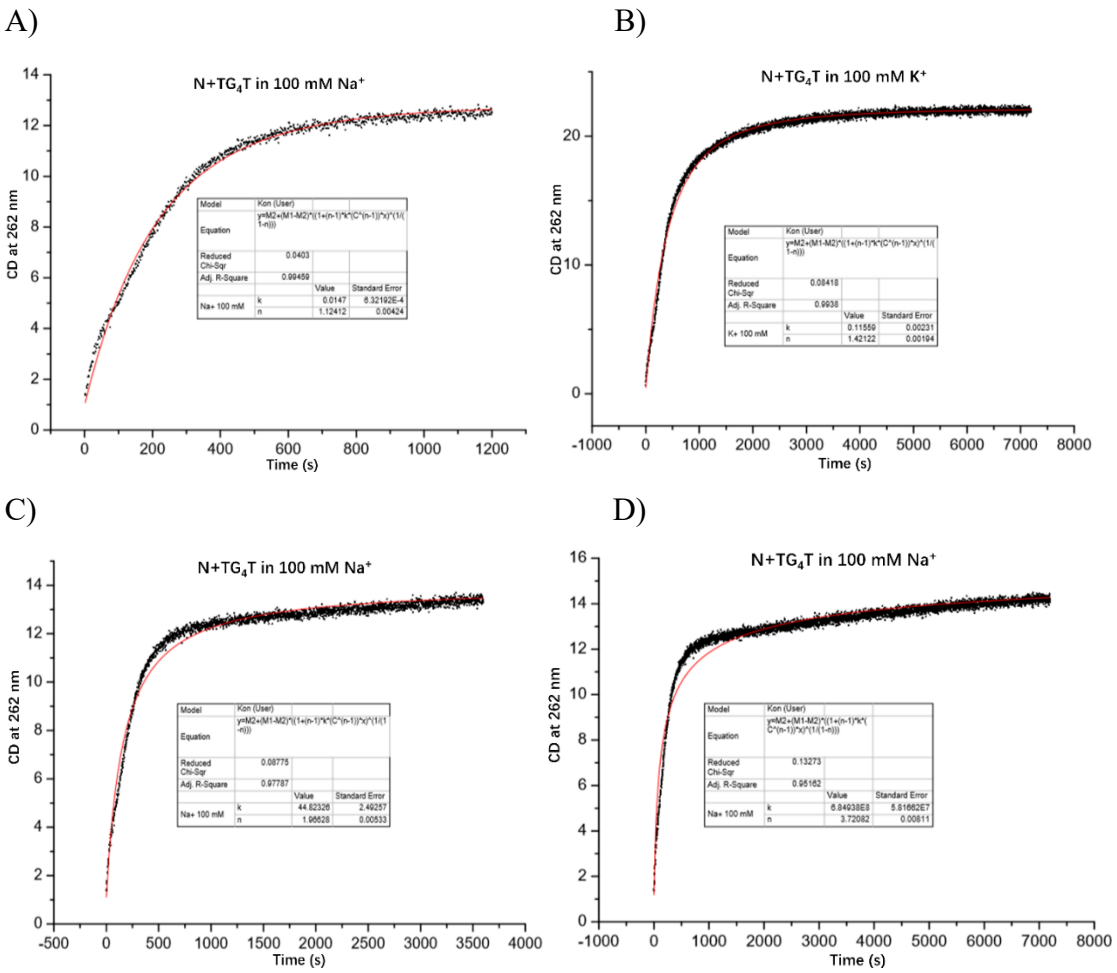


Figure D.5. G4 association at 5 °C (black line) and mathematical fits (red line) for complexes formed by N+TG$_4$T in 100 mM NaCl (A) and KCl (B). Poor fit ($R^2 < 0.99$) of association curves of N+TG$_4$T in 100 mM NaCl that was obtained for fitting 3600 s (C) and 7200 s (D) of association.

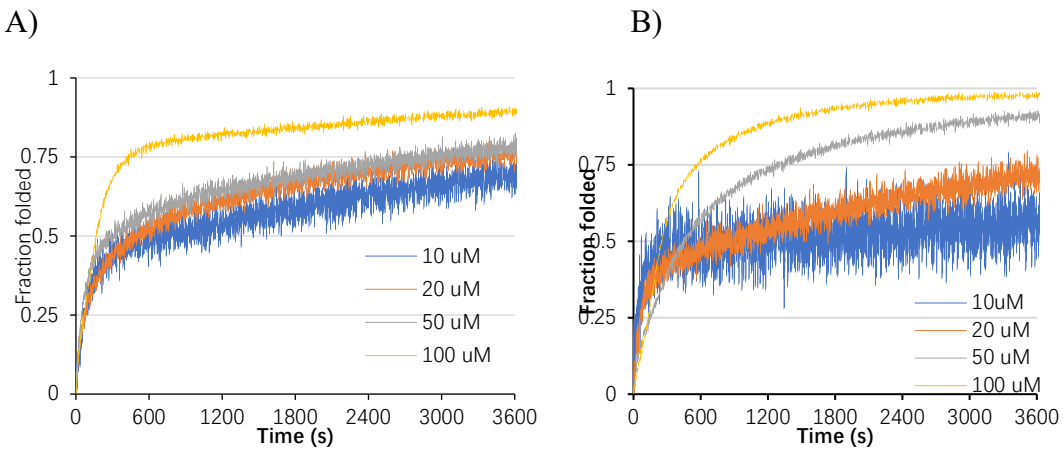


Figure D.6. G4 association at 5 °C for complexes formed by N+TG$_4$T at various strand concentrations (10, 20, 50, amd 100 µM) in 100 mM NaCl (A) and KCl (B).

5. Molecular modelling

Molecular modelling was performed using Mercury 3.10.3. As shown in Fig. D.7, $[TG_4T]_4$ crystal structure (pdb: 2O4F, Fig. D.7A),[239] the distance between the phosphate (orange ball) and the guanine O6 (red ball) in the centre of the G-tetrad is 9.14 Å. The distance between neighbouring phosphates located on different strands and within the same strand in $[TG_4T]_4$ are 8.28 and 6.89 Å, respectively. Molecular modelling of N+ phosphate group was conducted using Spartan '18 V1.4.4. The length of N+ modification between the phosphate and the quaternary amine was calculated as 8.11 Å (Fig. D.7B).

A) B)

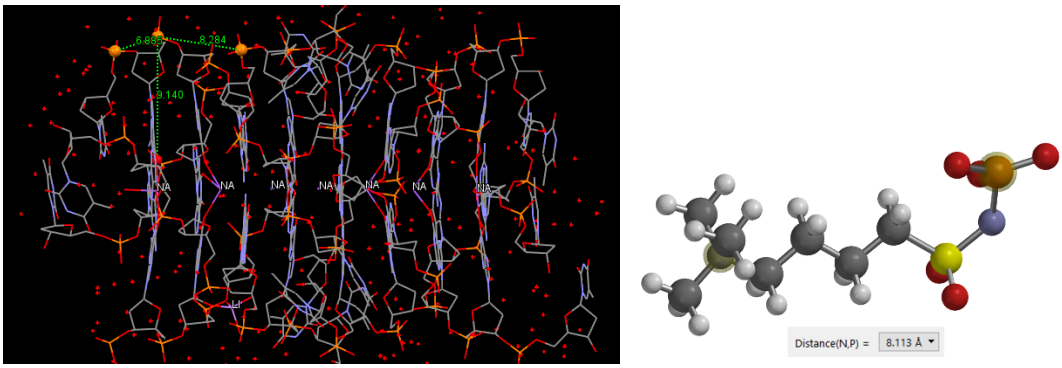


Figure D.7. Molecular modelling of A) parallel $[TG_4T]_4$ quadruplexes with Na^+ cations in the channel; and B) N+ phosphate group showing the distance between the phosphate and the quaternary amine.

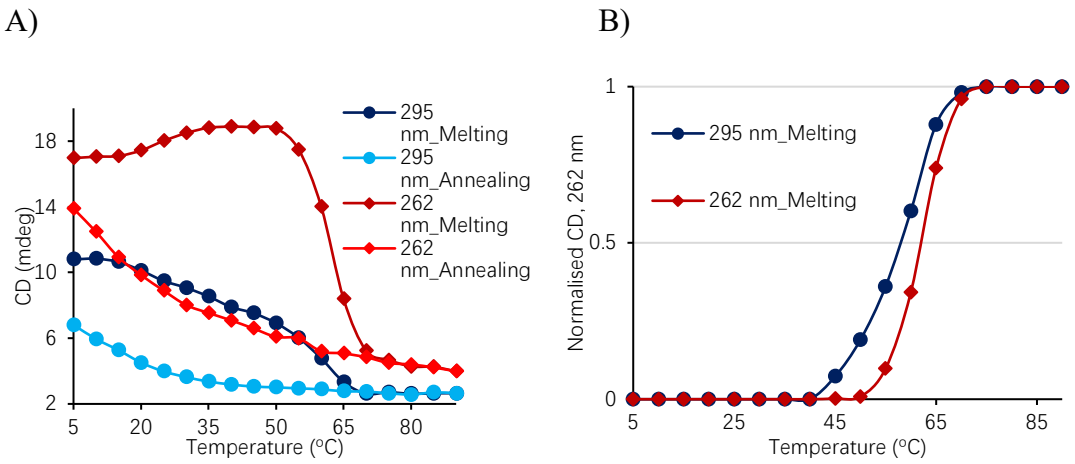


Figure D.8. A) CD melting and annealing curves for $[N^+TG_4T]_4$ at different wavelength, 262 and 295 nm. B) Normalised melting curves of $[N^+TG_4T]_4$ at 262 and 295 nm. $T_{1/2}$ for melting at 262 nm was 65 °C, whereas $T_{1/2}$ for melting at 295 nm was 60 °C.

