

Synthesis of Chemically Modified Guide RNAs and their Applications



Thanishta Mungur

St Catherine's College, University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Hilary 2023

Declarations

I declare that the work carried out in this thesis titled “Synthesis of Chemically Modified Guide RNAs and their Applications” was carried out and written by the author Thanishta Mungur under the supervision of Professor Tom Brown at the University of Oxford, United Kingdom. Colleagues’ collaborations are explicitly acknowledged and any material from the work of other people have been acknowledged in accordance with standard referencing practices.

The copyright of this thesis belongs to the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives License. Researchers are free to distribute or transmit the thesis, provided that they fully acknowledge it. This thesis should not be used for commercial purposes, altered, transformed or built upon.

Acknowledgements

I would to express my heartfelt gratitude to my supervisor Professor Tom Brown for providing me with the opportunity to work in his group and allowing me to scientifically contribute in such an interesting field. His support and guidance have been invaluable throughout my studies. I would like to thank Professor Afaf El-Sagheer and Dr Lapatrada Taemaitree for their contribution to my work and for always offering their insightful advice.

I would also like to thank the Boehringer Ingelheim Foundation for supporting me financially and without whom, this work would not have been possible. I would like to thank Professor Angela Russell for allowing me to use her tissue culture lab to carry out essential experiments for this work.

I have made friends for life in the Tom Brown group – Phoebe, Nicolo, Chunsen, Cam, Sven, Jerry, Diallo, Lillian, Brendan, Cheng and Sebastian. They got me through the ups and downs of research life. Thank you for celebrating good days with me and cheering me up on the not so good days.

I would most importantly like to thank my family – Mum, Dad, my brother and my partner for being my rock during the past 4 years. Their love and support are what get me through everything in life. Thank you for always believing in me!

Abstract

The genome editing tool CRISPR has transformed our understanding of how genetics affects human health. The Cas9 endonuclease, guided by a guide RNA (gRNA) has been used for genome editing, live cell imaging as well as for diagnostic purposes. The CRISPR-Cas9 system has been repurposed for live-cell imaging by mutating the Cas9 protein to inactivate the DNA cleavage activity (dCas9) and by appending green fluorescent proteins (GFP). This enables sgRNA-programmed localisation of GFP to any genomic loci. However, this method results in significant background fluorescence. To overcome this, we designed a metastable fluorescent gRNA that is only activated once it reaches the nucleus, to label its target. Live cell imaging showed a significant reduction in cytoplasmic accumulation of the fluorophore. To study cells and the genome, key modifications must be made on the sgRNA such as fluorophores or ligands that can aid cell delivery. These site-specific modifications cannot be introduced by transcription-mediated RNA synthesis and are expensive and difficult to source by solid-phase synthesis approach. We developed a novel method of introducing chemical modifications into the sgRNA; by extending the 3'-end and by crosslinking a short oligonucleotide that bears the modification. All the modifications were well tolerated by Cas9 *in vitro*. However, further work is required to test the modified sgRNAs in cells. Lastly, the R-loop structure was exploited as a means to introduce modifications into the CRISPR-Cas9 system. To test this hypothesis, we designed a short fluorescent oligonucleotide to bind to the R-loop structure formed after the sgRNA-Cas9 complex binds to their target, the telomeres. This method proved to be a successful tool for live cell imaging and will be applied to other areas in the future.

Abbreviations

A/C/G/T/U	Adenine / Cytosine / Guanine / Thymine / Uracil
ACE	Bis(acetoxyethoxy)-methyl ether
AF488	Alexa Fluor 488
AF594	Alexa Fluor 594
agRNA	Aptazyme-embedded guide RNA
APS	Ammonium persulfate
ASGPr	Asialoglycoprotein
ASO	Antisense oligonucleotide
ATP	Adenosine triphosphate
BCN	Bicyclo[6.1.0]nonyne
BHQ	Black Hole Quencher
BFP	Blue Fluorescent Protein
BiFC	Bimolecular fluorescence complementation
bp	base pair
BNA	Bridged Nucleic Acid
BTT	5-(Benzylmercapto)-1H-tetrazole or Benzylthiotetrazole
Cas9	CRISPR associated protein 9
Cas12a	CRISPR associated protein 12a
ChIP-seq	Chromatin immunoprecipitation followed by sequencing
CO ₂	Carbon dioxide
CPG	Controlled Pore Glass
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats

crRNA	CRISPR RNA
CTD	C-terminal domain
CTP	Cytidine triphosphate
Cy3	Cyanine-3
°C	Degrees Celcius
DAPI	4',6-diamidino-2-phenylindole
DBCO	Dibenzocyclooctyne
dCas9	Deactivated Cas9
DCM	Dichloromethane
DMEM	Dulbecco's Modified Eagle Medium
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DMT	4,4'-dimethoxytrityl
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	deoxyribonucleoside triphosphate
DPBS	Dulbecco's Phosphate-Buffered Saline
DSB	double-strand break
dsDNA	double-stranded DNA
DTT	Dithiothreitol
EDA	Ethylenediamine
EDTA	Ethylenediaminetetraacetic acid
EGFP	Enhanced Green Fluorescent Protein
EMSA	Electrophoretic Mobility Shift Assay
eOGN	External oligonucleotide

eq.	Equivalents
ETT	5-Ethylthio-1H-tetrazole or Ethylthiotetrazole
FACS	Fluorescence-Activated Cell Sorting
FBS	Fetal Bovine Serum
FISH	Fluorescence In Situ Hybridisation
FP	Fluorescence Protein
FRET	Fluorescence Resonance Energy Transfer
FS	Fluoro phosphorothioate
fgRNA	fluorencent guide RNA
GFP	Green Fluorescent Protein
gRNA	guide RNA
GTP	Guanosine triphosphate
h	Hour
HCl	Hydrochloric Acid
HDR	Homology-directed Repair
HEG	Hexaethylene glycol
HEK	Human Embryonic Kidney
HEPES	Hydroxyethyl piperazineethanesulfonic acid
HFIP	Hexafluoroisopropanol
HPLC	High Performance Liquid Chromatography
HR	Homologous recombination
IVT	<i>In vitro</i> transcription
^{CNV} K	Cyanovinylcarbazole
KRAB	Krüppel-associated box
LNA	Locked Nucleic Acid

LRMS	Low resolution mass spectrometry
LS	Large Subunit
M	Molar
MB	Molecular beacon
mg	Milligrams
min	Minutes
mL	Millilitres
mM	Millimolar
mmol	Millimoles
μL	Microlitres
μM	Micromolar
MOP	Methoxypsoralen
mRNA	Messenger RNA
MS	2'-O-methyl phosphorothioate
MSP	2'-O-methyl thiophosphonoacetate
ncRNAs	non-coding RNAs
ng	nanograms
NHEJ	Non-homologous End Joining
NHS	<i>N</i> -hydroxysuccinimide
NLS	Nuclear Localisation Signal
NMI	<i>N</i> -methylimidazole
NMR	Nuclear Magnetic Resonance
nt	Nucleotide
NTP	Nucleoside triphosphate
NUC	Nuclease (domain)

OD	Optical density
OGN	Oligonucleotide
PAGE	Polyacrylamide Gel Electrophoresis
PAM	Protospacer-Adjacent Motif
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PE	Prime Editing
pegRNA	Prime Editing guide RNA
PFS	Protospacer-flank site
pH	Potential of hydrogen
PNA	Peptide nucleic acid
PO	Phosphodiester
PS	Phosphorothioate
QD	Quantum Dot
qgRNA	Quenched guide RNA
QY	Quantum Yield
REC	Recognition (domain)
RFP	Red Fluorescent Protein
RNA	Ribonucleic acid
RNase	Ribonuclease
RNP	Ribonucleoprotein
RP-HPLC	Reverse Phase-HPLC
rpm	Revolutions per minute
rRNA	Ribosomal RNA
RT	Real Time

RT-PCR	Reverse transcription polymerase chain reaction
rt	Room temperature
SE	Succinimidyl ester
Sec	Seconds
sgRNA	Single guide RNA
siRNA	Short interfering RNA
smFRET	Small molecule FRET
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
SS	Small Subunit
TALE	Transcription Activator-like Effector
TALEN	Transcription Activator-like Effector Nuclease
TAMRA	Carboxytetramethylrhodamine
TBDMS	tert-butyldimethylsilyl
TBE	Tris-borate-EDTA
TC	Thionocarbamate
TCA	Trichloroacetic acid
TEA	Triethylamine
TEAB	Triethylammonium Bicarbonate
TEMED	<i>N,N,N',N'</i> -tetramethylethane-1,2-diamine
TF	Transcription Factors
TFIID	Transcription Factor II D
THF	Tetrahydrofuran
TOM	Triisopropylsilyloxymethyl
tracrRNA	Trans-activating CRISPR RNA

TRIS	Tris(hydroxymethyl)amino-methane
tRNA	Transfer RNA
TTP	Thymine Triphosphate
U	Units
^{cv} U	5-carboxyvinyl-2'-deoxyuridine
UPLC-MS	Ultra Performance Liquid Chromatography Mass
Chromatography	
UTP	Uridine triphosphate
UV	Ultraviolet
v/v	Volume by volume
vol	Volume
WED	Wedge (subdomain)
ZFN	Zinc-finger Nuclease

Table of Contents

Introduction	17
1.1 Basics of Nucleic Acids	17
1.1.1 Nucleic acid structure	17
1.1.2 Nucleic acid duplexes and forms	19
1.2 The Genome and its Functions	22
1.2.1 DNA replication	23
1.2.2 Transcription	25
1.2.3 Translation	27
1.3 Genome Editing	30
1.3.1 Early Days of Genome Editing	31
1.3.2 Double stranded breaks and DNA repair	31
1.3.3 Engineered nucleases	32
1.4 Discovery of CRISPR-Cas	34
1.4.1 Adaptive Immune Mechanism in Bacteria	34
1.4.2 Classification of CRISPR-Cas systems	37
1.4.3 CRISPR-Cas9 in Gene Editing	39
1.5 Structure of Cas9 enzyme	40
1.5.1 Bilobed Structure of Cas9	40
1.5.2 HNH and RuvC Nuclease Domains	42
1.6 DNA Targeting and Cleavage by CRISPR-Cas9	43
1.6.1 Interactions with the gRNA	43
1.6.2 Target Search and PAM Recognition	44
1.6.3 Local DNA Melting and R-loop Formation	46
1.6.4 DNA Cleavage	47
1.7 Modified Cas9 and its Applications	48

1.7.1	Cas9 Orthologs with Varying PAM Sequences	48
1.7.2	High-fidelity Cas9	50
1.7.3	Cas9 Nickases	51
1.7.4	Catalytically Inactive Cas9 (dead Cas9 or dCas9)	52
1.8	Chemical Modifications of the sgRNA	53
1.9	Synthesis of the gRNA	56
1.9.1	In vitro transcription	56
1.9.2	Solid-phase synthesis	58
1.10	Aims and Objectives	62
	Live Cell Imaging using fluorescently modified gRNAs	64
2.1	Background	64
2.1.1	Genomic Imaging Using Traditional Approaches	65
2.1.2	Live cell imaging using sequence specific methods	67
2.1.2.1	Live cell imaging using modified dCas9-CRISPR systems	67
2.1.2.2	Live cell imaging using modified sgRNA-CRISPR systems	70
2.2	Hypotheses and Aims	74
2.3	RNase H activated fgRNA concept	74
2.4	Fluorescent test probes	76
2.4.1	Design strategy and synthesis	76
2.4.2	Fluorescence Studies	78
2.5	RNase H Activated fgRNA Design I	81
2.5.1	Design Strategy and Synthesis	81
2.5.2	RNase H mediated in vitro Fluorescence Studies	84
2.5.3	Live Cell Imaging	86
2.6	Fluorescent Test Probe for Stability	90
2.6.1	Design Strategy	90
2.6.2	RNase H mediated in vitro Fluorescence Studies	92
2.6.3	Live Cell Imaging	93
2.7	RNase H Activated fgRNA Design II	94
2.7.1	Design Strategy	94

2.7.2 RNase H mediated in vitro Fluorescence Studies	95
2.8 Crosslinked Version of fgRNA Design II	99
2.8.1 Design Strategy	99
2.8.2 RNase H mediated in vitro Fluorescence Studies	102
2.8.3 Live Cell Imaging	103
2.9 Conclusions and Future Work	107
3'-Modified sgRNAs using crosslinking to improve synthesis	109
3.1 Background	109
3.1.1 Modifications to induce photo-crosslinking	110
3.1.1.1 Psoralen Modifications	110
3.1.1.2 3-Cyanovinylcarbazole modifications	113
3.1.2 Modified CRISPR Cas9 systems	119
3.1.2.1 Modifications on Cas9	119
3.1.2.2 Modifications on the gRNA	123
3.2 Hypotheses and Aims	125
3.3 Design and Synthesis of the Modified sgRNAs	126
3.5 In Vitro Cleavage Assay	129
3.5 In Cellulo Testing	130
3.5.1 Effect of the Constructs on Cells	130
3.5.2 In Cellulo Gene Editing Activity	132
3.5.2.1 T7E1 Assay	132
3.5.2.2 Gymnotic Delivery using the modified sgRNAs	132
3.5.3 Live Cell Imaging	134
3.6 Conclusion and Future Work	135
R-loop Structure Exploitation to introduce modifications	137
4.1 Background	137
4.1.1 Study of the R-loop structure	137
4.1.2 Role of the R-loop in 3'–3' Dual gRNAs	144
4.1.3 Role of the R-loop in Prime Editing	146
4.1.4 5'-extended sgRNAs	147

4.2 Hypotheses and Aims	148
4.3 Proof of Concept Using Live Cell Imaging	150
4.3.1 Design Strategy	150
4.3.2 Electrophoretic Mobility Shift Assay (EMSA)	151
4.3.3 Live Cell Imaging	153
4.4 Phosphorothioate linkages to improve nuclear localisation	156
4.4.1 Design Strategy & Synthesis	156
4.4.2 Live Cell Imaging	157
4.4.3 Toxicity of Phosphorothioates	160
4.4.4 Effect of Phosphorothioate Concentration on Aggregation	162
4.5 Optimisation of the fluorescent dye	164
4.5.1 Hydrophobicity of Cy3	164
4.5.2 Live Cell Imaging using AlexaFluor 594-labelled probes	166
4.6 Specificity of PS-eOGN probes	167
4.6.1 Scrambled probe design	167
4.6.2 Live cell using scrambled PS-eOGN	168
4.7 Conclusions and Future Work	169
Conclusions	171
Experimental	174
6.1 Oligonucleotide synthesis and Purification	174
6.1.1 DNA synthesis	174
6.1.2 RNA synthesis	175
6.1.3 Monomers and solid supports	176
6.2 Oligonucleotide Deprotection	177
6.2.1 Selective β -cyanoethyl removal	177
6.2.2 DNA deprotection	177
6.2.3 2'-O-TBDMS RNA deprotection	177
6.2.4 2'-O-TC RNA deprotection	178
6.3 Oligonucleotide purification	178
6.3.1 RP-HPLC	178

6.3.2 Desalting	178
6.3.3 Denaturing PAGE	179
6.4 Oligonucleotide mass spectrometry	180
6.5 Native PAGE	181
6.6 Agarose Gel Electrophoresis	181
6.7 In vitro Transcription	182
6.8 Post-synthetic oligonucleotide labelling	182
6.8.1 N-Hydroxysuccinimide (NHS) labelling	182
6.8.2 Click chemistry labelling (CuAAC)	183
6.9 Real-time fluorescence measurements	184
6.9.1 Real-time fluorescence with stoichiometric amount of quencher	184
6.9.2 Real-time fluorescence with increasing amount of quencher	184
6.9.3 Melting temperature studies	184
6.10 Crosslinking ligation of oligonucleotides	185
6.11 Electrophoretic Mobility Shift Assay	186
6.12 In vitro Cas9 Cleavage Assays	186
6.13 Cell Culture	186
6.14 Lipid Transfection for Live Cell Imaging	187
6.15 Confocal Microscopy	188
6.16 Gymnotic Delivery	188
6.17 Genomic DNA Extraction	189
6.18 T7E1 Assay	189
6.19 Chemical Synthesis	190
6.19.1 Synthesis of 6-Azido hexanoic acid NHS ester	190
6.19.2 Synthesis of Folic Acid NHS Ester	192
Appendix	194
References	204

Chapter 1

Introduction

1.1 Basics of Nucleic Acids

1.1.1 Nucleic acid structure

Nucleic acids are biopolymers, found in all living organisms, where they exist to process, encode and store information. They are composed of monomeric subunits known as nucleotides (nt) and each nucleotide consists of three components: a nitrogen heterocycle (nucleobase), a pentose sugar and a phosphate group (Figure 1.1).

In nature, two forms of nucleic acid polymers exist; DNA (deoxyribonucleic acid) and RNA (ribonucleic acids). The nucleobases combine with a deoxyribose sugar (2'-H) in DNA and a ribose sugar (2'-OH) in RNA. The four nucleobases in DNA are the purines:

adenine (A) and guanine (G), and the pyrimidines: cytosine (C) and thymine (T). In RNA, T is replaced by uracil (U) which lacks a methyl group at C5' position of the heterocycle.

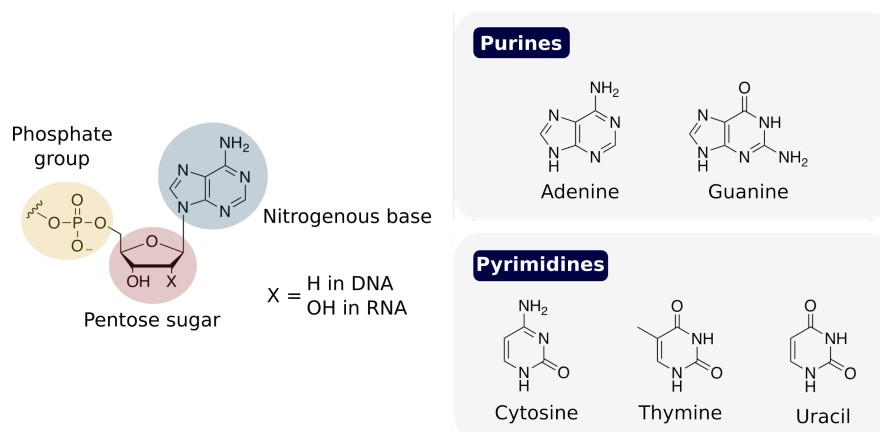


Figure 1.1 | Representation of the three components that make up a nucleotide and the structures of the five naturally occurring nitrogenous bases: Adenine (A), Guanine (G), Cytosine (C), Thymine (T) and Uracil (U).

The bond joining the C1' position of the sugar to the nucleobase is known as the *N*-glycosidic bond. Rotation about this bond can give rise to *syn*- and *anti*-conformations (Figure 1.2a). Rotation about the bond is restricted and due to steric reasons, the *anti*-conformation is generally favoured. The nucleobase can also be attached on the same face as the sugar, in the β -configuration, or on the opposite face, in the α -configuration (Figure 1.2b). However, in naturally occurring DNA, all the nucleotides are in the β -configuration.

The 5'-phosphate group of a nucleotide can form a covalent bond with a 3'-hydroxyl group of another nucleotide, to form a phosphodiester bond. A chain of nucleotides with several such bonds forms an oligonucleotide (Figure 1.2c). At physiological pH (7.4), phosphodiester bonds are negatively charged, making oligonucleotides highly

charged polyanionic molecules and soluble in aqueous environment. Moreover, the nucleic acid chain has directionality as one end has a 5'-hydroxyl group while the other end has a 3'-hydroxyl group. Conventionally, nucleic acid sequences are written in the 5' to 3' direction. Nucleic acids have hierarchical levels of complexity from linear chains to more complex structures where two or more oligonucleotides can interact with each other and adopt different forms.

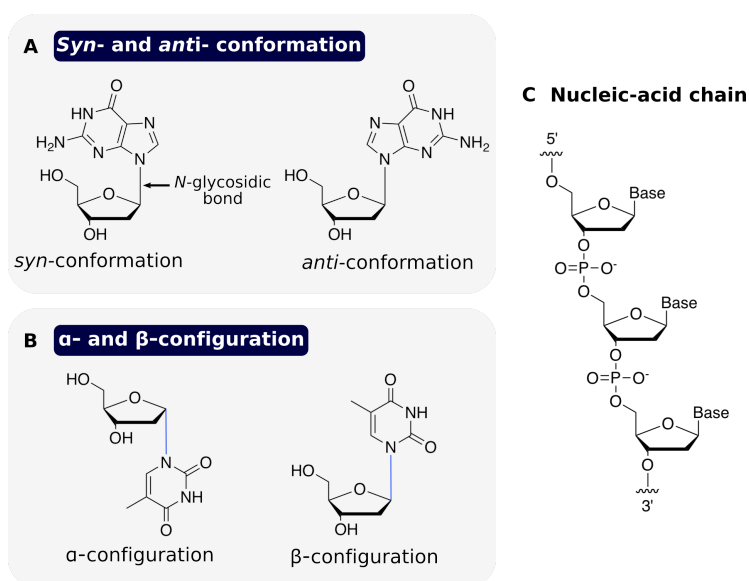


Figure 1.2 | A) *syn*- and *anti*-conformation of the nucleobase. B) α- and β- configurations of the nucleobase. C) Nucleic-acid chain, connecting individual nucleobase by phosphodiester bonds in the 5' to 3' direction.

1.1.2 Nucleic acid duplexes and forms

Oligonucleotides form secondary structures, known as duplexes, through the formation of hydrogen bonds between two complementary strands anti-parallel to each other. As modelled by Watson and Crick,¹ A pairs with T (or U in RNA) *via* the formation of two hydrogen bonds whereas C pairs with G *via* the formation of three hydrogen bonds and as such, the ratio of purines to pyrimidines is always 1:1 in duplexes (Figure 1.3a).² This results in the formation of a stable helical structure

running in opposite directions.³⁻⁴ In this structure, the hydrophobic nucleobases are internally protected while the hydrophilic phosphodiester backbone is exposed to the aqueous environment. Through the twisting of this helical structure, along with the geometry of the nucleobases, major grooves and minor grooves are formed along the length of the duplex. The helix can adopt the right-handed conformation A- or B-form or the less commonly adopted, left-handed Z-form (Figure 1.3b).

