

Design and Synthesis of Fluorogenic Oligonucleotide- Based Systems to Target Double- Stranded DNA



Sarah Walsh

Wolfson College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Hilary Term 2020

Declarations

I, Sarah Walsh, declare that the thesis titled “Design and Synthesis of Fluorogenic Oligonucleotide-Based Systems to Target Double-Stranded DNA” and the relating research presented in this thesis was performed and written by the author under the supervision of Prof. Tom Brown at the University of Oxford. The research is original and has not been submitted in full or in part for any other degree at this or at any other university. The copyright of this thesis rests with the author.

Acknowledgements

I am very thankful to Prof. Tom Brown, my supervisor, for all the guidance and support on my exciting scientific journey. These last four years would not have been possible without his tolerance of endless questions, countless meetings and out-of-hours emails. I would also like to thank Prof. Afaf El-Sagheer for her practical and level-headed outlook that helped keep me well-balanced. It is a testament to both of them for maintaining a positive environment to work in. I would also like to thank Dr. Tom Brown Jnr. and ATDBio for their guidance and support throughout my studies.

I would like to extend a big thank you to the post-docs who provided extensive advice and help with my thesis and, essentially, moulded me into the scientist I am today. Dr. Lapatrada Taemaitree, for her expertise, her humour and her harmonious ethos both inside the lab and out. Dr. Ysobel Baker, for her knowledge, her inspiring work ethic and her eagle-eyes. I would also like to thank Piotr for really sharing the whole experience with me. Dr. Arun Shivalingham and Dr. Sara Deornellas, who had a major role to play in making my research and social life truly great. Cam, Sven, Phoebe, Thanishta, Big Jack, Small Jack and the rest of the Brownies, both past and present, for truly making the research group into a family.

Lastly, I would like to thank my family and friends, for their love and support even if they didn't fully understand my work, and Joseph, for his encouragement and belief in me throughout the 4 years. It is thanks to all of these people that I am who I am.

“There are two ways to do something, the right way and again.”

Author Unknown.

Abbreviations

A/C/G/T/U	Adenosine/Cytosine/Guanine/Thymine/Uridine
Ac	Acetate
app.	Apparent
APS	Ammonium persulfate
BFP	Blue Fluorescent Protein
BHQ	Black Hole Quencher®
bp	Base pair(s)
br	Broad (IR/NMR)
BSA	Bovine Serum Albumin
bz	benzoyl
Cas9	CRISPR associated protein 9
CEP	2-Cyanoethyl <i>N,N</i> -diisopropyl phosphoramidite
CPG	Controlled Pore Glass
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
CuAAC	Copper(I)-catalysed Azide Alkyne Cycloaddition
Cy3	Cyanine-3
d	Doublet (NMR)
dCas9	Catalytically inactivated/dead/deactivated Cas9
DCM	Dichloromethane

dd	Doublet of doublets (NMR)
DIPEA	<i>N,N</i> -Diisopropylethylamine
DMAP	4-Dimethylaminopyridine
DMEM	Dulbecco's Modified Eagle Medium
DMF	<i>N,N</i> -Dimethylformamide
DMSO	Dimethyl sulfoxide
DMT	4,4'-Dimethoxytrityl
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
DPBS	Dulbecco's Phosphate-Buffered Saline
ds	Double-stranded
DSB	Double-Stranded Break
EDA	Ethylenediamine
EDTA	Ethylenediamine tetra-acetic acid
EdU	5-Ethylene-2'-deoxyuridine
EMSA	Electrophoretic Mobility Shift Assay
Eq	Equivalents
ESI	Electrospray ionisation
Et	Ethyl
F-gRNA	Fluorescent guide RNA
F-Q-gRNA	Fluorescent-quencher guide RNA
FBS	Fetal Bovine Serum

FISH	Fluorescence <i>In Situ</i> Hybridisation
FP	Fluorescent Protein
FRET	Förster Resonance Energy Transfer
GFP	Green Fluorescent Protein
gRNA	Guide RNA
HAA	Hexylamine acetate
HDR	Homology-Directed Repair
HEG	Hexaethylene glycol
HEPES	Hydroxyethyl piperazineethanesulfonic acid
HFIP	Hexafluoroisopropanol
HIV	Human Immunodeficiency Virus
HPLC	High Performance Liquid Chromatography
HRMS	High Resolution Mass Spectrometry
ib	isobutyl
IVT	<i>In Vitro</i> Transcription
LR	Low Resolution
M	Molar concentration
m	Multiplet (NMR)
Me	methyl
mg	Milligram
mL	Milliliter
MS	Mass spectrometry

MSNT	1-(2-Mesitylenesulfonyl)-3-nitro-1 <i>H</i> -1,2,4-triazole
NHEJ	Non-Homologous End Joining
NHS	<i>N</i> -Hydroxysuccinimide
NLS	Nuclear Localisation Sequence
nm	Nanometre
NMI	1-Methylimidazole
NMM	<i>N</i> -Methylmorpholine
NMR	Nuclear Magnetic Resonance
nt	Nucleotide(s)
ODN	Oligonucleotide
PAGE	Polyacrylamide Gel Electrophoresis
PAM	Protospacer-Adjacent Motif
pC	5-(1-Propynyl)-deoxycytidine
PCR	Polymerase Chain Reaction
pH	Potential of hydrogen
ppm	Parts per million
Q	Quencher
q	Quartet (NMR)
QD	Quantum dots
RFP	Red Fluorescent Protein
RNA	Ribonucleic Acid
RNase	Ribonuclease

RNP	Ribonucleoprotein
RP-HPLC	Reverse Phase-HPLC
RT	Room Temperature
s	Singlet (NMR)
sgRNA	Single-guide RNA
SPS	Solvent Purification System
ss	Single-stranded
t	Triplet (NMR)
tac	<i>tert</i> -butylphenoxyacetyl
TALE	Transcription activator-like effector(s)
TAMRA	Carboxytetramethylrhodamine
TBDMS	<i>Tert</i> -butyldimethylsilyl
TBE	Tris-borate-EDTA
TC	Thiomorpholine-4-carbothioate
TCA	Trichloroacetic acid
TEAB	Triethylammonium bicarbonate
TEMED	<i>N,N,N',N'</i> -Tetramethylethylene-1,2-diamine
TFO	Triplex-Forming Oligonucleotide
THF	Tetrahydrofuran
TLC	Thin Layer Chromatography
T_m	Thermal melting temperature
TO	Thiazole orange

tracrRNA	Trans-activating CRISPR-RNA
UPLC	Ultra-Performance Liquid Chromatography
UV	Ultra-Violet
vs	Versus
WC	Watson Crick
$\times g$	Relative centrifugal force times gravity
ZF	Zinc Fingers

Abstract

Fluorogenic oligonucleotide-based systems are important tools for sequence-specific DNA targeting. Fluorescence systems use the emission of a photon from an excited molecule (fluorophore) to produce a detectable signal. It is a highly sensitive process and its versatility has expanded its use to DNA targeting probes. The aim of the work herein is to investigate the design and synthesis of fluorogenic oligonucleotide-based systems specifically for targeting double-stranded DNA.

Triplex-forming oligonucleotides can bind in the major groove of dsDNA through Hoogsteen hydrogen bonding to form triplexes (triple helices). While other methods of dsDNA detection rely on denaturation of the double-stranded target, triplexes do not, and, hence, they provide a versatile approach to DNA recognition. However, the instability of triplexes, especially at neutral pH, limits their applications. The incorporation of a fluorogenic intercalator, thiazole orange, covalently attached to the TFO provides a triplex-stabilising moiety. The stabilising effect of thiazole orange is additive; multiple incorporations can produce triplexes with increased stability of +50 °C even at neutral pH. The extreme stability provided from the modified fluorogenic TFOs allows increased duplex binding affinity and can compensate for the limited target range of TFOs.

CRISPR-Cas is a programmable biological research tool that is invaluable in gene-editing and genomic imaging technologies. The introduction of fluorogenic moieties

both on the dCas9 enzyme and in guide RNA is crucial for imaging but often leads to high background fluorescence in biological systems. By modifying the gRNA to contain an 'on/off' fluorogenic system, genomic imaging in live cells of repetitive elements is improved and is cross-validated by GFP-tagged dCas9.

The efficient synthesis of long, modified RNA is fundamental, particularly in CRISPR research. The introduction of an azide moiety into RNA to provide a means of 'clicking' modified RNA strands together is an important potential approach. However, it is limited by the protecting groups used for RNA and, in turn, this hinders the length of efficient RNA synthesis. The synthesis of a clickable moiety compatible with 2'-O-TC RNA chemistry is investigated as it could provide a simple and cost-effective route to modified guide RNA and could be instrumental in applications such as modified guide RNA library generation.

Overall, the aims of this study were to synthesise and develop fluorogenic oligonucleotide-based systems and investigate their potential applications.

Table of Contents

Declarations	i
Acknowledgements	ii
Abbreviations	iv
Abstract	x
Table of Contents	xii
Chapter 1 Introduction.....	1
1.1 Basics of Nucleic Acids.....	2
1.1.1 Nucleic acid structure.....	2
1.2 Chemical Synthesis of Nucleic Acids	9
1.2.1 DNA chemical synthesis	11
1.2.2 RNA chemical synthesis	15
1.3 Fluorescence	18
1.3.1 Basics of fluorescence.....	18
1.3.2 Fluorescence quenching.....	19
1.3.3 Fluorescence hybridisation probes	21
1.4 Labelling Methods	22
1.5 Overall Hypotheses and Aims	23
Chapter 2 Fluorogenic Triplex-Forming Oligonucleotides	24
2.1 Background	25
2.1.1 Triplex-Forming Oligonucleotides (TFOs)	25
2.1.2 Tackling pH dependence of TFOs	29
2.1.3 Tackling the limited target range for TFOs	30
2.1.4 Other stabilisation techniques	33
2.1.5 Fluorescent TFOs.....	36
2.1.6 Thiazole orange (TO).....	37
2.1.7 TO and TFOs.....	38
2.2 Hypotheses and Aims.....	41

2.3	Triplex Design and Synthesis	42
2.4	Biophysical Studies of Triplex Formation	45
2.4.1	Triplex melting temperatures at pH 5.8	45
2.4.2	Triplex fluorescence melting.....	50
2.4.3	Triplex Surface Plasmon Resonance (SPR)	51
2.4.4	Triplex melting temperatures at pH 7	52
2.4.5	Triplex steady state fluorescence and quantum yield	57
2.4.6	Triplex PAGE at pH 7.....	59
2.5	Conclusion and Future Work.....	62
Chapter 3	Live Cell Imaging using CRISPR-dCas9 and Fluorogenic gRNA ..	64
3.1	Background	65
3.1.1	Fluorescence imaging of the genome	65
3.1.2	Live cell imaging	66
3.1.3	CRISPR-Cas9	68
3.2	Hypotheses and Aims.....	77
3.3	Hairpin F-Q-gRNA – Design I	77
3.3.1	Design strategy of the construct.....	77
3.3.2	Synthesis of the construct	79
3.3.3	<i>In vitro</i> analysis.....	81
3.4	HyBeacon FF-gRNA – Design II	85
3.4.1	Modifying the DNA-targeting region of crRNA	86
3.4.2	Extending the 5' end of crRNA	88
3.4.3	Design strategy of construct.....	88
3.4.4	Synthesis of the construct	89
3.4.5	<i>In vitro</i> analysis.....	91
3.5	RNase H QxF-gRNA – Design III	94
3.5.1	Design strategy of the construct.....	95
3.5.2	Synthesis of the construct	97
3.5.3	<i>In vitro</i> analysis.....	99
3.6	Live Cell Imaging	104
3.6.1	Cell delivery.....	104
3.6.2	Controls.....	105

3.7 RNase H QxF-gRNA – Design IV	108
3.7.1 Design and synthesis of the construct.....	108
3.7.2 <i>In vitro</i> analysis.....	110
3.7.3 Live cell imaging	111
3.8 Conclusion and Future Work.....	116
Chapter 4 Incorporation of Azides for RNA Synthesis.....	118
4.1 Background	119
4.1.1 RNA synthesis	119
4.1.2 Copper-catalysed Azide Alkyne Cyclo-addition (CuAAC)	123
4.1.3 CuAAC for CRISPR-Cas9 systems.....	124
4.1.4 Azide incorporation.....	125
4.2 Hypotheses and Aims.....	128
4.3 Results.....	129
4.3.1 Phosphotriester nucleosides	129
4.3.2 Reverse phosphorylation.....	133
4.3.3 Azide phosphoramidite	141
4.4 Conclusion and Future Work.....	149
Conclusion.....	151
Chapter 5 Experimental	154
5.1 Synthesis of phosphoramidites.....	155
5.1.1 General	155
5.1.2 Synthesis of 5-(1-propargylamino)-2'-deoxycytidine-3'-2-cyanoethyl <i>N,N</i> -diisopropyl phosphoramidite (O-CEP) (5).....	157
5.1.3 Synthesis of 5-(1-propargylamino)-2'-deoxyuridine-3'-O-CEP (7)	165
5.1.4 Synthesis of thiazole orange B6 (TO _{B6}) (10).....	167
5.1.5 Synthesis of 5'-azido-2'-deoxythymidine-3'-phosphotriester (12)	169
5.1.6 Synthesis of 5'-Azido-2'-deoxythymidine-3'-O-CEP (13).....	171
5.1.7 Synthesis of 2'-O-TC-5'-azido-uridine-3'-O-CEP (20)	173
5.2 Oligonucleotide Synthesis and Purification.....	180
5.2.1 General	180
5.2.2 Oligonucleotide synthesis	180

5.2.3	Selective β -cyanoethyl removal	181
5.2.4	Deprotection and purification	181
5.3	Oligonucleotide Labelling	183
5.3.1	Labelling oligonucleotides with NHS Ester	183
5.3.2	Labelling oligonucleotides with click chemistry (CuAAC).....	183
5.4	On-Resin Coupling of Monomers to Oligonucleotides.....	184
5.4.1	Azide-phosphotriester coupling.....	184
5.4.2	Reverse phosphitylation and coupling	185
5.5	Biophysical Studies	186
5.5.1	Thermal melting analysis	186
5.5.2	Fluorescence melting analysis using a CFX96™ Real Time PCR Detection System.....	187
5.5.3	Fluorescence steady state analysis	188
5.5.4	Quantum yield measurements using integrated sphere	189
5.5.5	Surface Plasmon Resonance (SPR)	190
5.5.6	Polyacrylamide gel electrophoresis (PAGE)	191
5.6	Cell Studies	193
5.6.1	Cell culture	193
5.6.2	<i>In Vitro</i> Transcription (IVT).....	194
5.6.3	Transfection	194
5.6.4	Confocal microscopy	195
Appendix		196
References		215

Chapter 1 Introduction

1.1 Basics of Nucleic Acids

1.1.1 Nucleic acid structure

1.1.1.1 Nucleic acid primary structure

Nucleic acids contain hierarchical levels of complexity from primary structures to tertiary structures. The primary structure of nucleic acids is made up of monomeric subunits called nucleotides (nt). There are two forms of nucleic acid polymers in nature, DNA (deoxyribonucleic acid) and RNA (ribonucleic acid) and their nucleotide subunits have a basic structure made up of three components: nitrogen heterocycles (nucleobases), a pentose sugar and a phosphate linker between the sugars. For DNA, the nucleobases include adenine (A) and guanine (G), which are classed as purine nucleobases, and cytosine (C) and thymine (T). In RNA, T is replaced by uracil (U), which does not contain the methyl group at the 5-position of the nucleobase and C, T and U are classed as pyrimidine bases (Figure 1.1 A). Other bases occur in both DNA and RNA, for example in DNA, derivatives of cytosine occur and these are involved in epigenetic control of gene expression and related processes.¹⁻³

Nucleobases combine with either a deoxyribose sugar (2'-H) to produce a DNA nucleoside or a ribose sugar (2'-OH) to produce an RNA nucleoside. The nucleobase is attached to the C1'-position of the sugar moiety through a *N*-glycosidic bond. The purine bases covalently bind to the sugar *via* the N9 position and pyrimidine bases bind *via* the N1 position. The nucleobase can be situated on the same side of the sugar as the 5'-hydroxyl group, in the β -configuration, or the opposite side, in the α -

configuration (Figure 1.1 B). In naturally occurring DNA, all the nucleobases are in β -configuration. The nucleobase can also rotate around the glycosidic bond to produce both *syn*- and *anti*-conformations (Figure 1.1 C). However, *anti*-conformations are generally more common.

A Nucleic Acid Structures

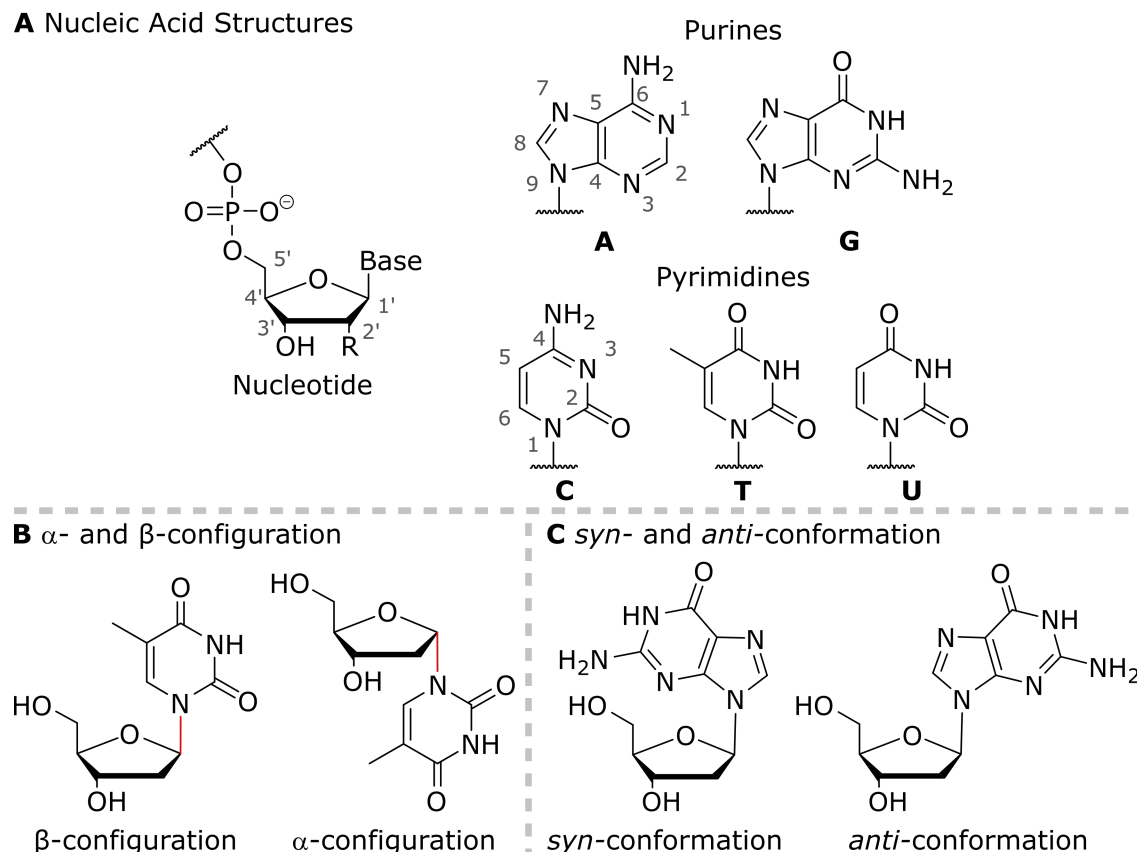


Figure 1.1: Basic of nucleic acid structures. A) Nucleotide and the corresponding nucleobases. A = adenine, G = guanine, C = cytosine, T = thymine, U = uracil, R = H or OH. B) α - and β - configurations of the nucleobase. C) *syn*- and *anti*-conformations of the nucleobase.

A phosphodiester group links the nucleosides together to form the chain of nucleotides in the oligonucleotide (ODN). The phosphate moiety links the 5'-position of one nucleotide to the 3'-position of the next nucleotide (Figure 1.2 A) to form a chain. At physiological pH (~ 7.4), the phosphodiester group is negatively charged which contributes to the solubility of the oligonucleotide in aqueous media. The

alternating arrangement of the sugar-phosphate provides the chain with directionality and the sequence is conventionally written from 5'- to 3'-direction.

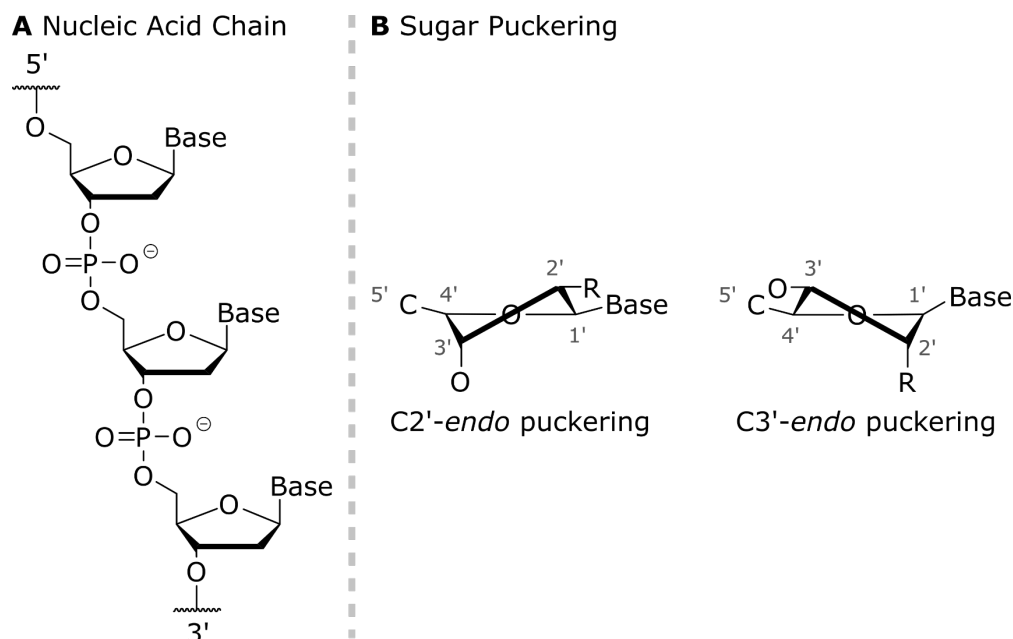


Figure 1.2: Nucleic acid chain and sugar pucker. A) Oligonucleotide chain in the 5' – 3' direction. B) Sugar pucker conformations: C2'-endo pucker (S-type) and C3'-endo pucker (N-type).

To obtain the most stable conformation, the sugar moiety in the chain can 'pucker' in which the ring can twist out of plane. The C2' and C3' can be displaced relative to the C4'–O–C1' plane. DNA adopts the S-type (Southern) conformation where there is C2'-endo pucker. RNA adopts a N-type (Northern) conformation where there is C3'-endo pucker (Figure 1.2 B). The pucker of the sugars determines the characteristics of the secondary structure of the chain.

1.1.1.2 Nucleic acid secondary structure

Two complementary strands of the nucleotide chain come together to produce a duplex (or double-stranded DNA, dsDNA). Chargaff determined that the ratio of

purines and pyrimidines in duplexes was always 1:1.⁴ This was followed by Franklin and Wilkins *et al.* with the determination, through X-ray crystallography, that dsDNA was helical.^{5,6} Finally, the duplex structure of DNA was determined by Watson and Crick who deduced that the two strands are held together through hydrogen bonding (H-bonding) and only specific bases would H-bond effectively with one another.⁷ The pattern of bonding, which became known as Watson-Crick (WC) base pairs (bp), requires an A base to pair with T (or U in RNA), using two inter-base H-bonds (Figure 1.3 A), and a G base to pair with C, using three inter-base H-bonds (Figure 1.3 B). It was also later determined that other base pairings could occur, although less commonly, such as Hoogsteen H-bonding (discussed in Chapter 2) or wobble bp, particularly in RNA.^{8,9}

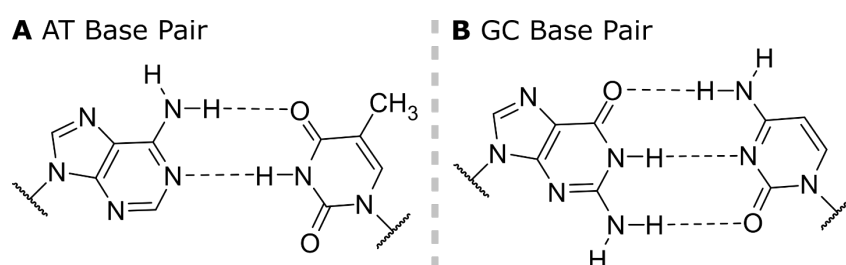
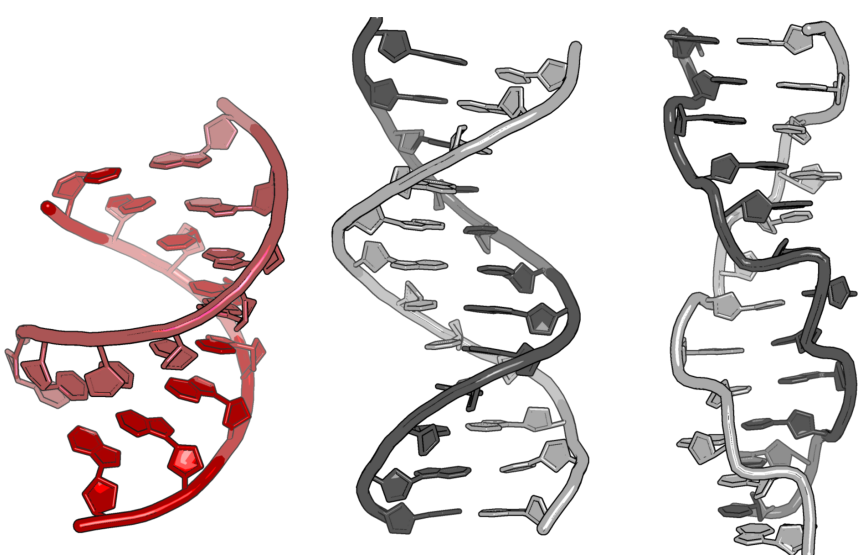


Figure 1.3: Watson-Crick base pairs. A) AT bp with two inter-base H-bonds. B) GC bp with three inter-base H-bonds. WC H-bonding indicated by dashed lines (---).

The two complementary strands of the duplex form an anti-parallel duplex (5'—3' of one strand and 3'—5' of the other) and the stacking of the base pairs on top of each other encourages the formation of the structure by stabilisation. The helical structure of the duplex hides the nucleobases within the structure to produce a hydrophobic core and exposes the negatively charged phosphate backbone (hydrophilic exterior) of the strands to the aqueous environment which also encourages formation along

with other factors such as ionic strength, H-bonding between bases and entropy of formation. With the twisting of the helix structure, the formation of a major groove and a minor groove can be seen through the full length of the duplex. The characteristics of the helix classify the nucleic acid structure into right-handed A-, B- and more uncommonly, left-handed Z-form conformations (Figure 1.4). Several factors determine the conformation of the helix including sugar pucker, helical turn and glycosidic bond conformation. A summary and comparison of the factors can be seen in Figure 1.4.



	A-RNA	B-DNA	Z-DNA
Helical Turn	32.7 Å 11 bp	35.4 Å 10.5 bp	45.6 Å 11.6 bp
Axial rise per bp	2.8 Å	3.4 Å	3.7 Å
Diameter	25.5 Å	23.7 Å	18.4 Å
Sugar Pucker	C3'-endo	C2'-endo	C3'-endo (G)/C2'-endo (C)
Glycosidic Bond	anti-	anti-	syn- (G)/anti- (C)
Major Groove	Narrow, deep	Wide	Flat
Minor Groove	Wide, shallow	Narrow	Narrow, deep

Figure 1.4: Comparison of helix conformations. A-RNA structure is an example of a 8-bp RNA:RNA duplex (red) (PDB: 1RXA).¹⁰ B-DNA structure is an example of a 12-bp DNA:DNA mixed sequence duplex (PDB: 1BNA).¹¹ Z-DNA structure is an example of a 12-bp DNA:DNA (GC)_n duplex (PDB: 4OCB).¹² Table adapted from previous studies.^{13–15}

dsDNA is predominantly in the B-conformation with a right-handed helix, the bases are perpendicular to the helical axis and the sugars in C2'-*endo* pucker. B-form DNA has a large major groove which can bind to proteins and other molecules such as triplex-forming oligonucleotides (Chapter 2). dsRNA tends to maintain an A-form conformation in a right-handed helix due to the steric bulk of the 2'-OH group and related factors. However, these conformations are not fixed as DNA can also be in A-form conformation under certain conditions such as high salt concentrations and low humidity. It is also seen in DNA:RNA heteroduplexes and is a crucial aspect of the Cas9-sgRNA-DNA complex (Chapter 3) or during transcription when DNA is converted to RNA in cells. The Z-form conformation was first discovered in 1979 with the analysis of a (CG)_n oligonucleotide.¹⁶ It is a left-handed helix and contains both *syn*- and *anti*- conformations which gives it its characteristic zig-zag structure. Although generally unfavourable, certain conditions such as super-helical stress and high salt concentrations or some cations can promote the Z-conformation.

While initially duplexes were thought to be rigid structures, nowadays it is understood that they are dynamic structures that can hybridise, anneal and denature which is key in their biological functions. The separation of two strands of DNA occurs by disruption of the H-bonds and base stacking, which can be caused by many factors including heat or pH. This transition can be analysed using various techniques including UV light absorption as a function of temperature (UV melting). The disruption of the base stacking in a dsDNA structure causes an increase in UV light

absorption, defined as hyperchromicity, and this cooperative transition from duplex to single-stranded DNA (ssDNA) can be conveniently monitored at 260 nm. The temperature at which 50% of the duplexes are completely un-hybridised is defined as the melting temperature (T_m), a parameter that is frequently referred to in this thesis.

1.1.1.3 Other secondary structures

The combination of WC base pairing in the duplex and Hoogsteen base pairing can give rise to another structure of nucleic acids called a triplex. A third strand of DNA (or RNA) can bind to the outer edges of the duplex to form a three-stranded structure made up of three-base triads called triplets (Figure 1.5 A). The concept of triplexes and their potential applications are discussed in more detail in Chapter 2.

Hoogsteen bonding between G bases can also lead to the formation of G-quadruplexes which are stacks of a square planar structure made up of four G bases called a G-tetrad (Figure 1.5 B). G-quadruplexes can form from one (intramolecular) or more strands (intermolecular) of oligonucleotides and can be parallel, anti-parallel or a mixture of both.¹⁷ C-rich ODNs can form a non-B secondary structure called an i-motif which sees the formation of hemi-protonated CC^+ base pairs in acidic pH (Figure 1.5 C).¹⁸ Multiple strands containing contiguous runs of C bases can form an intercalated structure between two parallel-stranded duplexes using this CC^+ base pair.¹⁹

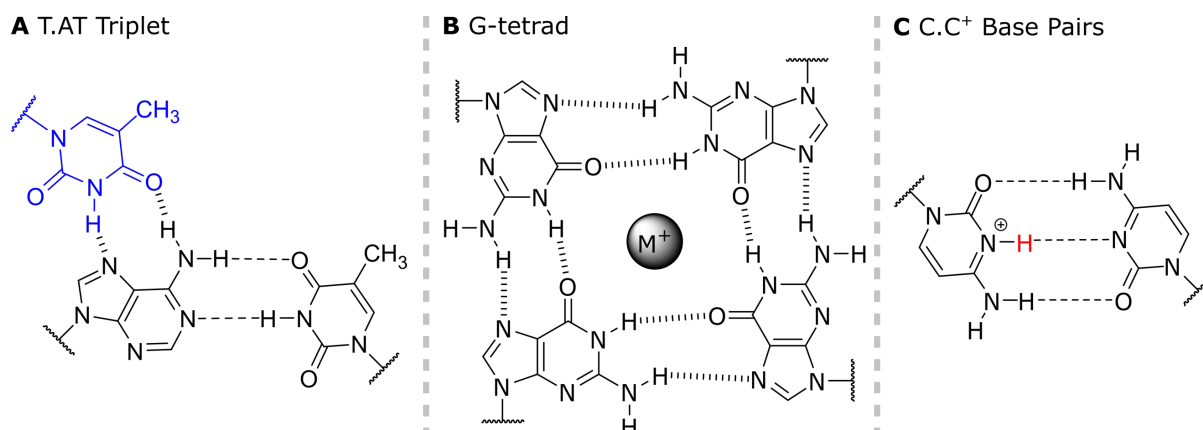


Figure 1.5: Other base pairing structures. A) T.AT triplet seen in a three-stranded triplex structure (see also Chapter 2 for more details). B) G-tetrad seen in a square planar sheet to make up G-quadruplexes. C) CC⁺ base pairs seen in i-motifs. WC H-bonding indicated by dashed lines (---) and Hoogsteen H-bonding indicated by hashed lines (||||).

1.2 Chemical Synthesis of Nucleic Acids

Oligonucleotides can be chemically synthesised using automated solid-phase synthesis methods. The use of a solid-phase compared to solution-based synthetic approach offers many advantages. Issues such as difficulties with solubility, purification, or incomplete reactions are more easily managed when the reactant is bound to a solid support (or resin). The oligonucleotide is built up stepwise from the solid support, from the 3'- to the 5'-end of the ODN, and a large excess of reagents can be washed through the solid support to push the reactions to completion, allowing for faster reaction time. The revolutionary process was established using peptides in the 1960's by Merrifield who later won the Nobel prize in Chemistry for his work.^{20,21} Today, the oligonucleotide synthesis process is automated, computer-controlled and a high-yielding process.²²⁻²⁴

The most common solid support used in the synthesis of oligonucleotides is an inorganic matrix composed of controlled pore glass (CPG) and silica gel. This solid support was pioneered by Caruthers *et al.* and its rigidity and lack of swelling in varying organic solvents made it desirable for oligonucleotide synthesis.²⁵ The pore size used with CPG resins (500 - 3000 Å) can be adjusted depending on the length of ODN synthesis required, with the larger pores generally being used for longer ODN synthesis (but have a lower loading of nucleoside). The solid support is functionalised with a reactive site and the first nucleoside required for the synthesis is attached *via* a succinyl ester through an amide bond to the resin. Polystyrene (PS) is another useful resin that contains moisture exclusion properties and is commonly used for oligonucleotide synthesis (40 nmol).²⁶

While there are many different approaches to oligonucleotide synthesis such as H-phosphonates and phosphotriester methods (discussed in Chapter 4), the current method of choice is the phosphoramidite approach, which is applicable to both DNA and RNA and was introduced in the 1980's.²⁴ A standard phosphoramidite monomer contains a nucleobase, sugar and a P(III)-group attached to the 3'-position of the sugar. The phosphoramidite is protected with a 2-cyanoethyl protecting group on the phosphite, a 4,4'-dimethoxytrityl protecting group (DMT) at the 5'-position of the sugar and, depending on the base being incorporated, a protecting group on any reactive amino groups of the nucleobase (Figure 1.6). DMT is the only acid-labile protecting group on the phosphoramidite and can, therefore, be removed before the

addition of each phosphoramidite to reveal the reactive 5'-OH group. The nucleobase protecting groups and phosphate protecting groups are all stable to acid and can be removed at the end of the synthesis using aqueous NH_3 . The final oligonucleotide product can also be cleaved from the resin using aqueous NH_3 which provides a straight-forward, one-step deprotection procedure after ODN synthesis.

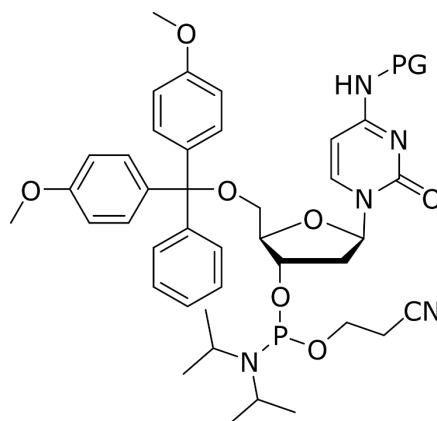


Figure 1.6: An example of protected dC phosphoramidite. dC phosphoramidite 5'-position protected with a DMT group and 3'-P(III) species protected with a 2-cyanoethyl group. A protecting group is also used on the nucleobase for the exocyclic amine (See Figure 1.9). PG = protecting group.

1.2.1 DNA chemical synthesis

1.2.1.1 Synthesis cycle

During the synthesis of an ODN, one nucleotide is added to the chain per cycle. A detailed summary of the cycle and the reagents used is given in Figure 1.7. To begin the cycle, the first step (detritylation) involves the removal of the DMT group from the 5'-position of the monomer attached to the resin. The presence of this group serves to avoid polymerisation during the functionalisation of the resin. The next step (activation and coupling) involves reacting an incoming excess of phosphoramidite (Base 2 in Figure 1.7) with an activator to convert the diisopropylamino group of the

phosphoramidite into a good leaving group. This is then readily displaced by the attack of 5'-OH group of the resin-bound monomer (Base 1 in Figure 1.7) to begin the chain. The next step (capping) is used to account for any unreacted 5'-OH group of the resin bound monomer. Unreacted 5'-OH groups could react with the phosphoramidite monomer in the next cycle (Base 3) and cause the incorrect sequence to be synthesised, so acetylation of the 5'-OH removes this possibility. This minimises the potential truncated sequences or 'failure sequences' that would otherwise accumulate through each successive cycle.

The next step (oxidation) converts the unstable phosphite triester (P(III) species), formed from the resin bound Base 1 monomer and Base 2 phosphoramidite, to a stable P(V) species through oxidation. This produces a 2-cyanoethyl protected DNA backbone. The final step (deprotection) before the next cycle removes the DMT group from the newly formed chain to prepare it for the next phosphoramidite (Base 3). The cation produced from deprotection (trityl) is an orange colour and absorbs at 495 nm. This can be monitored to determine the efficiency of the cycle in real time. Mechanisms for all steps involved in the cycle are given in Appendix A1.1.²⁷

Although each phosphoramidite coupling has a high efficiency, the repeated cycles reduce the overall yield of the oligonucleotide, making efficient DNA synthesis limited by the length of the oligonucleotide. The cumulative effect of poor coupling

would cause a poor overall yield and increase the difficulty of purification of the full-length product.

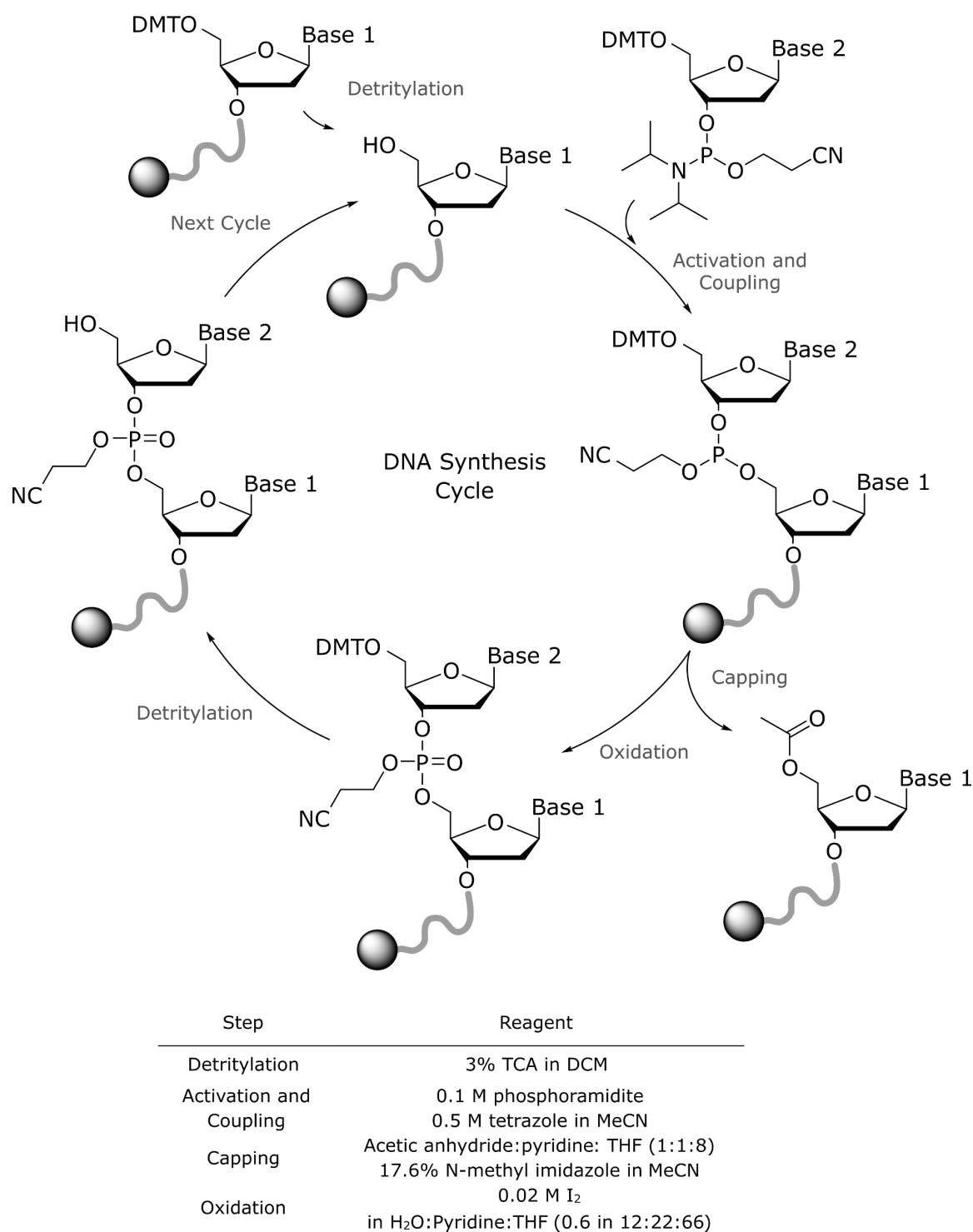


Figure 1.7: DNA solid-phase synthesis cycle and table of reagents. Solid support is indicated by sphere (grey).

1.2.1.2 Deprotection and removal from the solid support

As previously mentioned, the removal of the 2-cyanoethyl group and cleavage of the ODN from the resin occurs in aqueous NH_3 (Figure 1.8). However, the cyanoethyl group could react with the unprotected amines of the nucleobases to give an acrylonitrile adduct during the deprotection. To prevent this, the solid support bound ODN is treated with 20% diethylamine (DEA) in MeCN prior to deprotection in aqueous NH_3 . The DEA solution is not strong enough to remove the other protecting groups or cleave the ODN from the resin but will remove the 2-cyanoethyl group. Mechanisms for the deprotection and side reactions involved are given in Appendix A1.2.²⁷

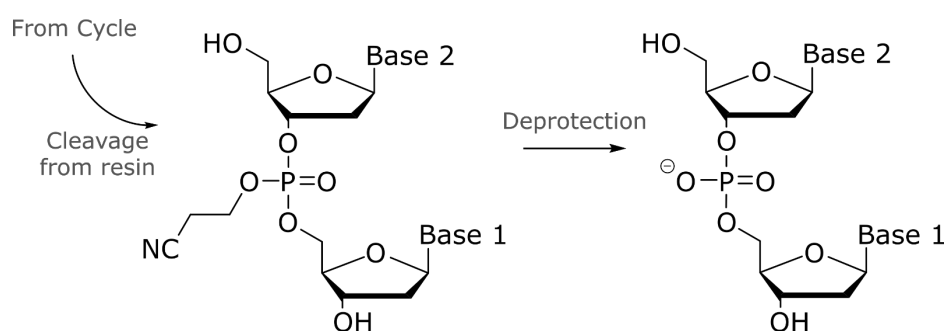


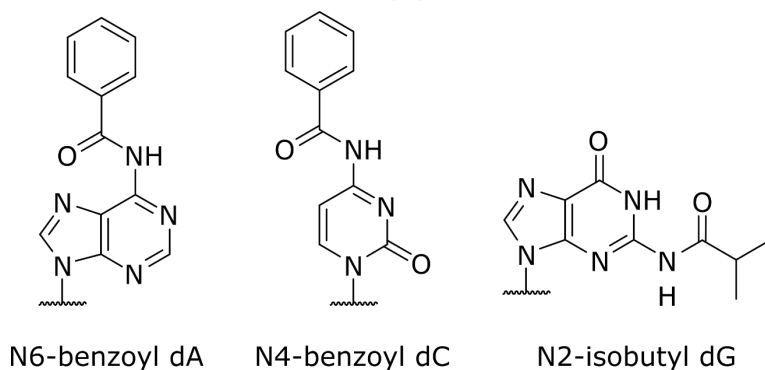
Figure 1.8: DNA cleavage and deprotection. Both steps can be carried out in aqueous NH_3 or the removal of 2-cyanoethyl (deprotection) can be carried out using 20% DEA in MeCN.

1.2.1.3 Nucleobase protecting groups and deprotection

The standard phosphoramidites for DNA synthesis contain a protecting group on each of the exocyclic amines of the nucleobase to avoid any unwanted side reactions. These include N6-benzoyl (bz) dA, N4-bz dC and N2-isobutyl (ib) dG (Figure 1.9 A) and can be removed using aqueous NH_3 . Thymine does not require base protection due to lack of a reactive exocyclic amine. Deprotection of N2-ib dG determines the time required

for deprotection for the ODNs (5 h at 55 °C) as the benzoyl groups on dA and dC are more readily cleaved by aqueous NH_3 . It is also possible to use N2-dimethylformamidyl dG which only requires 1 h at 55 °C for deprotection (Figure 1.9 B). However, its use can introduce other side-reactions.

A DNA nucleobase protecting groups (5 h)



B DNA nucleobase protecting group (1 h)

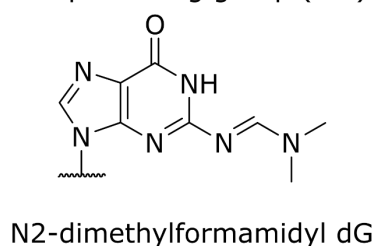


Figure 1.9: Standard DNA nucleobase protecting groups. A) Protecting groups for dA, dC and dG (N2-ib dG require 5 h in aqueous NH_3 at 55 °C for removal). B) Alternative protecting group for dG which requires 1 h in aqueous NH_3 at 55 °C for removal.

1.2.2 RNA chemical synthesis

For the synthesis of RNA, the methods are similar to DNA solid-phase synthesis except that the reactive 2'-OH group must be protected during synthesis. The 2'-OH group also increases RNA's susceptibility to degradation during and post-synthesis, which is discussed in more detail in Chapter 4. Moreover, the reagents and conditions used for the removal of the protecting group after synthesis must not degrade the RNA. The coupling times of RNA are also typically longer than DNA due to the steric

bulk of the 2'-OH protecting group adjacent to the 3'-phosphoramidite and also causes RNA coupling efficiencies to be lower.

1.2.2.1 2'-OH protecting groups and deprotection

The choice of protecting group during RNA synthesis is dictated by availability of high quality phosphoramidites. The most common choice is 2'-O-*tert*-butyldimethylsilyl (TBDMS) (Figure 1.10 A)²⁸ as all four phosphoramidites are readily available. The protecting group can be readily removed by treating the oligonucleotide with triethylamine trihydrofluoride (NEt₃·3HF) after aqueous NH₃ deprotection (Figure 1.10 B).^{29,30} However, since the TBDMS group is quite bulky, longer coupling times are needed compared to DNA coupling times during each cycle of the synthesis. The lower coupling efficiency of the monomers limits the length of efficient RNA synthesis to ~70-mer.³¹

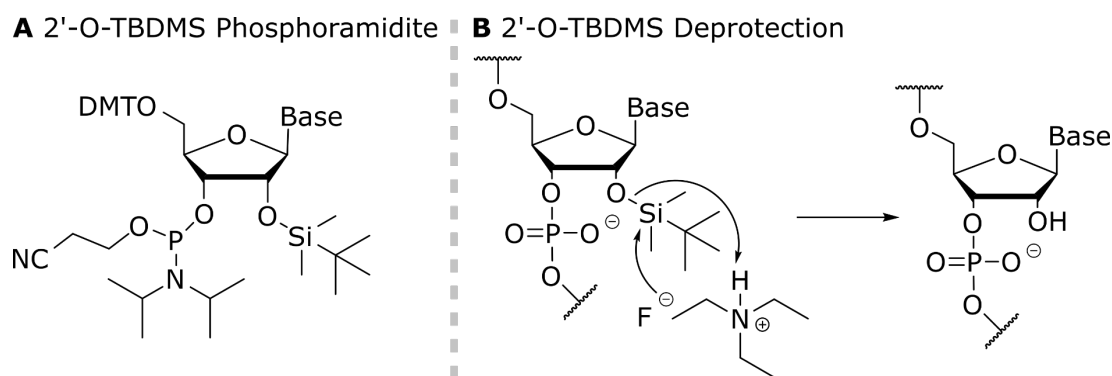


Figure 1.10: 2'-O-TBDMS RNA and deprotection. A) Phosphoramidite of 2'-O-TBDMS protected RNA. B) Deprotection of TBDMS group from the ODN using NEt₃·3HF after NH₃ deprotection.

Recently, another 2'-OH protecting group was introduced, 2'-O-thiomorpholine-4-carbothioate (TC) (Figure 1.11 A) that permits faster and more efficient RNA

synthesis. This protecting group, unlike TBDMS, can be removed in the same step as the nucleobase protecting groups and cleavage from the resin, using an ethylenediamine (EDA) solution in toluene (Figure 1.11 B), simplifying the workflow. EDA deprotection for TC is also faster than TBDMS deprotection (6 h in total vs ~14 h in total, respectively) and the coupling times for TC protected monomers are faster than TBDMS. All four TC phosphoramidites are commercially available and allow for the efficient synthesis of longer RNA (~100-mer).³²

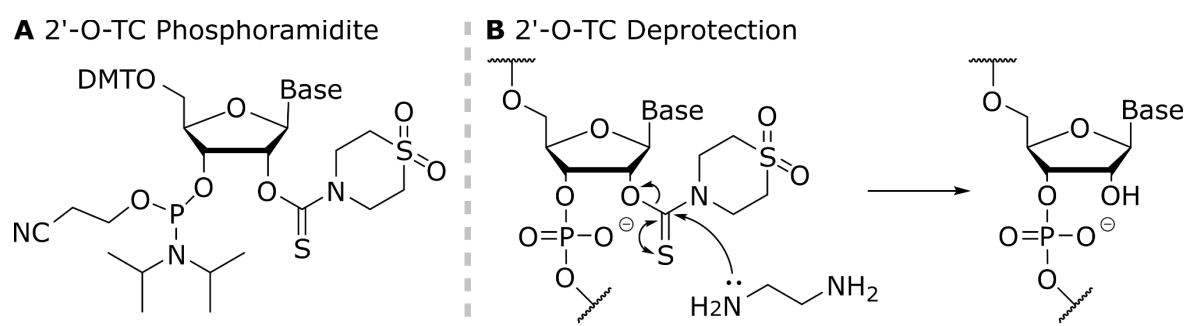


Figure 1.11: 2'-O-TC RNA and deprotection. A) Phosphoramidite of 2'-O-TC protected RNA. B) Deprotection of TC groups from the ODN using EDA.

1.2.2.2 Nucleobase protecting groups and deprotection

Different nucleobase protecting groups are used in RNA synthesis than those used in DNA synthesis. When deprotecting with aqueous NH_3 , due to RNA's susceptibility to chain cleavage or chain migration, faster deprotection times are needed. For 2'-O-TBDMS protected phosphoramidites, a *tert*-butylphenoxyacetyl (tac) group is used for A (N6), C (N4) and G (N2) (Figure 1.12). This group can be removed in aqueous NH_3 faster (2.5 h) than the DNA nucleobase protecting groups previously mentioned. The use of this group also helps the solubility of the phosphoramidite in MeCN. For 2'-O-TC protected phosphoramidites, the same protecting groups used in DNA synthesis

(benzoyl and isobutyl) can be used as they are readily removed by EDA in the 2'-O-TC deprotection.

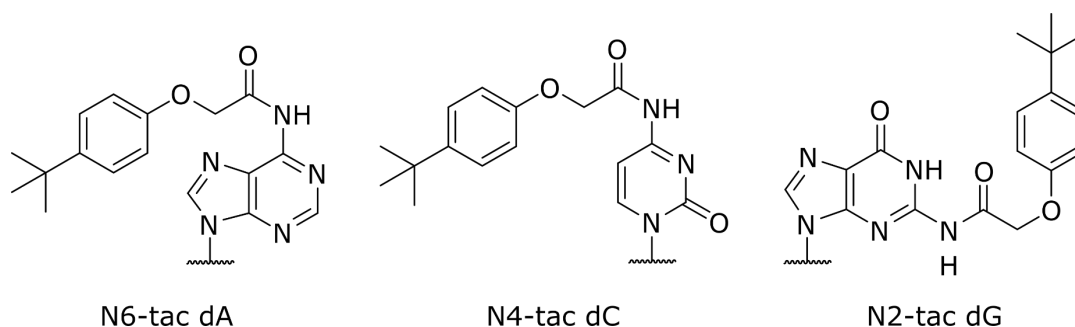


Figure 1.12: 2'-O-TBDMS RNA nucleobase protecting groups. Protecting groups for dA, dC and dG which require 2.5 h in aqueous NH_3 at 55 °C for removal.

1.3 Fluorescence

1.3.1 Basics of fluorescence

Nucleic acid research encompasses a wide range of analytical techniques and one such well-established process is fluorescence. The process of a molecule (fluorophore) absorbing a photon of light, at the excitation wavelength (λ_{ex}), then emitting a photon of light of lower energy, at the emission wavelength (λ_{em}), is called fluorescence. It is best described using a Jablonski diagram (Figure 1.13 A).³³ The phenomenon is initiated by the absorption of a photon ($h\nu_a$) and the molecule is excited from the ground state (S_0) to a higher vibrational energy level (excited singlet state, S_n). The molecule then undergoes a non-radiative relaxation to the lowest vibrational level of the excited singlet state (S_1) using processes called internal conversion (IC) or vibrational relaxation (VR). The excited molecule can then return to S_0 by the emission of a photon ($h\nu_f$), which is known as fluorescence. During fluorescence, $h\nu_f$ is lower in energy than $h\nu_a$ and so has a longer wavelength. The difference in energy

is lost as thermal energy and this is known as Stokes shift. The timescale for each process is: absorption $\sim 10^{-15}$ s, IC $\sim 10^{-12}$ s and fluorescence $\sim 10^{-8}$ s.³³

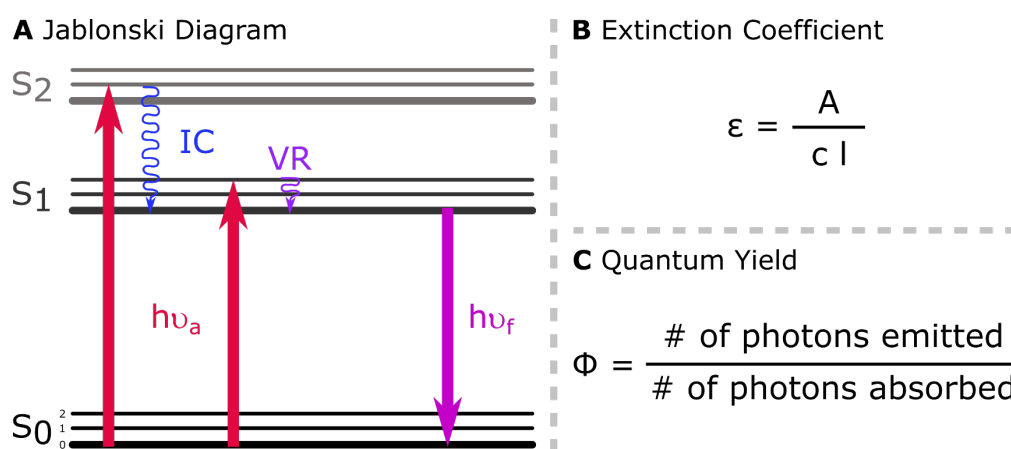


Figure 1.13: Fluorescence energy diagram and fluorescence equations. A) Adapted Jablonski energy level diagram illustrating the fluorescence process.³³ $h\nu_a$ = photon absorption, $h\nu_f$ = photon emission, S_0 = ground state, S_n = excited singlet states, IC = internal conversion, VR = vibrational relaxation. B) Equation used to determine ϵ (extinction coefficient) at a given λ ($M^{-1} \text{ cm}^{-1}$). A = absorbance at a given λ , c = concentration of the fluorophore (M) and l = path length (cm). C) Equation used to determine Φ (quantum yield).

The brightness of a fluorophore is usually defined by its extinction coefficient (ϵ) and its quantum yield (Φ). The extinction coefficient is a measure of the fluorophores efficiency in absorbing light and can be determined using the equation in Figure 1.13 B, which is derived from the Beer-Lambert Law. The quantum yield of a fluorophore determines the ratio of photons emitted relative to photons absorbed (Figure 1.13 C).

1.3.2 Fluorescence quenching

Fluorescence quenching is a common technique used in fluorescence systems and is based mainly on the energy transfer of two fluorophores or a fluorophore and quencher. Energy can be non-radiatively transferred between a donor (fluorophore) and acceptor (quencher) through quenching which is a decrease in fluorescence

emission. It occurs through two main mechanisms: dynamic quenching or static quenching and dynamic quenching can be further divided into collisional quenching and Förster Resonance Energy Transfer (FRET).³⁴

1.3.2.1 Static quenching

Static quenching or contact quenching occurs when a complex is formed between the donor and acceptor (Figure 1.14 A). The complex then absorbs light but there is no photon emission on relaxation of the excited complex to ground state. This process is also referred to as “ground-state complex formation” and the donor does not need to be in an excited state.³⁵

1.3.2.2 Collisional quenching

Collisional quenching is a dynamic form of quenching. This process occurs when fluorescence emission is reduced by intermolecular collisions with acceptor molecules and the deactivation of excited molecules without radioactive decay (Figure 1.14 B).³⁴

1.3.2.3 Förster Resonance Energy Transfer (FRET)

In this non-radiative process, energy is transferred from the donor (an excited fluorophore) to the acceptor (quencher). The donor then returns to the ground state without the emission of a photon. The acceptor can then emit the absorbed energy in the form of fluorescence. If the acceptor is a non-fluorescent molecule, the energy is

lost as heat. This process is highly distance-dependent (6^{th} power, donor-acceptor distance of typically 10 – 100 Å) and also requires sufficient spectral overlap of the donor emission spectrum and acceptor absorption spectra (Figure 1.14 C).^{36–38} As a result, the distance between donor-acceptor pairs can be measured using FRET and is commonly used in nucleic acid spatial and structural studies.^{39–41}

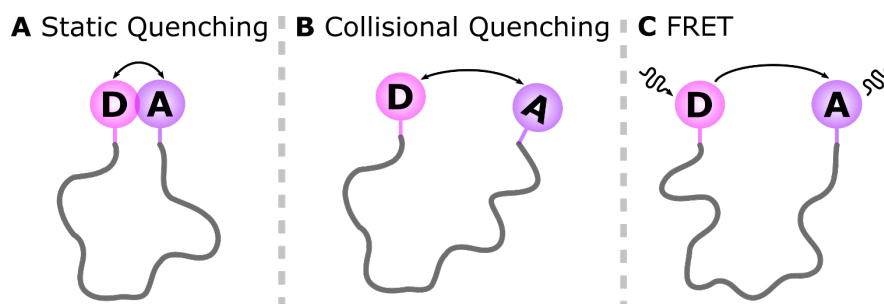


Figure 1.14: Comparison of fluorescence quenching. A) Static quenching occurs when a donor-acceptor pair form a complex. B) Collisional quenching occurs through intermolecular collisions of the donor-acceptor pair. C) FRET occurs when energy is transferred from the excited donor to the acceptor.

1.3.3 Fluorescence hybridisation probes

The use of the selective hybridisation of an oligonucleotide combined with the principles of fluorescence has led to the use of fluorescence probes in areas such as DNA diagnostics, sequencing, polymerase chain reactions and genomic imaging (discussed in greater detail in Chapter 2 and 3).^{42–47} Other applications take advantage of the sensitivity of fluorescence probes to detect genetic variations in DNA such as a single nucleotide polymorphisms.⁴⁶ Some commonly used fluorescence probes include molecular beacons and HyBeacons which are also further discussed in Chapter 3.^{48,49}

1.4 Labelling Methods

Fluorescence processes have been extensively used in nucleic acid studies both *in vitro* and *in vivo*. This is due to their high sensitivity and ease of incorporation. The incorporation of a fluorophore (covalently attached to an ODN) can be carried out during the solid-phase synthesis of the ODN or through post-synthetic labelling using a modified linker. Post-synthetic labelling is the more commonly used method for incorporation of a fluorophore as solid-phase synthesis requires the availability of a modified phosphoramidite that contains the dye attached to either the nucleobase, sugar or backbone. In addition, some fluorophores and quenchers are unstable to the conditions of oligonucleotide synthesis.⁵⁰ Although multiple post-synthetic labelling methods are available,⁵¹ the method used throughout these studies included the reaction of an amine modified ODN (through nucleobase, sugar or phosphate modifications) and a succinimidyl ester derivative (NHS-ester) of the required dye (Figure 1.15). This method of incorporation is straight-forward and generally high yielding.

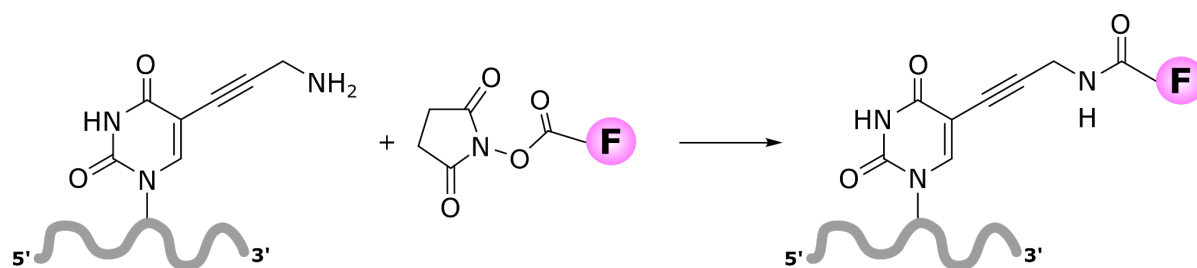


Figure 1.15: Post-synthetic labelling. Oligonucleotide containing an amine-modified nucleobase reacts with an NHS-ester dye to form an amide bond and covalently attached a fluorophore to an oligonucleotide.

1.5 Overall Hypotheses and Aims

The overall aim of the work discussed in this thesis is the incorporation of fluorescence moieties into DNA and RNA and the application of the fluorogenic probes in dsDNA detection. The specific aims of this work include;

Chapter 2 – To investigate the stabilisation of triplex-forming oligonucleotides using an intercalating fluorophore to resolve the limitations of triplex formation and expand the target range of triplex-forming oligonucleotide probes.

Chapter 3 – To investigate the incorporation of fluorogenic moieties into modified guide RNA to provide an efficient ‘on/off’ fluorescence system for imaging of genomic DNA in live cells with CRISPR-dCas9.

Chapter 4 – To develop an efficient synthetic method to incorporate an azide moiety that is compatible with 2'-O-TC chemistry with the intention of increasing the efficiency of the ‘split and click’ technique for the generation of long modified RNA needed for guide RNA studies and other applications such as guide RNA library generation.

Chapter 2 Fluorogenic Triplex- Forming Oligonucleotides

2.1 Background

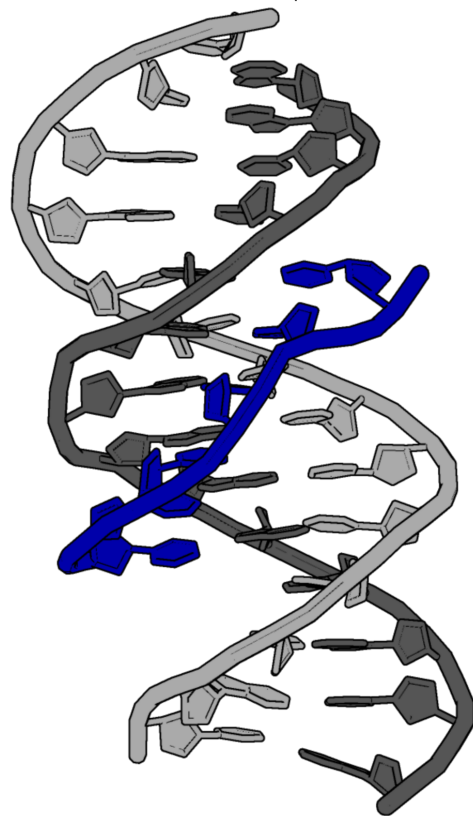
DNA is the code in which the genetic information of almost all living organisms is stored. The expression of this genetic information, in turn, determines an organism's response to environmental stimuli and is crucial for therapeutics. The ability to sequence-specifically target DNA is a necessity in therapeutics and diagnostics, and after the discovery of the double helical structure of DNA by Watson and Crick,⁷ a vast array of new technologies were developed. The characteristics of DNA as a target became well understood and the concept of an oligonucleotide probe was soon born. Unlike other DNA recognition moieties, oligonucleotides have the advantage of extended recognition sites (as long as the ODN itself) and non-target site binding is reduced.

2.1.1 Triplex-Forming Oligonucleotides (TFOs)

In 1957, Felsenfeld *et al.* showed that a third strand could interact with the double helix of DNA to form a triple-helical structure (triplex)⁵² but it was another 30 years before the potential of triplexes would truly be developed, when oligonucleotide synthesis became a more practical technology.⁵³ Triplex-forming oligonucleotides (TFOs) bind in the major groove of the double-stranded helix to form a triplex in a sequence-specific manner (Figure 2.1 A). The TFO maintains sequence specificity by interacting with the outer faces of the bases of the double helix using Hoogsteen or reverse-Hoogsteen H-bonding.^{8,54} A poly-pyrimidine (py) TFO can bind with a poly-purine (pu) stretch of DNA in a parallel orientation (both 5' — 3' direction, Figure 2.1

B). These interactions occur through the formation of three-base triads (triplets); T.AT triplets or C⁺.GC triplets (where in X.YZ, X is a TFO base and Y is a base from the poly-pu duplex strand, Figure 2.1 C). Comparably, a poly-pu TFO can interact with a poly-pu stretch of DNA in an anti-parallel direction (3' – 5' direction of the TFO, 5' – 3' direction of the target strand, Figure 2.1 D) through an A.AT, T.AT or G.GC triplet (Figure 2.1 E).^{55–57} For this chapter, I will focus on parallel triplexes due to their wider applicability.

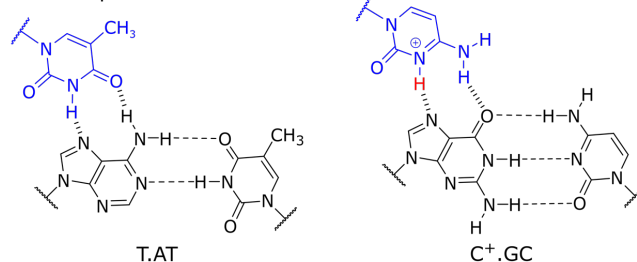
A NMR Structure of DNA Triplex



B Parallel Triplex Orientation



C Parallel Triplets



D Anti-parallel Triplex Orientation



E Anti-parallel Triplets

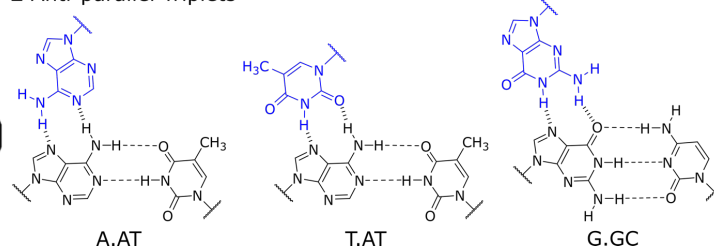


Figure 2.1: TFO structure and triplets. A) NMR structure of a parallel triplex adapted from Lane *et al.* (PDB = 1BWG).⁵⁸ TFO strand (blue) bound in the major groove of the duplex. B) Orientation of the TFO strand compared to the duplex in a parallel triplex. C) Triplet binding orientation in T.AT and C⁺.GC triplet for parallel triplexes. D) Orientation of the TFO strand compared to the duplex in an anti-parallel triplex. E) Triplet binding orientation in A.AT, T.AT and G.GC triplets for anti-parallel triplexes. TFO strand in blue and duplex strands in grey. H-bonding is indicated by dashed lines (---) and Hoogsteen bonding indicated by hashed lines (||||). Bases in TFO strand in blue and bases from duplex in black. Protonation is indicated by a red H atom.