6. Size-exclusion HPLC for the study of G4 structures

Stock solutions of ONs were prepared in HPLC grade water. Unless otherwise stated, ONs were prepared at 100 µM strand concentration in 10 mM Li-cacodylate buffer (pH 7.2, 100 mM NaCl/KCl), heated at 90 ºC for 5 min and allowed to slowly cool down to RT, then left at 4 ºC for at least overnight before SE-HPLC analysis. 50 mM Tris-HCl buffer (pH 7.5) supplemented with 100 mM NaCl or KCl was prepared in MQ water (18.2 Ω cm^{-1} resistivity) and used as a mobile phase. Preformed complexes were analysed on an Ultimate 3000 HPLC system, equipped with an autosampler, a diode array detector and a Thermo Acclaim SEC-300 column (4.6 × 300 mm; 5-µm hydrophilic polymethacrylate resin spherical particles, 300 Å pore size).

A)

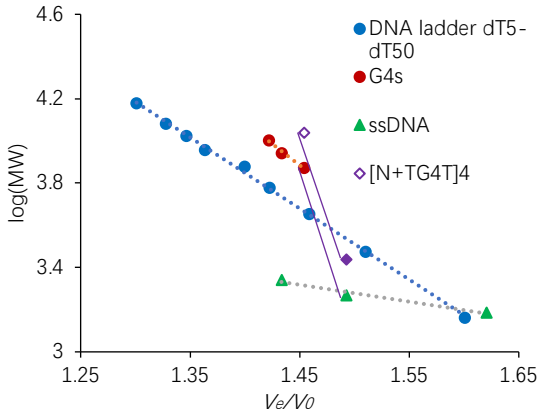


Figure D.9. Plot of the molecular weight decimal logarithm against relative elution volume for DNA ladder d(T_n), G4s (from left to right: $[TG_6T]_4$, $[TG_5T]_4$ and $[TG_4T]_4$), ssDNA (from left to right: TG_5T, TG_4T and TG_3T), $[N+TG_4T]_4$ and single-stranded $N+TG_4T$ in 10 mM Li-cacodylate buffer (pH 7.2, 100 mM Na^+). The line connecting $[N+TG_4T]_4$ and single-stranded $N+TG_4T$ is parallel with the line connecting $[TG_4T]_4$ and single-stranded TG_4T (purple line), indicating that $[N+TG_4T]_4$ forms a tetramolecular structure.

References

[1] Watson, J. D., and Crick, F. H. C. (1953), Molecular structure of nucleic acids: A structure for dioxyribose nucleic acid. *Nature*, *171*, 737-738.

[2] Filichev, V. V., and Pedersen, E. B. (2009), DNA-conjugated organic chromophores in DNA stacking interactions, In *Wiley Encyclopedia of Chemical Biology* (Begley, T., Ed.), pp 493-524, John Wiley & Sons, Inc., Hoboken, N. J.

[3] Lin, C., Ke, Y., Li, Z., Wang, J. H., Liu, Y., and Yan, H. (2009), Mirror image DNA nanostructures for chiral supramolecular assemblies. *Nano Lett.*, *9*, 433-436.

[4] Yakovchuk, P., Protozanova, E., and Frank-Kamenetskii, M. D. (2006), Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.*, *34*, 564-574.

[5] Nielsen., P. E., Egholm., M., Berg., R. H., and Buchardt, O. (1991), Sequence-selective eecorgnition of DNA by strand displacement with a thymine-substituted polyamide. *Science*, *254*, 1497-1500.

[6] Obika, S., Nanbu, D., Hari, Y., Morio, K., In, Y., Ishida, T., and Imanishi, T. (1997), Synthesis of 2'-O,4'-C-methyleneuridine and -cytidine. Novel bicyclic nucleosides having a fixed C-3,-endo sugar puckering. *Tetrahedron Lett.*, *38*, 8735-8738.

[7] Ueda, N., Kawabata, T., and Takemoto, K. (1971), Synthesis of N-(2,3-dihydroxypropyl) derivatives of nucleic bases. *J. Heterocycl. Chem.*, *8*, 827-829.

[8] Basham, B., Schroth, G. P., and Ho, P. S. (1995), An A-DNA triplet code: thermodynamic rules for predicting A- and B-DNA. *Proc. Natl. Acad. Sci. U. S. A.*, *92*, 6464-6468.

[9] Herbert, A., and Rich, A. (1999), Left-handed Z-DNA: structure and function. *Genetica*, *106*, 37-47.

[10] Dickerson, R. E. (1992), DNA structure from A to Z. *Methods Enzymol.*, *211*, 67-111.

[11] Rich, A., and Zhang, S. G. (2003), Z-DNA: the long road to biological function. *Nat. Rev. Genet.*, *4*, 566-572.

[12] Beal, P. A., and Dervan, P. B. (1991), 2nd Structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation. *Science*, *251*, 1360-1363.

[13] Moser, H. E., and Dervan, P. B. (1987), Sequence-specific cleavage of double helical DNA by triplex formation. *Science*, *238*, 645-650.

[14] Wu, Q., Gaddis, S. S., MacLeod, M. C., Walborg, E. F., Thames, H. D., DiGiovanni, J., and Vasquez, K. M. (2007), High-affinity triplex-forming oligonucleotide target sequences in mammalian genomes. *Mol. Carcinog.*, *46*, 15-23.

[15] Hoogsteen, K. (1963), The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Cryst.*, *16*, 907-916.

[16] Doluca, O., Withers, J. M., and Filichev, V. V. (2013), Molecular engineering of guanine-rich sequences: Z-DNA, DNA triplexes, and G-quadruplexes. *Chem. Rev.*, *113*, 3044-3083.

[17] Neidle, S., and Balasubramanian, S. (2006), *Quadruplex nucleic acids*, The Royal Society of Chemistry: London, Cambridge.

[18] Zhang, A. Y., Bugaut, A., and Balasubramanian, S. (2011), A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology. *Biochemistry*, *50*, 7251-7258.

[19] Hazel, P., Huppert, J., Balasubramanian, S., and Neidle, S. (2004), Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.*, *126*, 16405-16415.

[20] Xu, Y. (2011), Chemistry in human telomere biology: structure, function and targeting of telomere DNA/RNA. *Chem. Soc. Rev.*, *40*, 2719-2740.

[21] Huppert, J. L., and Balasubramanian, S. (2007), G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, *35*, 406-413.

[22] Biffi, G., Tannahill, D., Miller, J., Howat, W. J., and Balasubramanian, S. (2014), Elevated levels of G-quadruplex formation in human stomach and liver cancer tissues. *PLoS One*, *9*.

[23] Roach, R. J., Garavis, M., Gonzalez, C., Jameson, G. B., Filichev, V. V., and Hale, T. K. (2020), Heterochromatin protein 1 alpha interacts with parallel RNA and DNA G-quadruplexes. *Nucleic Acids Res.*, *48*, 682-693.

[24] Gehring, K., Leroy, J. L., and Gueron, M. (1993), A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature*, *363*, 561-565.

[25] Benabou, S., Avino, A., Eritja, R., Gonzalez, C., and Gargallo, R. (2014), Fundamental aspects of the nucleic acid i-motif structures. *RSC Adv.*, *4*, 26956-26980.

[26] Zeraati, M., Langley, D. B., Schofield, P., Moye, A. L., Rouet, R., Hughes, W. E., Bryan, T. M., Dinger, M. E., and Christ, D. (2018), I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.*, *10*, 631-637.

[27] Jasin, M. (1996), Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.*, *12*, 224-228.

[28] Stoddard, B. L. (2005), Homing endonuclease structure and function. *Q. Rev. Biophys.*, *38*, 49-95.

[29] de Souza, N. (2012), Primer: genome editing with engineered nucleases. *Nat. Methods*, *9*, 27-27.

[30] Smith, J., Grizot, S., Arnould, S., Duclert, A., Epinat, J. C., Chames, P., Prieto, J., Redondo, P., Blanco, F. J., Bravo, J., Montoya, G., Paques, F., and Duchateau, P. (2006), A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res.*, *34*, 12.

[31] Seligman, L. M., Chevalier, B. S., Chadsey, M. S., Edwards, S. T., Savage, J. H., and Veillet, A. L. (2002), Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Res.*, *30*, 3870-3879.

[32] Laforet, M., McMurrough, T. A., Vu, M., Brown, C. M., Zhang, K., Junop, M. S., Gloor, G. B., and Edgell, D. R. (2019), Modifying a covarying protein-DNA interaction changes substrate preference of a site-specific endonuclease. *Nucleic Acids Res.*, *47*, 10830-10841.

[33] Ashworth, J., Taylor, G. K., Havranek, J. J., Quadri, S. A., Stoddard, B. L., and Baker, D. (2010), Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. *Nucleic Acids Res.*, *38*, 5601-5608.

[34] Montoya, G. (2009), Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. *Hum. Gene Ther.*, *20*, 1027-1027.

[35] Foley, J. E., Maeder, M. L., Pearlberg, J., Joung, J. K., Peterson, R. T., and Yeh, J. R. J. (2009), Targeted mutagenesis in zebrafish using customized zinc-finger nucleases. *Nat. Protoc.*, *4*, 1855-1868.

[36] Kim, Y. G., Cha, J., and Chandrasegaran, S. (1996), Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U.S.A.*, *93*, 1156-1160.

[37] Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., and Gregory, P. D. (2010), Genome editing with engineered zinc finger nucleases. *Nature Rev. Genet.*, *11*, 636-646.

[38] Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y. Q., Fine, E. J., Wu, X. B., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G., and Zhang, F. (2013), DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.*, *31*, 827-832.

[39] Sundaresan, R., Parameshwaran, H. P., Yogesha, S. D., Keilbarth, M. W., and Rajan, R. (2017), RNA-independent DNA cleavage activities of Cas9 and Cas12a. *Cell Rep.*, *21*, 3728-3739.

[40] Lee, J. K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y. H., Lee, K., Jung, I., Kim, D., Kim, S., and Kim, J. S. (2018), Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.*, *9*, 10.