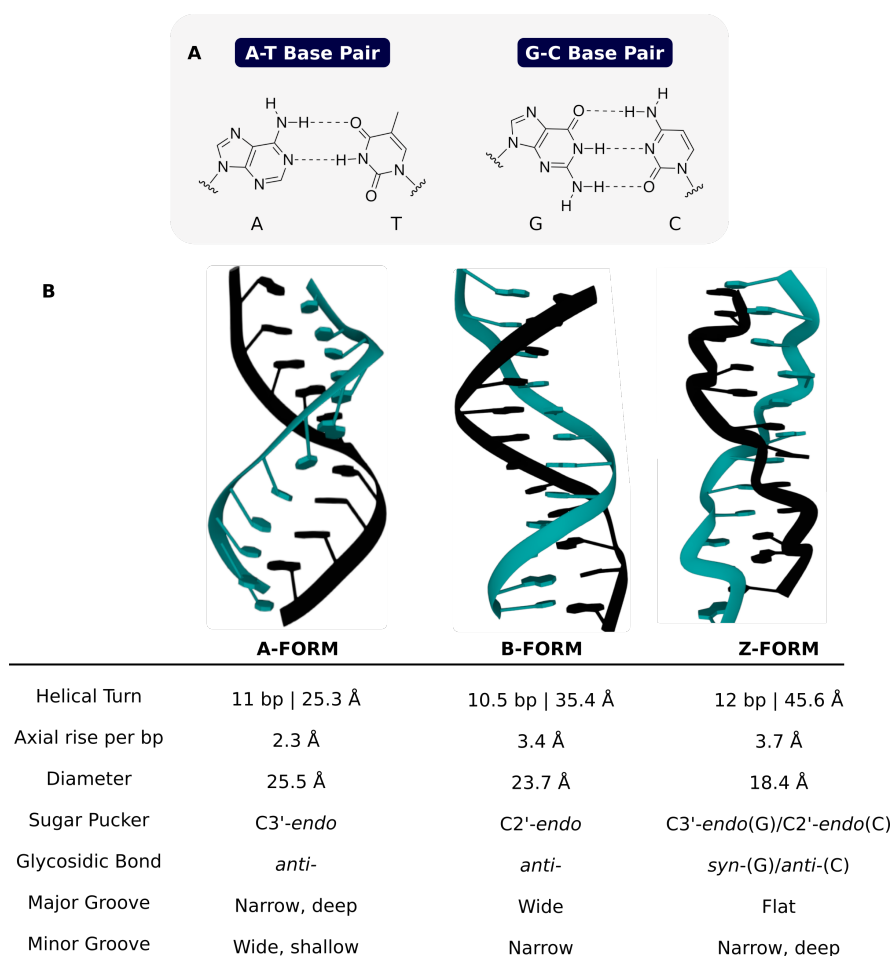


Figure 1.3 | Structures of the purines and pyrimidines and their interactions via hydrogen bonds, adapting to different forms of the nucleic acid. (A) The purine base A forms two hydrogen bonds with pyrimidine T (or U in RNA) while purine base G forms three hydrogen bonds with pyrimidine C. The purine bases consist of two heterocyclic rings while pyrimidines possess a single ring system. (B) shows the characteristics of the different nucleic acid conformations, A-form is displayed by an RNA-DNA heteroduplex (PDB: 5F9R) while the Z-form and B-form is depicted by a DNA duplex (PDB: 1BNA and 4OCB, respectively).⁵⁻⁷

The conformation adopted depends on various factors such as sugar pucker and *anti-syn* conformation of the nucleobase.⁸ By adopting different puckering conformations, sugar rings can alter the relative orientation of the phosphate backbone, the nucleobase and the ribonucleotide 2'-OH. The C2' and C3' of the sugar can be displaced from the C1'-O4'-C4' plane: The C2' can be displaced above the plane while the C3' displaced below, adopting a C2'-endo conformation (Figure 1.4). The opposite arrangement is also possible, with the C3' displaced above the plane and the C2' displaced below, adopting a C3'-endo conformation. The sugar pucker conformation affects the overall structure of the helix and enables DNA and RNA to adopt the most form.⁹ These structural effects translate into functional effects such as the differences in the major and minor grooves seen in the A- and B-form, which has crucial roles in protein recognition.¹⁰⁻¹¹

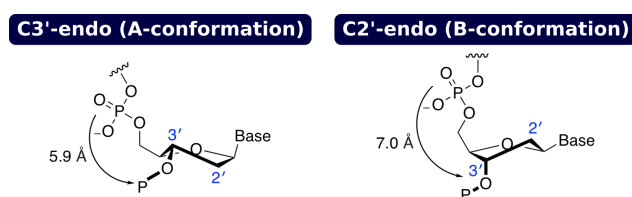


Figure 1.4 | Sugar pucker of the nucleobase showing the distance between the neighbouring phosphate atoms.

Double stranded DNA (dsDNA) predominantly adopts the B-form with C2'-endo sugar pucker while dsRNA favours C3'-endo pucker due to the steric bulk of the hydroxyl group and adopt the A-form. However, it is not uncommon to find DNA adopting the A-form when there is a small energy barrier between the two forms.¹² The A-form is also adopted by DNA-RNA heteroduplexes, a crucial aspect of transcription and sgRNA-DNA complexes in CRISPR systems (further discussed in section 1.6.3). Lastly, DNA can also adopt the Z-form which is significantly different from previously

discussed conformations; it is a left-handed helix and comprises of both *syn*- and *anti*- conformations resulting in a characteristic zig-zag structure.¹³ Specific sequences of these dsDNA are stored in cells, allowing fundamental processes to take place.

1.2 The Genome and its Functions

In most organisms, the genetic material is stored in the form of DNA. Approximately 3×10^9 base pairs of DNA are distributed over 23 pairs of chromosomes.¹⁴ Each of these chromosomes contains a single, very long dsDNA, wrapped around proteins called histones (Figure 1.5). This DNA-protein complex is called chromatin. The complex winds further, loops and condenses into the chromosome.¹⁵ This comprises the human genome which contains around 30 000 genes and which codes for proteins.¹⁴ Before DNA can be used for protein synthesis, an RNA transcript must be made by a process called transcription. The majority of the human genome is transcribed (at least 76%), however, only 1.2% of this transcribed RNA is protein-coding and the rest are non-coding RNAs (ncRNAs). Protein-coding RNA transcript is subsequently used in a process called translation to synthesise the protein. This is known as the 'Central dogma of molecular biology', which states that DNA makes RNA which then makes proteins.¹⁶ Some of the ncRNAs are functional and are responsible for enzymatic catalysis during splicing and translation, whereas others control gene expression at various levels, constituting an RNA-based regulatory network.

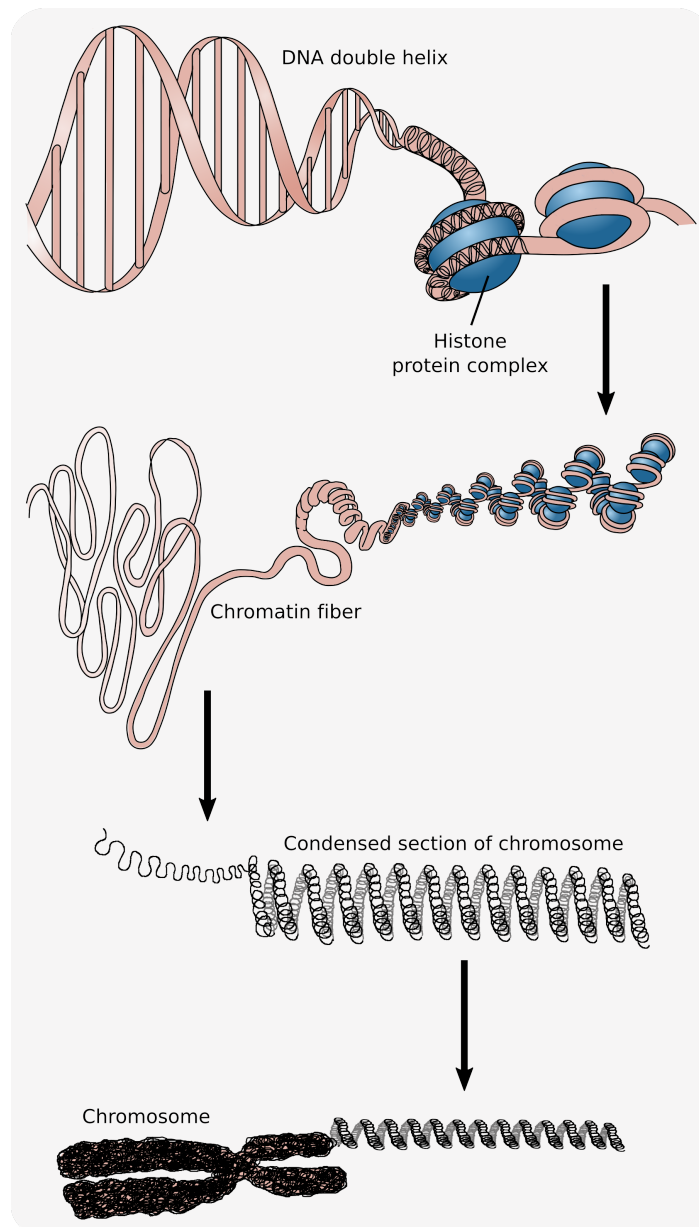


Figure 1.5 | DNA packaging: DNA wrapped around proteins called histone to form chromatin, followed by extensive coiling, looping and condensing to form the chromosome.

1.2.1 DNA replication

Before moving on to discuss the synthesis of RNA from DNA, it is important to acknowledge that cells regenerate themselves as often as every second. It is therefore imperative for the cells to preserve their genetic information so that the exact sequence can be passed on to the next cell. This process is known as DNA replication.

The template model for DNA replication was first proposed by Watson and Crick in 1953.¹⁷ When a cell divides, part of its dsDNA splits into two single strands, forming a Y-shaped structure called a 'replication fork' (Figure 1.6). Separation of the dsDNA is controlled by an enzyme known as helicase and one of the single stranded DNA is known as the leading strand while the other is called the lagging strand. The enzyme DNA polymerase uses the triphosphates dATP, dTTP, dCTP and dGTP to add nucleotides complementary to the parent strand one at a time, thereby synthesising DNA in the 5' to 3' direction.

Two DNA polymerase are present at the fork at any one time; one of them moves continuously to synthesise the new daughter DNA molecule on the leading strand while the other produces short bursts of around 1000 bases, known as 'Okazaki fragments',¹⁸ on the lagging strand. The polymerase on the lagging strand requires a short primer, with a free 3'-hydroxyl group, to initiate DNA synthesis. This is produced by an enzyme called DNA primase, which makes a short RNA molecule (the primer) onto which DNA polymerase adds the nucleotides. The 'Okazaki fragments' are sealed by the enzyme DNA ligase and the RNA is digested by an exonuclease at the end of replication. Thus, one parent strand is exactly copied into two daughter strands.¹⁹

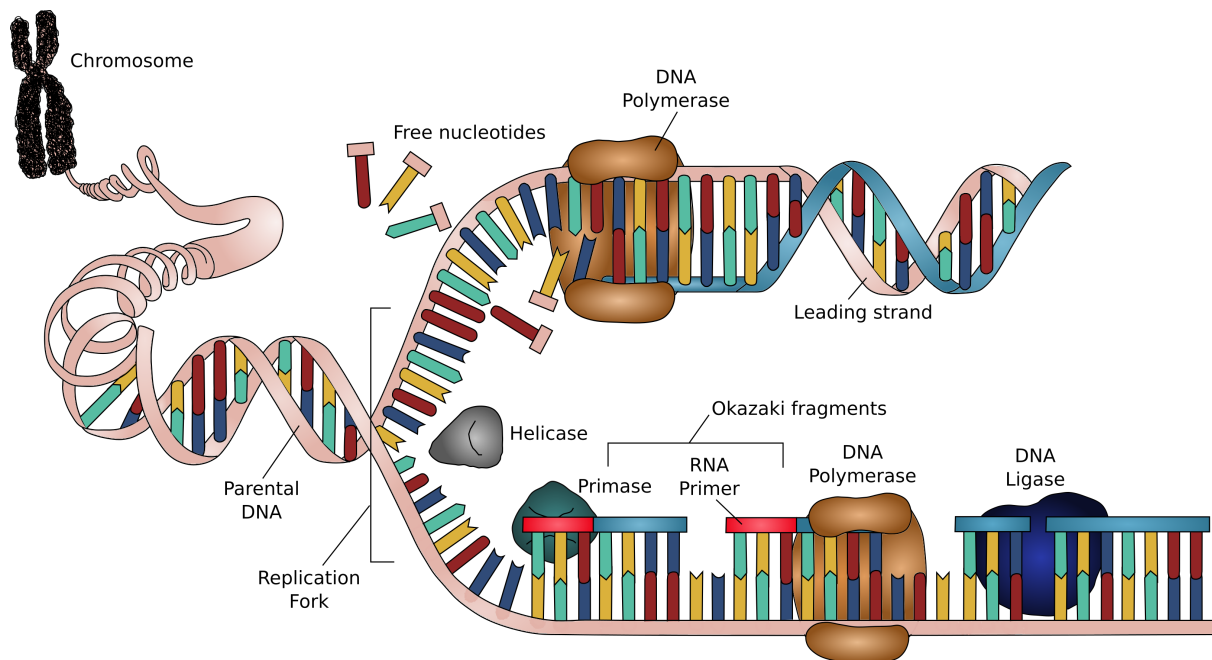


Figure 1.6 | Schematic of the process of DNA replication showing unwinding of the DNA helix by helicase at the replication fork, DNA polymerase synthesising the new daughter DNA at the leading strand and RNA primase, DNA ligase and DNA polymerase working together to join the Okazaki fragments.

1.2.2 Transcription

Transcription has various similarities to the process of DNA replication. During transcription, DNA copied into a messenger RNA (mRNA), which carries the information required for protein synthesis. Hence, the process of transcription is at the epicentre of gene expression.²⁰

Transcription begins with the partial unwinding of the dsDNA, by RNA polymerase. One of the two strands, known as the antisense strand, acts as a template for RNA synthesis. Watson-Crick base pairing between the exposed DNA nucleotides and ribonucleotide triphosphates occurs and RNA polymerase catalyses the formation of the phosphodiester bonds between the ribonucleotide. RNA polymerase moves stepwise along the DNA; it unwinds and exposes only the part of the dsDNA that is

just ahead of the active site for phosphodiester linkage (Figure 1.7a). As a result, the chain produced only grows by one nucleotide at a time, in a 5' to 3' direction, and is a copy of the DNA strand not acting as the template; the sense strand. Unlike DNA replication, the newly formed RNA chain does not remain hydrogen bonded to the DNA strand. The RNA dissociates just behind the region where the ribonucleotides are being added and the DNA double helix re-forms. In addition, compared to DNA which can be up to 250 million bp long, the mRNA is much shorter; around a few thousand nucleotides.

For RNA polymerase to initiate transcription, it does not require a primer. This is presumably because transcription does not need to be as accurate as DNA replication as RNA is short-lived and does not store genetic information in cells permanently. Initiation factors, called general transcription factors, are required to position RNA polymerase at transcription-initiation sites or the promoter region of the DNA (Figure 1.7b). The promoter has a TATA box in its sequence, located 25 bp upstream of the transcription site, with a conserved DNA sequence which is mostly TATAAA.²¹ The TATA box is recognised by one of the general transcription factors, TFIID, which in turn allows other transcription factors to bind, eventually recruiting RNA polymerase to bind, unwind and start synthesising the RNA transcript.

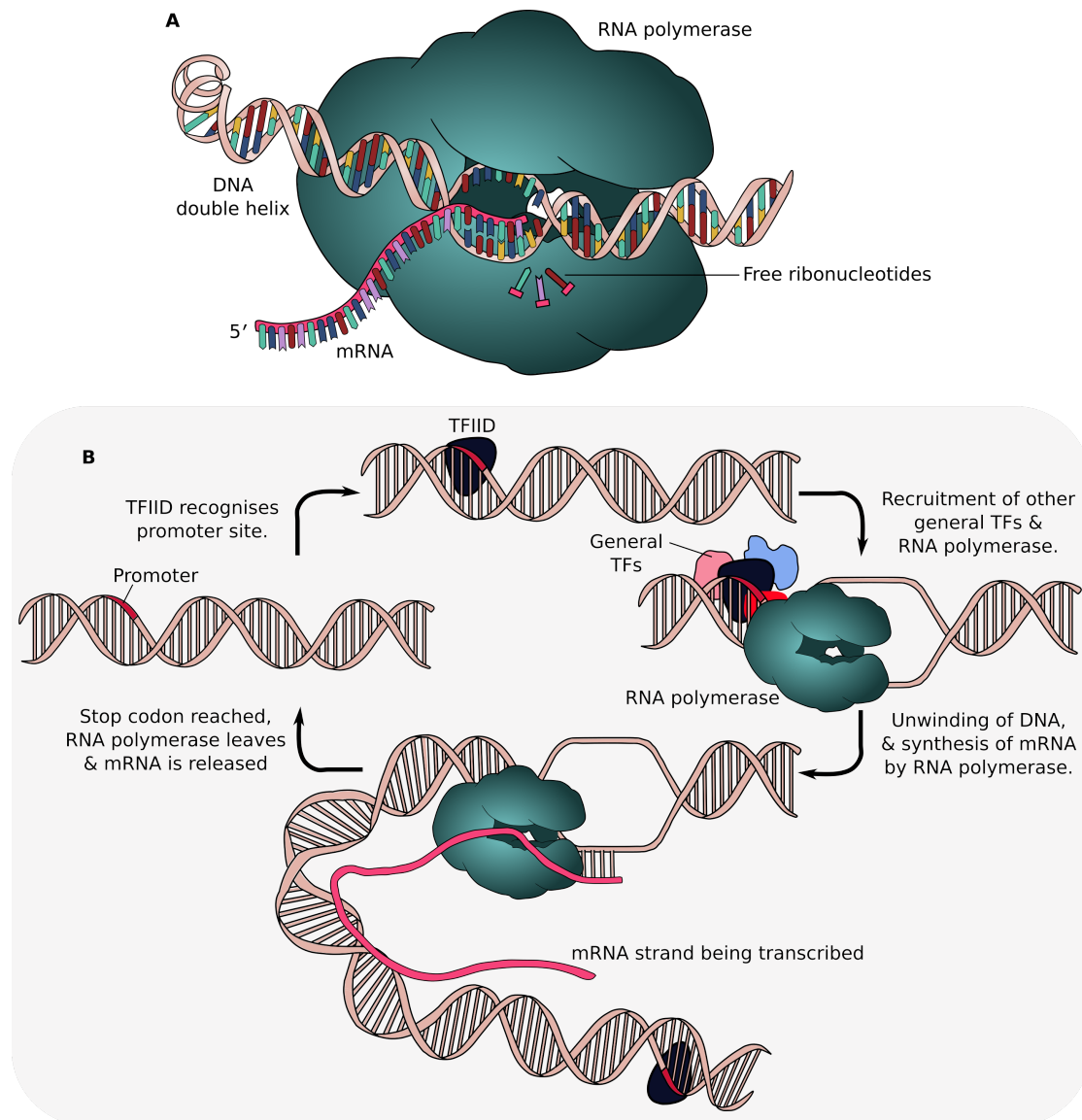


Figure 1.7 | Schematic of the transcription process. (A) Schematic of RNA polymerase partially unwinding and transcribing DNA. mRNA strand being made dissociates immediately after new ribonucleotide has been added. (B) Representation of the process of transcription (starting from middle left) where first the promoter site is recognised by TFIID, followed by recruitment of additional TFs and RNA polymerase. RNA polymerase then starts unwinding the DNA double helix and start transcription at the start codon. Upon reaching the stop codon, RNA polymerase dissociates and is recycled, releasing the mRNA.

1.2.3 Translation

The mRNA from the nucleus is more specifically transported to the ribosomes, constituting of the small ribosomal subunit (SS) and the large ribosomal subunit (LS), where it directs protein synthesis.²² The information encoded in the mRNA is read in

groups of three nucleotides, called codons. These codons represent amino acids and there are 61 different codons for amino acids. There are also three 'stop' codons to terminate the polypeptide chain and one 'start' codon, AUG that codes for the amino acid methionine, to initiate translation. Watson-Crick base pairing occurs between the codons of mRNA and the anticodon of another type of RNA called transfer RNA (tRNA). The tRNA is attached to an amino acid, specified by that codon.

Translation occurs in three key steps: initiation, elongation and termination. During initiation, a tRNA carrying the amino acid methionine binds to small ribosomal subunit (40S) (Figure 1.8a). This complex assembles on the 5'-end of the mRNA after recognising a 5'-guanine triphosphate (GTP) cap. The complex then begins to scan along the mRNA strand in the 5' to 3' direction until it reaches the start codon. Upon recognition of the start codon, the methionine-tRNA sits into the peptidyl site (P site) of the small ribosomal subunit. This is preceded by the binding of the large ribosomal subunit (60S) to form the initiation complex.²³

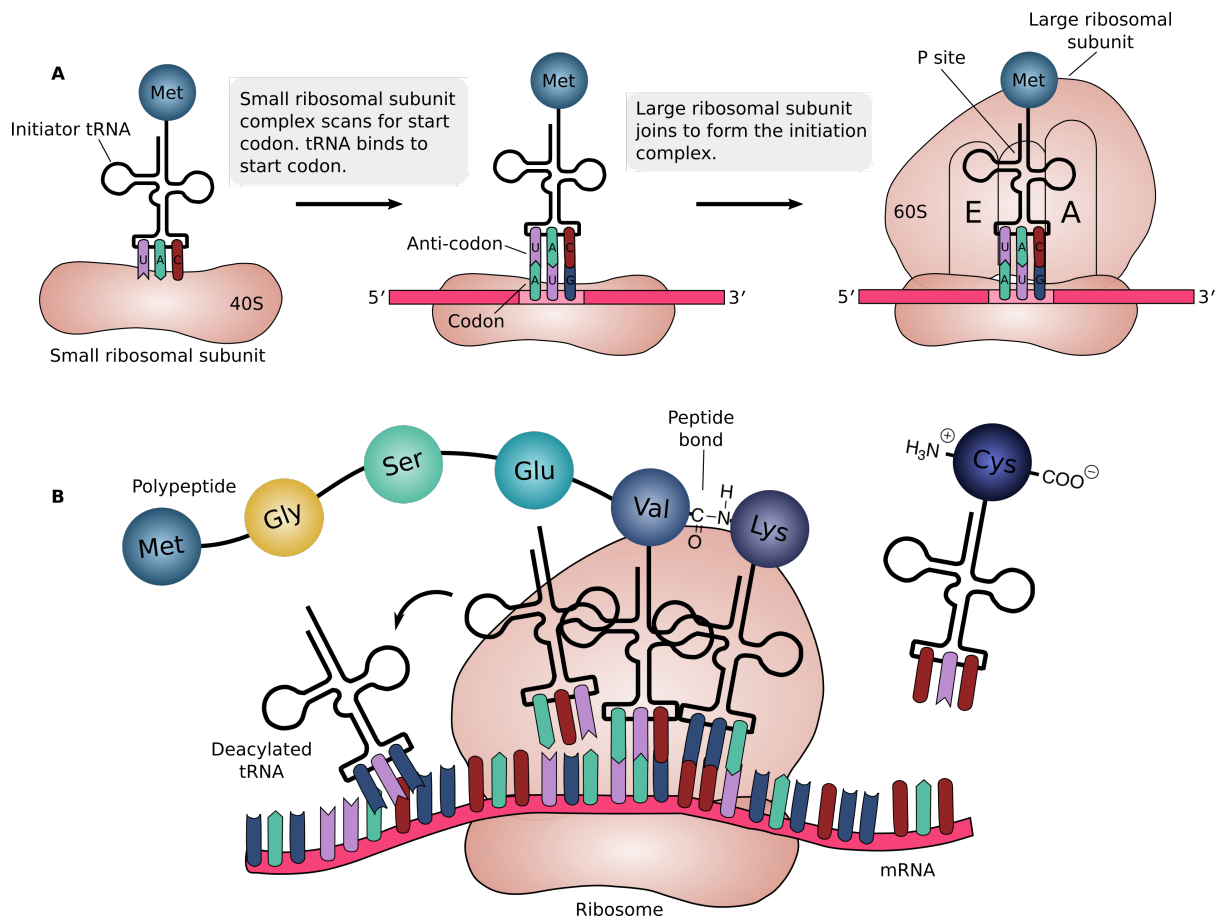


Figure 1.8 | (A) Representation of the processes occurring during the initiation of translation, where the initiator tRNA first binds to the small ribosome, followed by a scan to find the start codon and finally the large ribosome joins to form the initiation complex. (B) Formation of a polypeptide chain during translation where a peptide bond is being formed between the tRNA in the P and A site and the deacylated tRNA is leaving the ribosome.