2.1.1.1 Applications of TFOs

Triplexes have immense potential in many areas of nucleic acid research and have already been used for such. TFOs have been used in DNA mutagenesis,⁵⁹ gene inhibition,^{60–62} inhibition of DNA-protein binding,^{63,64} DNA repair and recombination⁵⁶ and nanotechnology.⁶⁵ One particular application that has received great attention is the attachment of artificial metallo-nucleases (AMNs) to TFOs for the cleavage of the dsDNA target. AMNs use redox metal ions to damage nucleotides to initiate strand excision.^{66,67} The use of a TFO allows for sequence-selective target cleavage. An extensive amount of work has been carried out with AMNs such as Fe(II)-EDTA complexes tethered to TFOs^{53,68–70} and other Fe(II) complexes.^{71–73} Cu(II) complexes have also been studied in great detail both as free cleavage agents^{74–76} and attached to a TFO.^{77,78}

2.1.1.2 Limitations of TFOs

Since their discovery, extensive biophysical, structural and thermodynamic studies have been carried out on triplex formation.^{58,79–82} Stable triplex formation is determined by a number of factors.⁸³ Firstly, to form a C⁺.GC triplet in parallel triplexes, protonation at the N3 position of the cytosine base (which has a pK_a of ~4.5 as a free nucleotide but is considered to be higher in a triplex state)⁸⁴ is required to form the Hoogsteen H-bond (Figure 2.1 C, red H and hashed lines). This means that parallel triplexes require an acidic environment to form.^{85,86} The range of acidic pH required for this protonation is dependent on the position of the C base within the

TFO. Terminal cytosines residues have a lower pK_a than internally positioned cytosines.⁸⁷

This necessity for protonation also induces another limitation on the sequence context of the target. The stability of the triplex is determined by the GC content of the target duplex, as a high ratio of G bases in the poly-pu target requires a corresponding number of C bases in the TFO. Contiguous runs of C bases or high cytosine content will destabilise the triplex, as close proximity of the positive charges leads to charge repulsion.⁸⁸ Another limitation associated with triplexes is the restricted target range. Figure 2.1 C and E show the most common and stable triplets that can form in the triplex, which means that the target is limited to a poly-pu target strand that contains only A and G nucleobases. This limits the target scope of triplexes. Finally, triplexes tend to have a lower binding affinity (TFO binding to duplex) compared to duplex formation which is partly due to the repulsion of three negatively charged strands in close proximity.⁸⁴

Much research has been devoted to solving the ‘triplex stability’ problem. The majority of these studies involve modifying the TFO through base, sugar and backbone modifications or introducing an external moiety to alleviate the triplex instability. These strategies are discussed below, and the modifications are designed to improve certain aspects of triplex formation, but the challenge remains to devise a simple, universal approach that can tackle them all.

2.1.2 Tackling pH dependence of TFOs

As previously discussed, one limitation is the need for an acidic environment for the formation of the C⁺.GC triplet, which causes triplexes to be unstable at high pH. Cytosine analogues, which form triplets with GC base pairs, have been studied (Figure 2.2 A). The T.AT and C⁺.GC triplets are isostructural and hence, do not cause any major distortions in the global structure of the triplex.⁸⁹ The use of a cytosine analogue rather than, for example, a bicyclic purine analogue maintains this isostructural characteristic in the modified triplet. Replacing cytosine with 5-methyl cytosine (^{Me}C) (Figure 2.2 B) promotes stability.⁸⁸ This is due to the contribution of the methyl group to the base stacking of the triplex and a slight increase in pK_a of the nucleobase.⁹⁰ However, this increase in stability is not enough for efficient triplex formation at higher pH. Subsequent studies used 5-bromo-substituted dC bases⁹¹ and other modifications used include 2-amino-pyridine (Figure 2.2 C), which has a higher pK_a than cytosine, and improved triplex stability by ~4 °C compared to cytosine.⁹²

Other modifications investigated were designed to remove the need for nucleobase protonation altogether by using bases that would form the second Hoogsteen H-bond without protonation. However, the contribution of protonation to the stability of the triplet must not be overlooked. The C⁺.GC triplet is more stable than the T.AT triplet and this is most likely due to a favourable electrostatic interaction of the positive charge with the anionic phosphate backbones of the triplex.^{87,93} These modifications include pseudoisocytosine (Figure 2.2 D) and its derivatives which can be used to

target contiguous guanine bases to form stable triplexes at neutral pH,^{94–97} or 6-oxocytosine (Figure 2.2 E), which is more stable at neutral pH but surprisingly, contiguous runs of this base are destabilising.^{98–101} A separate approach explored the use of purine analogues for GC recognition. These included 8-oxoadenine (Figure 2.2 F),^{102–105} N7-glycosylated-guanine (Figure 2.2 G)^{106–108} and isoguanine (Figure 2.2 H).^{109–112} In most cases, these modifications can contribute to the formation of triplexes at neutral pH but are more ideally suited to targeting contiguous runs of GC base pairs as the lack of isomorphism distorts the triplex structure.

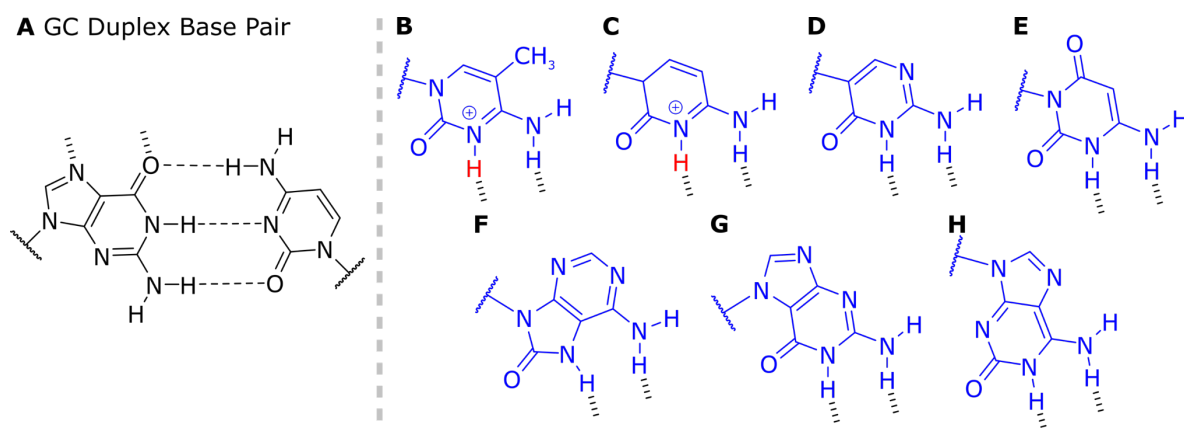


Figure 2.2: Modifications that target GC base pair to tackle pH dependence of triplexes. A) GC base pair of duplex with possible Hoogsteen binding sites indicated with hashed lines. B) ^{Me}C C) 2-amino-pyridine D) pseudocytosine E) 6-oxocytosine F) 8-oxoadenine G) N7-glycosylated guanine H) isoguanine. Hoogsteen binding sites indicated by hashed lines (||||). Protonation indicated with red H atom.

2.1.3 Tackling the limited target range for TFOs

To expand the sequence target range of TFOs, studies of triplets that form with ‘base inversions’ (CG base pairs (Figure 2.3 A) or TA base pairs (Figure 2.4 A)) have been carried out. An inversion is formed by a py base in a poly-pu strand. It is possible to form a triplet with these inversions using natural bases such as a T.CG triplet (Figure

2.3 B) but, the stability of this triplet was greatly reduced compared to the standard T.AT and C⁺.GC triplets.¹¹³ The G.TA triplet (Figure 2.4 B) showed the most promise when using natural bases but this triplet is still much less stable (-10 °C) than the other standard triplets.^{93,114} Other researchers investigated the use of modified bases in the TFO to encourage triplet formation with a TA or CG base pair.

2.1.3.1 CG-recognition monomers

Several modified bases that can recognise the CG inversion have been developed over the years. The rationale behind these modifications is described herein and broader reviews are available.^{115–118}

Cytosine analogues such as 2-aminopyrimidine (Figure 2.3 C) have been investigated based on the logic that the pK_a of N3 in this modification is lower (~3.3) than that of cytosine and remains unprotonated at triplex forming pH.¹¹⁹ This means the modified base will form a much weaker triplet with a GC base pair (only one Hoogsteen H-bond), which is desirable, as selectivity for CG over GC would be an issue when using C analogues. It also means that N3 becomes a much stronger H-bond acceptor for Hoogsteen H-bonding with the cytosine in the target strand and this is evident from triplex formation at neutral pH.

Brown *et al.*, have carried out a range of studies on CG-recognition monomers, substituted phenyl-*N*-methylpyrrolo dC nucleosides (^xP) (Figure 2.3 D).^{120,121} The best

of these was a $^{\text{Ph}}\text{P}$ modification since the modification, potentially, formed two bonds with the CG base pair, one of which is an unconventional C—H—O bond, whilst also contributing to base stacking due to the introduction of additional aromatic rings. Thymidine analogues such as $^{\text{4HT}}$ have also been studied; this also forms the C—H—O bond and has lower affinity for the competing AT base pair (Figure 2.3 E).¹²²

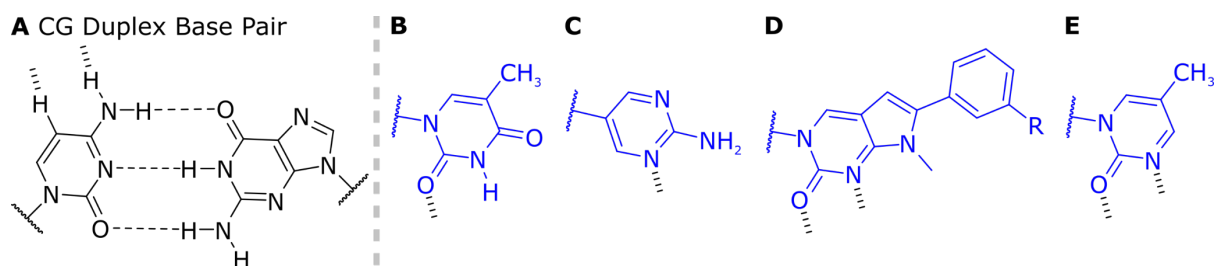


Figure 2.3: Modifications that target the CG base pair inversion to expand the target range of triplexes. A) CG base pair of duplex with possible Hoogsteen binding sites indicated B) Thymine C) 2-aminopyrimidine D) $^{\text{X}}\text{P}$ E) $^{\text{4HT}}$. Hoogsteen binding sites indicated by hashed lines (||||).

2.1.3.2 TA-recognition monomers

Similarly, work has been done on forming a stable X.TA triplet but this becomes more difficult than X.CG triplets since the methyl group of T protrudes into the expected binding area. A solution to this is the use of a longer modification that extends beyond the area that contains the methyl group and allows binding with both T and A bases in the base pair. One example of this is the thiazolyl aniline monomer, S (Figure 2.4 C).¹²³ This recognises TA inversions and has been used to form stable triplexes. However, it can also bind with CG inversions which reduces its selectivity as a TA recognition monomer.¹²⁴

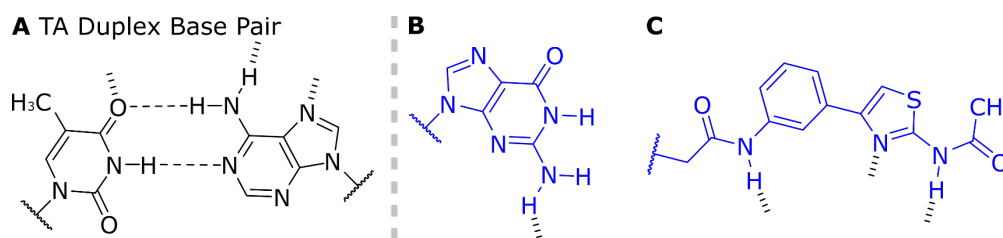


Figure 2.4: Modifications that target the TA base pair inversion to expand the target range of triplexes. A) TA base pair of duplex with possible Hoogsteen binding sites indicated B) Guanine C) S. Hoogsteen binding sites indicated by hashed lines (||||).

2.1.4 Other stabilisation techniques

2.1.4.1 Base stacking and hydrophobic interactions

In other approaches to improve the stability of triplexes, researchers investigated the contributing factors to the structural stability of the triplex. These include moieties that promote base stacking. Modified bases that contain extra aromatic systems such as the previously mentioned CG-recognition monomer, N-methylpyrrolo-dC, (Figure 2.3 D) take advantage of this.¹²¹ Introducing thiocarbonyl groups into the T and ^{Me}C bases (Figure 2.5 A and B, respectively) also results in stronger stacking effects.¹²⁵ As seen with the use of a methyl group on the base, modifications that enhance hydrophobic interactions can also have a stabilising effect, the most prominent being the use of propynyl groups in the base such as 5-propynyl deoxyuridine and its derivatives (Figure 2.5 C).^{126–128}

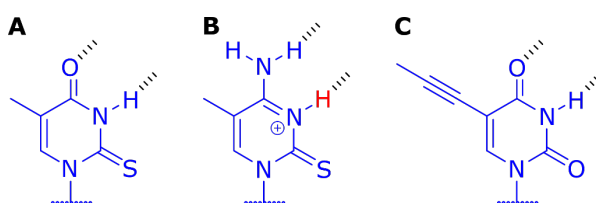


Figure 2.5: Modifications that contribute to the base stacking and hydrophobic interactions of the triplex. A) 2-thio T B) 2-thio ^{Me}C C) 5-propynyl U. Hoogsteen binding sites indicated by hashed lines (||||). Protonation indicated with red H atom.

2.1.4.2 Modified backbones and sugars

The introduction of sugar modifications can also afford more stable triplexes, such as 2'-O-Me modifications¹²⁹ (Figure 2.6 A) and Locked Nucleic Acids (LNA, Figure 2.6 B)¹³⁰ as these modifications force an N-type sugar pucker in the TFO which favours more stable triplexes. Completely modifying the TFOs with LNA moieties was shown to be destabilising and a general rule of alternating the LNA modification every 2-3 bases provides the best stabilisation.¹³¹ Other modifications to the phosphate backbone include the use of morpholino¹³² (Figure 2.6 C) or PNA¹³³ (Figure 2.6 D) backbones which do not have a negative charge and, therefore, reduced inter-strand repulsion in triplexes.

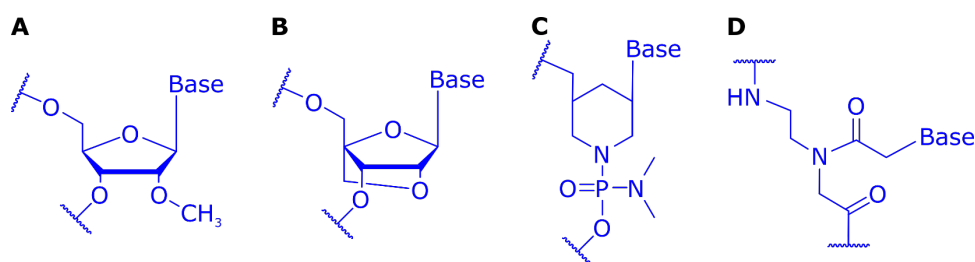


Figure 2.6: Modifications to the sugar or backbone to stabilise triplex formation. A) 2'-O-Me B) LNA which encourage N-type sugar pucker and C) Morpholino backbone and D) PNA backbone which reduce the overall negative charge of the triplex.

2.1.4.3 Cationic TFOs

Another form of stabilisation is the use of cationic moieties in triplex formation. This reduces the charge repulsion caused by the close proximity of the three anionic strands. Spermine, a polyamine found in the nuclei of cells,¹³⁴ can be tethered to the TFO to increase triplex stability.⁸⁴ Divalent cations such as Mg^{2+} are often used to help triplex formation, although the sequence context of the TFO can have an affect here.

There is more ionic dependence for T.AT triplets compared with C⁺.GC triplets due to the positive charge on the triplet. Monovalent cations such as Na⁺ are also used for stabilisation, however, a balance between the cationic moieties and the GC content of the target must be considered, as excessive positive charge could encourage destabilisation.¹³⁵

2.1.4.4 Cross-linking, binding ligands and intercalators

The use of small molecules for triplex stabilisation has also been explored. Cross-linking, in terms of triplex formation, is a technology that tethers a chemically reactive group to the TFO that will react with the target duplex to covalently attach the TFO to the duplex. The use of psoralen which can photo-crosslink with a TpA (T-phosphate-A) site is widely used for triplex stabilisation.^{136–138} Other reagents provide stabilisation through intercalation. Intercalation occurs when a planar molecule binds in between base pairs and contributes to base stacking (Figure 2.7 A) and is very effective in both duplex and triplex stabilisation. Binding ligands specific to triplexes such as benzopyridoindole,¹³⁹ BePI¹⁴⁰ and coralyne¹⁴¹ (Figure 2.7 B) encourage stabilisation through intercalation into the triplex. These molecules contain extended aromatic rings which contribute to stacking and their planar shape encourages intercalation into the triplex. Intercalators have been used both as moieties free in solution or tethered to the TFO to provide stability, with the latter being more useful.^{142–147} They provide a simple and versatile approach to triplex stabilisation.

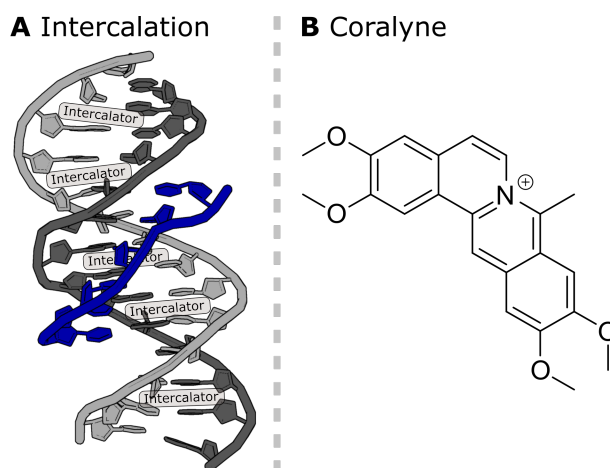


Figure 2.7: Binding ligands and intercalators. A) The intercalator, which is planar in conformation, can slot in between the stacks of base pairs to provide stabilisation by contributing to base stacking. TFO indicated in blue, duplex indicated in grey, intercalator indicated by grey box. B) Structure of triplex binding ligand Coralyne.

2.1.5 Fluorescent TFOs

The applications of triplexes are immense and have already been studied in great detail. One important example is the use of fluorescent TFOs for sequence specific detection and visualisation of dsDNA.^{42,148} This approach incorporates a fluorescent dye into the TFO and detects triplex formation through fluorescence analysis. However, these methods tend to also require a quencher to allow for a specific fluorescence enhancement upon hybridisation with the target (quenched system (TFO) to fluorescent system (triplex)). The most common approach is the use of molecular beacons.^{149,150} While triplex molecular beacons have been explored, a simpler approach may be more beneficial: the incorporation into the TFO of a dye that can be either quenched or fluorescent depending on its environment.

2.1.6 Thiazole orange (TO)

Thiazole orange (TO) is an asymmetric cyanine dye that is composed of two aromatic ring systems that are connected through a methine bridge (Figure 2.8 A and B). The specific characteristics of TO means it is a “*torsionally responsive fluorogenic dye*”.¹⁵¹ The most crucial characteristic is the environmental sensitivity of its fluorescence response. The fluorescence change depends on the conformation of the molecule, particularly the relative rotation of the heterocycles at the methine bridge.^{152–154} Restricted rotation and a planar conformation cause a major increase in fluorescence of the dye and, because of this, TO is commonly used in the detection of DNA.^{155–158}

TO can act as an intercalator in the presence of DNA and once intercalated, its planar conformation is stabilised by stacking with the base pairs and results in a fluorescence response. For this reason, TO is a significant tool in fluorescence detection of dsDNA^{159,160} and other secondary structures such as triplexes.^{161,162} It removes the need for a quencher as TO, when not intercalated, has low fluorescence. Derivatives of the molecule such as the homodimeric TO dye, TOTO, have been studied in great detail. TOTO is a symmetrical dimer of TO (Figure 2.8 C) and selectively binds to dsDNA using bis-intercalation as it contains two intercalating moieties. Bis-intercalation provides TOTO with a strong binding affinity towards nucleic acids.^{163–165}

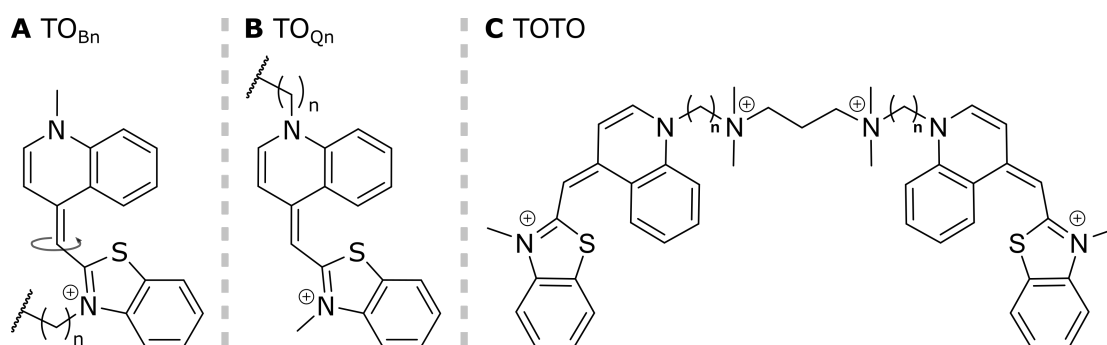


Figure 2.8: Structures of TO moieties. A) TO attached to DNA through the benzothiazole ring and B) through the quinoline ring. The arrow indicates the possible rotation around the methine bridge. n is number of carbons in the linker used in this study. B) Structure of TOTO homodimeric dye. n is the number of carbons used in the linker.

2.1.7 TO and TFOs

This intercalating fluorophore can be used as a free untethered molecule in solution for TFO stabilisation and detection.^{166,167} However, for sequence specificity, studies have focused on tethering TO directly to the TFOs. TO can be tethered to DNA through its benzothiazole moiety (TO_B) (Figure 2.8 A) or its quinoline moiety (TO_Q) (Figure 2.8 B). The fluorescence ‘light-switch’ ability of TO is adopted in many approaches, for example, tethering TO to a triplex-forming PNA strand¹⁶⁸ or using TO as a base surrogate in a triplex-forming PNA strand for the detection of dsRNA.^{169,170}

In terms of DNA TFOs, Asseline *et al.* conjugated a range of TO derivatives to the 5'-end of the TFO through a phosphothiolodiester bond and a varied-length carbon linker (Figure 2.9 A).¹⁷¹ They determined that a dye-linker length of at least 5/6 carbons or more is needed for successful intercalation of the dye, and that conjugating the dye through the benzothiazole moiety compared to the quinoline moiety was more effective for triplex stabilisation. While stabilisation and fluorescence

enhancement were achieved with the LU-TFOs (Light Up-TFOs), this stabilisation was not enough to form triplexes at neutral pH. As well as this, only a limited range of target sequences was studied in this work.

Filichev *et al.* also carried out similar work in which TO was tethered to parallel and anti-parallel TFOs in an analogue of TINA (Twisted Intercalating Nucleic Acid). TINA modifications are used to circumvent the self-aggregation of G-rich TFOs which hinders the formation of anti-parallel triplexes. These analogues were introduced to produce AFENA-TINA (Assembly Dependent Fluorescence Enhancing Nucleic Acids) (Figure 2.9 B).¹⁴³ However, the stabilisation was not as high as expected, and at neutral pH, triplexes could only be formed for anti-parallel TFOs. These studies were also limited in the range of sequences studied, and the authors did not investigate any triplex inversions.

Okamoto *et al.* explored two TO units incorporated at the 5-position of the thymine base in their ECHO probes (Exciton-Controlled Hybridisation-Sensitive Fluorescent Oligonucleotides) (Figure 2.9 C).¹⁷² The rationale behind the approach was that two molecules of TO would allow for suppressed fluorescence by utilising the excitonic interaction between them. The best example of ECHO probes provided a +17 °C increase in melting temperature at neutral pH and a ~67-fold increase in fluorescence compared to the unlabelled version. However, a second addition of the TO moiety did not provide an additive effect and, in fact, caused poorer fluorescence and weaker

stabilisation. The sequence context in this study also did not take into account base pair inversions, and triplexes were not observed at pH 8 (TFO applicability at this pH would include their use in enzymatic reactions such as polymerase chain reaction (PCR)).

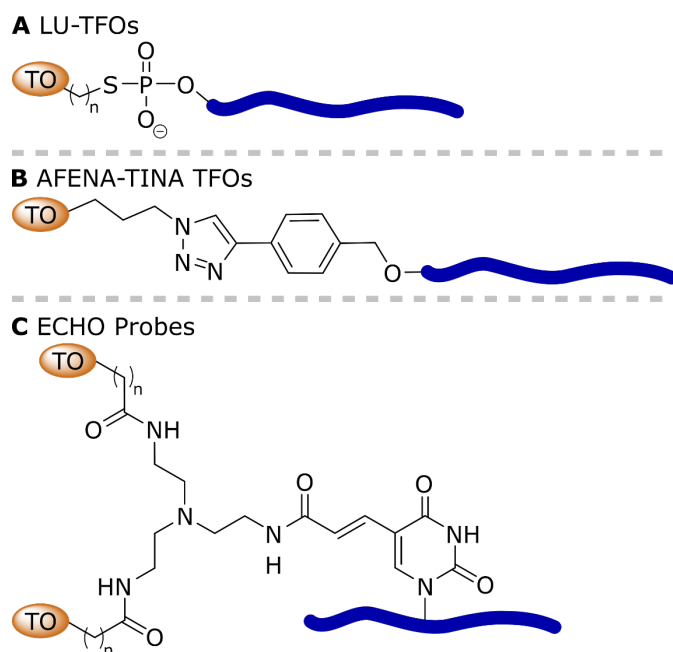


Figure 2.9: Studies of TO incorporation into TFOs. A) LU-TFOs that incorporated TO through a phosphothiolodiester linker at the 5'-end of the TFO. B) AFENA-TINA TFOs that use an intercalating twisted moiety to incorporate TO at the 5'-end. C) ECHO probes that introduce a TO dimer unit at the 5-position of a base of the TFO.

Although there have been some studies on the incorporation of TO into TFOs, there remains a need for a simple approach to triplex stabilisation using an intercalating fluorophore that can be translated across all sequence contexts and can expand the triplex target range to include base pair inversions at biologically relevant pH ranges.

2.2 Hypotheses and Aims

As previously mentioned, triplexes have great potential for diagnostic and therapeutic applications, but are limited by some issues; 1) limited target range – they require a poly-pu target to form stable canonical triplets, 2) necessity for low pH – protonation of cytosine to form the C⁺.GC triplet limits triplex formation to an acidic environment and 3) low binding affinity – the polyanionic nature of the triplex reduces its overall binding affinity.

The aim of this chapter is to investigate the stability of triplexes with the incorporation of a fluorogenic intercalator, TO, tethered to the TFO to produce TOTFOs (Thiazole Orange Triplex-Forming Oligonucleotides). The use of the fluorogenic intercalator would help improve the stability of the triplex and also provide a fluorescence detection method to indicate the formation of the triplex without major reduction of the selectivity of the TFOs. Inversions can be destabilising in triplex formation, so another aim is to counteract this destabilisation with a stronger binding affinity from the tethered intercalator. Increased stability could also allow triplex formation at neutral pH and higher. A combination of stability at biologically relevant pH, an expanded target range and sequence-selectivity are key for effective DNA targeting probes. Note, the studies described below have now been published¹⁷³ but they are expanded on in greater detail in this chapter.

2.3 Triplex Design and Synthesis

Previously studies have been carried out to incorporate TO into TFOs and evaluate the resultant probes, but only in a limited sequence environment. In the current work, to evaluate the versatility of the TOTFO probes, the TFOs were designed to target duplexes with varying degrees of GC content. As previously discussed, the number of cytosine bases in the TFO can have a substantial effect on the stability of the triplex. A total of five TFOs were designed to form parallel triplexes (TRIP-X, triplexes with the subscript “i” indicates the presence of an inversion). These comprised of 1) a T-rich TFO to target an A-rich purine strand (TFO-T in TRIP-T), 2) a T-rich TFO to target a duplex containing an CG inversion in the AT-rich sequence (TFO-T_i in TRIP-T_i), 3) a C-rich TFO targeting a G-rich purine strand (TFO-C in TRIP-C), 4) a C-rich TFO to target a duplex with a CG inversion in a GC-rich duplex (TFO-C_i in TRIP-C_i) and 5) a mixed T:C TFO that targeted a duplex with a 47% GC content (TFO-TC in TRIP-TC). This would provide a comprehensive overview of the stabilisation effect of the TOTFOs for all possible target sequence environment. The full set of triplex sequences is given in Figure 2.10 A-E.

As seen above, the ability of the TO stabilisation to compensate for the destabilising effect of a CG base inversion is investigated. Therefore, a CG-recognition monomer in the TFO is required to form a triplet with the base inversion. For this, 5-(1-propynyl)-deoxycytidine (pC, Glen Research) (Figure 2.10 F), a commercially-available C analogue was selected as the propynyl group reduces the pK_a at the N3

position of cytosine which will allow it to interact with CG and not GC to form a triplet, as shown in previous studies.¹¹⁹ Interestingly, pC has been previously investigated in TFOs as a GC-recognition monomer but causes destabilisation, suggesting it would be better suited for CG-recognition.¹⁷⁴ The alkyne group of pC also contributes to base stacking and it can potentially form the unconventional C—H—O bond seen in other CG-recognition monomers in Section 2.1.3.1.

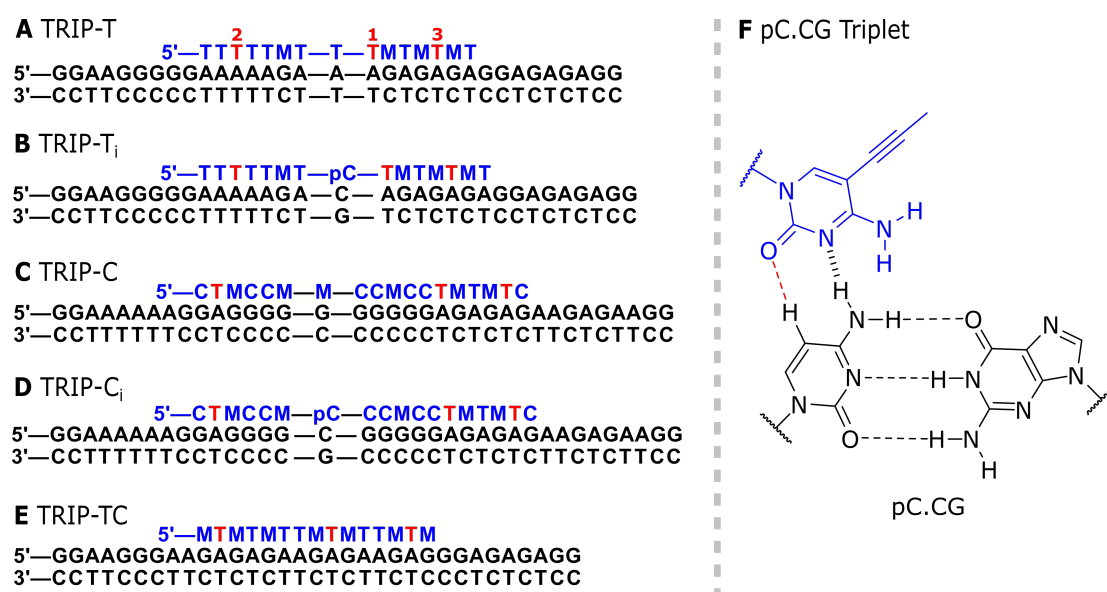


Figure 2.10: Sequences of triplexes studied. A-E) Sequences of triplexes and their corresponding target duplexes. TFOs are indicated in blue, **T** = position of TO incorporation through the linker modification and follow the order of 1, 2 and 3 indicated in A, M = ^{Me}C, pC = 5-(1-propynyl) dC. Triplexes with the subscript “i” indicates the presence of an inversion. For B) and D), site of inversion is denoted by —C— and in A) and C), the comparative triplet is marked similarly. F) Binding orientation for pC.CG triplet. Bases from TFO strand in blue, bases from duplex in black, C—H—O bond indicated in red.

All TFOs and their corresponding duplexes were prepared using automated solid-phase synthesis and good yields were obtained for all sequences. The modified nucleosides 5-(1-propargylamino)-dU (pdU, Experimental 5.1.3), 2'-aminoethoxy-T (AE, ATDBio Ltd), 5-amino-C6-dT (AC6, Glen Research) and pC were incorporated using phosphoramidite monomers and were purified according to Experimental 5.2.4.

All TFOs were labelled using active ester labelling of amines, and TO active esters were synthesised according to Experimental 5.1.4. Figure 2.11 A shows the experimental yields of each TFO after labelling. One (1xTO), two (2xTO) and three (3xTO) additions of TO were investigated. The positions of the TO incorporation were spaced out within the TFO sequence (as indicated by **T** in Figure 2.10). If the incorporations of TO were too close together, intercalation of the dye could be hindered by clustered TO units (incorporations follow the order of internal, towards 5'-end and towards the 3'-end as indicated by the numbering in Figure 2.10 A). Purity of all ODNs was determined using UPLC-MS as can be seen for all three additions in TFO-C_i in Figure 2.11 B (Full UPLC-MS data given in Appendix A4.1).

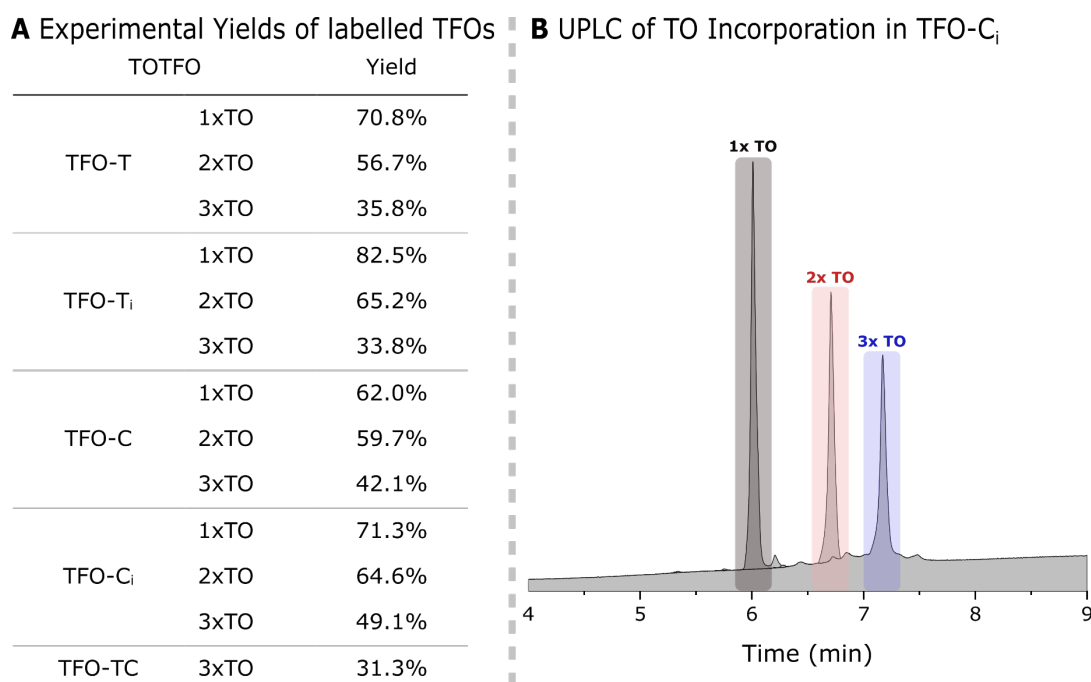


Figure 2.11: Yield and UPLC-MS for TO labelling of TFOs. A) Experimental yields of TFOs labelled with single/multiple TO through active ester labelling. B) UPLC chromatograms of pure 1xTO, 2xTO and 3xTO labelled TRIP-C_i. UPLC x-axis = time (min) and y-axis = UV absorbance at 260 nm.

2.4 Biophysical Studies of Triplex Formation

2.4.1 Triplex melting temperatures at pH 5.8

2.4.1.1 Poly-pu triplexes

Initially, the choice of linker used to conjugate TO to the TFO was investigated. Three linkers were examined, based on pdU, AC6 and AE. Thermal melting analysis was carried out to determine the melting temperature (T_m) of the triplex. To ensure reproducibility and avoid human error/artefacts in the experiments, the final measurement was determined from two independent experiments and each experiment was an average of five ramps. From Figure 2.12 A and B, we can see that in the case of 1xTO, incorporation through pdU modification provided the best stabilisation (higher T_m and sharper/more intense UV transition, $\Delta T_m = +10.6$ °C compared to the unlabelled control). Incorporation of TO through the other linkers was less triplex-stabilising (1xTO AC6 $\Delta T_m = +9.4$ °C and 1xTO AE $\Delta T_m = +9.0$ °C). This is most likely due to the contribution of the triple bond of pdU to base stacking. 5-(1-Propargylamino)-dC (Experimental 5.1.2) was also investigated as a linker. This modification would provide a TO linkage (amino group) and CG-recognition moiety (propynyl group) in a single monomer, but it resulted in a less stable triplex compared to when the amino group and propynyl group were on separate monomers ($\Delta T_m = -15.4$ °C compared to TRIP-T_i). pdU was used to for all further TO incorporations.

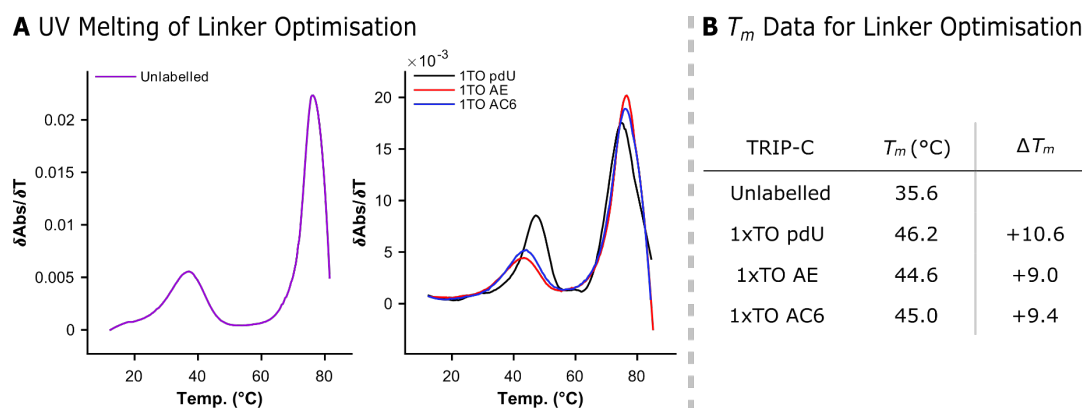


Figure 2.12: TO linker comparison. A) Smoothed first derivatives determined from the UV absorbance at 260 nm of unlabelled (left) and 1xTO pdU, 1xTO AE and 1xTO AC6 linkages for TRIP-C in 10 mM phosphate, 150 mM NaCl, 2 mM MgCl₂ at pH 5.8. The first transition is triplex denaturation and the second (higher temperature) is duplex denaturation. UV absorbance spectra can be seen in Figure A7. C) T_m data was obtained from the maxima of the triplex transition in the first derivative spectrum. Final T_m values are an average of two individual measurements in which each is an average of 5 ramps.

The triplexes were examined in two different buffers: a 10 mM MgCl₂ buffer ('high Mg²⁺ buffer') and one closer to cellular salt conditions using Na⁺ ions instead of K⁺ ions: 150 mM NaCl/2 mM MgCl₂ ('low Mg²⁺ buffer'). Both modes of TO attachment were examined at pH 5.8 to determine the most triplex-stabilising linkage; TO_{B6} – linked through the benzothiazole moiety, and TO_{Q6} – linked through the quinoline moiety (Figure 2.13 A and B, respectively). Although TO_{Q6} conjugated to TFO-T provided a large increase in stabilisation of $\Delta T_m = +14.2$ °C (compared with the unmodified TRIP-T in high Mg²⁺ buffer), the second addition of TO_{Q6} did not afford any additional triplex stabilisation (Figure 2.13 C). However, one TO_{B6} conjugated to TFO-T provided a larger increase ($\Delta T_m = +23.3$ °C compared to unmodified TRIP-T in the same buffer) and additional increases were seen with a second and third addition with 3xTO TRIP-T increasing by $\Delta T_m = +34.8$ °C. Therefore, TO_{B6} was used for the rest of this study.

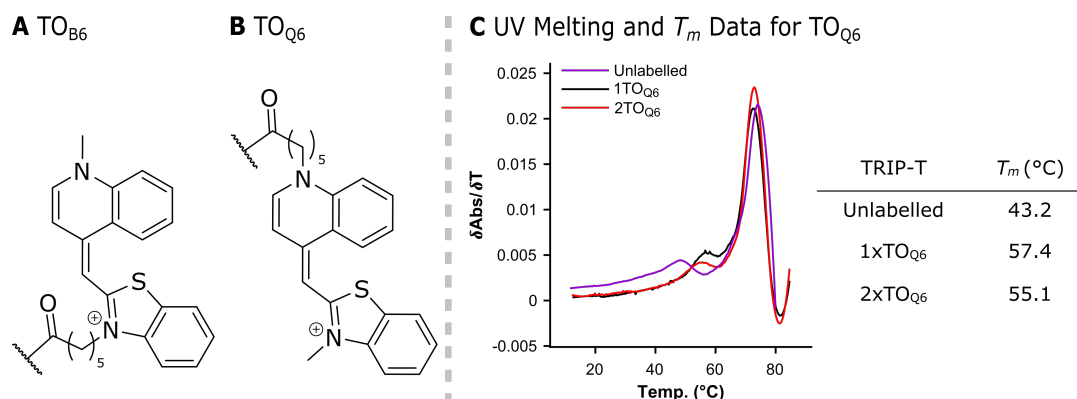


Figure 2.13: Linking TO through both moieties. A) Structure of TO linked through its benzothiazole moiety (TO_{B6}). B) Structure of TO linked through its quinoline moiety (TO_{Q6}). C) Smoothed first derivatives and T_m data determined from the UV absorbance at 260 nm of unlabelled (purple), 1xTO_{Q6} (black) and 2xTO_{Q6} (red) in TRIP-T in high Mg²⁺ buffer at pH 5.8. UV absorbance spectra can be seen in Figure A8.

All T_m data is given in Table 2.1. At pH 5.8, TO has an additive effect on the triplex stabilisation, even in low Mg²⁺ buffer. Three additions of TO in 3xTO TRIP-T resulted in an increase of $\Delta T_m = +44.8$ °C compared to the unlabelled TRIP-T in low Mg²⁺ buffer (entry 4). Similarly, for 3xTO TRIP-C, TO has a stabilising effect but not as large as 3xTO TRIP-T ($\Delta T_m = +26.2$ °C compared to TRIP-C, entry 13). Since TRIP-C contains more protonation sites, the introduction of TO (also positively charged) is not as effective. 3xTO TRIP-TC only saw a small overall increase in stabilisation with three additions of TO ($\Delta T_m = +9.7$ °C compared to TRIP-TC, entry 20).

Entry	Triplex	TFO Sequence (5'—3')	High Mg ²⁺	Low Mg ²⁺
<i>T_m</i> (°C)				
1	TRIP-T	TTT TTM TTT MTM TMT	43.2	31.9
2	1xTO TRIP-T	TTT TTM TTT MTM TMT	66.5	58.0
3	2xTO TRIP-T	TTT TTM TTT MTM TMT	76.4	65.1
4	3xTO TRIP-T	TTT TTM TTT MTM TMT	78.0	76.7
5	TRIP-T _i Control	TTT TTM TTT MTM TMT	49.0	29.6
6	TRIP-T _i	TTT TTM T ^p CT MTM TMT	50.3	31.2
7	1xTO TRIP-T _i	TTT TTM T ^p CT MTM TMT	66.3	48.7
8	2xTO TRIP-T _i	TTT TTM T ^p CT MTM TMT	74.6	66.9
9	3xTO TRIP-T _i	TTT TTM T ^p CT MTM TMT	76.1	73.2
10	TRIP-C	CTM CCM MCC MCC TMT MTC	44.0	35.6
11	1xTO TRIP-C	CTM CCM MCC MCC TMT MTC	50.6	46.2
12	2xTO TRIP-C	CTM CCM MCC MCC TMT MTC	54.3	55.1
13	3xTO TRIP-C	CTM CCM MCC MCC TMT MTC	62.5	61.8
14	TRIP-C _i Control	CTM CCT MCC MCC TMT MTC	31.5	21.0
15	TRIP-C _i	CTM CC ^p C MCC MCC TMT MTC	39.5	26.6
16	1xTO TRIP-C _i	CTM CC ^p C MCC MCC TMT MTC	42.9	29.7
17	2xTO TRIP-C _i	CTM CC ^p C MCC MCC TMT MTC	45.3	36.7
18	3xTO TRIP-C _i	CTM CC ^p C MCC MCC TMT MTC	56.4	44.7
19	TRIP-TC	MTM TMT TMT MTT MTM	77.0	72.7
20	3xTO TRIP-TC	MTM TMT TMT MTT MTM	86.7	76.3

Table 2.1: *T_m* data for all triplexes at pH 5.8. *T_m* values were obtained from the maxima of the triplex transition in first derivative spectra. Values in bold are determined by fluorescence melting. Fluorescence melting was used when the *T_m* could not be determined by UV melting. High Mg²⁺ buffer = 10 mM sodium phosphate, 10 mM MgCl₂ at pH 5.8. Low Mg²⁺ buffer = 10 mM sodium phosphate, 150 mM NaCl, 2 mM MgCl₂ at pH 5.8. *T_m* values are an average of two independent experiments, both of which are an average of 5 measurements. pC = 5-(1-propynyl) dC, M = ^{Me}C, T = TO attached to pdU. Full triplex sequences are given in Figure 2.10. *T_m* of the underlying duplexes are as follows: TRIP-T = 73.1 °C, TRIP-T_i = 73.9 °C, TRIP-TC = 73.2 °C, TRIP-C = 74.6 °C, TRIP-C_i = 77.3 °C.

2.4.1.2 Inversion triplexes

It is obvious that TO can provide strong stabilisation for uninterrupted poly-pu targets, so the focus was then turned to targets that contained base inversions, which are known to be destabilising. The use of the commercially available pC as a CG-

recognition monomer removes the necessity for the multi-step synthesis required for other modified CG-recognition monomers. As a control in each inversion triplex, a T base (in the TFO) was also paired with the CG base pair inversion for comparison (TRIP-X_i control). pC in TRIP-T_i provided slightly increased stabilisation compared to T in TRIP-T_i control, ($\Delta T_m = +1.2$ °C and $+1.6$ °C in high and low Mg²⁺ buffer, respectively, entry 5 and 6). Interestingly, pC in TRIP-C_i provided a larger stabilisation compared to T in TRIP-C_i control, ($\Delta T_m = +8.0$ °C and $+5.6$ °C in high and low Mg²⁺ buffer, respectively, entry 14 and 15). This is most likely due to the electron-withdrawing group (propynyl group) in the pC.CG triplet being stabilised by the positive charge of the C⁺.GC triplets surrounding it compared to pC.CG in TRIP-T_i (neighbouring T.AT triplets).

The destabilising effect of an inversion is evident as we compare TRIP-C (entry 10) with TRIP-C_i which has a $\Delta T_m = -9.0$ °C (entry 15) in low Mg²⁺ buffer. The effect pC has on the stability of the triplex is highlighted by comparing TRIP-C_i control which has a $\Delta T_m = -14.6$ °C (entry 14). The triplet stability is as follows: C⁺.GC (TRIP-C) > pC.CG (TRIP-C_i) > T.CG (TRIP-C_i control). With the incorporation of TO, we see an increase of $\Delta T_m = +18.1$ °C for 3xTO TRIP-C_i (low Mg²⁺ buffer, entry 18) (Figure 2.14 A) but a much more substantial increase of $\Delta T_m = +42.0$ °C for 3xTO TRIP-T_i (entry 9) (Figure 2.14 B). The difference in stability of both triplexes is evident by comparing the triplex transition intensities in Figure 2.14 A and B. This again highlights the effect of sequence environment on the stabilising effect of TO.

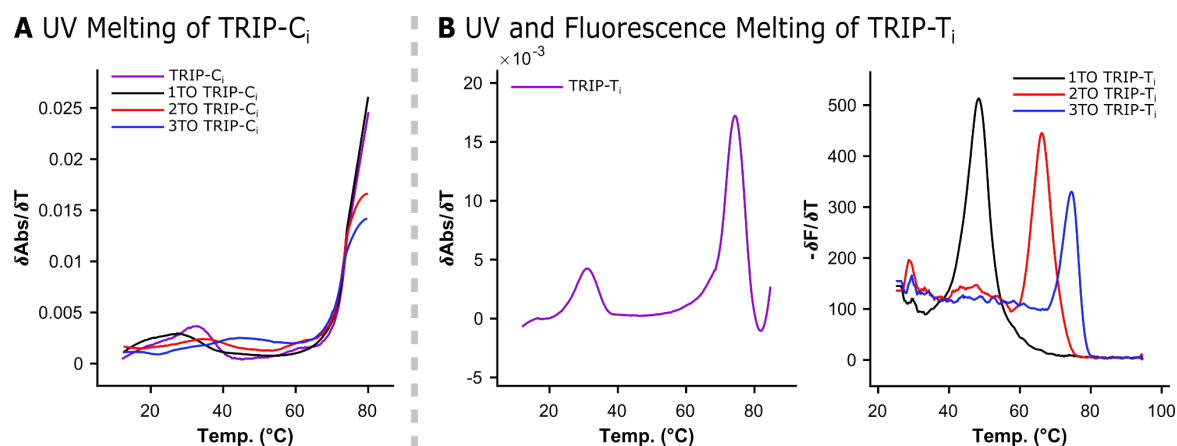


Figure 2.14: Derivatives of melting curves of inversion triplexes at pH 5.8. A) Smoothed first derivatives determined from the UV absorbance at 260 nm of unlabelled, 1xTO, 2xTO and 3xTO TRIP-C_i in low Mg²⁺ buffer. B) Smoothed first derivative determined from the UV absorbance at 260 nm of unlabelled (left) and smoothed first derivative determined from fluorescence spectra of 1xTO, 2xTO and 3xTO TRIP-T_i (right) in low Mg²⁺ buffer. Fluorescence measured at $\lambda_{\text{ex}} = 521$ nm and $\lambda_{\text{em}} = 536$ nm. UV absorbance spectra and fluorescence spectra can be seen in Figure A9.

2.4.2 Triplex fluorescence melting

As seen in Figure 2.14 B (right), fluorescence melting was also used to determine the T_m of the triplex because as the triplexes become more stable, the use of UV monitoring of thermal denaturation becomes less reliable. The transition of triplex melting becomes obscured by the transition for the duplex melting as the T_m of both complexes come closer together (Figure A10). Fluorescence melting analysis can help alleviate this problem as only the triplex melting transition would be detectable; the change in fluorescence from intercalated TO into freely rotating TO is measured. Table A1 emphasises where the differences between the thermal melting and fluorescence melting data for the triplexes occurs. Differences between the T_m results may be due to the difference in DNA concentrations used for both techniques (1 μM for thermal melting and 0.625 μM for fluorescence melting) or because both techniques analyse slightly different mechanisms of melting. Thermal melting is

sensitive to the dissociation of strands *via* the unstacking of the nucleobases whilst fluorescence melting monitors the intercalation of TO into the triplex.

2.4.3 Triplex Surface Plasmon Resonance (SPR)

Surface Plasmon Resonance (SPR) was used as an alternative analytical technique to corroborate the results from thermal and fluorescence melting. SPR can be used to measure real-time binding kinetics of molecular interactions. It uses a thin gold surface film coated with a dextran matrix to which oligonucleotides can be immobilised, using biotin — streptavidin interactions. The change in refractive index can be detected and provides quantitative binding analysis of the interactions. TRIP- T_i was used, and the unmodified triplex produced a K_a of 1.5 μM (Figure 2.15 A and E). This agrees with previous studies in which the triplex equilibrium constants are generally in the μM range.^{130,135,175} 1x, 2x and 3xTO TRIP- T_i were then analysed and the progressive increase in the K_a was measured (Figure 2.15 B, C, D and E). The 3xTO TRIP- T_i system gave binding in the nM range (3.9 nM) (Figure 2.15 D and E). The K_a of the triplex can also be calculated by measuring the T_m of the triplex at varying concentrations and plotting these measurements against each other in a Van't Hoff plot.¹⁷⁶ However, this was not carried out in these studies but it is evident that the relationship between the increasing T_m of the triplexes correlates to the increased binding affinities of the TOTFO probes.

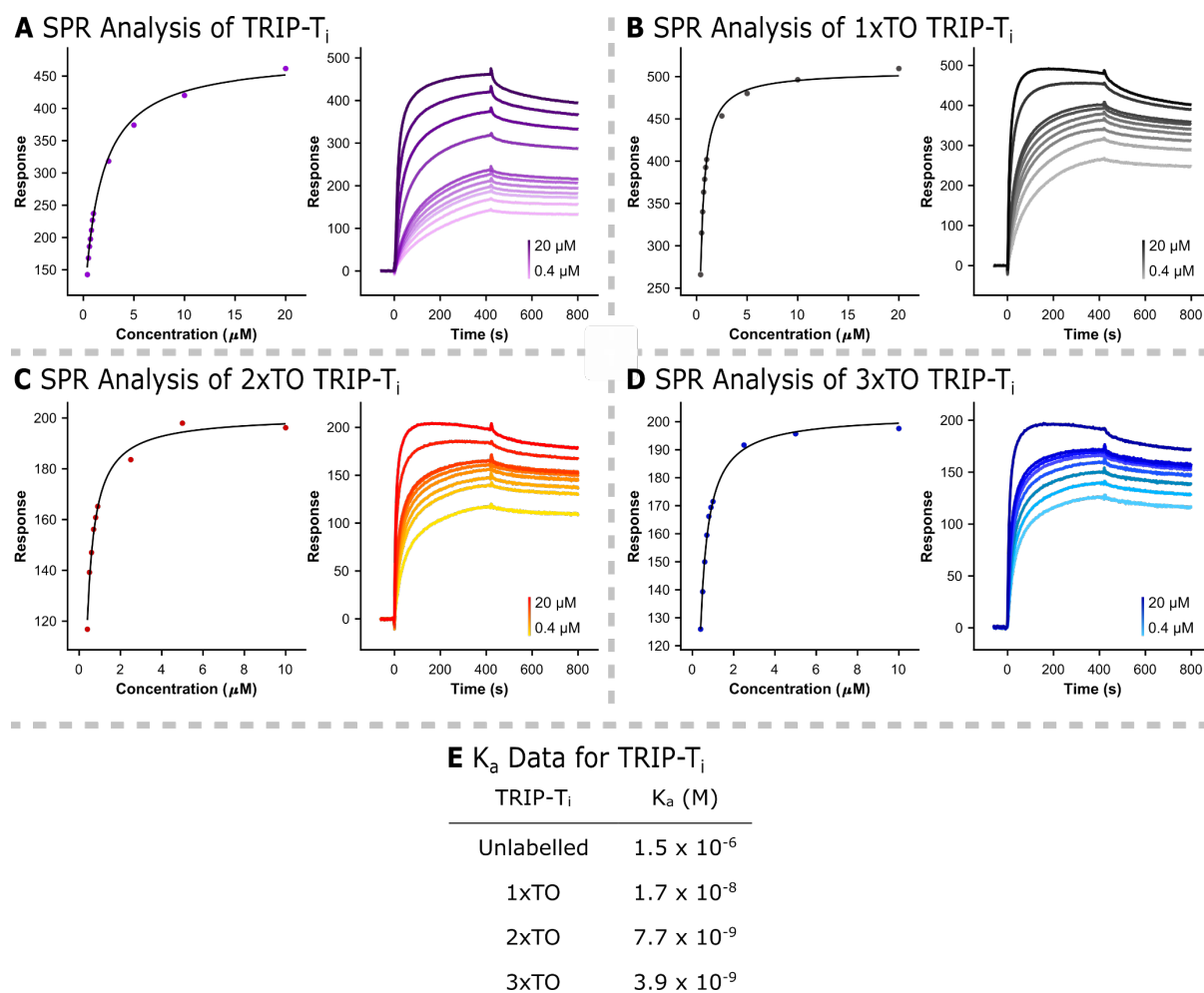


Figure 2.15: SPR analysis of: A) TRIP-T_i B) 1xTO TRIP-T_i C) 2xTO TRIP-T_i D) 3xTO TRIP-T_i. Binding response of triplex plotted against concentration of TFO used (affinity curve) (left) and binding response plotted against time (right). Buffer contains 10 mM MOPS, 10 mM MgCl₂, pH 5.8. The concentrations of injected TFOs ranged from 0.4 μM – 20 μM. E) K_a values were determined using 1:1 steady state equilibrium model of the triplex affinity.

2.4.4 Triplex melting temperatures at pH 7

2.4.4.1 Poly-pu triplexes

The triplexes were also analysed at pH 7. The stabilisation from TO would allow the triplexes to form at higher pH values in all sequence contexts. As expected, the increase in pH causes the T_m of the triplexes to drop, but we still observe a substantial increase in stabilisation with the addition of TO. All T_m data at pH 7 is given in Table 2.2 (note, data at pH 5.8 is also presented again for ease of comparison). 3xTO TRIP-

T provides a $\Delta T_m = +43.6$ °C (entry 4) compared to TRIP-T in low Mg^{2+} buffer. For the C-rich triplexes, the effects of pH on stability are severe. The T_m of TRIP-C (entry 10) is not detectable by thermal melting ($T_m < 18$ °C). This low stability is expected, as protonation is needed to form the $C^+.GC$ triplet. However, with the introduction of 3xTO, a $\Delta T_m \sim +17$ °C was reached in high Mg^{2+} buffer (entry 13).

Entry	Triplex	pH 5.8	Low Mg ²⁺	pH 7.0	Low Mg ²⁺
		High Mg ²⁺		High Mg ²⁺	
<i>T_m</i> (°C)					
1	TRIP-T	43.2	31.9	39.3	28.4
2	1xTO TRIP-T	66.5	58.0	60.0	50.6
3	2xTO TRIP-T	76.4	65.1	65.5	58.7
4	3xTO TRIP-T	78.0	76.7	74.0	72.0
5	TRIP-T _i Control	49.0	29.6	-	-
6	TRIP-T _i	50.3	31.2	23.0	-
7	1xTO TRIP-T _i	66.3	48.7	45.0	35.9
8	2xTO TRIP-T _i	74.6	66.9	57.5	51.3
9	3xTO TRIP-T _i	76.1	73.2	73.4	65.9
10	TRIP-C	44.0	35.6	<i>nt</i>	<i>nt</i>
11	1xTO TRIP-C	50.6	46.2	25.7	19.9
12	2xTO TRIP-C	54.3	55.1	30.4	26.0
13	3xTO TRIP-C	62.5	61.8	34.6	34.8
14	TRIP-C _i Control	31.5	21.0	-	-
15	TRIP-C _i	39.5	26.6	<i>nt</i>	<i>nt</i>
16	1xTO TRIP-C _i	42.9	29.7	<i>nt</i>	<i>nt</i>
17	2xTO TRIP-C _i	45.3	36.7	<i>nt</i>	<i>nt</i>
18	3xTO TRIP-C _i	56.4	44.7	25.4	24.7
19	TRIP-TC	77.0	72.7	52.6	49.0
20	3xTO TRIP-TC	86.7	76.3	73.2	73.4

Table 2.2: T_m data for all triplexes at pH 7. T_m values were obtained from the maxima of the triplex transition in the first derivative curves. Values in bold are determined by fluorescence melting. High Mg^{2+} buffer = 10 mM MOPS, 10 mM $MgCl_2$ at pH 7. Low Mg^{2+} buffer = 10 mM MOPS, 150 mM NaCl, 2 mM $MgCl_2$ at pH 7. T_m of the underlying duplexes are as follows: TRIP-T = 73.4 °C, TRIP-T_i = 74.3 °C, TRIP-C = 75.1 °C, TRIP-C_i = 77.5 °C, TRIP-TC = 74.1 °C.

For 3xTO TRIP-TC at pH 5.8, the increase in stability was not as high as the other triplexes but at pH 7, this increase becomes comparable to the others ($\Delta T_m = +24.5$ °C in low Mg^{2+} buffer, entry 20). This suggests that the stability of the underlying duplex contributes to the stability of the triplex. TO could only stabilise the already highly stable TRIP-TC at pH 5.8 by a small amount but at pH 7, where TRIP-TC became more unstable, the window for stabilisation expanded. The T_m of the duplex alone is 74.1 °C which potentially sets a limit on the possible stability of the triplex. The duplex contribution to stability of the triplex has previously been studied by others.⁷⁹

2.4.4.2 Inversion triplexes

The destabilising effect of the inversion is much more prominent at pH 7 as we compare TRIP-T (entry 1) to TRIP-T_i ($\Delta T_m = -16.3$ °C, entry 6) in high Mg^{2+} buffer. Regardless, TO is still capable of stabilising the inversion and this is strongly evident in TRIP-T_i (Figure 2.16) and TRIP-C_i. The T_m of TRIP-C_i, 1xTO and 2xTO are not detectable by thermal melting (entry 15, 16 and 17) but 3xTO TRIP-C_i forms a triplex with a $\Delta T_m \sim +8$ °C (entry 18). Even the most unstable triplexes can benefit from TO incorporation. In some sequence contexts, only one addition of TO is needed at pH 7 to form a stable triplex (1xTO TRIP-T_i, $T_m = 45.0$ °C, entry 7, Figure 2.16). These results show that TO intercalation combined with pC as a CG-recognition monomer compensates for the destabilisation of an inversion in the target, even at pH 7. This expands the triplex target range and makes TOTFOs more versatile. Stable triplexes

can even be formed above neutral pH, at pH 8, where 3xTO TRIP-T has a $T_m = 42.6$ °C in low Mg^{2+} buffer.

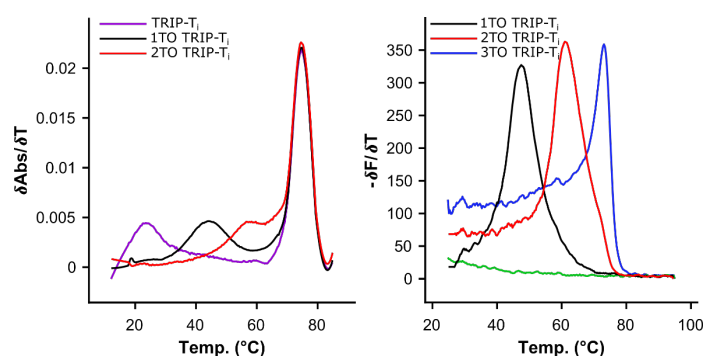


Figure 2.16: Derivatives of melting curves of TRIP-T_i at pH 7. Smoothed first derivatives determined from the UV absorbance at 260 nm of unlabelled, 1xTO, 2xTO TRIP-T_i (left) and smoothed first derivatives determined from fluorescence melting curves of 1xTO, 2xTO and 3xTO TRIP-T_i (right) in high Mg^{2+} buffer. Fluorescence measured at $\lambda_{ex} = 521$ nm and $\lambda_{em} = 536$ nm. UV absorbance spectra and fluorescence spectra can be seen in Figure A11.

2.4.4.3 Control triplexes

To confirm the sequence specificity of the TOTFOs, some controls were also investigated. As the human genome contains approximately 3×10^9 bases, selectivity is crucial to avoid off-target binding. A scrambled duplex (a duplex not complementary to the TFO) was used as a target with 1xTO TFO-T in low Mg^{2+} buffer at pH 7. With the correct target duplex, this triplex has a $T_m = 50.6$ °C. However, with the scrambled duplex, triplex formation was not detectable which confirms the necessity for efficient sequence-specific hybridisation of the TFO to the duplex target before TO stabilisation can be seen (Figure 2.17 A). To further investigate the base pair selectivity of the TOTFOs, 3xTO TFO-T was mixed with a duplex (low Mg^{2+} buffer, pH 7) containing a triplet mismatch to produce T.CG, T.TA and T.GC triplets and compared to the correct base pair triplet, T.AT. As expected, the mismatch causes a

decrease of $\Delta T_m = -9.8$ °C (T.CG), -10.0 °C (T.TA) and -13.8 °C (T.GC) even with 3 additions of TO (Figure 2.17 B). This provides evidence that TO intercalation does not destroy TFO selectivity even on a single triplet mismatch level.

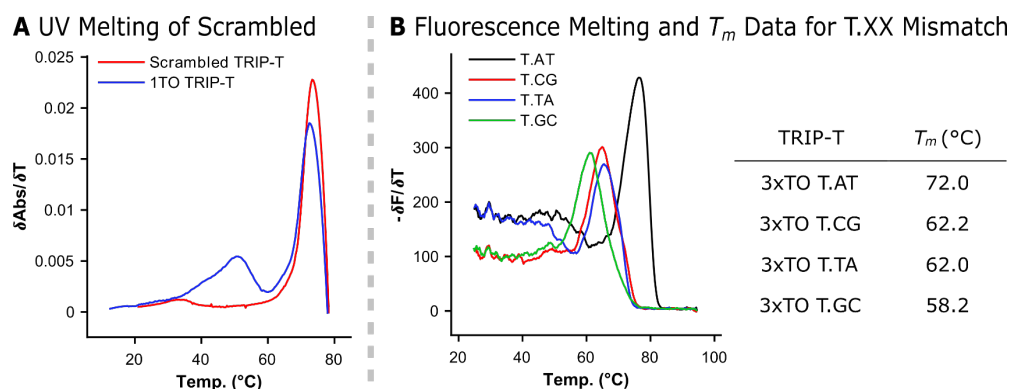


Figure 2.17: Derivative of melting curves of controls at pH 7. A) Smoothed first derivatives determined from the UV absorbance at 260 nm of TFO-T mixed with a scrambled duplex (red) and the correct target duplex (blue) in low Mg^{2+} buffer, pH 7. B) Smoothed first derivative and T_m data determined from fluorescence melting curves of 3xTO TFO-T mixed with duplexes containing triplet mismatches to produce T.AT, T.CG, T.TA and T.GC triplets in low Mg^{2+} buffer, pH 7. Fluorescence measured at $\lambda_{ex} = 521$ nm and $\lambda_{em} = 536$ nm. UV absorbance spectra and fluorescence spectra can be seen in Figure A12.

For the inversion triplexes, the selectivity of pC was investigated by mixing 1xTO TFO- T_i with duplexes containing a triplet mismatch to produce pC.GC, pC.AT and pC.TA triplets and compared to pC.CG at pH 5.8 (Figure 2.18 A) and pH 7 (Figure 2.18 B). The pC.AT triplet mismatch ($\Delta T_m = -12.3$ °C at pH 7 and $\Delta T_m = -9.4$ °C at pH 5.8) provided a less stable triplex compared to pC.CG which highlights the superiority in selectivity of pC over T as a CG-recognition monomer as T could also form a triplet with an AT base pair. pC.TA ($\Delta T_m = -10.5$ °C at pH 7) was also less stable compared to pC.CG. However, at pH 5.8, pC.GC ($\Delta T_m = +6.3$ °C at pH 5.8) was more stable compared to pC.CG. pC.GC could form a triplet with a single H-bond but this stability is pH dependent. At higher pH where protonation becomes more crucial for the formation

of a X.GC triplet, we see this stability diminishes. pC.CG is far less pH dependent ($\Delta T_m = -18.4$ °C, pH 5.8 vs pH 7) compared to the pC.GC triplet ($\Delta T_m = -24.6$ °C, pH 5.8 vs pH 7). Overall, pC is more selective at higher pH which is advantageous for the practical use of TOTFOs.

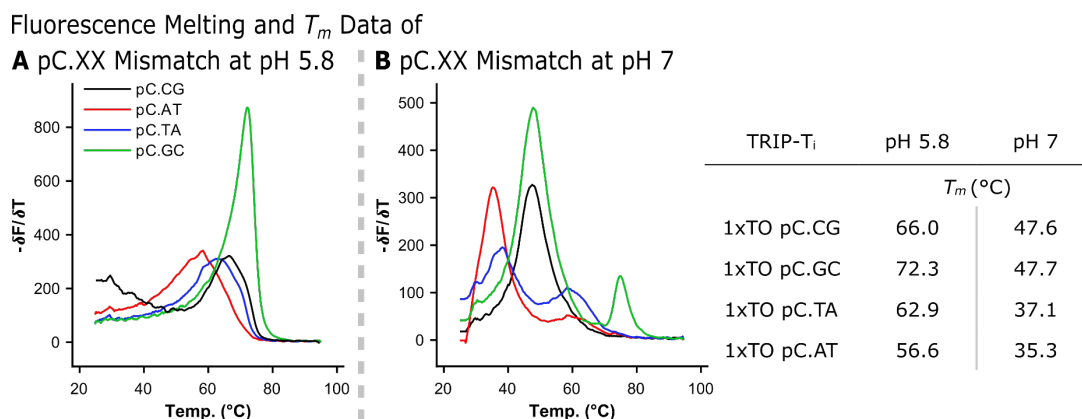


Figure 2.18: Derivative of melting curves of pC.XX mismatches at pH 5.8 and pH 7. Smoothed first derivative determined from fluorescence melting curves of 1xTO TFO-T_i mixed with duplexes containing triplet mismatches to produce pC.CG, pC.GC, pC.TA and pC.AT triplets in high Mg²⁺ buffer at A) pH 5.8 and B) pH 7. Fluorescence measured at $\lambda_{ex} = 521$ nm and $\lambda_{em} = 536$ nm. Fluorescence melting curves can be seen in Figure A13.