[41] Jensen, E. D., Ferreira, R., Jakociunas, T., Arsovska, D., Zhang, J., Ding, L., Smith, J. D., David, F., Nielsen, J., Jensen, M. K., and Keasling, J. D. (2017), Transcriptional reprogramming in yeast using dCas9 and combinatorial gRNA strategies. *Microb. Cell. Fact.*, *16*, 16.

[42] Saurabh, S., Vidyarthi, A. S., and Prasad, D. (2014), RNA interference: concept to reality in crop improvement. *Planta*, *239*, 543-564.

[43] Kim, D., Lim, K., Kim, S.-T., Yoon, S.-h., Kim, K., Ryu, S.-M., and Kim, J.-S. (2017), Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nature Biotech.*, *35*, 797-797.

[44] Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016), Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, *533*, 420-437.

[45] Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K. Y., Shimatani, Z., and Kondo, A. (2016), Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*, *353*, 8.

[46] Ma, Y. Q., Zhang, J. Y., Yin, W. J., Zhang, Z. C., Song, Y., and Chang, X. (2016), Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods.*, *13*, 1029-1037.

[47] Hess, G. T., Fresard, L., Han, K., Lee, C. H., Li, A., Cimprich, K. A., Montgomery, S. B., and Bassik, M. C. (2016), Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods*, *13*, 1036-1042.

[48] Yang, L. H., Briggs, A. W., Chew, W. L., Mali, P., Guell, M., Aach, J., Goodman, D. B., Cox, D., Kan, Y. A., Lesha, E., Soundararajan, V., Zhang, F., and Church, G. (2016), Engineering and optimising deaminase fusions for genome editing. *Nat. Commun.*, *7*, 11.

[49] Koblan, L. W., Erdos, M. R., Wilson, C., Cabral, W. A., Levy, J. M., Xiong, Z. M., Tavarez, U. L., Davison, L. M., Gete, Y. G., Mao, X., Newby, G. A., Doherty, S. P., Narisu, N., Sheng, Q., Krilow, C., Lin, C. Y., Gordon, L. B., Cao, K., Collins, F. S., Brown, J. D., and Liu, D. R. (2021), In vivo base editing rescues Hutchinson-Gilford progeria syndrome in mice. *Nature*, *589*, 608-614.

[50] Wyvekens, N., Topkar, V. V., Khayter, C., Joung, J. K., and Tsai, S. Q. (2015), Dimeric CRISPR RNA-guided FokI-dCas9 nucleases directed by truncated gRNAs for highly specific genome editing. *Hum. Gene Ther.*, *26*, 425-431.

[51] Woolf, T. M., Gurumurthy, C. B., Boyce, F., and Kmiec, E. B. (2017), To cleave or not to cleave: therapeutic gene editing with and without programmable nucleases. *Nat. Rev. Drug Discov.*, *16*, 296-296.

[52] Desjarlais, J. R., and Berg, J. M. (1993), Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA-binding proteins. *Proc. Natl. Acad. Sci. U.S.A.*, *90*, 2256-2260.

[53] Wu, J., Kandavelou, K., and Chandrasegaran, S. (2007), Custom-designed zinc finger nucleases: What is next? *Cell. Mol. Life Sci.*, *64*, 2933-2944.

[54] Dervan, P. B., and Burli, R. W. (1999), Sequence-specific DNA recognition by polyamides. *Curr. Opin. Chem. Biol.*, *3*, 688-693.

[55] Simon, P., Cannata, F., Perrouault, L., Halby, L., Concordet, J. P., Boutorine, A., Ryabinin, V., Sinyakov, A., and Giovannangeli, C. (2008), Sequence-specific DNA cleavage mediated by bipyridine polyamide conjugates. *Nucleic Acids Res.*, *36*, 3531-3538.

[56] Ghosh, I., Stains, C. I., Ooi, A. T., and Segal, D. J. (2006), Direct detection of double-stranded DNA: Molecular methods and applications for DNA diagnostics. *Mol. Biosyst.*, *2*, 551-560.

[57] Boutorine, A. S., Novopashina, D. S., Krasheninina, O. A., Nozeret, K., and Venyaminova, A. G. (2013), Fluorescent probes for nucleic acid visualization in fixed and live cells. *Molecules*, *18*, 15357-15397.

[58] Hsu, Patrick D., Lander, Eric S., and Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, *157*, 1262-1278.

[59] Buchini, S., and Leumann, C. J. (2003), Recent improvements in antigene technology. *Curr. Opin. Chem. Biol.*, *7*, 717-726.

[60] Gagnon, K. T., Watts, J. K., Pendergraff, H. M., Montaillier, C., Thai, D., Potier, P., and Corey, D. R. (2011), Antisense and antigene inhibition of gene expression by cell-permeable oligonucleotide-oligospermine conjugates. *J. Am. Chem. Soc.*, *133*, 8404-8407.

[61] Praseuth, D., Guieysse, A. L., and Helene, C. (1999), Triple helix formation and the antigene strategy for sequence-specific control of gene expression. *Biochim. Biophys. Acta-Gene Struct. Expr.*, *1489*, 181-206.

[62] Duca, M., Vekhoff, P., Oussedik, K., Halby, L., and Arimondo, P. B. (2008), The triple helix: 50 years later, the outcome. *Nucleic Acids Res.*, *36*, 5123-5138.

[63] Simon, P., Cannata, F., Concordet, J. P., and Giovannangeli, C. (2008), Targeting DNA with triplex-forming oligonucleotides to modify gene sequence. *Biochimie*, *90*, 1109-1116.

[64] Nielsen, P. E., Egholm, M., and Buchardt, O. (1994), Evidence for (PNA)2/DNA triplex structure upon binding of PNA to dsDNA by strand displacement. *J. Mol. Recog.*, *7*, 165-170.

[65] Hansen, G. I., Bentin, T., Larsen, H. J., and Nielsen, P. E. (2001), Structural isomers of bis-PNA bound to a target in duplex DNA11Edited by I. Tinoco. *J. Mol. Biol.*, *307*, 67-74.

[66] Holloman, W. K., Wiegand, R., Hoessli, C., and Radding, C. M. (1975), Uptake of homologous single-stranded fragments by superhelical DNA: A possible mechanism for initiation of genetic recombination. *Proc. Natl. Acad. Sci. U.S.A.*, *72*, 2394-2398.

[67] Lohse, J., Dahl, O., and Nielsen, P. E. (1999), Double duplex invasion by peptide nucleic acid: A general principle for sequence-specific targeting of double-stranded DNA. *Proc. Natl. Acad. Sci. U.S.A.*, *96*, 11804-11808.

[68] Palchaudhuri, R., and Hergenrother, P. J. (2007), DNA as a target for anticancer compounds: Methods to determine the mode of binding and the mechanism of action. *Curr. Opin. Biotechnol.*, *18*, 497-503.

[69] Neidle, S. (2001), DNA minor-groove recognition by small molecules. *Nat. Prod. Rep.*, *18*, 291-309.

[70] Tse, W. C., and Boger, D. L. (2004), Sequence-selective DNA recognition: Natural products and nature's lessons. *Chem. Biol.*, *11*, 1607-1617.

[71] Pelton, J. G., and Wemmer, D. E. (1990), Structure and dynamics of distamycin A with d(CGCAAATTGGC):d(GCCAATTTGCG) at low drug: DNA ratios. *J. Biomol. Struct. Dyn.*, *8*, 81-97.

[72] Kissinger, K., Krowicki, K., Dabrowiak, J. C., and Lown, J. W. (1987), Molecular recognition between oligopeptides and nucleic acids. Monocationic imidazole iexitropsins that display enhanced GC sequence dependent DNA binding. *Biochemistry*, *26*, 5590-5595.

[73] Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., and Dickerson, R. E. (1985), The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc. Natl. Acad. Sci. U.S.A.*, *82*, 1376-1380.

[74] Deng, J. P., Pan, B. C., and Sundaralingam, M. (2003), Structure of d(ITITACAC) complexed with distamycin at 1.6 angstrom resolution. *Acta Crystallogr. D, Biol. Crystallogr.*, *59*, 2342-2344.

[75] McHugh, M. M., Woynarowski, J. M., Sigmund, R. D., and Beerman, T. A. (1989), Effect of minor groove binding drugs on mammalian topoisomerase I activity. *Biochem. Pharmacol.*, *38*, 2323-2328.

[76] Bhaduri, S., Ranjan, N., and Arya, D. P. (2018), An overview of recent advances in duplex DNA recognition by small molecules. *Beilstein J. Org. Chem.*, *14*, 1051-1086.

[77] Gurova, K. (2009), New hopes from old drugs: revisiting DNA-binding small molecules as anticancer agents. *Future Oncol.*, *5*, 1685-1704.

[78] Zimmer, C., and Wahnert, U. (1986), Nonintercalating DNA-binding ligands: Specificity of the interaction and their use as tools in biophysical, biochemical and biological investigations of the genetic material. *Prog. Biophys. Mol. Biol.*, *47*, 31-112.

[79] Pezzoni, G., Grandi, M., Biasoli, G., Capolongo, L., Ballinari, D., Giuliani, F. C., Barbieri, B., Pastori, A., Pesenti, E., Mongelli, N., and Spreafico, F. (1991), Biological profile of FCE 24517, a novel benzoyl mustard analogue of distamycin A. *Br. J. Cancer*, *64*, 1047-1050.

[80] Dalessio, R., Geroni, C., Biasoli, G., Pesenti, E., Grandi, M., and Mongelli, N. (1994), Structure-activity relationship of novel distamycin-A derivatives - synthesis and antitumor-activity. *Bioorg. Med. Chem. Lett.*, *4*, 1467-1472.

[81] Cozzi, P. (2001), A new class of cytotoxic DNA minor groove binders: alpha-halogenoacrylic derivatives of pyrrolecarbamoyl oligomers. *Farmaco*, *56*, 57-65.