The tRNA in the P site sits next to a vacant landing site (A site), exposing another three-nucleotides long codon. A second tRNA with complementary anticodon is carried to the A site and the ribosomal peptidyl transferase center catalyses the formation of a peptide bond between the two amino acids. The ribosomal complex is translocated such that the deacylated tRNA falls into the exit site (E site), the peptide-containing tRNA is in the P site and the mRNA is moved by exactly three nucleotides so that another set of codons is exposed in the A site (Figure 1.8b).²³

This process keeps repeating and the polypeptide chain grows until a stop codon is reached. Stop codons are recognised by release factors which fit perfectly into the P sites of ribosomes although they are not tRNAs. These release factors cause the ribosomal peptidyl transferase centre to catalyse the hydrolysis of the ester bond between the polypeptide chain and the tRNA in the P site. The polypeptide chain is released and all the 'used' tRNA which exited through the E site are recycled.²³

The polypeptide is subsequently processed and folded to make functional proteins that control different mechanisms in the body. It was therefore understood that in order to make changes to these proteins, it was the genes that would need to be altered. For this reason, genome editing gained considerable attention in the 1970s and the 1980s and is still a subject undergoing intense study to date.²⁴

1.3 Genome Editing

Genome editing can be used to modify the sequence of the genome and the ability to specially target a particular DNA and make precise changes is the ultimate long-term goal of many scientist. Genome editing is a very promising way of finding treatments for genetic diseases with single point mutations such as cystic fibrosis, haemophilia and sickle cell anaemia. Moreover, it also has potential for the treatment of more complex diseases such as cancer. The applications of gene editing go beyond therapeutics where scientists are trying to 'knock-out' genes in order to study protein functions, inserting new sequences to make more resilient cells and eventually crops and even for genomic imaging.

1.3.1 Early Days of Genome Editing

The first attempt of gene editing involved the incorporation of a foreign DNA into yeast, using homologous recombination (HR).²⁵ Homologous incorporation was further exploited to insert DNA into the human genome. Examples include insertion of DNA into the β -globin locus, correction of a defective gene *via* the injection of copies of the same gene but carrying a different mutation and even the generation of mouse germ line chimeras from embryo-derived stem cells, which had been modified with a targeted gene disruption.²⁶⁻²⁸ Restriction enzymes were used to cut the DNA in all these experiments and they all suffered from the same problems: The rate of HR was extremely low (1 in 10^3 - 10^9 cells),²⁸ cell cycle dependence of HR and random insertions at undesired sites.²⁹⁻³⁰ These limitations hindered the application of this method. However, it was later discovered that HR events could be improved by several thousand folds if double stranded breaks (DSBs) were introduced at the site of insertion, a finding that made genome editing take a massive leap forward.³¹ DSBs occur when the phosphodiester bonds are broken in both strand of the helix at the same point. This results in the liberation of two DNA pieces with blunt ends.

1.3.2 Double stranded breaks and DNA repair

DSBs are one of the most damaging types of DNA lesions and the processing of repair of the break can cause loss of heterozygosity, sequence deletion and chromosome rearrangements, eventually resulting in cell death. DSBs can be repaired *via* two competing mechanisms: Non-homologous end joining (NHEJ) and Homology directed repair (HDR) (Figure 1.9).³² In NHEJ, the two broken ends of the DNA are ligated, without the need for a homologous DNA template.³³ This pathway is prone to errors

and often, substitutions or insertions/deletions (indels) of varying length are observed at the site of the DSB. This results in frameshift mutations and potentially loss of gene function (gene knockout). Compared to NHEJ, HDR is a more precise way of repairing DNA.³⁴ However, it required a homologous DNA template to function. The template has homologous flanking regions to the DSB and the correct or desired genetic sequence to be inserted into the genome through HR, leading to gene correction or precise gene knock-in.

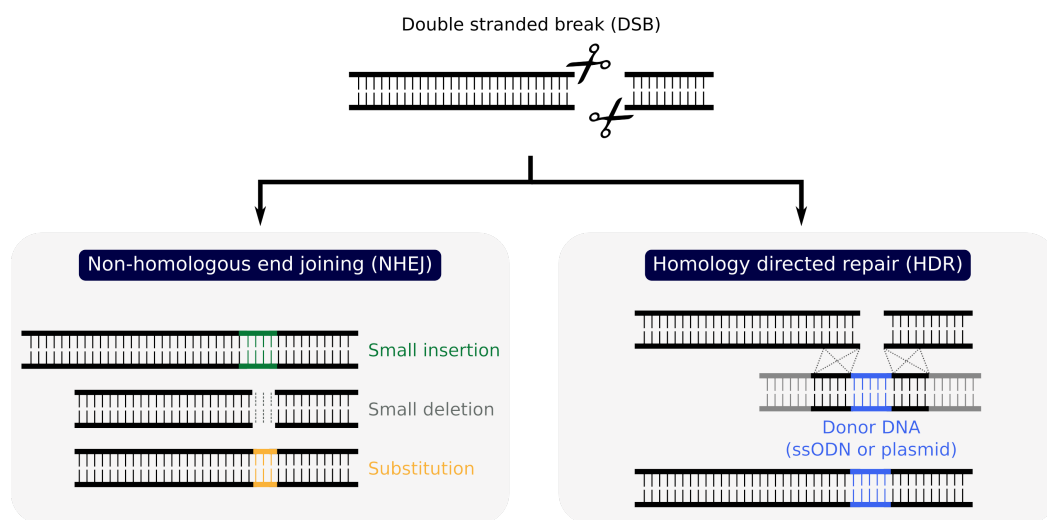


Figure 1.9 | Repair pathways of a DSB by either NHEJ or HDR.

1.3.3 Engineered nucleases

It was obvious that the key to efficient gene editing was the generation of DSBs. Therefore, there was an urge to find new ways of inducing targeted DSBs. Restriction enzymes such as EcoRI, EcoRV, or PvuII under the control of galactose-inducible promoters, were commonly used and they were effective at cutting DNA.³⁵⁻³⁶ However, their applications were limited because they could only cleave short, pre-determined sequences. In 1985,³⁷ a class of natural enzymes called meganucleases were

discovered with the ability to produce DSBs. Meganucleases recognise long DNA targets, approximately 14-40 bp. Although meganucleases successfully introduced sequence specific DSBs, it was very challenging to find the exact meganuclease with clinical or biotechnological relevance. In an attempt to improve the diversity of sequences meganucleases could target, multiple approaches were explored to reengineer meganucleases.³⁸⁻⁴⁰ Those included altering the cleavage specificity of the nuclease or engineering homing endonucleases with novel target specificities. Their successes were nonetheless only moderate, and the expertise needed to reengineer meganucleases hindered further innovations in the area. This led to the development of artificial proteins named zinc-finger proteins (ZFNs).⁴¹⁻⁴² ZFNs were generated by fusing a zinc finger DNA-binding protein to the catalytic domain of FokI, which is a bacterial restriction endonuclease. The zinc dependent domain is made up of individual fingers, which consists of around 30 amino-acids where each amino acid recognises 3 bp in the DNA.⁴³ By combining 3-6 individual zinc fingers, the target binding site can be expanded to 9-18 bp.⁴⁴ In addition, using an inverted pair of zinc fingers targeting the opposite strand of the dsDNA, two adjacent FokI domains can dimerise resulting in a sequence specific DSB. The discovery of ZFNs considerably increased target genome modification *via* HDR.⁴⁵⁻⁴⁶ A simpler way of generating DSBs was later discovered using transcription activator-like effector nucleases (TALENs). TALENs are very similar to ZFNs in that they both consist of an engineered DNA domain fused to FokI nuclease.⁴⁷ The DNA binding domain is derived from transcription activator-like effectors (TALEs) and consists of highly conserved amino acid repeats.⁴⁸ Individual TALE repeats recognise a single base of DNA, thus increasing the number of targetable DNA sequences by simply replacing the zinc-

finger DNA recognition domain of ZFN with a group of TALEs to create transcription activator-like effector nucleases (TALENs).

Despite huge advancements in the field of ZFNs and TALENs, both technologies still remain inaccessible to many users. This is because the level of protein engineering required to construct the DNA-binding domain is expensive, time-consuming and labour-intensive. The recent discovery of the novel CRISPR-Cas technology has simplified adapting nucleases to new targets in programmable manner.⁴⁹ The system is comprised of a nuclease (Cas enzymes), which is guided to a specific DNA sequence by an RNA molecule, called the guided RNA (gRNA). The gRNA follows simple Watson-Crick base pairing, offering an easier and more versatile way at targeting the genome. In addition, it has been shown that CRISPR-Cas attains higher levels of gene editing compared to TALEN in human cells.⁵⁰ Gene multiplexing has also been more accessible through the use of multiple gRNAs,⁵⁰⁻⁵¹ making CRISPR-Cas a widely adopted tool for gene editing.

1.4 Discovery of CRISPR-Cas

1.4.1 Adaptive Immune Mechanism in Bacteria

In 1987, when Nakata *et al.* were studying the *iap* enzyme, they reported the observation of an unusual set of 29 nt palindromic repeats, flanked by five intervening 32 nt nonrepetitive sequences., downstream of the *iap* gene.⁵² This was unlike other repetitive elements observed at the time, which normally took form of tandem repeats. Unfortunately, due to insufficient DNA sequence data, the biological function of those repeats could not be determined. Over the next ten years advancements in Sanger DNA sequencing resulted in several microbial genomes being sequenced. More and

more of these repeat elements were reported from various bacterial and archaeal strains and eventually it was concluded that these repeat elements were present in more than 40% of bacteria and 90% of archaea.⁵³

Ongoing investigations in the field led to coining the acronym CRISPR (**C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats) to describe the unusual repeat array and also led to the identification of CRISPR-associated (Cas) genes (Figure 1.10).⁵⁴ The Cas genes were well conserved and located adjacent to the repeat elements.⁵⁴ Sequence homology analysis showed that Cas genes encoded the enzyme helicase, responsible for unwinding DNA, and a nuclease, which cleaves DNA, suggesting they might be involved in DNA replication and gene maintenance.⁵⁵⁻

56

In 2005, a key discovery was made when analysing the 'spacer' sequences that separated the individual repeats. Those spacer sequences were identical to sequences found in bacteriophages and conjugative plasmids.⁵⁷⁻⁵⁹ Interestingly, bacteria or archaea carrying those spacer sequences were resistant to the corresponding viruses, but viral infection was successful in cells lacking the spacer. These observations led to speculations that the CRISPR spacer was involved in cell immunity and served as a defence mechanism against phage infection.

Despite several hypotheses being raised, there was no experimental evidence that linked CRISPR and cell immunity until Barrangou *et al.*⁶⁰ reported a nucleic acid based immune response. Specificity was dictated by the CRISPR spacers through complementary base pairing with the homologous sequence of the viral genome and

Cas enzymes were responsible for defence and invasion by the spacer in *Streptococcus thermophilus*.

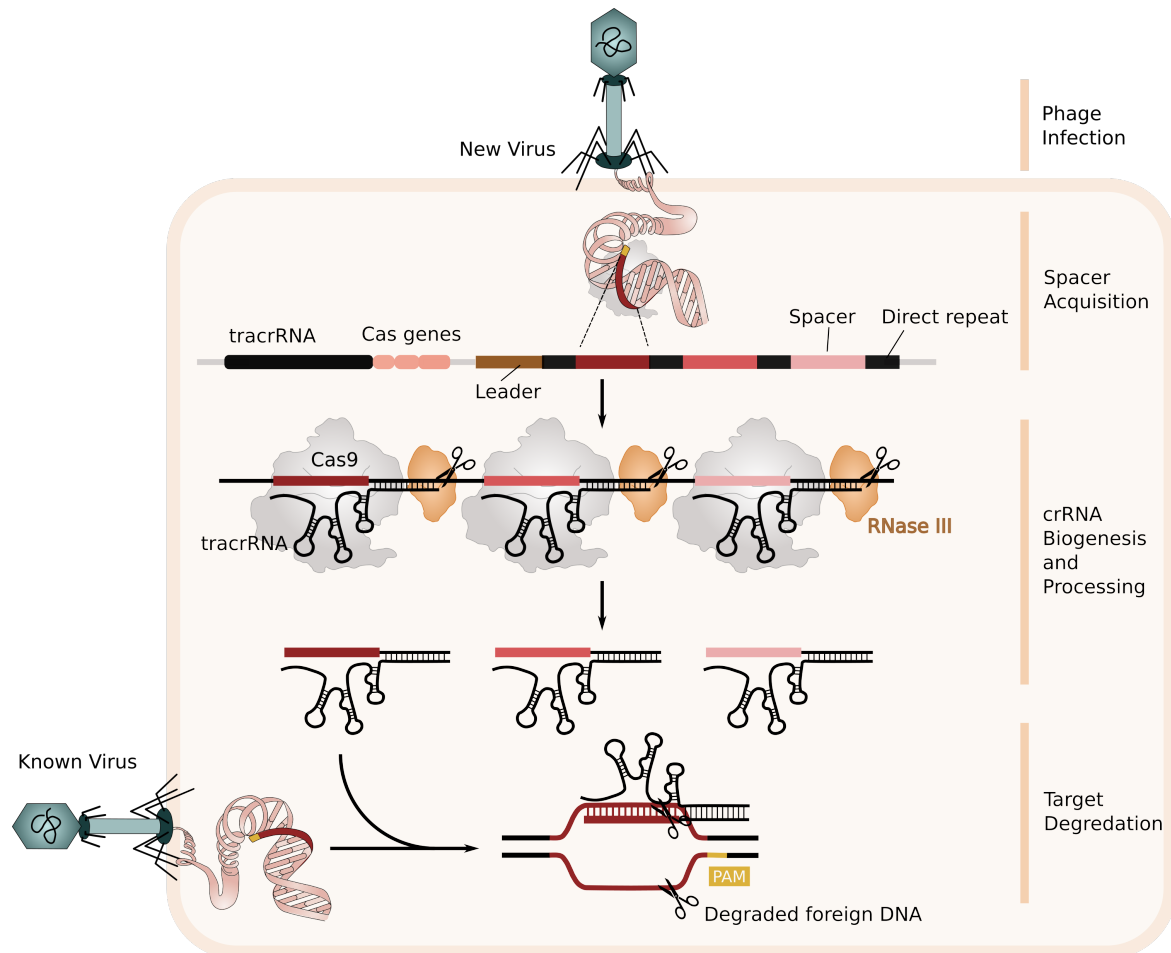


Figure 1.10 | CRISPR-Cas9 adaptive immunity in bacteria. Upon phage infection when exposed to a new virus, a series of events occur including spacer acquisition, crRNA biogenesis and processing and eventually target degradation when exposed to the same virus again.

This led to a series of experiment to elucidate the mechanism by which CRISPR works to provide immunity. It was found that firstly, the spacer sequences in the CRISPR locus are transcribed and converted to short CRISPR RNAs (crRNAs), each containing a spacer sequence to guide the Cas enzyme.⁵⁵ Another important observation made was the presence of a short motif, called the protospacer-adjacent motif (PAM), located next to the spacer on the target DNA and the absence of which

hindered all activity of CRISPR-Cas.^{59, 61} Thirdly, the DNA of the invading virus was cleaved by Cas enzymes in a double stranded manner and was site specific within the region of the spacer. Finally, in Cas9 dependent CRISPR systems, a dual RNA hybrid is formed involving crRNA and a non-coding *trans*-activating crRNA (tracrRNA).⁶² Following hybridisation of the tracrRNA with pre-crRNA transcripts on the CRISPR locus, Cas9 and endogenous RNase III process the CRISPR array into mature crRNAs, capable of digesting the dsDNA of invading viruses to inhibit infection. The system described is known as the type II CRISPR system, constituting of Cas9 enzyme, tracrRNA and the mature crRNA. Other CRISPR systems were also discovered simultaneously but the type II system stood out owing to its simplicity – it relies on a single Cas9 protein rather than a set of Cas proteins.⁶²⁻⁶³

1.4.2 Classification of CRISPR-Cas systems

CRISPR-Cas systems can be divided into two classes:⁶⁴ class 1 which requires multiple Cas proteins for crRNA processing and target cleavage and class 2 which only requires a single multidomain protein, for example Cas9 (Figure 1.11).^{49, 55, 65-66} These two classes are further divided into subtypes: Type I, III and IV in class 1 and II, V and VI in class 2, depending on the exact proteins involved in each functional module.

The target of type I, II and V effector nucleases is dsDNA. Target recognition is dictated by the crRNA and the presence of a PAM site.⁶⁷⁻⁷⁰ Type III nucleases on the other hand cleaves both ssDNA and RNA irrespective of a PAM motif being present (Table 1.1).⁷¹⁻⁷² In contrast to nucleases with a PAM dependency, Type VI effector nucleases instead recognise a motif called protospacer-flank site (PFS) to identify the

target. They also target ssRNA with the PFS located adjacent to the spacer sequence on the RNA strand.⁷³

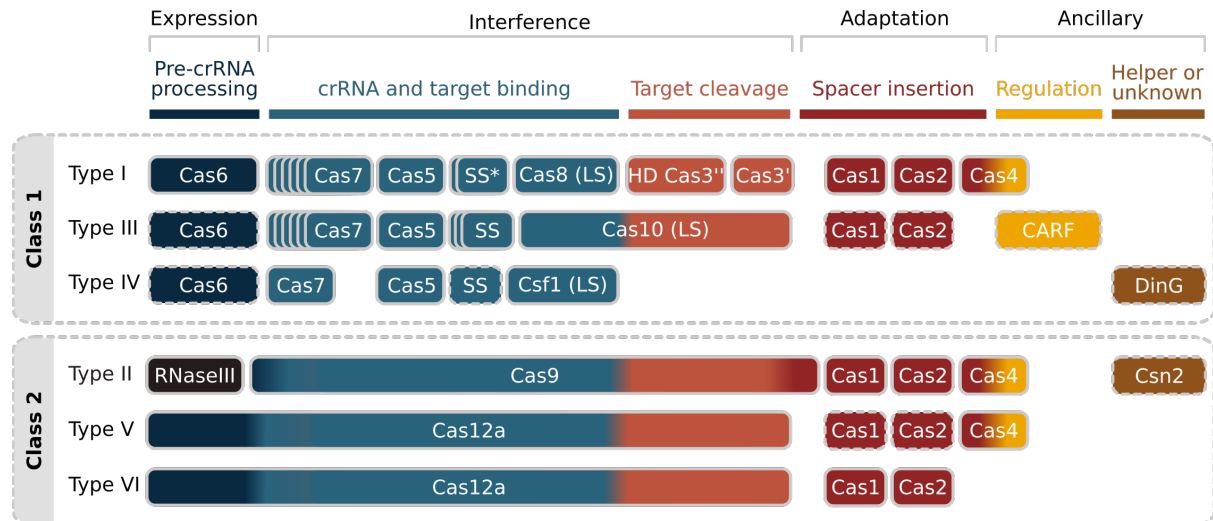


Figure 1.11 | Classification of CRISPR-Cas systems. SS stands for small subunit while LS stands for large subunit. SS* means that the putative small subunit is fused to the Cas8 large subunit for type I. Cas3 is a fusion of the nuclease HD Cas3'' and the helicase Cas3' which are encoded by two distinct genes. Cas10 and Csf1 forms part of the type III system LS and type IV system LS respectively. Dash lines indicate dispensable components. Cas6 with a dashed line for type III indicates that the protein is provided by other CRISPR-Cas loci. The two colours/three colours merge reflect that the proteins contribute to different stages of the CRISPR-Cas system.⁶⁴

Cleavage of the substrate occurs at specific sites for type I, II, III and V effector nucleases compared to type VI (Cas13a) where after the target RNA is specifically digested, non-specific degradation of surrounding RNAs also take place.⁷³⁻⁷⁴

The gRNA for the CRISPR-Cas9 (type II) system consists of a 40-mer crRNA, with a 20-mer spacer sequence for target specificity and an 86-mer tracrRNA, which binds to part of the crRNA by complementary base pairing to form a functional gRNA. The rest of the CRISPR subtypes are guided to their target by employing a single crRNA varying from 40-58 nt, with around 20-nt as the spacer for hybridisation with the target.

Class 2 CRISPR-systems are ideally suited for gene editing due to the need for only a single protein during target recognition and cleavage. Among the three subtypes, Cas9 in type II is the most widely adopted nuclease as it targets DNA and for its desirable 3-nt PAM sequence.

Table 1.1 | Characteristics of the different CRISPR-Cas systems.

	Type	Effector nuclease	gRNA (nt)		Target	PAM	Cleavage mechanism
			Spacer	All			
Class 1	I	Cascade-Cas3	32	61	dsDNA	5'-ATG-3'	Non-target ssDNA then 3'-5' exonucleolytic activity
	III	Csm/Cmr-Cas10 complex	39	45	ssRNA/ssDNA	-	ssRNA (transcript)-6-nt fragments; Non-template (transcription) ssDNA cleavage coupled to ssRNA cleavage
	IV	Csf1 complex	Uncharacterised				
Class 2	II	Cas9	20	> 100	dsDNA	5'-NGG-3'	Specific blunt-ended DSB 3bp after 5' of PAM
	V	Cas12a	20	42	dsDNA	5'-TTTN-3'	Specific DSB with 5' 5-nt overhangs distal to PAM
	VI	Cas13a	28	59	ssRNA	3'-A, U, or C	Specific RNA cleavage then non-specific RNase activity

1.4.3 CRISPR-Cas9 in Gene Editing

The idea of using CRISPR-Cas9 in the field of genome editing first arose when Siksnys *et al.*⁷⁵ demonstrated that the system was transferrable. This means that it was possible to transplant the type II CRISPR from one bacterium to another (*Streptococcus thermophilus* to *Escherichia coli*) without any loss of function. Crucial biochemical characterisation later on showed that DNA cleavage, guided by crRNA

can be carried out *in vitro* in the presence of purified Cas9 from *Streptococcus thermophilus* or *Streptococcus pyogenes*.^{49, 66} Furthermore, it came to light that the system could be made simpler by fusing the crRNA and the tracrRNA into a chimeric single guide RNA (sgRNA).⁴⁹ Mature crRNA-tracrRNA and sgRNAs were both shown to successfully accomplish gene editing in mammalian cells in 2013.^{50, 76} Since then, a plethora of structural, biophysical and biochemical studies have been carried out and today the molecular mechanism of DNA cleavage by CRISPR-cas9 is well documented.^{7, 77-81}

1.5 Structure of Cas9 enzyme

The multidomain DNA endonuclease *Streptococcus pyogenes* Cas9 (SpyCas9) is made up of 1368 amino acids.⁸² It has two distinct nuclease domains: An HNH-like domain which cleaves the target DNA strand and a RuvC-like domain which cleaves the non-target strand. The target DNA can be any 20-nt long DNA sequence as long as the sequence is unique compared to other genes to avoid off-target cleavage and as long as the target is present immediately adjacent to a PAM site. Cas9 cleaves DNA 3 bp upstream of the PAM site and also helps crRNA maturation and spacer acquisition.^{49, 66, 83-84} These processes are discussed in further detail below.