2.4.5 Triplex steady state fluorescence and quantum yield

To quantify the enhancement of fluorescence once triplex formation occurred, steady state fluorescence measurements were taken. The fluorescence of the TFO, in the absence and then in the presence of the target duplex, was measured and the intensities compared to determine the fold-increase of fluorescence between free TFO and triplex states. In general, TFOs, excited at 484 nm, showed reduced fluorescence intensities and with the introduction of the duplex (with sufficient time allowed for complex formation), an increase in the fluorescence intensity was observed. All fluorescence intensity data and ratios are given in Table 2.3.

Triplex	pH = 5.8			pH = 7.0		
	$F^{ss} @ \lambda_{em}^T$	$F^T @ \lambda_{em}^T$	F^T/F^{ss}	$F^{ss} @ \lambda_{em}^T$	$F^T @ \lambda_{em}^T$	F^T/F^{ss}
	a.u.					
1xTO TRIP-T	10.7	249	23.2	7.5	197	26.3
2xTO TRIP-T	72.2	606	8.4	52.4	524	10.0
3xTO TRIP-T	120	579	4.8	67.5	548	8.1
1xTO TRIP-T _i	87.7	705	8.0	51.5	643	12.5
2xTO TRIP-T _i	52.1	348	6.7	29.5	321	10.9
1xTO TRIP-C	62.9	485	7.7	100	668	6.7
2xTO TRIP-C	32.9	640	19.5	38.4	875	22.8
3xTO TRIP-C	18.7	468	25.1	19.9	711	35.7
1xTO TRIP-C _i	98.2	463	4.7	—	—	—
2xTO TRIP-C _i	80.8	400	5.0	—	—	—
3xTO TRIP-C _i	100	289	2.9	52.2	315	6.0

Table 2.3: Fluorescence intensity measurement of TOTFOs. Buffer = 10 mM sodium phosphate, 10 mM MgCl₂ at pH 5.8 and 10 mM MOPS, 10 mM MgCl₂ at pH 7. Fluorescence values are averages of two independent measurements. $F^{ss} @ \lambda_{em}^T$ = fluorescence intensity of TFO at the maximum emission wavelength of formed triplex, $F^T @ \lambda_{em}^T$ = fluorescence intensity of triplex at the maximum emission wavelength of formed triplex, F^T/F^{ss} = ratio of fluorescence intensity calculated by comparison of the TFO fluorescence intensity and the formed triplex fluorescence intensity at λ_{em}^T .

A fluorescence enhancement of 23-fold from TFO to triplex is evident for 1xTO TRIP-T at pH 5.8 (Figure 2.19 A). At pH 7, this increases to a 26-fold increase in intensity (Figure 2.19 B). The rest of the triplexes follow the same trend in that the fluorescence intensity ratios are generally higher at pH 7 compared to pH 5.8. With the second and third addition of TO, the fold-increase is reduced although the 2xTO TRIP-T and 3xTO TRIP-T fluorescence intensities are brighter. This is due to the single-stranded TFO containing multiple TO units. The ratios were determined using the fluorescence intensity at λ_{max} of the formed triplex. The area under the fluorescence intensity curve (Table A2) was also used to calculate the ratios, and these follow similar trends as those seen in Table 2.3. Quantum Yield (QY) analysis at pH 7 was also carried out for

TRIP-T. The QY ranged from 0.3 – 0.5% for the free TFOs and 5.4 – 23% for the triplexes, which is as expected. QY values and the ratios are given in Figure 2.19 C.

Fluorescence Intensities of

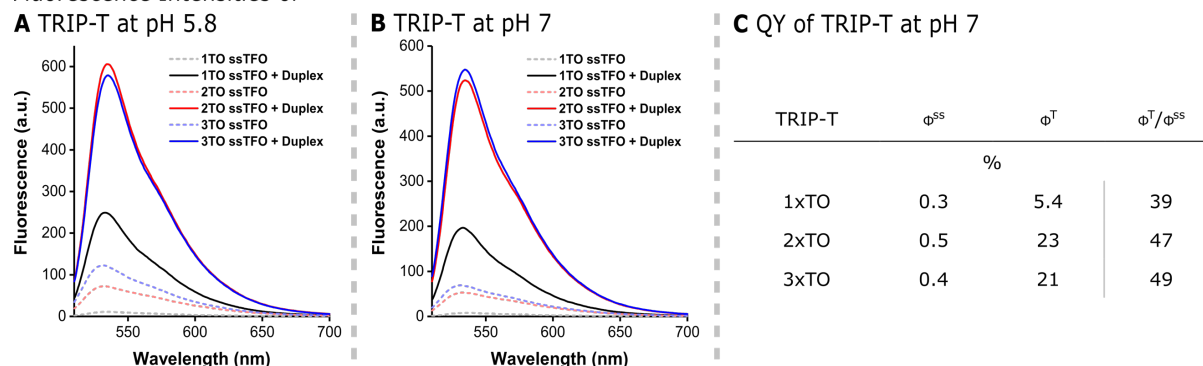


Figure 2.19: Fluorescence intensities and QY measurements of TRIP-T. Fluorescence intensity of 1xTO, 2xTO and 3xTO TRIP-T in high Mg^{2+} buffer at A) pH 5.8 and B) pH 7. 1xTO, 2xTO and 3xTO TFO-T measured in the absence (dashed line) and presence (solid line) of the duplex. C) QY data of 1xTO, 2xTO and 3xTO TRIP-T determined using an integrated sphere. Values are an average of two independent measurements. Buffer = 10 mM MOPS, 10 mM $MgCl_2$ at pH 7. Φ^{ss} = quantum yield of single-stranded TFO, Φ^T = quantum yield of TFO and duplex, Φ^T/Φ^{ss} = quantum yield upon triplex formation.

2.4.6 Triplex PAGE at pH 7

Another biophysical technique used to evaluate the triplexes was native polyacrylamide gel-electrophoresis (PAGE). The triplex is separated out on the gel based on charge and size and the bands can be visualised using both TO fluorescence and SYBR Gold staining. In Figure 2.20 A, Gel A shows two bands formed in the lanes containing 1x, 2x and 3xTO TRIP-T_i (construct 2, 3 and 4, respectively). The inherent effects of TO stabilisation are evident by comparing construct 1 (unlabelled TRIP-T_i) and construct 2. Since the gel is run at pH 7, construct 1 ($T_m = 23.0$ °C) will not form a triplex but with construct 2, 3 and 4, all containing TO, the formation of a triplex band is seen (the slower moving band).

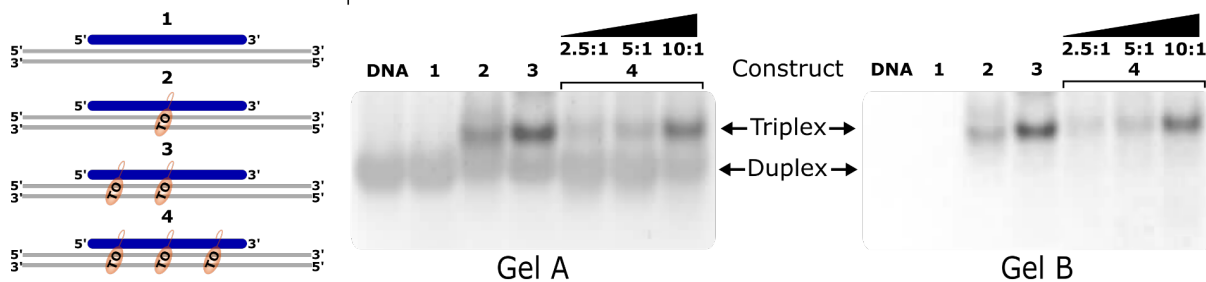
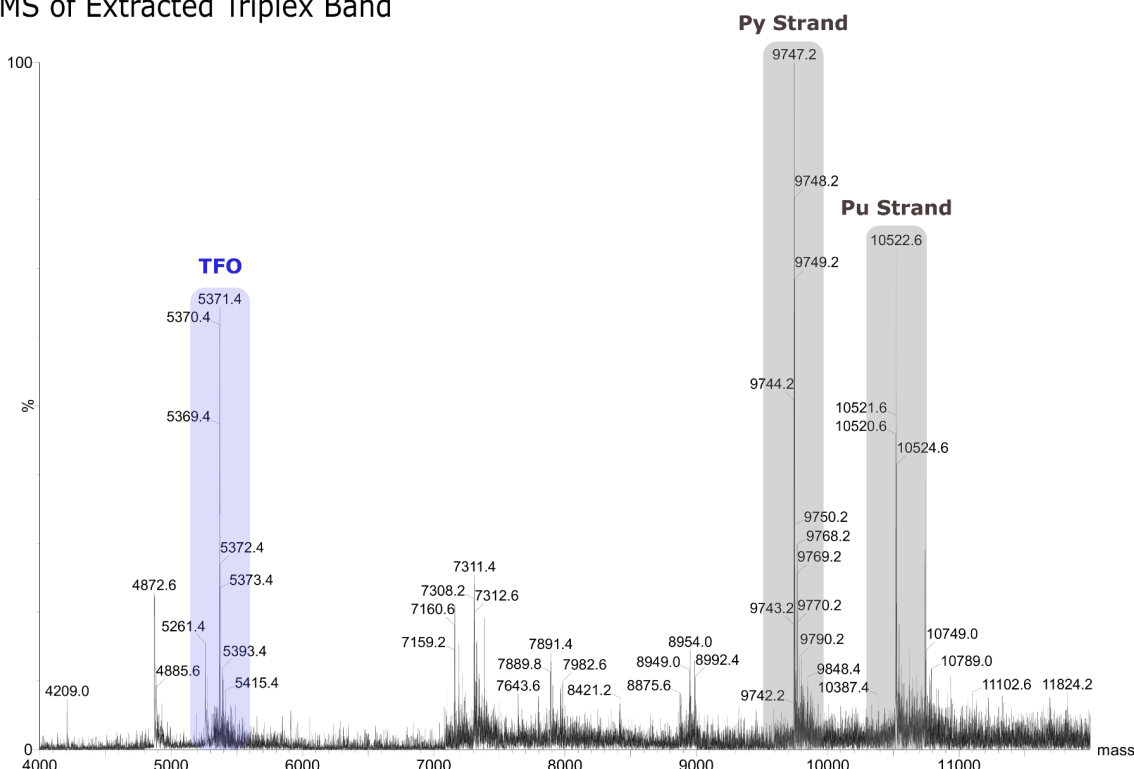
A Native PAGE of TRIP- T_i **B** MS of Extracted Triplex Band

Figure 2.20: Native PAGE analysis of triplex formation. A) 15% non-denaturing PAGE showing TRIP- T_i triplex formation. Buffer = 40 mM MOPS containing 10 mM NaOAc and 10 mM $MgCl_2$ at pH 7. Construct 1 = TRIP- T_i , construct 2 = 1xTO TRIP- T_i , construct 3 = 2xTO TRIP- T_i , construct 4 = 3xTO TRIP- T_i . Gel A captured using TO filter and Gel B stained using SYBR Gold. DNA lane contains target duplex for TFO- T_i . [10:1] TFO:duplex ratio used for all constructs. Additional [2.5:1] and [5:1] ratio used for construct 4. B) MS of gel-extracted 2xTO TRIP- T_i triplex band (construct 3). Expected mass of purine strand of duplex = 10522, actual mass = 10522.6. Expected mass of pyrimidine strand of duplex = 9746, actual mass = 9747.2 and expected mass of 2xTO TFO- T_i = 5372, actual mass = 5371.4.

An excess of TFO (10:1 TFO:duplex) was used for constructs 1-3 to ensure triplex formation. In construct 4, increasing the ratio of the TFO:duplex (2.5:1, 5:1 and 10:1) resulted in an increase in the intensity of the top band, which provides evidence that this is the triplex band. In the TO fluorescence channel (Figure 2.20 A, Gel B), only

the top band is visible indicating further that this band contains the triplex. To confirm this, this band for construct 3 was extracted from the gel and analysed by UPLC-MS. Analysis of the peaks indicated three peaks that corresponded to the masses of the TFO and both duplex strands (Figure 2.20 B).

2.5 Conclusion and Future Work

The aim of this work is to enhance the stability of triplexes by incorporation of an intercalating fluorophore thiazole orange into the TFO in multiple sequence contexts and to expand the target range of TFOs to include base pair inversions. In conclusion, triplex stability is greatly enhanced by incorporation of TO attached to pdU through the benzothiazole moiety of TO in triplex-forming oligonucleotides (TOTFOs) in all sequence contexts. This straightforward post-synthetic labelling method can provide very stable triplexes at (and above) neutral pH. The stabilising effects of TO are additive, with the most stable triplex at neutral pH being 3x TO TRIP-T whose stabilisation was evidenced by a $\Delta T_m = +45$ °C (compared to TRIP-T). Importantly, there is no major disruption to the selectivity of the TOTFOs which is evident from the probes capability to discriminate triplet mismatches.

This stability seen with TOTFOs is supported by a large enhancement in fluorescence upon triplex formation, TRIP-T has a 26-fold increase in fluorescence at pH 7. SPR analysis and native gel corroborated the results seen from thermal melting and fluorescence. 3xTO TRIP-T_i has a K_a of 3.9 nM (compared to a K_a of TRIP-T_i = 1.5 μ M) so it can be deduced that the substantial increase in T_m relates to the increased binding affinity of the triplex. Triplex formation was not detected in PAGE without the incorporation of TO with TRIP-T_i.

Previous studies also encounter stabilisation from TO tethered to the TFO, but this was not as substantial as the stabilisation evident in the TOTFO. In the work carried out by Asseline *et al.*, the LU-TFOs did not pertain enough stabilisation to form triplexes at neutral pH¹⁷¹ while this was not an issue for TRIP-T_i with just one incorporation of TO. Previous work was also limited by a selective target range, but we have shown TOTFOs can be effective in all sequence contexts without reduced selectivity. The ECHO probes¹⁷² previously discussed provided better fluorescence enhancements than the TOTFOs but again, this was within a very limited target range. TOTFOs provide an efficient and straight-forward system for the stabilisation of triplexes without reduced selectivity or limitations on the target range.

Some developments that could enhance the system even further would be the introduction of a second fluorophore along with TO at the modified site of the TFO. This would provide a system that was quenched substantially more when not bound to the triplex due to collisional quenching between the two dyes. Then, once triplex formation occurs and TO intercalates into the triplex, the second fluorophore would provide a greater increase in fluorescence and quantum yield. Once the system was optimised, the probes could be developed for detection of DNA in cells. The sequence-specificity of the TOTFO probe would be an advantage against the large genome target and the enhanced fluorescence would allow for visualisation.

Chapter 3 Live Cell Imaging using CRISPR-dCas9 and Fluorogenic gRNA

3.1 Background

We have seen in Chapter 2 that fluorescence probes provide great advances in detection of dsDNA. The organisation of DNA within an organism is a fundamental element to its functionality. Therefore, it is crucial to understand the spatiotemporal organisation of the genome, which plays an important role in the development of cells, differentiation and transcription factors.¹⁷⁷ Genomic loci mis-localisation or defects in the global organisation of the genome is commonly associated with disease^{178,179} and so, much research has been developed with the aim of understanding genomic dynamics and requires direct visualisation of the chromosomal loci *in situ*.

3.1.1 Fluorescence imaging of the genome

Non-fluorescent methods such as autoradiography^{180,181} and chromosome conformation capture (3C)¹⁸² help to provide an understanding of the genome but fluorescence imaging has been at the forefront of visualisation with just some of the many examples described herein. Over the years, chromatin in bacteria¹⁸³ to mammals¹⁸⁴ have been imaged using fluorescence imaging techniques and have opened the understanding of cell dynamics. These methods rely on illuminating the chromatin with fluorescence markers.

3.1.1.1 Fluorescently modified deoxynucleosides

One example of these methods is fluorescently modified deoxynucleosides. This technique incorporates modified 5-ethylene-2'-deoxyuridine (EdU) into dividing cells

and uses a copper-catalysed cycloaddition reaction between an azide-modified dye and the alkyne modified-EdU to stain the cells.¹⁸⁵ The stain highlights the incorporation of EdU into DNA in the cell. However, this method does not provide sequence specificity and works best in ‘fixed cells’ – the cell activity is stopped but the overall cell structures remains paused in place, including DNA and protein structures.

3.1.1.2 Fluorescence In Situ Hybridisation (FISH)

DNA Fluorescence *In Situ* Hybridisation (DNA FISH) is a sequence-specific technique that uses fluorescent oligonucleotide probes that are complementary to the target sequence (Figure 3.1 A).^{186,187} This technique also requires the cell to be ‘fixed’. The genomic DNA is then denatured partially, the probes can bind to the specific target and any unbound probes can be washed away. While DNA FISH has provided insights into genomic studies and is continuously being improved,^{188–190} it has some limitations.¹⁹¹ Fixed cells remove the ability for real-time dynamic studies and the harsh conditions (formamide and heat) run the risk of destroying the structural integrity of the cell structures, questioning the technique’s reliability.

3.1.2 Live cell imaging

Whilst DNA FISH techniques provide fixed spatial distances within cells, a need for a more versatile technique that incorporates live cell dynamics is required. Some methods used chromatin binding proteins to image chromosomes by tagging them

with fluorescent proteins (FP).^{179,192–195} However, chromatin binding proteins tend not to have any sequence-specificity.¹⁹¹

Other techniques with a similar approach include modules fused to fluorescent proteins such as zinc fingers (ZF) (Figure 3.1 C).¹⁹⁶ These artificial DNA-binding proteins allow for a more sequence specific approach. ZF recognise specific trinucleotides and so, by fusing ZF together they can be used to recognise longer sequences.¹⁹⁷ However, all possible trinucleotide combinations are not recognised with ZF, limiting their applications.

Transcription activator-like effectors (TALEs) (Figure 3.1 D) provided the next generation of DNA binding proteins. TALEs are an amino-acid chain that specifically bind to DNA sequences in living cells^{198,199} and provided a basis for dynamic imaging in live cells. However, the building-block assembly needed for these techniques provides a limitation for detection of varying sequences. With the discovery of bacterial clustered regularly interspaced short palindromic repeats (CRISPR) gene and its Cas (CRISPR-associated) proteins, a solution of the use a single protein, for both repetitive and non-repetitive sequences, became available.

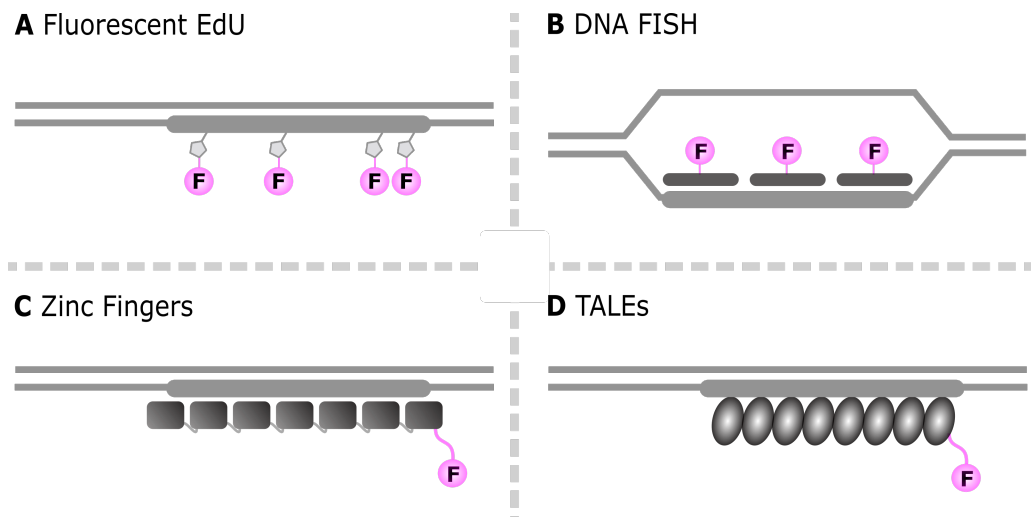


Figure 3.1: Schematics of live cell imaging techniques. A) Fluorescently stained DNA that contains incorporated EdU B) DNA Fluorescence *In Situ* Hybridisation probes bound to DNA in fixed cells C) Zinc Finger Proteins fused with a fluorophore that specifically bind to DNA sequences D) Transcription activator-like effectors fused with a fluorophore that specifically bind to DNA sequences.

3.1.3 CRISPR-Cas9

3.1.3.1 Discovery of CRISPR-Cas systems

CRISPR-Cas systems are derived from bacterial organisms that have developed them as a viral immune response.^{200–202} The first reports of CRISPR were made in the 1980s by Nakata *et al.*²⁰³ and from there, immense development of the system eventually led to its use in human cells in 2013.^{204–206} From there, a game-changing class of genome-editing proteins was born.^{206–210} It is a site-specific protein with nuclease activity that binds to a single-guide RNA (sgRNA), made up of CRISPR RNA (crRNA) and transactivating CRISPR RNA (tracrRNA). The protein can specifically cut dsDNA complementary to the crRNA segment of the sgRNA (Figure 3.2). Its development has led to extensive biophysical, structural and biochemical studies of the system.^{201,202,211–216} These studies will only focus on *Streptococcus pyogenes* Cas9 protein which is classified as a type II protein in class 2.²⁰⁷

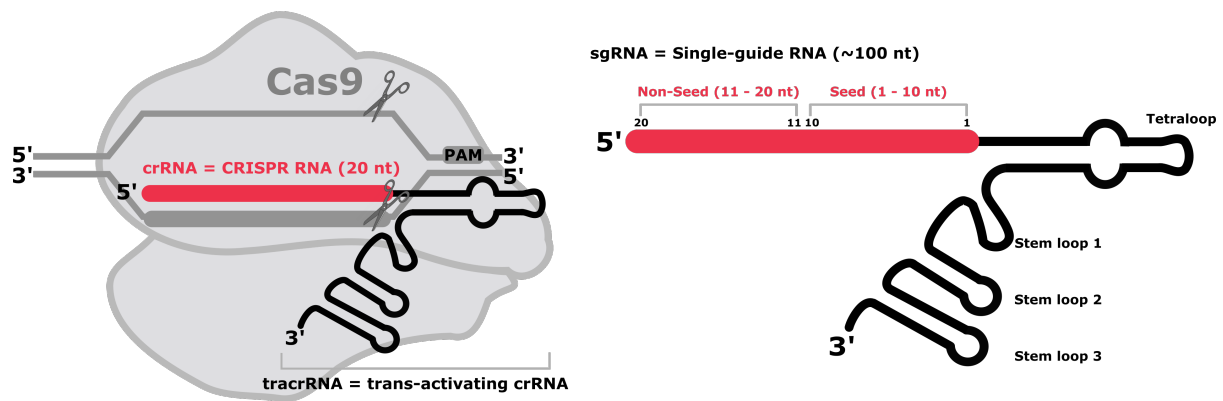


Figure 3.2: Schematics of CRISPR-Cas9. CRISPR-Cas9 complexed with sgRNA and opened dsDNA target (left). Cas9 cleaves a blunt-end double-stranded break in the target DNA between the 3rd and 4th base adjacent to the PAM region. crRNA = 20 nt section of sgRNA used to guide the Cas9 to its target, tracrRNA = secondary-structured segment of sgRNA that binds to Cas9 protein. More detailed schematic of sgRNA and its components (right).

Cas9 cleavage of DNA is a multi-step process.²⁰¹ 1) Cas9 protein binds to its sgRNA; crystal structures of the protein showed a disordered PAM (Protospacer-Adjacent Motif) recognition domain within the protein when sgRNA is not bound which inhibits non-specific DNA cleavage.^{214,215} 2) The ribonucleoprotein recognises the PAM region of the DNA target; the now-ordered PAM recognition domain will search for a 5'-NGG-3' motif in the non-target strand of the dsDNA target. DNA that does not contain this PAM region (even if it is complementary to the crRNA region) will disassociate immediately.²¹⁶ 3) Strand invasion and sgRNA-DNA duplex formation; following recognition of the PAM site, the DNA duplex is unwound by the protein which allows the sgRNA to strand invade and hybridise with the target strand of DNA. This forms a RNA-DNA heteroduplex which takes an A-form structure.²¹¹ It is the A-form structure of the RNA-DNA duplex that Cas9 recognises in the complex rather than its sequence, allowing the sgRNA sequence to be adjusted according to its needs. 4) Cleavage of both DNA strands; once the DNA is bound, the nuclease domain of the

protein aligns both strands into cleavage positions and cuts both strands simultaneously between the 3rd and 4th base from the PAM region (Figure 3.2). The HNH nuclease domain of the enzyme, responsible for cleavage of the target strand of the DNA, is suggested to employ a one-metal ion cleavage mechanism whereas the RuvC nuclease domain, responsible for the cleavage of the non-target strand, is likely to use a two-metal-ion mechanism. Both domains use a histidine and aspartate residue to carry out nucleic acid cleavage.^{201,217} This forms a blunt-end double-stranded break (DSB).^{214,218} Kinetic studies of Cas9 shows the release of the cleaved DNA product is extremely slow, essentially making it a single-turnover enzyme.²¹⁹

The cell machinery decides on whether this DSB will be repaired by non-homologous end joining (NHEJ) or homology-directed repair (HDR), two competing DNA repair pathways. NHEJ is an error-prone repair pathway that ligates both ends of the DNA without a homologous DNA template and the repair errors usually results in loss of gene function. HDR can introduce donor DNA into the damaged DNA or repair the cleavage to the correct sequence.²²⁰ With such a robust and easy-to-use system, it was not long before the Cas9 protein was used for genomic imaging.

3.1.3.2 Fluorescence imaging with CRISPR-dCas systems

Studies of the cleavage mechanisms of the Cas9 protein identified that by mutating the amino acids responsible for the cleavage in both the HNH nuclease domain and the RuvC nuclease domain, a catalytically deactivated version of the protein (dCas9)

resulted. The dCas9 enzyme could still bind to sgRNA and target DNA without causing cleavage.^{201,221,222} This was advantageous and necessary for imaging as nuclease activity would interfere with the cell's natural behaviour. The protein was then fused with a green fluorescent protein (dCas9-GFP) as well as a nuclear localisation sequence (NLS) and used sgRNA to target specific loci of the genome in human cells, both repetitive and non-repetitive regions, for the first imaging studies using dCas9.²²³ Repetitive regions of the genome act as an ideal target in imaging as it means the use of one sgRNA can provide multiple fluorescent moieties that concentrate to one locus and provide an intense fluorescent punctum. However, for non-repetitive regions, more elaborate techniques are required.

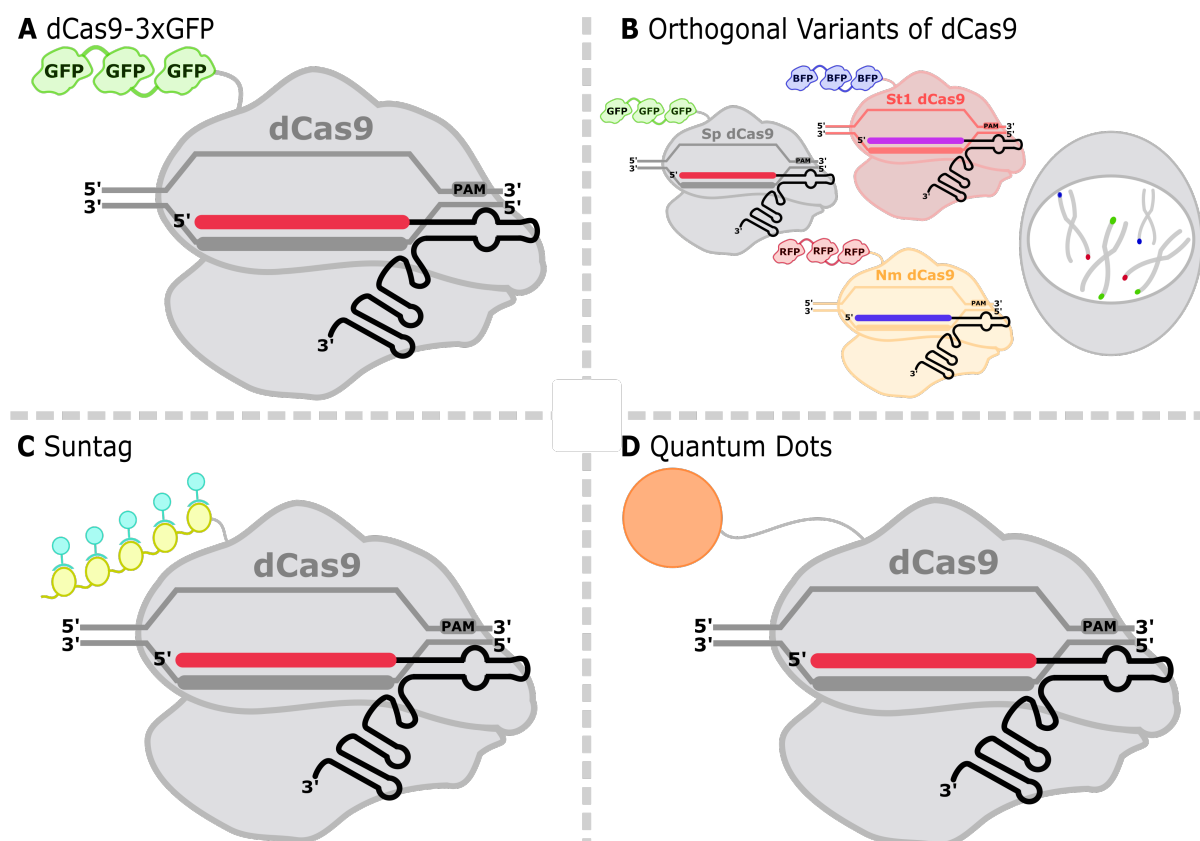


Figure 3.3: dCas9 modifications for imaging. A) dCas9 fused with 3xGFP B) Multi-colour labelling of genome using orthogonal dCas9 species fused to fluorescent proteins C) Suntag, a repeating peptide array fused to dCas9 to recruit fluorescent antibodies D) Quantum Dots fused to dCas9.

Further development of dCas9-GFP sought to increase the fluorescence intensity by fusing 3xGFP to the protein (Figure 3.3 A). Multi-colour imaging of different loci in the same cell was also established by fusing other FPs, such as red or blue, to orthogonal variants of dCas9 from different bacterial species (Figure 3.3 B).²²⁴ Nonetheless, this method is limited to the number of bacterial dCas9 variants. It is important for efficient imaging to achieve maximum target signal while minimising background noise and is estimated that 150–200 FP per genome loci are needed to provide detectable puncta (50–70 for dCas9-3xGFP).²²⁴ With this in mind, another approach focused on fusing a repeating peptide array (SunTag) that could recruit GFP-tagged antibodies to the dCas9 protein (Figure 3.3 C).²²⁵ However, this method was later shown to exhibit high background fluorescence.²²⁶ A solution to this was the bimolecular fluorescence complementation (BiFC) assay which split Venus, a FP, and fused the subunits to dCas9 and sgRNA. This means fluorescence can only occur once the two split segments were close in proximity. The versatility of this technique is limited by the need of a FP that can be split, however.²²⁶

Quantum dots (QDs) provided promise as they are luminescent nanoparticles that are much brighter and much more photostable than FP. In a recent study, QDs were conjugated to dCas9 proteins using a lipoic acid ligase-based method or a biotin-streptavidin based method in the nucleus of the cell and were capable of imaging Human Immunodeficiency Virus (HIV)-1 proviral DNA (Figure 3.3 D).²²⁷ However,

QDs are limited by problems such as cell delivery due to their large size and are also partial to aggregates and entrapment in lysosomes and endosomes.²²⁸

3.1.3.3 Fluorescence imaging with modified sgRNA

While modified dCas9-based methods have provided much insight into genomic imaging, alternative research focused on modifications to the sgRNA. To modify the sgRNA, it is important to maintain the sgRNA-dCas9 binding activity, which will be discussed further in Section 3.4.1 – 3.4.2. Previous studies on modified sgRNA showed the tetraloop (Figure 3.2) can be extensively altered without losing dCas9 activity.²²⁹

Modifying the sgRNA to include an aptamer sequence became a popular method. CRISPRainbow incorporates MS2, PP7 and boxB aptamers in various combinations to the stem loops or 3'-end of the sgRNA.²³⁰ This then recruits combinations of FP fused to RNA hairpin binding domains and the resultant array of FP provides a multi-colour spectrum for genomic imaging not limited to just blue, green and red (Figure 3.4 A). Another aptamer based-method includes extending the sgRNA to include a “Broccoli” aptamer²³¹ which becomes fluorescent upon binding to a specific small molecule that is added to the medium (Figure 3.4 B).²³² Other techniques, such as CRISPR-Sirius,²³³ used aptamers to enhance fluorescence by introducing multiple aptamer sites (Figure 3.4 C). This multiple-aptamer technique was also used in other studies.^{234–236}

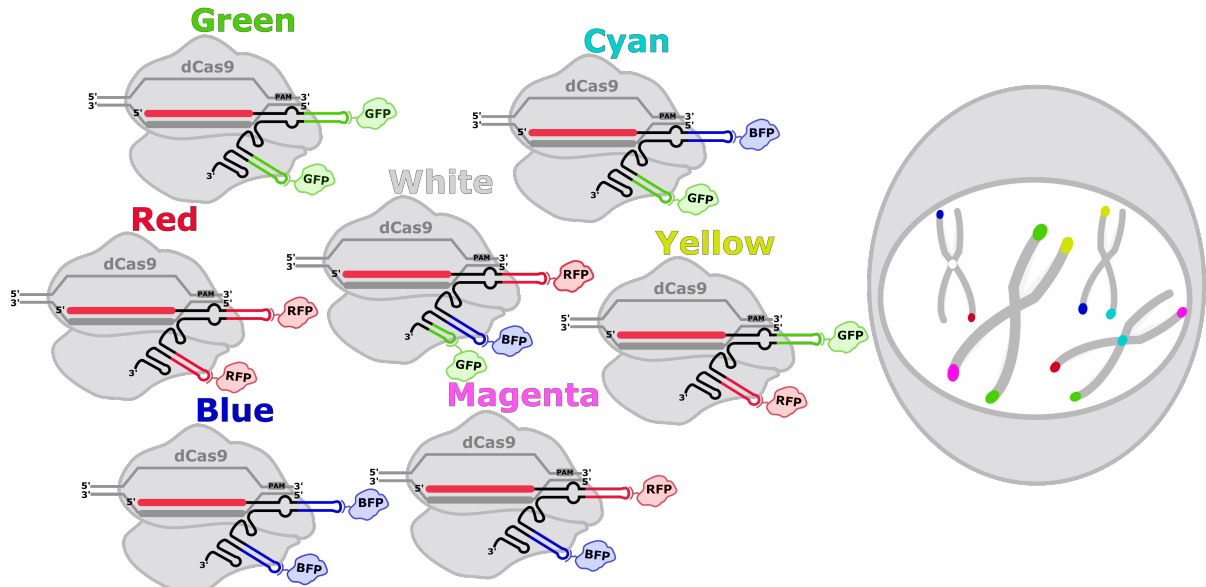
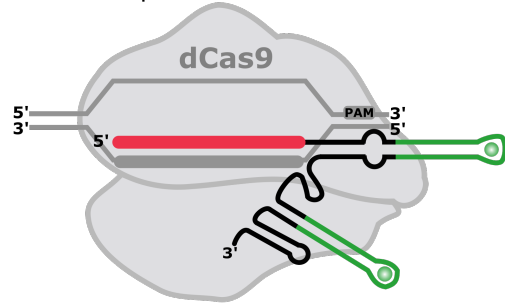
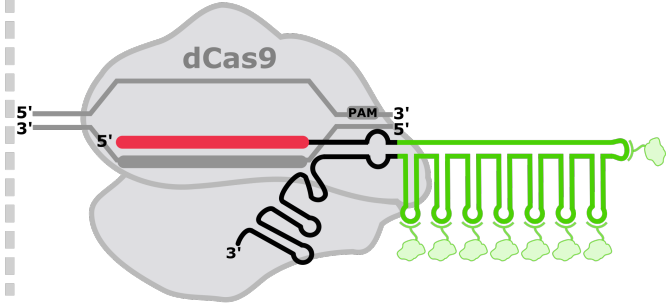
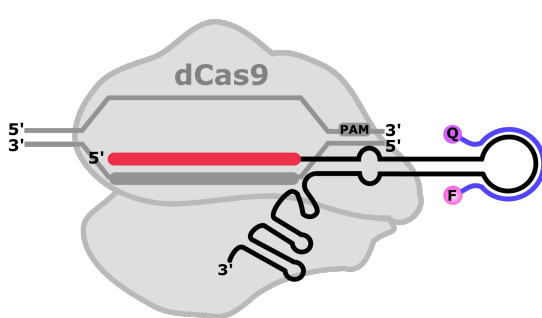
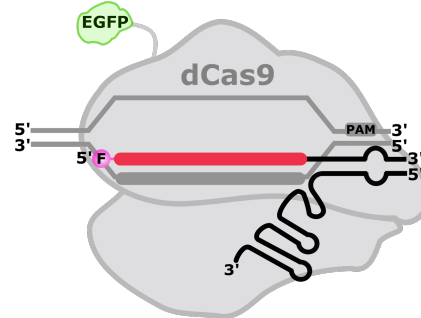
A CRISPRainbow**B Broccoli Aptamer****C CRISPR-Sirius****D Molecular Beacon****E LiveFISH**

Figure 3.4: sgRNA modifications for imaging. A) CRISPRainbow uses aptamers that recruit FP and combines red, blue and green FPs to increase the number of colours B) Broccoli aptamer uses an aptamer to bind a fluorescent small molecule C) CRISPR-Sirius uses multiple aptamers for a brighter fluorescence intensity D) Molecular Beacon probe that remains quenched until it binds to the sgRNA and increases fluorescence when the fluorophore and quencher separate from one another E) LiveFISH that uses a fluorescent sgRNA in combination with dCas9-EGFP.

While these approaches are encouraging, their limitations cannot be ignored. The heavily modified dCas9 fusion constructs require delivery of very large plasmids (~10

kilobases) into the cell which can be an issue. sgRNAs that are modified to recruit multiple FP have a tendency to aggregate and exhibit non-specific puncta,²²⁶ and these sgRNA with aptamer-recruiting sequences are usually very long (~400 nt in length) which impedes their production in cells and, potentially, their delivery efficiency.

The use of FP provides a good foundation for genomic DNA imaging, but the use of organic dyes is another alternative as they tend to generally be brighter, smaller in size and more photostable than FP.²³⁷ Wu *et al.* uses a molecular beacon (MB) design which incorporates a fluorophore and quencher on either end of a hairpin loop probe.²³⁸ This allows the probe to be a quenched system (since the fluorophore and the quencher are in close proximity) until the probe reaches the sgRNA, which hybridises to the MB's complementary sequence (Figure 3.4 D). When the MB hybridises, the fluorophore and quencher separate causing an increase in fluorescence. This approach could ensure a high signal-to-noise ratio. However, the dCas9 and sgRNA were introduced into cells together using an expression vector delivery method which requires excessive work if this approach was required for non-repetitive genome targeting (requires many sgRNAs). An alternative delivery method would provide a more robust methodology, but the use of electroporation or transfection would require synthesis of the sgRNA construct (~170 nt) *in vitro*, which could be challenging. To avoid false positive results in this assay, the MB and sgRNA would need guaranteed separation until reaching the genomic loci. Very recently, the

same group carried out a follow-up study in which they used two MBs and FRET between these MBs to image non-repetitive genomic targets but the issue of keeping the MB and sgRNA separated until they reach the target DNA still remains.²³⁹

LiveFISH, a study published during the work for this chapter, uses a dye attached directly to the gRNA and delivers the construct as a ribonucleoprotein (RNP) by electroporation to the cell (Figure 3.4 E).²⁴⁰ However, the versatility of RNP delivery is limited. Electroporation tends to be a harsh shock to cells which could change the morphology and levels of RNP will vary widely on a single cell level.²⁰²

3.2 Hypotheses and Aims

The most dominant issue in live cell imaging is finding the balance between maximum signal at the genomic loci while minimising background noise in the cell. While dCas9-3xGFP is a simple approach, the protein tends to localise into the nucleolus in live cells, providing a strong background noise. To tackle the signal-to-noise issue, a combination of a fluorescently labelled gRNA with dCas9-3xGFP could provide a solution. The issue of background fluorescence could be addressed with an 'on/off' fluorescent gRNA (F-gRNA) construct and the dCas9-3xGFP can be used for multi-colour co-localisation. This provides confirmation for on-target binding. This chapter discusses the development of a versatile gRNA construct that is bright, short in length and easy to synthesise and would allow for maximal fluorescence signal in cells whilst having reduced background signal from unbound probes.

3.3 Hairpin F-Q-gRNA – Design I

3.3.1 Design strategy of the construct

The design of the hairpin fluorophore-quencher gRNA (F-Q-gRNA) construct was based on a MB concept built into the gRNA itself. A quencher is placed in close proximity to the fluorophore by modifying the 5' end of the gRNA to carry a 2'-O-Me RNA:RNA hairpin, providing a quenched system (Figure 3.5). Once the gRNA binds to its DNA target, the hairpin, having a lower thermal stability due to its short length, would open, moving the quencher away from the fluorophore and allowing an increase in the fluorescence.

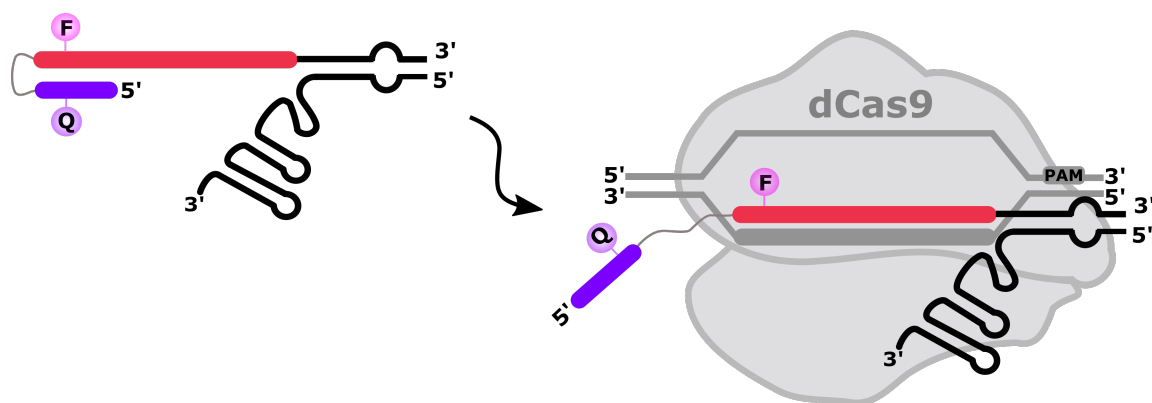


Figure 3.5: Hairpin F-Q-gRNA construct (Design I). Schematics showing the closed hairpin (quenched system, left) when the probe has not formed a complex with the dCas9 and the open hairpin (fluorescent system, right) when the gRNA–dCas9 complex encounters its correct DNA target.

The hairpin was made up of 2'-O-Me RNA bases to avoid degradation of the 5' end of the construct. It also was chosen to provide stability to the 8-mer hairpin. The hairpin linker was made up of 4x hexaethylene glycol (HEG) linkers (approximately ~10 Å in length) to allow easy hairpin formation, and more importantly, greater separation between the fluorophore and quencher upon target binding. The crRNA was slightly extended by 2 bases (22-nt targeting sequence) to allow the fluorophore to be attached to the 21st base. This was to avoid the fluorophore interfering with the gRNA–dCas9 interactions during the gRNA–dCas9–target DNA complex formation. Modifications within the non-seed region (11–20 bp region) are further explored in Section 3.4.1, but for initial studies, this was avoided.

Repetitive DNA sequences such as telomeres, centromeres and some genes (e.g. *MUC4*) are commonly used as a target in live cell imaging as the repetitive targets allow for additional fluorophores to collect at one genomic locus with only one gRNA sequence required.²²⁴ For the initial studies, the target chosen was the human mucin

gene, *MUC4*. This protein-coding gene has a 48-bp repeat region that occurs approximately 400 times on chromosome 3q29.^{223,241} For the hairpin F-Q-gRNA construct, *MUC4* was a more suitable target than a more repetitive target as the longer repeats provide greater spacing between each gRNA-dCas9 complex. When the hairpin is open, upon binding to the DNA target, the quencher from one complex will not quench the fluorescent signal from the neighbouring complex as illustrated in Figure 3.6.

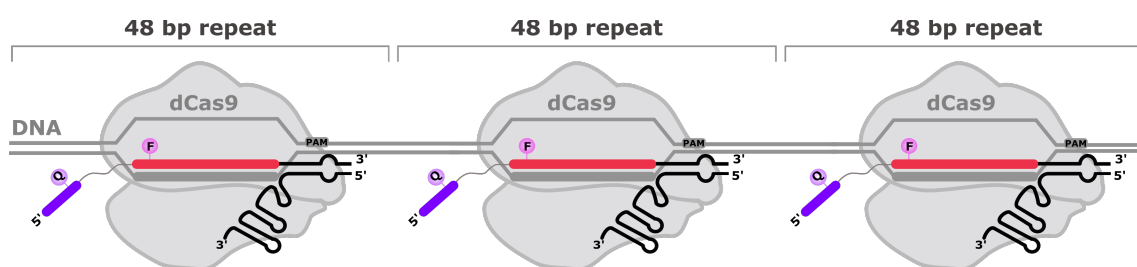


Figure 3.6: Hairpin F-Q-gRNA and dCas9 complexes bound to *MUC4* target. The 48-bp repeat allows sufficient spacing between each complex to avoid quenching of the fluorophore on the neighbour complex.

3.3.2 Synthesis of the construct

The synthesis of the hairpin F-Q-gRNA was divided up into two segments: a 44-mer crRNA, which contains the complementary target sequence, the fluorophore and the 2'-O-Me RNA hairpin carrying the quencher, and a constant 68-mer tracrRNA (Figure 3.7). By dividing the gRNA into fragments, the RNA synthesis becomes more efficient and higher yielding as the synthesis of short RNA is not hindered by its low coupling efficiencies compared to longer RNA. In addition, only crRNA will be made for each new DNA target as the tracrRNA sequence does not change. Note, reference to gRNA herein includes the crRNA construct and tracrRNA hybridised together.

A DNA Target

5'- GCCACGGGGCCTGCCAGT GCGGTGACCT—GTGGATGCTG AGGCTCATGAGCCCGAAGTGGC - 3'
 3'- CCGTGCCCCGGACGGTCA CCGCACTGGA—CACCTACGAC TCCGAGTACTCGGGCTTCACCG-NH₂ - 5'

22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

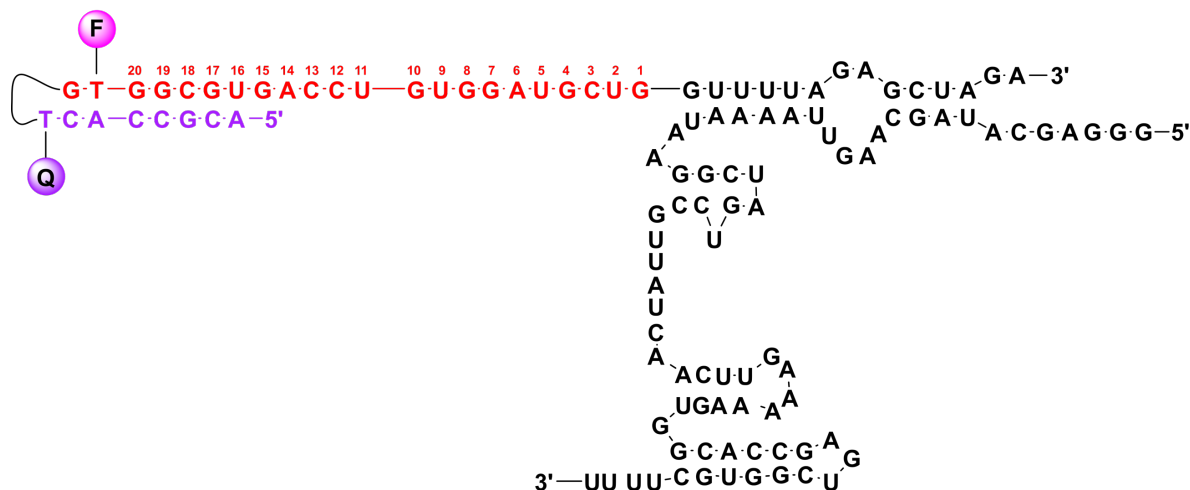
B Hairpin F-Q-gRNA

Figure 3.7: Sequences used in synthesis of Hairpin F-Q-gRNA. A) DNA target used. Bases in red are complementary to the DNA-targeting sequence of the gRNA, bases in black indicate the PAM region, NH₂ = 5'-Amino modification. B) Hairpin F-Q-gRNA construct. Bases in purple are 2'-O-Me RNA bases. T conjugated with Q = BHQ-2-deoxythymidine. Bases in red are complementary to target DNA, T conjugated with F = amino-C6-deoxythymidine labelled with a fluorophore, bases in black are the tracrRNA region. UPLC-MS data found in Appendix A4.1.

The crRNA was synthesised using solid-phase TBDMS synthesis with Black Hole Quencher® (BHQ-2) added to the construct *via* a commercially available modified T base phosphoramidite. The crRNA also contained an amino-C6-dT for Cyanine-3 (Cy3) labelling. 4x HEG modifications were also incorporated using a commercially available phosphoramidite. The crRNA was purified using reverse-phase (RP) HPLC and was subsequently labelled with Cy3 through NHS ester labelling (Figure 3.8 A). As a control, the crRNA without the HEG linker and the quencher-containing hairpin (F-crRNA) was also synthesised and labelled with Cy3. *In vitro* transcription (IVT) was used to synthesise the tracrRNA as this was unmodified RNA and gave sufficient yields for these studies (Figure 3.8 B). UPLC-MS data can be seen in Appendix A4.1.

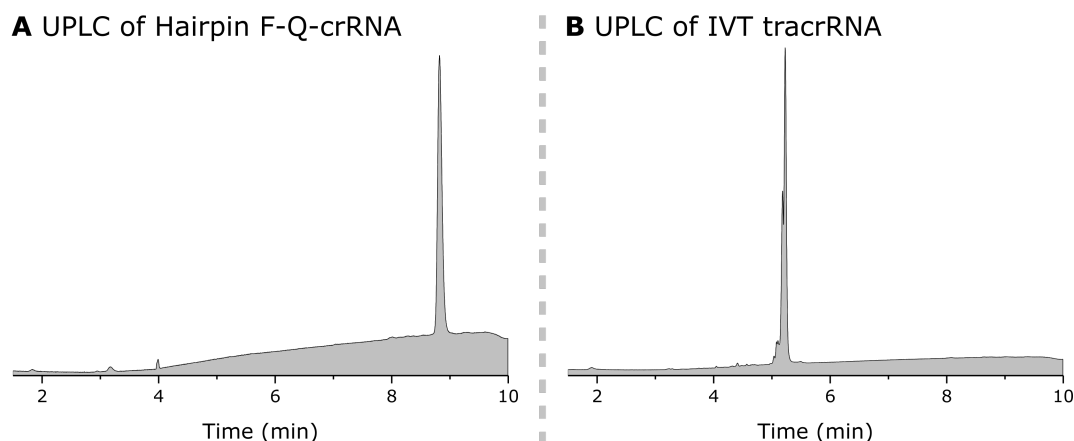


Figure 3.8: UPLC of Hairpin F-Q-gRNA construct. A) UPLC of Hairpin F-Q-crRNA labelled with Cy3. B) UPLC of IVT tracrRNA. Expected mass for A = 17224. Masses observed for A = 17224.0. Expected mass for B (+ 3x Mg²⁺) = 23217. Masses observed for B = 23215.4.

Cy3 and BHQ-2 were chosen as the fluorophore-quencher pair for this construct because Cy3 has previously been used in live cell imaging and shown to provide bright fluorescence.¹⁴ In addition, since the construct will be used in combination with a GFP-fused dCas9, it is important to choose a dye that is spectrally distant from GFP to minimise fluorescence overlap. BHQ-2 quenches in the range of 560–670 nm which should provide sufficient quenching of the Cy3 fluorophore.

3.3.3 *In vitro* analysis

3.3.3.1 *Fluorescence analysis*

Once the F-Q-gRNA and control F-gRNA were prepared, the efficiency of the MB concept was investigated by fluorescence intensity. From Figure 3.9 A, the effectiveness of the quenching system is evident with a ~50-fold increase in fluorescence for quencher-free F-gRNA compared F-Q-gRNA. Theoretically, when the F-Q-gRNA binds with the target and Cas9, the hairpin would open, and the

fluorescence change should be similar to the F-gRNA. However, when the Cas9-F-Q-gRNA complex was formed and incubated with the target DNA, the construct remains in its quenched state. This suggests that the hairpin has not opened. To evaluate this, UV melting studies were carried out to determine the melting temperature required to open the hairpin (Figure 3.9 B). The T_m of the hairpin F-Q-gRNA construct alone was 60 °C while the T_m of the construct and the ssDNA target strand was 78 °C. Since the fluorescence experiment was performed at 37 °C to mimic cellular conditions, it is possible that the hairpin has not opened, even though the formation of the crRNA:DNA duplex should be more favourable. The construct was labelled with another dye, TAMRA, to confirm the fluorophore used did not affect the results. The increase in fluorescence was only ~4 fold for the quencher-free F-gRNA compared F-Q-gRNA so Cy3 was used for further studies (Figure A14). Note that Cas9 is used instead of dCas9 for all studies *in vitro* due to its commercial availability.

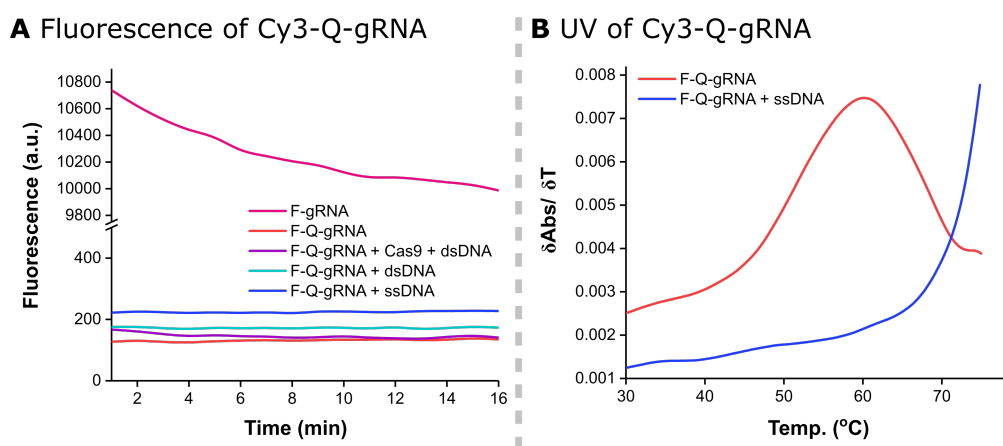


Figure 3.9: *In vitro* hairpin-opening study of F-Q-gRNA. A) Fluorescence measurements of hairpin Cy3-Q-gRNA in the presence and absence of Cas9 and DNA targets in Cas9 buffer at 37 °C over 15 min. Concentration of the construct is 0.4 μM in 20 μL and an equimolar ratio was used for all additional elements. Fluorescence measured at $\lambda_{\text{ex}} = 538 \text{ nm}$ and $\lambda_{\text{em}} = 554 \text{ nm}$. B) First derivative of UV absorbance at 260 nm of Cy3-Q-gRNA (red) and Cy3-Q-gRNA bound to the ssDNA target (blue) (UV absorbance spectra given in Figure A15) in Cas9 buffer. Concentration of the construct is 1 μM in 1000 μL and a 1:1 ratio was used for all additional elements.

3.3.3.2 DNA cleavage

To investigate if the presence of the hairpin prevents the binding of modified gRNA to Cas9 or inhibits Cas9-mediated unwinding of the DNA target (hence, hairpin opening), DNA cleavage was performed. For Cas9 to cut the target, binding between the protein and the construct must occur. However, this may be a short-lived binding. Therefore, cutting would not determine efficient binding for imaging but it would provide a useful indication of whether the modifications could be tolerated. The target strand of the dsDNA target was labelled with ATTO 647N dye to help visualisation of the cut product. Since the cut DNA is shorter than the uncut DNA, it should migrate faster on the gel.

From the gel in Figure 3.10, the gRNA with no dye, HEG linker or hairpin (Construct 1) was used as a comparison for the efficiency of DNA cleavage by Cas9 (39% cutting could be seen after 1 h, potentially due to the modified base). When a fluorophore is introduced, it is evident that it reduces the Cas9 cutting ability. The introduction of Cy3 (Construct 3) reduces the cutting ability to 14% compared with Construct 1 (39%). To see the effects of different dye, gRNA was labelled with TO instead of Cy3 (Construct 2) and it also reduces the cutting ability (8%). Even though the fluorophore was placed outside of the non-seed region (at 21th base), crystal structures have shown that the 2'-OH moiety at the 20th base in the crRNA has a H-bonding interaction with the Cas9 (Figure 3.11).²¹¹ This suggests that the presence of the fluorophore close to this region could interfere with Cas9 binding. When the quencher-containing hairpin

was included in the gRNA with no fluorophore (Construct 4), DNA cleavage seemed to be similar efficiency as Construct 2 (5%). With the combination of the hairpin and Cy3 (Construct 5), DNA cutting is abolished. Collectively, having a Cy3 fluorophore reduces the efficiency from 39% to 14% and having a hairpin only reduces it from 39% to 5%. With both modifications, the cutting efficiency is non-detectable. From this, we can deduce the presence of the fluorophore and hairpin hinders the binding of Cas9.

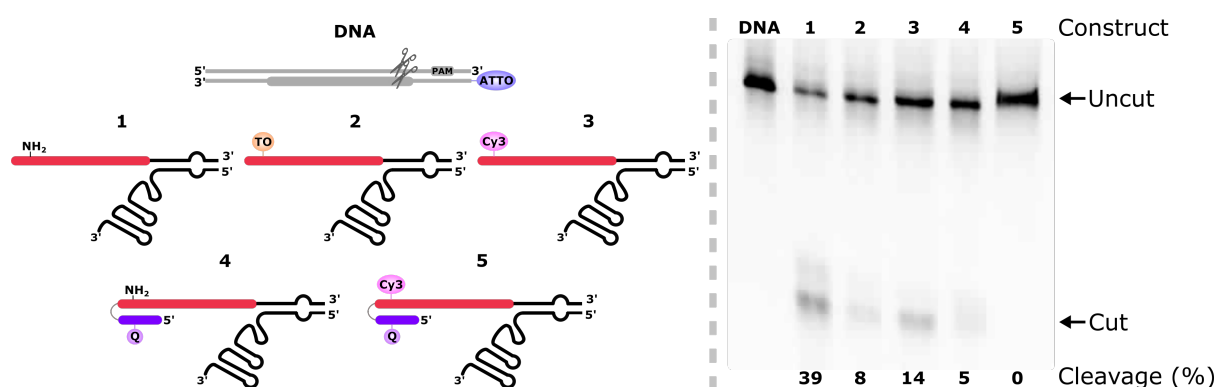


Figure 3.10: DNA cleavage assay of Hairpin F-Q-gRNA. Schematics of construct 1–5 are shown on the left. Construct 1 = Unlabelled gRNA, construct 2 = TO-gRNA, construct 3 = Cy3-gRNA, construct 4 = Unlabelled Q-gRNA, construct 5 = Cy3-Q-gRNA. 20% denaturing PAGE gel of DNA cleavage captured using ATTO-647N filter is shown on the right. DNA lane contains uncut ATTO 647N-labelled dsDNA target, all other lanes contain dsDNA target, gRNA, and Cas9 (1:1:1 ratio, 0.5 μ M in 20 μ L) incubated in Cas9 buffer at 37 °C for 1 h. % of cutting was calculated *via ImageJ*.

Therefore, from the gel and the fluorescence data, it can be concluded that: 1) the position of the fluorophore is not optimal and potentially interferes with Cas9 binding to the DNA target, 2) the hairpin is too stable and may not open efficiently, and/or 3) the hairpin may extend too far into the DNA-targeting region. It has been demonstrated that a hairpin which is fully complementary to the DNA-targeting region (a 20-mer hairpin) will block Cas9 activity.²⁴² However, in this case, approximately 8 bases appear to be sufficient to inhibit Cas9 activity.

3.4 HyBeacon FF-gRNA – Design II

For the second design, a HyBeacon approach was explored. A HyBeacon™ is a single-stranded, fluorophore-containing probe with no secondary structures. In a single-stranded state, the probe has a low fluorescence intensity as the flexible structure allows quenching of the fluorophore. However, when the probe hybridises with its target and forms a rigid duplex structure, there is a significant increase in fluorescence.⁴⁹ Based on this concept, other improved HyBeacons were developed which contained two fluorophores. Collisional interactions between the two fluorophores in the single-stranded state would help quench each other. Once the rigid double-stranded structure is formed, these interactions could no longer occur, allowing the quenched dyes to become fluorescent.^{243,244}

In comparison with the previous hairpin F-Q-gRNA design, the HyBeacon FF-gRNA was simpler to synthesise as the long HEG-linked 2'-O-Me hairpin was replaced by two nucleotide modifications. Having demonstrated that the position of the fluorophore on the crRNA has an effect on the Cas9's DNA cleavage activity, two questions were asked before designing the construct: 1) where in the DNA-targeting region of the crRNA could be modified with a fluorophore and 2) could the 5' end of the gRNA be extended and how long could this extension be so that the modifications would not interfere with Cas9 interactions?

3.4.1 Modifying the DNA-targeting region of crRNA

The crystal structure of the Cas9–sgRNA complex provided insight into mandatory Cas9–gRNA interactions which cannot be disrupted.²¹¹ This enabled a more rational chemical modification of the gRNA. Cas9 tolerates many modifications such as 2'-fluoro-RNA and phosphorothioate-modified backbones in the DNA-targeting region of crRNA.²⁴⁵ 2'-O-Me modifications are also used in modified gRNA and previous studies provided some insight on where to place the modifications for the fluorophores. Completely replacing the seed region (position 1–10) or the entire DNA-targeting region (position 1–20) with 2'-O-Me residues induces a total loss of gRNA activity.²⁴⁶ These results supports the crystallographic data that shows the 2'-OH RNA residues in the seed region are required for Cas9 binding and function (Figure 3.11 A).

A summary of some studies that modified the DNA-targeting region with 2'-O-Me and only cause a maximum of a 10% decrease in gRNA activity can be seen in Figure 3.11 B.^{211,245–250} Using 2'-O-Me modifications as an indicator would provide insight as to where a 2'-amino-modified nucleotide could be incorporated for fluorophore incorporation. Collectively, it is evident that positions 18–19 or 11–14 have potential to be modified with a fluorophore and still maintain sufficient gRNA activity.

A Cas9 and sgRNA interactions

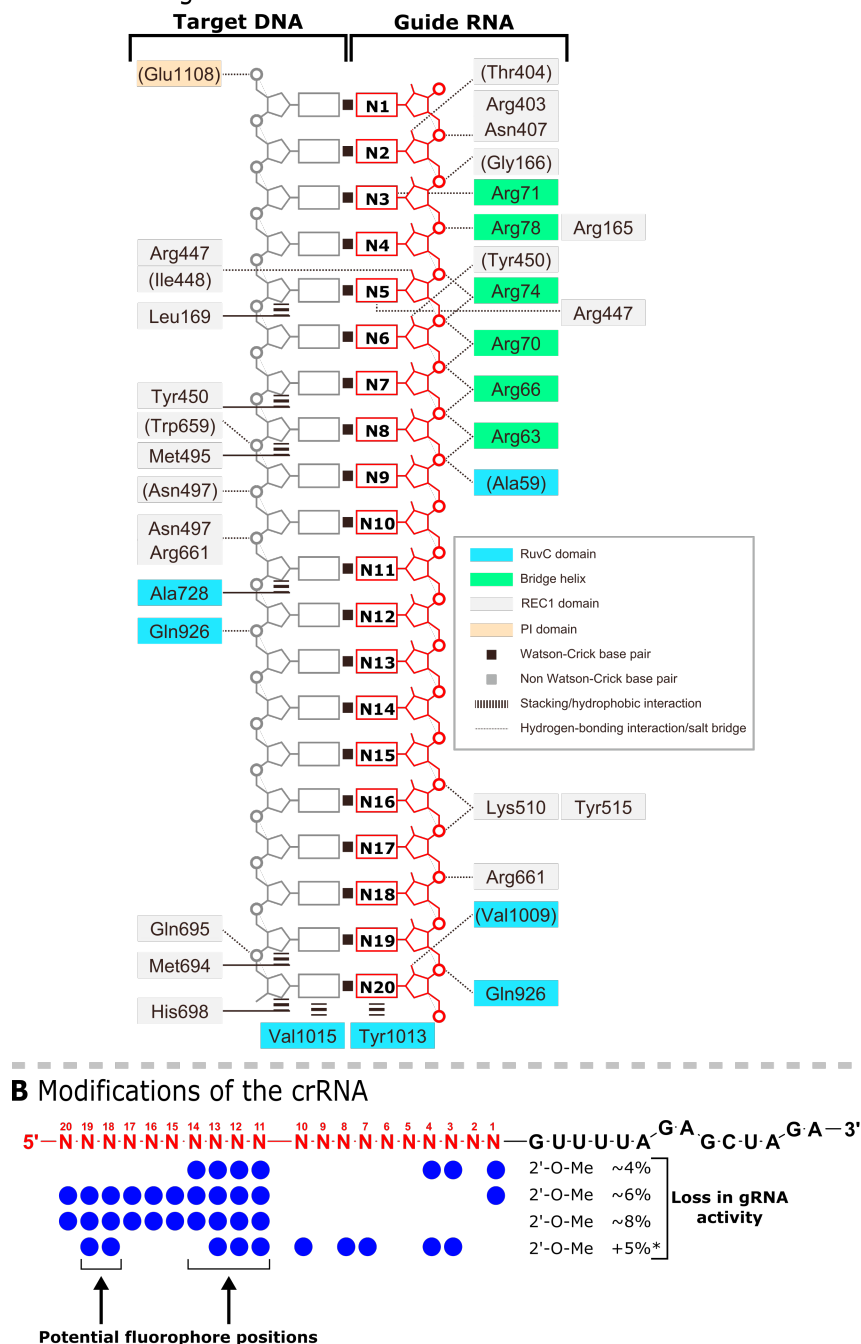


Figure 3.11: Interactions and modifications to the DNA-targeting region of crRNA. A) Schematic detailing the interactions of Cas9 within the DNA-targeting region (position 1–20) of sgRNA (red) and target DNA (grey) determined by crystal structures. Image adapted from Nishimasu, *et al.*²¹¹ B) Table summarising the various combinations of positions for 2'-O-Me. Only combinations that did not reduce the activity of the sgRNA lower than 10% compared to the unmodified control are shown. *Modifications that increase gRNA activity. DNA-targeting sequence is shown in red with N = nucleotide. 2'-O-Me modifications are indicated by a blue dot.

3.4.2 Extending the 5' end of crRNA

The other alternative is to place the modifications outside the DNA-targeting region to completely avoid interfering with Cas9–gRNA interactions. Previous studies have demonstrated that extending the 5' with DNA or RNA did not have an effect on Cas9 activity.^{229,251} Ferry *et al.*, has shown that the gRNA with up to a 25-nt 5'-extension of the DNA-targeting region has comparable activity to the non-extended gRNA.²⁴² From this, we could design a construct that contains a HyBeacon extending from the 5'-end of the DNA-hybridising region. However, with an extended gRNA, duplex formation beyond the 20-nt DNA-binding region will need to be investigated.

3.4.3 Design strategy of construct

Considering the literature, two new constructs were investigated; HyBeacon FF-gRNA (Hy FF-gRNA) and extended HyBeacon FF-gRNA (Ext Hy FF-gRNA). These gRNA constructs contain two fluorophores that will quench each other when the gRNA is in a single-stranded state. Once the gRNA forms a complex with Cas9 and subsequently hybridises to its DNA target, the rigid duplex structure in Cas9 will hinder the quenching interactions from occurring, resulting in an increase in fluorescence (Figure 3.12). The two constructs will be directly compared to assess whether introducing the HyBeacon concept within the DNA-targeting region or outside this region is more favourable.