[82] Fedier, A., Fowst, C., Tursi, J., Geroni, C., Haller, U., Marchini, S., and Fink, D. (2003), Brostallicin (PNU-166196) – a new DNA minor groove binder that retains sensitivity in DNA mismatch repair-deficient tumour cells. *Brit. J. Cancer*, *89*, 1559-1565.

[83] Rahman, A., O'Sullivan, P., and Rozas, I. (2019), Recent developments in compounds acting in the DNA minor groove. *MedChemComm*, *10*, 26-40.

[84] James M. Turner, Susanne E. Swalley, Eldon E. Baird, and Dervan, P. B. (1998), Aliphatic/aromatic amino acid pairings for polyamide recognition in the minor groove of DNA. *J. Am. Chem. Soc.*, *120*, 6219-6226.

[85] Sarah White, Jason W. Szewczyk, James M. Turner, Baird, E. E., and Dervan, P. B. (1997), Recognition of the four Watson–Crick base pairs in the DNA minor groove by synthetic ligands. *Nature*, *391*, 468-471.

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

[87] Clara L. Kielkopf, Sarah White, Jason W. Szewczyk, James M. Turner, Eldon E. Baird, Peter B. Dervan, and Rees, D. C. (1998), A structural basis for recognition of A·T and T·A base pairs in the minor groove of B-DNA. *Science*, *282*, 111-115.

[88] Hansen, M. E., Bentin, T., and Nielsen, P. E. (2009), High-affinity triplex targeting of double stranded DNA using chemically modified peptide nucleic acid oligomers. *Nucleic Acids Res.*, *37*, 4498-4507.

[89] Sun, B. W., Babu, B. R., Sorensen, M. D., Zakrzewska, K., Wengel, J., and Sun, J. S. (2004), Sequence and pH effects of LNA-containing triple helix-forming oligonucleotides: Physical chemistry, biochemistry, and modeling studies. *Biochemistry*, *43*, 4160-4169.

[90] Pellestor, F., Paulasova, P., Macek, M., and Hamamah, S. (2005), The use of peptide nucleic acids for in situ identification of human chromosomes. *J. Histochem. Cytochem.*, *53*, 395-400.

[91] Ray, A., and Nordén, B. (2000), Peptide nucleic acid (PNA): Its medical and biotechnical applications and promise for the future. *FASEB J.*, *14*, 1041-1060.

[92] Obikaa., S., Nanbua., D., Haria., Y., Morioa., K.-i., Inb., Y., Ishidab., T., and Imanishia., T. (1997), Synthesis of 2′-O,4′-C-methyleneuridine and -cytidine. Novel bicyclic nucleosides having a fixed C3, -endo sugar puckering. *Tetrahedron Lett.*, *38*, 8735 - 8738.

[93] Koshkin, A. A., Nielsen, P., Meldgaard, M., Rajwanshi, V. K., Singh, S. K., and Wengel, J. (1998), LNA (locked nucleic acid): An RNA mimic forming exceedingly stable LNA : LNA duplexes. *J. Am. Chem. Soc.*, *120*, 13252-13253.

[94] Torigoe, H., Hari, Y., Sekiguchi, M., Obika, S., and Imanishi, T. (2001), 2'-O,4'-C-methylene bridged nucleic acid modification promotes pyrimidine motif triplex DNA formation at physiological pH: thermodynamic and kinetic studies. *J. Biol. Chem.*, *276*, 2354-2360.

[95] Brunet, E., Alberti, P., Perrouault, L. C., Babu, R., Wengel, J., and Giovannangeli, C. (2005), Exploring cellular activity of locked nucleic acid-modified triplex-forming oligonucleotides and defining its molecular basis. *J. Biol. Chem.*, *280*, 20076-20085.

[96] Sato, T., Sato, Y., and Nishizawa, S. (2017), Optimization of the alkyl linker of TO base surrogate in triplex-forming PNA for enhanced binding to double-stranded RNA. *Chem. Eur. J.*, *23*, 4079-4088.

[97] Nielsen, P. E. (2010), Peptide nucleic acids (PNA) in chemical biology and drug discovery. *Chem. Biodiv.*, *7*, 786-804.

[98] Zaghloul, E. M., Madsen, A. S., Moreno, P. M., Oprea, II, El-Andaloussi, S., Bestas, B., Gupta, P., Pedersen, E. B., Lundin, K. E., Wengel, J., and Smith, C. I. (2011), Optimizing anti-gene oligonucleotide 'Zorro-LNA' for improved strand invasion into duplex DNA. *Nucleic Acids Res.*, *39*, 1142-1154.

[99] Geny, S., Moreno, Pedro M. D., Krzywkowski, T., Gissberg, O., Andersen, N. K., Isse, A. J., El-Madani, A. M., Lou, C., Pabon, Y. V., Anderson, B. A., Zaghloul, E. M., Zain, R., Hrdlicka, P. J., Jørgensen, P. T., Nilsson, M., Lundin, K. E., Pedersen, E. B., Wengel, J., and Smith, C. I. E. (2016), Next-generation bis-locked nucleic acids with stacking linker and 2′-glycylamino-LNA show enhanced DNA invasion into supercoiled duplexes. *Nucleic Acids Res.*, *44*, 2007-2019.

[100] Hansen, G. I., Bentin, T., Larsen, H. J., and Nielsen, P. E. (2001), Structural isomers of bis-PNA bound to a target in duplex DNA. *J. Mol. Biol.*, *307*, 67-74.

[101] Kaihatsu, K., Janowski, B. A., and Corey, D. R. (2004), Recognition of chromosomal DNA by PNAs. *Chem. Biol.*, *11*, 749-758.

[102] Ishizuka, T., Yoshida, J., Yamamoto, Y., Sumaoka, J., Tedeschi, T., Corradini, R., Sforza, S., and Komiyama, M. (2008), Chiral introduction of positive charges to PNA for double-duplex invasion to versatile sequences. *Nucleic Acids Res.*, *36*, 1464-1471.

[103] Ishizuka, T., Otani, K., Sumaoka, J., and Komiyama, M. (2009), Strand invasion of conventional PNA to arbitrary sequence in DNA assisted by single-stranded DNA binding protein. *Chem. Commun.*, 1225-1227.

[104] Asanuma, H., Niwa, R., Akahane, M., Murayama, K., Kashida, H., and Kamiya, Y. (2016), Strand-invading linear probe combined with unmodified PNA. *Bioorg. Med. Chem.*, *24*, 4129-4137.

[105] Nakamura, S., Kawabata, H., and Fujimoto, K. (2017), Double duplex invasion of DNA induced by ultrafast photo-cross-linking using 3-cyanovinylcarbazole for antigene methods. *Chem. Commun.*, *53*, 7616-7619.

[106] Manicardi, A., Gyssels, E., Corradini, R., and Madder, A. (2016), Furan-PNA: a mildly inducible irreversible interstrand crosslinking system targeting single and double stranded DNA. *Chem. Comm.*, *52*, 6930-6933.

[107] Tilani, N., De Costa, S., and Heemstra, J. M. (2013), Evaluating the effect of ionic strength on duplex stability for PNA having negatively or positively charged side chains. *PLoS One*, *8*, 8.

[108] Ganesh, K. N., and Nielsen, P. E. (2000), Peptide nucleic acids: Analogs and derivatives. *Curr. Org. Chem.*, *4*, 931-943.

[109] Bohländer, P. R., Vilaivan, T., and Wagenknecht, H.-A. (2015), Strand displacement and duplex invasion into double-stranded DNA by pyrrolidinyl peptide nucleic acids. *Org. Biomol. Chem.*, *13*, 9223-9230.

[110] Nolling, J., Rapireddy, S., Amburg, J. I., Crawford, E. M., Prakash, R. A., Rabson, A. R., Tang, Y. W., and Singer, A. (2016), Duplex DNA-invading gamma-modified peptide nucleic acids enable rapid identification of bloodstream infections in whole blood. *mBio.*, *7*, e00345-00316.

[111] Geny, S., Moreno, P. M., Krzywkowski, T., Gissberg, O., Andersen, N. K., Isse, A. J., El-Madani, A. M., Lou, C., Pabon, Y. V., Anderson, B. A., Zaghloul, E. M., Zain, R., Hrdlicka, P. J., Jorgensen, P. T., Nilsson, M., Lundin, K. E., Pedersen, E. B., Wengel, J., and Smith, C. I. (2016), Next-generation bis-locked nucleic acids with stacking linker and 2'-glycylamino-LNA show enhanced DNA invasion into supercoiled duplexes. *Nucleic Acids Res.*, *44*, 2007-2019.

[112] Didion, B. A., Karmakar, S., Guenther, D. C., Sau, S. P., Verstegen, J. P., and Hrdlicka, P. J. (2013), Invaders: Recognition of double-stranded DNA by using duplexes modified with interstrand zippers of 2′-O-(Pyren-1-yl)methyl-ribonucleotides. *ChemBioChem*, *14*, 1534-1538.

[113] Hrdlicka, P. J., Kumar, T. S., and Wengel, J. (2005), Targeting of mixed sequence double-stranded DNA using pyrene-functionalized 2'-amino-alpha-L-LNA. *Chem. Commun.*, 4279-4281.

[114] Kumar, T. S., Madsen, A. S., Ostergaard, M. E., Sau, S. P., Wengel, J., and Hrdlicka, P. J. (2009), Functionalized 2'-amino-alpha-L-LNA: directed positioning of intercalators for DNA targeting. *J. Org. Chem.*, *74*, 1070-1081.

[115] Sau, S. P., Kumar, T. S., and Hrdlicka, P. J. (2010), Invader LNA: Efficient targeting of short double stranded DNA. *Org. Biomol. Chem.*, *8*, 2028-2036.

[116] Karmakar, S., Guenther, D. C., and Hrdlicka, P. J. (2013), 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.