1.5.1 Bilobed Structure of Cas9

In the unbound state (apo-Cas9), two distinct lobes can be observed in the structure of Cas9. The first lobe is the alpha-helical recognition (REC) lobe, consisting of three alpha-helical domains which are responsible for the binding of DNA double helix (Figure 1.12a). The second lobe is the nuclease (NUC) lobe, consisting of the HNH

and the RuvC domains as well as a C-terminal domain (CTD).⁷⁸ The CTD is responsible for PAM recognition and interestingly in apo-Cas9, the PAM interacting sites are in a disordered state which means that the enzyme is unable to recognise the target DNA before binding of the gRNA. Experimental evidence of this includes DNA curtains assays (an *in-vitro* assay used to study protein-nucleic acid interactions through single-molecule visualisation) where apo-Cas9 has been shown to bind to DNA non-specifically but once the gRNA is introduced, the enzyme rapidly detaches from non-specific sites.⁸⁵ Crystallography data have also shown that Cas9 will adopt a catalytically inactive conformation until it binds to the gRNA, after which it will undergo structural activation for DNA recognition and cleavage (Figure 1.12b-c).⁸² This structural observation is further supported by biochemical findings that in the absence of gRNA, Cas9 are inactive as nucleases, making Cas9 an attractive candidate as an RNA-guided endonuclease.^{49, 78}

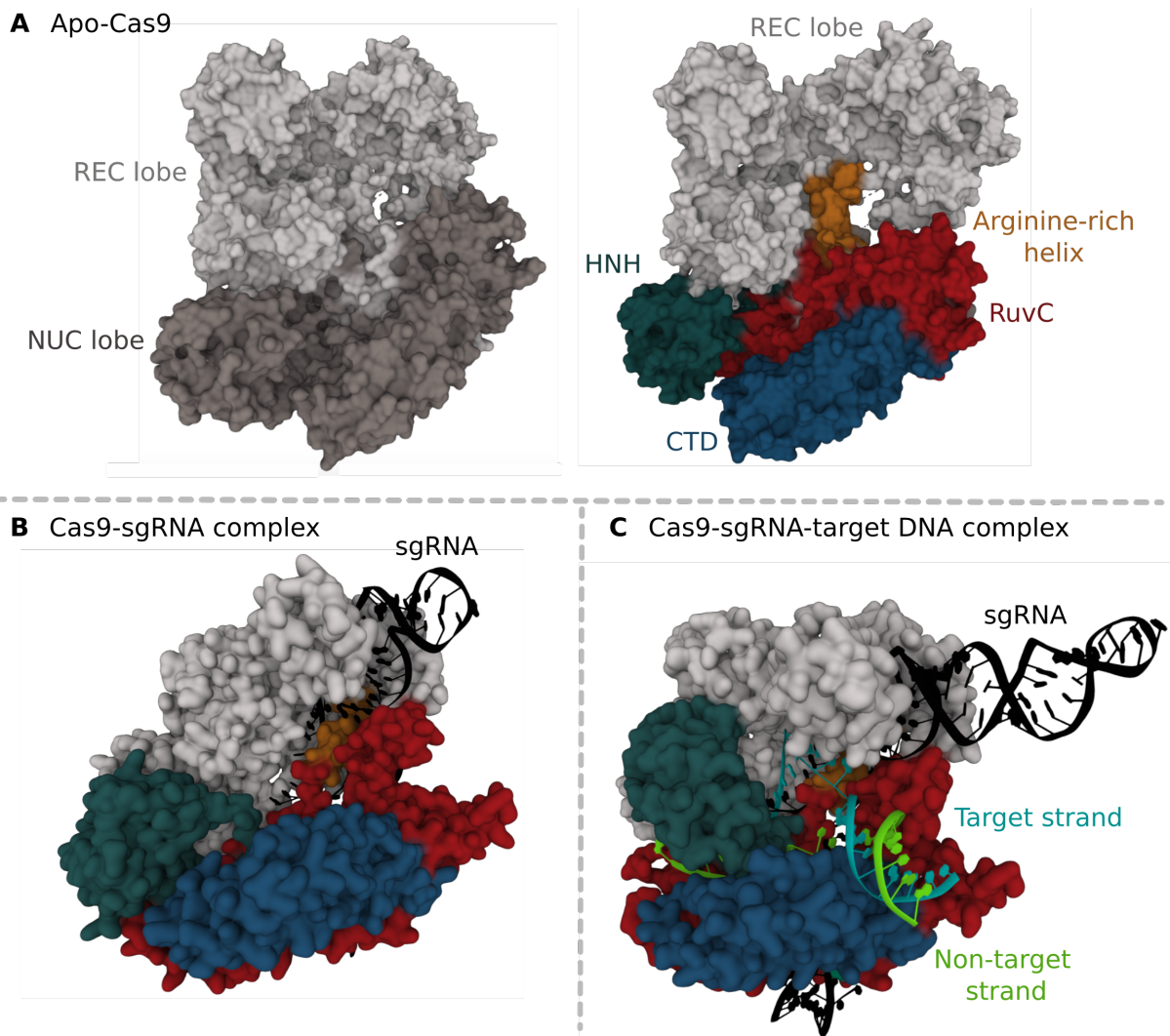


Figure 1.12 | 3D visualisation of Cas9. (A) shows Cas9 in the apo-state and shows the different parts that constitute the enzyme (PDB: 4CMP). (B) Cas9 undergoes a conformational change upon binding to the sgRNA (PDB: 4ZT0). (C) Cas9 undergoes a final conformational change after binding to the target DNA and is now “active” (PDB: 5F9R).

1.5.2 HNH and RuvC Nuclease Domains

The comparison of Cas9 to other DNA binding nucleases have revealed that the RuvC domain is similar to the one found in retroviral integrase, with a characteristic RNase H fold. This suggested that cleavage by RuvC proceeds *via* a two-metal ion catalytic mechanism on the non-target strand.⁷⁷⁻⁷⁸ On the contrary, the HNH domain is characterised by a $\beta\beta\alpha$ -metal fold, suggesting that cleavage occurs through a one-metal ion mechanism on the target strand.⁸² Mutating either of the two domains

converts the enzyme into a nickase and mutating both domains renders the enzyme catalytically inactive, called deadCas9 (dCas9) (further discussed in section 1.7.3 and 1.7.4).⁴⁹

1.6 DNA Targeting and Cleavage by CRISPR-Cas9

DNA cleavage by CRISPR-Cas9 is a multi-step process where the gRNA first needs to assemble onto the Cas9, then the ribonucleoprotein needs to recognise its target by finding the PAM, which is followed by strand invasion and the formation of gRNA-DNA duplex. Lastly, dsDNA is cleaved by Cas9.

1.6.1 Interactions with the gRNA

As discussed in section 1.5.1, it is crucial for the gRNA to bind to Cas9 for the latter to undergo a substantial structural rearrangement and change from an inactive state to an active endonuclease. Most of the structural changes occur in the REC lobe, which moves about 65 Å closer to the HNH domains.^{7, 82} Most of the conformational changes occur when the gRNA is loaded onto Cas9. Binding of the target DNA and PAM sequence results in significantly less conformational change, confirming that gRNA loading is a key regulator of Cas9 function.⁷⁸

The structure of the gRNA contains several motifs that interact with and are essential to the functioning of Cas9. This includes the anti-direct repeat where the crRNA and tracrRNA hybridize, as well as three stem loops formed by the tracrRNA. When the gRNA binds to Cas9, it makes numerous interactions including direct contacts with the stem loop 1, the repeat-anti-repeat duplex and the linker region between stem loop 1

and 2 (Figure 1.13). In contrast, considerably less contacts are made with stem loop 2 and stem loop 3.^{77, 80} Biochemical studies have shown that sgRNA without stem loop 2, stem loop 3 and the linker region are still able to cleave DNA but less efficiently.⁴⁹ However, removing stem loop 1 from the sgRNA completely abolishes cleavage. In addition, functional studies *in vivo* have demonstrated the need for stem loop 2 and 3 for Cas9 to perform at its best capacity.⁷⁷ Therefore, these observations suggest that the stem loop 1 and repeat-anti-repeat duplex are a vital part of the sgRNA, without which Cas9-sgRNA complex formation is impaired. On the other hand, while stem loop 2 and stem loop 3 are not as essential, they might still be required for stabilising the gRNA during complex formation and hence enhancing catalytic efficiency *in vivo*.^{50, 76-77, 86-88}

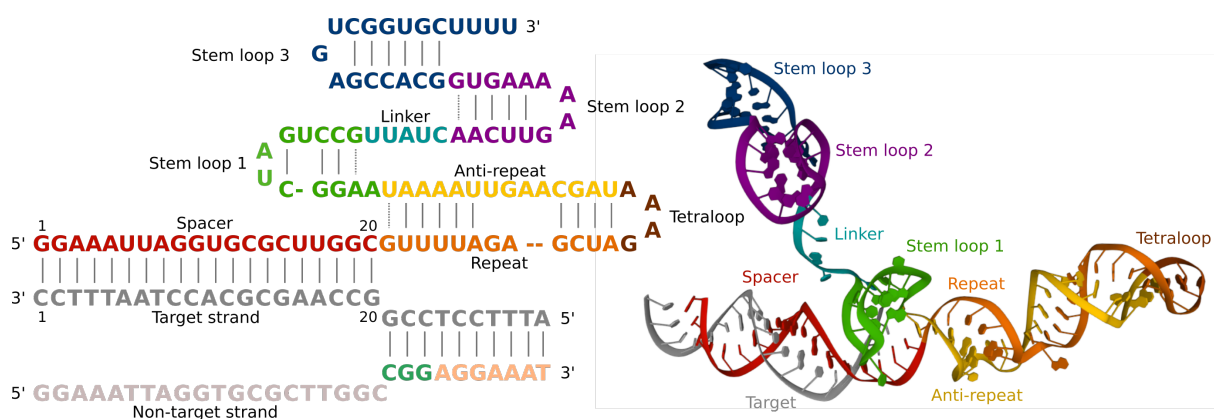


Figure 1.13 | Sequence and 3D representation of the sgRNA highlighting the different parts. The spacer of the sgRNA is shown in red, the repeat-anti-repeat in orange, tetraloop in brown, stem loop 1 in green, linker in teal, stem loop 2 in purple and stem loop 3 in blue (PDB: 5F9R).

1.6.2 Target Search and PAM Recognition

Once the ribonucleoprotein, consisting of the gRNA and the Cas9, is formed, the complex begins searching for its target.⁷ Searching and recognising the target depends on Watson-Crick base pairing with the 20 nt spacer on the gRNA and the presence of the conserved PAM sequence located next to the target site.^{49, 66}

The PAM site is vital for microorganisms to differentiate between self (in the CRISPR array) and non-self (viral) DNA.⁸⁹ Mutations in the PAM sequence is not tolerated *in vitro* and will not trigger an immune response in microorganisms.⁴⁹ For SpyCas9, the PAM sequence recognised is 5'-NGG-3', where N is any of the four bases.

Single molecule experiments demonstrated that Cas9 initially probes for the PAM sequence in the target DNA and after finding a match, it starts analysing the flanking DNA for RNA complementarity.⁸⁵ This process takes place through three-dimensional collisions where Cas9 rapidly releases DNA lacking the PAM sequence and dwells in time to verify if target DNA is complementary to the gRNA.^{85, 90-91}

There are two arginine residues R1333 and R1335 are responsible for PAM recognition motif of Cas9 (Figure 1.14). They form hydrogen bonds with the two guanines in the 5'-NGG-3' PAM sequence of the non-target DNA strand. Since N does not form any interactions with the PAM recognition motif of Cas9, it can be any of the four bases. Once the Cas9-gRNA complex finds the correct target, it triggers local DNA melting immediately upstream of the PAM followed by RNA strand invasion to form an RNA-DNA heteroduplex. This series of events leads to the displacement of the non-target DNA strand resulting in three-stranded structure known as an R-loop formation.^{85, 92}

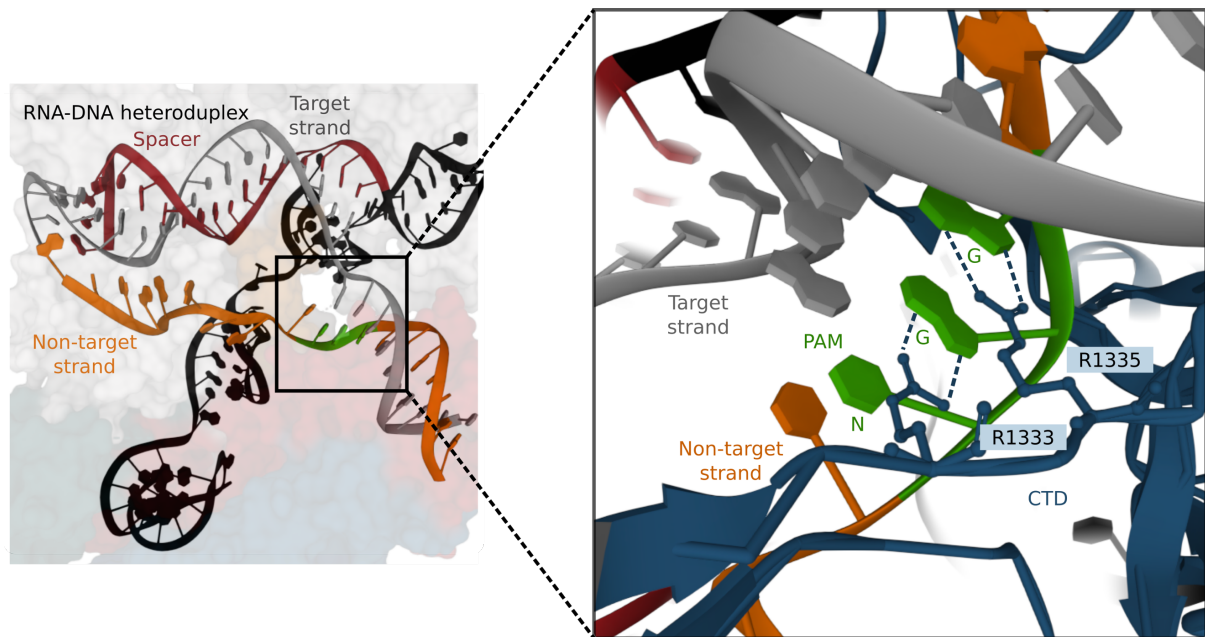


Figure 1.14 | Detailed view of the PAM dependence of Cas9. The close-up view shows interactions between residue R1333 and R1335 of the CTD of Cas9 and the two guanines of the in the 5'-NGG-3' PAM motif (PDB: 5F9R).

1.6.3 Local DNA Melting and R-loop Formation

Upon recognition of the PAM sequence, a sharp kink is formed in the target strand immediately upstream of the PAM (Figure 1.14). The phosphodiester group between the kink and the PAM (known as +1 phosphate) is stabilised through interactions with the K1107 and S1109 residues in Cas9, forming a phosphate lock loop.⁷⁹ (Figure 1.15). The formation of the kink is very important to allow the target strand DNA to transition from pairing with the non-target strand to pairing with the gRNA as it allows the target DNA to flip towards the crRNA for RNA-DNA heteroduplex formation. Biochemical and single molecule studies have shown that the interactions between the phosphate lock loop and the +1 phosphate triggers local DNA melting, allowing RNA strand invasion and stabilisation of the RNA-DNA duplex (A-form like structure).^{49, 79, 85}

The non-target strand is also displaced and forms another sharp kink at the –1 phosphate and extensive interactions occur between Cas9 and flipped bases at positions –2 and –3. A second kink is observed at position –4 on the non-target strand, which then extrudes through the positively charged tunnel between the RuvC and HNH domains. The PAM-proximal region is stabilised by base stacking onto the PAM duplex, resulting in the R-loop formation.⁸⁰ Biochemical studies with 2 bp mismatches introduced immediately adjacent to the PAM abolished Cas9 activity completely. However, the introduction of a small DNA bubble on the target DNA led to efficient Cas9 binding and cleavage without the need for RNA-DNA heteroduplex formation.⁸⁵

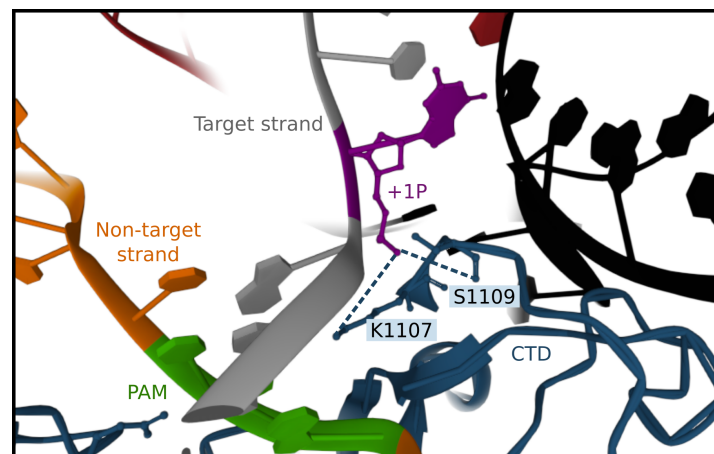


Figure 1.15 | Detailed visualisation of the phosphate lock loop. Once the PAM sequence is recognised, the phosphate upstream the PAM motif interacts with residues K1107 and S1109 in the CTD domain of Cas9, triggering local DNA melting and allowing RNA strand invasion (PDB: 5F9R).

1.6.4 DNA Cleavage

gRNA binding and RNA-DNA heteroduplex formation activates Cas9 for DNA cleavage. The two nuclease domains are employed to each cut one strand of the target DNA specifically 3 bp upstream from the NGG PAM sequence, resulting in a blunt-ended DSB.^{49, 66} However, Cas9-gRNA binding to DNA does not always trigger cleavage. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) has shown that sgRNA with as little as a 5-nt spacer can cause off-target DNA binding.

However, these sites are not cleaved by Cas9.⁹² This is because the active conformation of Cas9 is stabilised specifically *via* interactions between the 14th–17th bases of the gRNA spacer and the target DNA.⁹³ In addition, sgRNAs with spacers that are 17-18 nt in length have been found to significantly reduce off-target binding and cleavage.⁸⁹

1.7 Modified Cas9 and its Applications

1.7.1 Cas9 Orthologs with Varying PAM Sequences

A key limitation for using of CRISPR-Cas9 in gene editing is the requirement for a PAM sequence to be present at the target site. Naturally occurring Cas9 orthologs with different PAM requirements can be obtained from various species such as *Streptococcus thermophilus* CRISPR1 (St1), *Streptococcus thermophilus* CRISPR3 (St3), *Neisseria meningitidis* (Nm), *Francisella novicida* (Fn), *Staphylococcus aureus* (Sa), *Campylobacter jejuni* (Cj) and so on (Species with respective PAMs summarised in Table 1.2). However, the simple and least restrictive 5'-NGG-3' PAM sequence of SpCas9 made it a popular choice for gene editing. Therefore, attempting to engineer SpCas9 that recognise non-canonical PAM had been widely explored. For Cas9 to recognise a different PAM sequence, the PAM-recognition motif found in the PAM-interacting (PI) subdomain within the CTD domain must be modified. Single mutations in the PAM-recognition motif has been found to inhibit DNA cleavage but mutating the neighbouring Wedge (WED) subdomain further within the same CTD domain will restore Cas9 cleavage ability and alter the PAM requirements.⁹⁴⁻⁹⁵ There have been several SpCas9 variants reported, including VQR (D1135V/R1335Q/T1337R), EQR (D1135E/R1335Q/T1337R) and VRER ((D1135V/G1218R/R1335E/T1337R)) that recognise the PAM sequences 5'-NGAN/NGNG-3', 5'-NGAG-3' and 5'-NGCG-3'

respectively.⁹⁵ Crystal structures of the SpCas9 mutants have shown that altered sequences are recognised by an induced fit mechanism. The mutations introduced do not change the overall structure of the PI motif but instead enables minute distortions in the DNA backbone, which allows interactions between the altered amino acid residues and the new PAM sequence.⁹⁴ More recently, phage-assisted continuous evolution (PACE) has been used to mutate SpCas9 with the broadest PAM requirements, including NG, GAA and GAT, greatly increasing the scope of gene editing by CRISPR-Cas9.⁹⁶

Mutations in other species have also been carried out such as in SaCas9. The PAM requirement for wild-type SaCas9 is NNGRRN (R = A or G). Functional studies at target sites with several nucleotide combinations at positions 3–6 in the PAM sequence generated a group of chimeric Cas9 (cCas9), capable of recognising the PAM sequences NNVERRN, NNVACT, NNVAATG, NNVAATT, NNVGCT, NNVGTTG, and NNVGTT. The cCas9 have made almost a quarter of all genomic targets accessible in mammalian cells.⁹⁷ In addition, Chatterjee *et al.* reported a naturally occurring Cas9 endonuclease from *Streptococcus canis* (ScCas9) which had not been previously characterised. Interestingly, ScCas9 recognises an even more relaxed PAM sequence compared to SpCas9: 5'-NNG-3'. The off-targets assays conducted on ScCas9 also showed better results than SpCas9. However, more development is required for the broader application of ScCas9 as it was only more efficient than SpCas9 in non-5'-NNG-3' targets while SpCas9 remains more efficient in editing 5'-NNG-3' targets.⁹⁸

Table 1.2 | Characterised PAMs for Cas orthologs.

Cas9 System	Organism	PAM (5'-3')
SpCas9	<i>Streptococcus pyogenes</i>	NGG
FnCas9	<i>Francisella novicida</i>	NG
SaCas9	<i>Staphylococcus aureus</i>	NNGRRT
NmCas9	<i>Neisseria meningitidis</i>	NNNNGMTT
St1Cas9	<i>Streptococcus thermophilus CRISPR1</i>	NNAGAAW
St3Cas9	<i>Streptococcus thermophilus CRISPR3</i>	NGGNG
CjCas9	<i>Campylobacter jejuni</i>	NNNNACAC

1.7.2 High-fidelity Cas9

Even though SpCas9 is highly efficient at editing the genome, it can also cause off-target mutations, which can have drastic effects on organisms. It has been proposed that off-target cleavage occurs because the SpCas9-sgRNA complex has more energy than necessary for optimum recognition of its correct DNA target, therefore causing cleavage at mismatched sites as well.⁹⁹ The crystal structure of SpCas9-sgRNA-target DNA complex shows several interactions between the Cas9 and DNA, including hydrogen bonds between four Cas9 residues (N497, R661, Q695, Q926) and the phosphodiester backbone of the DNA.¹⁰⁰ To be able to use CRISPR-Cas9 in therapeutics, off-target editing needs to be eliminated and a way of doing so is to disrupt the hydrogen bonds so that the energy of the SpCas9-sgRNA is decreased but the complex still retains its ability for robust on-target activity. It has been reported that substituting all four Cas9 residues for alanine (N497A/R661A/Q695A/Q926A) results in undetectable off-target activities by genome-wide break capture. The new Cas9

variant possessed on-target activity comparable to wild-type SpCas9 and thereafter named SpCas9-HF1 for high-fidelity variant number 1.¹⁰⁰ However, it has recently been reported that when SpCas-HF1 is delivered as a ribonucleoprotein (RNP), which is the most therapeutically relevant method,¹⁰¹⁻¹⁰³ it has reduced on-target activity.¹⁰⁴ In an attempt to find mutations that would reduce off-targets of SpCas9 but retain its on-target activities as an RNP, a bacterial screen was carried out. This led to the identification of a single point mutation p.R691A, which when introduced in Cas9 (high-fidelity (HiFi) Cas9) mediates robust correction of the sickle cell disease and reduces off-target effects by approximately 20-fold compared to wild-type SpCas9.¹⁰⁴

1.7.3 Cas9 Nickases

It has been reported that up to five mismatches are tolerated in the sgRNA sequence, which can result in hundreds of potential off-target mutations.¹⁰⁵ In an attempt to curb off target effects, a Cas9 nickase (Cas9n) mutant was generated by substituting key catalytic residues with alanine in either the RuvC domain (D10A) or the HNH domain (N840A). Cas9n will only produce single-stranded nicks, which can be repaired by the endogenous base-excision repair pathway.¹⁰⁶⁻¹⁰⁷ Moreover, to improve efficiency a dual sgRNA system can be employed, generating double nicks on opposite strands and creating a 3' or 5' overhang along the target (Figure 1.16).^{49, 108} This system has been reported to reduce off-target activity by about 50 to 1500-fold in cell lines.¹⁰⁸

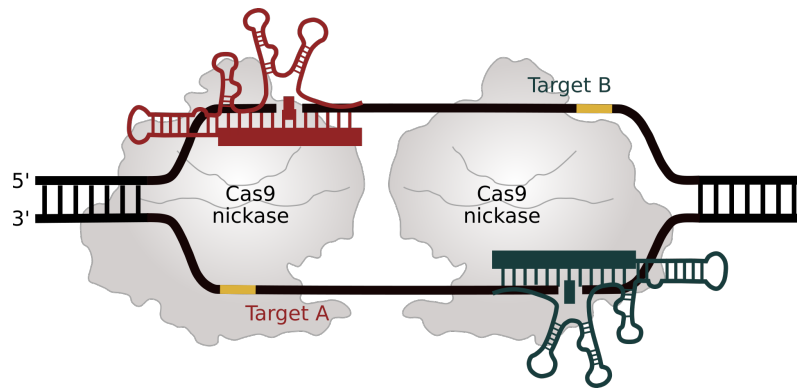


Figure 1.16 | Double nicks using Cas9n. Dual sgRNA system that generates 3' or 5' overhangs using two Cas9n and with a PAM dependency on both strands.