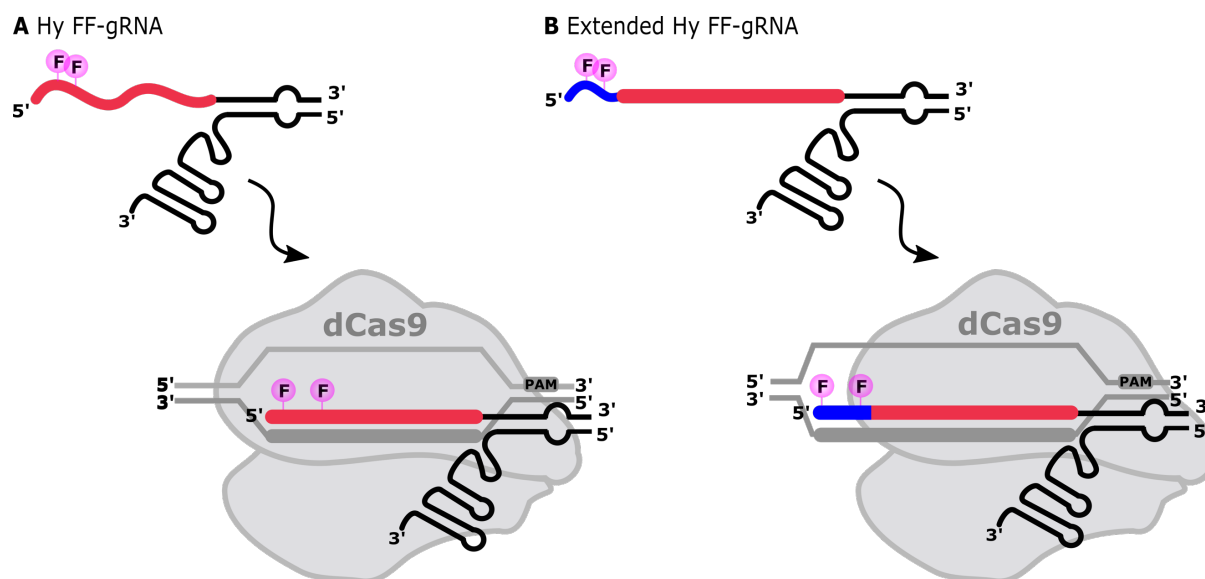


Figure 3.12: HyBeacon FF-gRNA Constructs (Design II). A) HyBeacon FF-gRNA and B) Extended HyBeacon FF-gRNA showing the quenched system (top) in the single-stranded state and the fluorescent system (bottom) when it forms the complex with dCas9 and the target DNA.

3.4.4 Synthesis of the construct

The same gRNA target, *MUC4*, was used in the analysis. Since the tracrRNA strand of the gRNA is identical to previous designs, only the 34-mer crRNA containing a HyBeacon moiety was synthesised using solid-phase TBDMS synthesis. For the Hy FF-gRNA construct (Figure 3.13 A), RNA nucleotides at position 12 and 18 were modified for fluorophore incorporation, giving a 6-base separation when the rigid duplex was formed. For the Ext FF-gRNA construct, a 10-base extension was added to the 5'-end of the crRNA, resulting in a 44-mer in length. Positions 24 and 29 were chosen for the fluorophore incorporation, providing a 5-base separation between two dyes (Figure 3.13 B). The extension was made up of 2'-O-Me RNA to avoid degradation of the gRNA and also to facilitate hybridisation with the DNA beyond the canonical 20-bp RNA:target DNA duplex.

A HyBeacon FF-gRNA**B** Extended HyBeacon FF-gRNA

Figure 3.13: Sequences used in synthesis of HyBeacon FF-gRNA. A) HyBeacon FF-crRNA B) Extended HyBeacon FF-crRNA. Bases in red are complementary to DNA target. Bases in blue are 2'-O-Me RNA bases. C conjugated with F = 2'-aminoethoxy^{Me}C labelled with a fluorophore (Cy3). DNA target and tracrRNA are the same as in Figure 3.7. UPLC-MS data found in Appendix A4.1.

For the incorporation of the fluorophore, a 2'-modified nucleotide, 2'-aminoethoxy^{Me}C, was chosen following the results of previous studies discussed in Section 3.4.1 which used 2'-OMe modifications. Both constructs were purified using the standard methods. Once purified, the construct was labelled with Cy3-NHS ester labelling, similar to previous methods, and pure products were confirmed by UPLC-MS (Figure 3.14 A and B).

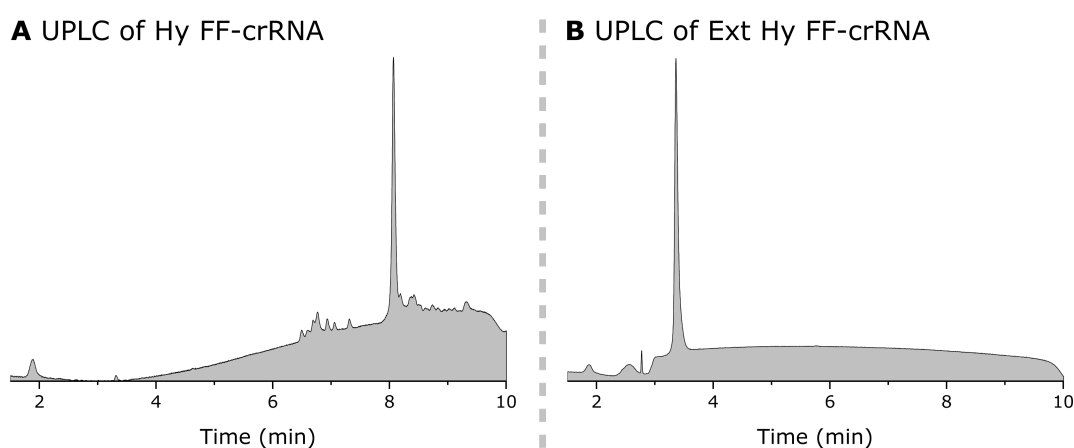


Figure 3.14: UPLC of HyBeacon FF-gRNA constructs. A) UPLC of Hy FF-crRNA labelled with 2 x Cy3 B) UPLC of Ext Hy FF-crRNA labelled with 2 x Cy3. Expected mass for A = 11981. Mass observed for A = 11981.0. Expected mass for B = 15291. Mass observed for B = 15292.0. Note, UPLC-MS gradient optimised for hydrophobic ODNs was used for B.

3.4.5 *In vitro* analysis

3.4.5.1 DNA cleavage assay

Although it has been shown that Cas9 tolerates 2'-O-Me groups at position 12 and 18 of crRNA, 2'-aminoethoxy ^{Me}C is slightly larger than 2'-O-Me and may affect Cas9 binding. Therefore, before labelling the construct with the fluorophores, the unlabelled constructs with 2'-aminoethoxy ^{Me}C were tested as a control in a Cas9-mediated DNA cleavage assay. In Figure 3.15 Gel A, a reduction of the uncut band and increase of a cut band confirms 42% cutting for Construct 2 and 58% cutting for Construct 4 has occurred after 1 h. The reaction was also carried out for 16 h and, as expected, DNA cleavage improved from 42% to 73% for Construct 2 and from 58% to 92% for Construct 4. Construct 4 (extended construct) is more efficient in DNA cleavage than Construct 2 which is expected as the modifications lie outside the Cas9 binding region.

Given that the gRNA construct allowed Cas9 to cut its DNA target, the constructs were labelled with Cy3 and the experiment was repeated (Figure 3.15, Gel B). With two bulky fluorophores now incorporated into the constructs, there is no major change to the DNA cleavage efficiency. Comparing construct 2 and 3, we see that both bands are similar (40% vs 42%). They are less efficient than construct 1 but this is expected as we move from unmodified to modified gRNAs. Similarly, the same phenomenon occurs for the extended constructs (4 and 5). DNA cleavage occurs for all constructs, denotating that binding does occur.

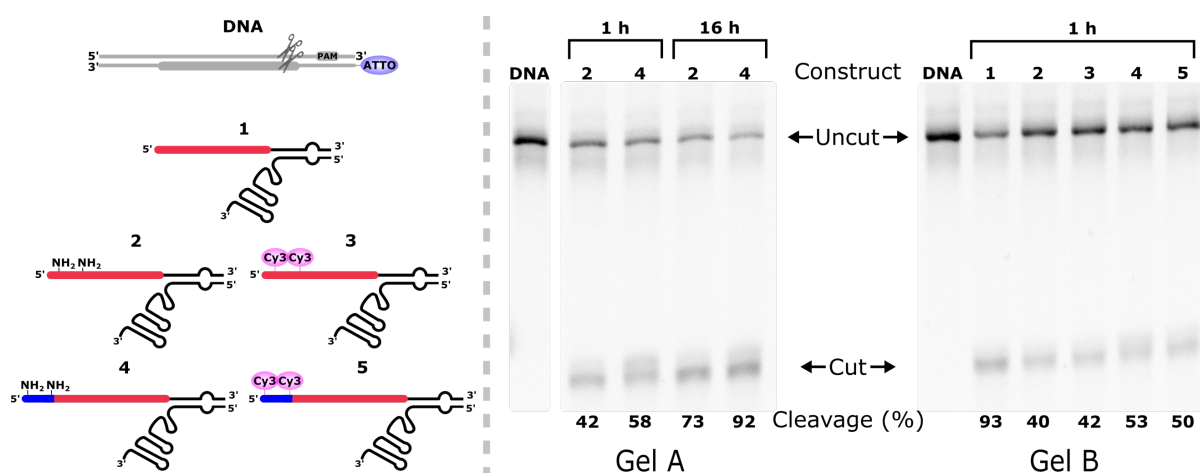


Figure 3.15: DNA cleavage assay for Hy FF-gRNA and Ext Hy FF-gRNA. Schematics of construct 1-5 are shown on the left. Construct 1 = unmodified gRNA, construct 2 = unlabelled Hy gRNA, construct 3 = Hy Cy3-Cy3-gRNA, construct 4 = unlabelled Ext Hy gRNA, construct 5 = Ext Hy Cy3-Cy3-gRNA. 20% denaturing PAGE gels of DNA cleavage captured using ATTO-647N filter are shown on the right. DNA lane contains uncut ATTO 647N-labelled dsDNA target, all other lanes contain dsDNA target, gRNA, and Cas9 (1:1:1 ratio, 0.5 μ M in 20 μ L) incubated in Cas9 buffer at 37 $^{\circ}$ C. Gel A contains the unlabelled versions of the construct with incubation times of 1 h and 16 h. Gel B contains all constructs including the unmodified control incubated for 1 h.

3.4.5.2 Fluorescence analysis

Next, fluorescence change was evaluated to investigate the feasibility of the HyBeacon approach; gRNA alone should provide low fluorescence and in the presence of Cas9 and the target DNA, a fluorescence increase should be seen. In Figure 3.16 A and B, there is a fluorescence increase from the single-stranded state of the gRNA to Cas9 complex formation with the dsDNA target. For the Hy FF-gRNA construct, this is a 2.7-fold increase in fluorescence and for the Ext Hy FF-gRNA, it is a 1.5-fold increase. When the gRNA construct meets dsDNA without the Cas9, the change in fluorescence is minimal confirming the theory that the gRNA construct will not open the dsDNA without the help of Cas9 which is beneficial for imaging. This is further seen in the controls in Figure A16.

Interestingly, when the complex between the gRNA and the Cas9 is formed (no dsDNA), there is an increase in fluorescence of 2.2-fold for Hy FF-gRNA (Figure 3.16 A). This could be attributed to Cas9 forming a rigid complex structure with the gRNA and separating the two fluorophores to allow an increase in fluorescence. This suggests that the turn-on response of the HyBeacon in the presence of DNA would not be very effective. Also, Cas9 induces a ‘bubble’ opening in the dsDNA but it is not established whether the bubble can be extended out of the Cas9 binding region so that the target strand can form a duplex with the extended section of the gRNA. The minimal change in fluorescence for the Ext Hy FF-gRNA construct does not help to confirm that duplex formation outside of the bubble is occurring.

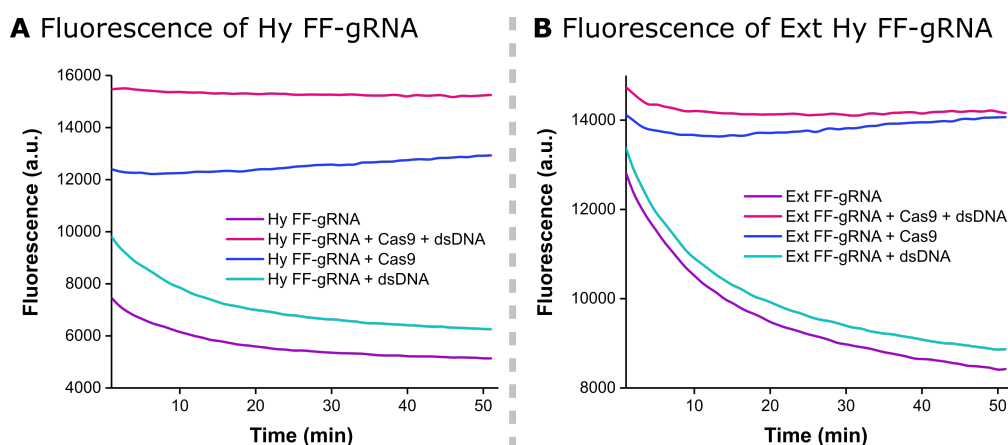


Figure 3.16: Fluorescence measurements of Hy FF-gRNA and Ext Hy-gRNA. Fluorescence data of A) Hy FF-gRNA and B) Ext Hy FF-gRNA labelled with Cy3 in Cas9 buffer at 37 °C over 60 min. Concentrations of the construct is 0.5 μ M in 20 μ L and a 1:1:1 ratio of gRNA:Cas9:DNA target was used for all additional elements. Fluorescence measured at λ_{ex} = 538 nm and λ_{em} = 554 nm. Controls can be found in Figure A16.

The HyBeacon systems have benefits, however, some issues with the system need to be considered. Although there is a change in fluorescence, the fold increase is not very large. The increase in fluorescence when the gRNA forms a complex with the

Cas9 alone could introduce some issue when imaging. There is potential for a false positive result as the initial change in fluorescence might not necessarily mean that the gRNA has bound to its target as seen in Figure 3.16 B where the change in fluorescence between Ext FF-gRNA and Ext FF-gRNA + Cas9 + dsDNA is minimal. A system that would require all three components (gRNA, Cas9 and dsDNA) of the complex to be together before a change in fluorescence was detected could provide an even more effective imaging system.

3.5 RNase H QxF-gRNA – Design III

Whilst the previous results showed some promise, the fluorescence change would need to be much more distinct to be effective in cell imaging, so the use of a quencher system was re-visited. Previous work has been carried out using a cleavable moiety to activate a sgRNA-dCas9 complex by nucleases²⁴² and the use of a selective ribonuclease, available in the cellular environment, is an interesting concept. RNase H1 is a ribonuclease that will selectively hydrolyse RNA in a RNA:DNA heteroduplex.^{252,253} Thus, providing the gRNA with a segment that could be cleaved by RNase H1 in the cell could provide a solution for a detachable quencher system. Studies have been previously carried out with this concept and this new design is based on those results.¹⁴

3.5.1 Design strategy of the construct

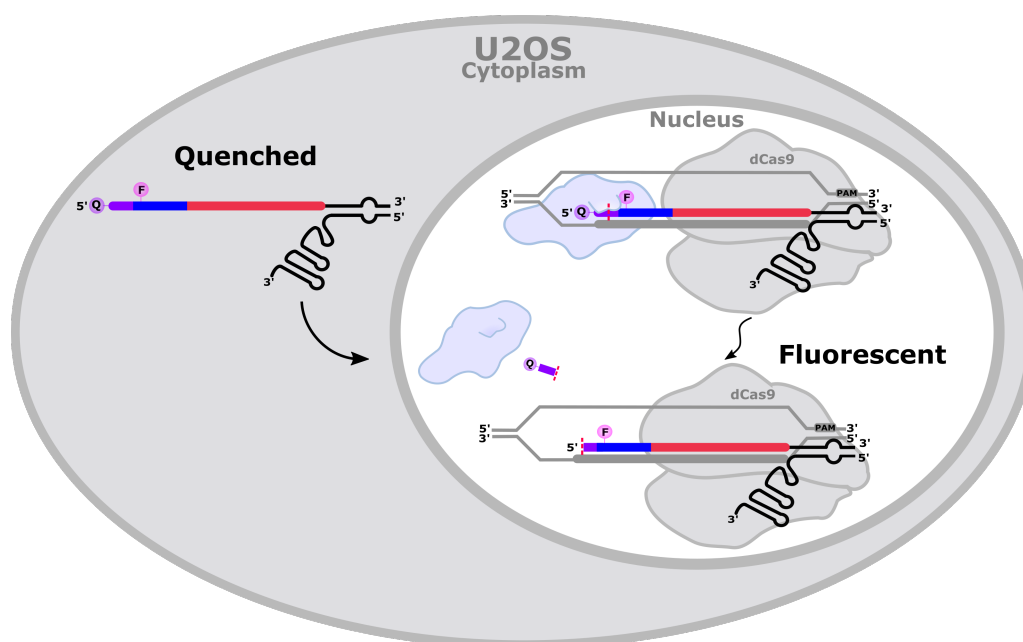


Figure 3.17: Design of RNase QxF-gRNA (Design III). Schematics showing QxF-gRNA (quenched system) in a cell setting and the RNase H-cleaved xF-gRNA (the fluorescent system) that is generated only when all three components are present: gRNA, dCas9 and DNA target.

As with the first design, the gRNA contains a quencher to provide an initially quenched fluorescent state. Once the dCas9–gRNA recognises its DNA target and an RNA:DNA heteroduplex is formed, RNase H1 could then cleave the segment of RNA that contains a quencher, resulting in an increase in fluorescence (Figure 3.17). From our previous studies, we can deduce that extending the 5'-end is tolerated by dCas9. Therefore, this RNase H-cleavable section could lie outside the dCas9 binding region. There is some debate about the location of RNase H in the cell; while most believe it is mainly found in the nucleus of the cell,²⁵⁴ others have found that it can also be found in the cytoplasm.²⁵⁵ Therefore, the system would need to be cleavable only when it is at the genomic loci and would not be cleavable if it binds with RNase H outside of the nucleus. By designing a system that requires binding to the target DNA before it can

be cleaved, the system would only be effective when all three components of the complex have formed (gRNA, dCas9 and dsDNA target). This RNase H-activated system provides an 'on/off' imaging approach.

For this design, various aspects would need to be considered. 1) The heteroduplex needs to be accessible to the RNase H enzyme; the cleavable segment needs to be further away from the dCas9 binding region to allow the RNase H to bind. 2) The cleavable RNA section cannot be too long otherwise, once the RNA is cut, it is stable enough to remain hybridised in the duplex. 3) The distance between the fluorophore and the quencher needs to be taken into account and the fluorophore must not be in a section of the construct that could be cleaved by RNase H.

A lot of studies have characterised the activity of RNase H, providing guidelines to the design of the new construct. Generally, RNase H needs a minimum of 4 viable bases in a RNA:DNA duplex to cleave^{256,257} although cleavage with only 3 bases have been observed in some studies.^{258,259} Mechanistically, RNase H is considered not to be sequence-selective, but instead recognises the mixture of A- and B-form conformations of the heteroduplex. RNase H binds along the minor groove of the heteroduplex and interacts with the 2'-OH group of 4 consecutive RNA nucleotides that are in C3'-*endo* sugar pucker for cleavage. It also interacts with 7 bases of the DNA strand which is mainly in C2'-*endo* sugar puckers.^{255,260}

Modifications to the 2' of the DNA strand in the DNA:RNA heteroduplex such as 2'-O-Me diminishes RNase H activity²⁶¹ but introducing these modifications into the RNA strand is also shown to stop cleavage at these sites.^{256,259} *E. Coli* RNase H is standardly used *in vitro* as it is readily available and its preferences for RNA targets is practically identical to human RNase H.²⁵⁶ The catalytic domains of both RNase H show a strong similarity²⁶² and so, for these studies, *E. Coli* RNase H was used for *in vitro* analysis.

From the literature, we concluded that the extension needs to be at least 7 bases long to allow the RNase H to bind with a minimum of 4 sites for potential cleavage. 2'-O-Me can be used to direct the cleavage towards the region where the quencher is attached and also used to protect the region where the fluorophore is attached.

3.5.2 Synthesis of the construct

With these requirements, three versions of the RNase H constructs were synthesised to test the viability of a number of factors: 1) RNase H binding site – the constructs contain an 8-base binding region or a 15-base binding region (denoted by Y in the construct name X:Y, Figure 3.18) to compare how much RNA is needed to bind and cut; 2) RNase H cutting sites – the constructs contained 4, 5 and 6 possible RNase H cutting sites (denoted by X in the construct name X:Y) to allow maximum cutting but also to allow the segments to be short enough that once cut, they are not stable enough to remain hybridised in the heteroduplex; 3) Position of the fluorophore – the

fluorophore is either attached to a nucleotide within the DNA-targeting region (Figure 3.18 B) or attached to a nucleotide in the extension (Figure 3.18 A and C).

A RNase 4:8 QxF-gRNA



B RNase 6:8 QxF-gRNA



C RNase 5:15 QxF-gRNA

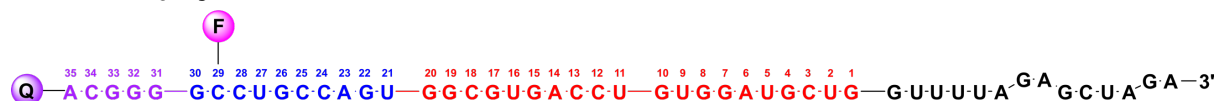


Figure 3.18: Sequences used for the RNase QxF-gRNA design. A) RNase 4:8 QxF-gRNA construct (where possible cleavage sites = 4 bases and RNase H binding site = 8 bases) B) RNase 6:8 QxF-gRNA construct (where possible cleavage sites = 6 bases and RNase H binding site = 8 bases) C) RNase 5:15 QxF-gRNA construct (where possible cleavage sites = 5 bases and RNase H binding site = 15 bases). Bases in red are complementary to DNA target. Bases in blue are 2'-O-Me RNA bases. C conjugated with F = 2'-aminoethoxy^{Me}C labelled with a fluorophore (Cy3). DNA target and tracrRNA are the same as in Figure 3.7. UPLC-MS data found in Appendix A4.1.

The constructs were synthesised using solid-phase synthesis and TC chemistry. BHQ-2 and Cy3 were chosen as the detachable quencher–fluorophore pair again (QxF where ‘Qx’ denotes a cleavable quencher). BHQ-2 was incorporated as a 5'-BHQ-2 phosphoramidite and 2'-aminoethoxy^{Me}C was used for the fluorophore incorporation. The 4:8 QxF-gRNA construct was a 42-mer that contains a fluorophore at position 24 and has a QxF distance of 4 bases (Figure 3.18 A). The 6:8 QxF-gRNA construct was also a 42-mer that contains a fluorophore at position 18 and has a QxF distance of 10 bases (Figure 3.18 B). The 5:15 QxF-gRNA construct was a 49-mer that contains a fluorophore at position 29 and has a QxF distance of 6 bases (Figure 3.18 C). 2'-O-Me bases are used to direct the RNase H cleavage and also, to protect the fluorophore

when it lies outside the DNA-targeting region. As controls, the constructs without the 5'-BHQ-2 quencher were also synthesised. Once purified, all constructs were labelled with Cy3-NHS ester (Figure 3.19 A-C).

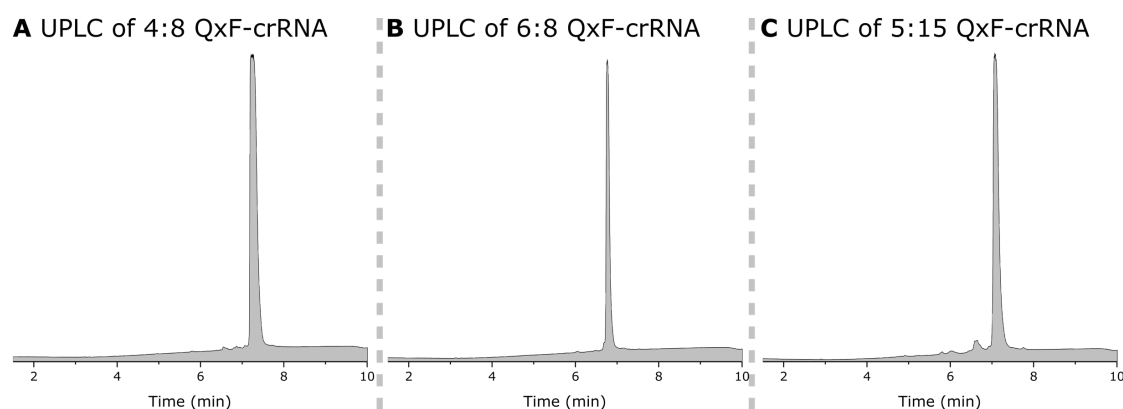


Figure 3.19: UPLC of RNase QxF-gRNA constructs (Design III). A) UPLC of 4:8 QxF-crRNA labelled with Cy3 B) UPLC of 6:8 QxF-crRNA labelled with Cy3 C) UPLC of 5:15 QxF-crRNA labelled with Cy3. Expected mass for A = 14585. Mass observed for A = 14586.0. Expected mass for B = 14571. Mass observed for B = 14572.0. Expected mass for C = 16990. Mass observed for C = 16990.0.

3.5.3 *In vitro* analysis

3.5.3.1 RNase H cleavage

To confirm whether the binding region for RNase H was sufficient when the Cas9 protein was present in the complex, an RNase H cleavage assay was performed. From the gels in Figure 3.20 A-C, we can see cleavage of the gRNA is occurring for all three QxF-gRNAs. The construct with a quencher was used to provide the most accurate determination of cutting but since the quencher is present, the band of the construct alone is faint in the fluorescence channel. The distance between the uncut and cut bands is larger than expected but this could be due to the larger size and/or hydrophobic nature of the quencher which could retard the QxF-gRNA complex compared to the cut xF-gRNA.

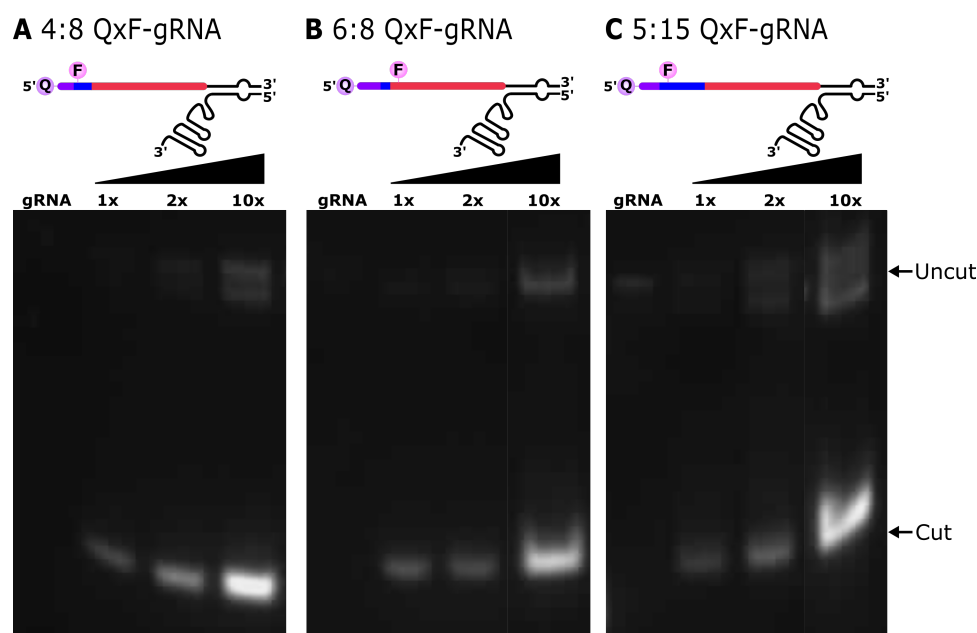


Figure 3.20: RNase H cleavage assay of the all QxF-gRNA construct. 20% denaturing PAGE gel was used. gRNA lane contains 0.5 μ M in 20 μ L of A) 4:8 QxF-gRNA, B) 6:8 QxF-gRNA C) 5:15 QxF-gRNA construct. All other lanes contain 1x (0.5 μ M), 2x and 10x gRNA–Cas9–dsDNA complex and 1x (10 U) RNase H incubated in RNase H buffer at 37 $^{\circ}$ C for 1 h. Cas9 was added to the gRNA and incubated for 15 min for complex formation before RNase H was added. Gel image captured using a Cy3 Filter.

3.5.3.2 Fluorescence

Each construct was investigated using fluorescence by comparing 1) QxF-gRNA (the quenched system), 2) xF-gRNA (the maximum potential fluorescence); fluorescence increase if no quencher is present and 3) QxF-gRNA–Cas9–dsDNA complex and RNase H (the experimental fluorescence); fluorescence increase once the quencher is removed by RNase H with Cas9 present. The QxF-gRNA with ssDNA target and RNase H (the maximum RNase fluorescence) was also measured to compare how much of the experimental fluorescence is hindered by the presence of Cas9.

For all QxF-gRNA (Figure 3.21 A-C), quenching was effective as the gRNA construct alone has very low fluorescence. As expected, we can conclude that more RNA sites

available for RNase H cleavage encourages a better increase in fluorescence; 6 sites > 5 sites > 4 sites (Figure 3.21 D). This can be seen by comparing the maximum RNase fluorescence (solid lines) of all three gRNA. While the experimental fluorescence (dotted lines) follows the same trend, the presence of Cas9 hinders all three gRNA as their fluorescence does not fully reach its maximum RNase fluorescence (solid line vs dotted line).

4:8 QxF-gRNA was closest to its maximum RNase fluorescence (77% of RNase fluorescence) compared to 6:8 QxF-gRNA (55% of RNase fluorescence) although both have the same binding space (8 bases) and 6:8 QxF-gRNA has more cutting options. One hypothesis for this is that the 5th and 6th option in 6:8 QxF-gRNA are too close to the Cas9 binding region and become less viable options. This suggests accessibility is key for effective cutting. The position of the fluorophore could also contribute to the low fluorescence as it is within the DNA-targeting region of the gRNA. This could reduce the binding ability of Cas9 and hence, the formation of the DNA:RNA duplex. This position of the fluorophore is also an issue for fluorescence of xF-gRNA. 4:8 xF-gRNA and 5:15 xF-gRNA (fluorophore outside of the DNA-targeting region) have a generally higher fluorescence compared to 6:8 xF-gRNA. The fluorescence measurement was also carried out for a further 4 h and the experimental cutting continues to increase over this 4 h period (Figure A17 (left)). In cellular conditions, however, it is difficult to know how long the RNase H will be bound to the complex.

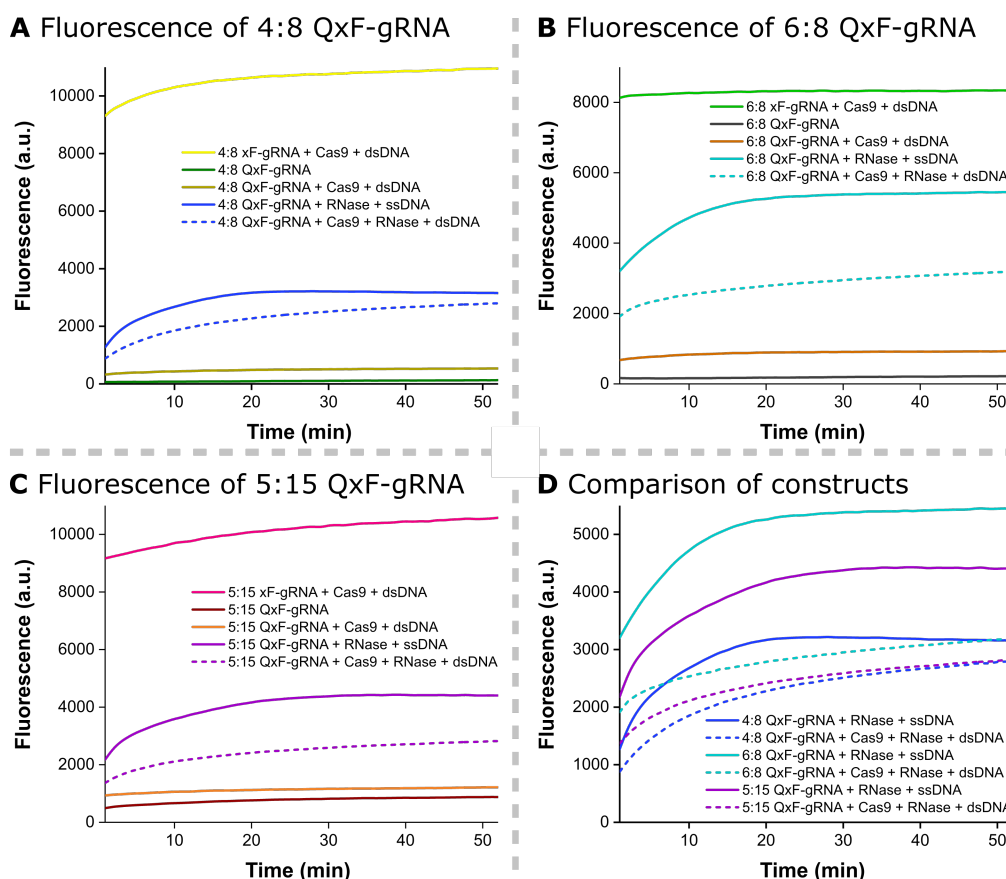


Figure 3.21: Fluorescence of RNase constructs (Design III). Fluorescence data of A) RNase 4:8 QxF-gRNA construct, B) RNase 6:8 QxF-gRNA construct and C) RNase 5:15 QxF-gRNA construct labelled with Cy3 in Cas9 buffer at 37 °C over 60 min. Concentrations of the construct is 0.25 μM in 20 μL and a 1:1 ratio was used for all additional elements except for Cas9 which was used in excess (5 eq, 1.25 μM in 20 μL). Cas9 was added to the sample and allowed 15 min to form the complex before RNase H (5U) was added. D) A comparison of the maximum RNase fluorescence (solid line) and the experimental fluorescence (dashed line) of each construct. Fluorescence measured at $\lambda_{\text{ex}} = 538 \text{ nm}$ and $\lambda_{\text{em}} = 554 \text{ nm}$. Controls can be found in Figure A17.

A combination of the maximum number of cutting sites (6 sites), distance from the DNA-targeting region (10 bases) and fluorophore position (outside the DNA-binding region) would lead to the generation of an optimised gRNA which is discussed in Section 3.7.

3.5.3.3 Electrophoretic Mobility Shift Assay (EMSA)

Electrophoretic mobility shift assay (EMSA) was employed to investigate the binding of the RNase H QxF-gRNA construct to Cas9. Here, a non-denaturing PAGE was used to detect the ribonucleoprotein complex formation. Since the complex is larger than the pore of the gel and the net charge of gRNA is reduced when shielded by Cas9, the bound gRNA will remain in the well whereas the unbound species will migrate into the gel. The binding of the gRNA is assessed by increasing the equivalence (eq) of Cas9 to see where the gRNA band disappears. The binding specificity is further investigated by introducing a competing IVT sgRNA, known to bind efficiently with a Cas9 protein. If displacement occurs with 1 eq of the competitor, the binding is not very efficient. This is the case for the 5:15 QxF-gRNA construct (Figure 3.22, right gel). At least 2 eq of Cas9 is needed to bind to 5:15 QxF-gRNA. The reasoning for this inefficient binding is explored in Section 3.6.2.

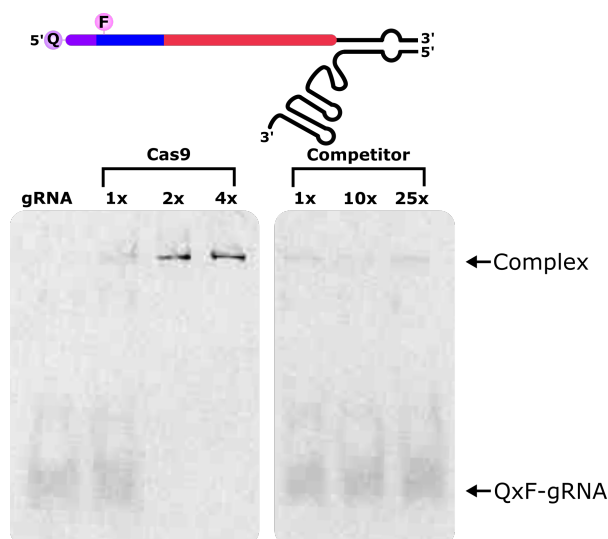


Figure 3.22: EMSA of the 5:15 QxF-gRNA construct. 10% native PAGE gel was used. gRNA lane contains 5:15 QxF-gRNA construct (1 μ M in 10 μ L), Cas9 lanes contain gRNA (1 μ M) and Cas9 (1 μ M, 2 μ M and 4 μ M). Competitor lanes contain gRNA (1 μ M), Cas9 (1 μ M) and IVT sgRNA as a competitor (1 μ M, 10 μ M and 25 μ M) in Cas9 buffer supplemented with (0.01 μ g) BSA. Complex formation proceeded for 15 min at 37 $^{\circ}$ C before adding a competitor and further incubated for 5 min at 37 $^{\circ}$ C. Gel image captured using a Cy3 Filter.

3.6 Live Cell Imaging

3.6.1 Cell delivery

For the delivery of the modified system into live cells, lipid-based transfection was explored for this study as an electroporation system was not readily available. Lipid-based transfection does, however, have drawbacks. Labelled gRNA can get trapped in the endosomes or lysosomes after transfection which would cause increased background fluorescence in the study. However, the design of the new quenched construct aims to directly address this issue.

The constructs and a plasmid encoding for dCas9-3xGFP were transfected using Lipofectamine reagents which are cationic lipids that facilitate delivery of oligonucleotides into the cells.^{263,264} A complex between the lipid and oligonucleotide is formed and the complex can enter the cell in vesicles called endosomes using endocytosis. These lipoplexes (cationic) electrostatically interact with the lipid bilayer of the endosomal membrane (anionic) and allows for the release of the oligonucleotide into the cytoplasm of the cell.^{265,266} This process must occur before the endosomes fuse with lysosomes in the cell where the oligonucleotides are degraded.^{264,267} The plasmid for dCas9-3xGFP is transfected prior to the construct to allow time for the plasmid to enter into the cell, be transcribed into mRNA and be translated at the ribosome into the dCas9 protein. The protein will then bind with the transfected gRNA construct and the NLS component of protein should direct it towards the nucleus where it can bind with the target DNA.

3.6.2 Controls

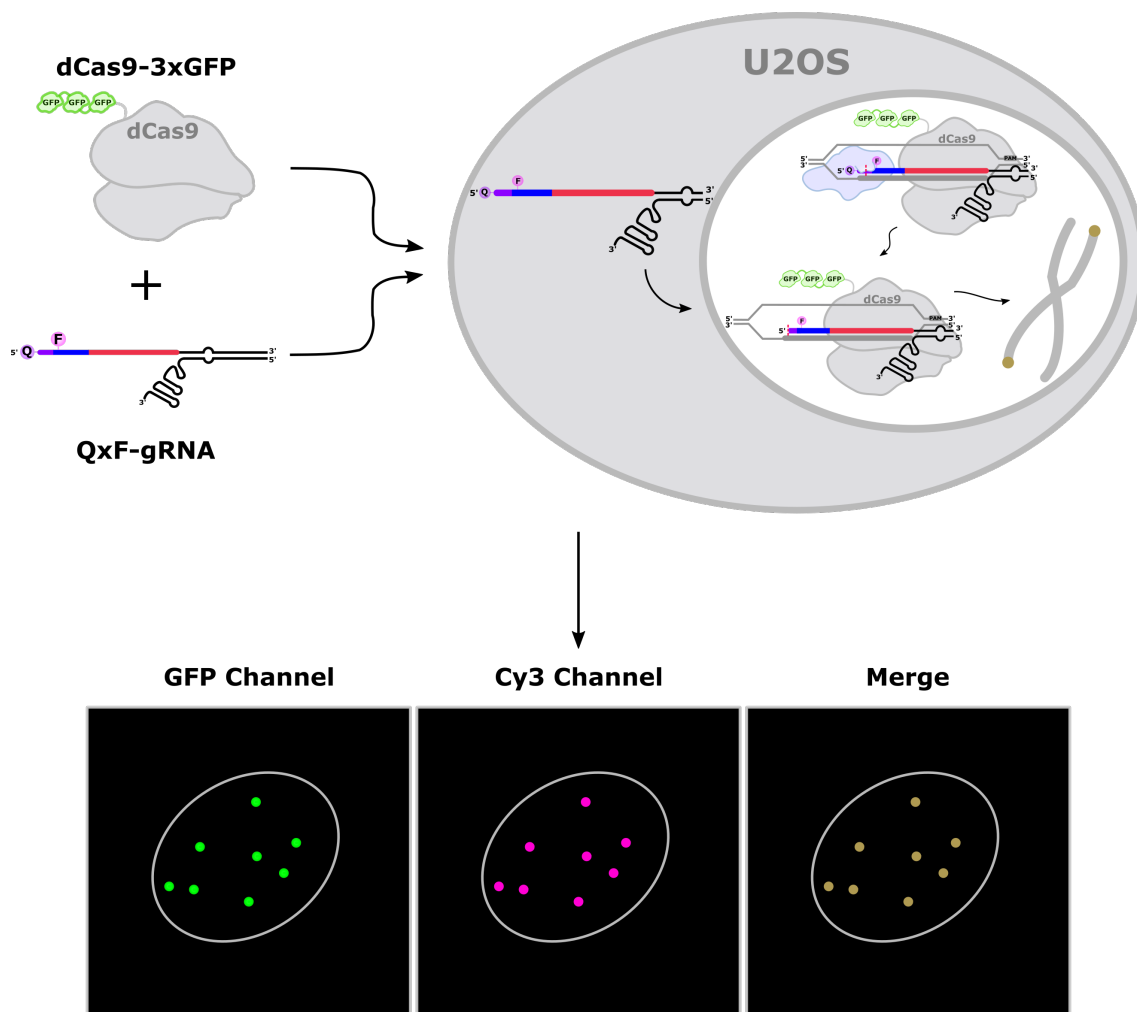


Figure 3.23: Schematic of live cell imaging using dCas9-3xGFP and QxF-gRNA constructs. dCas9-3xGFP and the construct are delivered into U2OS cells. They form a complex with the target DNA in the genome and RNase H removes the quencher (Q) from the construct. Unbound constructs remained quenched. This provides fluorescent puncta in the nucleus of the cell which can be imaged using confocal fluorescence microscopy.

Telomeres are a common target in live cell imaging as the highly repetitive regions of the genome facilitate imaging. Telomeres contain multiple TTAGGG repeats and so a large number of dCas9-gRNA complexes would be recruited to the same region providing a detectable signal. The cells used in this study were U2OS cells (human bone osteosarcoma cells) which contain ~65 chromosomes²⁶⁸ to even further facilitate genomic imaging. Since the dCas9 protein is fused with GFP, imaging the cells in both

the GFP channel and the Cy3 channel would provide evidence of co-localisation of the two components of the complex (Figure 3.23). It also can be used to determine binding efficiencies of constructs that do not contain a fluorophore.

These studies were carried out in tandem with the *in vitro* analysis from design III. Initially, the delivery method was confirmed using a standard unmodified sgRNA construct, synthesised using IVT, that is known to bind to dCas9. The bright puncta in the nucleus of Figure 3.24 shows the successful transfection of both components of the complex (dCas9 and IVT sgRNA) and the binding to genomic DNA. For every transfection carried here after, IVT sgRNA was also transfected as a positive control.

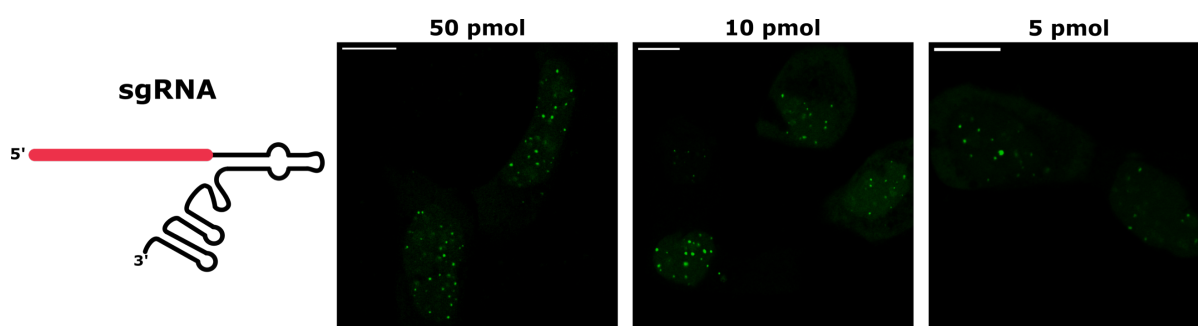


Figure 3.24: Live cell imaging of dCas9-3xGFP and IVT sgRNA. Images were taken using confocal fluorescence microscopy of U2OS cells transfected with IVT sgRNA (quantities indicated) and a dCas9-3xGFP plasmid. Identical excitation and emission settings were used for each image. Laser used to visualise GFP (green) is described in Section 5.6.4. Scale bar = 10 μ m.

Next, the separation of the crRNA and tracrRNA was investigated by transfecting a native, split version of the gRNA (Figure 3.25). This construct did not generate clear puncta suggesting the dCas9–gRNA–DNA complex kinetic stability is insufficient even at higher quantities (up to 100 pmol). The reasoning for this is most likely due to the limited hybridisation of the gRNA components; 9 WC bp and 1 wobble bp may

not be sufficient. The images in Figure 3.25 also highlight the issue involved with using dCas9-3xGFP alone as an imaging method as it accumulates in the nucleolus.

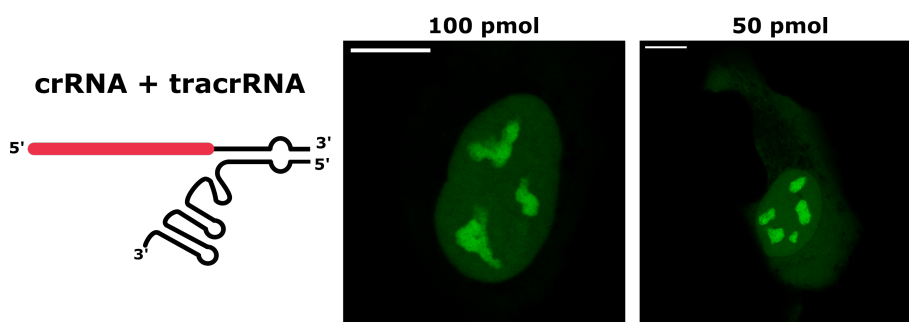


Figure 3.25: Live cell imaging of dCas9-3xGFP and split gRNA. Images were taken using the GFP channel (green) as previously described. Scale bar = 10 μ m.

To eliminate this problem, an extension of the 3'-end of the crRNA was investigated so that a total of 32 bases hybridise to a 5'-extended tracrRNA (31 WC bp and 1 wobble bp), extending the construct by 22 bases. This should enable stronger hybridisation of the crRNA and tracrRNA to enable stable binding. Firstly, the construct was tested *in vitro* (Figure 3.26) and no major difference was seen by comparing DNA cleavage of the non-extended 3'-version (left) with the 3'-extended version (right).

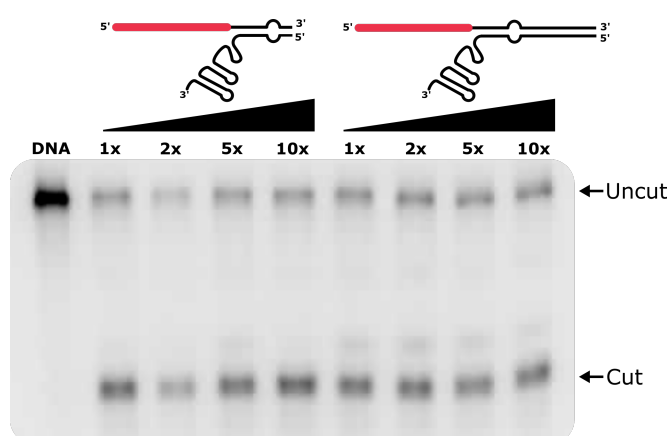


Figure 3.26: DNA cleavage of gRNA vs Ext gRNA control. 10% denaturing PAGE gel was used. DNA lane contains uncut dsDNA, all other lanes contain dsDNA (0.25 μ M in 20 μ L) with 1x, 2x, 5x and 10x construct and Cas9 enzyme incubated at 37 $^{\circ}$ C for 1 h in Cas9 buffer. Gel image captured using an ATTO Filter. DNA = dsDNA target with target strand labelled with ATTO 647N.

Following this, the construct was examined in cells. Bright puncta can be seen in Figure 3.27, confirming efficient binding, even at quantities similar to what is needed for the IVT sgRNA. This 3'-extension was incorporated into the final design of the construct.

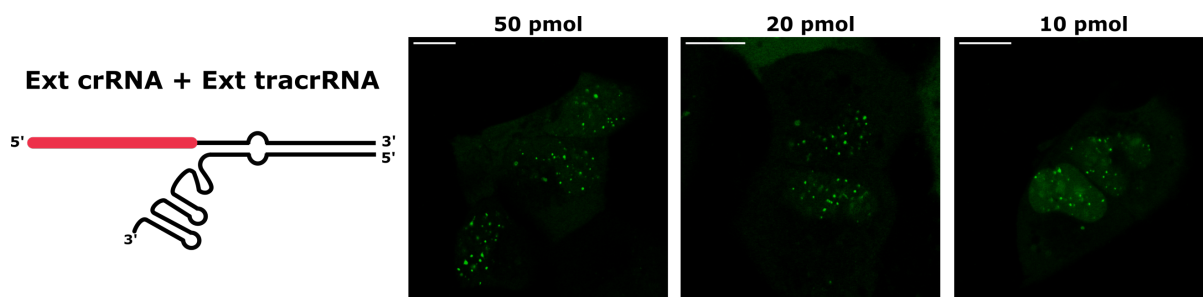


Figure 3.27: Live cell imaging of dCas9-3xGFP and extended gRNA. Images were taken using the GFP channel (green) as previously described. Scale bar = 10 μ m.

3.7 RNase H QxF-gRNA – Design IV

3.7.1 Design and synthesis of the construct

For the final design of the RNase H construct, the results from the constructs of design III and the results of the live cell imaging controls were all considered. Firstly, the binding of the construct was addressed. EMSA gel (Figure 3.22) and live cell imaging (Figure 3.25) confirm that the previous constructs would not form stable dCas9–gRNA–DNA complexes, but extending the crRNA at the 3'-end and the tracrRNA at the 5'-end is sufficient to overcome this problem for the unmodified control (Figure 3.27). This was incorporated into the new design.

A DNA Target

5' - ACCTCGCTGATGGCCACGCG GGGTTA GGGTTAGGGT TAGGGTTAGG —GTTAGGGTTA AGGCTCATGAGCCCGAAGTGGC - 3'
 3' - TGGAGCGACTACCGGTGCGC CCCAAT CCCAATCCCA ATCCCAATCC —CAATCCCAAT TCCGAGTACTCGGGCTTCACCG - 5'

36 35 34 33 32 31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

B RNase 6:16 QxF-gRNA

Figure 3.28: Sequences used in RNase 6:16 QxF-gRNA construct. A) DNA target used. Bases in red are complementary to the DNA-targeting region of the gRNA, bases in black indicate the PAM region B) RNase 6:16 QxF-gRNA construct. Bases in blue are 2'-O-Me RNA bases. Bases in purple are cleavable RNA bases. Q = 5'-BHQ-2. Bases in red are complementary to target DNA, U conjugated with F = amino-C6-deoxyuridine labelled with a fluorophore, bases in black are the tracrRNA region. UPLC-MS data found in Appendix A4.1.

The target for the construct was changed to a telomeric target as the quencher is no longer covalently attached to the gRNA, as in Design I, if RNase H cleavage occurs. Therefore, the spacing between different dCas9-gRNA-DNA complexes is no longer important. Therefore, a new DNA target was synthesised (Figure 3.28 A). A combination of the constructs from design III were used in design IV. The optimum number of cutting sites of 6 was used and a longer distance from the DNA-targeting region was also used. The fluorophore was kept outside the DNA-targeting region and a 10-base QxF distance was used (incorporated at 26th base) (Figure 3.28 B).

The 70-mer crRNA and its controls were synthesised and labelled using the same methods as the previous constructs (Figure 3.29 A). For the fluorophore incorporation, an amino-C6-deoxyuridine modification was used since in the

telomeric target, C was no longer in the sequence. An extended tracrRNA was also synthesised (Figure 3.29 B).

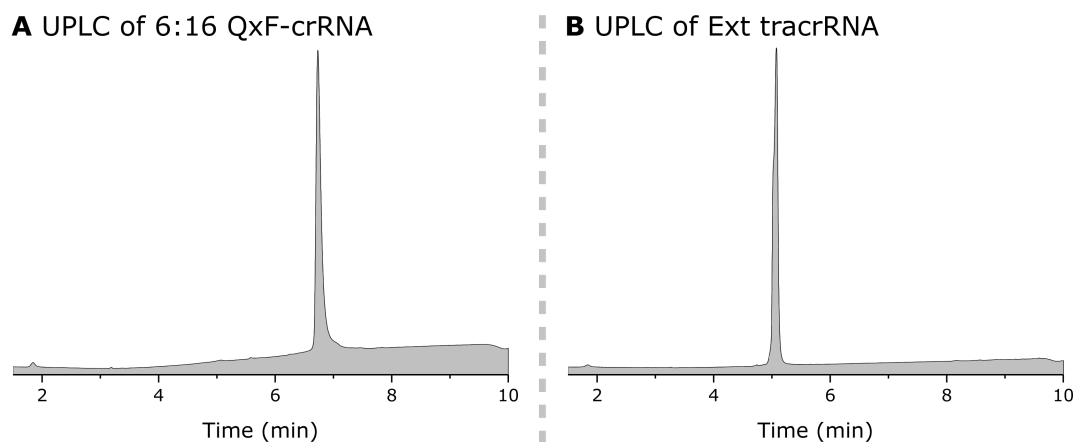


Figure 3.29: UPLC of RNase 6:16 QxF-gRNA construct (Design IV). A) UPLC of 6:16 QxF-crRNA labelled with Cy3 B) UPLC of Ext tracrRNA. Expected mass for A = 23908. Mass observed for A = 23907.0. Expected mass for B = 27842. Mass observed for B = 27843.0.

3.7.2 *In vitro* analysis

3.7.2.1 EMSA

An EMSA PAGE gel was carried out to confirm that the construct could form a stable Cas9–gRNA complex. The quencher-free 6:16 xF-gRNA control was used for the binding assay to facilitate visualisation. From Figure 3.30, we can see that the complex is forming in the well and only 1 eq of Cas9 protein is needed for complete binding. The binding is also specific as 10 eq of the competitor was required before the xF-gRNA construct could be displaced. Note, in the gRNA lane, a faint band can be seen where the complex band should be. This is due to some sample diffusing from the neighbouring lane.

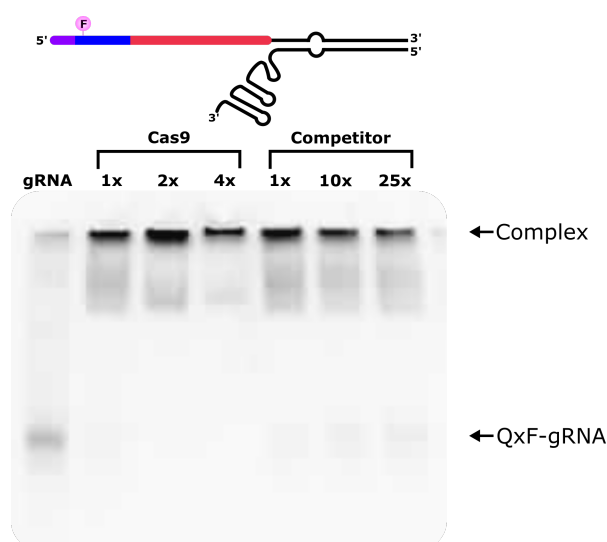


Figure 3.30: EMSA PAGE gel of RNase 6:16 construct. 10% native PAGE gel was used. gRNA lane contains 6:16 xF-gRNA construct (1 μ M in 10 μ L), Cas9 lanes contain gRNA (1 μ M) and Cas9 (1 μ M, 2 μ M and 4 μ M). Competitor lanes contain gRNA (1 μ M), Cas9 (1 μ M) and IVT sgRNA as a competitor (1 μ M, 10 μ M and 25 μ M) in Cas9 buffer supplemented with (0.01 μ g) BSA. Complex formation proceeded for 15 min at 37 $^{\circ}$ C before adding a competitor and further incubated for 5 min at 37 $^{\circ}$ C. Gel image captured using a Cy3 Filter.

3.7.3 Live cell imaging

Next, the construct was investigated in live cells. The construct was transfected and was compared to the control, which did not contain the quencher, to investigate the ‘on/off’ applicability of the system. At a wide-field level, as shown in Figure 3.31, there is a substantial difference in background fluorescence between the two constructs (6:16 xF-gRNA (top) vs 6:16 QxF-gRNA (bottom)) when using the same microscope excitation and emission settings for all images. Although background fluorescence is evident in both xF-gRNA images and QxF-gRNA images, in the cells that contain a quencher, it is greatly reduced. The residual background fluorescence could be caused by unsuccessful endosomal escape of the construct which could lead to degradation of the gRNA. This could be optimised further by introducing a modified backbone or nucleotides known to be resistant to degradation.

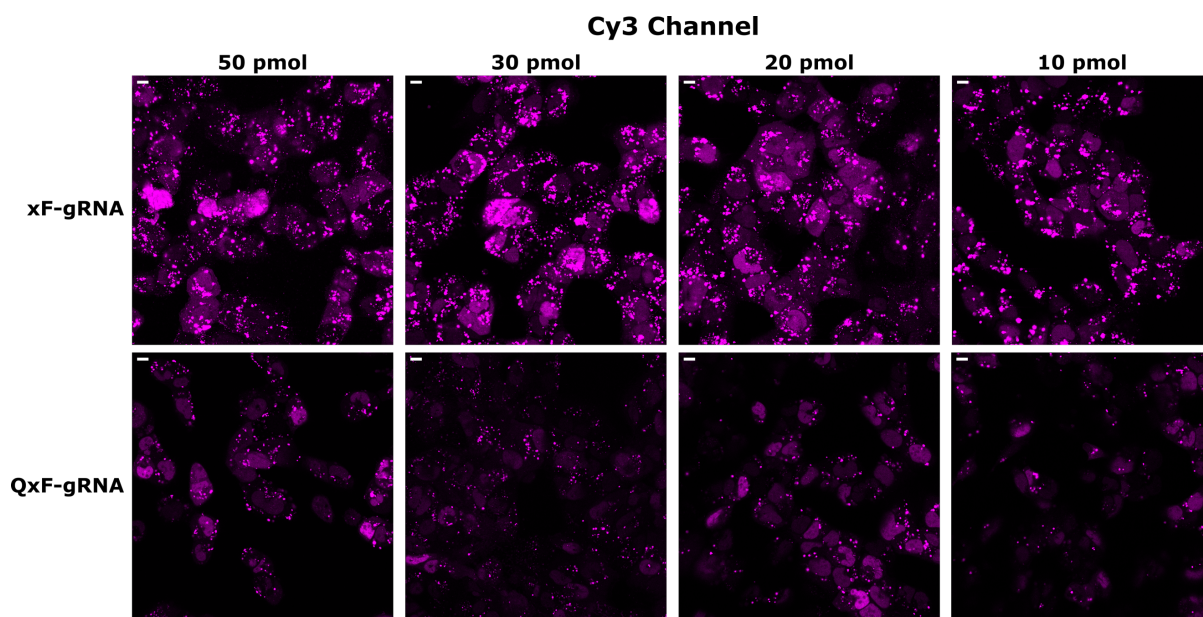


Figure 3.31: Wide-field live cell imaging of dCas9-3xGFP and 6:16 QxF-gRNA vs 6:16 xF-gRNA. Images were taken using confocal fluorescence microscopy of U2OS cells transfected with the gRNA constructs (quantities indicated) and a dCas9-3xGFP plasmid. Identical excitation and emission settings were used for each image. Laser used to visualise Cy3 (pink) is described in Section 5.6.4. Scale bar = 10 μ m.

At a single-cell level (Figure 3.32 - Figure 3.33), in the GFP channel, the green puncta highlight the dCas9-3xGFP binding with the genomic DNA. In the Cy3 channel, 6:16 QxF-gRNA bound to the target provides bright pink puncta in the nucleus. In the merge channel, the co-localisation of the GFP signal and the Cy3 signal produce the yellow/off-white puncta indicating that the QxF-gRNA, complexed with the dCas9-3xGFP, binds to the DNA target. Comparing 6:16 QxF-gRNA (bottom) to 6:16 xF-gRNA (top), the overall background fluorescence is reduced indicating the presence of the quencher but once the QxF-gRNA construct reaches the DNA target, the quencher is removed. This is indicated by the puncta of both systems being of comparable intensity. The ‘on/off’ fluorescence system is effective providing reduced fluorescence until the Cas9-gRNA-DNA complex is formed. This agrees with the *in*

in vitro results. The system is effective at higher quantities (50–30 pmol, Figure 3.32 A and B). At lower quantities (10 pmol, Figure 3.33 B), the background is minimal, but the intensity of the puncta is also reduced for both xF and QxF constructs. This is expected as lower quantities of the gRNA would not provide as strong an intensity.

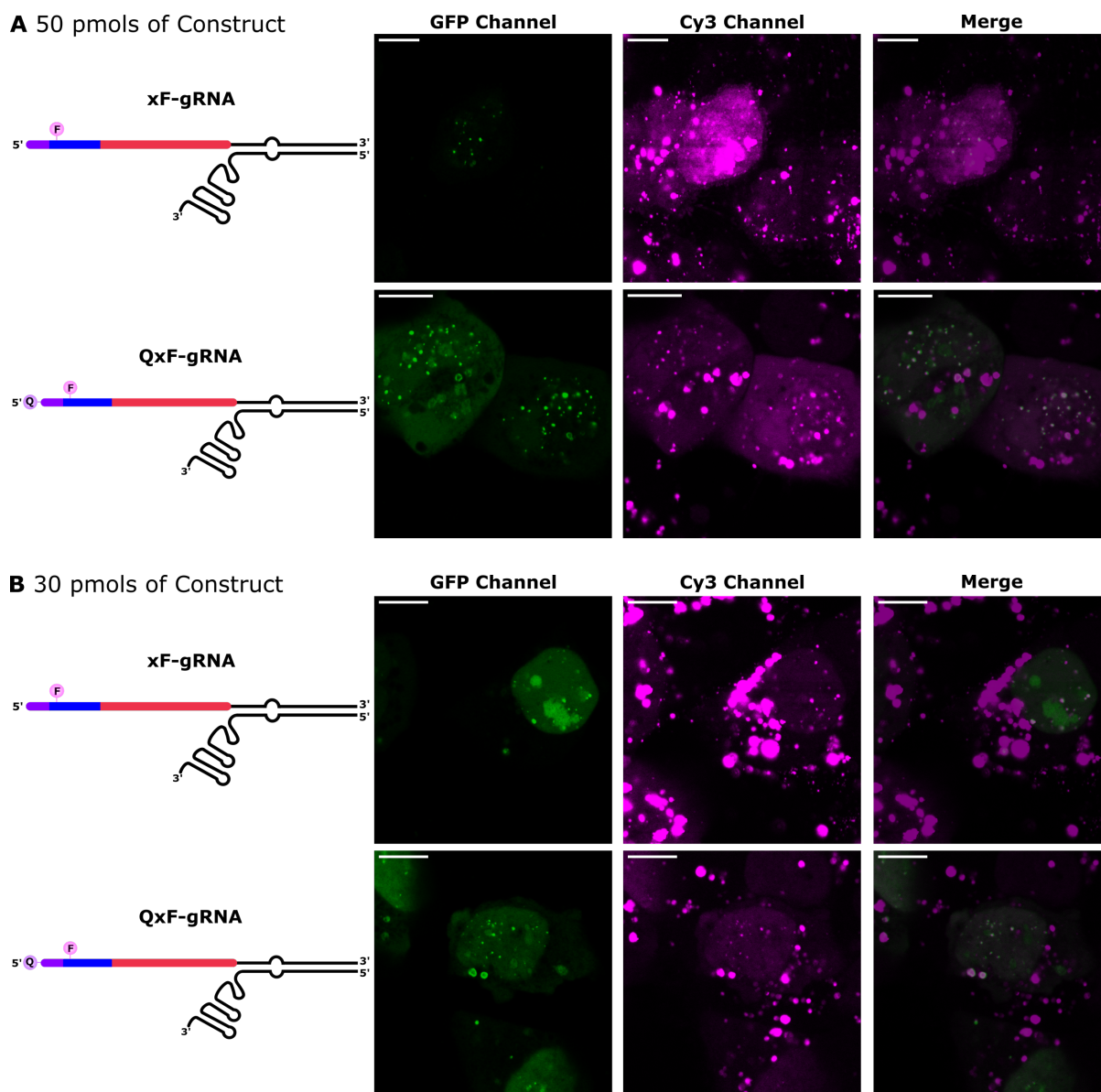


Figure 3.32: Live cell imaging of dCas9-3xGFP and 6:16 QxF-gRNA vs 6:16 xF-gRNA. Images were taken using confocal fluorescence microscopy of U2OS cells transfected with A) 50 pmols and B) 30 pmols of gRNA constructs and a dCas9-3xGFP plasmid. Identical excitation and emission settings were used for each image. Laser used to visualise GFP (green) and Cy3 (pink) (described in Section 5.6.4) as indicated. The co-localisation of both dyes can be seen by off-white/yellow spots in the merged images. Scale bar = 10 μ m.

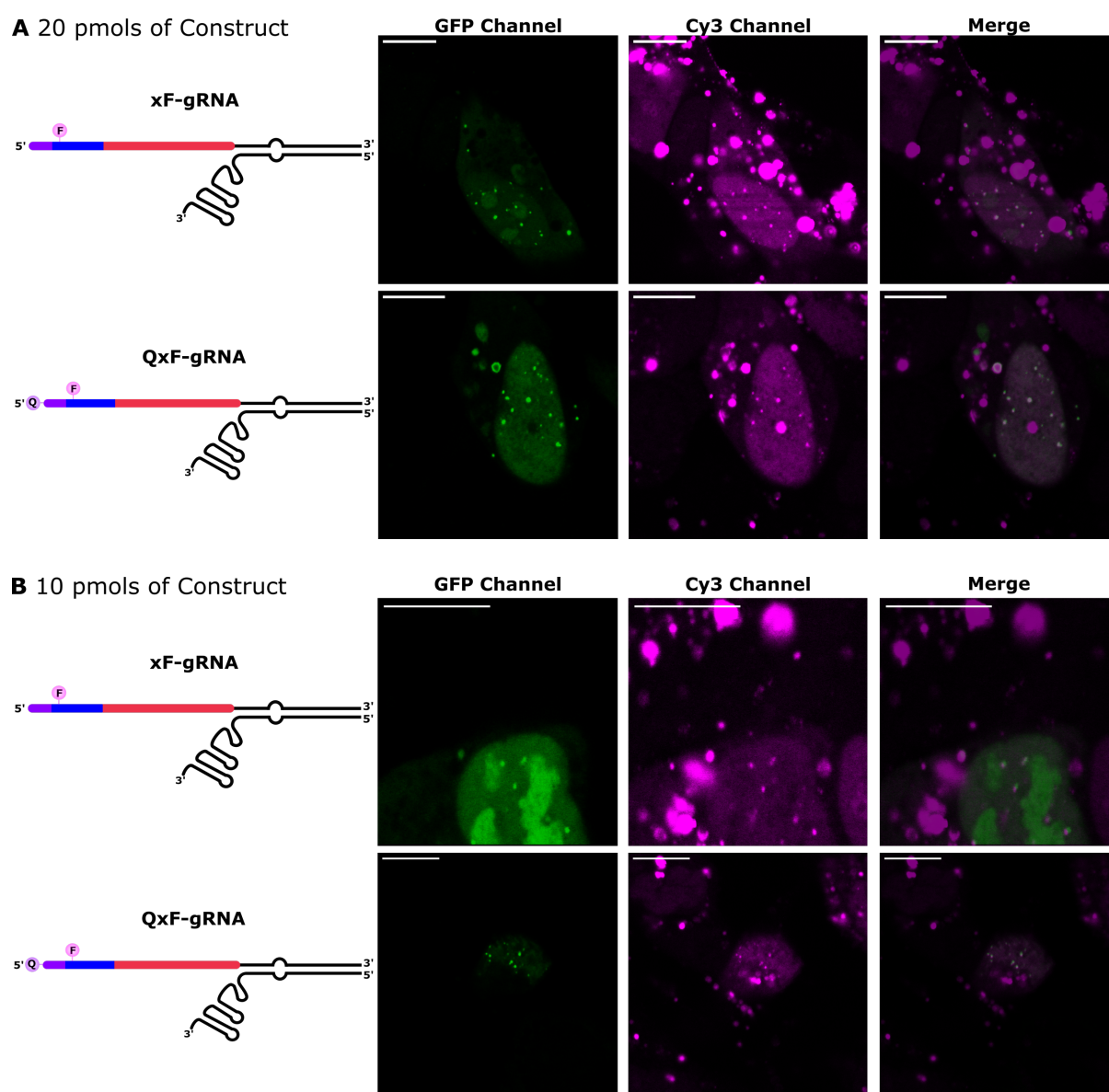


Figure 3.33: Live cell imaging of dCas9-3xGFP and 6:16 QxF-gRNA vs 6:16 xF-gRNA. Cells transfected with A) 20 pmols and B) 10 pmols of gRNA constructs. Images were captured and analysed as described in Figure 3.32. The co-localisation of both dyes can be seen by off-white/yellow spots in the merged images. Scale bar = 10 μ m.