[117] Stirchak, E. P., Summerton, J. E., and Weller, D. D. (1989), Uncharged stereoregular nucleic-acid analogs: 2. Morpholino nucleoside oligomers with carbamate internucleoside linkages. *Nucleic Acids Res.*, *17*, 6129-6141.

[118] Alter, J., Lou, F., Rabinowitz, A., Yin, H. F., Rosenfeld, J., Wilton, S. D., Partridge, T. A., and Lu, Q. L. (2006), Systemic delivery of morpholino oligonucleotide restores dystrophin expression bodywide and improves dystrophic pathology. *Nat. Med.*, *12*, 175-177.

[119] Summerton, J. E. (2007), Morpholino, siRNA, and S-DNA compared: Impact of structure and mechanism of action on off-target effects and sequence specificity. *Curr. Top. Med. Chem.*, *7*, 651-660.

[120] Nan, Y. C., and Zhang, Y. J. (2018), Antisense Phosphorodiamidate Morpholino Oligomers as Novel Antiviral Compounds. *Front. Microbiol.*, *9*, 15.

[121] Wu, B., Moulton, H. M., Iversen, P. L., Jiang, J. G., Li, J., Li, J. B., Spurney, C. F., Sali, A., Guerron, A. D., Nagaraju, K., Doran, T., Lu, P. J., Xiao, X., and Lu, Q. L. (2008), Effective rescue of dystrophin improves cardiac function in dystrophin-deficient mice by a modified morpholino oligomer. *Proc. Natl. Acad. Sci. U. S. A.*, *105*, 14814-14819.

[122] Amantana, A., and Iversen, P. L. (2005), Pharmacokinetics and biodistribution of phosphorodiamidate morpholino antisense oligomers. *Curr. Opin. Pharm.*, *5*, 550-555.

[123] Nielsen, C. B., Petersen, M., Pedersen, E. B., Hansen, P. E., and Christensen, U. B. (2004), NMR structure determination of a modified DNA oligonucleotide containing a new intercalating nucleic acid. *Bioconjug. Chem.*, *15*, 260-269.

[124] Christensen, U. B., and Pedersen, E. B. (2002), Intercalating nucleic acids containing insertions of 1-*O*-(1-pyrenylmethyl)glycerol: stabilisation of dsDNA and discrimination of DNA over RNA. *Nucleic Acids Res.*, *30*, 4918-4925.

[125] Filichev, V. V., Christensen, U. B., Pedersen, E. B., Babu, B. R., and Wengel, J. (2004), Locked nucleic acids and intercalating nucleic acids in the design of easily denaturing nucleic acids: Thermal stability studies. *ChemBioChem*, *5*, 1673-1679.

[126] Filichev, V. V., Vester, B., Hansen, L. H., and Pedersen, E. B. (2005), Easily denaturing nucleic acids derived from intercalating nucleic acids: thermal stability studies, dual duplex invasion and inhibition of transcription start. *Nucleic Acids Res.*, *33*, 7129-7137.

[127] Filichev, V. V., and Pedersen, E. B. (2005), Stable and selective formation of Hoogsteen-type triplexes and duplexes using twisted intercalating nucleic acids (TINA) prepared via postsynthetic Sonogashira solid-phase coupling reactions. *J. Am. Chem. Soc.*, *127*, 14849-14858.

[128] Filichev, V. V., Nielsen, M. C., Bomholt, N., Jessen, C. H., and Pedersen, E. B. (2006), High thermal stability of 5'-5'-linked alternate Hoogsteen triplexes at physiological pH. *Angew. Chem. Int. Ed.*, *45*, 5311-5315.

[129] Paramasivam, M., Cogoi, S., Filichev, V. V., Bomholt, N., Pedersen, E. B., and Xodo, L. E. (2008), Purine twisted-intercalating nucleic acids: a new class of anti-gene molecules resistant to potassium-induced aggregation. *Nucleic Acids Res.*, *36*, 3494-3507.

[130] Doluca, O., Boutorine, A. S., and Filichev, V. V. (2011), Triplex-forming twisted intercalating nucleic acids (TINAs): Design rules, stabilization of antiparallel DNA triplexes and inhibition of G-quartet-dependent self-association. *ChemBioChem*, *12*, 2365-2374.

[131] Doluca, O., Hale, T. K., Edwards, P. J. B., González, C., and Filichev, V. V. (2014), Assembly dependent fluorescence enhancing nucleic acids in sequence-specific detection of double-stranded DNA. *ChemPlusChem*, *79*, 58-66.

[132] Filichev, V. V., Astakhova, I. V., Malakhov, A. D., Korshun, V. A., and Pedersen, E. B. (2008), 1-, 2-, and 4-Ethynylpyrenes in the structure of twisted intercalating nucleic acids: Structure, thermal stability, and fluorescence relationship. *Chem. Eur. J.*, *14*, 9968-9980.

[133] Zhao, J., and Zhai, Q. (2020), Recent advances in the development of ligands specifically targeting telomeric multimeric G-quadruplexes. *Bioorg. Chem.*, *103*.

[134] Saha, S., Cai, J. F., Eiler, D., and Hamilton, A. D. (2010), Programing the formation of DNA and PNA quadruplexes by pi-pi-stacking interactions. *Chem. Commun.*, *46*, 1685-1687.

[135] Franceschin, M., Borbone, N., Oliviero, G., Casagrande, V., Scuotto, M., Coppola, T., Borioni, S., Mayol, L., Ortaggi, G., Bianco, A., Amato, J., and Varra, M. (2011), Synthesis of a dibromoperylene phosphoramidite building block and its incorporation at the 5'- end of a G-quadruplex forming oligonucleotide: Spectroscopic properties and structural studies of the resulting dibromoperylene conjugate. *Bioconjug. Chem.*, *22*, 1309-1319.

[136] Pagano, B., Amato, J., Iaccarino, N., Cingolani, C., Zizza, P., Biroccio, A., Novellino, E., and Randazzo, A. (2015), Looking for efficient G-quadruplex ligands: evidence for selective stabilizing properties and telomere damage by drug-like molecules. *ChemMedChem*, *10*, 640-649.

[137] Zhu, L. N., Wu, B., and Kong, D. M. (2013), Specific recognition and stabilization of monomeric and multimeric G-quadruplexes by cationic porphyrin TMPipEOPP under molecular crowding conditions. *Nucleic Acids Res.*, *41*, 4324-4335.

[138] Dixon, I. M., Lopez, F., Tejera, A. M., Esteve, J. P., Blasco, M. A., Pratviel, G., and Meunier, B. (2007), A G-quadruplex ligand with 10000-fold selectivity over duplex DNA. *J. Am. Chem. Soc.*, *129*, 1502-1503.

[139] Shinohara, K., Sannohe, Y., Kaieda, S., Tanaka, K., Osuga, H., Tahara, H., Xu, Y., Kawase, T., Bando, T., and Sugiyama, H. (2010), A chiral wedge molecule inhibits telomerase activity. *J. Am. Chem. Soc.*, *132*, 3778-3782.

[140] Cousins, A. R. O., Ritson, D., Sharma, P., Stevens, M. F. G., Moses, J. E., and Searle, M. S. (2014), Ligand selectivity in stabilising tandem parallel folded G-quadruplex motifs in human telomeric DNA sequences. *Chem. Commun.*, *50*, 15202-15205.

[141] Kormuth, K. A., Woolford, J. L., and Armitage, B. A. (2016), Homologous PNA hybridization to noncanonical DNA G-quadruplexes. *Biochemistry*, *55*, 1749-1757.

[142] Ishizuka, T., Yang, J., Komiyama, M., and Xu, Y. (2012), G-rich sequence-specific recognition and scission of human genome by PNA/DNA hybrid G-quadruplex formation. *Angew. Chem. Int. Ed.*, *51*, 7198-7202.

[143] Paul, A., Sengupta, P., Krishnan, Y., and Ladame, S. (2008), Combining G-quadruplex targeting motifs on a single peptide nucleic acid scaffold: A hybrid (3+1) PNA-DNA bimolecular quadruplex. *Chem. Eur. J.*, *14*, 8682-8689.

[144] Oyaghire, S. N., Cherubim, C. J., Telmer, C. A., Martinez, J. A., Bruchez, M. P., and Armitage, B. A. (2016), RNA G-quadruplex invasion and translation inhibition by antisense gamma-peptide nucleic acid oligomers. *Biochemistry*, *55*, 1977-1988.

[145] Xu, Y., Suzuki, Y., Ito, K., and Komiyama, M. (2010), Telomeric repeat-containing RNA structure in living cells. *Proc. Natl. Acad. Sci. U. S. A.*, *107*, 14579-14584.

[146] Takahashi, S., Kim, K. T., Podbevsek, P., Plavec, J., Kim, B. H., and Sugimoto, N. (2018), Recovery of the formation and function of oxidized G-quadruplexes by a pyrene-modified guanine tract. *J. Am. Chem. Soc.*, *140*, 5774-5783.

[147] Biffi, G., Tannahill, D., Miller, J., Howat, W. J., and Balasubramanian, S. (2014), Elevated levels of G-quadruplex formation in human stomach and liver cancer tissues. *PLoS One*, *9*, e102711-e102719.

[148] Schmid, F. X. (2001), Biological macromolecules: UV-visible spectrophotometry. *e LS*.

[149] Allawi, H. T., and SantaLucia, J. (1997), Thermodynamics and NMR of internal GT mismatches in DNA. *Biochemistry*, *36*, 10581-10594.

[150] Mergny, J. L., De Cian, A., Ghelab, A., Sacca, B., and Lacroix, L. (2005), Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res.*, *33*, 81-94.

[151] Mergny, J. L., and Lacroix, L. (2003), Analysis of thermal melting curves. *Oligonucleotides*, *13*, 515-537.

[152] Kypr, J., Kejnovska, I., Renciuk, D., and Vorlickova, M. (2009), Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.*, *37*, 1713-1725.