1.7.4 Catalytically Inactive Cas9 (dead Cas9 or dCas9)

Mutating both the RuvC and HNH domains in Cas9 (D10A and H840A) results in a catalytically inactive Cas9, called dCas9, which lacks nuclease activity but retains its ability to bind with the gRNA and target DNA specifically. dCas9 expanded the scope of CRISPR beyond gene editing.¹⁰⁹ CRISPR/dCas9 systems have been used for sequence specific gene regulation, by fusing dCas9 to transcriptional activators such as VP64 as or transcriptional repressors such as Krüppel-associated box (KRAB).^{69, 110-113} Similarly, CRISPR/dCas9 can be used for epigenome editing by fusing dCas9 with epigenetic modifier proteins such as histone demethylase, DNA demethylase and histone acetyltransferase.¹¹⁴⁻¹¹⁵ An affinity-tagged dCas9 can also be used in ChIP assays to study endogenous protein-DNA interactions at specific locations. After reaching its target, the tagged dCas9 can be used to pull down proteins associated with that region and mass spectrometry can then be used to identify those proteins.¹¹⁶⁻¹¹⁷ In addition, dCas9 fused with fluorescent proteins can be used for live cell genomic imaging at specific loci, allowing the study of genomic organisation and dynamics in living cell. This application is explored in further details in Chapter 2.

1.8 Chemical Modifications of the sgRNA

Unmodified RNAs are subject to rapid cellular degradation due to the 2'-hydroxyl group and therefore protecting RNA chemically to be able to achieve efficient editing *in vivo* has become very desirable. Sugar modifications such as 2'-O-methyl RNA (2'-OMe-RNA) and 2'-Fluoro RNA (2'-F-RNA) have been shown to considerably reduce susceptibility to nucleases.¹¹⁸⁻¹²⁰ Moreover, backbone modifications, through the inclusion of thioester groups such as phosphorothioate (PS) groups or in combination with 2'-F or 2'-OMe such as 2'-OMe-phosphorothioate (MS), 2'-F phosphorothioate (FS) and 2'-OMe-thiophosphonoacetate (MSP) groups have also been employed to provide resistance against nucleases *in vivo*.¹¹⁸⁻¹²¹ The sgRNA can tolerate up to 70% of 2'-OMe or 2'-F sugar modifications without losing its genome editing ability.¹¹⁹ By studying in details the residues in Cas9 that interact with the sgRNA, the regions of the sgRNA that can be modified have been identified and shown in Figure 1.17. These studies have shown that modified sgRNA or gRNA are not only more efficient but also more specific than their unmodified counterpart.^{118, 122-123}

As well as incorporating chemical modification in the sgRNA, a partial DNA guided-Cas9 has also been developed. Compared to RNA monomers, DNA nucleotides are 10-fold cheaper and chemically more stable, making them an attractive substitution for ribonucleotides in the gRNA. To perform efficiently *in vitro*, the crRNA only requires five RNA bases (37 out of 42 can be DNA bases) and 62% of the tracrRNA can be coded as DNA.¹²⁴ However, these high contents of DNA hinder cleavage efficiency *in vivo*, presumably because of kinetic destabilisation of the Cas9-gRNA-target DNA complex in cells.¹²⁴ Reducing the number of DNA nucleotides to less than 12 DNA

nucleotides at the 5' end or less than 4 DNA nucleotides at the 3' end of the crRNA begins to restore Cas9 guided cleavage activity in cells.¹²⁵

Other sugar modifications involving the formation of a second ring with from the 2'-carbon to the 3'-carbon have also been explored, such as the 2',4'-constrained ethyl bicyclic nucleic acids (2',4'-cEt) can be included in the crRNA, strengthening its interaction with the tracrRNA and enhancing Cas9 cleavage activity.^{120, 122} Locked nucleic acids (LNAs) and next generation *N*-methyl substituted bridged nucleic acids (2',4'-BNA^{NC}[N-Me]) have also been incorporated in the crRNA and interestingly, they have been found to improve cleavage specificity both *in vitro* and *in cellulo*.¹²⁶ The increased specificity is due to BNAs and LNAs impairing the formation of the stable “zipped” conformation when hybridised to an off-target DNA sequence.¹²⁶

Furthermore, to increase the functionality of CRISPR/Cas9, chemical handles have also been conjugated at the 5' and 3'-end of the crRNA.¹²⁷⁻¹²⁸ Synthetic DNA to enhance HDR have been attached onto the crRNA as well as chemical molecules such as fluorophores, which is further discussed in the upcoming chapters.

1.9 Synthesis of the gRNA

1.9.1 *In vitro* transcription

RNA is conveniently synthesized by *in vitro* transcription using the components of bacteriophage systems. An RNA polymerase and a template with a promoter sequence are the two simple components required to synthesise transcripts, with sizes ranging from less than 30 nt to those above 10^4 nt and in scales from μg to mg amounts.¹²⁹ The most commonly employed systems are the T3, T7, and SP6 systems. During the course of this work, the T7 system, derived from the T7 phage of *E. coli* was used and will be detailed out below.

T7-based RNA amplification was first introduced by Van Gelder, Eberwine, and his co-workers in 1990 and is now one of the most widely used amplification protocol.¹³⁰ To amplify the RNA sequence of interest, the target sequence is placed downstream from the T7 promoter site. The minimum promoter sequence needed for efficient transcription is from -17 to $+6$, with the $+1$ base being the first base incorporated into RNA during transcription. Therefore, there is not complete freedom of choice of the sequence at the 5'-end of the transcribed RNA. Most T7 promoters, like class III promoters, have G's at $+1$, $+2$, and $+3$, and the first two G's are critical for transcriptional yield.¹³¹⁻¹³²

The template for *in vitro* transcription can either be a linearised plasmid DNA that contains the T7 promoter sequence, or it can be PCR products, with the T7 promoter as part of the 5'-oligonucleotide primer in the PCR reaction or generally, it can be any highly pure DNA, with a T7 promoter sequence consisting of two annealed

oligonucleotides (only the T7 promoter part of the template needs to be double-stranded) (Figure 1.18).¹²⁹

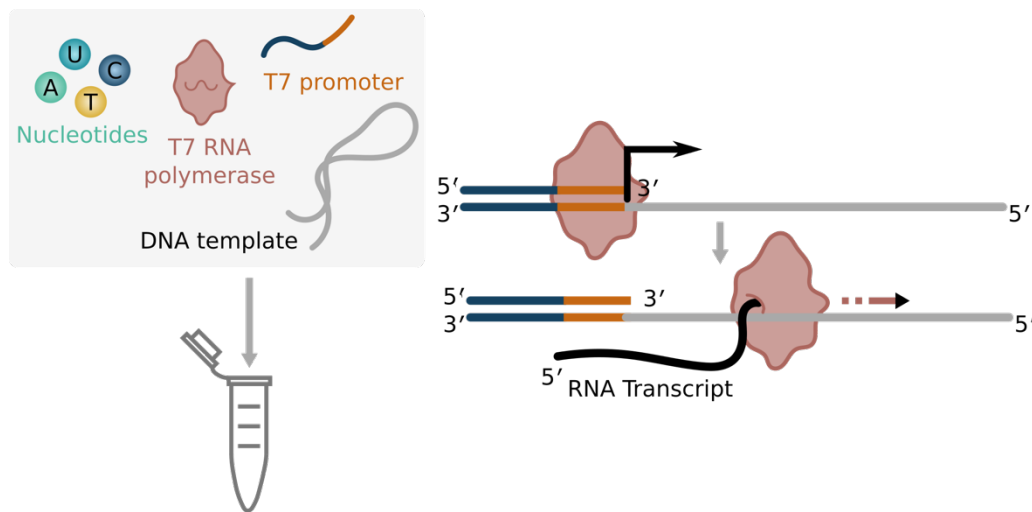


Figure 1.18 | Schematic representation of the generation of sgRNA by *in vitro* transcription. sgRNA can easily be made by mixing the DNA template, the promoter sequence, nucleotides and bacteriophage promoters, followed by incubation at 37 °C for 2–16 h.

T7 RNA polymerase is added to the mixture, along with the NTPs and the polymerase binds to the promoter site and starts transcribing by adding complementary nucleotides and making the phosphodiester backbone. The termination step *in vitro* occurs when the polymerase falls off at the very end of the template. The average rate of *in vitro* transcription is 200–260 nt/s and the frequency error about 6×10^{-6} .¹³³ This means large amounts of RNA can be generated quickly at an acceptable purity for CRISPR applications. The resulting RNA is very easily purified by adding DNase to digest the template, proteinase K to digest the polymerase (and DNase) and desalting to get rid of salts and digested reactants. The resulting solution contains pure RNA, which is ready to use. The main drawback of using *in vitro* transcription is the inability to introduce modifications specifically into the RNA.

1.9.2 Solid-phase synthesis

The gRNA can alternatively be made by using automated solid-phase synthesis, a process that allows the four different RNA monomers to be added stepwise from the 3' to 5' direction (Figure 1.19).¹³⁴ As opposed to *in vitro* transcription, solid-phase synthesis does not require a template and modified nucleotides can easily be incorporated during the synthesis. The process involves a cycle consisting of 1) activation and coupling, 2) capping, 3) oxidation and 4) detritylation, incorporating one RNA monomer at a time.

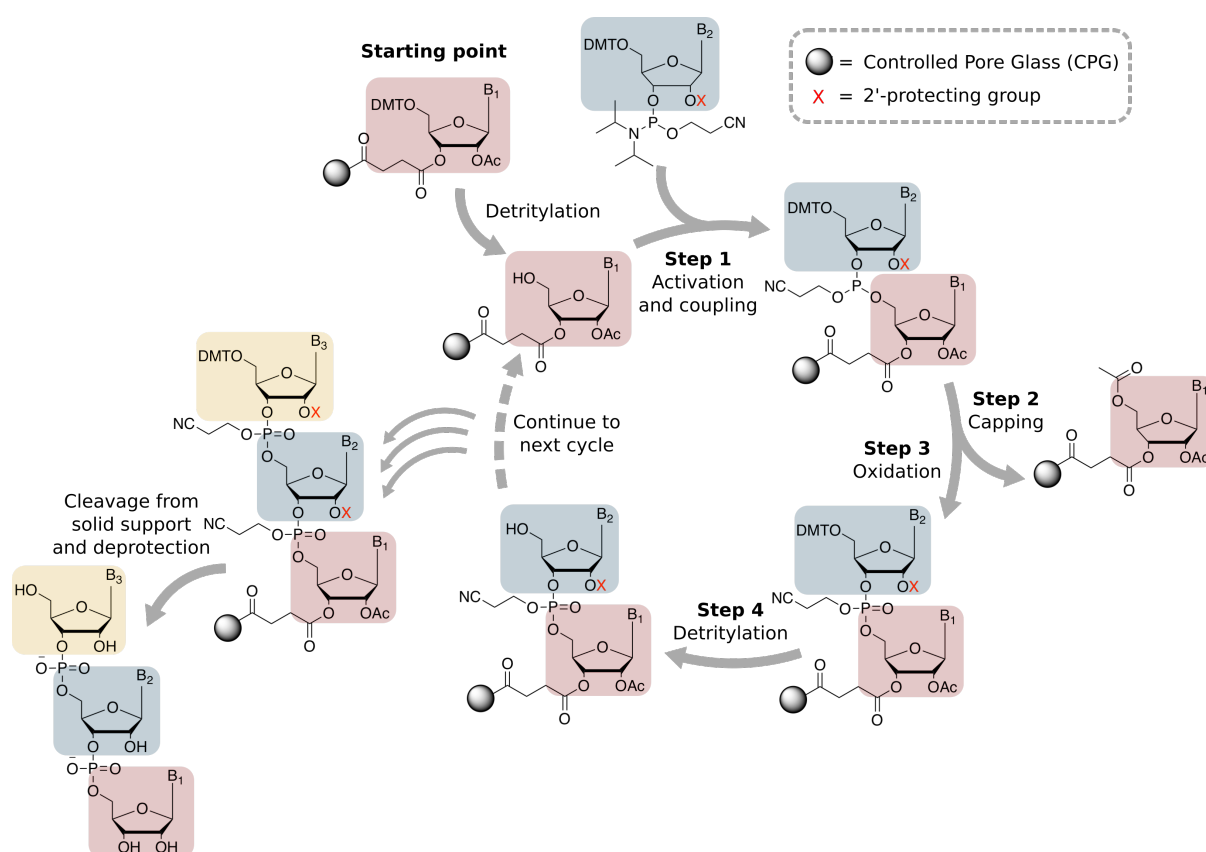


Figure 1.19 | Solid-phase synthesis of RNA.

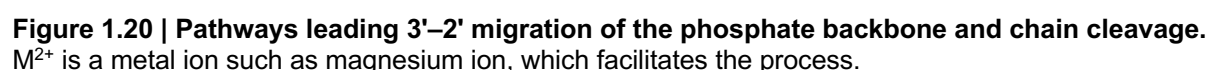
The first step of the synthesis cycle is removing the 5'-DMT protecting group (detritylation) from a nucleoside attached to a controlled pore glass (CPG) solid support by treating it with dichloroacetic acid or another weak acid. Detritylation exposes the 5'-OH, which is ready to react with the next base. The appropriate nucleoside phosphoramidite is mixed with an activator (tetrazole or a derivative) to

convert the diisopropylamino group of the nucleoside phosphoramidite into a better leaving group. Next, the activated nucleoside phosphoramidite is coupled with the CPG bound nucleoside, forming a phosphite triester linkage (P(III)).

Any unreacted 5'-OH is capped by using two capping solutions: acetic anhydride and *N*-methylimidazole (NMI). If left untreated, any unreacted 5'-OH will participate in the next coupling step, reacting with the incoming phosphoramidite. The resulting oligonucleotide would lack one base and would correspond to a deletion mutation of the desired oligo. Deletion mutations will accumulate after each successive cycle, leading to a complex mixture of several incorrect oligonucleotides.

The phosphite triester linkage (P(III)) is then converted to phosphate triester (P(V)) *via* iodine mediated oxidation. This is done prior to the next TCA detritylation reaction as P(III) is unstable to acid. This cycle of steps is repeated until the desired oligonucleotide is obtained. At the end of the synthesis, the oligonucleotide is cleaved from the solid support with concentrated ammonium hydroxide at room temperature. The RNA is then deprotected and purified.

Solid-phase synthesis of RNA is based on solid-phase synthesis of DNA, with the exception of the 2'-hydroxyl group on the ribose sugar, which has to be protected to prevent migration from 3'–2' of the phosphate backbone or base-mediated 2'-OH attack of the phosphate backbone during deprotection, resulting in degradation (Figure 1.20).



DMTO Base

NC-CH₂-O-P(=O)(N(CH₃)₂)-O-Si(CH₃)₂-C(CH₃)₃

TBDMS Protecting group

DMTO Base

NC-CH₂-O-P(=O)(N(CH₃)₂)-O-Si(CH₃)₂-C(CH₃)₃

TC Protecting group

DMTO Base

NC-CH₂-O-P(=O)(N(CH₃)₂)-O-Si(CH₃)₂-C(CH₃)₃

ACE Protecting group

DMTO Base

NC-CH₂-O-P(=O)(N(CH₃)₂)-O-Si(CH₃)₂-C(CH₃)₃

TOM Protecting group

R¹ = or

TBDMS protected phosphoramidites are perhaps the most commonly employed ones due to their widespread commercial availability and their reasonable cost. The protecting group is selectively removed using a source of fluoride ion (e.g. triethylamine trihydrofluoride).¹³⁵ However, the TBDMS group is not 100% stable when exposed to the basic conditions used to remove the protecting groups on the bases (tert-butylphenoxyacetyl protecting groups). Losing the TBDMS group at that stage

can lead to phosphodiester chain cleavage or 3'–2' phosphate migration. Moreover, steric hinderance due to the bulky 2'-group reduces coupling efficiency compared to DNA synthesis. Considering all the above factors, the length of RNA that can be obtained in good yields and good purity is limited to approximately 70 nucleotides.¹³⁶

The TC protecting group is becoming increasingly popular in the synthesis of RNA. Although the monomers are more expensive, the TC group can be removed simultaneously with the other protecting nucleobase protecting groups using ethylenediamine (EDA), allowing a one-step deprotection of all protecting groups.¹³⁷ This greatly simplifies the workflow. Moreover, it was observed that RNA with a free 2'-hydroxyl is remarkably resistant to degradation during treatment with certain amine bases (such as EDA) under nonaqueous conditions. The cleaved and deprotected RNA remains insoluble and adsorbed to the solid-phase matrix and is released from the solid support by simply washing with a few drops of RNase-free water.

While both methods of RNA synthesis (*in vitro* transcription and solid phase synthesis) are promising ways to make RNA more accessible, they both have limitations. Several factors need to be taken into account, for example whether the RNA is modified or the length or sequence of the RNA, before choosing the method of synthesis. The pros and cons associated with each method are summarised in Table 1.3.

Table 1.3 | Pros and Cons of Solid Phase RNA synthesis compared to *in vitro* transcription.

Pros	Cons
Solid Phase Synthesis	
Easy to incorporate chemical modifications	Costly
Can be made on large scale	Limited to <100-mers due to coupling efficiency
Efficient synthesis for <100-mers	Requires HPLC or PAGE purification
<i>In Vitro</i> Transcription	
No limit on length of RNA	Difficult to scale up
Easy purification process	Suffer from polymerase-bias
	Challenging/not possible to incorporate chemical modifications

1.10 Aims and Objectives

While extensive work has been carried out on modifying the Cas9 protein to widen the scope of CRISPR, there are limited studies where the gRNA is used to introduce modifications in the system. This is often the case due to the high costs associated with synthesising gRNA.

The work carried out in project aims to:

Chapter 2 | develop fluorescent guide RNAs which efficiently label genomic DNA, without interfering signals or background fluorescence.

Chapter 3 | improve modified sgRNA production *via* a simple and high yielding chemical ligation and testing the applications of the modified sgRNA *in cellulo*.

Chapter 4 | introduce modifications into the CRISPR-Cas9 system through a novel way involving the R-loop structure.

Chapter 2

Live Cell Imaging

using fluorescently modified gRNAs

2.1 Background

We have seen in Chapter 1 that the DNA double helix is complexly folded into several layers to form a chromosome, such that it can fit the limited space available in the cell nucleus. The spatial arrangement of the chromatin fibre highly affects the function of DNA and therefore, simply knowing the sequence of a gene is insufficient to determine its physiological function. In addition, nuclear processes such as transcription and replication take place at very specific spatially defined locations in the nucleus. The organisational properties of the genomes and the machineries that act on them create an elaborate architectural environment in which genomes must function. The mechanism and the series of actions that take place to enable them to do so remain one of the greatest challenges in modern cell biology.¹³⁸ Labelling the genomes and

visualisation of the chromosomal loci *in situ* is key to understanding many of these biological processes.

2.1.1 Genomic Imaging Using Traditional Approaches

Small molecule fluorophores such as DAPI and Hoechst 33342 have been extensively used to non-specifically stain DNA and generate fluorescence, creating a global image of DNA structure in cells.¹³⁹⁻¹⁴⁰ These dyes are cell permeable and they work by binding to the minor groove of double stranded DNA, with a preference for A-T rich regions. Binding to double stranded DNA significantly increases the dyes' quantum yield and they can be visualised by fluorescence microscopy. However, there are many other brighter fluorophores that do not possess DNA binding properties. In order to harness their fluorescent properties, these fluorophores have been attached to oligonucleotides that hybridise to DNA or RNA targets *via* Watson-Crick base pairing Figure 2.1.¹⁴¹⁻¹⁴² The labelled oligonucleotides are called Fluorescence *In Situ* Hybridisation (FISH) probes.

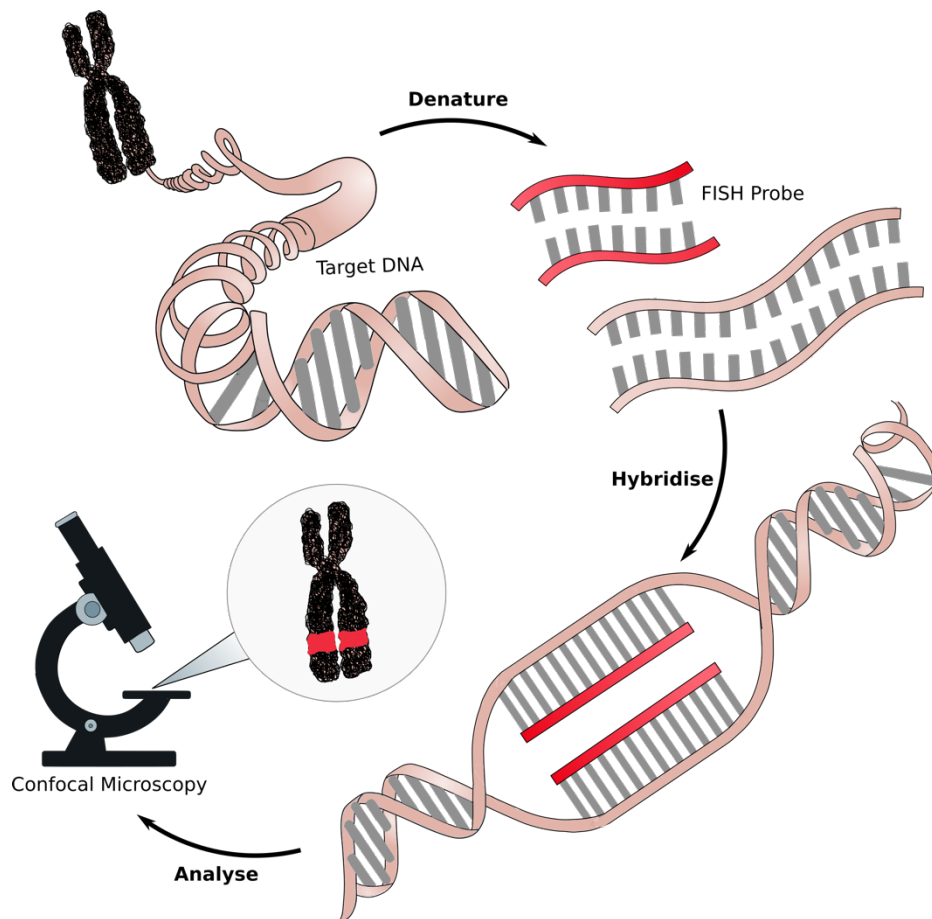


Figure 2.1 | Steps involved in Fluorescent *In Situ* Hybridisation (FISH). Target DNA is first isolated and then denatured using formaldehyde. FISH probe is added, which hybridises to the denatured DNA. Fluorescently labelled DNA is then viewed under a microscope. This method allows for detecting the target DNA but does not allow studying of the DNA in its dynamic environment.