The optimum quantity is 20 pmol (Figure 3.33 A); the 6:16 xF-gRNA construct has background fluorescence in both the nucleus and in the cytoplasm whereas with the QxF-gRNA construct, the background fluorescence is reduced in the nucleus. This highlights the need for the quencher in the construct. The puncta in both constructs are of similar intensity. While it is clear from Figure 3.32 - Figure 3.33 that the

quencher is cleaved from the QxF-gRNA, one experiment that could provide more evidence for RNase H activation is the use of a scrambled 5'-extension in the QxF-gRNA. This would ensure the RNA:DNA heteroduplex could not form and the recognition site for RNase H would not be available for cleavage. Another option is the use of DNA or 2'-O-Me modifications which should disrupt the RNA:DNA heteroduplex and help provide confirmation that RNase H is responsible for the removal of the quencher.

3.8 Conclusion and Future Work

The overall aim of this work was to develop an 'on/off' fluorescence system using modified gRNA that could allow for effective imaging of genomic DNA in live cells. To conclude, a fluorophore-quencher gRNA construct can be used for live cell imaging that reduces background fluorescence signal and retains on-target signals. The use of an 'on/off' system in the RNase H activated gRNAs provides an alternative approach to studies that use modified sgRNAs to recruit multiple FPs as these systems tend to aggregate and provide non-specific puncta.^{226,233} Since the RNase H activated system requires all components of the complex (gRNA, Cas9 enzyme and target DNA) before fluorescence enhancement is observed, this ensures the puncta evident in the nucleus are positive results. Other studies such as the use of a MB and a modified sgRNA by Wu *et al.*, contain a risk of false positive results as the MB and sgRNA can bind and cause fluorescence enhancement before meeting the target.²³⁸

In the course of these studies, the development of the design of the gRNA constructs provided insight into modifying the gRNA systems. It was determined that the DNA-targeting region of the crRNA can be modified and sufficiently maintain the DNA cleavage activity of Cas9 *in vitro*. A fluorophore can also be incorporated through these modifications without major disruption to the DNA cleavage activity of Cas9. The gRNA can also be extended through the 5'-end and mediated slightly better Cas9 DNA cleavage *in vitro* than a gRNA modified in the DNA-binding region.

More importantly, the 5'-extension gRNA can be further developed to contain a fluorophore and detachable quencher and can successfully bind with Cas9 *in vitro* and in cells. It can be used to image genomic DNA with reduced background fluorescence compared to gRNA containing no quencher. The detachable quencher can be successfully cleaved *in vitro*, however, further studies are needed to determine if RNase H is responsible for the removal of the quencher in the cell. As discussed previously, the use of modifications and/or a scrambled sequence in the 5'-extension of QxF-gRNA that would disrupt the RNA:DNA heteroduplex would provide more evidence for the RNase H activation of the QxF-gRNA. Further studies could also be carried out with the 5'-extension to understand if mismatches are tolerated or other modifications which would improve the synthesis of the gRNA.

Another development of QxF-gRNA could include tackling the cytoplasmic background seen in the cell images. This fluorescence background could be caused by degradation of unbound gRNA in the cell. The introduction of more quenchers into the system would provide an interesting study. Two quenchers would provide more steric hindrance at the 5'-end of the gRNA which could reduce the amount of degradation. It would also provide even more quenching to the gRNA. Since lysosome-trapping is a potential issue and one of the causes could be from lipid-transfection, a comparison with other delivery methods such as electroporation could provide more insight. Once the construct is fully developed, it could be investigated with non-telomeric targets such as *MUC4* and non-repetitive genome targets.

Chapter 4 Incorporation of Azides for RNA Synthesis

4.1 Background

RNA plays a fundamental role in many cellular processes, ranging from DNA transcription to cell differentiation and so, the biological functions of RNA and their potential applications has been studied in great detail. The field of molecular biology relies heavily on the efficient production of high-yielding, long oligonucleotides. This is not such an issue for DNA synthesis as optimisation of the automated solid-phase phosphoramidite method (described in Chapter 1) allows for the assembly of DNA ODNs up to ~150 bases in length.²⁶⁹ RNA synthesis, however, has an upper limitation of 70-100 bases in length with the standard chemical RNA synthesis methods.^{31,32,270} This can be an issue for CRISPR-Cas9 research as the synthesis of chemically modified sgRNA (usually 100+ nt in length) is central and non-chemical methods such as IVT cannot facilitate the incorporations of modifications.

4.1.1 RNA synthesis

The most common methods of RNA synthesis are *in vitro* transcription (IVT) and automated solid-phase synthesis. Whilst these methods are conventionally used in modern research labs, both methods come with their own limitations, which will be discussed in further detail below.

4.1.1.1 *In vitro* transcription (IVT)

In a biological system, RNA is made through a process called transcription. DNA acts as a template for the complementary RNA and enzymes called RNA polymerases carry

out the generation of new RNA. This system has been mimicked for *in vitro* applications which use T7, SP6 or T3 RNA polymerase, purified from bacteriophage, to produce RNA from a linear DNA template.²⁷¹ The template contains a DNA promoter sequence which forms a duplex with the complement T7 promotor and the polymerase specifically binds to this duplex to begin transcription. It reads from the DNA template to synthesise RNA using NTPs added to the mixture (Figure 4.1). Studies have shown that a ssDNA template is sufficient for successful transcription and the transcription process will terminate when the polymerase reaches the end of the DNA template (“run-off” transcription).²⁷² The product can be purified using commercially available kits and confirmation of the correct product must be carried out (*via* sequencing or mass spectrometry).

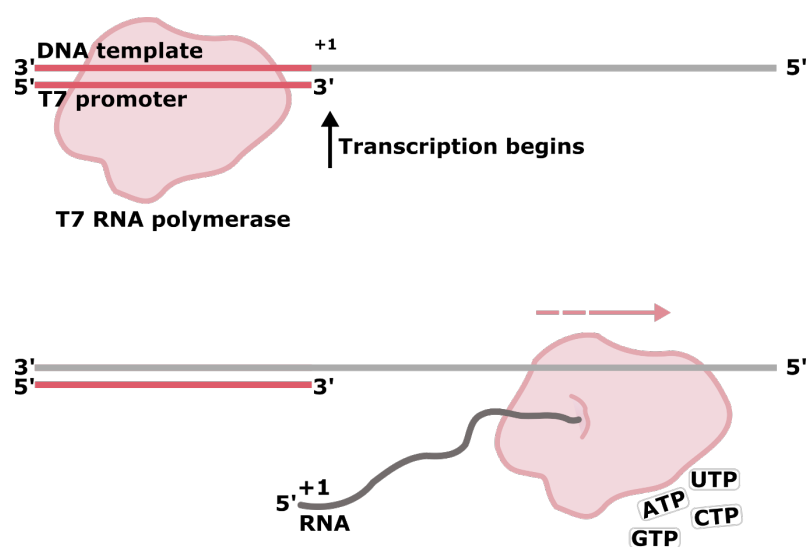


Figure 4.1: Mechanism of *in vitro* transcription. RNA polymerase binds to dsDNA duplex containing the T7 promoter and its complement in the DNA template. Transcription begins at position +1. RNA is transcribed from the ssDNA template by incorporating NTPs (A/U/G/C). Transcription finishes when the promoter reaches the end of the ssDNA template.

IVT can be used to synthesis RNA from as small as 30 nt to 10^4 nt but has clear limitations. The method requires the use of enzymes which can be expensive, and it

is difficult to scale up. It also requires a complete DNA template which requires additional synthesis prior to RNA synthesis and can also lead to issues when preparing long RNA (>150 nt). IVT requires the incorporation of the T7 promoter into the DNA template, extending the length of the template. Additionally, there is not complete freedom in the choice of the 5'-position of the RNA as the polymerase requires a specific nucleotide at the 5' of the RNA to initiate transcription. For example, T7 polymerase, class III, requires G in the +1, +2 and +3 position for efficient transcriptional yield. IVT can result in errors at both the 5'- and 3'-end of the RNA. Considering that the application of the modified RNA synthesis discussed in this chapter is for sgRNA for CRISPR-Cas9, errors to the 5'-end can lead to off-target binding, since this is the region of DNA hybridisation in the Cas9-gRNA complex. Finally, while studies have shown incorporation of modifications during IVT using chemically modified triphosphates is possible,²⁷³ the efficiencies tend to be lower and dependent on the polymerase ability to incorporate the dNTP into the growing RNA strand. There is also limited positional selectivity to add these modifications when the natural base is to be incorporated as well.

4.1.1.2 Solid-phase synthesis of RNA

RNA can also be prepared using automated solid-phase synthesis, as described in Chapter 1.²⁴ However, the yield and efficiency of RNA synthesis *via* this method does not correspond to that available for DNA. This is due to the presence of the reactive 2'-OH group. It requires an additional selective protecting group step in the synthesis

of the phosphoramidites, making them more expensive. The hydroxy group also increases RNAs susceptibility to degradation during synthesis. Migration of the phosphate backbone from the 3'-position to the available 2'-position (chain migration) or chain cleavage all lead to by-products, reducing the yield of the RNA synthesis and complicating its final purification (Figure 4.2). The coupling efficiencies between consecutive monomers are reduced in chemical synthesis and in RNA synthesis, the presence of the bulky protected 2'-OH group contributes to this. This results in more “failed sequences” and also limits the synthesis of RNA to ~70-100 nt in length, depending on what protecting group chemistry is used.^{31,32,270} Moreover, the choice of protecting group for RNA controls the chemistry used in the synthesis and some modifications are not compatible with certain reagents.

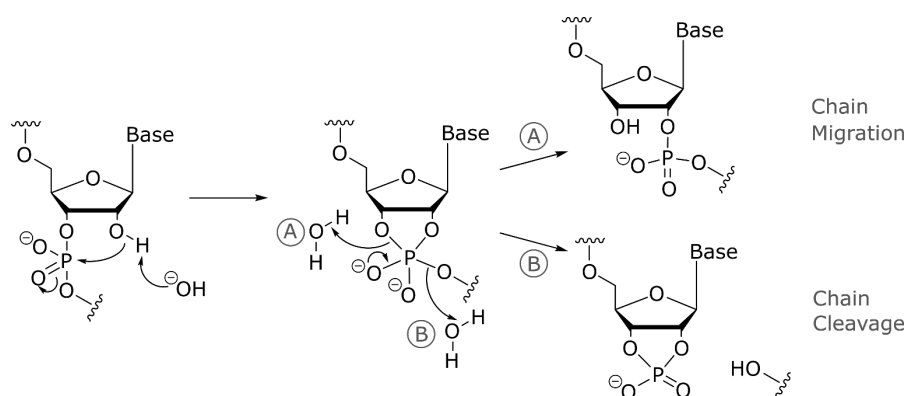


Figure 4.2: RNA degradation. Degradation of RNA chain through A) chain migration and B) chain cleavage in basic conditions.

Nonetheless, solid-phase synthesis has several advantages over IVT as it can be performed on larger scale, does not require a template or enzymes, and modifications can be easily incorporated at precise positions within the RNA or at the 3'- or 5'-end. However, the restriction on length and coupling efficiencies poses a problem. One

solution to the problem of length is to use a chemical linkage to ligate RNA covalently. Shorter RNA containing this reactive group can be produced with high efficiency and yields and ligated together to produce a larger construct needed for sgRNA.

4.1.2 Copper-catalysed Azide Alkyne Cyclo-addition (CuAAC)

The ligation of oligonucleotides has previously been studied using a triazole linkage between ODNs for bioconjugation.^{274,275} The cyclo-addition of an alkyne and azide (Copper-catalysed Azide Alkyne Cyclo-addition, CuAAC) to produce a triazole linkage is a common example of click chemistry. The reaction is catalysed by Cu(I) and is robust and high yielding. Originally, the cyclo-addition reaction produced two isomers and required high temperature²⁷⁶ but in the presence of Cu(I), the reaction can be directed to produce the 1,4-triazole isomer.²⁷⁷ The click mechanism is shown in Figure 4.3 A.

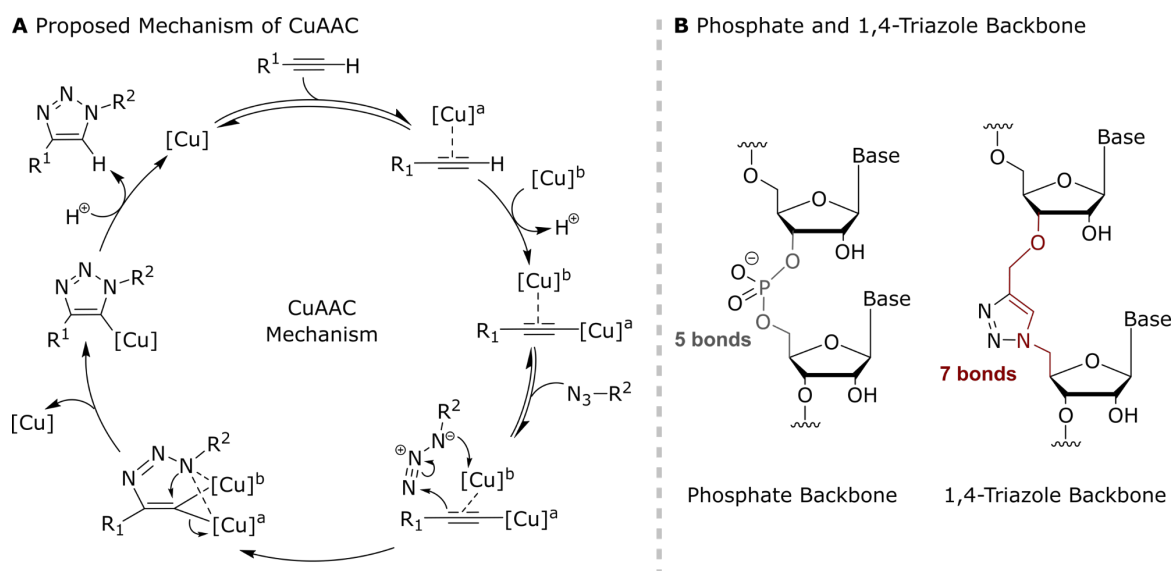


Figure 4.3: Mechanism of CuAAC and comparison of triazole linkage. A) Mechanism of CuAAC reaction to produce a 1,4-triazole isomer. B) Comparison of the natural phosphate backbone (5 bonds in grey) and 1,4-triazole linkage (7 bonds in red).

Pioneered by our group, extensive studies have been carried out using various versions of the triazole linkage as a substitute for the phosphate backbone.^{278–283} Our group particularly focused on the biocompatibility of a 7-bond triazole linker (Figure 4.3 B compares it with the phosphate backbone) which is tolerated in biological systems and can be read through by DNA and RNA polymerases *in vitro*.^{275,284–287}

Other ligation methods such as enzymatic ligation²⁸⁸ and condensation²⁸⁹ exist and have made sufficient progress but are restricted by the enzymes' compatibility with buffer, pH and salt concentrations. The efficiency and scalability of click chemistry is a desirable advantage of this linker.

4.1.3 CuAAC for CRISPR-Cas9 systems

The introduction of a triazole linkage into sgRNA has previously been explored by Taemaitree et. al.,²⁹⁰ who divided the sgRNA into two sections: the spacer region and the rest of the crRNA and tracrRNA. Their 'split and click' methodology then used a click linkage (a chemical triazole linkage) between the spacer region and the rest of the sgRNA to provide a high throughput method for synthesising, in high yields, multiple sgRNAs (Figure 4.4). The focus of the work was to provide efficient synthesis of multiple spacer RNAs and the production of only one crRNA/tracrRNA section. This could be used to produce the library of sgRNAs required to target non-repetitive genome sequences.

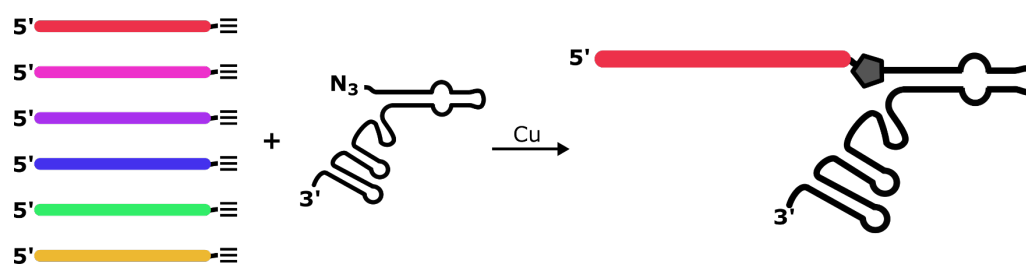


Figure 4.4: Split and Click concept. Taemaitree *et al.*²⁹⁰ used CuAAC to ligate alkyne functionalised spacer RNA (multi-coloured) with azide modified crRNA/tracrRNA (black) resulting in a triazole linkage (pentagon). This was used to produce a library of sgRNA in high efficiency and purity with applications in non-repetitive genome targeting.

We know from Dr. Taemaitree's study that the 1,4-triazole linkage is well tolerated by the CRISPR-Cas9 system and provides a suitable linkage between the sections of the sgRNA. However, the major limitation of this high-throughput methodology is the production of 5'-azide ODNs.

4.1.4 Azide incorporation

The incorporation of an azide into an ODN can be achieved using many methods. Enzymatic incorporation of azide-modified triphosphates has successfully been studied but difficulty to control the site of incorporation limits this method.²⁹¹ Azide-modified linkers can be successfully incorporated using post-synthetic labelling and modified nucleosides functionalised to the resin²⁹² all provide azide-modified ODNs but not at the 5'-position. However, to maintain the biocompatibility and minor changes to the phosphate backbone structure, the azide preferentially requires incorporation at the 5'-position of the ODN.²⁹³

While a phosphoramidite approach would provide a straight-forward method to incorporate an azide into an ODN at the 5'-position, it faces a major obstacle. Azides readily reacts with P(III) species, referred to as a Staudinger reaction (Figure 4.5),^{294,295} and so difficulty arises during the phosphitylation step, where a P(III) reagent is used to synthesise the phosphoramidite. Alternative methods to incorporate azides in solid-phase synthesis use a P(V) species in the nucleoside such as H-phosphonates^{296,297} or phosphotriesters.^{279,298–302} These methods provide an alternative to the phosphoramidite method by coupling the P(V)-nucleotide to the ODN.

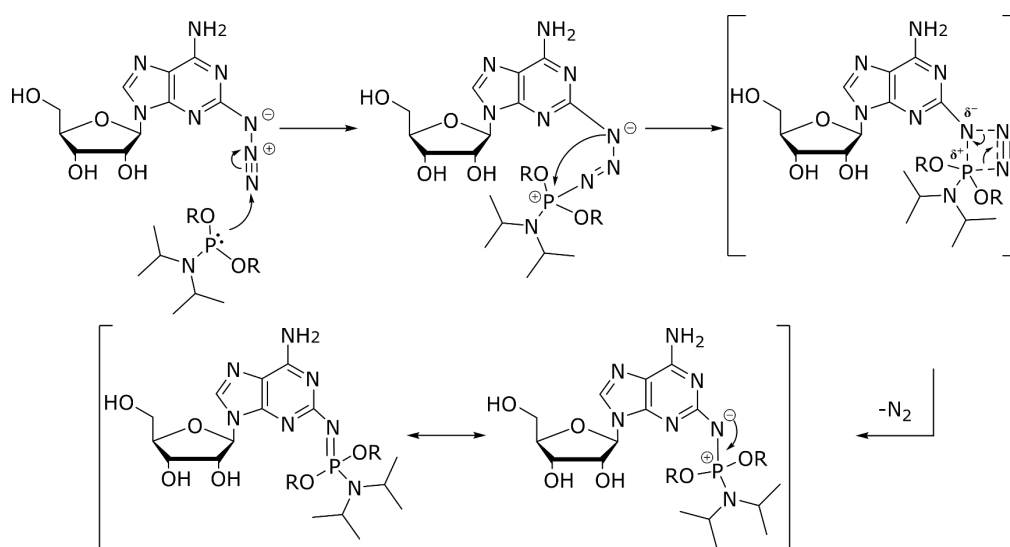


Figure 4.5: Staudinger reaction. Mechanism of Staudinger reaction between a P(III) species from the phosphitylating reagent and 2-azide-adenosine nucleoside. Adapted from Wada *et al.*²⁹⁵

In the “split and click” study, the 5'-OH was converted to the 5'-azide RNA on the solid support.^{275,290} Before the RNA is deprotected, the ODN, bound to the resin, is reacted with methyltriphenoxy-phosphonium iodide ($[(\text{CH}_3\text{P}(\text{OPh})_3]^+\text{I}^-)$ to convert the 5'-OH to 5'-I. This 5'-I RNA was then converted to 5'-N₃ using NaN₃ (Figure 4.6 A). This is followed by standard deprotection methods. However, this method has some

drawbacks. The conversion worked relatively efficiently with a 66-mer RNA synthesised using 2'-O-TBDMS chemistry.¹⁴ The conversion was also tried on RNA synthesised using 2'-O-TC chemistry. As previously discussed in Section 1.2.2, the use of 2'-O-TC chemistry for RNA synthesis provides better overall yields and is a preferred method for RNA synthesis, especially for the generation of longer RNA. However, it is reported by Sigma-Aldrich that the 2'-O-TC protecting group is liable to oxidation and can produce a carbamate protecting group that cannot be removed during deprotection (Figure 4.6 B).³⁰³ To perform the conversion of 5'-OH to 5'-I, 0.5 M of iodine is needed and, although this does not cause major degradation when the RNA is shorter (10-mer), it causes complete degradation and incomplete deprotection of the RNA when attempted on a longer RNA (66-mer). For this reason, a new approach to incorporate a 5'-azide moiety into RNA that is compatible with 2'-O-TC RNA is needed.

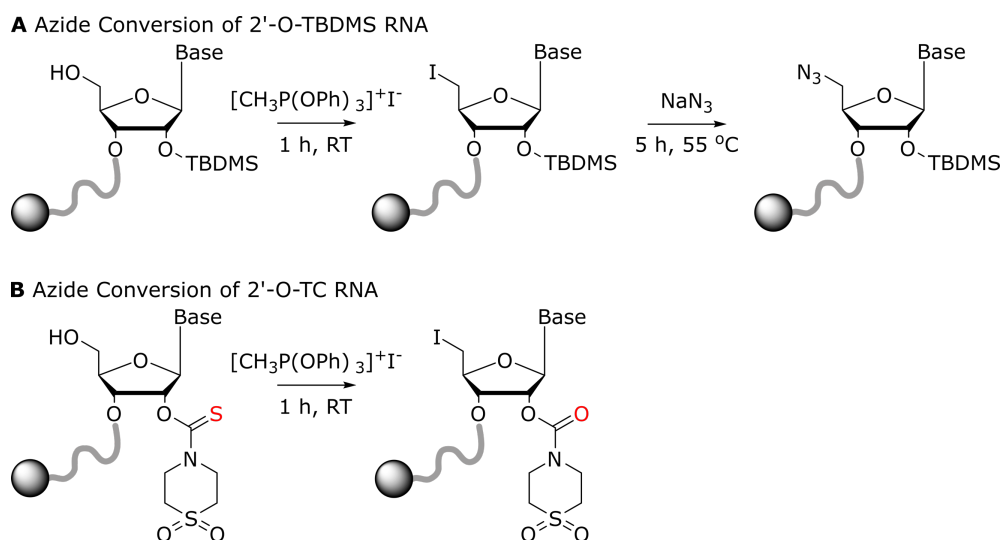


Figure 4.6: Conversion of 5'-OH RNA to 5'-azide RNA using different 2'-protecting groups. A) Azide conversion reaction conditions with 2'-O-TBDMS RNA and B) Unwanted oxidation of 2'-O-TC group to a carbamate group (red) during iodo-conversion of 2'-O-TC RNA. Azide conversion not completed as the carbamate protecting group cannot be removed during deprotection.

4.2 Hypotheses and Aims

The efficient synthesis of long, modified RNA is fundamental, particularly in CRISPR research. The introduction of an azide moiety into RNA to provide a means of ‘clicking’ modified RNA strands together is a desired approach. However, this approach is limited by the protecting groups used for RNA which, in turn, hinders the length of efficient RNA synthesis. The aim of this chapter is to develop an efficient synthesis of an azide RNA monomer that is compatible with 2'-O-TC RNA synthesis. This could provide a simple and cost-effective route to modified gRNA synthesis. We have already seen a potential application of this approach in the ‘split and click’ studies and this could be extended to many other bioconjugation applications or ligation applications since the triazole ligation linker is biocompatible.

To produce the 5'-azide RNA, a variety of different approaches will be undertaken: a phosphotriester method, a “reverse phosphitylation” method and a phosphoramidite method. The H-phosphonate approach was not undertaken as this reaction entails an oxidation step. To incorporate the H-phosphonate into an ODN, an oxidation step is required after the coupling step to produce the $[\text{PO}_4]^-$ backbone between the ODN and the monomer. As we have seen from the 5'-OH to 5'-I conversion, this oxidation causes difficulties when using 2'-O-TC RNA chemistry and, therefore, was not attempted.

4.3 Results

4.3.1 Phosphotriester nucleosides

Phosphotriesters provide an alternative approach for the synthesis of modified ODNs^{23,304} where the modification cannot be incorporated using the standard phosphoramidite method. Many previous studies have taken advantage of this alternative approach to incorporate an azide at the 2'-position^{299,300,302} and 3'-position²⁷⁹ of DNA and RNA. Another study has also incorporated an azide at the 5'-position of the sugar in DNA.²⁹³ Based on this article, a 5'-azido RNA monomer protected with 2'-O-TC was envisioned.

Since the aim is to introduce just one azide at the 5'-end of the RNA, the ODN can be synthesised using standard 5'-DMT-phosphoramidites (since the solid-phase synthesis builds the ODN from the 3' to the 5'). Here, the final step of the assembly would be the addition of the azide-containing phosphotriester *via* a phosphotriester coupling. This takes advantage of the automation and speed of the solid-phase synthesis for the rest of the RNA synthesis.

4.3.1.1 5'-Azido DNA phosphotriester

The synthesis of the phosphotriester was initially tested on a DNA monomer to confirm effective coupling before considering RNA. 5'-azido-deoxythymidine (**12**) phosphotriester was synthesised from 5'-azido-deoxythymidine (**11**) following a published protocol (Figure 4.7).^{279,293} The reaction provided 82% of the pure product.

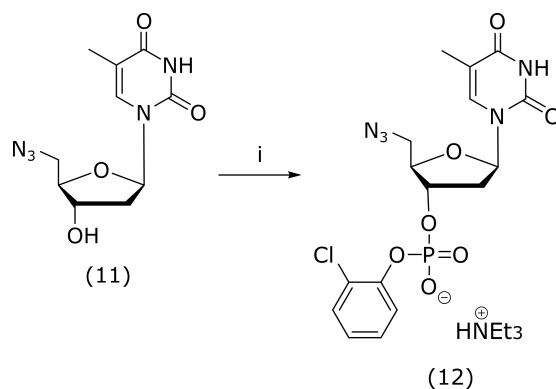


Figure 4.7: Synthesis of 5'-azido dT phosphotriester, 12. Reaction was carried out for 1 h in anhydrous THF. i) 1,2,4-triazole (5.5 eq), NEt_3 (5 eq) and 2-chlorophenyl phosphorodichloridate (2.5 eq) were reacted for 30 min before compound **11** was reacted with 1-methylimidazole (4 eq) and added to the solution to give 82% of compound **12**.

The DNA phosphotriester **12** was then coupled to a fully protected 10-mer DNA and 11-mer RNA immobilised on CPG resin (Figure 4.8). Initially, 1 eq of compound **12** was used, similar to the conditions used in the paper. An excess of pyridine was used to ensure the entire resin was immersed in the reacting solution. This resulted in no effective coupling, so the reaction conditions were altered.

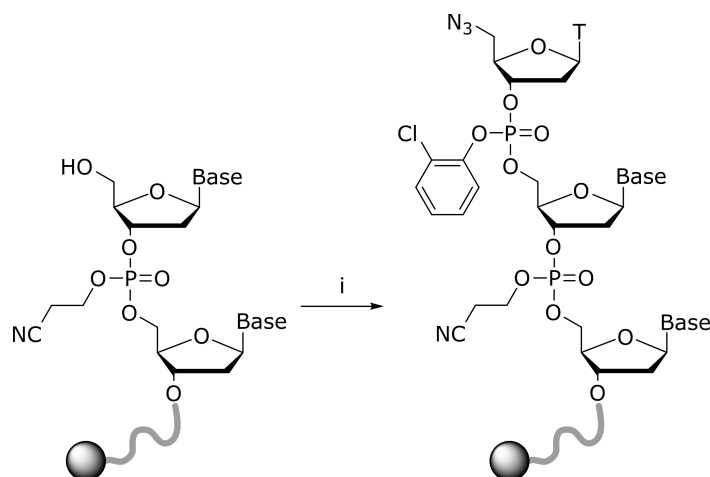


Figure 4.8: Phosphotriester coupling. i) The phosphotriester triethylamine salt (**12**) is reacted with an ODN attached to the resin in the presence of MSNT (5 eq) and 1-methyl-imidazole (8.4 eq). Following steps include standard deprotection of the ODN to remove nucleobase protecting groups and cleave the ODN from the resin for analysis.

Next, 2 eq of the phosphotriester **12** was added to the column and left for 2 h. This also showed no coupling, so 4 eq of compound **12** was reacted. The reaction time was also increased to overnight. This resulted in coupling of the phosphotriester to the 10-mer DNA (Figure 4.9 A), confirmed by UPLC-MS. However, the presence of the starting material (SM: 10-mer DNA) highlights the low efficiency of the coupling. No coupling of the monomer could be seen to the 11-mer RNA ODN (Figure 4.9 B) and only SM was observed. For analysis purposes, the protecting group of the phosphotriester was not removed but we anticipate that this could be done following reported protocol.²⁹³

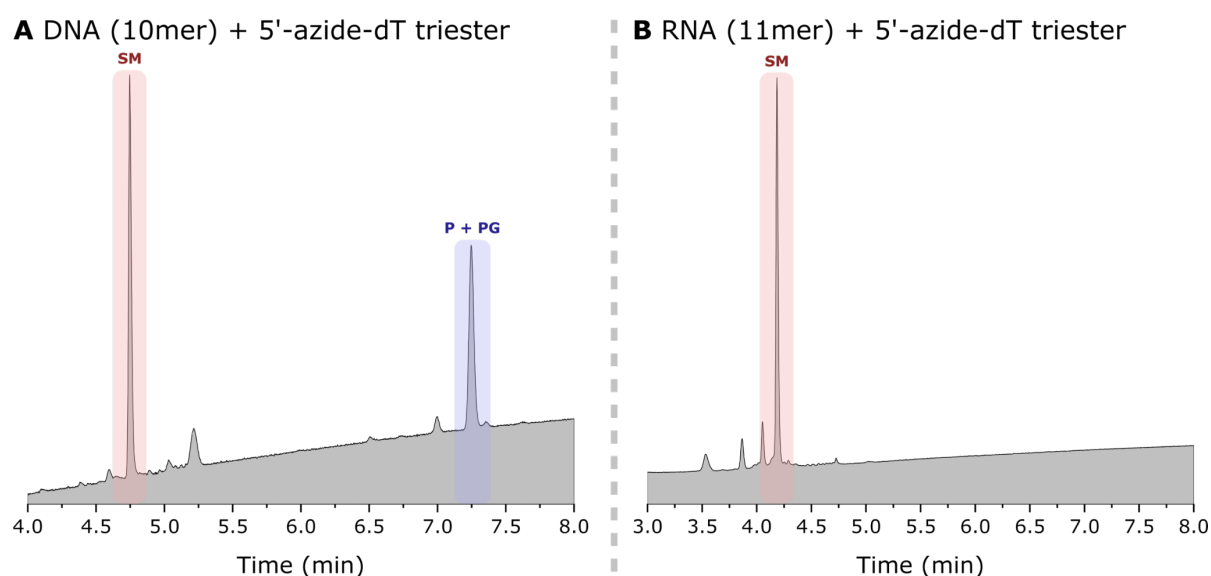


Figure 4.9: UPLC analysis of 5'-azide dT phosphotriester coupled to DNA/RNA taken from UPLC-MS. A) UPLC of DNA (10-mer) coupled with 5'-azide dT phosphotriester. B) UPLC of RNA (11-mer) coupled with 5'-azide dT phosphotriester. Expected masses for A: starting material = 2980, product + protecting group = 3420. Masses observed for A: 2979.5 (SM), 3419.5 (P + PG). Expected masses for B: SM = 3328, P + PG = 3768. Masses observed for B: 3328.2 (SM).

4.3.1.2 5'-DMT DNA phosphotriester

To confirm that the azide moiety was not an issue during the coupling, a standard 5'-DMT-deoxythymidine phosphotriester, was synthesised (similar to Figure 4.7). Successful synthesis of the compound provided 62% of the pure product. To ensure moisture did not affect the coupling reaction, the resin and all the reagents were dried overnight at 0.1 mbar. The reaction was also attempted in an Eppendorf allowing a smaller reaction volume thereby increasing the concentration of the reacting phosphotriester.

2 eq of 5'-DMT-deoxythymidine phosphotriester was used (0.27 M compared to 0.15 M used in the article). The reaction was left for 1 h with a DNA (10-mer) resin, similar to previous reactions. This reaction was unsuccessful and only the SM peak was seen in the UPLC-MS (Figure 4.10 A). Two more reactions using 0.14 M of 5'-DMT-deoxythymidine phosphotriester were also carried out (on a column) for 2 h and overnight, respectively. However, none of the reactions successfully incorporated the phosphotriester and only led to the production of an impurity with a mass of SM +67 (Figure 4.10 B and C).

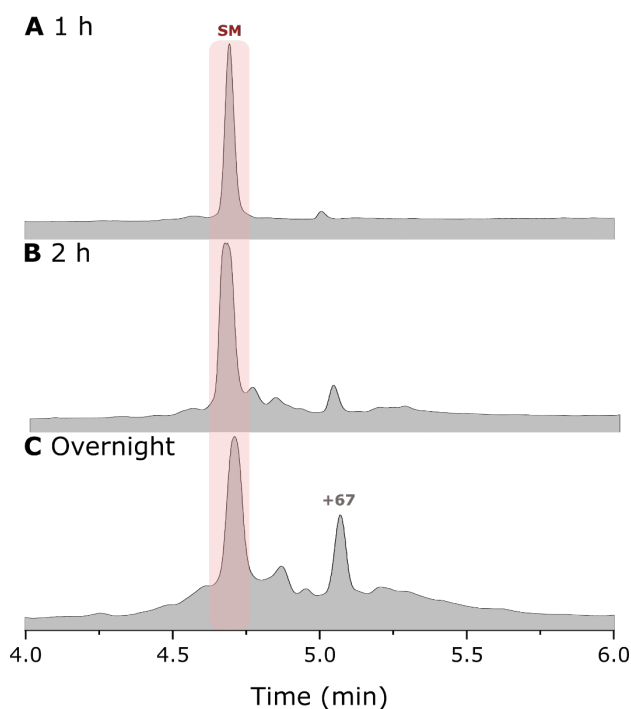


Figure 4.10: UPLC analysis of 5'-DMT dT phosphotriester coupled to DNA taken from UPLC-MS. UPLC of DNA (10-mer) coupled with 5'-DMT dT phosphotriester for A) 1 h, B) 2 h and C) overnight. Expected masses for A/B/C: starting material = 2954, product + protecting group = 3369. Masses observed: 2952.5 (SM), 3019.5 (impurity).

Since the phosphotriester approach was unsuccessful for most reactions and only produced a low coupling efficiency for one reaction, another approach to the synthesis of a 5'-azide RNA ODN was attempted.

4.3.2 Reverse phosphitylation

Since the Staudinger reaction causes issue with incorporating an azide into a phosphoramidite directly, an alternative approach was investigated in which the phosphitylation was attempted on the ODN on solid support. This would mean the 5'-position of the ODN would be phosphitylated to produce an ODN with a 5'-P(III) reactive species that could then be reacted with a 5'-azide-3'-OH monomer to form a phospho-linkage between the 5'-azide monomer and the ODN (Figure 4.11).

Separating the azide and P(III) species until the reaction should reduce the amount of side reactions caused from both intra- and intermolecular Staudinger reaction.

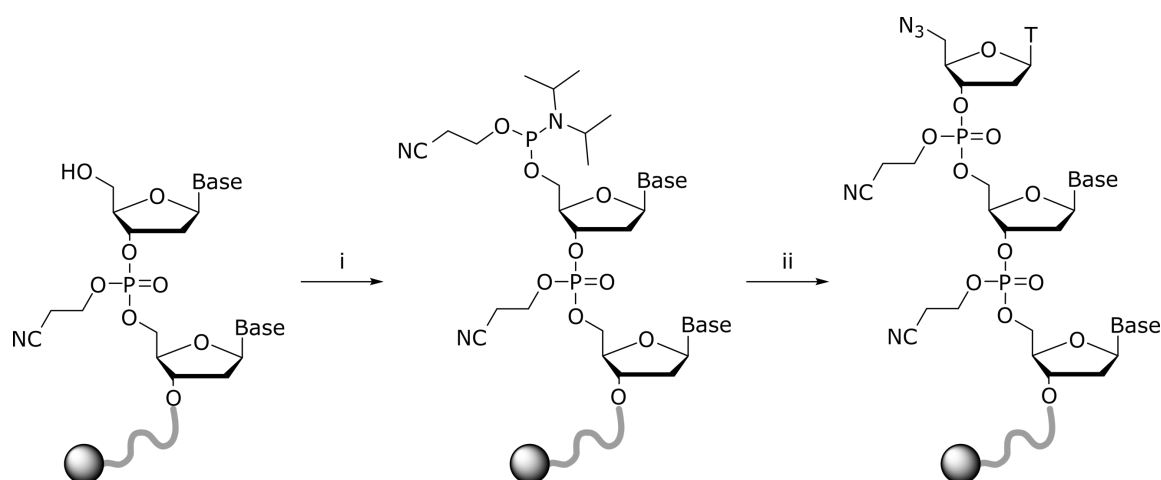


Figure 4.11: Schematic of “reverse phosphitylation” of an ODN on the resin. i) The free 5'-OH on the resin bound-ODN is reacted with a phosphitylating reagent and base to produce a reactive 5'-phosphite. ii) This is then reacted with the free 3'-OH of the 5'-azide nucleoside (**11**) and tetrazole to couple the monomer to the ODN. Following steps include standard oxidation and deprotection of the ODN to remove nucleobase protecting groups and cleave the ODN from the resin for analysis.

Many conditions for ‘reverse phosphitylation’ exist.^{146,305–309} In this study, different variables were tested which include reaction flask, reaction solvent, equivalents, concentration, phosphitylating reagent, reaction time and coupling reagents. For simplicity, each reaction is defined with a letter-number system (e.g. **A1**).

To start, a model system using a commercially available 5'-DMT-protected thymidine monomer and an immobilised DNA ODN (10-mer) was used to identify the best reaction time and solvent for the reaction (Table 4.1). The reactions were carried out on a column containing the CPG resin which was dried in a desiccator overnight prior to the reaction. The solvents and reagents used in the reaction were dried over 3 Å molecular sieves and bubbled with Argon prior to use. Reactions **A1** and **A2** were

carried out in DMF and reactions **B1** – **B4** were carried out in MeCN. The reactions were left for varying times according to the table and 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (Cl-CEP) was used as a phosphitylating reagent. The reacted ODNs were deprotected using standard methods and analysed by UPLC-MS (Appendix A4.1). However, the results showed no coupling of the monomer.

	Flask	Solvent	Base	CEP	Conc (M)	Vol (uL)	Time	Monomer	Solvent	Conc (M)	Vol (uL)	Time	Prod
A1	Column	DMF	DIPEA	Cl-CEP	0.1	500	45 min	DMT-T	MeCN	0.1	500	30 min	N
A2					0.1			DMT-T	DMF	0.1		o/n	N
B1	Column	MeCN	DIPEA	Cl-CEP	0.1	500	15 min	DMT-T	MeCN	0.1	500	o/n	N
B2					0.1		45 min	DMT-T		0.1		30 min	N
B3					0.1		45 min	DMT-T		0.1		o/n	N
B4					0.1		2 h	DMT-T		0.1		o/n	N

Table 4.1: Reaction conditions for “reverse phosphitylation” for reaction time and solvent. Reactions **A1**, **A2** and **B1** – **B4**. Cl-CEP = 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite, o/n = overnight, Prod = successful incorporation of the monomer (Y/N = Yes/No).

For reactions **B1**, **B2**, **B3** and **B4** (Figure 4.12 A-D), the starting material (SM) mass was 2954. The expected coupled product would have a mass of 3258. However, the products formed from these reactions had masses of SM +80, +64 and +164. The expected product structures of these impurities are theorised in Figure 4.12 E and are suspected to be incomplete reaction intermediates or oxidised products. A major impurity, with a molecular weight of 3034 (which corresponds to the SM +80), was observed for all the reactions. We hypothesise that this is the oxidised, phosphitylated ODN which cannot react with the monomer (Figure 4.12 E). The other impurities could also form from oxidation (+64) and/or incomplete protonation (+164) of the phosphitylated ODN. However, these results did suggest that a reaction time of at least 45 min was required for a substantial conversion of the starting material.

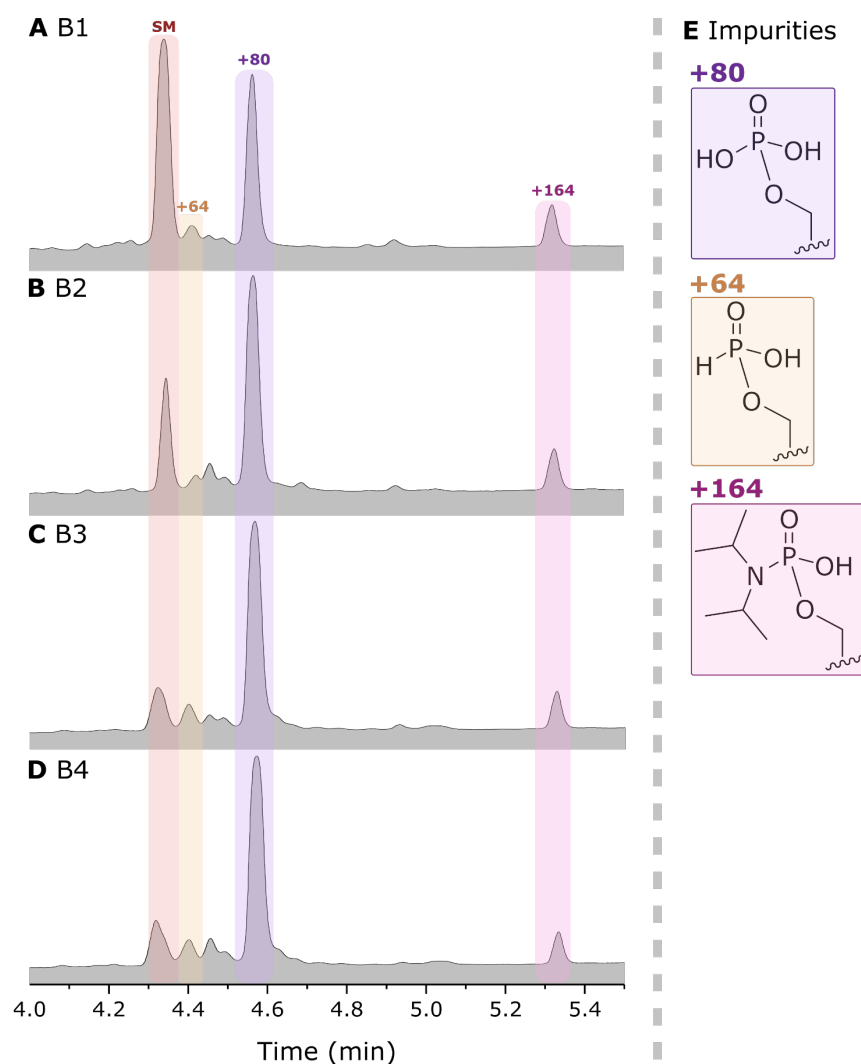


Figure 4.12: UPLC analysis of “reverse phosphitylation” of DNA (10-mer) and 5'-DMT-dT taken from UPLC-MS and structures of impurities. A) **B1**: phosphitylation reaction time = 15 min, B) **B2**: phosphitylation reaction time = 45 min, C) **B3**: phosphitylation reaction time = 45 min and D) **B4**: phosphitylation reaction time = 2 h. Expected masses for **B1 - B4**: starting material = 2954, product = 3258. Masses observed: 2953.7 (SM), 3017.7 (+64), 3033.7 (+80), 31117.7 (+164). E) Potential structures of other peaks seen in UPLC corresponding to the SM mass + value.

Next, a more concentrated reaction was attempted to help encourage both steps of the reactions and it was carried out in a vial to ensure a sealed argon system (Table 4.2). For reactions **C1-C3**, the SM mass was 3283. The expected coupled product would have a mass of 3586. We can see, in Figure 4.13, the production of some of the product.

	Flask	Solvent	Base	CEP	Conc (M)	Vol (uL)	Time	Monomer	Solvent	Conc (M)	Vol (uL)	Time	Prod
C1		DCM	NMM	Cl-CEP	1	200	45 min	DMT-T		0.5	200		Y
C2	Vial	MeCN	NMM	Cl-CEP	1	200	45 min	DMT-T	MeCN	0.5	200	o/n	Y
C3		MeCN	Tetrazole	Bis-CEP	0.5	500	3 h	DMT-T		0.5	500		N

Table 4.2: Reaction conditions for “reverse phosphitylation” for reaction concentration. Reactions **C1** – **C3**. Cl-CEP = 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite, Bis-CEP = 2-cyanoethoxy-bis(*N,N*-diisopropylamino)-phosphine, o/n = overnight, Prod = successful incorporation of the monomer (Y/N = Yes/No).

In **C1** (Figure 4.13 A), using DCM as a solvent provided more impurities whereas in **C2**, less impurities along with the product can be seen (Figure 4.13 B). Another phosphitylating reagent was also used, 2-cyanoethoxy-bis(*N,N*-diisopropylamino)-phosphine (bis-CEP) in **C3**, but this did not provide a product peak. Therefore, **C2** reaction conditions were used as a basis for the following reactions.

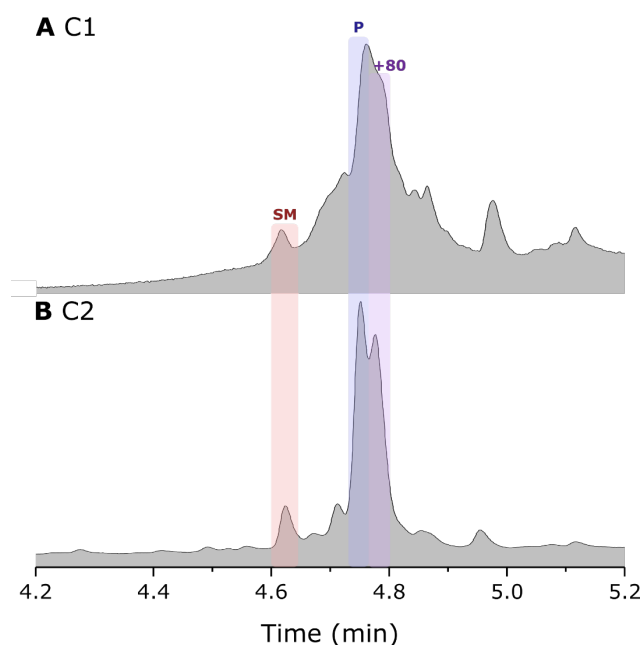


Figure 4.13: UPLC analysis of “reverse phosphitylation” of DNA (11-mer) and 5'-DMT-dT taken from UPLC-MS. A) **C1**: phosphitylation solvent = DCM, B) **C2**: phosphitylation solvent = MeCN. Expected masses for **C1** – **C2**: starting material = 3283, product = 3586. Masses observed: 3283.0 (SM), 3587.0 (P), 3363.0 (SM + 80).

Although product formation was seen in reaction **C2**, it was not very efficient so the use of a different monomer for the reaction was attempted (Table 4.3). However, the use of 5'-DMT-deoxycytidine in **D1**, provided a decrease in product formation (product mass = 3571) (Figure 4.14 B) compared to **C2** which used 5'-DMT-dT (Figure 4.14 A).

	Flask	Solvent	Base	CEP	Conc (M)	Vol (uL)	Time	Monomer	Solvent	Conc (M)	Vol (uL)	Time	Prod
D1					1		45 min	DMT-C		0.5		o/n	Y
D2					1		15 min	DMT-C		0.5		1 h	Y
D3	Vial	MeCN	NMM	Cl-CEP	1	200	15 min	DMT-C	MeCN	0.5	200	30 min	Y
D4					1		5 min	DMT-C		0.5		1 h	Y
D5					1		5 min	DMT-C		0.5		10 min	Y

Table 4.3: Reaction conditions for “reverse phosphitylation” for 5'-DMT dC. Reactions **D1 – D5**. Cl-CEP = 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite, o/n = overnight, Prod = successful incorporation of the monomer (Y/N = Yes/No).

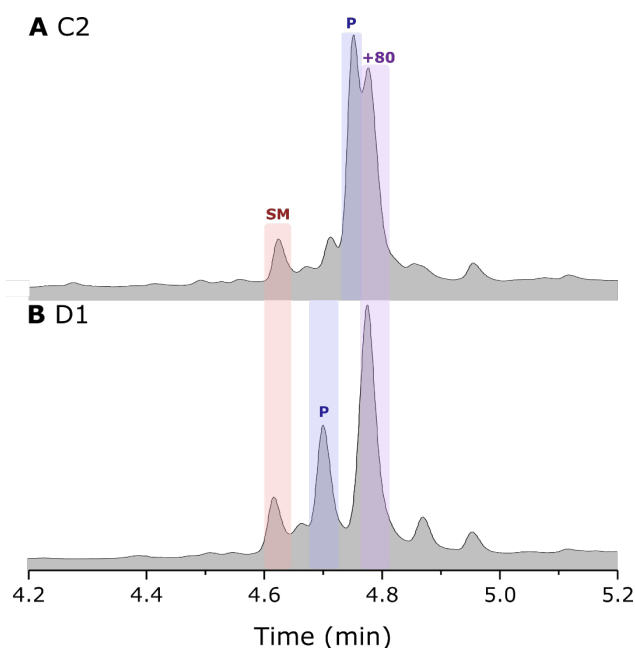


Figure 4.14: UPLC analysis of “reverse phosphitylation” of DNA (11-mer) and 5'-DMT-dT or 5'-DMT-dC taken from UPLC-MS. A) **C2**: monomer = 5'-DMT-dT, B) **D1**: monomer = 5'-DMT-dC. Expected mass for **C2**: starting material = 3283, product = 3586. Masses observed for **C2**: 3283.0 (SM), 3587.0 (P), 3363.0 (+80). Expected mass for **D1**: SM = 3283, P = 3571. Masses observed for **D1**: 3282.5 (SM), 3571.2 (P), 3362.5 (+80).

These results were interesting because both monomers have the same accessibility to the 3'-OH and therefore, would be expected to couple with similar efficiencies. The solubility of both monomers may cause an effect on the coupling efficiency. To further investigate this, varying the reaction times was also investigated (Table 4.3) but this seemed to be ineffective to the product peak with 5'-DMT dC in **D2** – **D5** (Figure 4.15 A-D).

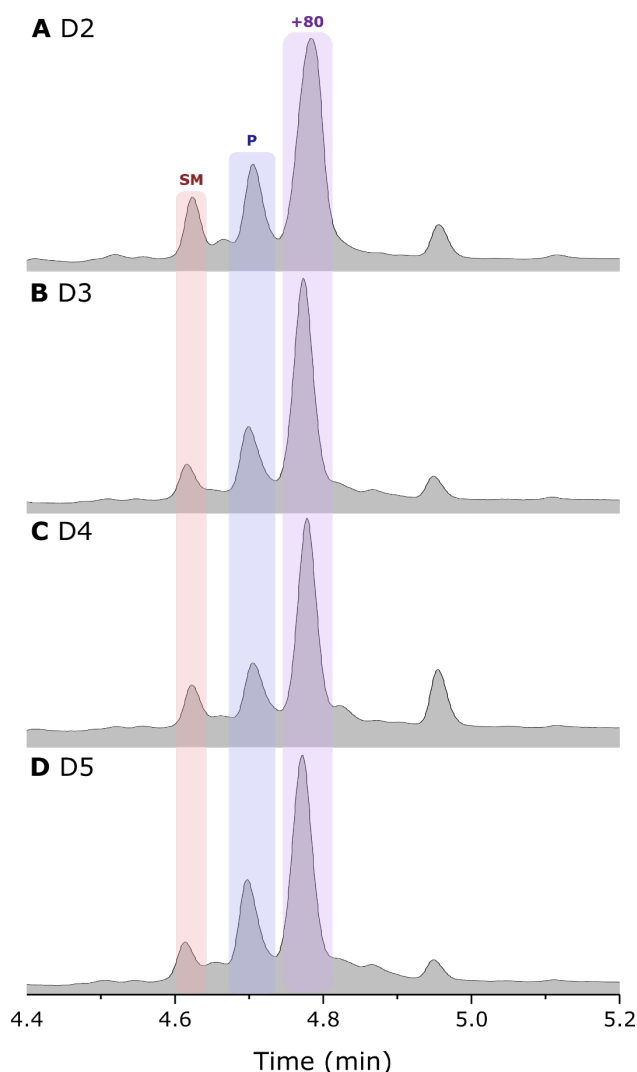


Figure 4.15: UPLC analysis of “reverse phosphitylation” of mixed DNA (11-mer) and 5'-DMT-dC taken from UPLC-MS. A) **D2**: phosphitylation reaction time = 15 min, coupling reaction time = 1 h, B) **D3**: phosphitylation reaction time = 15 min, coupling reaction time = 30 min, C) **D4**: phosphitylation reaction time = 5 min, coupling reaction time = 1 h and D) **D5**: phosphitylation reaction time = 5 min, coupling reaction time = 10 min. Expected masses for **D2** – **D5**: starting material = 3283, product = 3571. Masses observed: 3282.5 (SM), 3571.0 (P), 3362.5 (+80).

Since the reaction seemed to be dependent on the monomer and potentially, the solubility of the monomer, 5'-azido-deoxythymidine (**11**) was also attempted (**E - G** reactions, Table 4.4). However, all the reaction conditions resulted in no product formation for 5'-azido-deoxythymidine (**11**). The monomer was not very soluble in the tetrazole solution and so larger volumes of reaction solution needed to be used. Next, a mixture of MeCN and DMF (2:1) as a solvent was attempted as well as heating the reaction (**F** reactions, Table 4.4). Bis-CEP was also attempted (**G** reactions, Table 4.4) but all conditions were unsuccessful. The conditions were also attempted on 5'-DMT dT to encourage more product formation compared to previous reactions but no increase in product was seen for these reactions.

	Flask	Solvent	Base	CEP	Conc (M)	Vol (uL)	Time	Monomer	Solvent	Conc (M)	Vol (uL)	Time	Prod
E1	Vial	MeCN	NMM	Cl-CEP	1	200	45 min	DMT-T	MeCN	0.2	600	2 h	Y
E2				Cl-CEP	1			Azido-T		0.2	700		N
F1	Vial	MeCN	NMM	Cl-CEP	0.5	300	45 min	DMT-T	MeCN DMF	0.4	300 2:1	o/n	Y
F2					0.5			Azido-T		0.4			N
F3					1	200	45 min	DMT-T		0.4			Y
F4					1		40 °C	DMT-T		0.4			Y
G1	Column	MeCN	Tetrazole	Bis-CEP	3.1	200	2 h	Azido-T	MeCN	0.1	250	2 h	N
G2					3.1			Azido-T	DMF	0.4	1.5:1	o/n	N

Table 4.4: Reaction conditions for “reverse phosphitylation” for reaction concentration. Reactions **E - G**. Cl-CEP = 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite, Bis-CEP = 2-cyanoethoxy-bis(*N,N*-diisopropylamino)-phosphine, o/n = overnight, Prod = successful incorporation of the monomer (Y/N = Yes/No).

The solubility of the monomer is a determining factor for the efficiency of the coupling and a balance between solubility and low reaction volume is needed for efficiency of the reaction. 5'-azide dT (**11**) is the most difficult to dissolve of the monomers used and results in no product formation. After the lack of success of the reverse phosphitylation approach, a final approach towards incorporating azide into

ODNs was attempted. However, if more time was available, the use of another tetrazole solution or a wider use of solvent mixtures could help increase the coupling efficiency.

4.3.3 Azide phosphoramidite

As previously discussed, the incorporation of an azide into a phosphoramidite is rarely studied due to concerns relating to an undesired Staudinger reaction between the phosphitylating reagent and the nucleoside containing the azide (See Figure 4.5). However, recent studies have provided a solution to this issue. Shmanai *et al.* incorporated an azide into a phosphoramidite using hydroxypyrrolidines.³¹⁰ Whilst this modification did not contain a nucleobase, the pyrrolidine moiety mimics a ribose unit. Following this study, Lingala *et. al.* also incorporated an azide into a phosphoramidite (modifying the nucleobase to contain the azide) by following similar reaction conditions to the previous work.³¹¹

The phosphoramidite method would provide the most straight-forward method of azide incorporation so this approach was attempted next. Some key conditions were determined for the synthesis of azido-phosphoramidites from the previous studies.^{310,311} Concentrating or evaporating the compound during the reaction always led to decomposition and because of this, the phosphoramidite would also need to be stored in solution after the reaction or used immediately. The use of certain solvents such as MeCN also encouraged decomposition so the optimum solvent used for

storage was determined to be benzene. This provided a premise for the next attempt for azide incorporation.

4.3.3.1 5'-Azido deoxythymidine phosphoramidite

To confirm 5'-azide nucleosides could be incorporated using the methods described by Lingala *et al.*,³¹¹ the synthesis of the azide phosphoramidite was initially tried with 5'-azide-deoxythymidine (**11**, Figure 4.16). The reaction was carried out in CHCl_3 as compound **11** was not soluble in benzene. Excess equivalents of the phosphitylating reagent (2 eq) were used to support a faster reaction time. Once the reaction was complete, as judged by TLC, the reaction mixture was washed with KCl, and the organic layer was directly loaded onto the column without solvent evaporation. Following column chromatography, the solution was concentrated at low temperatures (25 °C) and re-dissolved in benzene (10 x 5 mL), taking care not to evaporate the product to dryness, to produce product (**13**).

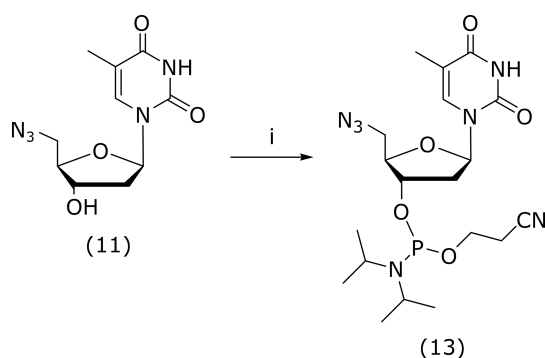


Figure 4.16: Synthesis of 5'-azido-deoxythymidine phosphoramidite, 13. Reaction was carried out for 1 h in anhydrous CHCl_3 . i) Compound **11** was reacted with 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (2 eq) and DIPEA (5 eq). CHCl_3 was replaced with C_6H_6 for use on the synthesiser and for storage. NMR analysis was not performed as the compound decomposed when concentrated, as determined by TLC. Purity was confirmed by UPLC-MS after successful incorporation into oligonucleotides. See Appendix A4.1.

Next, we evaluated the stability of phosphoramidite **13**. To do this, **13** was used to incorporate an azide at the 5'-end of an ODN straight after column chromatography (0 h) and this was compared to phosphoramidite **13** used for the same synthesis after being left for 5 h at RT (5 h). Both a 11-mer DNA (5'-azide-T₁₁) and 11-mer DNA (5'-azide-A/G/C/T) were synthesised. Successful incorporation of the azide phosphoramidite **13** was confirmed by UPLC-MS (Figure 4.17 A and B) for both ODNs. Very minimal amounts of the undesired products were seen, and incorporation was greater than 90%.

As shown in Figure 4.17 A, **13** is not very stable at RT for prolonged periods of time. A number of reactions can occur during this time to produce more of the impurities seen in 5 h compared to 0 h. The P(III) species in the phosphoramidite can react with an azide, inter- or intramolecularly, and therefore, will not react with the tetrazole to incorporate the nucleoside into the ODN. The increase in the amount of the 10-mer SM (mass = 2980) highlights this side reaction. The azide-phosphoramidite can be successfully incorporated into the ODN but the azide can then react with other P(III) species and form the amine product (mass = 3283) which is also seen (NH₂-DNA). Both impurities elute together in UPLC, so it is difficult to determine which is the major impurity.

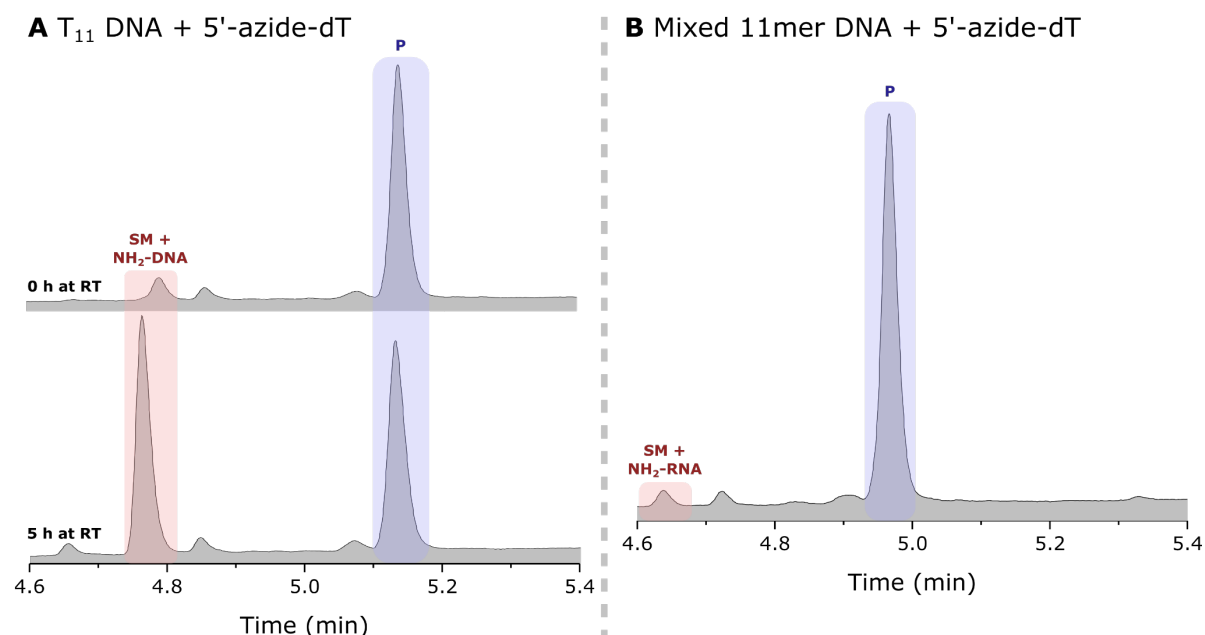


Figure 4.17: UPLC analysis of DNA synthesis using 5'-azide dT phosphoramidite (13). A) Synthesis of 5'-azide T₁₁ DNA (11-mer) using 5'-azide dT phosphoramidite (13) left at RT for 0 h and 5h. Expected masses for A: starting material = 2980, product = 3309. Masses observed for A: 2980.0 (SM), 3309.0 (P), 3283.0 (NH₂-DNA). B) Synthesis of 5'-azide mixed DNA (11-mer) using 5'-azide dT phosphoramidite (13) left at RT for 0 h. Expected masses for B: SM = 2954, P = 3283. Masses observed for B: 2954.0 (SM), 3284.0 (P), 3257.0 (NH₂-DNA). NH₂-DNA = incorporation of 5'-azide dT phosphoramidite followed by reduction of azide to amine.

4.3.3.2 5'-Azido-2'-O-TC-uridine phosphoramidite

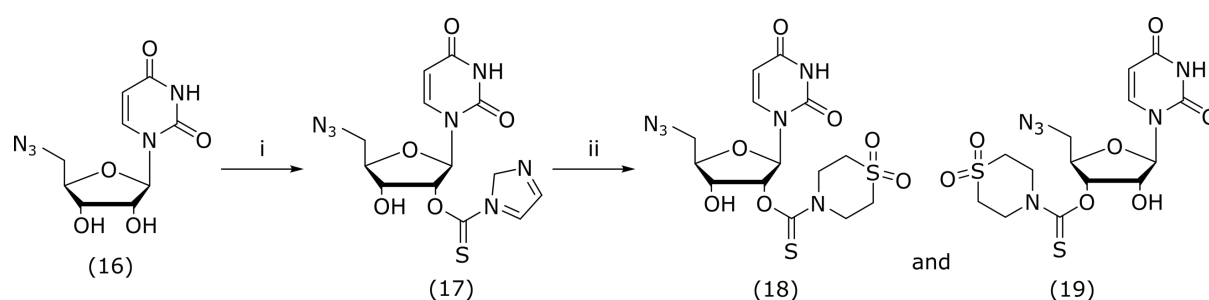
Having successfully incorporated the 5'-azide dT phosphoramidite (13), we focused on the synthesis of a 2'-O-TC U phosphoramidite (20). The synthetic route investigated is shown in Figure 4.18 and Figure 4.19. The 5'-OH is first converted to a 5'-N₃ and the resulting azido compound **16** is then protected with a TC group to give a mixture of the 2' and 3' protected compounds (**18** and **19**, respectively), that we anticipated would be separable by column chromatography. Other synthetic routes such as the use of 3',5'-O-(tetraisopropylidisiloxane-1,3-diyl) protecting group can be used to protect the 3'- and 5'-position of the monomer and to allow for selective protection of the 2'-position when synthesising RNA monomers.³² However, since the

Chemical reaction scheme showing the synthesis of compound 16 from compound 14:

Compound 14 (a ribose derivative with a uracil base and a hydroxymethyl group at the 2' position) reacts with reagent **i** to form compound 15 (where the 2' hydroxyl is replaced by a methanesulfonyl group, MsO).

Compound 15 then reacts with reagent **ii** to form compound 16 (where the MsO group is replaced by an azide group, N₃).

The synthesis started from uridine (**14**) and was carried out according to literature (Figure 4.18).²⁹³ The reaction provided 51% of compound **15** followed by a 63% yield for compound **16**. For the protection of the 5'-azide-uridine (**16**), a thiocarbonyl-diimidazole compound (**17**) was synthesised first without isolation then the group was converted to a thiomorpholine-dioxide as a mixture of 2' and 3' isomers (**18** and **19**) (Figure 4.19).



145

The separation of the two isomers proved more challenging than anticipated. On TLC, the isomer spots ran close together and multiple attempts at column chromatography, using various solvent systems, were needed to separate the isomers. Eventually, the final compound contained a mixture of isomer **18:19** in a ratio of 77:23 and this mixture provided a 41% yield for the reaction.

Whilst a pure sample was not obtained, isomer **18** was carried through to the next step of the synthesis. Since this was only a test reaction, the production of an ODN with a mixture of the 2'-chain extension or 3'-chain extension was not relevant. Phosphitylation was carried similarly to Section 4.3.3.1. However, the reaction time was much longer for compound **20** (5 h) and more equivalents of phosphitylating reagent were needed (3 eq) (Figure 4.20). This is most likely due to the 3'-position of an RNA monomer being more sterically hindered than in DNA and so reacts slower.

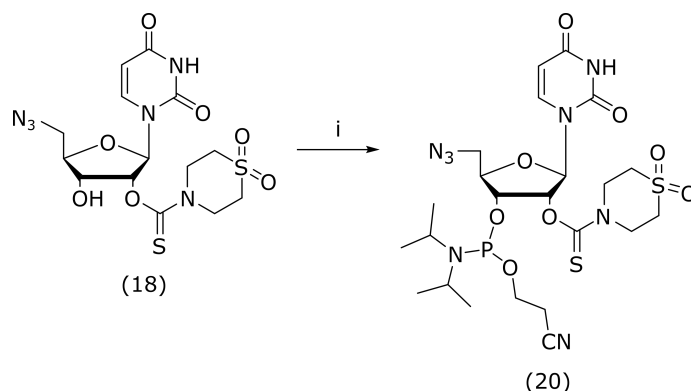


Figure 4.20: Synthesis of 5'-azido-2'-O-TC uridine phosphoramidite, 20. Reaction was carried out for 5 h in anhydrous CHCl_3 . i) Compound **18** was reacted with 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (3 eq) and NMM (5 eq). CHCl_3 was replaced with C_6H_6 for use on the synthesiser and for storage. NMR analysis was not performed as the compound decomposed when concentrated, as determined by TLC. Purity was confirmed by UPLC-MS after successful incorporation into oligonucleotides. See Appendix A4.1.

Once synthesised, compound **20** was used on the synthesiser to produce a 12-mer 2'-O-TC RNA and a 21-mer 2'-O-TC RNA. From Figure 4.21 A and B, overall, the incorporation was not as successful as the DNA monomer with only approx. 20-25% incorporation in the shortest RNA. There is a greater mixture of the side products in the synthesis of RNA making it difficult to determine accurately how much product is produced as the peaks all elute together in the UPLC.