[153] Karsisiotis, A. I., Hessari, N. M., Novellino, E., Spada, G. P., Randazzo, A., and Webba da Silva, M. (2011), Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angew. Chem. Int. Ed.*, *50*, 10645-10648.

[154] Birren, B. W., Simon, M. I., and Lai, E. (1990), The basis of high resolution separation of small DNAs by asymmetric-voltage field inversion electrophoresis and its application to DNA sequencing gels. *Nucleic Acids Res.*, *18*, 1481-1487.

[155] Zídek, L., Štefl, R., and Sklenár, V. (2001), NMR methodology for the study of nucleic acids. *Curr. Opin. Struct. Biol.*, *11*, 275-281.

[156] Webba da Silva, M. (2007), NMR methods for studying quadruplex nucleic acids. *Methods*, *43*, 264-277.

[157] Banoub, J. H., Newton, R. P., Esmans, E., Ewing, D. F., and Mackenzie, G. (2005), Recent developments in mass spectrometry for the characterization of nucleosides, nucleotides, oligonucleotides, and nucleic acids. *Chem. Rev.*, *105*, 1869-1915.

[158] Il'ina, E. N., and Govorun, V. M. (2009), Mass spectrometry of nucleic acids in molecular medicine. *Russ. J. Bioorg. Chem.*, *35*, 135-149.

[159] Bryld, T., Hojland, T., and Wengel, J. (2004), DNA-selective hybridization and dual strand invasion of short double-stranded DNA using pyren-1-ylcarbonyl-functionalized 4'-C-piperazinomethyl-DNA. *Chem. Comm.*, 1064-1065.

[160] Sau, S. P., Madsen, A. S., Podhevsek, P., Andersen, N. K., Kumar, T. S., Andersen, S., Rathje, R. L., Anderson, B. A., Guenther, D. C., Karmakar, S., Kumar, P., Plavec, J., Wengel, J., and Hrdlicka, P. J. (2013), 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.

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

[162] Anderson, B. A., Onley, J. J., and Hrdlicka, P. J. (2015), Recognition of double-stranded DNA using energetically activated duplexes modified with N2'-Pyrene-, Perylene-, or coronene-functionalized 2'-N-Methyl-2'-amino-DNA monomers. *J. Org. Chem.*, *80*, 5395–5406.

[163] Guenther, D. C., Anderson, G. H., Karmakar, S., Anderson, B. A., Didion, B. A., Guo, W., Verstegen, J. P., and Hrdlicka, P. J. (2015), Invader probes: Harnessing the energy of intercalation to facilitate recognition of chromosomal DNA for diagnostic applications. *Chem. Sci.*, *6*, 5006-5015.

[164] Guenther, D. C., Karmakar, S., and Hrdlicka, P. J. (2015), Bulged Invader probes: Activated duplexes for mixed-sequence dsDNA recognition with improved thermodynamic and kinetic profiles. *Chem. Commun.*, *51*, 15051-15054.

[165] Anderson, B. A., and Hrdlicka, P. J. (2016), Merging two strategies for mixed-sequence recognition of double-stranded DNA: Pseudocomplementary invader probes. *J. Org. Chem.*, *81*, 3335-3346.

[166] Mills, M., Arimondo, P. B., Lacroix, L., Garestier, T., Hélène, C., Klump, H., and Mergny, J.-L. (1999), Energetics of strand-displacement reactions in triple helices: A spectroscopic study. *J. Mol. Biol.*, *291*, 1035-1054.

[167] Puglisi, J. D., and Tinoco, I., Jr. (1989), Absorbance melting curves of RNA. *Methods Enzymol.*, *180*, 304-325.

[168] Malakhov, A. D., Skorobogatyi, M. V., Prokhorenko, I. A., Gontarev, S. V., Kozhich, D. T., Stetsenko, D. A., Stepanova, I. A., Shenkarev, Z. O., Berlin, Y. A., and Korshun, V. A. (2004), 1-(Phenylethynyl)pyrene and 9,10-bis(phenylethynyl)anthracene, useful fluorescent dyes for DNA labeling: Excimer formation and energy transfer. *Eur. J. Org. Chem.*, 1298-1307.

[169] Astakhova, I. V., Malakhov, A. D., Stepanova, I. A., Ustinov, A. V., Bondarev, S. L., Paramonov, A. S., and Korshun, V. A. (2007), 1-phenylethynylpyrene (1-PEPy) as refined excimer forming alternative to pyrene: Case of DNA major groove excimer. *Bioconjug. Chem.*, *18*, 1972-1980.

[170] Xia, T. B., SantaLucia, J., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X. Q., Cox, C., and Turner, D. H. (1998), Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, *37*, 14719-14735.

[171] Doluca, O., Withers, J. M., Loo, T. S., Edwards, P. J. B., Gonzalez, C., and Filichev, V. V. (2015), Interdependence of pyrene interactions and tetramolecular G4-DNA assembly. *Org. Biomol. Chem.*, *13*, 3742-3748.

[172] Schneider, U. V., Mikkelsen, N. D., Lindqvist, A., Okkels, L. M., Johnk, N., and Lisby, G. (2012), Improved efficiency and robustness in qPCR and multiplex end-point PCR by twisted intercalating nucleic acid modified primers. *PLoS One*, *7*, e38451.

[173] Miller, P. S., McParland, K. B., Jayaraman, K., and Tso, P. O. P. (1981), Biochemical and biological effects of nonionic nucleic acid methylphosphonates. *Biochemistry*, *20*, 1874-1880.

[174] Shoji, Y., Akhtar, S., Periasamy, A., Herman, B., and Juliano, R. L. (1991), Mechansism of cellular uptake of modified oligodeoxynucleotides containing methylphosphonate linkages. *Nucleic Acids Res.*, *19*, 5543-5550.

[175] Nagahama, K., Veedu, R. N., and Wengel, J. (2009), Nuclease resistant methylphosphonate-DNA/LNA chimeric oligonucleotides. *Bioorg. Med. Chem. Lett.*, *19*, 2707-2709.

[176] Sacca, B., Lacroix, L., and Mergny, J. L. (2005), The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Res.*, *33*, 1182-1192.

[177] Kuo, T. C., Wu, M. W., Lin, W. C., Matulis, D., Yang, Y. S., Li, S. Y., and Chen, W. Y. (2020), Reduction of interstrand charge repulsion of DNA duplexes by salts and by neutral phosphotriesters - Contrary effects for harnessing duplex formation. *J. Taiwan Inst. Chem. Eng.*, *110*, 1-7.

[178] Kupryushkin, M. S., Pyshnyi, D. V., and Stetsenko, D. A. (2014), Phosphoryl guanidines: A new type of nucleic acid analogues. *Acta Naturae*, *6*, 116-118.

[179] Korotkovs, V., Reichenbach, L. F., Pescheteau, C., Burley, G. A., and Liskamp, R. M. J. (2019), Molecular construction of sulfonamide antisense oligonucleotides. *J. Org. Chem.*, *84*, 10635-10648.

[180] Yanachkov, I., Zavizion, B., Metelev, V., Stevens, L. J., Tabatadze, Y., Yanachkova, M., Wright, G., Krichevsky, A. M., and Tabatadze, D. R. (2017), Self-neutralizing oligonucleotides with enhanced cellular uptake. *Org. Biomol. Chem.*, *15*, 1363-1380.

[181] Prokhorova, D. V., Chelobanov, B. P., Burakova, E. A., Fokina, A. A., and Stetsenko, D. A. (2017), New oligodeoxyribonucleotide derivatives bearing internucleotide N-tosyl phosphoramidate groups: Synthesis and complementary binding to DNA and RNA. *Russ. J. Bioorg. Chem.*, *43*, 38-42.

[182] Chelobanov, B. P., Burakova, E. A., Prokhorova, D. V., Fokina, A. A., and Stetsenko, D. A. (2018), New oligodeoxynucleotide derivatives containing N-(methanesulfonyl)-phosphoramidate (mesyl phosphoramidate) internucleotide group. *Russ. J. Bioorg. Chem.*, *43*, 664-668.

[183] Schwertz, G., Witschel, M. C., Rottmann, M., Bonnert, R., Leartsakulpanich, U., Chitnumsub, P., Jaruwat, A., Ittarat, W., Schafer, A., Aponte, R. A., Charman, S. A., White, K. L., Kundu, A., Sadhulchan, S., Lloyd, M., Freiberg, G. M., Srikumaran, M., Siggel, M., Zwyssig, A., Chaiyen, P., and Diederich, F. (2017), Antimalarial inhibitors targeting serine hydroxymethyltransferase (SHMT) with in vivo efficacy and analysis of their binding mode based on X-ray cocrystal structures. *J. Med. Chem.*, *60*, 4840-4860.

[184] Finkelstein, H. (1910), Darstellung organischer Jodide aus den entsprechenden Bromiden und Chloriden. *Ber. Dtsch. Chem. Ges.*, *43*, 1528-1532.

[185] Brasen, W., and Hauser, C. (1954), Ortho-methylbenzyl alcohol. *Org. Synth.*, *34*, 58-60.

[186] Hazen, G. G., Bollinger, F. W., Roberts, F. E., Russ, W. K., Seman, J. J., and Staskiewicz, S. (1996), 4-dodecylbenzenesulfonyl azides - (Benzenesulfonyl azides, 4-dodecyl-). *Org. Synth.*, *73*, 144-151.

[187] Felsenfeld.G, and Miles, H. T. (1967), The physical and chemical properties of nucleic acids. *Annu. Rev. Biochem.*, *36*, 407-448.

[188] ElHassan, M. A., and Calladine, C. R. (1996), Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *J. Mol. Biol.*, *259*, 95-103.

[189] Kotin, L. (1963), On the effect of ionic strength on the melting temperature of DNA. *J. Mol. Biol.*, *7*, 309-311.

[190] Rayburn, E. R., and Zhang, R. W. (2008), Antisense, RNAi, and gene silencing strategies for therapy: Mission possible or impossible? *Drug Discov. Today*, *13*, 513-521.