FISH probes are used on ‘fixed’ cells, whereby all cellular activity is stopped but DNA structure and protein interactions to some extent remain intact. The double stranded DNA in cells is then partially denatured, allowing the fluorescent probes to hybridise. The unbound fluorescent probes are removed by washing the cells. This provides excellent spatial and temporal information on genome architecture. For example, FISH has enabled enhancer-promoter interactions in transcription activation to be studied and allowed the 24-colour karyotyping of human chromosomes for detection of chromosome aberrations.¹⁴³⁻¹⁴⁵ The downside to using FISH is that fixed cells are preserved in a static environment, preventing the study of real-time dynamics. The harsh conditions (formamide and heat) also run the risk of destroying the structural

integrity of the cell structures, questioning the technique's reliability. Live cell imaging on the other hand are representative of dynamic cells, that can move and change over time.

2.1.2 Live cell imaging using sequence specific methods

The main factor contributing to the success of live cell imaging is a high signal to background ratio. This is because in live cells, it is not possible to wash away excess probe. As a result, we need probes that can efficiently and specifically bind to its respective target. This means that highly repetitive genomic loci such as the centromeres or telomeres are an attractive target; a single DNA-binding protein can be labelled with a fluorescent protein (FP) to bind to the same region, enriching the fluorescence signal. Engineered DNA-binding modules such as zinc fingers and TALEs have been fused to fluorescent proteins and have successfully been employed in imaging repetitive loci.¹⁴⁶⁻¹⁴⁷ However, imaging of non-repetitive genes is incredibly challenging using these DNA binding modules. This is because several sequence specific DNA-binding proteins are required and engineering multiple zinc fingers or TALEs is not straightforward.

2.1.2.1 Live cell imaging using modified dCas9-CRISPR systems

Similar to genome editing, CRISPR-Cas9 provides a solution to this problem; various DNA sites can be targeted by using complementary sgRNAs and a single protein. The catalytically deactivated version of Cas9 (dCas9, D10A/H840A mutations) can be fused to green fluorescent protein (GFP) and guided by the sgRNA to specifically label either repetitive or non-repetitive genes in living human cells (Figure 2.2a).¹⁴⁸ However, this method results in accumulation of GFP in the nucleus and a high

nuclear background fluorescence is often observed, resulting in poor signal to background ratio.

It has also been reported that 26–36 different sgRNAs are required to detect a non-repetitive genomic locus using CRISPR.¹⁴⁸ The number of sgRNAs required can be significantly reduced by improving the signal to background ratio. In an attempt to do so, three GFP proteins have been fused to dCas9 (dCas9-3xGFP).¹⁴⁹ The problem of nucleolar accumulation of GFP still persisted despite having brighter signals. In addition, fusing more than three FPs can inhibit the efficiency of imaging due to the massive size of the dCas9-sgRNA complex and the difficulty of its production. Multi-colour imaging of different loci in the same cell was also established by fusing other fluorescent proteins, such as red or blue, to orthogonal variants of dCas9 from different bacterial species (Figure 2.2b).¹⁴⁹ Unfortunately, this method is limited to the number of bacterial dCas9 variants.

With this in mind, efforts were diverted into engineering dCas9 so that it can carry a repeating peptide array called SunTag; this way, multiple copies of GFP-tagged antibodies can be recruited (Figure 2.2c).¹⁵⁰ Nonetheless, it was later discovered that this method exhibited high background fluorescence.¹⁵¹ An alternate solution to this problem was the bimolecular fluorescence complementation (BiFC) strategy, whereby a FP is split into two parts and fluorescence can only occur once the two split segments are in close proximity. The BiFC system consisted of *Venus* C-terminal (FP) fused to SunTag-dCas9 and a gRNA modified with an aptamer that can recruit *Venus* N-terminal (Figure 2.2d). Once the two *Venus* subunits assembled, this method allows the signal to be considerably enriched by the SunTag while the background

fluorescence from spontaneously assembled *Venus* proteins that is diffusely localized in the nucleus would be considerably lower.¹⁵¹ However, the versatility of this technique was limited by the need of a FP that can be split.

Quantum dots (QDs) provided a promising alternative due to their highly bright and photostable nature. In fact, the intensity of fluorescent signal of a single QD is 20 times higher and 100 times more stable than a small molecule fluorophore.¹⁵² In a study in 2017, a dual-colour CRISPR system was developed, whereby a pair of QD-labelled dCas9 and sgRNA were successfully used to image Human Immunodeficiency Virus (HIV)-1 proviral DNA (Figure 2.2e).¹⁵³ However, QDs are limited by problems such as cell delivery due to their large size and they are also prone to aggregation and entrapment in lysosomes and endosomes.¹⁵²

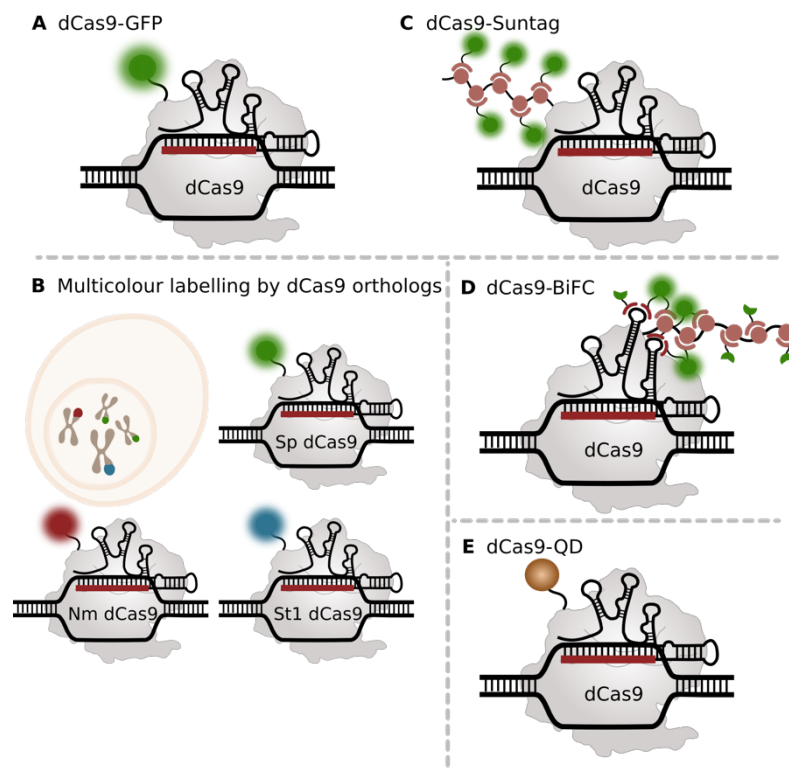


Figure 2.2 | Live cell imaging approaches using modified dCas9. (A) shows the first attempt using a dCas9-GFP fusion protein. (B) shows the incorporation of a Suntag array onto the dCas9, which binds to multiple GFP-tagged antibodies, amplifying the fluorescent signal (C) shows different species of dCas9 fused to different fluorescent proteins for multicolour labelling (D) shows the BiFC strategy where the fluorescent protein is split into two parts and fluorescence only occurs once the two split segments

are in close proximity, reducing background fluorescence (E) shows dCas9 conjugated to a quantum dot as an alternative to fluorescent protein, due to their bright and photostable nature

2.1.2.2 Live cell imaging using modified sgRNA-CRISPR systems

The main limitations of modifying the dCas9 was the growing size of the protein, making it challenging to deliver or to express. Therefore, attention was focused on modifying the sgRNA instead. As mentioned in Section 1.8, there are only parts of the sgRNA where modifications are tolerated.

A popular way of imaging using modified sgRNA is to include an aptamer in the design of the sgRNA.¹⁵⁴ In 2016, the sgRNA was modified to carry the ‘Broccoli’ aptamer, which binds to a small molecule fluorophore that is added to the medium and generates fluorescence (Figure 2.3a).⁹¹ The aptamer-based technique did not stop there as the need for imaging different genomic loci with different colours emerged. This would provide an opportunity to measure the distance between inter- and intrachromosomal loci, which would give valuable information on chromosome structure and dynamics.^{149, 155} A combination of different aptamer hairpins that recruit different fluorescent proteins can be included into the design of the sgRNA. The spectral overlapping of individual fluorescence signals generates new colours in addition to the primary colours (secondary colours: white = red + blue + green, cyan = blue + green, magenta = red + blue, yellow = green + red). This technique was named CRISPRainbow (Figure 2.3b).¹⁵⁵ Improvements on the technique led to the development of a more sensitive multicolour system known as CRISPR-Sirius. In CRISPR-Sirius, the sgRNA was designed to include thermostable octets of aptamers, which were optimised to prevent misfolding. This meant that a single CRISPR-Sirius-

gRNA could recruit 8 FPs compared to only 2 FPs in CRISPRainbow (Figure 2.3c).¹⁵⁶⁻

159

While these approaches are highly encouraging, their limitations cannot be ignored. sgRNA recruiting only 1 or two FPs tend to have low brightness while those recruiting several FPs have a tendency to cause aggregation, leading to non-specific labelling.¹⁵¹ Moreover, the size of the sgRNA is restrictive; one sgRNA fused to aptamers approaches 400 nt, which impedes their production in cells and also interferes with cell delivery.

While the use of FP laid a good foundation for genomic imaging, small molecule fluorophores offer the advantage of being very bright and highly photostable. Moreover, they can easily be conjugated on the sgRNA, greatly reducing its size.¹⁶⁰ The biggest concern when using chemical fluorophores conjugated to sgRNA was entrapment in lysosomes and endosomes, contributing to background fluorescence. To overcome this issue, fluorophores have been combined with quenchers to create a metastable system that is non-fluorescent outside the nucleus and fluorescent inside.

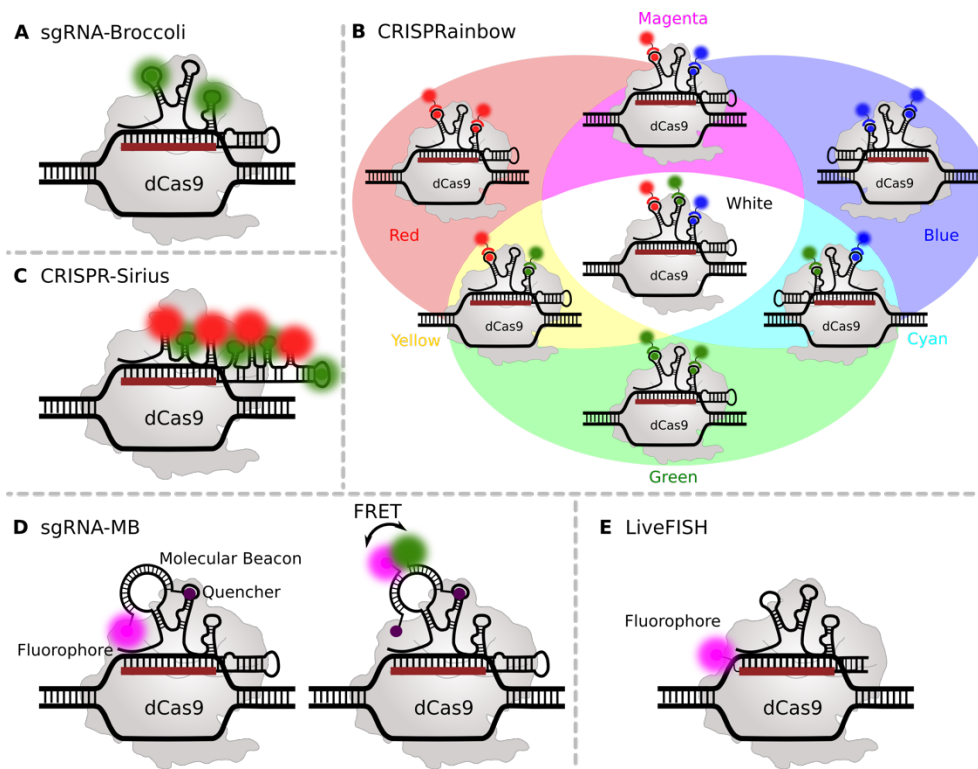


Figure 2.3 | Live cell imaging approaches using modified sgRNA. (A) shows the incorporation of Broccoli aptamers into sgRNA and their binding to fluorophores. (B) shows a technique described as CRISPRainbow where the fluorescent colour palette is extended to seven colours by combining two or three different aptamers on the sgRNA. (C) shows sgRNA designed to include thermostable octets of aptamers, capable of recruiting eight fluorescent proteins (D) shows a molecular beacon approach, combining the use of a fluorophore and a quencher to minimise background fluorescence (E) shows a technique known as LiveFISH, whereby a fluorescently tagged gRNA is electroporated into cells.

In the construct developed by Wu *et al.*,¹⁶¹ a molecular beacon (MB) is employed whereby a fluorophore and a quencher was incorporated on either side of the hairpin loop probe (Figure 2.3d). A sgRNA scaffold was also designed to harbour the unique sequence of the MB. Consequently, any unbound MB will be inactivated and when the MB hybridises to the sgRNA scaffold, the quencher separates from the fluorophore, resulting in an increase in fluorescence. While this study was used to image repetitive genomic sequences, recently, the group reported the use of a larger sgRNA scaffold, two MBs and FRET between the two MBs to image non repetitive genomic sequences (Figure 2.3d).¹⁶² However, the systems delivers the dCas9 and sgRNA scaffold using an expression vector delivery method, which requires excessive work, especially when

targeting several genes simultaneously. The size of the sgRNA (~170 nt) scaffold restricts the delivery by any other method. Moreover, the dual-FRET MB system cannot be used to visualise multiple genomic loci simultaneously as it entails two optically distinct MBs to elicit a FRET signal.

In 2019, Wang *et al.* reported their study, LiveFISH, while work was being carried out for this chapter (Figure 2.3e).¹⁶³ The study involved using a fluorophore conjugated to the gRNA and delivering the construct as a ribonucleoprotein (RNP) by electroporation to the cell. While repeating the study, we found several limitations such as low cell viability and changes in cell morphology after electroporation. Moreover, the study required high levels of dCas9 protein (modified with EGFP to verify target through co-localisation), which is considerably more expensive than using plasmid and lastly although there has been development in the area, electroporation *in vivo* is still very restrictive.¹⁶⁴

In 2022, Clow *et al.* developed casilio imaging, whereby the sgRNA is appended with multiple binding sites for Pumilio/FBF (PUF).¹²⁸ The modified sgRNA is capable of recruiting up to 15 or more fluorescent proteins (PUF-tethered fluorescent proteins) to a target as opposed to requiring target site tiling of dCas9-GFP direct fusions using 12 or more individual gRNAs in previous studies. They have successfully demonstrated live cell imaging of non-repetitive genomic loci with one guide RNA per locus. However, the problem associated with large RNA size and challenging transfection still prevails.

2.2 Hypotheses and Aims

We propose a novel chemical solution to improve live-cell DNA imaging results; sgRNAs can be sequence-specifically tagged with very bright, highly stable small molecule fluorophore(s), greatly reducing the size of the sgRNA, significantly increasing the number of colours available and enabling multiple loci to be imaged. In addition, we will only require a dCas9 plasmid, making the process both simpler and more cost effective. There have only been a handful of experiments in which the fluorophore was incorporated into gRNAs. They have either been used for FACS enrichment of transfected cells,¹²⁷ for fixed cell imaging (CASFISH) or they have been used in conjunction with Cas9 protein and electroporated into cells.^{163, 165} Our method will make use of a much simpler delivery method; transfection using cationic lipids. Not only will this highly reduce the cost of the experiment, it will also expand the applications of CRISPR mediated live cell imaging. The main challenge with using cationic lipids to deliver fluorescent gRNA is the tendency to 'get trapped' and accumulate in the endosomes and lysosomes, contributing to background fluorescence. Our unique approach will combine a fluorophore and a quencher on the gRNA to turn the fluorescence 'on/off' in an attempt to reduce background fluorescence.

2.3 RNase H activated fgRNA concept

Our main goal is to create a system that will lead to activation of fluorescence upon gRNA nuclear localisation. In doing so, we will be keeping the background fluorescence in the cytoplasm to a minimum and we will only have active fluorescence at the site of genomic labelling. The gRNA will have a fluorophore-linked 15-base

DNA-2'-OMe extension *via* a HEG (hexaethylene glycol) linker, which will hybridise to a quencher-linked 15-base RNA-2'-OMe oligo (Figure 2.4a). A stable RNA-DNA heteroduplex will be formed which will be susceptible to native cellular RNase H cleavage. Upon RNase H cleavage, the quencher will be released, and fluorescence will be restored (Figure 2.4b). It has been reported that RNase H is found primarily in the nucleus, hence only the DNA-RNA heteroduplex of the fgRNA that has been successfully delivered into the nucleus will be activated. FgRNA outside the nucleus will remain inactive and invisible (dimmer).¹⁴⁸ 2'-OMe ribonucleotides are incompatible with RNase H cleavage and thus, we are able to vary the number of 2'-OMe ribonucleotides to find the optimum window for efficient RNase H cleavage and melting of the cleaved strand with the quencher.¹⁶⁶ 2'-OMe ribonucleotides also improves cellular stability of oligos, preventing premature degradation *via* exo- or endonucleases.¹⁶⁷⁻¹⁶⁸

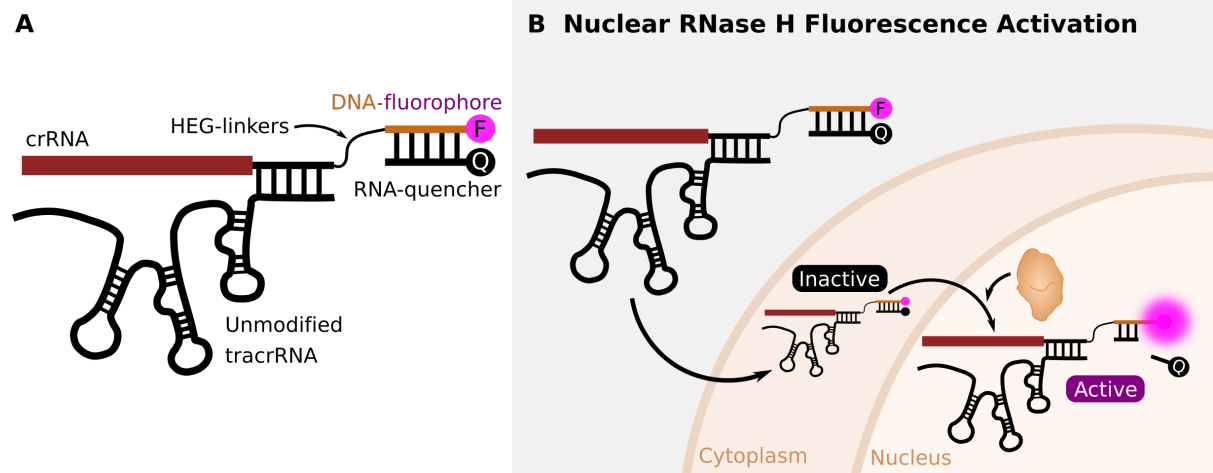


Figure 2.4 | Design of the fgRNA probe and schematic representation of RNase H activation. (A) fgRNA consisting of unmodified tracrRNA and crRNA conjugated to a small duplex via HEG (hexaethylene glycol) linker at the 3'-end. One strand of the duplex is all DNA bases and it is conjugated to a fluorophore. The complementary strand of the small duplex is made of RNA bases and it is conjugated to a quencher. (B) explains the RNase H activation process. In the cytoplasm, the fluorescence is being quenched by the quencher and the probe is inactive. Once in the nucleus, where RNase H is predominantly present, The RNA part of the small duplex containing the quencher is cleaved off and the probe is activated.

2.4 Fluorescent test probes

2.4.1 Design strategy and synthesis

In order to achieve the best results, the design of the fgRNA needs to be optimised. The fluorophore-quencher pair needs to be optimised and also the number and positions of 2'-OMe needs to be optimised so that we get efficient melting of the quencher bearing strand under physiological conditions after RNase H cleavage. However, the cost of synthesising a modified gRNA is quite high. As a result, three different 15-mer test probes were designed in order to optimise the fgRNA (Figure 2.5). The test probes were representative of the small DNA-RNA duplex attached to the gRNA *via* a HEG-linker. As shown in Figure 2.5a, the RNA complement design was standard, comprising of eleven RNA bases and four 2'-OMe sugars at the 5'-wing. The RNA was also designed with a 5'-amino hexyl for post-synthetic labelling with a quencher or BHQ-2 was incorporated directly. To vary the RNase H cleavage window, different number of 2'-OMe sugars were incorporated in the DNA strand. As well as controlling the RNase H cleavage window (RNase H will not cleave a 2'-OMe-RNA hybrid), the 2'-OMe sugars in the DNA strand also increases the stability of the probe in cells. All fifteen bases in probe 1 were unmodified DNA nucleotides. Probe 2 had seven 2'-OMe sugars at the 5'-end, and eight unmodified DNA nucleotides, providing a window of four bases for RNase H cleavage. Finally, Probe 3 had nine 2'-OMe sugars at the 5'-end and six unmodified DNA nucleotides, providing a window of only two bases for RNase H cleavage. The DNA strands were made with a 3'-amino group for post-synthetic labelling with fluorophores.

As well as finding the optimum RNase H cleavage window, finding the correct fluorophore quencher pair is very important for the probe to work efficiently. Ideally,

we would like the probe to have a very low fluorescence in the presence of the quencher (less than 1000 a.u.) and a very high fluorescence once treated with RNase H (5000 a.u. and above). Probe 1 was therefore labelled with different fluorophore-quencher pairs such as Cy3-BHQ-2, Alexa Fluor 594-BHQ-2, TAMRA-BHQ-1 and Alexa Fluor 488-BHQ-1 to screen the best one (Figure 2.5b).