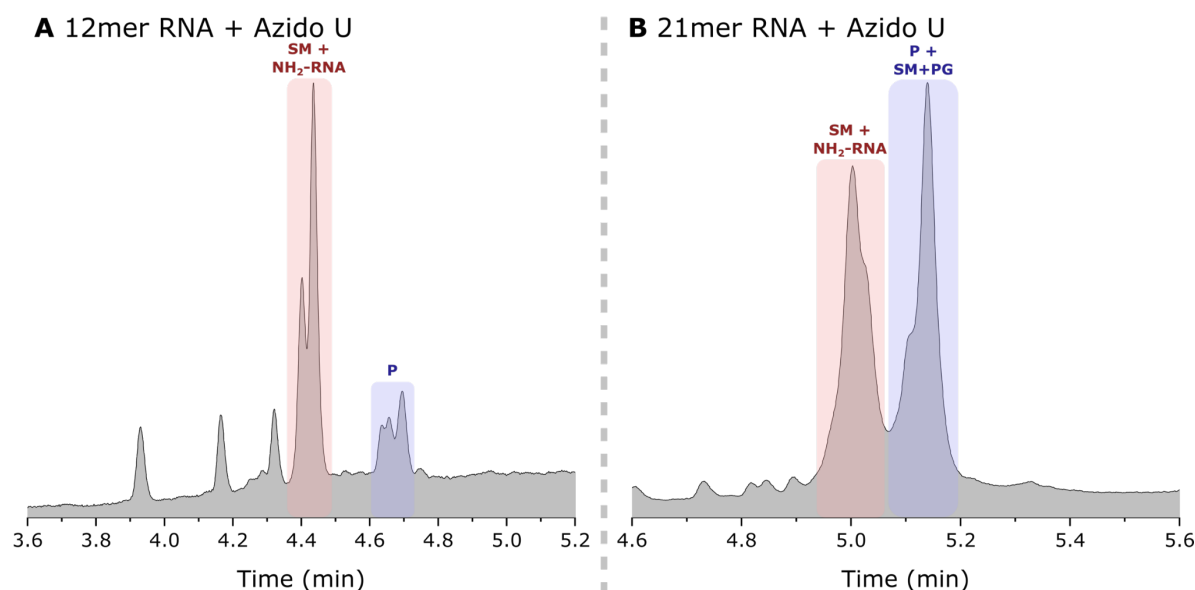


Figure 4.21: UPLC analysis of RNA synthesis using 5'-azide U phosphoramidite (20**).** A) Synthesis of 5'-azide RNA (12-mer) using 5'-azide U phosphoramidite (**20**). Expected masses for A: starting material = 3344, product = 3673. Masses observed for A: 3344.2 (SM), 3673.0 (P), 3648.0 (NH₂-RNA). B) Synthesis of 5'-azide RNA (21-mer) using 5'-azide U phosphoramidite (**20**). Expected masses for B: SM = 6468, P = 6797. Masses observed for B: 6468.1 (SM), 6793.0 (P), 6770.1 (NH₂-RNA), 6576.0 (SM + PG). NH₂-RNA = incorporation of 5'-azide dT phosphoramidite followed by reduction of azide to amine, SM + PG = starting material with unsuccessful deprotection.

The limited success of the azide-RNA monomer **20** when compared to the azide-DNA monomer **13** as judged by the reaction time and purity of the final products was likely due to the lower coupling efficiencies of RNA monomers compared to DNA monomers. Although not completely successful, the production of the 5'-azide RNA could be optimised if more time permitted. The efficiency of the synthesis could

potentially be increased by reducing the reaction time with the use of more phosphitylating reagent. However, a balance between the two would need to be determined because too much reagent can also lead to unwanted side reactions.

4.4 Conclusion and Future Work

The overall aim of this study is to develop an efficient method for the synthesis of an azide moiety in RNA that is compatible with 2'-O-TC RNA chemistry. This would enhance the 'clickable' system used in the 'split and click' technique for the generation of long, modified RNA which could have applications in CRISPR gRNA library generation. Overall, three different approaches were investigated in an attempt to incorporate a 5'-azide into an RNA ODN: the phosphotriester method, the 'reverse phosphitylation' method and the phosphoramidite method. In all cases, incorporation of a 5'-azide DNA monomer was more successful than the corresponding 5'-azide RNA monomer. Many conditions were tested for all three approaches providing some useful information on what is not favourable for these reactions.

For the phosphotriester, conditions were not completely successful even for the standard couplings. For the future of this work, attempts could be made on a small molecular level. Attempts could be made to couple a 5'-azide phosphotriester with a 5'-OH monomer and this could help pinpoint where in the reaction is the difficulty. For the reverse phosphitylation system, the solubility of the monomer was the limiting factor so the use of an alternative tetrazole solution in a more polar solvent could help to produce more effective coupling and may also help to reduce the impurities formed.

The phosphoramidite method provided the most promise for the incorporation of a 5'-azide RNA. The phosphitylation of the 5'-azide RNA monomer requires optimisation by reducing the reaction time and, potentially, with low heat or more phosphitylating reagent. Carrying out a study on some other solvents for the reaction that may be more suitable for RNA monomers would also provide some interesting information.

Another approach that could be investigated is a H-phosphonate approach. As discussed previously, H-phosphonates were not attempted here due the oxidation step required during the coupling. However, determining the minimum concentration required to oxidise the phosphate backbone after coupling of the H-phosphonate without oxidising the 2'-O-TC protecting group to the unwanted carbamate could allow for another approach to incorporate a 5'-azide into RNA.

Conclusion

The aim of the research undertaken here was to synthesis and develop fluorogenic oligonucleotide-based systems to target dsDNA.

In Chapter 2, the aim of the work was to stabilise TFOs in all sequence contexts using an intercalating fluorophore and expand the target range of TFOs. With the incorporation of thiazole orange (TOTFOs), triplex stability is greatly enhanced. A straightforward, post-synthetic labelling method can provide ease of synthesis and very stable triplexes at (and above) neutral pH in all sequence contexts. The stabilising effects are additive and extends the range of target duplexes to contain a base inversion. There is no major disruption to the sequence selectivity of the TOTFOs even with three additions of TO. This stability is corroborated by native gel, and there is an evident relationship between increased stability of the TOTFOs, a large enhancement in fluorescence upon triplex formation and increased binding affinity.

In Chapter 3, the aim of the work was to develop an 'on/off' fluorescence system using modified gRNA that could allow for effective imaging of genomic DNA in live cells. Whilst modifying the gRNA, a range of characteristics were determined. Modifications to the DNA-targeting region of the crRNA can be carried out and still sufficiently maintain the DNA cleavage activity of Cas9 *in vitro* but, more importantly,

the gRNA can be extended and modified through the 5'-end to contain a fluorophore and detachable quencher (QxF-gRNA). This design can successfully bind with Cas9 *in vitro* and in cells and the system can be used to reduce background fluorescence signal compared to gRNA containing no quencher. The 'on/off' system retains strong on-target signals while imaging genomic DNA. Since the system requires all components of the complex (gRNA, Cas9 enzyme and target DNA) before fluorescence enhancement is observed, this can help alleviate misleading results seen in other studies.

In Chapter 4, the aim of this work is to develop a simple and cost-efficient method for azide incorporation in RNA synthesis that is compatible with 2'-O-TC RNA protecting groups. This approach would enhance the current 'split and click' technique for the generation of long modified RNA. Three different approaches were investigated in an attempt to incorporate a 5'-azide: the phosphotriester method, the 'reverse phosphitylation' method and the phosphoramidite method. Incorporation of a 5'-azide DNA monomer was more successful than the corresponding 5'-azide RNA monomer. For the phosphotriester and the 'reverse phosphitylation' method, reactions were not completely successful even for the standard couplings. More extensive studies would be required for these approaches. The phosphoramidite method provided the most promise. The phosphitylation of the 5'-azide RNA monomer requires optimisation but it is evident that there is potential with this

approach. Many conditions were tested for all three approaches providing some useful information on what is not favourable for these reactions.

In summary, the work described here shows modified oligonucleotide-based probes that contain a fluorogenic aspect that have potential in many applications and provides concrete foundations for future work to be carried out.

Chapter 5 Experimental

5.1 Synthesis of phosphoramidites

5.1.1 General

All reagents were purchased from Sigma-Aldrich, Acros Organics, Fluka, Alfa Aesar, Fisher Scientific and were used without purification. The following solvents, CH₂Cl₂, DMF, NEt₃, and DIPEA were of analytical grade and were used without purification. Dry solvents were obtained using the MBraun SPS bench top solvent purification system.

Thin layer Chromatography (TLC) was performed using Merck TLC silica gel 60 F₂₅₄ plates (0.22 mm thickness, aluminium coated). Compounds were visualised using UV irradiation (254 nm and 365 nm) and also stained with the following solutions: p-anisaldehyde (used for alcohols, sugars and DMT-protected compounds) or potassium permanganate (KMnO₄) (used for compounds with an oxidisable group). Column chromatography was carried out under pressure (air/argon) using Merck Geduran® 60Å (40-63 µm) silica-gel.

¹H NMR spectra were measured at 400 MHz on a Bruker DPX400 (AVIIIHD 400) spectrometer. ¹³C NMR spectra were measured at 101 MHz on a Bruker DPX400 spectrometer. ³¹P NMR spectra were recorded on a Bruker AVIIIHD 400 spectrometer at 162 MHz. Chemical shifts are given in ppm and *J* coupling values are quoted in Hz. ¹H and ¹³C NMR spectra were internally referenced to the residual undeuterated

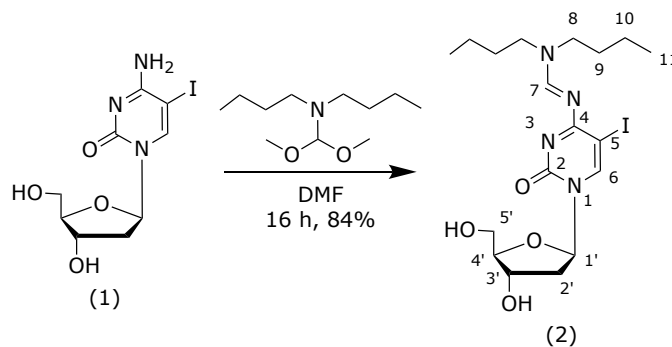
solvent signal.³¹² Assignment of the compounds was aided by COSY (^1H - ^1H) and HSQC (^1H - ^{13}C) experiments. Spectra were reprocessed using Mestre Nova software.

Low-resolution mass spectra (LRMS) were recorded using electrospray ionisation (ESI^+ or ESI^-) on a Waters ZMD quadrupole mass spectrometer in HPLC grade methanol or acetonitrile. High-resolution mass spectra (HRMS) were recorded in HPLC grade methanol/acetonitrile using electrospray ionisation (ESI) on a Bruker APEX III FT-ICR mass spectrometer.

Infrared spectra (IR) were recorded using a Bruker Tensor 27 FT-IR spectrometer. All absorptions are measured in cm^{-1} and are described as broad (br), weak (w), medium (m) or strong (s).

5.1.2 Synthesis of 5-(1-propargylamino)-2'-deoxycytidine-3'-2-cyanoethyl *N,N*-diisopropyl phosphoramidite (O-CEP) (5)

5.1.2.1 2'-Deoxy-N4-[(di-*n*-butylamino)methylidene]-5-iodo-cytidine (2)



2'-Deoxy-N4-[(di-*n*-butylamino)methylidene]-5-iodo-cytidine (**2**) was synthesised following published procedures.³¹³ Compound **1** (0.69 g, 1.96 mmol, 1 eq) was dissolved in anhydrous DMF (7.5 ml). *N,N*-di-*n*-butylformamide dimethyl acetal (0.82 g, 4.04 mmol, 2 eq) was added and the reaction was stirred for 16 h at RT. The solution was evaporated to give an orange oily solid and purified using flash silica gel column chromatography eluting with 10% MeOH in CH₂Cl₂. This gave **2** as an off-white solid (0.806 g, 1.64 mmol, 84%).

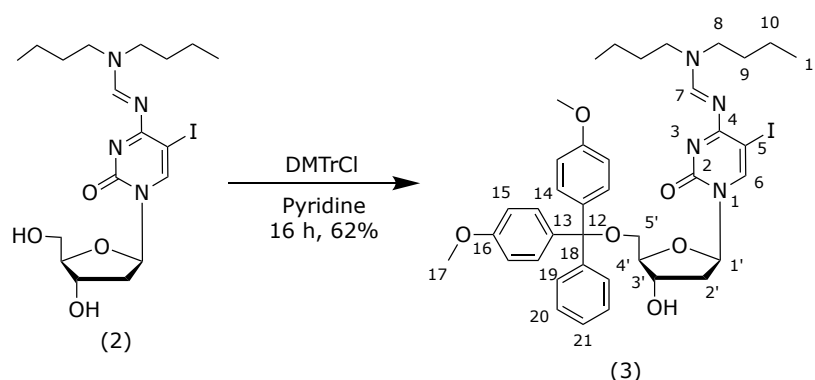
R_f: 0.39 (MeOH:CH₂Cl₂; 10:90)

LRMS (ESI⁺) m/z (%): 493.2 [M+H]⁺ (100), (Calculated mass = 493.36)

¹H NMR (400 MHz, CDCl₃): δ 8.61 (s, 1H, **H**⁷), 8.23 (s, 1H, **H**⁶), 6.02 (t, *J* = 6.0 Hz, 1H, **H**^{1'}), 4.51 (q, *J* = 5.5 Hz, 1H, **H**^{3'}), 4.01 – 3.93 (m, 2H, OH^{3'}, **H**^{4'}), 3.91 – 3.75 (m, 3H, OH^{5'}, **H**^{5'}), 3.53 (t, *J* = 7.4 Hz, 2H, **H**⁸), 3.29 (t, *J* = 7.4 Hz, 4H, **H**⁸), 2.45 – 2.28 (m, 2H, **H**^{2'}), 1.56 (dq, *J* = 7.4 Hz, 4H, **H**⁹), 1.28 (dq, *J* = 7.4 Hz, 4H, **H**¹⁰), 0.88 (dt, *J* = 9.8, 7.4 Hz, 6H, **H**¹¹).

^{13}C NMR (101 MHz, CDCl_3): δ 168.8 (C^4), 158.4 (C^7), 155.8 (C^2), 148.0 (C^6), 88.4 (C^1), 87.5 (C^4), 70.3 (C^3) 69.9 (C^5), 61.7 (C^5), 52.7 (C^8), 46.5 (C^8), 40.8 (C^2), 31.0 (C^9), 29.0 (C^9), 20.3 (C^{10}), 19.8 (C^{10}), 13.9 (C^{11}), 13.7 (C^{11}).

5.1.2.2 2'-Deoxy-5'-O-(4,4'-diethoxytrityl)-N4-[(di-n-butylamino) methylidene]-5-iodo-cytidine (3)



2'-Deoxy-5'-O-(4,4'-diethoxytrityl)-N4-[(di-n-butylamino) methylidene]-5-iodo-cytidine (**3**) was synthesised following published procedures.³¹³ Compound **2** (2.70 g, 5.49 mmol, 1 eq) was dissolved in dry pyridine (30 mL). 4,4'-Dimethoxytrityl chloride (2.23 g, 6.59 mmol, 1.2 eq) was added in small portions over a 10 min period and the reaction was allowed stir for 16 h at RT. Once completed, the reaction was quenched with MeOH (15 mL). The solvent was removed by evaporation and the residue was dissolved in CH₂Cl₂ (80 mL). The reaction mixture was washed with a 10% aqueous NaHCO₃ solution (2 x 50 mL) and sat. aqueous KCl solution (2 x 50 mL). The organic layer was dried over Na₂SO₄ and evaporated to dryness. It was purified using flash silica gel column chromatography using a gradient of 0-7% MeOH in CH₂Cl₂ with 0.1% NEt₃. This gave **3** as a white foam (2.72 g, 3.42 mmol, 62%).

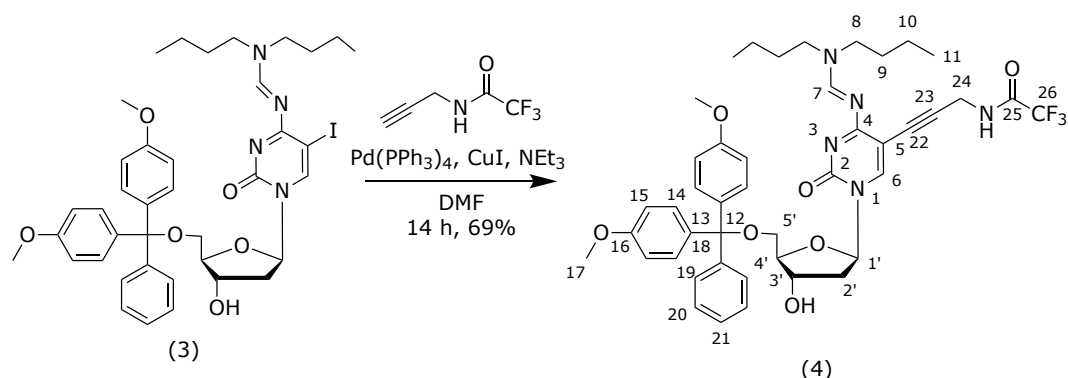
R_f: 0.39 (MeOH:CH₂Cl₂:NEt₃; 7:92:1)

LRMS (ESI⁺) m/z (%): 795.1 [M+H]⁺ (100), (Calculated mass = 795.73)

¹H NMR (400 MHz, CDCl₃): δ 8.68 (s, 1H, **H**⁷), 8.14 (s, 1H, **H**⁶), 7.41 – 7.32 (m, 2H, **H**¹⁹), 7.32 – 7.10 (m, 7H, **H**¹⁴, **H**²⁰ and **H**²¹), 6.83 – 6.74 (m, 4H, **H**¹⁵), 6.25 (dd, *J* = 7.0, 6.1 Hz, 1H, **H**¹), 4.41 – 4.34 (m, 1H, **H**³), 4.03 (q, *J* = 3.7 Hz, 1H, **H**⁴), 3.72 (s, 6H, **H**¹⁷), 3.57 – 3.48 (m, 2H, **H**⁸), 3.38 – 3.17 (m, 4H, **H**^{8, 5}), 2.58 (ddd, *J* = 13.6, 6.1, 3.7 Hz, 1H, **H**²), 2.14 (dt, *J* = 13.6, 6.1 Hz, 1H, **H**²), 1.64 – 1.49 (m, 4H, **H**⁹), 1.36 – 1.17 (m, 4H, **H**¹⁰), 0.93 – 0.83 (m, 6H, **H**¹¹).

¹³C NMR (101 MHz, CDCl₃-*d*): δ 168.8 (**C**⁴), 158.6 (**C**⁷), 158.3 (**C**¹⁶), 155.7 (**C**²), 146.2 (**C**⁶), 144.5 (**C**¹⁸), 135.7 (**C**¹³), 130.1 (**C**²⁰), 130.0 (**C**¹⁹), 128.1 (**C**¹⁴), 126.9 (**C**²¹), 113.4 (**C**¹⁵), 86.8 (**C**¹²), 86.2 (**C**⁴), 86.0 (**C**¹), 72.5 (**C**³), 70.0 (**C**⁵), 63.6 (**C**⁵), 55.3 (**C**¹⁷), 52.6 (**C**⁸), 46.4 (**C**⁸), 42.2 (**C**²), 31.0 (**C**⁹), 29.0 (**C**⁹), 20.2 (**C**¹⁰), 19.8 (**C**¹⁰), 13.9 (**C**¹¹), 13.7 (**C**¹¹).

5.1.2.3 2'-Deoxy-5'-O-(4,4'-diethoxytrityl)-N4-[(di-n-butylamino) methylidene]-5-[N-trifluoroacetylpropargylamine]-cytidine (4)



2'-Deoxy-5'-O-(4,4'-diethoxytrityl)-N4-[(di-n-butylamino) methylidene]-5-[N-trifluoroacetylpropargylamine]-cytidine (**4**) was synthesised following published procedures.³¹³ Compound **3** (2.50 g, 3.15 mmol, 1 eq) was dissolved in anhydrous DMF (36 mL). CuI (0.12 g, 0.63 mmol, 0.2 eq) was added and then NEt₃ (1.2 mL, 15.7 mmol, 5 eq) followed by N-trifluoroacetylpropargylamine (720 μL, 3.5 mmol, 1.1 eq). The reaction was stirred for 10 min in the dark before adding Pd(PPh₃)₄ (0.36 g, 0.32 mmol, 0.1 eq). The yellow reaction mixture was stirred in the dark for 14 h. Once completed, the solvent was evaporated to produce a brown oil which was dissolved in CH₂Cl₂ (100 mL). The reaction mixture was washed with a 5% aqueous EDTA solution (1 x 100 mL), 5% aqueous NaHCO₃ solution (1 x 100 mL) and sat. aqueous KCl solution (1 x 100 mL). The organic layer was dried over Na₂SO₄ and evaporated to dryness. It was purified using flash silica gel column chromatography eluting with 4% MeOH and 0.1% pyridine in CH₂Cl₂. This gave **4** as a yellow foam (1.76 g, 2.15 mmol, 69%).

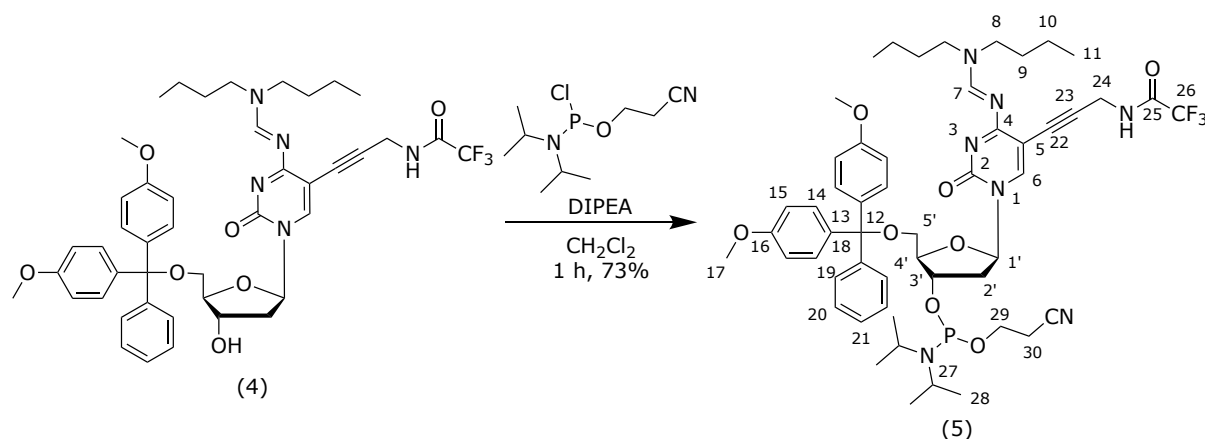
R_f: 0.28 (MeOH:CH₂Cl₂:Py; 4:95:1)

LRMS (ESI⁺) m/z (%): 818.6 [M+H]⁺ (100), (Calculated mass = 818.91)

¹H NMR (400 MHz, CDCl₃): δ 8.75 (s, 1H, **H**⁷), 8.30 (s, 1H, **H**⁶), 7.44 – 7.32 (m, 2H, **H**¹⁹), 7.32 – 7.10 (m, 7H, **H**²⁰, **H**²¹ and **H**¹⁴), 6.82 – 6.73 (m, 4H, **H**¹⁵), 6.29 (t and broad s, *J* = 6.3 Hz, 2H, NH, **H**¹), 4.45 (dt, *J* = 6.3, 3.9 Hz, 1H, **H**³), 4.11 – 4.00 (m, 2H, OH, **H**⁴), 4.00 – 3.81 (m, 2H, **H**²⁴), 3.72 (s, 6H, **H**¹⁷), 3.52 (td, *J* = 6.9, 1.5 Hz, 2H, **H**⁸), 3.36 (dd, *J* = 10.7, 2.8 Hz, 1H, **H**⁵), 3.32 – 3.19 (m, 3H, **H**⁵ and **H**⁸), 2.61 (ddd, *J* = 13.7, 6.3, 3.9 Hz, 1H, **H**²), 2.19 (dt, *J* = 13.7, 6.3 Hz, 1H, **H**²), 1.54 (tt, *J* = 9.0, 6.9 Hz, 4H, **H**⁹), 1.34 – 1.15 (m, 4H, **H**¹⁰), 0.87 (t, *J* = 7.3 Hz, 6H, **H**¹¹).

¹³C NMR (101 MHz, CDCl₃): δ 170.6 (**C**⁴), 158.4 (**C**⁷), 158.3 (**C**¹⁶), 154.7 (**C**²), 145.4 (**C**⁶), 144.7 (**C**¹⁸), 135.9 (**C**²⁵), 135.7 (**C**¹³), 130.1 (**C**²⁰), 130.0 (**C**¹⁹), 128.0 (**C**¹⁴), 126.9 (**C**²¹), 113.3 (**C**¹⁵), 98.3 (**C**⁵), 86.8 (**C**¹²), 86.8 (**C**⁴), 86.3 (**C**¹), 85.2 (**C**²²), 78.8 (**C**²³), 71.8 (**C**³), 63.5 (**C**⁵), 55.3 (**C**¹⁷), 52.4 (**C**⁸), 45.8 (**C**⁸), 42.5 (**C**²), 31.0 (**C**⁹), 30.6 (**C**²⁴), 29.0 (**C**⁹), 20.0 (**C**¹⁰), 19.8 (**C**¹⁰), 13.7 (**C**¹¹), 13.6 (**C**¹¹).

5.1.2.4 5-(1-Propargylamino)-2'-deoxycytidine-3'-O-CEP (5)



5-(1-Propargylamino)-2'-deoxycytidine-3'-O-CEP (**5**) was synthesised following published procedures.³¹³ Compound **4** (923 mg, 1.13 mmol, 1 eq) was dissolved in anhydrous CH_2Cl_2 (10 mL). DIPEA (491 μL , 2.82 mmol, 2.5 eq) was added to the reaction mixture followed by 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (302 μL , 1.35 mmol, 1.2 eq). The reaction was allowed stir under argon at RT for 1 h and monitored by TLC. Once completed, anhydrous CH_2Cl_2 (10 mL) was added and the reaction mixture was washed with degassed sat. KCl (10 mL). The organic layer was dried over Na_2SO_4 and evaporated to dryness under reduced pressure, taking care to keep the product under argon when removing from the büchi. The resulting yellow foam was purified using flash silica gel column chromatography eluting with 90% EtOAc and 0.05% NEt_3 in hexane and under argon. This gave **5** as a pale-yellow foam (837 mg, 0.82 mmol, 73%).

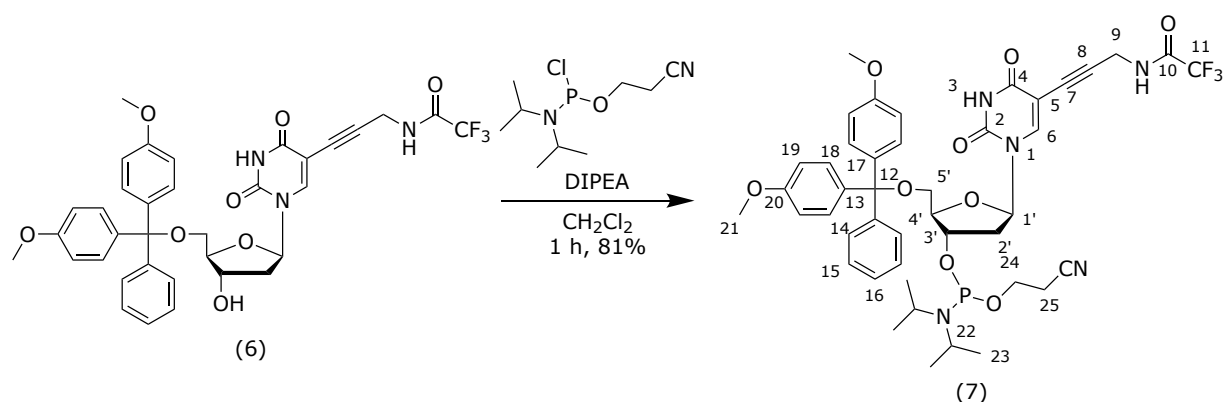
R_f: 0.40, 0.53 (EtOAc:hexane: NEt_3 ; 89:10:1), Product has two isomers.

LRMS (ESI⁺) m/z (%): 1019.1 $[\text{M}+\text{H}]^+$ (100), (Calculated mass = 1019.13)

¹H NMR (400 MHz, CD₂Cl₂): δ 8.68 (s, 1H, **H**⁷), 8.17 (s, 1H, **H**⁶), 7.44 – 7.34 (m, 2H, **H**¹⁹), 7.29 (ddd, $J = 8.5, 5.0, 2.8$ Hz, 4H, **H**²⁰), 7.23 (tdd, $J = 8.5, 5.0, 2.8$ Hz, 2H, **H**¹⁴), 7.20 – 7.10 (m, 1H, **H**²¹), 6.78 (ddd, $J = 8.5, 5.0, 2.8$ Hz, 4H, **H**¹⁵), 6.36 (br s, 1H, NH), 6.15 (dt, $J = 9.0, 6.3$ Hz, 1H, **H**¹), 4.48 (tdd, $J = 10.5, 7.9, 4.1$ Hz, 1H, **H**³), 4.11 (dt, $J = 10.5, 3.4$ Hz, 1H, **H**⁴), 4.05 – 3.85 (m, 2H, **H**²⁴), 3.80 – 3.62 (m, 7H, **H**¹⁷ and **H**²⁷), 3.62 – 3.54 (m, 1H, **H**²⁷), 3.54 – 3.39 (m, 4H, **H**²⁹ and **H**⁸), 3.37 – 3.18 (m, 4H, **H**⁸ and **H**⁵), 2.69 – 2.52 (m, 2H, **H**² and **H**³⁰), 2.36 (t, $J = 6.3$ Hz, 1H, **H**³⁰), 2.18 (ddt, $J = 15.3, 7.9, 6.3$ Hz, 1H, **H**²), 1.60 – 1.49 (m, 4H, **H**⁹), 1.26 (m, 4H, **H**¹⁰), 1.08 (dd, $J = 6.8, 3.3$ Hz, 9H, **H**²⁸), 0.98 (d, $J = 6.8$ Hz, 3H, **H**²⁸), 0.87 (t, $J = 7.3$ Hz, 6H, **H**¹¹).

5.1.3 Synthesis of 5-(1-propargylamino)-2'-deoxyuridine-3'-O-CEP (7)

5.1.3.1 5-(1-Propargylamino)-2'-deoxyuridine-3'-O-CEP (7)



5-(1-Propargylamino)-2'-deoxyuridine-3'-O-CEP (**7**) was synthesised following published procedures.^{313,314} Compound **6**³¹⁵ (1.50 g, 2.20 mmol, 1 eq) was dissolved in anhydrous CH₂Cl₂ (15 mL). DIPEA (958 μL, 5.50 mmol, 2.5 eq) was added to the reaction mixture followed by 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (689 μL, 3.09 mmol, 1.4 eq). The reaction was allowed stir under argon at RT for 1 h and monitored by TLC. Once completed, anhydrous CH₂Cl₂ (10 mL) was added and the reaction mixture was washed with degassed sat. KCl (10 mL). The organic layer was dried over Na₂SO₄ and evaporated to dryness under reduced pressure, taking care to keep the product under argon when removing from the büchi. The resulting foam was purified using flash silica gel column chromatography eluting with 60% EtOAc and 0.05% NEt₃ in hexane and under argon. This gave **7** as a white solid (1.57 g, 1.78 mmol, 81%).

R_f: 0.53, 0.66 (EtOAc:hexane:NEt₃; 59:40:1), Product has two isomers.

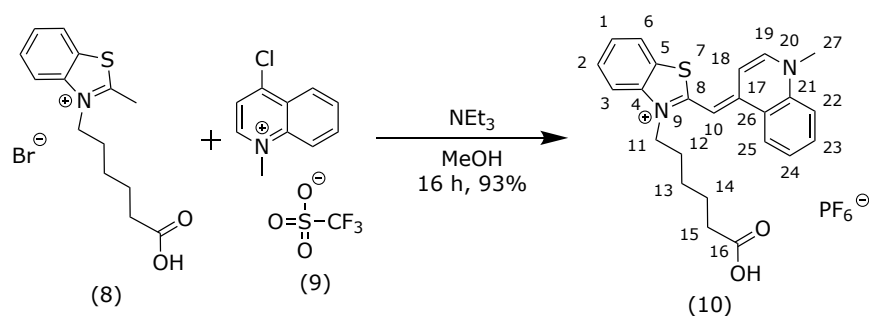
LRMS (ESI⁺) m/z (%): 881.0 [M+H]⁺ (100), (Calculated mass = 880.87)

¹H NMR (400 MHz, CDCl₃): δ 8.23 (d, J = 15.4 Hz, 1H, **H⁶**), 7.37 (dt, J = 8.1, 1.6 Hz, 2H, **H¹⁴**), 7.31 – 7.13 (m, 7H, **H¹⁵**, **H¹⁶** and **H¹⁹**), 6.78 (ddd, J = 9.0, 3.6, 1.6 Hz, 4H, **H¹⁸**), 6.23 (q, J = 6.1 Hz, 1H, **H¹**), 4.59 (dq, J = 9.8, 3.1 Hz, 1H, **H³**), 4.12 (dd, J = 14.4, 3.1 Hz, 1H, **H⁴**), 3.89 – 3.75 (m, 2H, **H²⁴** and **H⁵**), 3.72 – 3.26 (m, 12H, **H²⁴**, **H⁵**, **H⁹**, **H²¹** and **H²²**), 2.63 – 2.46 (m, 2H, **H²** and **H²⁵**), 2.34 (dt, J = 9.8, 6.1 Hz, 2H, **H²** and **H²⁵**), 1.27 – 0.93 (m, 12H, **CH²³**).

³¹P NMR (162 MHz, CDCl₃): δ 149.00, 148.74.

5.1.4 Synthesis of thiazole orange B6 (TO_{B6}) (10)

5.1.4.1 3-(5-Carboxypentyl)-2-((1-methylquinolin-4(1H)ylidene) methyl) benzo-[d]-thiazol-3-ium hexafluorophosphate (TOB6) (10)



TO_{B6} (**10**) was synthesised following published procedures.³¹⁶ Compound **8**³¹⁶ (0.93 g, 2.7 mmol, 1.0 eq) was mixed with compound **9**³¹⁶ (1.05 g, 3.2 mmol, 1.2 eq) in a solution of anhydrous MeOH (34 mL) and anhydrous CH₂Cl₂ (34 mL) under argon. Anhydrous Et₃N (0.94 mL, 6.75 mmol, 2.5 eq) was added and the solution was stirred for 16 h at RT. Once completed, the reaction was precipitated with Et₂O (400 mL), followed by filtration and washed with Et₂O (2 x 20 mL). The solvent was evaporated and the reaction re-dissolved in a mixture of THF (26 mL) and H₂O (24 mL). LiOH (207 mg, 8.4 mmol, 2.5 eq) was added and the solution was stirred for an additional 8 h. The reaction was washed with a 37% aqueous HCl solution (1 x 15 mL), the aqueous layer was extracted with 25% 2-propanol in CH₂Cl₂ (5 x 80 mL). The organic layers were combined and dried over Na₂SO₄, filtered and evaporated to dryness. The crude product was dissolved in a mixture of 2-propanol (25 mL) and CH₂Cl₂ (75 mL) and NaPF₆ (0.8 g, 2.1 eq) was added. The precipitate was filtered, washed with 25% 2-propanol in CH₂Cl₂ (2 x 10 mL) and H₂O (2 x 15 mL) to give the product (**10**) as a red solid (0.59 g, 1.07 mmol, 93%).

R_f: 0.37 (MeOH:CH₂Cl₂; 15:85)

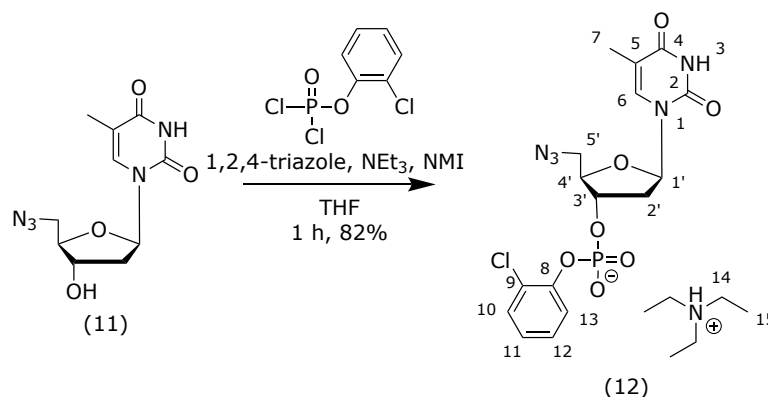
LRMS (ESI⁺) m/z (%): 405.1 [M-PF₆]⁺ (100), (Calculated mass = 405.54)

¹H NMR (400 MHz, d₆-DMSO): δ 8.74 (dd, *J* = 8.7, 1.2 Hz, 1H, **H²²**), 8.62 (d, *J* = 7.2 Hz, 1H, **H¹⁹**), 8.11 – 7.95 (m, 3H, **H²⁴**, **H²⁵** and **H³**), 7.85 – 7.73 (m, 2H, **H²³** and **H⁶**), 7.70 – 7.57 (m, 1H, **H¹**), 7.43 (dd, *J* = 8.7, 7.2, 2H, **H²** and **H¹⁸**), 6.94 (s, 1H, **H¹⁰**), 4.63 (t, *J* = 7.4 Hz, 2H, **H¹¹**), 4.19 (s, 3H, **H²⁷**), 2.21 (t, *J* = 7.2 Hz, 2H, **H¹⁵**), 1.82 (m, 2H, **H¹²**), 1.61 (m, 2H, **H¹⁴**), 1.49 (m, 2H, **H¹³**).

¹³C NMR (101 MHz, d₆-DMSO): δ 174.8 (**C¹⁶**), 159.1 (**C⁸**), 148.6 (**C¹⁷**), 145.0 (**C¹⁹**), 139.9 (**C⁴**), 138.0 (**C²¹**), 133.2 (**C²⁴** or **C²⁵** or **C³**), 128.2 (**C¹**), 127.1 (**C²³**), 125.3 (**C²²**), 124.5 (**C²**), 124.0 (**C²⁶** or **C⁵**), 123.8 (**C²⁶** or **C⁵**), 122.9 (**C²⁴** or **C²⁵** or **C³**), 118.2 (**C²⁴** or **C²⁵** or **C³**), 112.9 (**C⁶**), 107.9 (**C¹⁸**), 87.4 (**C¹⁰**), 45.6 (**C¹¹**), 42.4 (**C²⁷**), 34.6 (**C¹⁵**), 26.7 (**C¹²**), 25.8 (**C¹⁴**), 24.6 (**C¹³**).

5.1.5 Synthesis of 5'-azido-2'-deoxythymidine-3'-phosphotriester (12)

5.1.5.1 5'-Azido-2'-deoxythymidine-3'-O-(2-chlorophenyl)monophosphate (12)



5'-Azido-2'-deoxythymidine-3'-O-(2-chlorophenyl)monophosphate (**12**) was synthesised following published procedures.^{279,317} 1,2,4-Triazole (142 mg, 2.06 mmol, 5.5 eq) and NEt₃ (260 μ L, 1.87 mmol, 5 eq) were dissolved in anhydrous THF (5 mL). 2-Chlorophenyl phosphorodichloridate (154 μ L, 0.94 mmol, 2.5 eq) was added to the reaction mixture and allowed stir for 15 min at RT. Compound **11** (100 mg, 0.37 mmol, 1 eq) and 1-methylimidazole (120 μ L, 1.50 mmol, 4 eq) were dissolved in anhydrous THF (2 mL) added to the reaction mixture. The reaction was stirred under argon at RT for 1 h and monitored by TLC. Once completed, the reaction mixture was quenched with H₂O (80 μ L) followed by NEt₃ (300 μ L). The solvent was evaporated and re-dissolved in CH₂Cl₂ (10 mL) and sat. NaHCO₃ solution (10 mL). The reaction mixture was washed with CH₂Cl₂ (3 x 20 mL), and the combined organic layers were washed with brine solution (1 x 20 mL) and dried over Na₂SO₄ before being evaporated. The compound was purified using flash silica gel column chromatography eluting with a gradient of 10-25% MeOH:CH₂Cl₂ and 0.05% NEt₃. This gave **12** as a white solid (140 mg, 0.30 mmol, 82%).

R_f: 0.3 (MeOH:CH₂Cl₂:NEt₃; 20:79:1)

LRMS (ESI⁺) m/z (%): 458.0 [M+H]⁺ (100), (Calculated mass = 457.76)

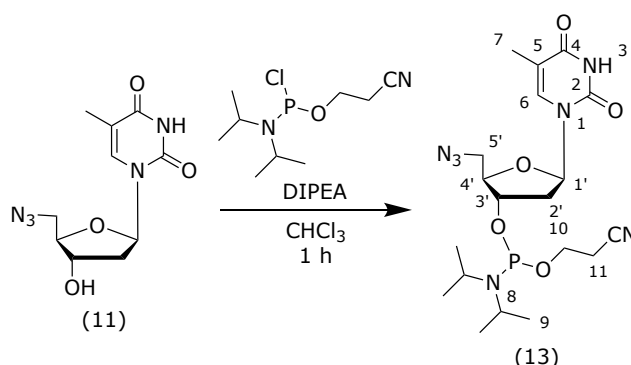
¹H NMR (400 MHz, d₆-DMSO): δ 11.35 (s, 1H, NH), 7.63 – 7.53 (m, 1H, H¹⁰), 7.55 – 7.49 (m, 1H, H⁶), 7.37 (dddd, *J* = 11.7, 7.9, 1.6, 0.9 Hz, 1H, H¹²), 7.29 – 7.15 (m, 1H, H¹¹), 7.06 – 6.93 (m, 1H, H¹³), 6.17 (dd, *J* = 8.1, 6.1 Hz, 1H, H¹), 4.68 (ddt, *J* = 9.4, 6.7, 3.5 Hz, 1H, H³), 4.06 (dt, *J* = 6.7, 3.5 Hz, 1H, H⁴), 3.62 – 3.50 (m, 2H, H⁵), 3.02 (q, *J* = 7.3 Hz, 6H, H¹⁴), 2.38 – 2.19 (m, 2H, H²), 1.79 (d, *J* = 1.2 Hz, 3H, H⁷), 1.16 (t, *J* = 7.3 Hz, 9H, H¹⁵).

¹³C NMR (101 MHz, d₆-DMSO): δ 164.1 (C⁴), 150.9 (C²), 150.1 (C⁸), 136.4 (C⁶), 130.1 (C¹⁰), 130.0 (C¹²), 128.0 (C¹¹), 123.2 (C⁹), 121.6 (C¹³), 110.3 (C⁵), 84.4 (C¹), 84.0 (C⁴), 75.3 (C³), 52.3 (C⁵), 46.1 (C¹⁴), 37.3 (C²), 12.5 (C⁷), 9.3 (C¹⁵).

³¹P NMR (162 MHz, d₆-DMSO): δ -6.9.

5.1.6 Synthesis of 5'-Azido-2'-deoxythymidine-3'-O-CEP (13)

5.1.6.1 5'-Azido-2'-deoxythymidine-3'-O-CEP (13)



Synthesis of 5'-azido 2'-deoxythymidine-3'-O-CEP (**13**) was adapted from previously published procedures.^{310,311,313} Compound **11** (150 mg, 0.56 mmol, 1 eq) was dissolved in anhydrous CHCl_3 (5 mL). DIPEA (489 μL , 2.81 mmol, 5 eq) was added to the reaction mixture followed by 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (250 μL , 1.12 mmol, 2 eq). The reaction was allowed stir under argon at RT for 1 h and monitored by TLC. Once completed, the reaction mixture was washed with degassed sat. KCl (10 mL). The organic layer was dried over Na_2SO_4 and added to a silica column without concentration. The compound was purified by flash silica gel column chromatography eluting with 50% acetone:hexane. The isomeric mixture of products was confirmed by TLC. The solution was concentrated at low temperature and re-dissolved in benzene (10 x 5 mL) taking care not to evaporate the product to dryness. The product (**13**) was stored at $-20\text{ }^\circ\text{C}$ in solution to avoid Staudinger by-product formation.

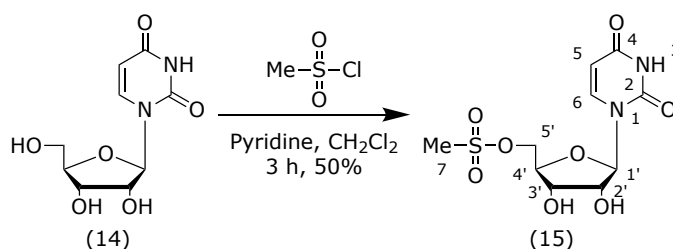
R_f: 0.29 (acetone:hexane; 50:50)

LRMS (ESI⁺) m/z (%): 466.3 $[\text{M}+\text{H}]^+$ (100), (Calculated mass = 466.51)

NMR analysis was not performed as the compound decomposed when concentrated, as determined by TLC. Purity was confirmed by UPLC-MS after successful incorporation into oligonucleotides. See Appendix A4.1.

5.1.7 Synthesis of 2'-O-TC-5'-azido-uridine-3'-O-CEP (20)

5.1.7.1 5'-O-methanesulfonyluridine (15)



Synthesis of 5'-O-methanesulfonyluridine (**15**) was adapted from previously published procedures.³¹⁸ Compound **14** (2.0 g, 8.19 mmol, 1 eq) was co-evaporated with anhydrous pyridine (3 x 5 mL) and dissolved in anhydrous pyridine (50 mL). Methanesulfonyl chloride (668 μ L, 8.60 mmol, 1.05 eq) was dissolved in CH_2Cl_2 (10 mL) and added, dropwise, to the reaction over 1 h. The reaction mixture was allowed stir for an additional 2 h at RT and monitored by TLC. Once completed, the reaction mixture was quenched with MeOH (5 mL) and evaporated. The compound was purified using flash silica gel column chromatography eluting with a gradient of 0–10% MeOH: CH_2Cl_2 . This gave **15** as a white foam (1.32 g, 4.10 mmol, 50%).

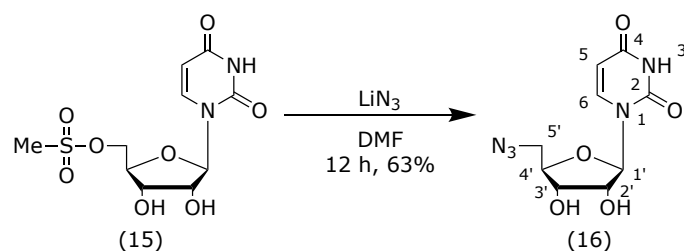
R_f: 0.31 (MeOH: CH_2Cl_2 ; 10:90)

LRMS (ESI⁺) m/z (%): 323.0 $[\text{M}+\text{H}]^+$ (100), (Calculated mass = 323.29)

¹H NMR (400 MHz, *d*₆-DMSO): δ 11.44 – 11.37 (m, 1H, **NH³**), 7.64 (d, *J* = 8.1 Hz, 1H, **H⁶**), 5.79 (d, *J* = 5.4 Hz, 1H, **H^{1'}**), 5.66 (dd, *J* = 8.1, 1.9 Hz, 1H, **H⁵**), 4.47 – 4.32 (m, 2H, **H^{5'}**), 4.11 (qd, *J* = 5.4, 1.1 Hz, 1H, **H^{2'}**), 4.07 – 4.02 (m, 1H, **H^{4'}**), 3.97 (q, *J* = 5.4 Hz, 1H, **H^{3'}**), 3.23 (s, 3H, **H⁷**).

^{13}C NMR (101 MHz, $\text{d}_6\text{-DMSO}$): δ 163.5 (C^4), 151.2 (C^2), 141.3 (C^6), 102.6 (C^5), 89.0 (C^1), 81.5 (C^4), 72.7 (C^2), 70.1 (C^5), 70.0 (C^3), 37.3 (C^7).

5.1.7.2 5'-Azido-uridine (16)



Synthesis of 5'-azido-uridine (**16**) was adapted from previously published procedures.^{317,319} Compound **15** (1.3 g, 4.03 mmol, 1 eq) was dissolved in DMF (15 mL). LiN_3 (20% wt. in H_2O) (5.65 mL, 23.0 mmol, 5.7 eq) was added and the reaction was stirred for 12 h at 60 °C and monitored by TLC. Once completed, the reaction mixture was evaporated and re-dissolved in MeOH (15 mL), then vacuum filtered over Celite to remove excess LiN_3 . The compound was purified using flash silica gel column chromatography eluting with a gradient of 0-10% MeOH: CH_2Cl_2 . This gave **16** (724 mg, 2.69 mmol, 63%) as a white foam.

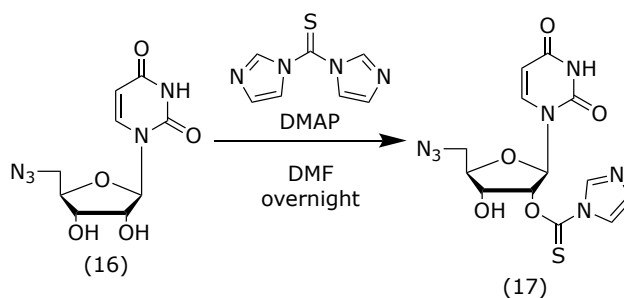
R_f: 0.38 (MeOH: CH_2Cl_2 ; 10:90)

LRMS (ESI⁺) m/z (%): 270.0 $[\text{M}+\text{H}]^+$ (100), (Calculated mass = 270.22)

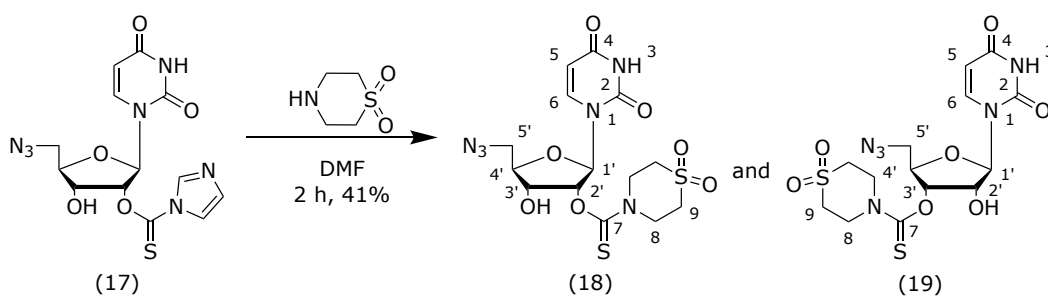
¹H NMR (400 MHz, d₆-DMSO): δ 11.37 (s, 1H, NH^3), 7.69 (d, $J = 8.0$ Hz, 1H, CH^6), 5.77 (d, $J = 5.3$ Hz, 1H, $\text{H}^{1'}$), 5.67 (d, $J = 8.0$ Hz, 1H, H^5), 4.14 (m, 1H, $\text{H}^{2'}$), 3.93 (t, $J = 3.8$ Hz, 1H, H^4), 3.63 – 3.57 (m, 2H, H^5), 3.51 (m, 1H, $\text{H}^{3'}$).

¹³C NMR (101 MHz, d₆-DMSO): δ 163.5 (C^4), 151.2 (C^2), 141.6 (C^6), 102.6 (C^5), 89.2 ($\text{C}^{1'}$), 82.6 (C^4), 72.8 ($\text{C}^{2'}$), 70.8 ($\text{C}^{3'}$), 52.1 (C^5).

IR: 3370.5 (m, O-H), 2105.0 (s, N_3), 1659.2 (s, C=O) cm^{-1} .

5.1.7.3 5'-Azido-2'-O-(1,1-dioxo-1 λ^6 -thiomorpholine-4-carbothioate)-uridine (18)

Compound **16** (470 mg, 1.75 mmol, 1 eq) was dissolved in DMF (15 mL). 1,1'-Thiocarbonyldiimidazole (342 mg, 1.92 mmol, 1.1 eq) was added along with 4-dimethylaminopyridine (DMAP) (21 mg, 0.175 mmol, 0.1 eq). The reaction was stirred overnight at RT and monitored by TLC to produce **17**.



Once completed, thiomorpholine 1,1-dioxide (283 mg, 2.10 mmol, 1.2 eq) was added to the crude reaction mixture (**17**) and left to stir for 2 h. The reaction was analysed by TLC and the solvent evaporated. The oil was co-evaporated with toluene (3 x 5 mL) to remove excess DMF. The compound was purified using flash silica gel column chromatography eluting with a gradient of 0-4% 2-propanol:CH₂Cl₂. The mixture now contained a mixture of 2'-O-TC (**18**) and 3'-O-TC (**19**) isomers. Two more columns eluting with a gradient of 50-80% hexane:EtOAc were carried out to produce an enriched mixture of isomer **18** (327 mg, 0.73 mmol, 41%, 77:23 of **18:19**).

R_f 17: 0.62 (2-propanol:CH₂Cl₂; 5:95), **R_f 18:** 0.48 (2-propanol:CH₂Cl₂; 5:95), **R_f 19:** 0.29 (2-propanol:CH₂Cl₂; 5:95).

LRMS (ESI⁺) m/z (%): 448.0 [M+H]⁺ (100), (Calculated mass = 447.85)

HRMS (ESI⁺, MeOH) m/z: 469.0570 [M+Na]⁺, (Calculated mass: 469.0576 for C₁₄H₁₈N₆O₇NaS₂)

IR: 2104.5 (m, N₃), 1683.8 (s, C=O), 1380.6 (w, S=O), 1120.1 (s, C-O) cm⁻¹.

¹H NMR Isomer 18 (400 MHz, MeOD): δ 7.60 (d, *J* = 8.1 Hz, 1H, H⁶), 5.94 (d, *J* = 5.3 Hz, 1H, H^{1'}), 5.80 – 5.75 (m, 1H, H^{2'}), 5.67 (d, *J* = 8.1 Hz, 1H, H⁵), 4.43 (t, *J* = 6.1, 4.9 Hz, 1H, H^{3'}), 4.36 – 4.25 (m, 2H, H⁸ or H⁹), 4.12 – 4.06 (m, 2H, H⁸ or H⁹), 4.04 (app. td, *J* = 4.9, 3.3 Hz, 1H, H^{4'}), 3.69 – 3.59 (m, 2H, H^{5'}), 3.22 – 3.05 (m, 4H, H⁸ or H⁹).

¹³C NMR Isomer 18 (101 MHz, MeOD): δ 187.3 (C⁷), 164.6 (C⁴), 150.7 (C²), 141.5 (C⁶), 102.1 (C⁵), 88.4 (C^{1'}), 83.4 (C^{4'}), 81.0 (C^{2'}), 69.5 (C^{3'}), 51.7 (C^{5'}), 51.0 (C⁸ or C⁹), 50.5 (C⁸ or C⁹), 48.4 (C⁸ or C⁹), 43.6 (C⁸ or C⁹).

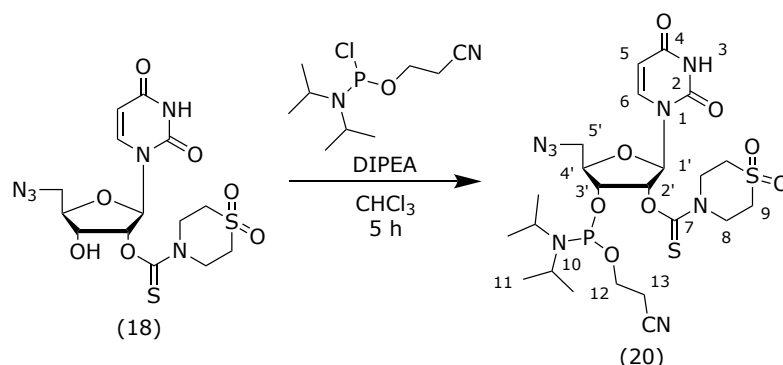
¹H NMR Isomer 19 (400 MHz, MeOD): δ 7.63 (d, *J* = 8.1 Hz, 1H, H⁶), 5.83 (d, *J* = 6.3 Hz, 1H, H^{1'}), 5.69 (d, *J* = 8.1 Hz, 1H, H⁵), 5.64 (app. dd, *J* = 6.3, 3.7 Hz, 1H, H^{3'}), 4.49 (t, *J* = 6.3 Hz, 1H, H^{2'}), 4.28 (app. dt, *J* = 4.4, 3.7 Hz, 3H, H^{4'}, H⁸ or H⁹), 4.11 – 4.02 (m, 2H, H⁸ or H⁹), 3.75 – 3.59 (m, 2H, H^{5'}), 3.21 – 3.08 (m, 4H, H⁸ or H⁹).

¹³C NMR Isomer 19 (101 MHz, MeOD): δ 187.5 (C⁷), 164.4 (C⁴), 140.9 (C⁶), 102.2 (C⁵), 89.4 (C^{1'}), 80.7 (C^{4'}), 79.1 (C^{3'}), 71.7 (C^{2'}), 51.8 (C^{5'}), 50.5 (C⁸ or C⁹), 50.4 (C⁸ or C⁹), 49.3 (C⁸ or C⁹), 43.6 (C⁸ or C⁹).

Spectra available in Appendix A4.2.

5.1.7.4 5'-Azido-2'-O-(1,1-dioxo-1λ⁶-thiomorpholine-4-carbothioate)-uridine-3'-

O-CEP (20)



Compound **18** (100 mg, 0.22 mmol, 1 eq, 77:33 of isomers **18:19** as determined by NMR) was dissolved in anhydrous CHCl_3 (6 mL). NMM (123 μL , 1.12 mmol, 5 eq) was added to the reaction mixture followed by 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (100 μL , 0.45 mmol, 2 eq). The reaction was allowed stir under argon at RT for 5 h and monitored by TLC. Once complete, the reaction mixture was washed with degassed sat. KCl (10 mL). The organic layer was dried over Na_2SO_4 and added to a silica column without concentration. The compound was purified by flash silica gel column chromatography eluting with 50% acetone:hexane. The isomeric mixture of product **20** was confirmed by TLC. The solution was concentrated at low temperature and re-dissolved in benzene (10 x 5mL) taking care not to evaporate the product to dryness.

R_f: 0.58, 0.64 (acetone:hexane; 50:50)

LRMS (ESI⁺) m/z (%): 648.0 $[\text{M}+\text{H}]^+$ (100), (Calculated mass = 648.7)

NMR analysis was not performed as the compound decomposed when concentrated, as determined by TLC. Purity was confirmed by UPLC-MS after successful incorporation into oligonucleotides. See Appendix A4.1.

5.2 Oligonucleotide Synthesis and Purification

5.2.1 General

Solid supports, standard DNA/RNA phosphoramidite and all other reagents used in the synthesis were purchased from Link Technologies, Glen Research and Applied Biosystems Ltd. Modified phosphoramidites such as 5-Methyl deoxycytidine phosphoramidite (^{Me}C), 5-(1-propynyl)-deoxycytidine phosphoramidite (pC), amino C6 deoxythymidine phosphoramidite (AC6), 2'-O-Me A/G/C/U phosphoramidites, 5'-BHQ-2 phosphoramidite, BHQ-2-deoxythymidine phosphoramidite, 5'-amino-modifier phosphoramidite, 5'-alkyne-modifier phosphoramidite, 5'-biotynylated TEG phosphoramidite and HEG phosphoramidite were purchased from Glen Research, Link Technologies and Berry and Associates. NAP Sephadex gel-filtration columns were purchased from GE Healthcare and used according to the manufacturer's instructions.

5.2.2 Oligonucleotide synthesis

All oligonucleotides (ODNs) were synthesised using an Applied Biosystems 394 automated DNA/RNA synthesiser using the standard 1.0 µmol phosphoramidite cycle of detritylation, coupling, capping and oxidation. Stepwise coupling efficiencies and overall yields were monitored using the automated trityl cation conductivity monitoring facility and for all ODNs were >98.0%. Standard dA-bz, dG-ib, dC-bz and dT monomers were coupled for 35 s for DNA synthesis, A-tac, G-tac, C-tac and U monomers were coupled for 6 min for 2'-O-TBDMS RNA synthesis and A-bz, G-ib, C-

bz and U monomers were coupled for 3 mins for 2'-O-TC RNA synthesis. Non-standard monomers were coupled for 7 min.

5.2.3 Selective β -cyanoethyl removal

ODNs that contained a primary amine were treated with 20% diethylamine solution in MeCN (20 min at RT) to avoid unwanted reactions with between cyanoethyl and the amine during the ammonia cleavage.

5.2.4 Deprotection and purification

DNA ODNs were deprotected and cleaved from the solid support using concentrated aqueous NH_3 solution (1 h, RT) followed by heating in sealed tubes (5 h at 55 °C). ODNs that contained pC phosphoramidite monomers were deprotected for 12 h at 55 °C.

RNA ODNs protected with 2'-O-TBDMS were cleaved from the resin using a concentrated aqueous NH_3 :EtOH solution (3:1) by heating in sealed tubes (2.5 h at 55 °C). NH_3 was removed *in vacuo* and the solution was lyophilised before being re-dissolved in a solution of DMSO (300 μL) and triethylamine trihydrofluoride (300 μL) and heated (2.5 h at 65 °C). The solution was allowed to cool to room temperature before butanol (3 mL) and 3M sodium acetate (50 μL , pH = 5.2) was added. The solution was cooled (30 min at -80 °C) and then pelleted by centrifugation (30 min at 4 °C, 12,000 $\times g$). The supernatant was removed, and the pellet washed with 70%

ethanol twice. The pellet was lyophilised and re-dissolved in nuclease-free water before purification.

RNA ODNs protected with 2'-O-TC were deprotected and cleaved from the resin using a solution of anhydrous ethylenediamine:toluene (1:1) (6 h at RT). The resin was carefully washed with MeCN before the cleaved ODN was eluted using 1 mL of nuclease-free water, ready for purification.

ODNs were purified using reverse-phase HPLC on a Gilson system using a Luna® 10 µm C8 100 Å 250 x 10 mm column. For unmodified ODNs, the purification gradient used was 10% - 40% buffer B in buffer A over 20 min, flow rate of 4 mL/min (buffer A: 0.1 M triethylammonium bicarbonate (TEAB); buffer B: 0.1 M TEAB with 50% acetonitrile). ODNs with multiple modifications and those that were labelled with fluorescent dyes were purified with a gradient of 20 - 65% of buffer B in buffer A over 30 mins or with hexylammonium acetate (HAA) buffer with a gradient of 60% - 80% buffer B in buffer A over 30 min, flow rate of 4 mL/min (buffer A: 0.1 M HAA; buffer B: 0.1 M HAA with 50% acetonitrile). Fractions were desalted using two NAP-10 gel filtration columns according to the manufacturer's instructions.

For RNA ODNs, the column was heated to 40 °C to disrupt any secondary structures. Elution of ODNs was monitored by ultraviolet absorption at a suitable wavelength in the range 260 – 300 nm. Negative ion electrospray mass spectra were recorded using

UPLC-MS Xevo G2-QTOF (Waters) mass spectrometer with an ACQUITY UPLC system and an Oligonucleotide BEH C18 reversed-phase column (130Å, 1.7 µm, 2.1 mm x 50 mm). The gradient used was 0 – 70% buffer B in buffer A over 7.5 min, flow rate of 0.2 mL/min (buffer A: 8.6 mM NEt₃, 200 mM HFIP, 5 % methanol; buffer B: 20% buffer A in methanol). Raw data was processed using deconvolution software Mass Lynx v4.1 and is given in Appendix A4.1.

5.3 Oligonucleotide Labelling

5.3.1 Labelling oligonucleotides with NHS Ester

Freeze-dried ODN (100 nmol, 1 eq) was dissolved in 0.5 M NaHCO₃ labelling buffer (30 µL, pH 8.6) and an NHS active ester of the dye (2 µmol, 20 eq) was dissolved in DMSO (20 µL) and both solutions added together. The final solution was shaken for 4 h at 40 °C. The solution was desalted to remove excess dye using NAP-10 gel filtration columns and purified according to Experimental 5.2.4.

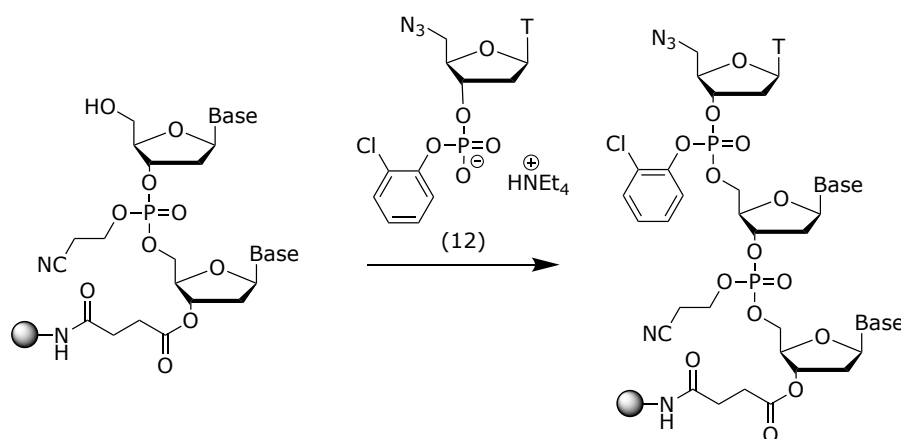
5.3.2 Labelling oligonucleotides with click chemistry (CuAAC)

A solution of click catalyst was prepared from tris(3-hydroxypropyltriazolylmethyl) amine ligand (THPTA, 10.5 µmol, 70 eq in H₂O) and CuSO₄·5H₂O (1.5 µmol, 10 eq in H₂O). The solution was mixed with an azide modified dye (0.5 µmol, 3.3 eq) in 100 µL of DMSO. The mixture was added to the ODN (150 nmol, 1 eq) dissolved in 10 µL of H₂O and sodium ascorbate (15 µmol, 100 eq in H₂O) was added last to begin the reaction. The final solution was shaken for 3 h at 30 °C (850 rpm). The solution was

desalted to remove excess dye using a NAP-10 followed by a NAP-25 gel filtration column and purified according to Experimental 5.2.4.

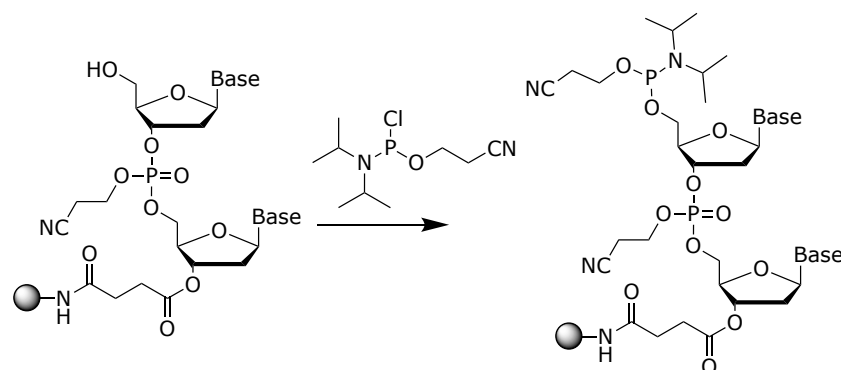
5.4 On-Resin Coupling of Monomers to Oligonucleotides

5.4.1 Azide-phosphotriester coupling

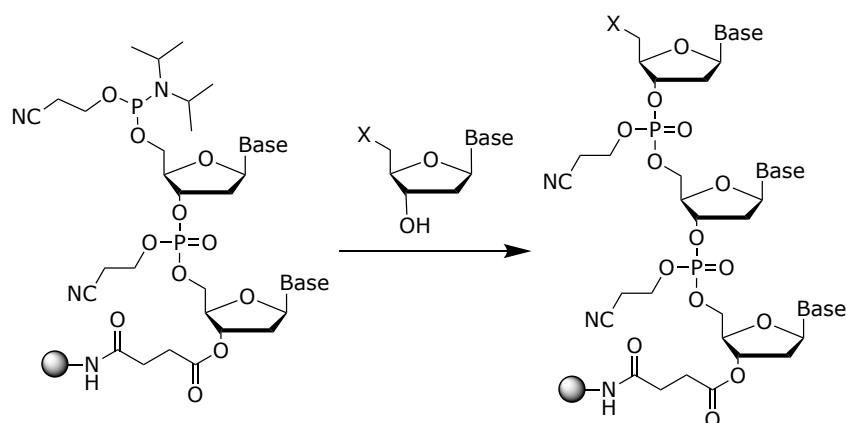


Compound **12** was co-evaporated in anhydrous pyridine (3 x 5 mL) and dried using high vacuum. 1-(2-Mesitylenesulfonyl)-3-nitro-1*H*-1,2,4-triazole (MSNT, 104 mg, 350 μ mol, 5 eq) was added to **12** (40 mg, 70 μ mol, 1 eq) in anhydrous pyridine (500 μ L) to mix for 15 min. N-methyl-imidazole (49 μ L, 588 μ mol, 8.4 eq) was added and the solution transferred to a column containing ODN on solid support (CPG resin). The resin and solution were mixed and allowed to react (overnight, RT). The reaction mix was removed from the column and the resin was washed with anhydrous pyridine (3 x 1 mL) and anhydrous MeCN (3 x 1 mL). For analysis, the phosphate protecting group was not removed but could be done so using published protocols.²⁹³

5.4.2 Reverse phosphitylation and coupling



ODNs on solid support (CPG resin) were dried in a desiccator overnight along with the nucleoside and N-methylmorpholine (NMM) and the solvent was dried over molecular sieves and degassed before use. Conditions used were varied according to Section 4.3.2. MeCN (200 μ L) was added to the resin followed by NMM (55 μ L, 2.5 M) and the solution mixed. Phosphitylating reagent (45 μ L, 1 M) was added to the solution and allowed to react under argon for 45 min – overnight. The reaction solution was removed, and the resin washed with DMF (1 x 1 mL) and MeCN (3 x 1 mL).