[191] Robles, J., Grandas, A., and Pedroso, E. (2001), Synthesis of modified oligonucleotides containing 4-guanidino-2-pyrimidinone nucleobases. *Tetrahedron*, *57*, 179-194.

[192] Roig, V., and Asseline, U. (2003), Oligo-2 '-deoxyribonucleotides containing uracil modified at the 5-position with linkers ending with guanidinium groups. *J. Am. Chem. Soc.*, *125*, 4416-4417.

[193] Kanazaki, M., Ueno, Y., Shuto, S., and Matsuda, A. (2000), Highly nuclease-resistant phosphodiester-type oligodeoxynucleotides containing 4 'alpha-C-aminoalkylthymidines form thermally stable duplexes with DNA and RNA. A candidate for potent antisense molecules. *J. Am. Chem. Soc.*, *122*, 2422-2432.

[194] Puri, N., Majumdar, A., Cuenoud, B., Natt, F., Martin, P., Boyd, A., Miller, P. S., and Seidman, M. M. (2002), Minimum number of 2 '-O-(2-aminoethyl) residues required for gene knockout activity by triple helix forming oligonucleotides. *Biochemistry*, *41*, 7716-7724.

[195] Bailey, P., Sartorelli, V., Hamamori, Y., and Muscat, G. E. O. (1998), The orphan nuclear receptor, COUP-TF II, inhibits myogenesis by post-transcriptional regulation of MyoD function: COUP-TF II directly interacts with p300 and MyoD. *Nucleic Acids Res.*, *26*, 5501-5510.

[196] Michel, T., Debart, F., and Vasseur, J. J. (2003), Efficient guanidination of the phosphate linkage towards cationic phosphoramidate oligonucleotides. *Tetrahedron Lett.*, *44*, 6579-6582.

[197] Vasquez, K. M., Dagle, J. M., Weeks, D. L., and Glazer, P. M. (2001), Chromosome targeting at short polypurine sites by cationic triplex-forming oligonucleotides. *J. Biol. Chem.*, *276*, 38536-38541.

[198] Fox, K. R. (2000), Targeting DNA with triplexes. *Curr. Med. Chem.*, *7*, 17-37.

[199] Hashimoto, H., Nelson, M. G., and Switzer, C. (1993), Zwitterionic DNA. *J. Am. Chem. Soc.*, *206*, 7128-7134.

[200] Bentley, J., Brazier, J. A., Fisher, J., and Cosstick, R. (2007), Duplex stability of DNA center dot DNA and DNA center dot RNA duplexes containing 3 '-S-phosphorothiolate linkages. *Org. Biomol. Chem.*, *5*, 3698-3702.

[201] Tomac, S., Sarkar, M., Ratilainen, T., Wittung, P., Nielsen, P. E., Norden, B., and Graslund, A. (1996), Ionic effects on the stability and conformation of peptide nucleic acid complexes. *J. Am. Chem. Soc.*, *118*, 5544-5552.

[202] Collie, G. W., and Parkinson, G. N. (2011), The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chem.Soc. Rev.*, *40*, 5867-5892.

[203] Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., and Neidle, S. (2006), Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Res.*, *34*, 5402-5415.

[204] Williamson, J. R., Raghuraman, M. K., and Cech, T. R. (1989), Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell*, *59*, 871.

[205] Lane, A. N., Chaires, J. B., Gray, R. D., and Trent, J. O. (2008), Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.*, *36*, 5482-5515.

[206] Nielsen, J. T., Arar, K., and Petersen, M. (2009), Solution structure of a locked nucleic acid modified quadruplex: introducing the V4 folding topology. *Angew. Chem. Int. Ed.*, *48*, 3099-3103.

[207] Haase, L., Karg, B., and Weisz, K. (2019), Manipulating DNA G-quadruplex structures by using guanosine analogues. *ChemBioChem*, *20*, 985-993.

[208] Filitcheva, J., Edwards, P. J. B., Norris, G. E., and Filichev, V. V. (2019), alpha-2 '-Deoxyguanosine can switch DNA G-quadruplex topologies from antiparallel to parallel. *Org. Biomol. Chem.*, *17*, 4031-4042.

[209] Sacca, B., Lacroix, L., and Mergny, J. L. (2005), The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Res.*, *33*, 1182-1192.

[210] Ecker, D. J., Wyatt, J. R., Vickers, T., Buckheit, R., Roberson, J., and Imbach, J. L. (1995), Novel guanosine quartet structure binds to the Hiv envelope and inhibits envelope mediated cell-fusion. *Nucleosides Nucleotides Nucleic Acids*, *14*, 1117-1127.

[211] Pozmogova, G. E., Zaitseva, M. A., Smirnov, I. P., Shvachko, A. G., Murina, M. A., and Sergeenko, V. I. (2010), Anticoagulant effects of thioanalogs of thrombin-binding DNA-aptamer and their stability in the plasma. *Bull. Exp. Biol. Med.*, *150*, 180-184.

[212] Datta, B., and Armitage, B. A. (2001), Hybridization of PNA to structured DNA targets: G-quadruplex invasion and the overhang effect. *J. Am. Chem. Soc.*, *123*, 9612-9619.

[213] Datta, B., Schmitt, C., and Armitage, B. A. (2003), Formation of a PNA(2)-DNA(2) hybrid quadruplex. *J. Am. Chem. Soc.*, *125*, 4111-4118.

[214] Petraccone, L., Pagano, B., Esposito, V., Randazzo, A., Piccialli, G., Barone, G., Mattia, C. A., and Giancola, C. (2005), Thermodynamics and kinetics of PNA−DNA quadruplex-forming chimeras. *J. Am. Chem. Soc.*, *127*, 16215-16223.

[215] Marin, V. L., and Armitage, B. A. (2006), Hybridization of complementary and homologous peptide nucleic acid oligomers to a guanine quadruplex-forming RNA. *Biochemistry*, *45*, 1745-1754.

[216] Pinto, B., Rusciano, G., D'Errico, S., Borbone, N., Sasso, A., Piccialli, V., Mayol, L., Oliviero, G., and Piccialli, G. (2017), Synthesis and label free characterization of a bimolecular PNA homo quadruplex. *Biochim. Biophys. Acta (BBA) - General Subj.*, 1222-1228.

[217] Krishnan-Ghosh, Y., Stephens, E., and Balasubramanian, S. (2004), A PNA4 quadruplex. *J. Am. Chem. Soc.*, *126*, 5944-5945.

[218] Datta, B., Bier, M. E., Roy, S., and Armitage, B. A. (2005), Quadruplex formation by a guanine-rich PNA oligomer. *J. Am. Chem. Soc.*, *127*, 4199-4207.

[219] Amato, J., Pagano, B., Borbone, N., Oliviero, G., Gabelica, V., De Pauw, E., D'Errico, S., Piccialli, V., Varra, M., Giancola, C., Piccialli, G., and Mayol, L. (2011), Targeting G-quadruplex structure in the human c-Kit promoter with short PNA sequences. *Bioconjug. Chem.*, *22*, 654-663.

[220] Englund, E. A., Gupta, P., Micklitsch, C. M., Onyshchenko, M. I., Remeeva, E., Neumann, R. D., Panyutin, I. G., and Appella, D. H. (2014), PPG peptide nucleic acids that promote DNA guanine quadruplexes. *ChemBioChem*, *15*, 1887-1890.

[221] Marin, V. L., and Armitage, B. A. (2005), RNA guanine quadruplex invasion by complementary and homologous PNA probes. *J. Am. Chem. Soc.*, *127*, 8032-8033.

[222] Pontinha, A. D. R., Chiorcea-Paquim, A. M., Eritja, R., and Oliveira-Brett, A. M. (2014), Quadruplex nanostructures of d(TGGGGT): Influence of sodium and potassium ions. *Anal. Chem.*, *86*, 5851-5857.

[223] Stetsenko, D. A., Kupryushkin, M. S., and Pyshnyi, D. V. (2016), Preparation of modified DNA and their use for the treatment of a disease and as diagnostic agents, p 106pp., Noogen LLC, Russia.

[224] Largy, E., and Mergny, J. L. (2014), Shape matters: size-exclusion HPLC for the study of nucleic acid structural polymorphism. *Nucleic Acids Res.*, *42*, e149.

[225] Karsisiotis, A. I., Hessari, N. M. a., Novellino, E., Spada, G. P., Randazzo, A., and da Silva, M. W. (2011), Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angew. Chem. Int. Ed.*, *50*, 10645-10648.

[226] Masiero, S., Trotta, R., Pieraccini, S., De Tito, S., Perone, R., Randazzo, A., and Spada, G. P. (2010), A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. *Org. Biomol. Chem.*, *8*, 2683-2692.

[227] Gray, D. M., Wen, J. D., Gray, C. W., Repges, R., Repges, C., Raabe, G., and Fleischhauer, J. (2008), Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities. *Chirality*, *20*, 431-440.

[228] Šket, P., Virgilio, A., Esposito, V., Galeone, A., and Plavec, J. (2012), Strand directionality affects cation binding and movement within tetramolecular G-quadruplexes. *Nucleic Acids Res.*, *40*, 11047-11057.

[229] Largy, E., and Mergny, J.-L. (2014), Shape matters: size-exclusion HPLC for the study of nucleic acid structural polymorphism. *Nucleic Acids Res.*, *42*, e149-e149.

[230] Wyatt, J. R., Davis, P. W., and Freier, S. M. (1996), Kinetics of G-quartet-mediated tetramer formation. *Biochemistry*, *35*, 8002-8008.

[231] Ida, J., Chan, S. K., Glokler, J., Lim, Y. Y., Choong, Y. S., and Lim, T. S. (2019), G-quadruplexes as an alternative recognition element in disease-related target sensing. *Molecules*, *24*, 30.