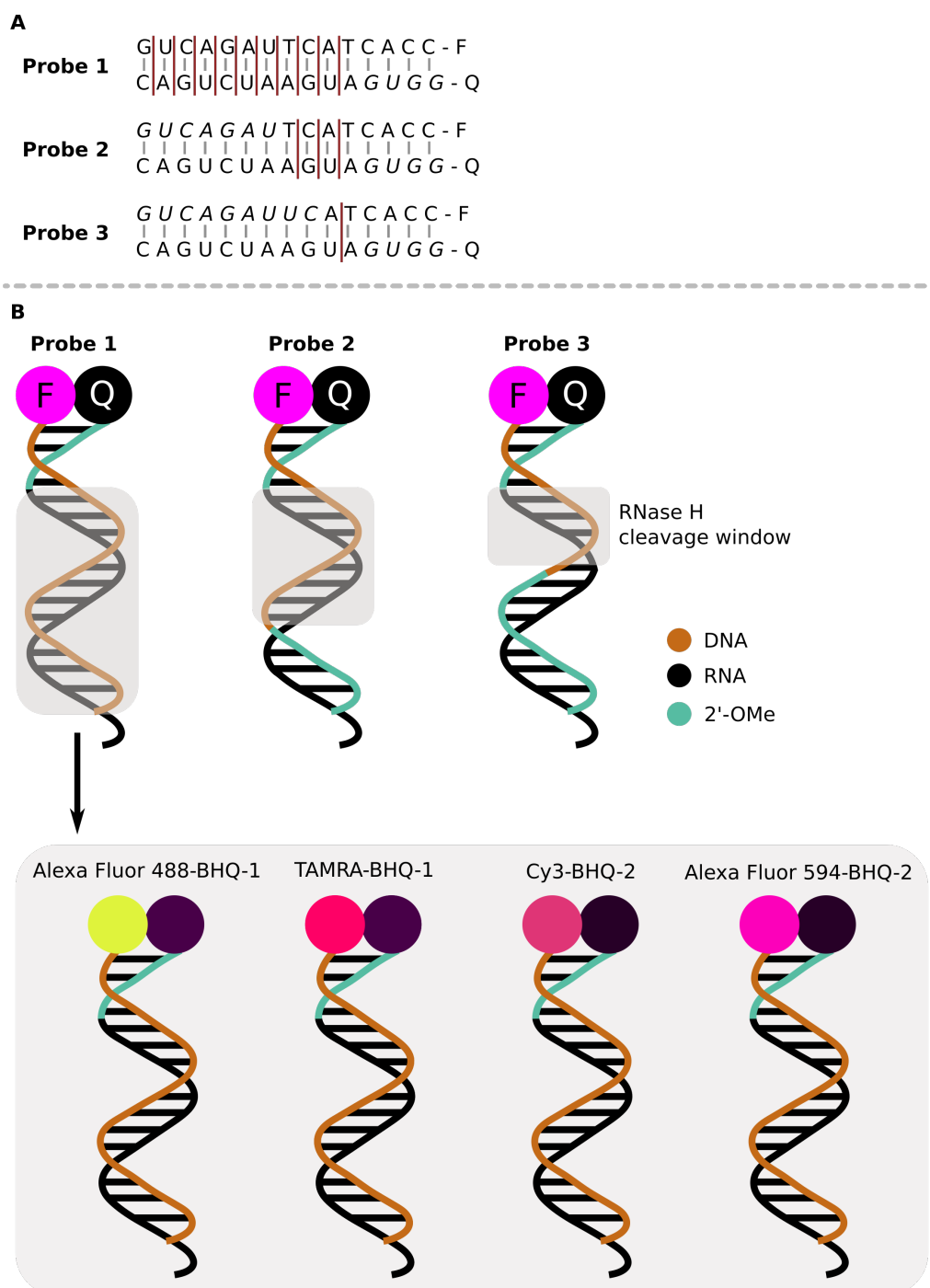


Figure 2.5 | Optimisation of the probe. (A) Sequence of probe 1–3. Red lines represent RNase H cleavable nucleotides. Italicised nucleotides are 2'-OMe. F is fluorophore and Q is quencher. (B) Design

of probe 1–3 with varying length of RNA-DNA heteroduplex region, where RNase H can effectively cut. The fluorophore-quencher pair was also optimised by testing probes labelled with Alexa Fluor 488 and BHQ-1, TAMRA and BHQ-1, Cy3 and BHQ-2 and Alexa Fluor 594 and BHQ-2. The oligonucleotide sequences and structures of the fluorophores and quenchers can be found in Appendix Table 1.

2.4.2 Fluorescence Studies

The fluorescence intensities of the four fluorophore-quencher pair were measured *in vitro*. From Figure 2.6, it is evident that the most effective fluorophore-quencher pair is Alexa Fluor 594-BHQ-2, with a quenched fluorescence of only 650 a.u and a rise in fluorescence of up to 10,500 a.u upon RNase H treatment (16-fold increase). Cy3-BHQ-2 in comparison have a much higher quenched fluorescence and lower rise in fluorescence when RNase H was added. Both probes with TAMRA-BHQ-1 and Alexa Fluor 488-BHQ-1 had very inefficient quenching and gave negligible rise in fluorescence in the presence of RNase H. One possibility accounting for the inefficient quenching could be contrary to BHQ-2, BHQ-1 quencher was incorporated into the probes by post synthetic labelling instead of coupling the phosphoramidite. Therefore, it could be that BHQ-1 degraded or was photobleached during handling, explaining the inefficient quenching. Hence, probe 2 and probe 3 were labelled with Alexa Fluor 594 and BHQ-2 and the fluorescence measurements of the three probes were compared (Figure 2.7).

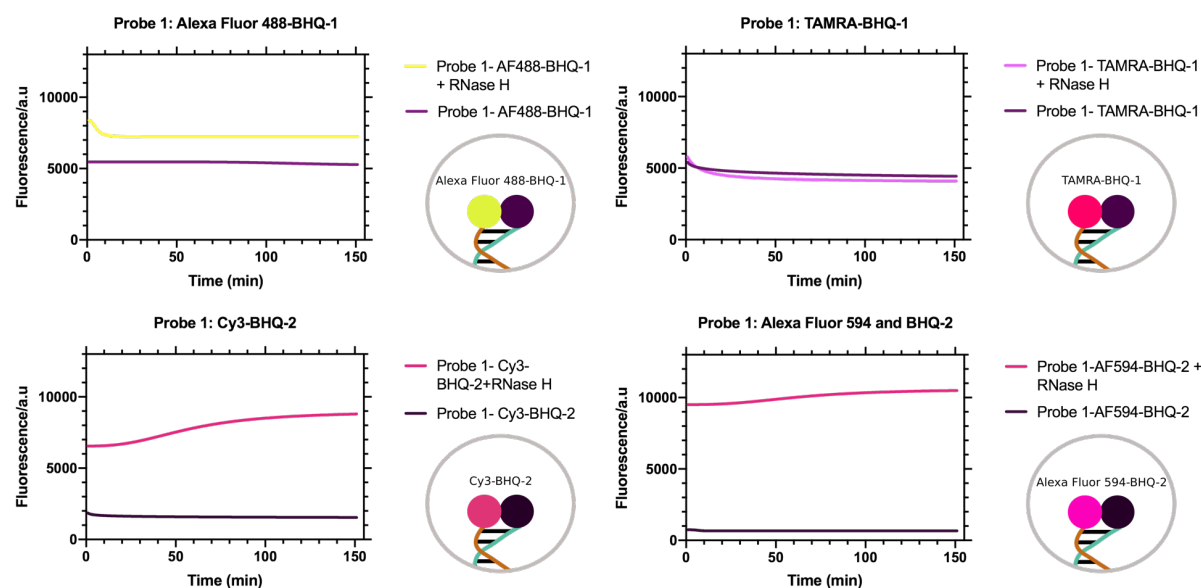


Figure 2.6 | Fluorescence measurements of Probe 1, labelled with different fluorophore-quencher pair. The differently labelled probes were resuspended in buffer containing 50 mM Tris-HCl, 7.5 mM KCl, 3 mM MgCl₂ and 10 mM DTT. RNase H was added to the mixture and the rise in fluorescence due to RNase H cleavage of the RNA-DNA heteroduplex was measured over time. The oligonucleotide sequences can be found in Appendix Table 1.

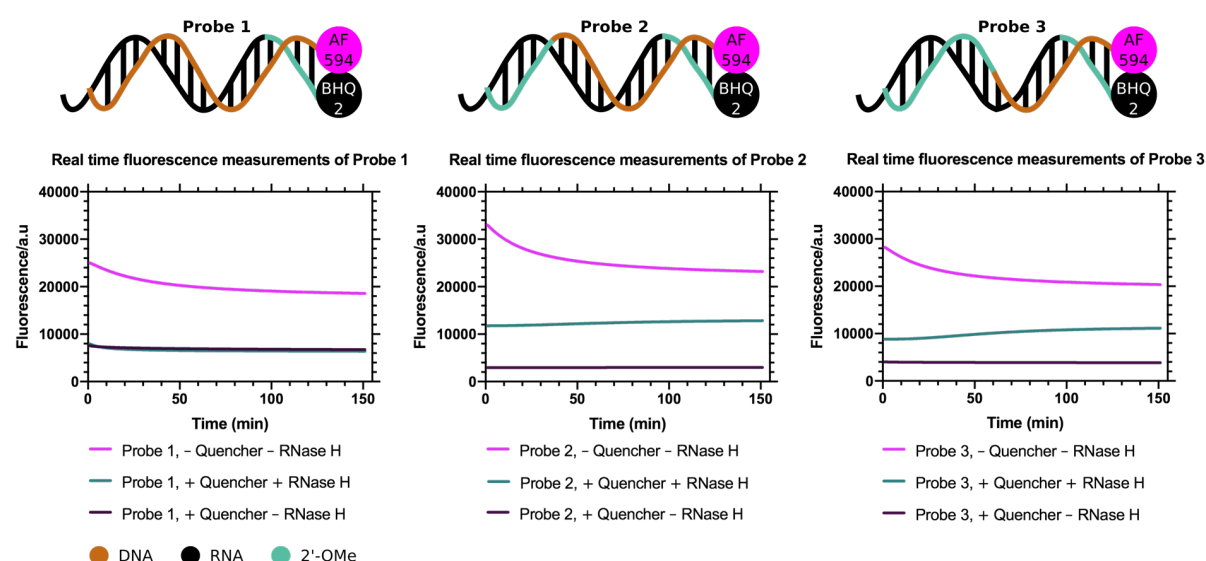


Figure 2.7 | Fluorescence measurements of Probe 1–3, with RNase H cleavage window. Probes 1–3 labelled with Alexa Fluor 594 and BHQ-2 were resuspended in buffer containing 50 mM Tris-HCl, 7.5 mM KCl, 3 mM MgCl₂ and 10 mM DTT. RNase H was added to the mixture and the rise in fluorescence due to RNase H cleavage of the RNA-DNA heteroduplex was measured over time. The oligonucleotide sequences can be found in Appendix Table 1.

Fluorescence measurements showed that probe 2 performed the best, with the lowest quenched fluorescence (2950 a.u) and highest rise in fluorescence in the presence of

RNase H (12801 a.u). Therefore, it was concluded from these results that probe 2 performed the best, with the lowest quenched fluorescence and highest rise in fluorescence in the presence of RNase H. While probe 2 only had two RNase H cleavable nucleotides, probe 1 had more than four. This means that a longer cleaved strand bearing the quencher might have formed. Generally, longer duplexes are more stable and therefore less likely to dissociate.^{5,8} This presumably means that the majority of the cleaved strand in probe 1 is still hybridised to its complementary strand and hence still quenched. In comparison to Probe 2, Probe 3 only had one RNase H cleavable nucleotide and the rise in fluorescence after treatment with RNase H is slightly less. The increase in fluorescence for probe 2 in the presence of RNase H was approximately 3.33-fold while that for probe 3 was 1.81-fold. This might be due to longer reaction times because of the presence of only one cleavage site. The difference in reaction time was analysed by treating probe 2 and 3 with RNase H and loading the samples onto a 20% denaturing PAGE at intervals of 15 s, 30 s, 1 min, 5 min, 15 min and 30 min (Figure 2.8). While the experiment proved that RNase H fully cleaved probe 2 faster than probe 3, it also showed that probe 3 was fully cleaved after 5 min. It is therefore unclear why the rise in fluorescence for probe 3 was slightly less than probe 2.

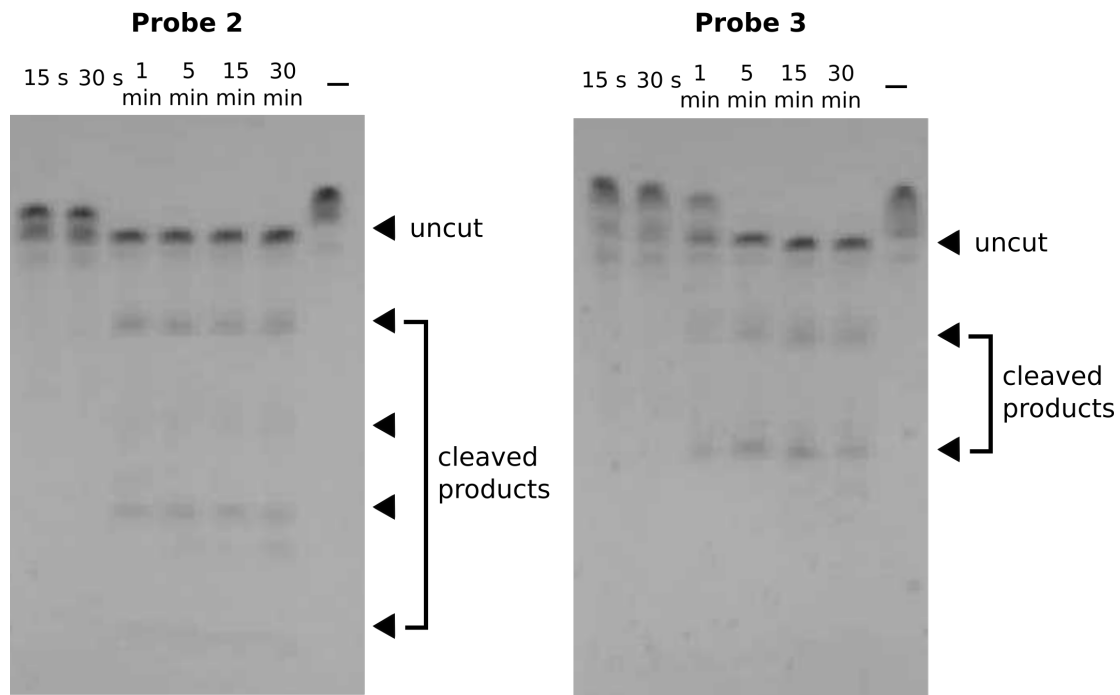


Figure 2.8 | PAGE analysis of probe 2 and 3 in the presence of RNase H as a function of time. Probe 2 was fully cleaved after 30 s while probe 3 took more than 1 min to be completely cleaved by RNase H. The reaction was loaded on an 8% denaturing PAGE gel and the efficiency of the reaction was determined by visualising under the UV channel. The oligonucleotide sequences can be found in Appendix Table 1.

The series of experiments on the different probes allowed us to determine the optimum RNase H cleavage window and also the choice of fluorophore and quencher that will give the most effective results. The design of probe 2 was therefore adapted to make a fgRNA-probe (fluorescent gRNA).

2.5 RNase H Activated fgRNA Design I

2.5.1 Design Strategy and Synthesis

The fgRNA was designed to target the telomeres, which are a common target for preliminary live cell imaging studies due to their repetitive nature. The synthesis of the fgRNA was divided into three segments: the crRNA (58-mer), which contains the complementary target sequence and the fluorophore, the tracrRNA (78-mer) and a

small oligo complementary to part of the crRNA (15-mer), containing the quencher (Figure 2.9). The gRNA is composed of a 79-mer tracrRNA and a 58-mer crRNA and making them separately ensures purer RNA products due to the limitations in the synthesis of long RNA. By dividing the gRNA into fragments, the RNA synthesis becomes more efficient and higher yielding as the synthesis of shorter RNA is not hindered by its low coupling efficiencies compared to longer RNA. In addition, only the crRNA will be made for each new DNA target as the tracrRNA sequence does not change and the short quencher-complement can also be kept constant.

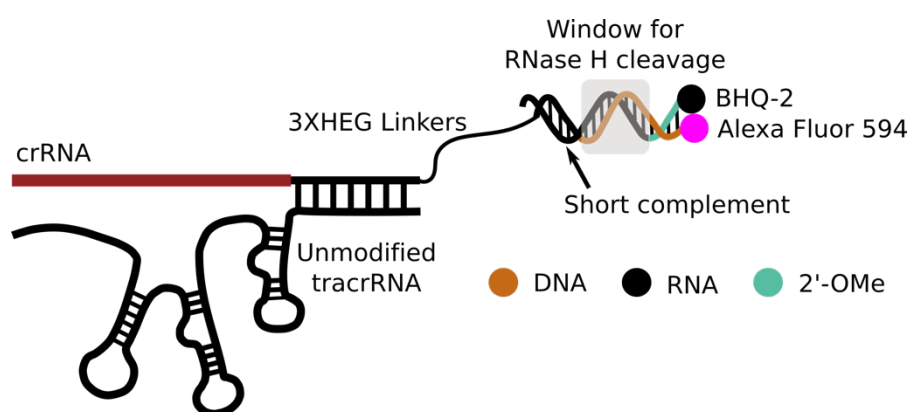


Figure 2.9 | RNase H Activated fgRNA Design I. The gRNA consists of an unmodified tracrRNA and crRNA conjugated to a small duplex via HEG linker at the 3'-end. One strand of the duplex is part DNA and part RNA bases and it is conjugated to a fluorophore. The complementary strand of the small duplex is made of RNA and 2'-OMe ribonucleotides and it is conjugated to a quencher. This allows control over the site of cleavage by RNase H. The oligonucleotide sequences can be found in Appendix Table 2. The RNase H cleavable window is 4 nucleotides in length

The crRNA was synthesised using solid-phase TC-synthesis. Three HEG modifications were incorporated using a commercially available hexaethylene glycol phosphoramidite. The last 8 nucleotides (5' → 3') were deoxyribosyl (DNA) bases. The crRNA also contained an amino-C7 at the 3'-end, incorporated using a commercially available amino-modified controlled pore glass (CPG). In comparison to test probe 2, 2'-OMe ribonucleotides were not incorporated in the strand bearing the fluorophore in the short duplex. This was for ease of synthesis when only RNA and DNA nucleotides

are used; adding another modification would further complicate the design. Nevertheless, the window for RNase H cleavage was kept the same, as 4 nucleotides. The same short quencher-complement that was used in the test probes was also used in the fgRNA. The complement was synthesised using solid-phase TBDMS RNA synthesis with a Black Hole Quencher[®] (BHQ-2) added to the construct *via* a commercially available phosphoramidite (Figure 2.9). Both the crRNA and the complement were purified using reverse-phase high performance liquid chromatography (RP-HPLC). The crRNA was subsequently labelled with AlexaFluor594 *via* NHS ester labelling. The tracrRNA was unmodified and therefore *in vitro* transcription was used for its synthesis. UPLC-MS data are attached in Appendix 2. The purity of the modified RNAs used are also depicted in Figure 2.10.

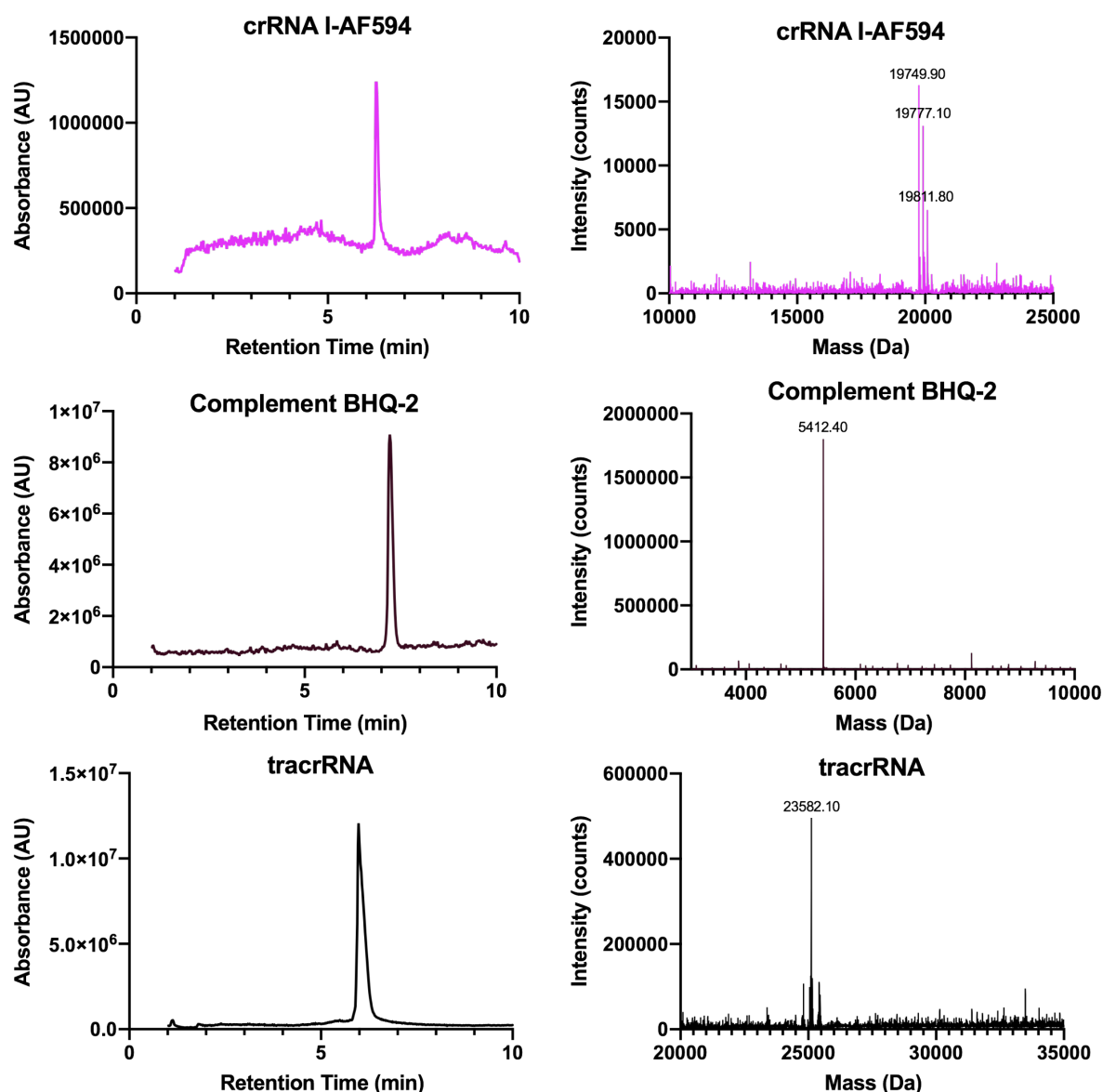


Figure 2.10 | Typical HPLC and MS traces of purified modified RNA. Modified RNAs used to assemble fgRNA I.

2.5.2 RNase H mediated *in vitro* Fluorescence Studies

Once all the components were prepared, the crRNA, the quencher-complement and the tracrRNA were mixed in a 1:1:1 ratio. Real-time fluorescence measurements of the construct were then taken in the absence and presence of RNase H. As a control, the gRNA was also assembled using a short complement that lacked the quencher.

This control will give us an estimate on the maximum fluorescence that the probe can reach.

As shown in Figure 2.11a, the probe was very efficiently quenched when no RNase H was present (690 a.u). In the absence of the quencher, the probe also displayed very high fluorescence of about 19,400 a.u. When treated with RNase H, the fluorescence almost immediately reached its maximum fluorescence. Decreasing the amount of RNase H did not decrease the rate of cleavage. Instead, the probe failed to reach maximum fluorescence, even when left to react for 12 h, and plateaued at a much lower fluorescence intensity, indicating that more RNase H was required. Adding Cas9 to the reaction mixture resulted in a general drop in fluorescence (Figure 2.11b). The main concern was RNase H would be unable to bind efficiently to the DNA-RNA heteroduplex after Cas9 binds to the gRNA. However, it did not greatly affect the rate of RNase H cleavage and the probe was still able to reach close to its maximum fluorescence.