5'-DMT-dT or DMT-dC(Bz) (0.5 M) were dissolved in 0.5 M tetrazole in MeCN (Thermo Fisher Scientific) and compound **11** was dissolved in a mix of 0.5 M tetrazole in MeCN and DMF and reacted overnight under argon. The resin was washed with

MeCN (3 x 1 mL) and reacted with a solution of 0.02 M I₂ in H₂O:Pyridine:THF (0.6 in 12:22:66) (Thermo Fisher Scientific) for 10 min. After oxidation, the resin was washed with MeCN (3 x 1 mL) and for nucleosides containing 5'-DMT protection, the resin was treated with trichloroacetic acid (3% TCA solution) for acid removal of the protecting group before being fully deprotected in aqueous NH₃. Effective coupling was confirmed using UPLC-MS (Appendix A4.1).

5.5 Biophysical Studies

5.5.1 Thermal melting analysis

Thermal denaturation studies were carried out on a Cary 4000 UV-Visible Spectrophotometer from Varian in Hellma® SUPRASIL synthetic quartz cuvettes of 10 mm path-length and 1 mL sample volume. The measurements were monitored at 260 nm using Cary WinUV thermal application software. For triplex studies, the TFO and duplex were combined in a concentration/ratio of 2.5:1 μ M and dissolved in the appropriate filtered buffer. The first ramp equilibrated the samples by initial denaturation by heating to 85 °C at 10 °C/min. The samples were held at this temperature for 2 min before annealing was carried out by cooling to 20 °C at a rate of 0.5 °C/min and holding for 20 min. Heating was at a rate of 0.5 °C/min and samples were held at 85 °C for 2 min. A total of six ramps are carried out but the first (fast) ramp was excluded to ensure uniform results. No evidence of hysteresis (greater than ± 1 °C) was observed. Hysteresis can occur in triplex melting analysis as the rate of annealing is generally slow. An average T_m was determined from the first derivatives

of the melting curves. Final T_m values are an average of two individual measurements in which each is an average of 5 ramps and produced errors within ± 0.6 °C (calculated using standard deviation).

5.5.2 Fluorescence melting analysis using a CFX96™ Real Time PCR Detection System

Fluorescence melting experiments were conducted on a CFX96™ Real-Time PCR Detection System (Bio-Rad Laboratories Inc.) using Bio-Rad CFX Manager 3.0 software. For triplex studies, samples of 0.25 μ M of the duplex in 20 μ L filtered buffer were used with TFO concentration of 0.625 μ M to provide a ratio of [2.5:1]. The samples were pre-incubated at 25 °C for 1 min. They were denatured by heating to 95 °C at a rate of 0.1 °C/s then held at 95 °C for 5 min before cooling to 25 °C at the same rate. They were held at 25 °C for 5 min then denatured by heating to 95 °C at a rate of 0.6 °C/min before holding at 95 °C for 5 min. The samples were then cooled to 25 °C at the same rate. The fluorescence melting curve data was converted to the first derivatives to give the melting temperatures ($-d(F)/dT$ where F is the fluorescence and T is temperature in °C). Step 3 (slow melting steps) and step 4 (slow annealing steps) were used for data analysis. The excitation and emission channel used on the CFX96™ Real-Time PCR Detection System for fluorophores are below. All melting experiments were carried out in duplicates and the average values used.

Fluorophore	λ_{ex} (nm)	λ_{em} (nm)	Dye Chanel	λ_{ex} (nm)	λ_{em} (nm)
TO	480	530	TET [™]	521	536
Cy3	550	570	VIC	538	554
TAMRA	557	583	ROX	575	602

5.5.3 Fluorescence steady state analysis

5.5.3.1 Fluorescence steady state analysis using CFX96[™]

For CRISPR-Cas9 studies, steady state fluorescence studies were carried out also using a CFX96[™] Real-Time PCR Detection System and Bio-Rad CFX Manager 3.0 software. Samples contained 0.5 μM of construct (gRNA and tracrRNA), target DNA and enzyme in 20 μL of Cas9 buffer (New England BioLab) were used unless otherwise stated. For RNase H fluorescence studies, samples contained 0.25 μM of construct (gRNA and tracrRNA, 1 eq), 1 eq of target DNA, 5 eq of Cas9 enzyme and 5 Units (1 μL) of RNase H (5000 U/mL stock) enzyme in 20 μL of RNase H buffer (New England BioLab) unless otherwise stated. Cas9 enzyme was added 15 min prior to RNase H to allow time for the complex to form. These studies measured fluorescence intensities in the dye channels indicated in Experimental 5.5.2 at a set temperature of 37 °C for 60 min using 60 cycles, unless otherwise stated. An average of all cycles was taken, and background fluorescence of the buffer was subtracted to give a final fluorescence intensity in arbitrary units.

5.5.3.2 Fluorescence scan analysis

For triplex analysis, fluorescence studies were performed on a Perkin Elmer LS50B Luminescence Spectrometer fitted with Perkin Elmer PTP-1 Peltier temperature controller. *FLWinlabTempScan* software was used with 400 nm/s scan speed. For ODNs containing TO, the emission was recorded from 510 nm to 700 nm, excitation wavelength = 484 nm, gain = high (900V), excitation slit width = 7 nm and emission slit width = 7 nm. Final concentrations of the samples were 1.8 μM of the duplex in 250 μL buffer and 1 μM of TFO. Duplex samples were pre-annealed at 95 $^{\circ}\text{C}$ for 5 min before being allowed to cool to RT overnight. On addition of the duplex to the TFO, samples were left for 400 s to allow complete formation of the triplex (Surface Plasmon Resonance analysis provided the formation time of the triplex). Spectra were recorded for the single-stranded TFO and then titrated with 1.8 eq of the target duplex and recorded again at 20 $^{\circ}\text{C}$. Fluorescence intensity ratios were calculated by comparison of the fluorescence intensity at the maximum emission wavelength of the formed triplex ($\lambda_{\text{em}}^{\text{T}}$). Final fluorescence values were calculated from an average of two individual measurements in which both the ssTFO and formed triplex were measured and produced errors within $\pm 5.4\%$.

5.5.4 Quantum yield measurements using integrated sphere

The absolute fluorescence quantum yields were measured using a SC-30 integrating sphere module (Edinburgh Instruments) and the re-absorption effect was corrected when possible. For all TFOs and formed triplex complexes, the excitation wavelength

was 488 nm, the wavelength step size 0.1 nm and integration dwell time 0.2 s. All fluorescence samples were prepared with optical densities under 0.1 under ambient conditions (1 μ M of TFO in 3000 μ L of buffer). Sample spectra were recorded for the single-stranded TFO and then titrated with 1.8 eq of the target duplex, allowed anneal for 400 s and recorded again at 20 °C. Three scans were repeated for both the TFO and triplex sample solutions as well as the buffer. The scattering region between 483 nm and 593 nm, and emission region between 500 and 700 nm were chosen for the calculation of the observed quantum yields.

5.5.5 Surface Plasmon Resonance (SPR)

SPR measurements were carried out on a BIAcore X100 instrument (GE Healthcare). Biosensor chips, pre-coated with streptavidin (GE Healthcare), were used throughout the experiment and the running buffer was 10 mM MOPS, 10 mM MgCl_2 (pH 5.8). The sensor chip surface was pre-treated prior to immobilisation with 3 serial injections of 1 M NaCl in 50 mM NaOH followed by a prime with running buffer at 25 °C. The 3'-biotinylated py strand of the duplex was passed over the surface at 10 μ L/min until the target immobilisation response (Imm RU) was achieved, corresponding to 700 response units (RU). A solution of the pu strand of the duplex (10 μ M for 240 s contact time) was passed over the chip surface at 5 μ L/min to form a duplex on the surface and allowing a 360 s stabilisation period. Solutions containing 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 2.5, 5, 10 or 20 μ M of the TFO (for 420 s contact time) were then injected over the immobilized duplex at a flow rate of 5 μ L/min to generate the triplex. 1800 s were

allowed for dissociation of the triplex. The surface was regenerated using NaOH at a flow rate of 10 $\mu\text{L}/\text{min}$ (10 mM for 240 s contact time) and re-equilibrated with running buffer (360 s) before the next cycle was begun. Equilibrium constants (K_a) were determined for the triplexes using the BIAcore X100 Evaluation Software 2.0.1 Plus Package. The model used assumes 1:1 binding and that the kinetics are first-order, and not limited by mass transport effects.⁵⁸ Accurate k_{diss} could not be directly calculated from the disassociation curves as disassociation of the triplex was very slow so, they were calculated using the association data according to literature.¹³⁰

5.5.6 Polyacrylamide gel electrophoresis (PAGE)

Native reagents	8% (10 mL)	15% (70 mL)
40% acrylamide:bis-acrylamide (29:1)	2 mL	26.25 mL
10x Tris-borate-EDTA (TBE)	0.5 mL	-
10x MOPS-acetate	-	7 mL
Milli-Q water	7.45 mL	36.15 mL
10% ammonium persulfate (APS)	80 μL	560 μL
Tetramethylethylenediamine (TEMED)	8 μL	56 μL

Denaturing reagents	10% (60 mL)	20% (60 mL)
40% acrylamide:bis-acrylamide (29:1)	15 mL	30 mL
10x TBE	6 mL	6 mL
Milli-Q water	14 mL	5 mL
7M Urea	25.2 g	25.2 g
10% APS	560 μL	560 μL
TEMED	56 μL	56 μL

15% Native PAGE and 10-20% denaturing PAGE (W x D x H = 180 mm x 1.5 mm x 240 mm) were run in 1x buffer and 8% native PAGE (W x D x H = 80 mm x 1 mm x 75 mm) was run in 0.5x buffer and all gels were visualised using Syngene G:Box imager. Unlabelled samples were stained using 1x SYBR Gold.

5.5.6.1 Native PAGE

Triplex formation was analysed by 15% native PAGE. Gels were run according to the recipe in the table in Experimental 5.5.6 at RT at 125 V for 18 h using 1x MOPS-acetate buffer (40 mM MOPS buffer containing 10 mM NaOAc and 10 mM MgCl₂, pH 7). Samples were heated (90 °C for 5 min) and slowly cooled to RT overnight before loading. The concentration of each duplex strands was 1 µM in 25 µL (25 pmol) with varying ratios of TFO strand. Samples were loaded in 5% glycerol and the same MOPS running buffer.

5.5.6.2 Electrophoretic Mobility Shift Assay (EMSA) native PAGE

Cas9 binding was analysed using 8% electrophoretic mobility shift assay (EMSA) native PAGE. Gels were run according to the recipe in the table in Experimental 5.5.6 at 4 °C at 200 V for 1 h using 0.5x TBE. Samples contained 1 µM in 10 µL (10 pmol) of gRNA/tracrRNA, 1x Cas9 buffer and 1 µL BSA (10 mg/mL stock) with increasing eq of Cas9 (1 eq = 1 µM, 2 eq and 4 eq) for lanes that did not contain a competitor and 1 µM of Cas9 when a competitor was added. The competitor used was IVT sgRNA in increasing eq (1 eq, 10 eq and 25 eq). Sample were incubated (37 °C for 15 min) before

being loaded onto the gel with 5% glycerol and in 0.5x TBE running buffer with 5 mM MgCl_2 .

5.5.6.3 Denaturing PAGE

Cleavage efficiencies were analysed by constant watts denaturing PAGE. 10-20% gels were run according to the recipe in the table in Experimental 5.5.6 at RT at 20 W for 2-3 h using 1x TBE. Samples contained 0.5 μM in 20 μL (10 pmol) of dsDNA target, gRNA/tracrRNA and Cas9 in a 1:1 ratio in 1x Cas9 buffer unless otherwise stated. Gels analysing RNase H cutting efficiencies contained 0.25 μM in 20 μL (5 pmol) of dsDNA with 1x, 2x, 5x and 10x gRNA/tracrRNA, Cas9 enzyme and RNase H enzyme (1x = 5U) in 1x RNase H buffer. Cas9 was added to the sample and allowed 15 min to form the complex before adding RNase H enzyme. All samples were incubated (37 °C for 1 h) followed by incubation (37 °C for 1 h) with Proteinase K (1 μL , 20 μM). Samples were then mixed with formamide (50% v/v) and the same 1x TBE running buffer and loaded onto the gel. Quantifications of the bands in the gel was performed using ImageJ.

5.6 Cell Studies

5.6.1 Cell culture

U2OS (human osteosarcoma cell line) were cultured in DMEM (no phenol red, HEPES-buffered, with glutamine, high glucose) (Thermo Fisher Scientific) supplemented with 10% FBS (Thermo Fisher Scientific) in a humidified incubator (5% CO_2 at 37 °C).

5.6.2 *In Vitro* Transcription (IVT)

TracrRNA was synthesised using IVT. A dsDNA template was formed from adding a single-stranded DNA template (5 μ M, 1 eq) to a single-stranded DNA complementary to the T7 promoter sequence (5 μ M, 1 eq) followed by H₂O to make a final concentration of 5 μ M. The MEGAScript™ T7 *in vitro*-transcription kit (Thermo Fisher Scientific) was used for the transcription reaction and contained ATP (75 mM, 2 μ L), CTP (75 mM, 2 μ L), GTP (75 mM, 2 μ L), UTP (75 mM, 2 μ L), T7 enzyme mix (2 μ L), hybridisation buffer (2 μ L, 10x: 100 mM Tris-HCl pH 8.5, 500 mM NaCl and 10 mM EDTA), dsDNA template-T7 mix (0.25 μ M, 1 μ L) and H₂O (7 μ L) to make a total volume of 20 μ L. The reaction was then left for 16 h at 37 °C and purified using the MEGAClear™ transcription Clean-up kit (Thermo Fisher Scientific). The reaction was mixed with Elution Solution (100 μ L), Binding Solution (350 μ L) and EtOH (100%, 250 μ L) and then transferred to the filter cartridge. After centrifugation (14,000 x *g* for 1 min), the filtrate was discarded, and the filter cartridge was washed and centrifuged twice more with Wash Solution (500 μ L). Elution Solution (50 μ L) was applied to the filter cartridge and heated (10 min at 70 °C). After centrifugation (14,000 x *g* for 1 min), the eluted IVT RNA was collected and analysed by UPLC-MS (Appendix A4.1).

5.6.3 Transfection

4 x 10⁴ U2OS cells were seeded on a Nunc™ Lab-Tek™ II 8-well chambered plates. After 24 h, the cells were washed with DPBS and the media was replaced with fresh Opti-MEM® medium. Lipofectamine® 2000 reagent (Invitrogen) (1 μ L) was diluted in Opti-

MEM® medium (25 µL). The dCas9-encoded DNA plasmid (0.5 µg, 0.5 µL from 1 µg/µL stock, gifted from Dr. Taemaitree) was diluted in Opti-MEM® medium (25 µL). The DNA solution was added to the lipid solution (1:1) and incubated (5 min at RT) before being added to the cells. After 8 h, the cells were washed again with DPBS and the media was replaced with fresh Opti-MEM® medium. Lipofectamine® RNAiMAX Reagent (Invitrogen) (1.5 µL) was diluted in Opti-MEM® medium (25 µL). The appropriate gRNA (10-100 pmols, 1-10 µL of 10 µM stock) was diluted in Opti-MEM® medium (25 µL). The RNA solution was added to the lipid solution (1:1) and incubated (5 min at RT) before being added to the cells. Cells were then allowed to recover and grow for 24 - 48 h with media replacement when necessary.

5.6.4 Confocal microscopy

Live-cell imaging was carried out on a Zeiss LSM 780 confocal microscope with a 63x 1.4 Numerical Aperture (NA) oil immersion objective at 37 °C with 5% CO₂ (Zeiss Ltd). The cells were excited using internal lasers and emission detected as follows; for dCas9-3xGFP, GFP was excited at 488 nm and the emission recorded from 493–542 nm. For Cy3-gRNA, Cy3 was excited at 543 nm and the emission recorded from 548–604 nm. Images were processed using ImageJ software and scale bar are set to 10 µm.

Appendix

A1 Appendix for Chapter 1

A1.1 Mechanisms of DNA synthesis steps

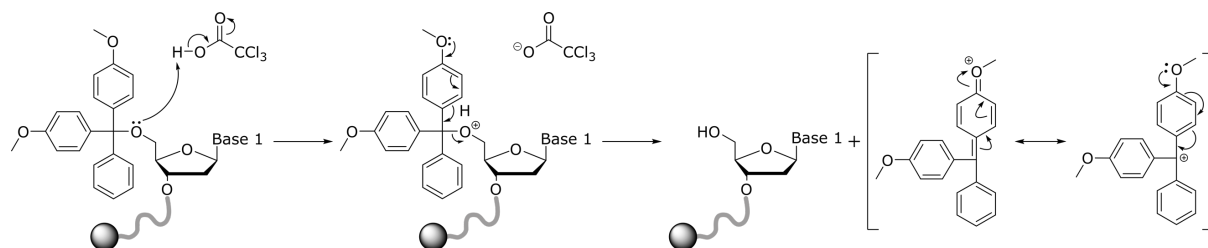


Figure A1: Detritylation. The removal of the DMT group from the 5'-position of one nucleotide using TCA to produce a free OH group for reaction with the next nucleotide. Reaction supported by production of DMT cation.

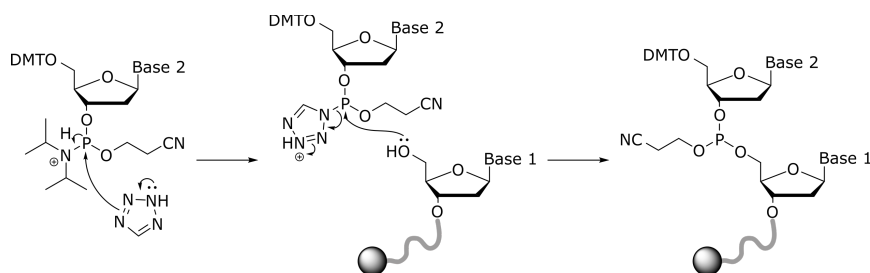


Figure A2: Activation and Coupling. Activation of the phosphoramidite using an activator by protonation of the diisopropylamino group before reaction with tetrazole to provide a good leaving group for the incoming nucleotide. Coupling of the activated nucleotide with the incoming nucleotide.

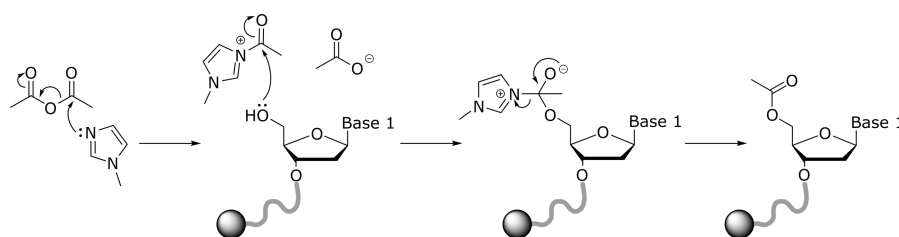


Figure A3: Capping. Capping of any free 5'-OH groups with an acetyl group produced from the reaction of acetic anhydride and *N*-methyl imidazole to avoid the production of truncated sequences.

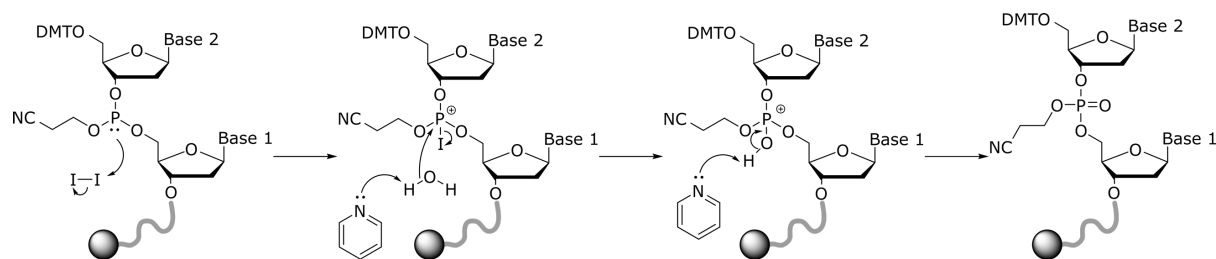


Figure A4: Oxidation. Oxidising the reactive P(III) species into a stable P(V) species using I_2 , pyridine and water to produce the protected DNA backbone.

A1.2 Mechanisms of DNA deprotection steps

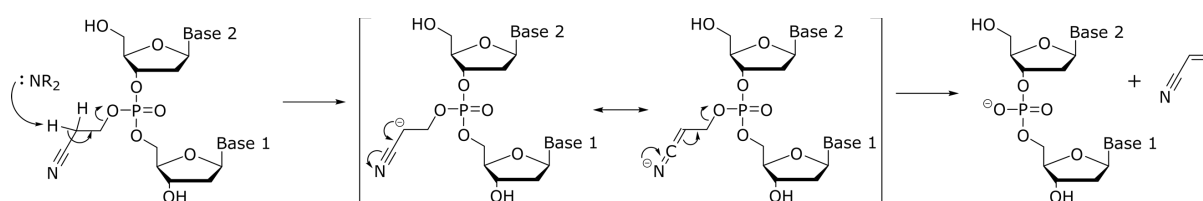


Figure A5: 2-Cyanoethyl deprotection. Removal of 2-cyanoethyl in basic conditions (NH_3 or 20% DEA) to produce the DNA backbone. Reaction is supported by production of anion through E1CB mechanism.

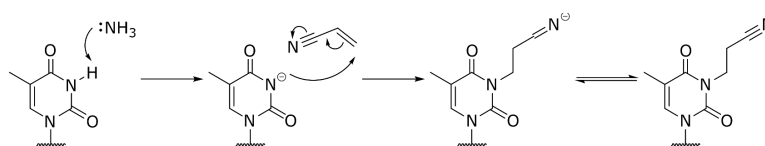


Figure A6: Formation of cyanoethyl adduct. Potential side reactions of cyanoethyl deprotection by-product reacting with bases in the ODN. This can be avoided by using less basic deprotection conditions (20% DEA).

A2 Appendix for Chapter 2

A2.1 UV absorbance and fluorescence spectra

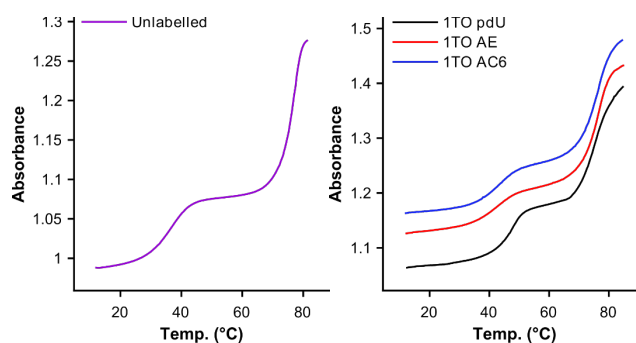


Figure A7: UV absorbance of linker comparison. Smoothed UV absorbance at 260 nm of unlabelled (left) and 1xTO pdU, 1xTO AE and 1xTO AC6 linkages (right) for TRIP-C in low Mg^{2+} buffer at pH 5.8. The first transition is triplex denaturation and the second (higher temperature) is duplex denaturation. Derivatives can be seen in Figure 2.12. Final T_m values are an average of two individual measurements in which each is an average of 5 ramps.

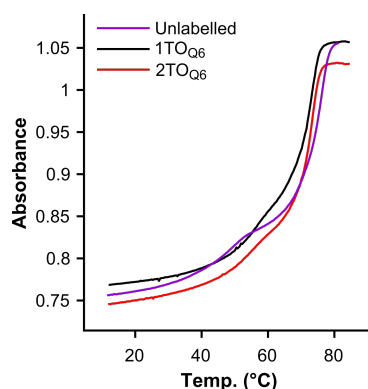


Figure A8: UV absorbance of TO_{Q6}. Smoothed UV absorbance at 260 nm of unlabelled (purple), 1xTO_{Q6} (black) and 2xTO_{Q6} (red) in TRIP-T in high Mg^{2+} buffer at pH 5.8. Derivatives can be seen in Figure 2.13.

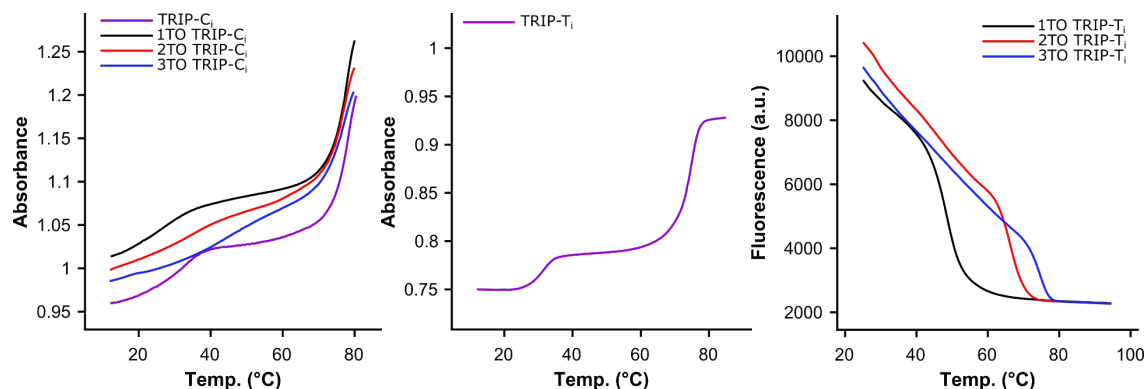


Figure A9: UV absorbance of TRIP-C_i and TRIP-T_i and fluorescence melting of TRIP-T_i at pH 5.8. Smoothed UV absorbance at 260 nm of unlabelled, 1xTO, 2xTO and 3xTO TRIP-C_i (left) and TRIP-T_i (centre) and smoothed fluorescence melting curve of 1xTO, 2xTO and 3xTO TRIP-T_i (right) in low Mg²⁺ buffer at pH 5.8. Fluorescence measured at $\lambda_{\text{ex}} = 521$ nm and $\lambda_{\text{em}} = 536$ nm. Derivatives can be seen in Figure 2.14.

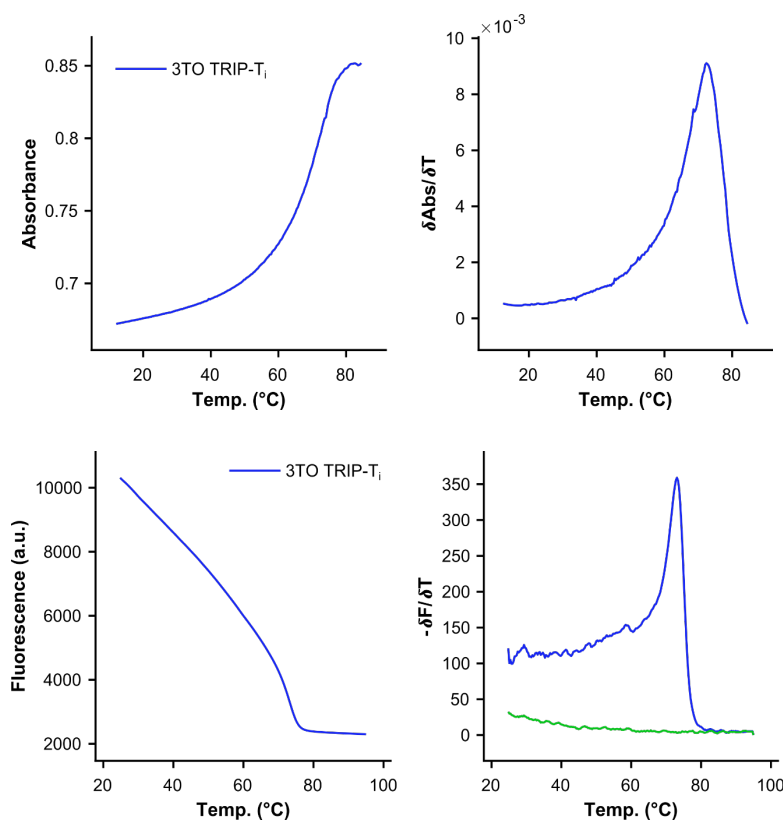


Figure A10: UV absorbance compared to fluorescence melting of 3xTO TRIP-T_i. Top: broad smoothed UV absorbance at 260 nm (left) and smoothed derivative (right) of 3xTO TRIP-T_i in high Mg²⁺ buffer at pH 7. Bottom: smoothed fluorescence melting curves (left) and smoothed derivative (right) of 3xTO TRIP-T_i in high Mg²⁺ buffer at pH 7. Fluorescence measured at $\lambda_{\text{ex}} = 521$ nm and $\lambda_{\text{em}} = 536$ nm.

Triplex	pH 5.8				pH 7			
	UV		Fluorescence		UV		Fluorescence	
	High Mg ²⁺	Low Mg ²⁺	High Mg ²⁺	Low Mg ²⁺	High Mg ²⁺	Low Mg ²⁺	High Mg ²⁺	Low Mg ²⁺
	T_m (°C)				T_m (°C)			
TRIP-T	43.2	31.9	-	-	39.3	28.4	-	-
1xTO TRIP-T	66.5	58.0	72.2	53.9	60.0	50.6	51.6	42.6
2xTO TRIP-T	n.d.	65.1	76.4	71.7	65.5	58.7	71.9	61.2
3xTO TRIP-T	n.d.	n.d.	78.0	76.7	n.d.	n.d.	74.0	72.0
TRIP-T _i Control	49.0 ^a	29.6 ^a	-	-	-	-	-	-
TRIP-T _i	50.3	31.2	-	-	23.0	-	-	-
1xTO TRIP-T _i	66.3	48.7	66.0	47.0	45.0	35.9	47.6	32.3
2xTO TRIP-T _i	n.d.	66.9	74.6	65.1	57.5	51.3	61.8	50.0
3xTO TRIP-T _i	n.d.	n.d.	76.1	73.2	n.d.	n.d.	73.4	65.9
TRIP-C	44.0	35.6	-	-	<i>nt</i>	<i>nt</i>	-	-
1xTO TRIP-C	50.6	46.2	49.1	44.1	25.7	19.9	n.d.	n.d.
2xTO TRIP-C	n.d.	55.1	54.3	52.4	30.4	26.0	n.d.	n.d.
3xTO TRIP-C	n.d.	61.8	62.5	64.0	34.6	34.8	n.d.	n.d.
TRIP-C _i Control	31.5 ^a	21.0 ^a	-	-	-	-	-	-
TRIP-C _i	39.5	26.6	-	-	<i>nt</i>	<i>nt</i>	-	-
1xTO TRIP-C _i	42.9	29.7	48.0	30.5	<i>nt</i>	<i>nt</i>	<i>nt</i>	<i>nt</i>
2xTO TRIP-C _i	45.3	36.7	n.d.	42.4	<i>nt</i>	<i>nt</i>	<i>nt</i>	<i>nt</i>
3xTO TRIP-C _i	56.4	44.7	n.d.	48.1	25.4	24.7	n.d.	n.d.
TRIP-TC	77.0	72.7	-	-	52.6	49.0	-	-
3xTO TRIP-TC	80+	80+	86.7	76.3	75.3	80+	73.2	73.4

Table A1: UV T_m data compared to fluorescence T_m data. T_m data were obtained from the maxima of the first derivatives of the UV or fluorescence (**bold**) melting curves. Final T_m values for UV melting are an average of two individual measurements in which each is an average of 5 ramps. Final T_m values for fluorescence melting are an average of two individual measurements. n.d. = not determined. *nt* = no triplex formation observed.

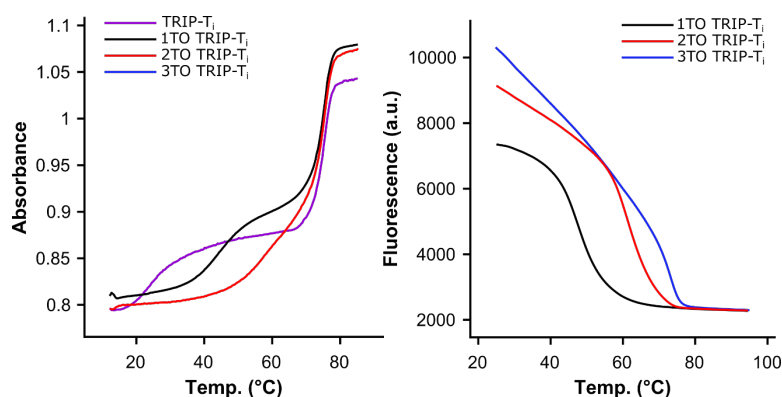


Figure A11: UV absorbance and fluorescence melting of TRIP-T_i in pH 7. Smoothed UV absorbance at 260 nm of unlabelled, 1xTO, 2xTO and 3xTO TRIP-T_i (left) and smoothed fluorescence melting curve of 1xTO, 2xTO and 3xTO TRIP-T_i (right) in high Mg²⁺ buffer at pH 7. Fluorescence measured at $\lambda_{\text{ex}} = 521$ nm and $\lambda_{\text{em}} = 536$ nm. Derivatives can be seen in Figure 2.16.

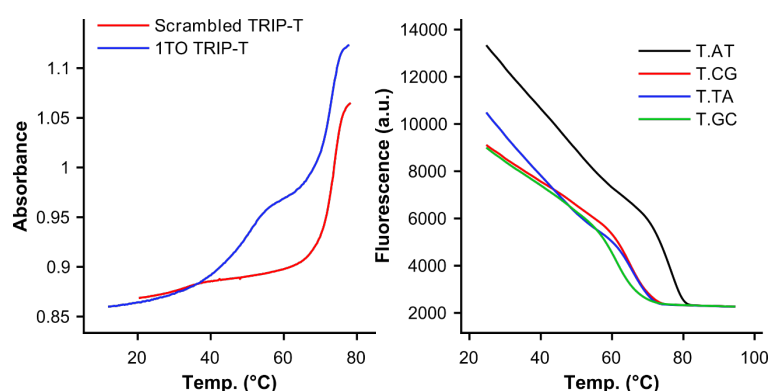


Figure A12: UV absorbance and fluorescence melting of controls at pH 7. Smoothed UV absorbance at 260 nm (left) of TFO-T mixed with a scrambled duplex (red) and the correct target duplex (blue) in low Mg²⁺ buffer, pH 7. Smoothed fluorescence melting curves (right) of 3xTO TFO-T mixed with duplexes containing mismatches to produce T.AT, T.CG, T.TA and T.GC triplets in low Mg²⁺ buffer, pH 7. Fluorescence measured at $\lambda_{\text{ex}} = 521$ nm and $\lambda_{\text{em}} = 536$ nm. Derivatives can be seen in Figure 2.17.

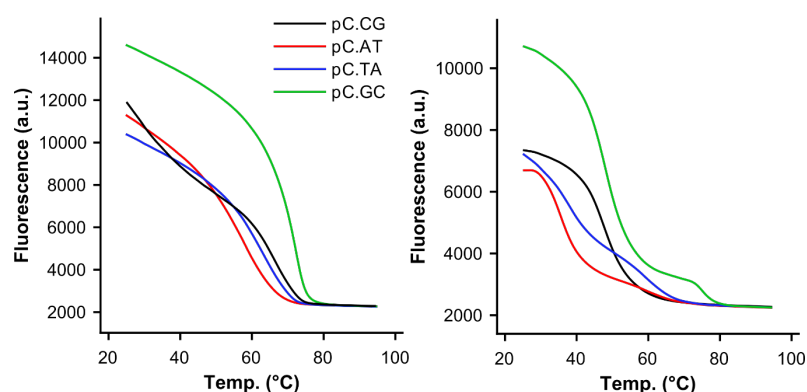


Figure A13: Fluorescence melting curves of pC.XX mismatches at pH 5.8 and pH 7. Smoothed fluorescence melting curves of 1xTO TFO-T_i mixed with duplexes containing triplet mismatches to produce pC.CG, pC.GC, pC.TA and pC.AT triplets in high Mg²⁺ buffer at pH 5.8 (left) and pH 7 (right). Fluorescence measured at $\lambda_{\text{ex}} = 521$ nm and $\lambda_{\text{em}} = 536$ nm. Derivatives can be seen in Figure 2.18.

Triplex	pH = 5.8					pH = 7.0				
	λ_{em}^{ss}	λ_{em}^T	F ^{ss} Area	F ^T Area	F ^T /F ^{ss}	λ_{em}^{ss}	λ_{em}^T	F ^{ss} Area	F ^T Area	F ^T /F ^{ss}
	nm		a.u.			nm		a.u.		
1xTO TRIP-T	534	533	757	15213	20.1	536	533	539	11977	22.2
2xTO TRIP-T	532	535	5420	37068	6.8	533	535	4002	32602	8.1
3xTO TRIP-T	532	535	8285	35821	4.3	531	535	4917	34086	6.9
1xTO TRIP-T _i	532	535	5922	42691	7.2	530	532	3467	38824	11.2
2xTO TRIP-T _i	535	535	4000	21601	5.4	532	535	2517	20097	8.0
1xTO TRIP-C	538	536	6011	30493	5.1	532	532	6695	41322	6.2
2xTO TRIP-C	564	538	4949	41512	8.4	533	535	5913	57494	9.7
3xTO TRIP-C	585	539	3845	33505	8.7	534	535	4438	49113	11.1
1xTO TRIP-C _i	533	535	7470	28682	3.8	—	—	—	—	—
2xTO TRIP-C _i	536	536	6550	25213	3.8	—	—	—	—	—
3xTO TRIP-C _i	533	536	7171	18004	2.5	583	538	4118	19425	4.7

Table A2: Fluorescence intensities measured via area under graph. Data from fluorescence intensity measurements calculated by integrating the total area under the fluorescence curve from 510 nm to 700 nm. λ_{em}^{ss} = maximum emission wavelength of ssTFO, λ_{em}^T = maximum emission wavelength of formed triplex, F^{ss} Area = fluorescence curve area of the ssTFO, F^T Area = fluorescence curve area of the triplex, F^T/F^{ss} = fluorescence intensity ratio calculated by comparison of the ssTFO and the formed triplex area. (See also Table 2.3).

A3 Appendix for Chapter 3

A3.1 UV absorbance and fluorescence spectra

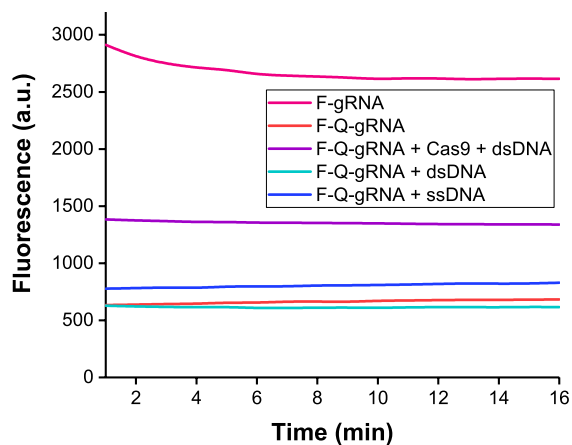


Figure A14: Fluorescence of the hairpin TAMRA-Q-gRNA (Design I). Fluorescence data of hairpin F-Q-gRNA labelled with TAMRA in Cas9 buffer at 37 °C over 15 min. Concentrations of the construct is 0.4 μ M in 20 μ L and an equimolar ratio was used for all additional elements. Fluorescence measured at $\lambda_{\text{ex}} = 575$ nm and $\lambda_{\text{em}} = 602$ nm.

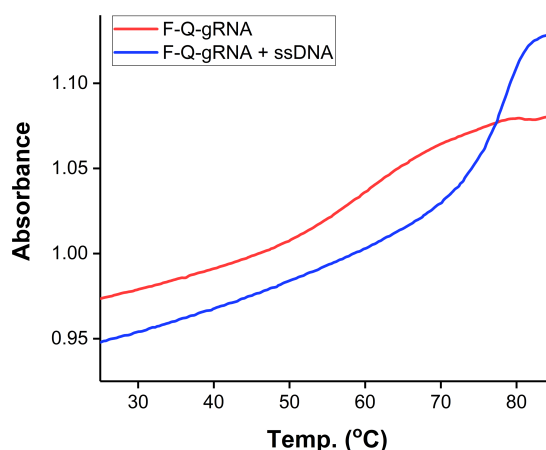


Figure A15: UV absorbance of the hairpin Cy3-Q-gRNA (Design I). UV absorbance measured at 260 nm. The measurements were performed in Cas9 buffer. Concentrations of the construct is 1 μ M in 1000 μ L and a 1:1 ratio was used for all additional elements. Derivatives can be seen in Figure 3.9 B.

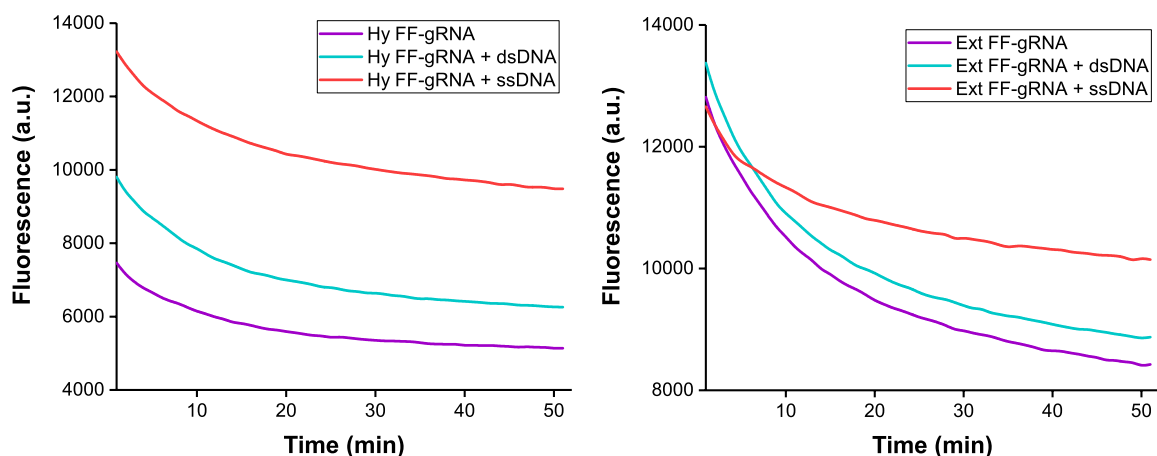


Figure A16: Fluorescence Controls for Hy and Ext Hy FF-gRNA (Design II). Fluorescence data of Hy FF-gRNA (left) and Ext Hy FF-gRNA (right) labelled with Cy3 in Cas9 buffer at 37 °C over 60 min. Concentrations of the construct is 0.5 μ M in 20 μ L and an equimolar ratio was used for all additional elements. Fluorescence measured at $\lambda_{\text{ex}} = 538$ nm and $\lambda_{\text{em}} = 554$ nm. (See also Figure 3.16).

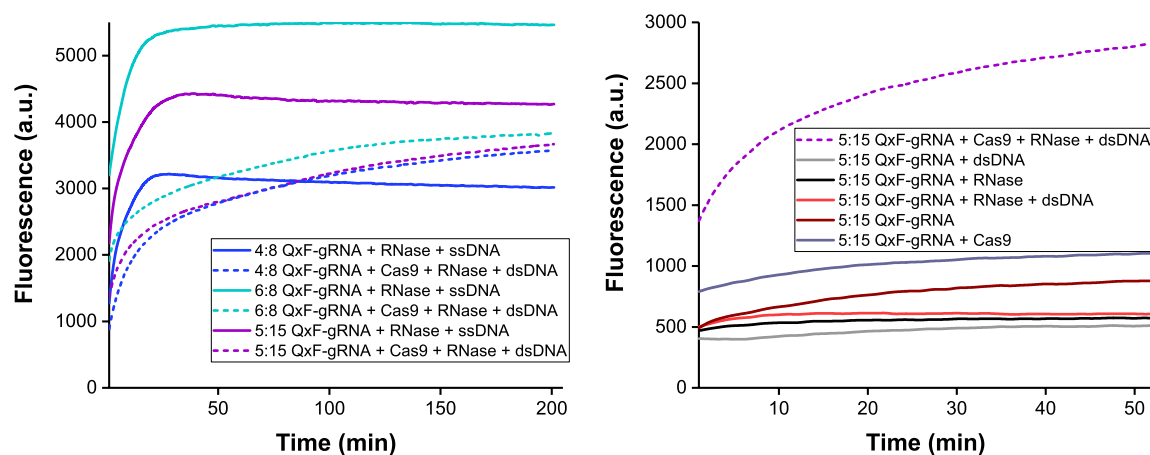


Figure A17: Fluorescence of RNase H QxF-gRNA constructs (Design III). Fluorescence data comparing maximum RNase fluorescence (solid line) and the experimental fluorescence (dashed line) of each construct in 4 h reaction time (left) and controls for RNase 5:15 construct labelled with Cy3 in Cas9 buffer at 37 °C over 60 min (right). Concentrations of the construct is 0.25 μ M in 20 μ L and an equimolar ratio was used for all additional elements except for Cas9 (5 eq, 1.25 μ M in 20 μ L). Cas9 was added to the gRNA and allowed 15 min to form the complex before RNase H (5U) was added. Fluorescence measured at $\lambda_{\text{ex}} = 538$ nm and $\lambda_{\text{em}} = 554$ nm. (See also Figure 3.21).

A4 Appendix for Chapter 5

A4.1 Mass Spectrometry data for synthesised oligonucleotides

ODN	Modification	Sequence (5' – 3')	Cal. Mass	Found Mass
TRIP-T	Unlabeled	TTTTTMTTMTMTMT	4497	4497.1
	1x pdU	TTTTTMTT ^P UMTMTMT	4536	4538.0
	1x TO	TTTTTMTT ^{TO} TMTMTMT	4923	4923.7
	2x pdU	TT ^P UTTMTT ^P UMTMTMT	4575	4576.1
	2x TO	TT ^{TO} TTTMTT ^{TO} TMTMTMT	5349	5349.0
	3x pdU	TT ^P UTTMTT ^P UMTM ^P UMT	4614	4615.2
	3x TO	TT ^{TO} TTTMTT ^{TO} TMTM ^{TO} TMT	5772	5773.2
	Pu Target	GGAAGGGGGAAAAAGAGAGAGAGGA GAGAGG	10546	10547.0
	Py Target	CCTCTCTCCTCTCTCTTTTCCCCC TTCC	9721	9722.2
TRIP-T _i	Control	TTTTTMTTMTMTMT	4193	4192.1
	Unlabeled	TTTTTMT ^P CTMTMTMT	4698	4699.3
	1x pdU	TTTTTMT ^P C ^P UMTMTMT	4559	4561.1
	1x TO	TTTTTMT ^P C ^{TO} TMTMTMT	4946	4947.0
	2x pdU	TT ^P UTTMT ^P C ^P UMTMTMT	4598	4600.2
	2x TO	TT ^{TO} TTTMT ^P C ^{TO} TMTMTMT	5372	5372.0
	3x pdU	TT ^P UTTMT ^P C ^P UMTM ^P UMT	4637	4638.4
	3x TO	TT ^{TO} TTTMT ^P C ^{TO} TMTM ^{TO} TMT	5798	5799.5
	Pu Target	GGAAGGGGGAAAAAGACAGAGAGAGGA GAGAGG	10522	10522.1
	Py Target	CCTCTCTCCTCTCTCTGTCTTTTCCCCC TTCC	9746	9746.4
TRIP-C	Unlabeled	CTMCCMMCCMCTMTMTC	5287	5288.2
	1x pdU	CTMCCMMCCMCC ^P UMTMTC	5326	5327.0
	1x TO	CTMCCMMCCMCC ^{TO} TMTMTC	5711	5712.4
	2x pdU	C ^P UMCCMMCCMCC ^P UMTMTC	5366	5363.5
	2x TO	C ^{TO} TMCCMMCCMCC ^{TO} TMTMTC	6138	6137.0
	3x pdU	C ^P UMCCMMCCMCC ^P UMTM ^P UC	5405	5405.0
	3x TO	C ^{TO} TMCCMMCCMCC ^{TO} TMTM ^{TO} TC	6564	6562.6
	Pu Target	GGAAAAAAGGAGGGGGGGGGGAGAGA GAAGAGAAGG	11549	11550.0

	Py Target	CCTTCTCTTCTCTCTCCCCCCCCCTCC TTTTTTCC	10574	10574.1
TRIP-C _i	Control	CTMCCTMCCMCCTMTMTC	5530	5530.5
	Unlabelled	CTMCC ^P CMCCMCCTMTMTC	5328	5328.4
	1x pdU	CTMCC ^P CMCCMCC ^P UMTMTC	5351	5352.0
	1x TO	CTMCC ^P CMCCMCC ^{TO} MTMTC	5738	5739.0
	2x pdU	C ^P UMCC ^P CMCCMCC ^P UMTMTC	5390	5391.0
	2x TO	C ^{TO} TMCC ^P CMCCMCC ^{TO} MTMTC	6164	6165.0
	3x pdU	C ^P UMCC ^P CMCCMCC ^P UMTM ^P UC	5429	5429.9
	3x TO	C ^{TO} TMCC ^P CMCCMCC ^{TO} MTM ^{TO} TC	6590	6591.4
	Pu Target	GGAAAAAAGGAGGGCGGGGGGAGAGA GAAGAGAAGG	11484	11485.3
	Py Target	CCTTCTCTTCTCTCTCCCCCGCCCTCC TTTTTTCC	10638	10637.5
TRIP-TC	3x pdU	M ^P UMTM ^P TTM ^P UMTTM ^P UM	4611	4611.0
	3x TO	M ^{TO} TM ^{TO} MTM ^{TO} TTM ^{TO} MTM ^{TO} TM	5770	5770.3
	Pu Target	GGAAGGGGAAGAGAGAAGAGAAGAGGGA GAGAGG	10562	10563.5
	Py Target	CCTCTCTCCCTCTTCTCTTCTCTCTTCCC TTCC	9706	9707.5
TRIP-C Varied Linker	1x AE	CTMCCMMCCMCC ^{ae} TMTMTC	5347	5345.6
	1x TO	CTMCCMMCCMCC ^{TO} TMTMTC	5734	5732.6
	1x AC6	CTMCCMMCCMCC ^{ac} TMTMTC	5442	5442.6
	1x TO	CTMCCMMCCMCC ^{TO} TMTMTC	5829	5828.4
Mismatch Duplexes	Pu-T	GGAAGGGGGAAAAAGATAGAGAGAGGA GAGAGG	10537	10538.5
	Py-A	CCTCTCTCCTCTCTCTCTATCTTTTTCCCC TTCC	9730	9731.7
	Pu-G	GGAAGGGGGAAAAAGAGGAGAGAGAGGA GAGAGG	10562	10563.0
	Py-C	CCTCTCTCCTCTCTCTCTCTTTTTCCCC TTCC	9706	9707.8
ATTO- TOTFO TRIP-T _i	5'-alkyne 1x pdU	alk-TTTTTMT ^P C ^P UMTM ^P MT	4719	4720.0
	5'-ATTO 1x TO	atto-TTTTTMT ^P C ^{TO} TM ^{TO} MTMT	5952	5950.6
	3'-alkyne 1x pdU	TTTTTMT ^P C ^P UMTM ^P MT-alk	4894	4894.0
	3'-ATTO 1x TO	TTTTTMT ^P C ^{TO} TM ^{TO} MTMT-atto	6126	6127.0
TRIP-T _i	Biotinylated Py	Bio- CCTCTCTCCCTCTCTCTGCTTTTTCCCC TTCC	10315	10317.1

Scrambled Duplex	Pu	GGAAGGGGG <u>AGAAGAAGAAGAGAAGGA</u> GAAAGG	10546	10547.0
	Py	CCTTTCTCCTTCTCTTCTTCTTCTCCCCC TTCC	9721	9722.0

Table A3: Mass spectrometry data for Chapter 2. A/G/C/T = DNA nucleotide, Pu = polypurine duplex strand, Py = polypyrimidine duplex strand, ^pU = 5-(1-propargylamino)-2'-deoxyuridine (pdU), ^pC = 5-(1-propynyl)-2'-deoxycytidine (pC), ^{TO}T = Thiazole orange attached, ^{ae}T = 2'-aminoethoxy-thymidine, ^{ac}T = Amino-C6-deoxythymidine, M = 5-methyl deoxycytidine, alk = 5'-alkyne linker, atto = ATTO647N, Bio = Biotin TEG, A/G/C/T = Triplex binding site, ~~A/G/C/T~~ = Mismatch site.

ODN	Modification	Sequence (5' – 3')	Cal. Mass	Found Mass
IVT Templates	IVT DNA Template (sgRNA-telomere)	AAAAGCACCGACTCGGTGCCACTTTTTTC AAGTTGATAACGGACTAGCCTTATTTTAA CTTGCTATTTCTAGCTCTAAACTAACCC TAACCCTAACCCCTACCCTATAGTGAGTC GTATTAGA	37318	37324.0
	T7 Promoter	TCTAATACGACTCACTATAG	6060	6059.0
	IVT DNA template (tracrRNA)	AAAAGCACCGACTCGGTGCCACTTTTTTC AAGTTGATAACGGACTAGCCTTATTTTAA CTTGCTATGCTCCCTATAGTGAGTCGTA TTAGA	27656	27657.0
Telomere	IVT sgRNA	pppGGGUAGGGUUAGGGUUAGGGUUAG UUUUAGAGCUAGAAAAUAGCAAGUUAAA AUAAGGCUAGUCCGUUAUCAACUUGAA AAAGUGGCACCGAGUCGGUGCUUUU	33677 (+3xMg)	33663.0
	Native gRNA	GGGUAGGGUUAGGGUUAGGGUUAGUU UUAGAGCUAGA	12072	12074.0
	IVT tracrRNA	pppGGGAGCAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUG GCACCGAGUCGGUGCUUUU	23217 (+3xMg)	23215.4
	Extended gRNA	GGGUAGGGUUAGGGUUAGGGUUAGUU UUAGAGCUAUGCUGUUUUUAAUCGACU GGCA-alk	18752	18753.0
	Extended tracrRNA	NH ₂ - UGCCAGUCGAUUAAAAACAGCAUAGCA AGUUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUG CUUUU	27842	27843.0
MUC4	Native gRNA	GGCGUGACCUGUGGAUGCUGGUUUUA GAGCUAGA	10987	10989.1
	Extended gRNA	GGCGUGACCUGUGGAUGCUGGUUUUA GAGCUAUGCUGUUUUUAAUCGACUGGC A-alk	17667	17668.0
	DNA Target	NH ₂ - GCCACTTCGGGCTCATGAGCCTCAGCA TCCACAGGTCACGCCACTGGCAGGCC CGTGGC	18523	18524.0
	ATTO647N DNA Target	atto- GCCACTTCGGGCTCATGAGCCTCAGCA TCCACAGGTCACGCCACTGGCAGGCC CGTGGC	19152	19153.2
	DNA PAM Strand	GCCACGGGGCCTGCCAGTGGCGTGACC TGTGGATGCTGAGGCTCATGAGCCCGA AGTGGC	18616	18616.0
	Hairpin Q-crRNA	acgccac ^{QT} -HHHH- G ^{ac} TGGCGUGACCUGUGGAUGCUGGUU UUAGAGCUAGA	16785	16787.0
	crRNA	G ^{ac} TGGCGUGACCUGUGGAUGCUGGUU UUAGAGCUAGA	12125	12126.0

MUC4 Hairpin F-Q	Hairpin Cy3-Q-crRNA	acgccac ^{QT} -HHHH- G ^{Cy3} TGGCGUGACCUGUGGAUGCUGGU UUUAGAGCUAGA	17224	17224.0
	Cy3-crRNA	G ^{Cy3} TGGCGUGACCUGUGGAUGCUGGU UUUAGAGCUAGA	12564	12564.0
	Hairpin TAMRA-Q-crRNA	acgccac ^{QT} -HHHH- G ^{TAM} TGGCGUGACCUGUGGAUGCUGGU UUUAGAGCUAGA	17199	17198.0
	TAMRA-crRNA	G ^{TAM} TGGCGUGACCUGUGGAUGCUGGU UUUAGAGCUAGA	12538	12538.0
	TO-crRNA	G ^{TO} TGGCGUGACCUGUGGAUGCUGGU UUUAGAGCUAGA	12512	12513.0
MUC4 HyBeacon	Hy crRNA	GG ^{ae} CGUGAC ^{ae} CUGUGGAUGCUGGUUU UAGAGCUAGA	11103	11102.3
	Hy FF-crRNA	GG ^{Cy3} CGUGAC ^{Cy3} CUGUGGAUGCUGGUUU UUUAGAGCUAGA	11981	11981.0
	Extended Hy crRNA	g ^{ae} Ccugc ^{ae} CaguGGCGUGACCUGUGGAU GCUGGUUUUAGAGCUAGA	14413	14413.4
	Extended Hy FF-crRNA	g ^{Cy3} Ccugc ^{Cy3} CaguGGCGUGACCUGUGGA UGCUGGUUUUAGAGCUAGA	15291	15292.0
	4:8 crRNA	CUGC ^{ae} CaguGGCGUGACCUGUGGAUGC UGGUUUUAGAGCUAGA	13634	13635.0
MUC4 RNase H	4:8 xF-crRNA	CUGC ^{Cy3} CaguGGCGUGACCUGUGGAUG CUGGUUUUAGAGCUAGA	14073	14074.0
	4:8 QxF-crRNA	Q- CUGC ^{Cy3} CaguGGCGUGACCUGUGGAUG CUGGUUUUAGAGCUAGA	14585	14586.0
	6:8 crRNA	CUGCCAg ^{ae} GG ^{ae} CGUGACCUGUGGAUGC UGGUUUUAGAGCUAGA	13620	13621.0
	6:8 xF-crRNA	CUGCCAg ^{ae} GG ^{Cy3} CGUGACCUGUGGAUG CUGGUUUUAGAGCUAGA	14059	14060.0
	6:8 QxF-crRNA	Q- CUGCCAg ^{ae} GG ^{Cy3} CGUGACCUGUGGAUG CUGGUUUUAGAGCUAGA	14571	14572.0
	5:15 crRNA	ACGGGg ^{ae} CcugccaguGGCGUGACCUGU GGAUGCUGGUUUUAGAGCUAGA	16039	16040.0
	5:15 xF-crRNA	ACGGGg ^{Cy3} CcugccaguGGCGUGACCUGU GGAUGCUGGUUUUAGAGCUAGA	16477	16479.0
	5:15 QxF-crRNA	Q- ACGGGg ^{Cy3} CcugccaguGGCGUGACCUGU GGAUGCUGGUUUUAGAGCUAGA	16990	16990.0
	6:16 crRNA	GGGUUAgggu ^{ae} TaggguUAGGGUUAGGG UUAGGGUUAGUUUUAGAGCUAUGCUG UUUUUAAUCGACUGGCA	22954	22959.0
	6:16 xF-crRNA	GGGUUAgggu ^{Cy3} TaggguUAGGGUUAGGG UUAGGGUUAGUUUUAGAGCUAUGCUG UUUUUAAUCGACUGGCA	23398	23396.0

Telomere RNase H	Q-			
	6:16 QxF- crRNA	GGGUUAagggu ^{Cy3} TaggguUAGGGUUAGGG UUAGGGUUAGUUUUAGAGCUAUGCUG UUUUUAAUCGACUGGCA	23908	23907.0
	DNA Target	GCCACTTCGGGCTCATGAGCCTTAACCC TAACCCTAACCCCTAACCCCTAACCCCTAAC CCCGCGTGGCCATCAGCGAGGT	23697	23697.0
DNA PAM Strand	DNA PAM	ACCTCGCTGATGGCCACGCGGGGTTAG GGTTAGGGTTAGGGTTAGGGTTAGGGT	24381	24382.0
	Strand	TAAGGCTCATGAGCCCGAAGTGGC		

Table A4: Mass spectrometry data for Chapter 3. A/G/C/T = DNA nucleotide, A/G/C/U = RNA nucleotide, a/g/c/u = 2'-O-Me nucleotide, alk = 5'-alkyne linker, NH₂ = 5'-amino linker, atto = ATTO647N, ppp = triphosphate, H = HEG, ^{ae}C = 2'-aminoethoxy methyl cytosine, ^{ac}T = Amino-C6-deoxythymidine, Q = 5'-BHQ-2, ^QT = 5-BHQ2-deoxythymidine, ^{Cy3}T/C = Cy3 attached, ^{TO}T = Thiazole orange attached, A/G/C/T = crRNA binding site, **AGG** = PAM region.

ODN	Modification	Sequence (5' – 3')	Cal. Mass	Found Mass
Azido dT-Triester	10-mer DNA Crude	TTTTTTTTTTT	2980	2979.5
	11-mer RNA (TC) Crude	UUUUUAUUUUU	3328	3328.2
	10-mer DNA + Az dT	N -TTTTTTTTTTTT	3420 (+PG)	3419.5
	11-mer RNA + Az dT	N-T UUUUUAUUUUU	3768 (+PG)	3328.0
DMT dT-Triester	Mixed 10-mer DNA Crude	CGTCTACTCT	2954	2952.5
	Mixed 10-mer DNA + DMT dT	TCGTCTACTCT	3369 (+PG)	2954.6
Reverse Phos.	10-mer DNA Crude	CGTCTACTCT	2954	2953.7
	11-mer DNA Crude	CGTCTACTGCT	3283	3282.5
Azido dT-CEP	T-rich DNA + Az dT	N -TTTTTTTTTTTT	3309	3309.0
	Mixed DNA + Az dT	N-TC GTCTACTCT	3283	3284.0
Azido U-CEP	U-rich RNA (TC) Crude	UUUUUGUUUUU	3344	3344.2
	Mixed RNA (TC) + Az U	CUGAGCAAGAGCGAGCAGCU	6468	6468.0
	U-rich RNA + Az U	N-U UUUUUGUUUUU	3673	See Section 4.3.3.2
	Mixed RNA (TC) + Az U	N-UC UGAGCAAGAGCGAGCAGCU	6797	

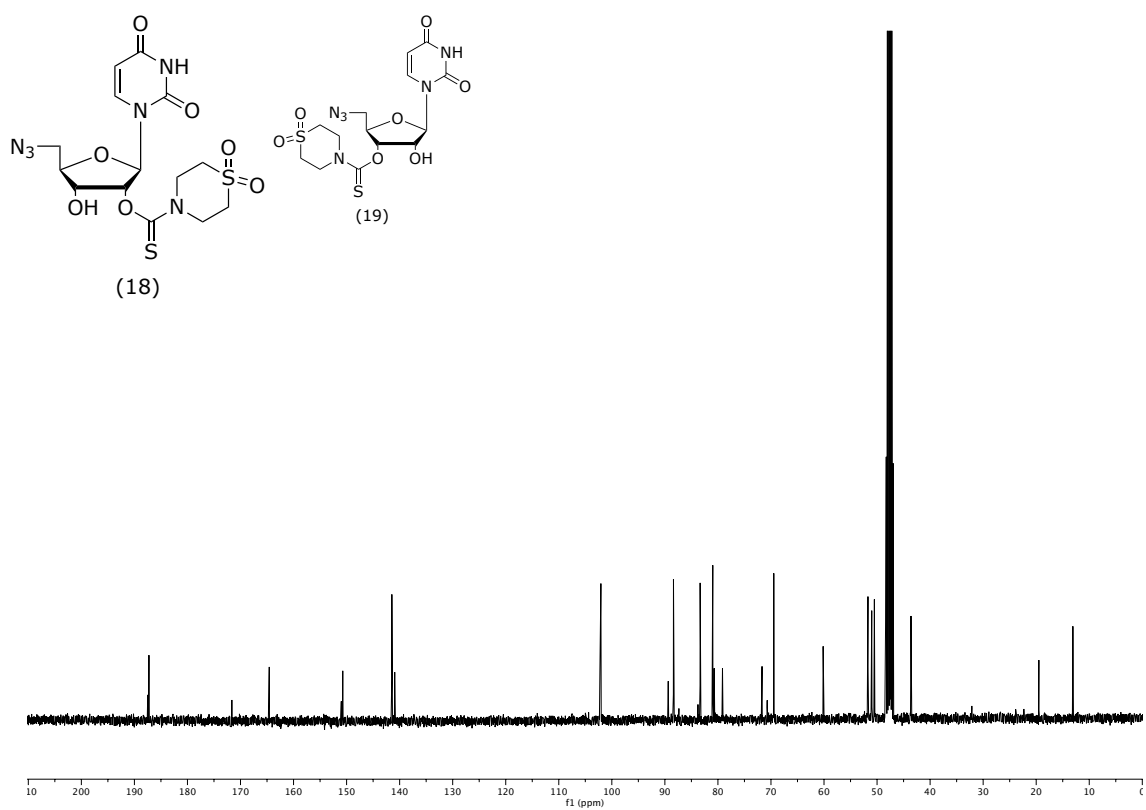
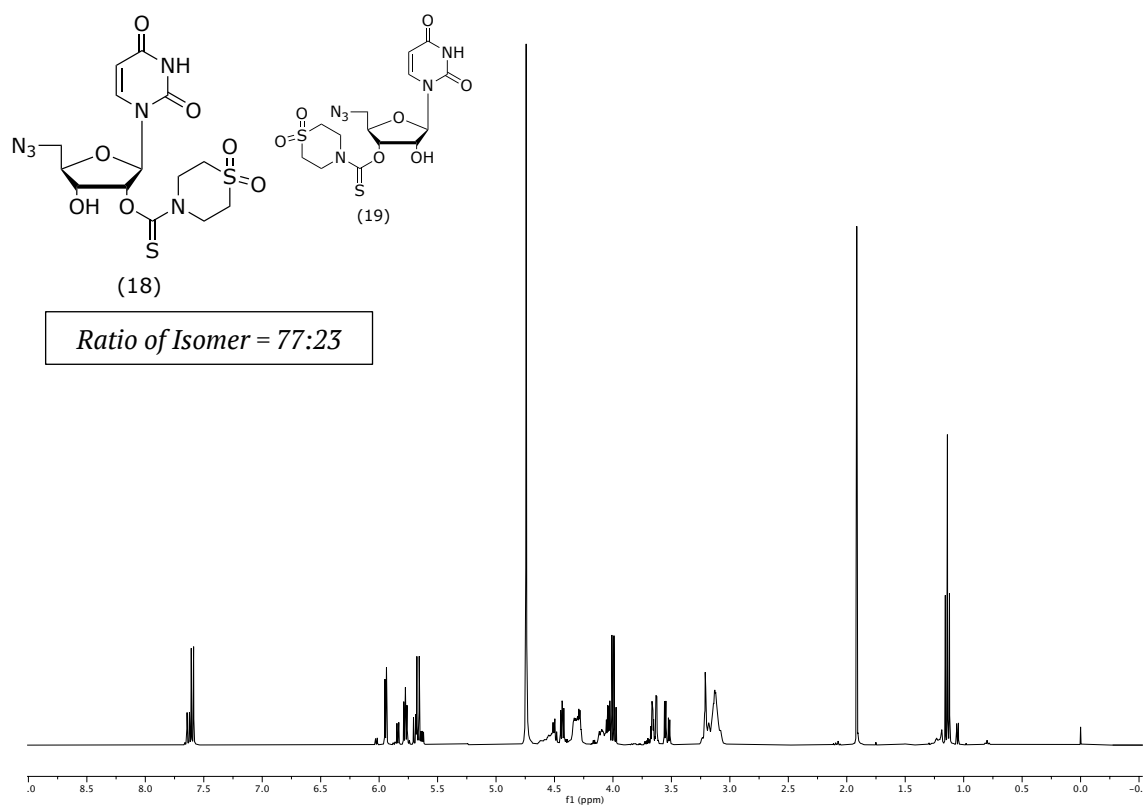
Table A5: Mass spectrometry data for Chapter 4. A/G/C/T = DNA nucleotide, A/G/C/U = RNA nucleotide, N-T or Az dT= 5'-N₃-deoxythymidine, N-U or Az U = 5'-N₃-uridine, +PG = Mass including phosphotriester protecting group.