[232] Virgilio, A., Esposito, V., Randazzo, A., Mayol, L., and Galeone, A. (2005), 8-Methyl-2'-deoxyguanosine incorporation into parallel DNA quadruplex structures. *Nucleic Acids Res.*, *33*, 6188-6195.

[233] Tran, P. L. T., Virgilio, A., Esposito, V., Citarella, G., Mergny, J.-L., and Galeone, A. (2011), Effects of 8-methylguanine on structure, stability and kinetics of formation of tetramolecular quadruplexes. *Biochimie*, *93*, 399-408.

[234] Virgilio, A., Esposito, V., Citarella, G., Pepe, A., Mayol, L., and Galeone, A. (2012), The insertion of two 8-methyl-2'-deoxyguanosine residues in tetramolecular quadruplex structures: trying to orientate the strands. *Nucleic Acids Res.*, *40*, 461-475.

[235] Petraccone, L., Erra, E., Randazzo, A., and Giancola, C. (2006), Energetic aspects of Locked nucleic acids quadruplex association and dissociation. *Biopolymers*, *83*, 584-594.

[236] Pedersen, E. B., Nielsen, J. T., Nielsen, C., and Filichev, V. V. (2011), Enhanced anti-HIV-1 activity of G-quadruplexes comprising locked nucleic acids and intercalating nucleic acids. *Nucleic Acids Res.*, *39*, 2470-2481.

[237] Wyatt, J. R., Vickers, T. A., Roberson, J. L., Buckheit, R. W., Jr., Klimkait, T., DeBaets, E., Davis, P. W., Rayner, B., Imbach, J. L., and Ecker, D. J. (1994), Combinatorially selected guanosine-quartet structure is a potent inhibitor of human immunodeficiency virus envelope-mediated cell fusion. *Proc. Natl. Acad. Sci. U. S. A.*, *91*, 1356-1360.

[238] Nagesh, N., and Chatterji, D. (1995), Ammonium ion at low concentration stabilizes the G-quadruplex formation by telomeric sequence. *J. Biochem. Biophys.l Meth.*, *30*, 1-8.

[239] Creze, C., Rinaldi, B., Haser, R., Bouvet, P., and Gouet, P. (2007), Structure of a d(TGGGGT) quadruplex crystallized in the presence of Li+ ions. *Acta Crystallogr. D, Biol. Crystallogr.*, *63*, 682-688.

[240] Wan, W. B., and Seth, P. P. (2016), The medicinal chemistry of therapeutic oligonucleotides. *J. Med. Chem.*, *59*, 9645-9667.

[241] Emehiser, R. G., Hall, E., Guenther, D. C., Karmakar, S., and Hrdlicka, P. J. (2020), Head-to-head comparison of LNA, (MP)gamma PNA, INA and invader probes targeting mixed-sequence double-stranded DNA. *Org. Biomol. Chem.*, *18*, 56-65.

[242] Miroshnichenko, S. K., Patutina, O. A., Burakova, E. A., Chelobanov, B. P., Fokina, A. A., Vlassov, V. V., Altman, S., Zenkova, M. A., and Stetsenko, D. A. (2019), Mesyl phosphoramidate antisense oligonucleotides as an alternative to phosphorothioates with improved biochemical and biological properties. *Proc. Natl. Acad. Sci. U.S.A.*, *116*, 1229-1234.

[243] Kaltenboeck, B., and Wang, C. M. (2005), Advances in real-time PCR: Application to clinical laboratory diagnostics, In *Adv. Clinical Chem.* (Makowski, G. S., Ed.), pp 219-259.

[244] Yager, P., Domingo, G. J., and Gerdes, J. (2008), Point-of-care diagnostics for global health. *Annu. Rev. Biomed. Eng.*, *10*, 107-144.

[245] Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., and Hase, T. (2000), Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res.*, *28*, 7.

[246] Brown, P. O., and Botstein, D. (1999), Exploring the new world of the genome with DNA microarrays. *Nature Genet.*, *21*, 33-37.

[247] Geinguenaud, F., Souissi, I., Fagard, R., Motte, L., and Lalatonne, Y. (2012), Electrostatic assembly of a DNA superparamagnetic nano-tool for simultaneous intracellular delivery and in situ monitoring. *Nanomedicine*, *8*, 1106-1115.

[248] Neuberger, T., Schöpf, B., Hofmann, H., Hofmann, M., and von Rechenberg, B. (2005), Superparamagnetic nanoparticles for biomedical applications: Possibilities and limitations of a new drug delivery system. *J. Magnetism Magnetic Mat.*, *293*, 483-496.

[249] Tran, T. V., Nguyen, B. V., Nguyen, T. T. P., Tran, T. T., Pham, K. G., Le, Q. B., Do, B. N., Pham, H. N., Nguyen, C. V., Van, D. P. H., Ha, V., Doan, T. H. T., and Le, H. Q. (2019), Development of a highly sensitive magneto-enzyme lateral flow immunoassay for dengue NS1 detection. *PeerJ*, *7*, 18.

[250] Filichev, V. V., Gaber, H., Olsen, T. R., Jorgensen, P. T., Jessen, C. H., and Pedersen, E. B. (2006), Twisted intercalating nucleic acids - Intercalator influence on parallel triplex stabilities. *Eur. J. Org. Chem.*, *2006*, 3960-3968.

DRC 16

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	YONGDONG SU
Name/title of Primary Supervisor:	ASSOCIATE PROFESSOR VYACHESLAV V. FILICHEV
In which chapter is the manuscript /published work:	CHAPTER 4

Please select one of the following three options:

(●) The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

Su, Y., Bayarjargal, M., Hale, T. K., & Filichev, V. V. DNA with zwitterionic and negatively charged phosphate modifications: Formation of DNA triplexes, duplexes and cell uptake studies, (2021). Beilstein J. Org. Chem., 17, 749–761.

( ) The manuscript is currently under review for publication – please indicate:

- The name of the journal:

Beilstein Journal of Organic Chemistry

- The percentage of the manuscript/published work that was contributed by the candidate: 70.00

- Describe the contribution that the candidate has made to the manuscript/published work:

The candidate synthesised N+ and Ts- modified ONs, evaluated properties of parallel triplexes, antiparallel duplexes with DNA and RNA and performed enzymatic digestion assay.
The candidate wrote the first and all subsequent drafts of the manuscript according to suggestions of co-authors, prepared the supporting information, was drafting responses to referees' comments.

( ) It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	YongdongSu 数字签名者：YongdongSu 日期：2021.04.20 10:04:35 +12'00'
Date:	20-Apr-2021
Primary Supervisor's Signature:	Vyacheslav V. Filichev Digitally signed by Vyacheslav V. Filichev Date: 2021.04.20 12:57:01 +12'00'
Date:	20-Apr-2021

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

GRS Version 5 – 13 December 2019
DRC 19/09/10

DRC 16

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	YONGDONG SU
Name/title of Primary Supervisor:	ASSOCIATE PROFESSOR VYACHESLAV V. FILICHEV
In which chapter is the manuscript /published work:	CHAPTER 5

Please select one of the following three options:

(●) The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

Su, Y., Fujii, H., Burakova, E. A., Chelobanov, B. P., Fujii, M., Stetsenko, D. A., & Filichev, V. V. (2019). Neutral and Negatively Charged Phosphate Modifications Altering Thermal Stability, Kinetics of Formation and Monovalent Ion Dependence of DNA G-Quadruplexes. Chemistry - An Asian Journal. 14(8). 1212-1220.

( ) The manuscript is currently under review for publication – please indicate:

- The name of the journal:

Chemistry - An Asian Journal

- The percentage of the manuscript/published work that was contributed by the candidate: 50.00

- Describe the contribution that the candidate has made to the manuscript/published work:

The candidate conducted evaluation of G-quadruplex formation of chemically modified G-rich ONs, which were synthesised in the laboratory of Dr. Stetsenko (Novosibirsk State University, Novosibirsk, Russia). The candidate prepared the first draft of the manuscript, supporting information and was involved in drafting responses to referees' comments.

( ) It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	YongdongSu 数字签名者：YongdongSu 日期：2021.04.20 10:14:41 +12'00'
Date:	20-Apr-2021
Primary Supervisor's Signature:	Vyacheslav V. Filichev Digitally signed by Vyacheslav V. Filichev Date: 2021.04.20 12:59:48 +12'00'
Date:	20-Apr-2021

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

GRS Version 5 – 13 December 2019
DRC 19/09/10

DRC 16

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	YONGDONG SU
Name/title of Primary Supervisor:	ASSOCIATE PROFESSOR VYACHESLAV V. FILICHEV
In which chapter is the manuscript /published work:	CHAPTER 5

Please select one of the following three options:

(●) The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

Su, Y., Edwards, P. J. B., Stetsenko, D. A., & Filichev, V. V. (2020). The Importance of Phosphates for DNA G-Quadruplex Formation: Evaluation of Zwitterionic G-Rich Oligodeoxynucleotides. ChemBioChem, 21(17), 2455-2466.

( ) The manuscript is currently under review for publication – please indicate:

- The name of the journal:

ChemBioChem

- The percentage of the manuscript/published work that was contributed by the candidate: 80.00

- Describe the contribution that the candidate has made to the manuscript/published work:

The candidate synthesised and purified zwitterionic N+TG4T sequence and evaluated the property of G4s formed by N+TG4T. The candidate performed all analysis of the derived data and prepared the first and all subsequent drafts of the manuscript and supporting information according to suggestions of co-authors.
The candidate prepared a response to referees' comments and revised the manuscript accordingly.

( ) It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	YongdongSu 数字签名者：YongdongSu 日期：2021.04.20 10:19:13 +12'00'
Date:	20-Apr-2021
Primary Supervisor's Signature:	Vyacheslav V. Filichev Digitally signed by Vyacheslav V. Filichev Date: 2021.04.20 13:02:09 +12'00'
Date:	20-Apr-2021

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

GRS Version 5 – 13 December 2019
DRC 19/09/10