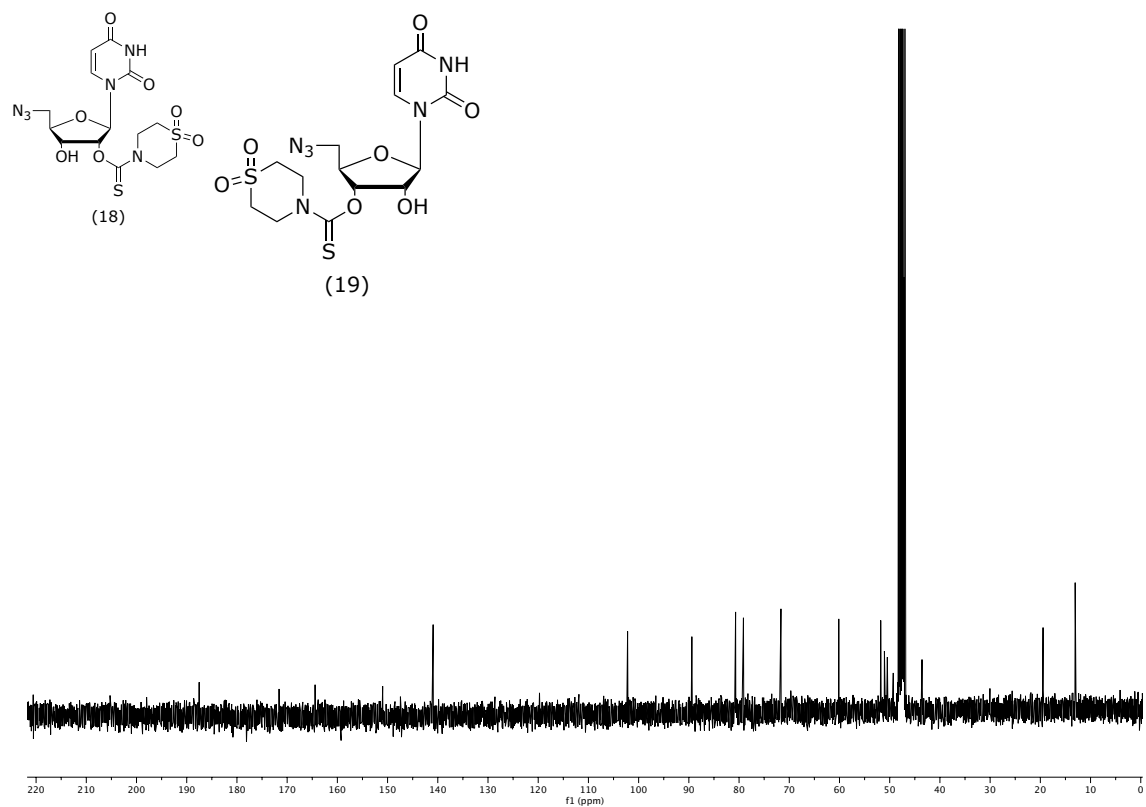
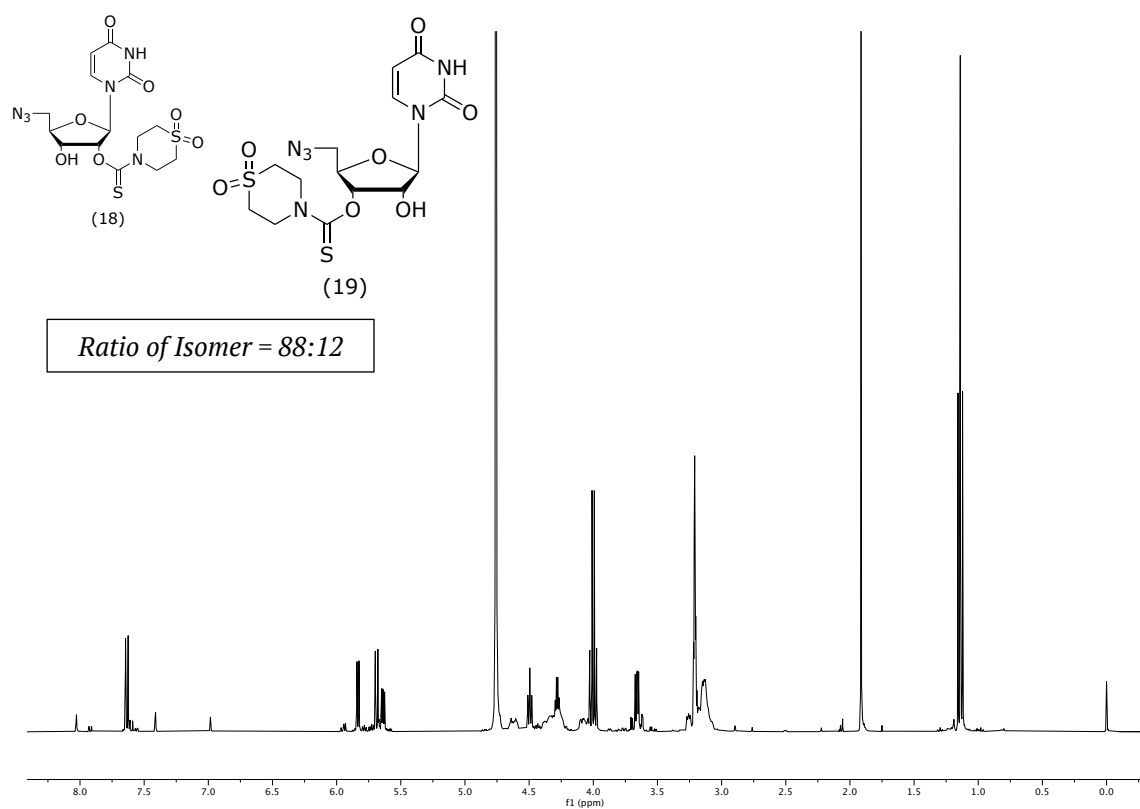
Reaction	Sequence (5' – 3')	Monomer	SM Mass	Prod Mass
A1	CGTCTACTCT	DMT-T	2954	3258
A2	CGTCTACTCT	DMT-T		
B1	CGTCTACTCT	DMT-T	2954	3258
B2	CGTCTACTCT	DMT-T		
B3	CGTCTACTCT	DMT-T		
B4	CGTCTACTCT	DMT-T		
C1	CGTCTACTGCT	DMT-T	3283	3586
C2	CGTCTACTGCT	DMT-T		
C3	CGTCTACTGCT	DMT-T		
D1	CGTCTACTGCT	DMT-C	3283	3571
D2	CGTCTACTGCT	DMT-C		
D3	CGTCTACTGCT	DMT-C		
D4	CGTCTACTGCT	DMT-C		
D5	CGTCTACTGCT	DMT-C		
E1	CGTCTACTGCT	DMT-T	3283	3586
E2	CGTCTACTGCT	Azido-T		3613
F1	CGTCTACTGCT	DMT-T	3283	3586
F2	CGTCTACTGCT	Azido-T		3613
F3	CGTCTACTGCT	DMT-T		3586
F4	CGTCTACTGCT	DMT-T		3586
G1	CGTCTACTGCT	Azido-T	3283	3613
G2	CGTCTACTGCT	Azido-T		

Table A6: Mass spectrometry data for Chapter 4 (reverse phosphitylation). For observed mass, see Section 4.3.2. A/G/C/T = DNA nucleotide. SM Mass = calculated mass of DNA without reaction. Prod Mass = calculated mass of DNA reacted with corresponding monomer.

A4.2 NMR Spectra of Novel Compounds

A4.2.1 ^1H NMR spectra and ^{13}C NMR spectra of 18



A4.2.2 ^1H NMR spectra and ^{13}C NMR spectra of 19

References

1. Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat. Rev. Genet.* **3**, 662–673 (2002).
2. Laszlo, A. H. *et al.* Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 18904–9 (2013).
3. Arand, J. *et al.* In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet.* **8**, e1002750 (2012).
4. Chargaff, E. & Vischer, E. The composition of the desoxypentose nucleic acids of thymus and spleen. *J. Biol. Chem.* **177**, 405–16 (1949).
5. Franklin, R. E. & Gosling, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature* **171**, 740–741 (1953).
6. Wilkins, M. H. F., Seeds, W. E., Stokes, A. R. & Wilson, H. R. Helical Structure of Crystalline Deoxypentose Nucleic Acid. *Nature* **172**, 759–762 (1953).
7. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).
8. Hoogsteen, K. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta. Cryst.* **16**, 907–916 (1963).
9. Crick, F. H. C. Codon—anticodon pairing: The wobble hypothesis. *J. Mol. Biol.* **19**, 548–555 (1966).
10. Portmann, S., Usman, N. & Egli, M. The Crystal Structure of r(CCCCGGGG) in

- Two Distinct Lattices. *Biochemistry* **34**, 7569–7575 (1995).
11. Drew, H. R. *et al.* Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci.* **78**, 2179–2183 (1981).
 12. Luo, Z., Dauter, M. & Dauter, Z. Phosphates in the Z-DNA dodecamer are flexible, but their P-SAD signal is sufficient for structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 1790–1800 (2014).
 13. Belmont, P., Constant, J. F. & Demeunynck, M. Nucleic acid conformation diversity: From structure to function and regulation. *Chem. Soc. Rev.* **30**, 70–81 (2001).
 14. Taemaitree, L. Chemically Modified Guide RNAs to Expand CRISPR Functionality. (University of Oxford, 2018).
 15. Neidle, S. RNA Structures and Their Diversity. in *Principles of Nucleic Acid Structure* 204–217 (Elsevier, 2008).
 16. Wang, A. H. J. *et al.* Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680–686 (1979).
 17. Doluca, O., Withers, J. M. & Filichev, V. V. Molecular engineering of guanine-rich sequences: Z-DNA, DNA triplexes, and G-quadruplexes. *Chem. Rev.* **113**, 3044–3083 (2013).
 18. Marsh, R. E., Bierstedt, R. & Eichhorn, E. L. The crystal structure of cytosine-5-acetic acid. *Acta Crystallogr.* **15**, 310–316 (1962).
 19. Mergny, J. L. Fluorescence energy transfer as a probe for tetraplex formation: The i- motif. *Biochemistry* **38**, 1573–1581 (1999).

20. Merrifield, R. B. Solid-Phase Peptide Synthesis. in *Advances in Enzymology and Related Areas of Molecular Biology, Volume 32* 221–296 (1969).
21. Merrifield, R. B. Solid-Phase Peptide Synthesis. III. An Improved Synthesis of Bradykinin. *Biochemistry* **3**, 1385–1390 (1964).
22. Gait, M. J. & Sheppard, R. C. Rapid synthesis of oligodeoxyribonucleotides: a new solid-phase method. *Nucleic Acids Res.* **4**, 1135–1158 (1977).
23. Gait, M. J. *Oligonucleotide Synthesis: a Practical Approach*. (IRL Press, 1984).
24. Beaucage, S. L. & Caruthers, M. H. Deoxynucleoside phosphoramidites-A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* **22**, 1859–1862 (1981).
25. Matteucci, M. D. & Caruthers, M. H. Synthesis of deoxyoligonucleotides on a polymer support. *J. Am. Chem. Soc.* **103**, 3185–3191 (1981).
26. Blackburn, G. M., Gait, M. J., Loakes, D. & Williams, D. M. Synthesis of Oligonucleotides. in *Nucleic Acids in Chemistry and Biology* 143–166 (Royal Society of Chemistry, 2007).
27. ATDBio Ltd. Solid-phase oligonucleotide synthesis. (2005). Available at: <https://www.atdbio.com/content/17/Solid-phase-oligonucleotide-synthesis>. (Accessed: 18th December 2019)
28. Usman, N., Pon, R. T. & Ogilvie, K. K. Preparation of ribonucleoside 3'-O-phosphoramidites and their application to the automated solid phase synthesis of oligonucleotides. *Tetrahedron Lett.* **26**, 4567–4570 (1985).
29. Gasparutto, D. *et al.* Chemical synthesis of a biologically active natural tRNA

- with its minor bases. *Nucleic Acids Res.* **20**, 5159–5166 (1992).
30. Westman, E. & Stromberg, R. Removal of f-butyldimethylsilyl protection in RNA-synthesis. Triethylamine trihydrofluoride (TEA, 3HF) is a more reliable alternative to tetrabutylammonium fluoride (TBAF). *Nucleic Acids Res.* **22**, 2430–2431 (1994).
 31. Ogilvie, K. K., Usman, N., Nicoghossian, K. & Cedergren, R. J. Total chemical synthesis of a 77-nucleotide-long RNA sequence having methionine-acceptance activity (solid-phase synthesis of tRNA/ribonucleoside phosphoramidites/amino-acylation). *Proc. Natl. Acad. Sci. USA* **85**, 5764–5768 (1988).
 32. Dellinger, D. J. *et al.* Streamlined process for the chemical synthesis of RNA using 2'-O-thionocarbamate-protected nucleoside phosphoramidites in the solid phase. *J. Am. Chem. Soc.* **133**, 11540–11556 (2011).
 33. Lakowicz, J. *Principles of Fluorescence Spectroscopy*. (Springer US, 2006).
 34. Johansson, M. K. & Cook, R. M. Intramolecular dimers: A new design strategy for fluorescence-quenched probes. *Chem. - A Eur. J.* **9**, 3466–3471 (2003).
 35. Marras, S. A. E. Selection of fluorophore and quencher pairs for fluorescent nucleic acid hybridization probes. *Methods Mol. Biol.* **335**, 3–16 (2006).
 36. Clegg, R. M. Fluorescence resonance energy transfer. *Curr. Opin. Biotechnol.* **6**, 103–110 (1995).
 37. Vladimir, D. V. *Fluorescent Energy Transfer Nucleic Acid Probes*. (2006).
 38. Haugland, R. P., Yguerabide, J. & Stryer, L. Dependence of the Kinetics of

- singlet-singlet energy transfer on spectral overlap. *Proc. Natl. Acad. Sci.* **63**, 23–30 (1969).
39. Lilley, D. M. & Wilson, T. J. Fluorescence resonance energy transfer as a structural tool for nucleic acids. *Curr. Opin. Chem. Biol.* **4**, 507–517 (2000).
40. Jares-Erijman, E. & Jovin, T. M. Determination of DNA helical handedness by fluorescence resonance energy transfer. *J. Mol. Biol.* **257**, 597–617 (1996).
41. Clegg, R. M. FRET tells us about proximities, distances, orientations and dynamic properties. *Rev. Mol. Biotechnol.* **82**, 177–179 (2002).
42. Boutorine, A. S., Novopashina, D. S., Krasheninina, O. A., Nozeret, K. & Venyaminova, A. G. Fluorescent probes for nucleic acid visualization in fixed and live cells. *Molecules* **18**, 15357–15397 (2013).
43. Fernández-Suárez, M. & Ting, A. Y. Fluorescent probes for super-resolution imaging in living cells. *Nature Reviews Molecular Cell Biology* **9**, 929–943 (2008).
44. Bao, G., Rhee, W. J. & Tsourkas, A. Fluorescent Probes for Live-Cell RNA Detection. *Annu. Rev. Biomed. Eng.* **11**, 25–47 (2009).
45. Whitcombe, D., Theaker, J., Guy, S. P., Brown, T. & Little, S. Detection of PCR products using self-probing amplicons and fluorescence. *Nat. Biotechnol.* **17**, 804–807 (1999).
46. Ranasinghe, R. T. & Brown, T. Fluorescence based strategies for genetic analysis. *Chem. Commun.* **11**, 5487–5502 (2005).
47. Wittwer, C. T., Herrmann, M. G., Moss, A. A. & Rasmussen, R. P. Continuous Fluorescence Monitoring of Rapid Cycle DNA Amplification. *Biotechniques* **22**,

- 130–138 (1997).
48. Tyagi, S. & Kramer, F. R. Molecular Beacons: Probes that Fluoresce upon Hybridization. *Nat Biotech* **14**, 303–308 (1996).
 49. French, D. J., Archard, C. L., Brown, T. & McDowell, D. G. HyBeacon™ probes: A new tool for DNA sequence detection and allele discrimination. *Mol. Cell. Probes* **15**, 363–374 (2001).
 50. Sanborn, M. E., Connolly, B. K., Gurunathan, K. & Levitus, M. M. Fluorescence properties and photophysics of the sulfoindocyanine Cy3 linked covalently to DNA. *J. Phys. Chem. B* **111**, 11064–11074 (2007).
 51. Weisbrod, S. H. & Marx, A. Novel strategies for the site-specific covalent labelling of nucleic acids. *Chem. Commun.* **44**, 5675 (2008).
 52. Felsenfeld, G., Davies, D. R. & Rich, A. Formation of a three stranded polynucleotide molecule. *J. Am. Chem. Soc.* **79**, 2023–2024 (1957).
 53. Moser, H. E. & Dervan, P. B. Sequence-specific cleavage of double helical DNA by triple helix formation. *Science* **238**, 645–650 (1987).
 54. Hoogsteen, K. The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. *Acta. Cryst.* **12**, 822–823 (1959).
 55. Chan, P. P. & Glazer, P. M. Triplex DNA: Fundamentals, advances, and potential applications for gene therapy. *J. Mol. Med.* **75**, 267–282 (1997).
 56. Mukherjee, A. & Vasquez, K. M. Triplex technology in studies of DNA damage, DNA repair, and mutagenesis. *Biochimie* **93**, 1197–1208 (2011).
 57. Frank-Kamenetskii, M. D. & Mirkin, S. M. Triplex DNA structures. *Annu. Rev.*

- Biochem.* **64**, 65–95 (1995).
58. Asensio, J. L., Brown, T. & Lane, A. N. Solution conformation of a parallel DNA triple helix with 5' and 3' triplex–duplex junctions. *Structure* **7**, 1–11 (1999).
59. Zain, R. *et al.* Design of a triple-helix-specific cleaving reagent. *Chem. Biol.* **6**, 771–777 (1999).
60. Morgan, A. R. & Wells, R. D. Specificity of the three-stranded complex formation between double-stranded DNA and single-stranded RNA containing repeating nucleotide sequences. *J. Mol. Biol.* **37**, 63–80 (1968).
61. Hanvey, J. C., Shimizu, M. & Wells, R. D. Site-specific inhibition of EcoRI restriction/modification enzymes by a DNA triple helix. *Nucleic Acids Res.* **18**, 157–161 (1990).
62. Taniguchi, Y. & Sasaki, S. An efficient antigene activity and antiproliferative effect by targeting the Bcl-2 or survivin gene with triplex forming oligonucleotides containing a W-shaped nucleoside analogue (WNA-βT). *Org. Biomol. Chem.* **10**, 8336 (2012).
63. Maher, L., Wold, B. & Dervan, P. Inhibition of DNA binding proteins by oligonucleotide-directed triple helix formation. *Science* **245**, 725–730 (1989).
64. Noonberg, S. B., Scott, G. K., Hunt, C. A., Hogan, M. E. & Benz, C. C. Inhibition of transcription factor binding to the HER2 promoter by triplex-forming oligodeoxyribonucleotides. *Gene* **149**, 123–126 (1994).
65. Hu, Y., Cecconello, A., Idili, A., Ricci, F. & Willner, I. Triplex DNA Nanostructures: From Basic Properties to Applications. *Angew. Chem., Int. Ed.*

- 56, 15210–15233 (2017).
66. Cadet, J. & Wagner, J. R. DNA Base Damage by Reactive Oxygen Species, Oxidizing Agents, and UV Radiation. *Cold Spring Harb. Perspect. Biol.* **5**, a012559–a012559 (2013).
67. Kellett, A., Molphy, Z., Slator, C., McKee, V. & Farrell, N. P. Molecular methods for assessment of non-covalent metallodrug-DNA interactions. *Chem. Soc. Rev.* **48**, 971–988 (2019).
68. Strobel, S. A. & Dervan, P. B. Cooperative site specific binding of oligonucleotides to duplex DNA. *J. Am. Chem. Soc.* **111**, 7286–7287 (1989).
69. Strobel, S. a & Dervan, P. B. Site-specific cleavage of a yeast chromosome by oligonucleotide-directed triple-helix formation. *Science* **249**, 73–75 (1990).
70. Boidot-Forget, M., Chassignol, M., Takasugi, M., Thuong, N. T. & Hélène, C. Site-specific cleavage of single-stranded and double-stranded DNA sequences by oligodeoxyribonucleotides covalently linked to an intercalating agent and an EDTA-Fe chelate. *Gene* **72**, 361–371 (1988).
71. Marchand, C. *et al.* A new family of sequence-specific DNA-cleaving agents directed by triple-helical structures: benzopyridoindole-EDTA conjugates. *Chem. - A Eur. J.* **6**, 1559–63 (2000).
72. Vorobjev, P., Tchaika, O. & Zarytova, V. Efficient cleavage of DS DNA by bleomycin conjugated via hexaethylene glycol linker to triplex-forming oligonucleotides. *Nucleosides. Nucleotides Nucleic Acids* **23**, 1047–1051 (2004).
73. Pavlova, A. S., Vorobyev, P. E. & Zarytova, V. F. Direct site-specific cleavage of

- double-stranded DNA by conjugates of bleomycin A5 with triplex-forming oligonucleotide. *Russ. J. Bioorganic Chem.* **35**, 197–206 (2009).
74. Bales, B. C., Pitié, M., Meunier, B. & Greenberg, M. M. A minor groove binding copper-phenanthroline conjugate produces direct strand breaks via β -elimination of 2-deoxyribonolactone. *J. Am. Chem. Soc.* **124**, 9062–9063 (2002).
75. Sigman, D. S., Graham, D. R., D'Aurora, V. & Stern, A. M. Oxygen-dependent cleavage of DNA by the 1,10-phenanthroline . cuprous complex. Inhibition of Escherichia coli DNA polymerase I. *J. Biol. Chem.* **254**, 12269–72 (1979).
76. Fantoni, N., Lauria, T. & Kellett, A. Genome Engineering with Synthetic Copper Nucleases. *Synlett* **26**, 2623–2626 (2015).
77. Francois, J. C. *et al.* Sequence-specific recognition and cleavage of duplex DNA via triple-helix formation by oligonucleotides covalently linked to a phenanthroline-copper chelate. *Proc. Natl. Acad. Sci.* **86**, 9702–9706 (1989).
78. Francois, J. C., Saison-Behmoaras, T., Chassignol, M., Thuong, N. T. & Helene, C. Sequence-targeted cleavage of single- and double-stranded DNA by oligothymidylates covalently linked to 1,10-phenanthroline. *J. Biol. Chem.* **264**, 5891–5898 (1989).
79. Rusling, D. A., Rachwal, P. A., Brown, T. & Fox, K. R. The stability of triplex DNA is affected by the stability of the underlying duplex. *Biophys. Chem.* **145**, 105–110 (2009).
80. Asensio, J. L., Dosanjh, H. S., Jenkins, T. C. & Lane, A. N. Thermodynamic, kinetic, and conformational properties of a parallel intermolecular DNA triplex

- containing 5' and 3' junctions. *Biochemistry* **37**, 15188–15198 (1998).
81. Plum, G. E., Park, Y. W., Singleton, S. F., Dervan, P. B. & Breslauer, K. J. Thermodynamic characterization of the stability and the melting behavior of a DNA triplex: a spectroscopic and calorimetric study. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 9436–9440 (1990).
 82. Faria, M. & Giovannangeli, C. Triplex-forming molecules: From concepts to applications. *J. Gene Med.* **3**, 299–310 (2001).
 83. Pilch, D. S., Brousseau, R. & Shafer, R. H. Thermodynamics of triple helix formation: Spectrophotometric studies on the d(A)₁₀-2d(T)₁₀ and d(C+3T3C+3)·d(G3A4G3)·d(C3T4C3) triple helices. *Nucleic Acids Res.* **18**, 5743–5750 (1990).
 84. Hampel, K. J., Crosson, P. & Lee, J. S. Polyamines Favor DNA Triplex Formation at Neutral pH. *Biochemistry* **30**, 4455–4459 (1991).
 85. Soto, A. M., Loo, J. & Marky, L. A. Energetic contributions for the formation of TAT/TAT, TAT/CGC⁺, and CGC⁺/CGC⁺ base triplet stacks. *J. Am. Chem. Soc.* **124**, 14355–14363 (2002).
 86. Sugimoto, N., Wu, P., Hara, H. & Kawamoto, Y. pH and cation effects on the properties of parallel pyrimidine motif DNA triplexes. *Biochemistry* **40**, 9396–9405 (2001).
 87. Asensio, J. L., Lane, A. N., Dhesi, J., Bergqvist, S. & Brown, T. The contribution of cytosine protonation to the stability of parallel DNA triple helices. *J. Mol. Biol.* **275**, 811–22 (1998).

88. Lee, J. S., Woodsworth, M. L., Latimer, L. J. & Morgan, A. R. Poly(pyrimidine).poly(purine) synthetic DNAs containing 5-methylcytosine form stable triplexes at neutral pH. *Nucleic Acids Res.* **12**, 6603–6614 (1984).
89. Thuong, N. T. & Hélène, C. Sequence-Specific Recognition and Modification of Double-Helical DNA by Oligonucleotides. *Angew. Chemie Int. Ed. English* **32**, 666–690 (1993).
90. Singleton, S. F. & Dervan, P. B. Influence of pH on the Equilibrium Association Constants for Oligodeoxyribonucleotide-Directed Triple Helix Formation at Single DNA Sites. *Biochemistry* **31**, 10995–11003 (1992).
91. Yang, L. *et al.* Influence of 5-bromodeoxycytosine substitution on triplex DNA stability and conformation. *Biophys. Chem.* **76**, 25–34 (1999).
92. Bates, P. J. *et al.* Efficient triple helix formation by oligodeoxyribonucleotides containing α - or β -2-amino-5-(2-deoxy-D-ribofuranosyl) pyridine residues. *Nucleic Acids Res.* **24**, 4176–4184 (1996).
93. Wang, E., Malek, S. & Feigon, J. Structure of a G.T.A triplet in an intramolecular DNA triplex. *Biochemistry* **31**, 4838–4846 (1992).
94. Ono, A., Ts'o, P. O. P. & Kan, L. Triplex Formation of an Oligonucleotide Containing 2'-O-Methylpseudoisocytidine with a DNA Duplex at Neutral pH. *J. Org. Chem.* **57**, 3225–3230 (1992).
95. Ono, A., Ts'o, P. O. P. & Kan, L. S. Triplex formation of oligonucleotides containing 2'-O-methylpseudoisocytidine in substitution for 2'-deoxycytidine. *J. Am. Chem. Soc.* **113**, 4032–4033 (1991).

96. Chin, T. M. *et al.* 'Paper-clip' type triple helix formation by 5'-d-(TC)3T(a)(CT)3C(b)(AG)3 (a and b = 0-4) as a function of loop size with and without the pseudoisocytosine base in the Hoogsteen strand. *Biochemistry* **39**, 12457–12464 (2000).
97. Mayer, A., Häberli, A. & Leumann, C. J. Synthesis and triplex forming properties of pyrrolidino pseudoisocytidine containing oligodeoxynucleotides. *Org. Biomol. Chem.* **3**, 1653–8 (2005).
98. Xiang, G., Bogacki, R. & McLaughlin, L. W. Use of a pyrimidine nucleoside that functions as a bidentate hydrogen bond donor for the recognition of isolated or contiguous G-C base pairs by oligonucleotide-directed triplex formation. *Nucleic Acids Res.* **24**, 1963–1970 (1996).
99. Xiang, G. & McLaughlin, L. W. A cytosine analogue containing a conformationally flexible acyclic linker for triplex formation at sites with contiguous G-C base pairs. *Tetrahedron* **54**, 375–392 (1998).
100. Parsch, U. & Engels, J. W. pH-independent triple-helix formation with 6-oxocytidine as cytidine analogue. *Chemistry* **6**, 2409–24 (2000).
101. Berressem, R. & Engels, J. W. 6-Oxocytidine a novel protonated C-base analogue for stable triple helix formation. *Nucleic Acids Res.* **23**, 3465–3472 (1995).
102. Krawczyk, S. H. *et al.* Oligonucleotide-mediated triple helix formation using an N3-protonated deoxycytidine analog exhibiting pH-independent binding within the physiological range. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 3761–4 (1992).

103. Jetter, M. C. & Hobbs, F. W. 7,8-Dihydro-8-oxoadenine as a Replacement for Cytosine in the Third Strand of Triple Helices. Triplex Formation without Hypochromicity. *Biochemistry* **32**, 3249–3254 (1993).
104. Miller, P. S., Bi, G., Kipp, S. A., Fok, V. & DeLong, R. K. Triplex formation by a psoralen-conjugated oligodeoxyribonucleotide containing the base analog 8-oxo-adenine. *Nucleic Acids Res.* **24**, 730–736 (1996).
105. Miller, P. S., Bhan, P., Cushman, C. D. & Trapane, T. L. Recognition of a Guanine-Cytosine Base Pair by 8-Oxoadenine. *Biochemistry* **31**, 6788–6793 (1992).
106. St Clair, A., Xiang, G. & McLaughlin, L. W. Synthesis and triplex forming properties of an acyclic N7-glycosylated guanine nucleoside. *Nucleosides Nucleotides* **17**, 925–937 (1998).
107. Koshlap, K. M., Schultze, P., Brunar, H., Dervan, P. B. & Feigon, J. Solution structure of an intramolecular DNA triplex containing an N7-glycosylated guanine which mimics a protonated cytosine. *Biochemistry* **36**, 2659–2668 (1997).
108. Hunziker, J., Priestley, E. S., Brunar, H. & Dervan, P. B. Design of an N7-Glycosylated Purine Nucleoside for Recognition of GC Base Pairs by Triple Helix Formation. *J. Am. Chem. Soc.* **117**, 2661–2662 (1995).
109. Seela, F. & Wei, C. Oligonucleotides Containing Consecutive 2'-Deoxyisoguanosine Residues: Synthesis, duplexes with parallel chain orientation, and aggregation. *Helv. Chim. Acta* **80**, 73–85 (1997).

110. Seela, F. & Shaikh, K. I. pH-Independent triplex formation: hairpin DNA containing isoguanine or 9-deaza-9-propynylguanine in place of protonated cytosine. *Org. Biomol. Chem.* **4**, 3993 (2006).
111. Seela, F., Shaikh, K. I. & Wiglenda, T. pH-Independent Recognition of the dG·dC Base Pair in Triplex DNA: 9-Deazaguanine N7-(2'-Deoxyribonucleoside) and Halogenated Derivatives Replacing Protonated dC. *Helv. Chim. Acta* **89**, 598–613 (2006).
112. Seela, F., Jiang, D. & Budow, S. Triplexes with 8-aza-2'-deoxyisoguanosine replacing protonated dC: Probing third strand stability with a fluorescent nucleobase targeting duplex DNA. *ChemBioChem* **11**, 1443–1450 (2010).
113. Yoon, K. *et al.* Elucidation of the sequence-specific third-strand recognition of four Watson-Crick base pairs in a pyrimidine triple-helix motif: T.AT, C.GC, T.CG, and G.TA. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 3840–4 (1992).
114. Radhakrishnan, I. & Patel, D. J. Solution structure of a pyrimidine.purine.pyrimidine DNA triplex containing T.AT, C+.GC and G.TA triples. *Structure* **2**, 17–32 (1994).
115. Fox, K. R. & Brown, T. An extra dimension in nucleic acid sequence recognition. *Q. Rev. Biophys.* **38**, 311–320 (2005).
116. Hari, Y., Obika, S. & Imanishi, T. Towards the Sequence-Selective Recognition of Double-Stranded DNA Containing Pyrimidine-Purine Interruptions by Triplex-Forming Oligonucleotides. *Eur. J. Org. Chem.* **2012**, 2875–2887 (2012).
117. Fox, K. R., Brown, T. & Rusling, D. A. DNA Recognition by Parallel Triplex

- Formation. in *DNA-targeting Molecules as Therapeutic Agents* 1–32 (The Royal Society of Chemistry, 2018).
118. Malnuit, V., Duca, M. & Benhida, R. Targeting DNA base pair mismatch with artificial nucleobases. Advances and perspectives in triple helix strategy. *Org. Biomol. Chem.* **9**, 326–336 (2011).
119. Dong Li Chen & McLaughlin, L. W. Use of pK(a) differences to enhance the formation of base triplets involving C-G and G-C base pairs. *J. Org. Chem.* **65**, 7468–7474 (2000).
120. Gerrard, S. R., Edrees, M. M., Bouamaied, I., Fox, K. R. & Brown, T. CG base pair recognition within DNA triple helices by modified N-methylpyrrolo-dC nucleosides. *Org. Biomol. Chem.* **8**, 5087–5096 (2010).
121. Johannsen, M. W., Gerrard, S. R., Melvin, T. & Brown, T. Triplex-mediated analysis of cytosine methylation at CpA sites in DNA. *Chem. Commun.* **50**, 551–553 (2014).
122. Prévot-Halter, I. & Leumann, C. J. Selective recognition of a C-G base-pair in the parallel DNA triple-helical binding motif. *Bioorganic Med. Chem. Lett.* **9**, 2657–2660 (1999).
123. Guianvarc'h, D., Benhida, R., Fourrey, J. L., Maurisse, R. & Sun, J. S. Incorporation of a novel nucleobase allows stable oligonucleotide-directed triple helix formation at the target sequence containing a purine-pyrimidine interruption. *Chem. Commun.* **1**, 1814–1815 (2001).
124. Wang, Y. *et al.* Stable recognition of TA interruptions by triplex forming

- oligonucleotides containing a novel nucleoside. *Biochemistry* **44**, 5884–5892 (2005).
125. Okamoto, I., Seio, K. & Sekine, M. Triplex forming ability of oligonucleotides containing 2'-O-methyl-2-thiouridine or 2-thiothymidine. *Bioorganic Med. Chem. Lett.* **16**, 3334–3336 (2006).
126. Fox, K. R. Targeting DNA with Triplexes. *Curr. Med. Chem.* **7**, 17–37 (2000).
127. Bijapur, J., Keppler, M. D., Bergqvist, S., Brown, T. & Fox, K. R. 5-(1-propargylamino)-2'-deoxyuridine (U(P)): A novel thymidine analogue for generating DNA triplexes with increased stability. *Nucleic Acids Res.* **27**, 1802–1809 (1999).
128. Sollogoub, M., Dominguez, B., Brown, T. & Fox, K. R. Synthesis of a novel bis-amino-modified thymidine monomer for use in DNA triplex stabilisation. *Chem. Commun.* **3**, 2315–2316 (2000).
129. Shimizu, M., Konishi, A., Shimada, Y., Inoue, H. & Ohtsuka, E. Oligo (2'-O-methyl) ribonucleotides Effective probes for duplex DNA. *FEBS Lett.* **302**, 155–158 (1992).
130. Torigoe, H., Hari, Y., Sekiguchi, M., Obika, S. & Imanishi, T. 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 (2001).
131. Sun, B.-W. *et al.* Sequence and pH Effects of LNA-Containing Triple Helix-Forming Oligonucleotides: Physical Chemistry, Biochemistry, and Modeling

- Studies. *Biochemistry* **43**, 4160–4169 (2004).
132. Lacroix, L., Arimondo, P. B., Takasugi, M., Hélène, C. & Mergny, J. L. Pyrimidine morpholino oligonucleotides form a stable triple helix in the absence of magnesium ions. *Biochem. Biophys. Res. Commun.* **270**, 363–369 (2000).
133. Hansen, M. E., Bentin, T. & Nielsen, P. E. High-affinity triplex targeting of double stranded DNA using chemically modified peptide nucleic acid oligomers. *Nucleic Acids Res.* **37**, 4498–4507 (2009).
134. Sarhan, S. & Seiler, N. On the Subcellular Localization of the Polyamines. *Biol. Chem. Hoppe. Seyler.* **370**, 1279–1284 (1989).
135. Rougée, M. *et al.* Kinetics and thermodynamics of triple-helix formation: effects of ionic strength and mismatches. *Biochemistry* **31**, 9269–78 (1992).
136. Li, H. *et al.* Triplex staples: DNA double-strand cross-linking at internal and terminal sites using psoralen-containing triplex-forming oligonucleotides. *Bioconjug. Chem.* **17**, 1561–1567 (2006).
137. Cassidy, R. A., Kondo, N. S. & Miller, P. S. Triplex formation by psoralen-conjugated chimeric oligonucleoside methylphosphonates. *Biochemistry* **39**, 8683–8691 (2000).
138. Vasquez, K. M., Wensel, T. G., Hogan, M. E. & Wilson, J. H. High-efficiency triple-helix-mediated photo-cross-linking at a targeted site within a selectable mammalian gene. *Biochemistry* **35**, 10712–10719 (1996).
139. Silver, G. C. *et al.* Conjugates of oligonucleotides with triplex-specific intercalating agents. Stabilization of triple-helical DNA in the promoter region

- of the gene for the α -subunit of interleukin 2 (IL-2R α). *Bioconjug. Chem.* **8**, 15–22 (1997).
140. Pilch, D. S. *et al.* Characterization of a Triple Helix-specific Ligand. *J. Mol. Biol.* **232**, 926–946 (1993).
141. Keppler, M. D., James, P. L., Neidle, S., Brown, T. & Fox, K. R. DNA sequence specificity of triplex-binding ligands. *Eur. J. Biochem.* **270**, 4982–4992 (2003).
142. Ejlersen, M., Langkjær, N. & Wengel, J. 3'-Pyrene-modified unlocked nucleic acids: synthesis, fluorescence properties and a surprising stabilization effect on duplexes and triplexes. *Org. Biomol. Chem.* **15**, 2073–2085 (2017).
143. Doluca, O., Hale, T. K., Edwards, P. J. B., Gonzalez, C. & Filichev, V. V. Assembly dependent fluorescence enhancing nucleic acids in sequence-specific detection of double-stranded DNA. *Chempluschem* **79**, 58–66 (2014).
144. Filichev, V. V. & Pedersen, E. B. 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 (2005).
145. Streckowski, L. *et al.* New triple-helix DNA stabilizing agents. *Bioorganic Med. Chem. Lett.* **15**, 1097–1100 (2005).
146. Gianolio, D. a, Segismundo, J. M. & McLaughlin, L. W. Tethered naphthalene diimide-based intercalators for DNA triplex stabilization. *Nucleic Acids Res.* **28**, 2128–2134 (2000).
147. Fox, K. R. & Brown, T. Small Molecule – Oligonucleotide Conjugates. in *DNA*

- Conjugates and Sensors* 75–102 (2012).
148. Grossmann, T. N., Röglin, L. & Seitz, O. Triplex molecular beacons as modular probes for DNA detection. *Angew. Chem., Int. Ed.* **46**, 5223–5225 (2007).
 149. Darby, R. J. *et al.* High throughput measurement of duplex, triplex and quadruplex melting curves using molecular beacons and a LightCycler. *Nucleic Acids Res.* **30**, e39 (2002).
 150. Wang, Y. *et al.* Triplex molecular beacons for sensitive recognition of melamine based on abasic-site-containing DNA and fluorescent silver nanoclusters. *Chem. Commun.* **51**, 7958–7961 (2015).
 151. Silva, G. L., Ediz, V., Yaron, D. & Armitage, B. A. Experimental and computational investigation of unsymmetrical cyanine dyes: Understanding torsionally responsive fluorogenic dyes. *J. Am. Chem. Soc.* **129**, 5710–5718 (2007).
 152. Privat, E. *et al.* Fluorescent properties of oligonucleotide-conjugated thiazole orange probes. *Photochem. Photobiol.* **75**, 201–210 (2002).
 153. Netzel, T. L., Nafisi, K., Zhao, M., Lenhard, J. R. & Johnson, I. Base-content dependence of emission enhancements, quantum yields, and lifetimes for cyanine dyes bound to double-strand DNA: Photophysical properties of monomeric and bichromophoric DNA stains. *J. Phys. Chem.* **99**, 17936–17947 (1995).
 154. Isacsson, J. & Westman, G. Solid-phase synthesis of asymmetric cyanine dyes. *Tetrahedron Lett.* **42**, 3207–3210 (2001).

155. Nygren, J., Svanvik, N. & Kubista, M. The Interactions Between the Fluorescent Dye Thiazole Orange and DNA. *Biopolymers* **46**, 39–51 (1998).
156. Lartia, R. & Asseline, U. New cyanine-oligonucleotide conjugates: Relationships between chemical structures and properties. *Chem. - A Eur. J.* **12**, 2270–2281 (2006).
157. Rye, H. S., Quesada, M. A., Peck, K., Mathies, R. A. & Giazar, A. N. High-sensitivity two-color detection of double-stranded DNA with a confocal fluorescence gel scanner using ethidium homodimer and thiazole orange. *Nucleic Acids Res.* **19**, 327–333 (1991).
158. Rye, H. S. *et al.* Picogram detection of stable dye-DNA intercalation complexes with two-color laser-excited confocal fluorescence gel scanner. in *Methods in Enzymology* **217**, 414–431 (1993).
159. Ghosh, I., Stains, C. I., Ooi, A. T. & Segal, D. J. Direct detection of double-stranded DNA: molecular methods and applications for DNA diagnostics. *Mol. Biosyst.* **2**, 551–560 (2006).
160. Pei, R., Rothman, J., Xie, Y. & Stojanovic, M. N. Light-up properties of complexes between thiazole orange-small molecule conjugates and aptamers. *Nucleic Acids Res.* **37**, 2–7 (2009).
161. Lubitz, I., Zikich, D. & Kotlyar, A. Specific high-affinity binding of thiazole orange to triplex and g-quadruplex DNA. *Biochemistry* **49**, 3567–3574 (2010).
162. Xu, L. *et al.* Thiazole Orange as a Fluorescent Light-Up Probe for the i-Motif and its Application to the Development of a Molecular Logic System. *Asian J. Org.*

- Chem.* **4**, 1375–1378 (2015).
163. Glazer, A. N. & Rye, H. S. Stable dye–DNA intercalation complexes as reagents for high-sensitivity fluorescence detection. *Nature* **359**, 859–861 (1992).
164. Bunkenborg, J., Stidsen, M. M. & Jacobsen, J. P. On the sequence selective bis-intercalation of a homodimeric thiazole orange dye in DNA. *Bioconjug. Chem.* **10**, 824–831 (1999).
165. Petersen, M. & Jacobsen, J. P. Solution Structure of a DNA Complex with the Fluorescent Bis-Intercalator TOTO Modified on the Benzothiazole Ring. *Bioconjug. Chem.* **9**, 331–340 (1998).
166. Li, K. & Liu, B. Conjugated Polyelectrolyte Amplified Thiazole Orange Emission for Label Free Sequence Specific DNA Detection with Single Nucleotide Polymorphism Selectivity. *Polymer (Guildf)*. **81**, 4099–4105 (2009).
167. Gorab, E. & Pearson, P. L. Thiazole Orange as an Alternative to Antibody Binding for Detecting Triple-helical DNA in Heterochromatin of *Drosophila* and *Rhynchosciara*. *J. Histochem. Cytochem.* **66**, 143–154 (2018).
168. Svanvik, N., Westman, G., Wang, D. & Kubista, M. Light-up probes: thiazole orange-conjugated peptide nucleic acid for detection of target nucleic acid in homogeneous solution. *Anal. Biochem.* **281**, 26–35 (2000).
169. Sato, T., Sato, Y. & Nishizawa, S. Triplex-Forming Peptide Nucleic Acid Probe Having Thiazole Orange as a Base Surrogate for Fluorescence Sensing of Double-stranded RNA. *J. Am. Chem. Soc.* **138**, 9397–9400 (2016).
170. Sato, T., Sato, Y. & Nishizawa, S. Optimization of the Alkyl Linker of TO Base

- Surrogate in Triplex-Forming PNA for Enhanced Binding to Double-Stranded RNA. *Chem. - A Eur. J.* **23**, 4079–4088 (2017).
171. Renard, B.-L., Lartia, R. & Asseline, U. Targeting DNA with “light-up” pyrimidine triple-helical forming oligonucleotides conjugated to stabilizing fluorophores (LU-TFOs). *Org. Biomol. Chem.* **6**, 4413–4425 (2008).
 172. Ikeda, S., Yanagisawa, H., Yuki, M. & Okamoto, A. Fluorescent triplex-forming DNA oligonucleotides labeled with a thiazole orange dimer unit. *Artif. DNA PNA XNA* **4**, 19–27 (2013).
 173. Walsh, S., El-Sagheer, A. H. & Brown, T. Fluorogenic thiazole orange TOTFO probes stabilise parallel DNA triplexes at pH 7 and above. *Chem. Sci.* **9**, 7681–7687 (2018).
 174. Froehler, B. C., Wadwani, S., Terhorst, T. J. & Gerrard, S. R. Oligodeoxynucleotides containing C-5 propyne analogs of 2-deoxyuridine and 2-deoxycytidine. *Tetrahedron Lett.* **33**, 5307–5310 (1992).
 175. Bates, P. J. *et al.* Detection and kinetic studies of triplex formation by oligodeoxynucleotides using real-time biomolecular interaction analysis (BIA). *Nucleic Acids Res.* **23**, 3627–3632 (1995).
 176. Schroeder, S. J. & Turner, D. H. Optical Melting Measurements of Nucleic Acid Thermodynamics. in *Methods in Enzymology* 371–387 (2009).
 177. Misteli, T. Beyond the Sequence: Cellular Organization of Genome Function. *Cell* **128**, 787–800 (2007).
 178. Misteli, T. Higher-order genome organization in human disease. *Cold Spring*

- Harb. Perspect. Biol.* **2**, a000794 (2010).
179. Roukos, V. *et al.* Spatial dynamics of chromosome translocations in living cells. *Science* **341**, 660–664 (2013).
 180. Callan, H. G. Review Lecture: Replication of DNA in the Chromosomes of Eukaryotes. *Proc. R. Soc. London. Ser. B, Biol. Sci.* **181**, 19–41 (1972).
 181. Bick, M. D. & Davidson, R. L. Total substitution of bromodeoxyuridine for thymidine in the DNA of a bromodeoxyuridine dependent cell line. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 2082–2086 (1974).
 182. Dekker, J. *et al.* The 4D nucleome project. *Nature* **549**, 219–226 (2017).
 183. Viollier, P. H. *et al.* Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9257–9262 (2004).
 184. Saad, H. *et al.* DNA Dynamics during Early Double-Strand Break Processing Revealed by Non-Intrusive Imaging of Living Cells. *PLoS Genet.* **10**, (2014).
 185. Salic, A. & Mitchison, T. J. A chemical method for fast and sensitive detection of DNA synthesis in vivo. *Proc. Natl. Acad. Sci.* **105**, 2415–2420 (2008).
 186. Speicher, M. R., Ballard, S. G. & Ward, D. C. Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nat. Genet.* **12**, 368–375 (1996).
 187. Langer-Safer, P. R., Levine, M. & Ward, D. C. Immunological methods for mapping genes on *Drosophila* polytene chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 4381–4385 (1982).
 188. Beliveau, B. J. *et al.* Single-molecule super-resolution imaging of chromosomes

- and in situ haplotype visualization using Oligopaint FISH probes. *Nat. Commun.* **6**, 7147 (2015).
189. Beliveau, B. J. *et al.* Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 21301–21306 (2012).
 190. Deng, W., Shi, X., Tjian, R., Lionnet, T. & Singer, R. H. CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11870–11875 (2015).
 191. Shao, S., Chang, L., Hou, Y. & Sun, Y. Illuminating the structure and dynamics of chromatin by fluorescence labeling. *Front. Biol. (Beijing)*. **12**, 241–257 (2017).
 192. Matzke, A. J. M., Huettel, B., Van Der Winden, J. & Matzke, M. Use of two-color fluorescence-tagged transgenes to study interphase chromosomes in living plants. *Plant Physiol.* **139**, 1586–1596 (2005).
 193. Robinett, C. C. *et al.* In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J. Cell Biol.* **135**, 1685–1700 (1996).
 194. Belmont, A. S. Visualizing chromosome dynamics with GFP. *Trends Cell Biol.* **11**, 250–257 (2001).
 195. Kanda, T., Sullivan, K. F. & Wahl, G. M. Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. *Curr. Biol.* **8**, 377–385 (1998).
 196. Lindhout, B. I. *et al.* Live cell imaging of repetitive DNA sequences via GFP-

- tagged polydactyl zinc finger proteins. *Nucleic Acids Res.* **35**, (2007).
197. Segal, D. J., Dreier, B., Beerli, R. R. & Barbas, C. F. Toward controlling gene expression at will: Selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 2758–2763 (1999).
 198. Miyanari, Y., Ziegler-Birling, C. & Torres-Padilla, M. E. Live visualization of chromatin dynamics with fluorescent TALEs. *Nat. Struct. Mol. Biol.* **20**, 1321–1324 (2013).
 199. Ma, H., Reyes-Gutierrez, P. & Pederson, T. Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. *Proc. Natl. Acad. Sci.* **110**, 21048–21053 (2013).
 200. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
 201. Jiang, F. & Doudna, J. A. CRISPR–Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* **46**, 505–529 (2017).
 202. Knight, S. C., Tjian, R. & Doudna, J. A. Genomes in Focus: Development and Applications of CRISPR-Cas9 Imaging Technologies. *Angew. Chemie - Int. Ed.* **57**, 4329–4337 (2018).
 203. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakamura, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isoenzyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433 (1987).

204. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–823 (2013).
205. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2013**, 1–9 (2013).
206. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
207. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).
208. Mali, P. *et al.* Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 (2013).
209. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–23 (2013).
210. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, (2014).
211. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
212. Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P. & Lu, T. K. Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells. *Mol. Cell* **54**, 698–710 (2014).
213. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).

214. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481 (2015).
215. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, (2014).
216. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).
217. Yang, W. An equivalent metal ion in one- and two-metal-ion catalysis. *Nat. Struct. Mol. Biol.* **15**, 1228–1231 (2008).
218. Jiang, F. *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867–871 (2016).
219. Raper, A. T., Stephenson, A. A. & Suo, Z. Functional Insights Revealed by the Kinetic Mechanism of CRISPR-Cas9. *J. Am. Chem. Soc.* **140**, 2971–2984 (2018).
220. Bassing, C. H. & Alt, F. W. The cellular response to general and programmed DNA double strand breaks. *DNA Repair (Amst)*. **3**, 781–796 (2004).
221. Gilbert, L. A. *et al.* CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154**, 442–451 (2013).
222. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).
223. Chen, B. *et al.* Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* **155**, 1479–1491 (2013).

224. Ma, H. *et al.* Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3002–3007 (2015).
225. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646 (2014).
226. Hong, Y., Lu, G., Duan, J., Liu, W. & Zhang, Y. Comparison and optimization of CRISPR/dCas9/gRNA genome-labeling systems for live cell imaging. *Genome Biol.* **19**, 1–10 (2018).
227. Ma, Y. *et al.* Live Visualization of HIV-1 Proviral DNA Using a Dual-Color-Labeled CRISPR System. *Anal. Chem.* **89**, 12896–12901 (2017).
228. Delehanty, J. B. *et al.* Delivering quantum dot-peptide bioconjugates to the cellular cytosol: Escaping from the endolysosomal system. *Integr. Biol.* **2**, 265–277 (2010).
229. Shechner, D. M., Hacisuleyman, E., Younger, S. T. & Rinn, J. L. Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat. Methods* **12**, 664–670 (2015).
230. Ma, H. *et al.* Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nat. Biotechnol.* **34**, 528–530 (2016).
231. Filonov, G. S., Moon, J. D., Svensen, N. & Jaffrey, S. R. Broccoli: Rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J. Am. Chem. Soc.* **136**, 16299–16308 (2014).
232. Ma, H. *et al.* CRISPR-Cas9 nuclear dynamics and target recognition in living

- cells. *J. Cell Biol.* **214**, 529–537 (2016).
233. Ma, H. *et al.* CRISPR-Sirius: RNA scaffolds for signal amplification in genome imaging. *Nat. Methods* **15**, 928–931 (2018).
 234. Wu, B. *et al.* Synonymous modification results in high-fidelity gene expression of repetitive protein and nucleotide sequences. *Genes Dev.* **29**, 876–886 (2015).
 235. Shao, S. *et al.* Long-term dual-color tracking of genomic loci by modified sgRNAs of the CRISPR/Cas9 system. *Nucleic Acids Res.* **44**, (2016).
 236. Qin, P. *et al.* Live cell imaging of low- and non-repetitive chromosome loci using CRISPR-Cas9. *Nat. Commun.* **8**, 1–10 (2017).
 237. Wu, X., Mao, S., Ying, Y., Krueger, C. J. & Chen, A. K. Progress and Challenges for Live-cell Imaging of Genomic Loci Using CRISPR-based Platforms. *Genomics. Proteomics Bioinformatics* **17**, 119–128 (2018).
 238. Wu, X. *et al.* A CRISPR/molecular beacon hybrid system for live-cell genomic imaging. *Nucleic Acids Res.* **46**, e80 (2018).
 239. Mao, S., Ying, Y., Wu, X., Krueger, C. J. & Chen, A. K. CRISPR/dual-FRET molecular beacon for sensitive live-cell imaging of non-repetitive genomic loci. *Nucleic Acids Res.* **9**, 1–8 (2019).
 240. Wang, H. *et al.* CRISPR-mediated live imaging of genome editing and transcription. *Science* **365**, 1301–1305 (2019).
 241. Nollet, S. *et al.* Human mucin gene MUC4: Organization of its 5'-region and polymorphism of its central tandem repeat array. *Biochem. J.* **332**, 739–748 (1998).

242. Ferry, Q. R. V., Lyutova, R. & Fulga, T. A. Rational design of inducible CRISPR guide RNAs for de novo assembly of transcriptional programs. *Nat. Commun.* **8**, 1–10 (2017).
243. Hall, L. M., Gerowska, M. & Brown, T. A highly fluorescent DNA toolkit: Synthesis and properties of oligonucleotides containing new Cy3, Cy5 and Cy3B monomers. *Nucleic Acids Res.* **40**, 1–10 (2012).
244. Gerowska, M., Hall, L., Richardson, J., Shelbourne, M. & Brown, T. Efficient reverse click labeling of azide oligonucleotides with multiple alkynyl Cy-Dyes applied to the synthesis of HyBeacon probes for genetic analysis. *Tetrahedron* **68**, 857–864 (2012).
245. Mir, A. *et al.* Heavily and fully modified RNAs guide efficient SpyCas9-mediated genome editing. *Nat. Commun.* **9**, 1–9 (2018).
246. Yin, H. *et al.* Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing. *Nat. Biotechnol.* **35**, 1179–1187 (2017).
247. Rahdar, M. *et al.* Synthetic CRISPR RNA-Cas9-guided genome editing in human cells. *Proc. Natl. Acad. Sci.* **112**, E7110–E7117 (2015).
248. Rueda, F. O. *et al.* Mapping the sugar dependency for rational generation of a DNA-RNA hybrid-guided Cas9 endonuclease. *Nat. Commun.* **8**, 1–11 (2017).
249. Yin, H. *et al.* Partial DNA-guided Cas9 enables genome editing with reduced off-target activity. *Nat. Chem. Biol.* **14**, 311–316 (2018).
250. O'Reilly, D. *et al.* Extensive CRISPR RNA modification reveals chemical compatibility and structure-activity relationships for Cas9 biochemical activity.

- Nucleic Acids Res.* **47**, 546–558 (2019).
251. Lee, K. *et al.* Synthetically modified guide RNA and donor DNA are a versatile platform for CRISPR-Cas9 engineering. *Elife* **6**, 1–17 (2017).
 252. Stein, H. & Hausen, P. Enzyme from Calf Thymus Degrading the RNA Moiety of DNA-RNA Hybrids: Effect on DNA-Dependent RNA Polymerase. *Science* **166**, 393–395 (1969).
 253. Nowotny, M. *et al.* Structure of Human RNase H1 Complexed with an RNA/DNA Hybrid: Insight into HIV Reverse Transcription. *Mol. Cell* **28**, 264–276 (2007).
 254. Shen, W. *et al.* Dynamic nucleoplasmic and nucleolar localization of mammalian RNase H1 in response to RNAP I transcriptional R-loops. *Nucleic Acids Res.* **45**, 10672–10692 (2017).
 255. Liang, X. H., Sun, H., Nichols, J. G. & Crooke, S. T. RNase H1-Dependent Antisense Oligonucleotides Are Robustly Active in Directing RNA Cleavage in Both the Cytoplasm and the Nucleus. *Mol. Ther.* **25**, 2075–2092 (2017).
 256. Kiełpiński, Ł. J., Hagedorn, P. H., Lindow, M. & Vinther, J. RNase H sequence preferences influence antisense oligonucleotide efficiency. *Nucleic Acids Res.* **45**, 12932–12944 (2017).
 257. Donis-Keller, H. Site specific enzymatic cleavage of RNA. *Nucleic Acids Res.* **7**, 179–192 (1979).
 258. Lima, W. F. & Crooke, S. T. Cleavage of Single Strand RNA Adjacent to RNA-DNA Duplex Regions by. *Mol. Biol.* **272**, 1–4 (1997).
 259. Hogrefe, H. H., Hogrefe, R. I., Walder, R. Y. & Walder, J. A. Kinetic analysis of

- Escherichia coli RNase H using DNA-RNA-DNA/DNA substrates. *J. Biol. Chem.* **265**, 5561–5566 (1990).
260. Pallan, P. S. & Egli, M. Insights into RNA/DNA hybrid recognition and processing by RNase H from the crystal structure of a non-specific enzyme-dsDNA complex. *Cell Cycle* **7**, 2562–2569 (2008).
 261. Venkateswarlu, D., Lind, K. E., Mohan, V., Manoharan, M. & Ferguson, D. M. Structural properties of DNA:RNA duplexes containing 2'-O-methyl and 2'-S-methyl substitutions: A molecular dynamics investigation. *Nucleic Acids Res.* **27**, 2189–2195 (1999).
 262. Cerritelli, S. M. & Crouch, R. J. Ribonuclease H: The enzymes in eukaryotes. *FEBS J.* **276**, 1494–1505 (2009).
 263. Chesnoy, S. & Huang, L. Structure and Function of Lipid-DNA Complexes for Gene Delivery. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 27–47 (2000).
 264. Cardarelli, F. *et al.* The intracellular trafficking mechanism of Lipofectamine-based transfection reagents and its implication for gene delivery. *Sci. Rep.* **6**, 1–8 (2016).
 265. Xu, Y. & Szoka, F. C. Mechanism of DNA release from cationic liposome/DNA complexes used in cell transfection. *Biochemistry* **35**, 5616–5623 (1996).
 266. Zelphati, O. & Szoka, F. C. Mechanism of oligonucleotide release from cationic liposomes. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 11493–11498 (1996).
 267. Hoekstra, D., Rejman, J., Wasungu, L., Shi, F. & Zuhorn, I. Gene delivery by cationic lipids: In and out of an endosome. *Biochem. Soc. Trans.* **35**, 68–71

- (2007).
268. Janssen, A. & Medema, R. H. Genetic instability: Tipping the balance. *Oncogene* **32**, 4459–4470 (2013).
269. Caruthers, M. H. Chemical Synthesis of DNA and DNA Analogues. *Acc. Chem. Res.* **24**, 278–284 (1991).
270. Pitsch, S., Weiss, P. A., Jenny, L., Stutz, A. & Wu, X. Reliable chemical synthesis of oligoribonucleotides (RNA) with 2'-O-[(triisopropylsilyl)oxy]methyl(2'-O-tom)-protected phosphoramidites. *Helv. Chim. Acta* **84**, 3773–3795 (2001).
271. Beckert, B. & Masquida, B. Synthesis of RNA by In Vitro Transcription. in *RNA Methods and Protocols* 29–41 (Humana Press, 2011).
272. Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**, 8783–8798 (1987).
273. Milisavljevič, N., Perlíková, P., Pohl, R. & Hocek, M. Enzymatic synthesis of base-modified RNA by T7 RNA polymerase. A systematic study and comparison of 5-substituted pyrimidine and 7-substituted 7-deazapurine nucleoside triphosphates as substrates. *Org. Biomol. Chem.* **16**, 5800–5807 (2018).
274. Kumar, R. *et al.* Template-Directed Oligonucleotide Strand Ligation, Covalent Intramolecular DNA Circularization and Catenation Using Click Chemistry. *J. Am. Chem. Soc.* **129**, 6859–6864 (2007).
275. El-Sagheer, A. H. & Brown, T. New strategy for the synthesis of chemically modified RNA constructs exemplified by hairpin and hammerhead ribozymes.

- Proc. Natl. Acad. Sci. U. S. A.* **107**, 15329–15334 (2010).
276. Huisgen, R. 1,3-Dipolar Cycloadditions. Past and Future. *Angew. Chemie Int. Ed. English* **2**, 565–598 (1963).
277. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chemie - Int. Ed.* **40**, 2004–2021 (2001).
278. Isobe, H., Fujino, T., Yamazaki, N., Guillot-Nieckowski, M. & Nakamura, E. Triazole-linked analogue of deoxyribonucleic acid (TLDNA): Design, synthesis, and double-strand formation with natural DNA. *Org. Lett.* **10**, 3729–3732 (2008).
279. El-Sagheer, A. H. & Brown, T. Synthesis and polymerase chain reaction amplification of DNA strands containing an unnatural triazole linkage. *J. Am. Chem. Soc.* **131**, 3958–3964 (2009).
280. El-Sagheer, A. H. & Brown, T. Click nucleic acid ligation: Applications in biology and nanotechnology. *Acc. Chem. Res.* **45**, 1258–1267 (2012).
281. Isobe, H. & Fujino, T. Triazole-linked analogues of DNA and RNA (TLDNA and TLRNA): Synthesis and functions. *Chem. Rec.* **14**, 41–51 (2014).
282. Shivalingam, A., Tyburn, A. E. S., El-Sagheer, A. H. & Brown, T. Molecular requirements of high-fidelity replication-competent DNA backbones for orthogonal chemical ligation. *J. Am. Chem. Soc.* **139**, 1575–1583 (2017).
283. Sharma, V. K. *et al.* Synthesis and biological properties of triazole-linked locked nucleic acid. *Chem. Commun.* **53**, 8906–8909 (2017).

284. El-Sagheer, A. H., Sanzone, A. P., Gao, R., Tavassoli, A. & Brown, T. Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11338–11343 (2011).
285. Birts, C. N. *et al.* Transcription of click-linked DNA in human cells. *Angew. Chemie - Int. Ed.* **53**, 2362–2365 (2014).
286. El-Sagheer, A. H. & Brown, T. Efficient RNA synthesis by in vitro transcription of a triazole-modified DNA template. *Chem. Commun.* **47**, 12057–12058 (2011).
287. Dallmann, A. *et al.* Structure and dynamics of triazole-linked DNA: Biocompatibility explained. *Chem. - A Eur. J.* **17**, 14714–14717 (2011).
288. Cost, G. J. Enzymatic ligation assisted by nucleases: Simultaneous ligation and digestion promote the ordered assembly of DNA. *Nat. Protoc.* **2**, 2198–2202 (2007).
289. Sokolova, N. I., Ashirbekova, D. T., Dolinnaya, N. G. & Shabarova, Z. A. Chemical reactions within DNA duplexes Cyanogen bromide as an effective oligodeoxyribonucleotide coupling agent. *FEBS Lett.* **232**, 153–155 (1988).
290. Taemaitree, L., Shivalingam, A., El-Sagheer, A. H. & Brown, T. An artificial triazole backbone linkage provides a split-and-click strategy to bioactive chemically modified CRISPR sgRNA. *Nat. Commun.* **10**, 1610 (2019).
291. Rao, H., Tanpure, A. A., Sawant, A. A. & Srivatsan, S. G. Enzymatic incorporation of an azide-modified UTP analog into oligoribonucleotides for post-transcriptional chemical functionalization. *Nat. Protoc.* **7**, 1097–1112

- (2012).
292. Santner, T., Hartl, M., Bister, K. & Micura, R. Efficient Access to 3'-Terminal Azide-Modified RNA for Inverse Click-Labeling Patterns. *Bioconjug. Chem.* **25**, 188–195 (2014).
293. Said, H., Guttroff, C., El-Sagheer, A. H. & Richert, C. Two-step synthesis of a 5-azidothymidine building block for the assembly of oligonucleotides for triazole-forming ligations. *Synlett* **23**, 2923–2926 (2012).
294. Uzagare, M. C., Claußnitzer, I., Gerrits, M. & Bannwarth, W. A Novel Method for the Labelling of Peptides and Proteins through a Bioorthogonal Staudinger Reaction by Using 2-Cyanoethyl Phosphoramidites. *ChemBioChem* **13**, 2204–2208 (2012).
295. Wada, T. *et al.* Synthesis and properties of 2-azidodeoxyadenosine and its incorporation into oligodeoxynucleotides. *Tetrahedron Lett.* **42**, 9215–9219 (2001).
296. Jankowska, J., Sobkowski, M., Stawiński, J. & Kraszewski, A. Studies on aryl H-phosphonates. I. An efficient method for the preparation of deoxyribo- and ribonucleoside 3'-H-phosphonate monoesters by transesterification of diphenyl H-phosphonate. *Tetrahedron Lett.* **35**, 3355–3358 (1994).
297. Ötvös, L., Bajor, Z., Kraicsovits, F., Sági, G. & Tegye, Z. Synthesis and enzymatic characterization of P1-thio-P2-oxo trideoxynucleoside diphosphates having AZT, FdU, or dT at the 3'-position. *Nucleosides, Nucleotides and Nucleic Acids* **21**, 79–92 (2002).

298. Efimov, V. A. *et al.* Application of new catalytic phosphate protecting groups for the highly efficient phosphotriester oligonucleotide synthesis. *Nucleic Acids Res.* **14**, 6525–6540 (1986).
299. Polushin, N. N., Smirnov, I. P., Verentchikov, A. N. & Coull, J. M. Synthesis of oligonucleotides containing 2'-azido-and 2'-amino-2'-deoxyuridine using phosphotriester chemistry. *Tetrahedron Lett.* **37**, 3227–3230 (1996).
300. Efimov, V. A., Molchanova, N. S. & Chakhmakhcheva, O. G. Approach to the synthesis of natural and modified oligonucleotides by the phosphotriester method using o-nucleophilic intramolecular catalysis. *Nucleosides, Nucleotides and Nucleic Acids* **26**, 1087–1093 (2007).
301. Efimov, V. A., Aralov, A. V., Klykov, V. N. & Chakhmakhcheva, O. G. Synthesis of RNA by the rapid phosphotriester method using azido-based 2'-O-protecting groups. *Nucleosides, Nucleotides and Nucleic Acids* **28**, 846–865 (2009).
302. Aigner, M. *et al.* Chemical synthesis of site-specifically 2'-azido-modified RNA and potential applications for bioconjugation and RNA interference. *ChemBioChem* **12**, 47–51 (2011).
303. Wolter, A., Studte, S. & Leuck, M. TC RNA Phosphoramidites Properties and Application. Available at: <https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/SAFC/Posters/1/tc-rna-phosphoramidites.pdf>. (Accessed: 3rd November 2019)
304. Letsinger, R. L. & Lunsford, W. B. Synthesis of thymidine oligonucleotides by phosphite triester intermediates. *J. Am. Chem. Soc.* **98**, 3655–3661 (1976).

305. Rajur, S. B., Robles, J., Wiederholt, K., Kuimelis, R. G. & McLaughlin, L. W. Hoechst 33258 Tethered by a Hexa(ethylene glycol) Linker to the 5'-Termini of Oligodeoxynucleotide 15-Mers: Duplex Stabilization and Fluorescence Properties. *J. Org. Chem.* **62**, 523–529 (1997).
306. Stewart, K. M., Rojo, J. & McLaughlin, L. W. Ru(II) Tris(bipyridyl) Complexes with Six Oligonucleotide Arms as Precursors for the Generation of Supramolecular Assemblies. *Angew. Chemie* **116**, 5932–5935 (2004).
307. Stewart, K. M. & McLaughlin, L. W. Four-Arm Oligonucleotide Ni(II) - Cyclam-Centered Complexes as Precursors for the Generation of Supramolecular Periodic Assemblies. *J. Am. Chem. Soc.* **126**, 2050–2057 (2004).
308. Meng, M. *et al.* Two base pair duplexes suffice to build a novel material. *ChemBioChem* **10**, 1335–1339 (2009).
309. Plietzsch, O. *et al.* Four-fold click reactions: Generation of tetrahedral methane- and adamantane-based building blocks for higher-order molecular assemblies. *Org. Biomol. Chem.* **7**, 4734–4743 (2009).
310. Fomich, M. A. *et al.* Azide phosphoramidite in direct synthesis of azide-modified oligonucleotides. *Org. Lett.* **16**, 4590–4593 (2014).
311. Lingala, S., Nordstrøm, L. U. & Mallikaratchy, P. R. Synthesis of stable azide and alkyne functionalized phosphoramidite nucleosides. *Tetrahedron Lett.* **60**, 211–213 (2019).
312. Fulmer, G. R. *et al.* NMR chemical shifts of trace impurities: Common laboratory solvents, organics, and gases in deuterated solvents relevant to the

- organometallic chemist. *Organometallics* **29**, 2176–2179 (2010).
313. Booth, J. Propargylamino Modified Bases for DNA Stabilisation. (University of Southampton, 2004).
314. Cruickshank, K. A. & Stockwell, D. L. Oligonucleotide labelling: A concise synthesis of a modified thymidine phosphoramidite. *Tetrahedron Lett.* **29**, 5221–5224 (1988).
315. Ren, X. Enzymatic synthesis of chemically modified DNA for analytical and diagnostic applications. (University of Oxford, 2015).
316. Klimkowski, P., De Ornellas, S., Singleton, D., El-Sagheer, A. H. & Brown, T. Design of thiazole orange oligonucleotide probes for detection of DNA and RNA by fluorescence and duplex melting. *Org. Biomol. Chem.* **17**, 5943–5950 (2019).
317. Said, H., Guttroff, C., El-Sagheer, A. H. & Richert, C. Two-step synthesis of a 5-azidothymidine building block for the assembly of oligonucleotides for triazole-forming ligations. *Synlett* **23**, 2923–2926 (2012).
318. Tiecco, M., Di Profio, P., Germani, R. & Savelli, G. Synthesis of novel 5'-uridine-head amphiphiles as model for DNA molecular recognition. *Nucleosides, Nucleotides and Nucleic Acids* **28**, 911–923 (2009).
319. Abramova, T. V., Koroleva, L. S. & Silnikov, V. N. New orthogonally trifunctionalized morpholine nucleosides. *Mendeleev Commun.* **29**, 169–171 (2019).