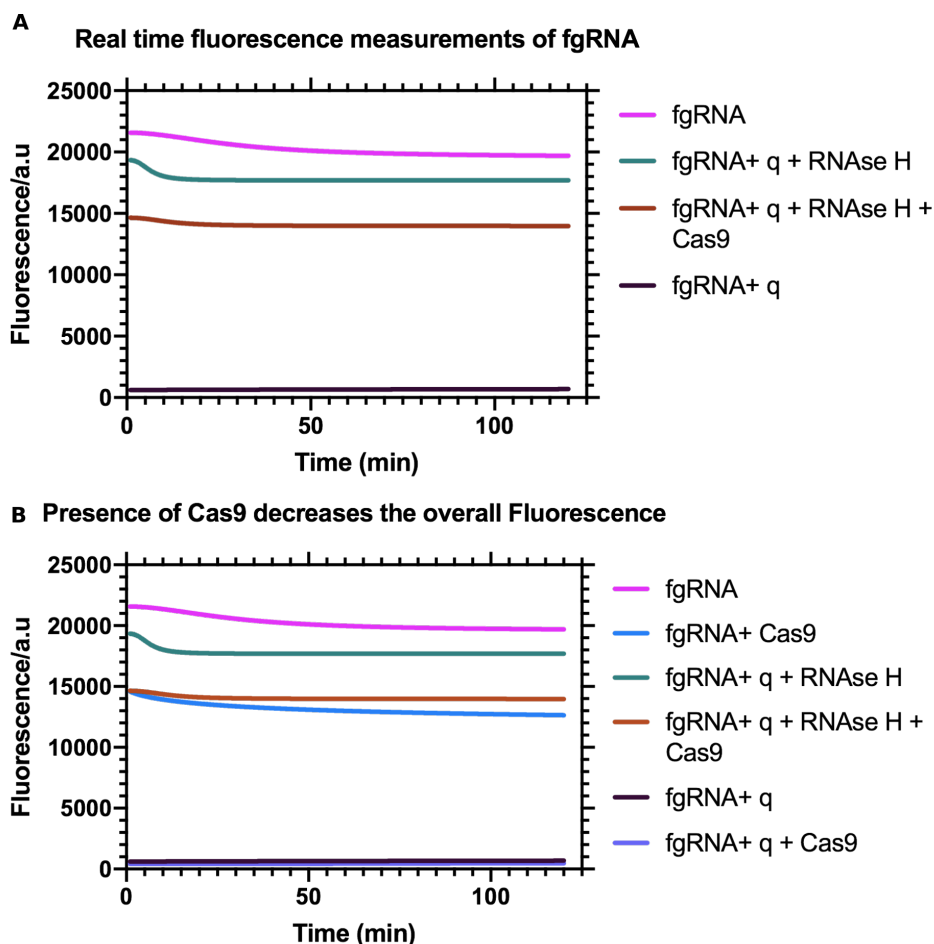


Figure 2.11 | Real time fluorescence measurements of fgRNA I *in vitro*. fgRNA I was resuspended in buffer containing 50 mM Tris-HCl, 7.5 mM KCl, 3 mM MgCl₂ and 10 mM DTT. Cas9 was added to the mixture followed by RNase H and the rise in fluorescence due to RNase H cleavage of the RNA-DNA heteroduplex was measured over time. (A) shows a considerable rise in fluorescence when RNase H was added to the quenched probe. (B) shows an overall decrease in fluorescence in the presence of Cas9. The oligonucleotide sequences can be found in Appendix Table 2.

2.5.3 Live Cell Imaging

After successful *in vitro* results, the next step was to test the fgRNA probe in cells. As mentioned previously, in this study we will be using cationic lipids, Lipofectamine reagents, to deliver the probe into cells. Lipofectamine has widely been used to successfully deliver DNA and siRNA into cells.¹⁶⁹⁻¹⁷¹ The cationic lipids is made up of positively charged heads, which interacts with and surrounds the negatively charged oligonucleotide. The lipid-oligo complex is then engulfed by the cell *via* endocytosis and is stored within vesicles known as endosomes. To escape from the endosomes,

the cationic lipids need to interact with the anionic lipids from the endosomal membrane to form ion pairs, which eventually destabilises the lipid bilayer of the endosomal membrane and releases the oligo into the cytoplasm.^{158, 172-173} For transfection to be successful, this process must happen before the endosome with the lipid-oligo complex fuses with lysosomes, which will result in degradation of the oligonucleotide.¹⁷⁴

After entering the cells and escaping from the endosomes, the fgRNA will need to form a complex with dCas9, which has a nuclear localisation signal. For the purpose of this study, dCas9-3xGFP plasmid was transfected and expressed in cells before the transfection of the fgRNA (Figure 2.12a). This will allow the puncta generated by the fgRNA to be cross validated as both the fgRNA and the dCas9 are fluorescently tagged. the same spot should be seen as two colours when using different lasers, confirming the colocalisation of dCas9 and fgRNA (Figure 2.12b). For successful genomic labelling, the fgRNA-dCas9-3xGFP complex must enter the nucleus and reach their target: telomeres. Since the telomeres are made up of highly repetitive sequences, several fgRNA-dCas9-3xGFP complexes are recruited to the same region, increasing the intensity of the fluorescent signal and resulting in a very bright punctum.

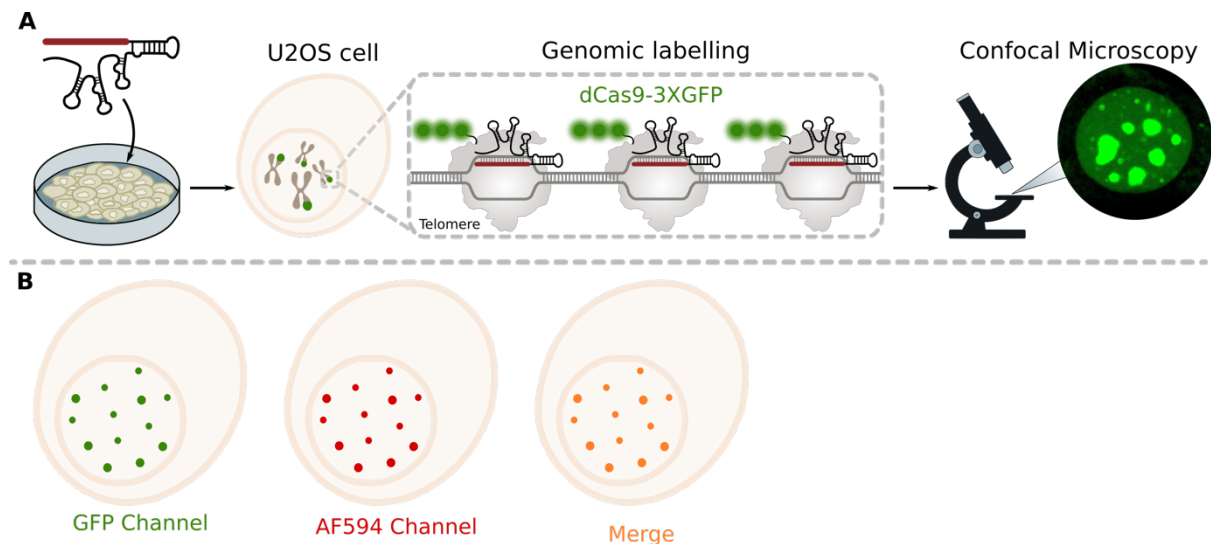
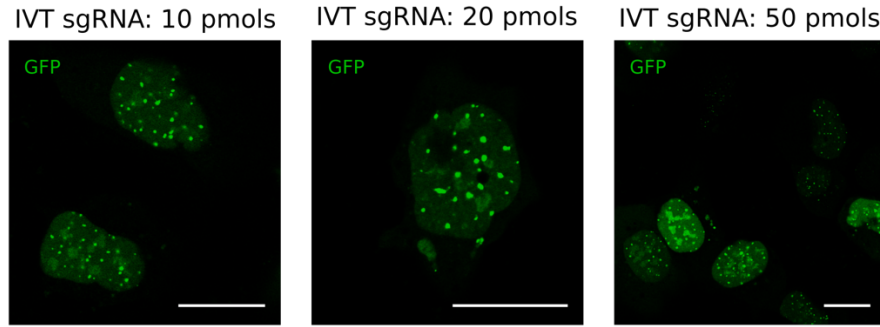


Figure 2.12 | Schematic representation of telomeric labelling and colocalization between GFP and AlexaFluor 594. (A) shows a schematic representation of transfection, telomeric labelling and visualisation. (B) illustrates labelled telomeres appear as green spots in the GFP channel, as red spots in the AlexaFluor594 channel and as orange spots when the two images are merged.

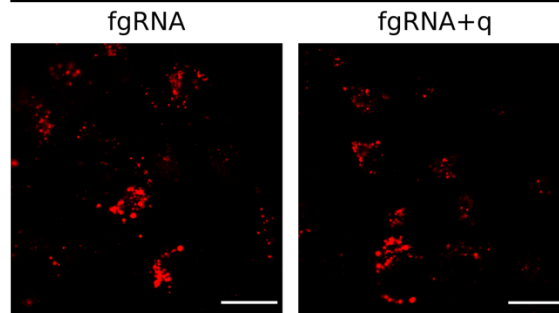
Initially, unmodified *in vitro* transcribed sgRNA was introduced into the cells as well as dCas9-3xGFP plasmid. This was used as a control as the sgRNA is known to bind to dCas9 very well. Furthermore, we also screened different concentrations of the sgRNA to find the optimum amount, that gave high transfection rates and good genomic labelling. The bright spots in Figure 2.13a shows successful transfection of both the dCas9-3xGFP plasmid and the sgRNA and also shows successful binding to genomic DNA. Transfecting 50 pmols of the sgRNA gave the highest transfection and target binding rates. For every transfection carried out here after, 50 pmols of fgRNA will be used.

A Transfection efficiency of IVT sgRNA at various concentrations



B

AF594 Background Signal



C

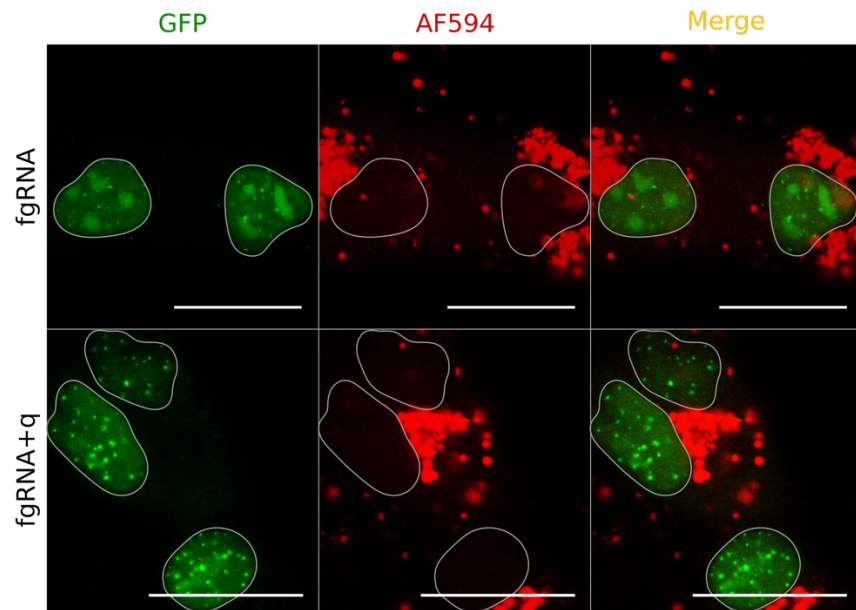


Figure 2.13 | Live cell imaging probe concentration optimisation and fluorescence labelling of fgRNA I. (A) shows transfection of 10, 20 and 50 pmols of sgRNA respectively and shows that the most efficient transfection was with 50 pmols of sgRNA. (B) shows the background signal of fgRNA I and fgRNA I+ q and shows there was no significant reduction of background even with the use of a quencher. (C) shows the labelling of the telomeres, which can only be observed in the GFP channel. Telomeric labelling is not observed in the AF594 channel and instead, the fluorophore was observed accumulated in high concentration in the cytoplasm. fgRNA I was transfected into dCas9-3XGFP expressing U2OS cells using Lipofectamine reagents. After 12–24 h, live-cell imaging was performed with different lasers used to visualise GFP (green) and Alexa Fluor 594 (red). Scale bar = 25 μ m. The oligonucleotide sequences can be found in Appendix Table 2.

Next, the fgRNA+q (fluorescent gRNA + quencher) was transfected into cells. As a control, the same fgRNA without the quencher was also transfected, to compare the background fluorescence as well as the intensity of genomic labelling. After visualisation under the confocal microscope, there was no significant improvement in background fluorescence when a quencher was used (Figure 2.13b). Upon closer inspection of individual cells, bright green spots representing telomeric labelling could be seen in the nucleus in the GFP channel (Figure 2.13c). However, those telomeric spots were not seen in the AlexaFluor 594 channel. Instead, large bright patches could be observed outside the nucleus (background fluorescence).

It was speculated that part of the probe where the fluorophore and the quencher is present was degraded in the cytoplasm, possibly by exonucleases and endonucleases, in the cytoplasm. Only part of the probe that was still functional and lacking a fluorophore and a quencher was making its way into the nucleus. The fluorescently tagged degraded oligo accumulated in the cytoplasm, which could be seen as bright red patches in Figure 2.13c. From this experiment, it was clear that the probe needed extra protection against exo- and endonucleases. As a result, the structure of the fgRNA was further optimised.

2.6 Fluorescent Test Probe for Stability

2.6.1 Design Strategy

To make the probe resistant to nucleases in the cytoplasm, more modified nucleotides (modified sugar-phosphate backbones) were incorporated, in the region surrounding the fluorophore, the quencher and the linker. As mentioned in section 2.3, 2'-OMe modifications greatly improve chemical and biological stability of nucleic acids *in*

vivo.¹⁷⁵ This is because, 2'-OMe changes the sugar pucker conformation and the overall structure of the helix, making it unrecognisable by *exo*- and *endonucleases*. The most commonly used phosphodiester backbone mimic is the phosphorothioate linkage. It is known to improve metabolic stability and also enhance pharmacodynamic properties through interaction with plasma proteins (Section 4.4.1). Unlike 2'-OMe sugar modifications, phosphorothioate linkages are compatible with RNase H cleavage.

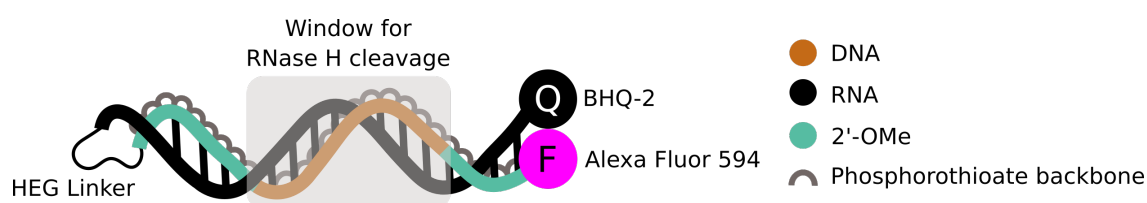


Figure 2.14 | Design of test probe with modifications to prevent exonuclease degradation. Small duplex to mimic the duplex conjugated to the gRNA, with 2'-OMe and phosphorothioate modifications to improve the stability of the probe *in cellulo*. The duplex is fluorescently labelled with Alexa Fluor 594 and is also conjugated to BHQ-2. There is a region of RNA-DNA heteroduplex, for efficient RNaseH degradation. The window for RNase H cleavage is 4 nucleotides. The oligonucleotide sequence can be found in Appendix Table 3.

With this in mind, a new test probe was designed. The probe was designed like a hairpin, to represent the overhanging duplex in the gRNA probe. At the 3'-end, an amino-C7 modification was incorporated *via* a commercially available CPG, which will be post synthetically labelled with AlexaFluor 594 (Figure 2.14). The proceeding fifteen nucleotides all contained 2'-OMe sugars with phosphorothioate linkages, except for four nucleotides in the middle of the strand, which contained unmodified DNA sugars with phosphorothioate linkages. These four nucleotides constitute the window on which RNase H can act. The complementary strand was synthesised with a commercially available BHQ-2 phosphoramidite at the 5'-end. All the nucleotides in

the strand were RNA bases and the two strands were joined together with a HEG linker, with a phosphorothioate backbone.

2.6.2 RNase H mediated *in vitro* Fluorescence Studies

Firstly, the compatibility of the heavily modified probe with RNase H was tested *in vitro*. When the fluorescence of the probe was measured in the absence of RNase H, it gave a very low fluorescence reading, averaging around 400 a.u (Figure 2.15). This shows that the probe was very efficiently quenched. Treating the probe with one or two units of RNase H resulted in a slow rate of reaction and the fluorescence failed to plateau, which means the cleavage reaction was incomplete. Upon addition of three units of RNase H, a gradual increase in fluorescence is observed, which plateaus at around 12,000 a.u. This is equivalent to a 30-fold rise in fluorescence, which was very promising.

Nonetheless, these observations show that the cleavage is concentration dependent as seen in Figure 2.15. The lower the concentration of RNase H, the slower the rate of reaction, taking more than 16 h for the fluorescence to become constant. This is an important observation, since we cannot control the concentration of RNase H in cells, we need to make sure that the probe is stable enough to avoid degradation in the cytoplasm but also sufficiently susceptible to RNase H fast enough to ensure efficient genomic labelling.

Even with three units of RNase H, the rate of reaction was slower than the gRNA probe, taking around 11 h to become constant. This is perhaps desirable as it would

mean rate of degradation in the cytoplasm will also be slower. However, it can also lead to inefficient target labelling if the probe takes too long to be activated

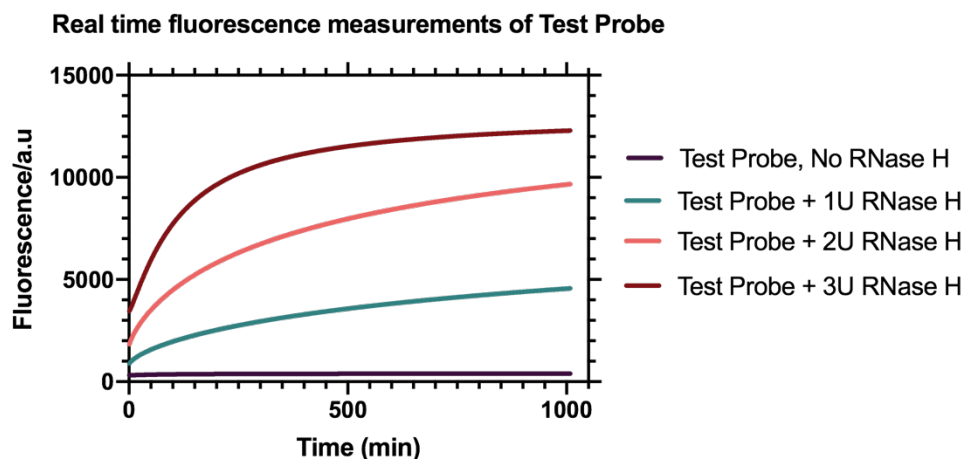


Figure 2.15 | Real time fluorescence measurements of the Test Probe in the presence and absence of RNase H. The test probe labelled with Alexa Fluor 594 and BHQ-2 was resuspended in buffer containing 50 mM Tris-HCl, 7.5 mM KCl, 3 mM MgCl₂ and 10 mM DTT. RNase H (1–3 units) was added to the mixture and the rise in fluorescence due to RNase H cleavage of the RNA-DNA heteroduplex was measured over time. The oligonucleotide sequences can be found in Appendix Table 3.

2.6.3 Live Cell Imaging

To test the stability of the probe in cells, it was transfected using Lipofectamine reagents. Since the probe did not code for any gene, it was not expected to see any specific labelling. Instead, absence of signal in the cytoplasm and an accumulation of fluorescence in the nucleus was expected, where RNase H are predominantly found. *In cellulo* results showed that there was no accumulation of fluorescence signal in the cytoplasm, which means the probe was stable to exonucleases (Figure 2.16). The red signal from the AlexaFluor 594 dye could only be observed in the nucleus, where RNase H cleaved the quencher part of the probe and activated fluorescence. Based on these promising results, the fluorescent gRNA was designed again and tested in cells again.

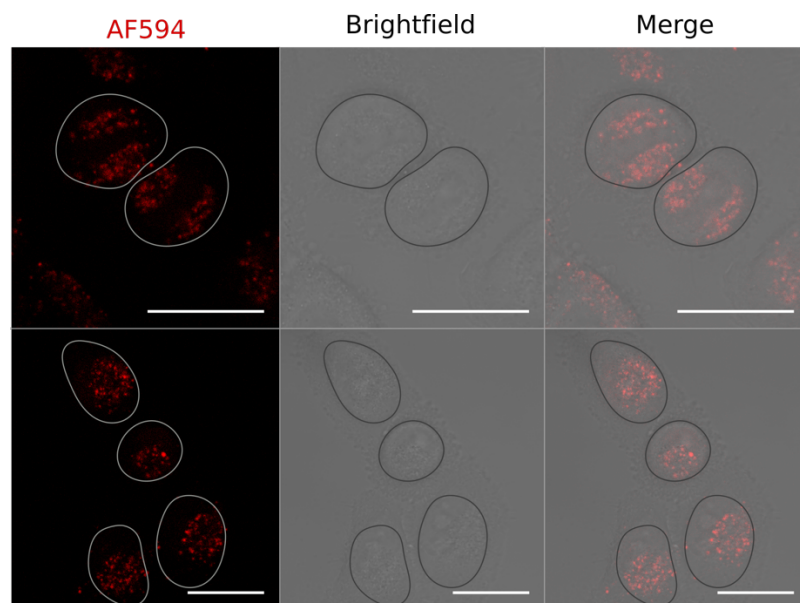


Figure 2.16 | Live cell imaging using the test probe. The test probe was designed with the intention of increasing the stability and live cell imaging proved that the test probe was more stable than fgRNA I as the majority of fluorescence was concentrated in the nucleus. There was a negligible amount of fluorescence in the cytoplasm. Transfection was performed in exactly the same way as the previous fgRNA constructs using Lipofectamine reagents. After 12–24 h, live-cell imaging was performed with different lasers used to visualise GFP (green) and Alexa Fluor 594 (red). Scale bar = 25 μ m. The oligonucleotide sequences can be found in Appendix Table 3.

2.7 RNase H Activated fgRNA Design II

2.7.1 Design Strategy

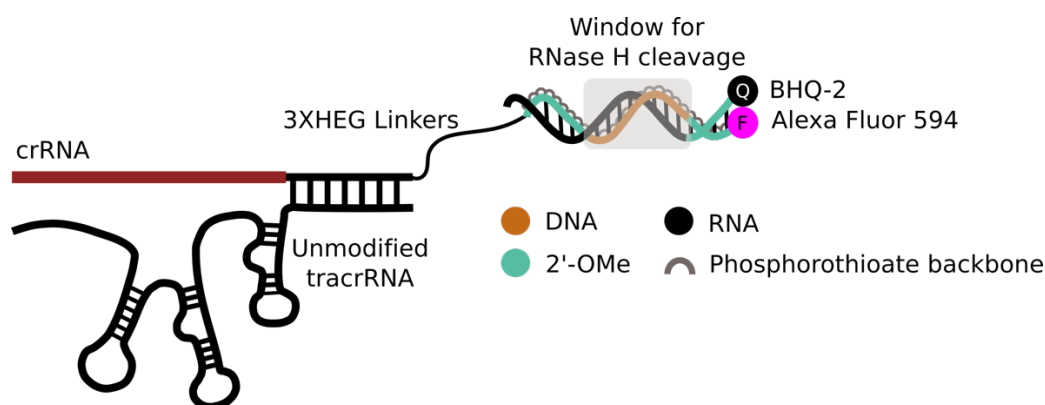


Figure 2.17 | Design of fgRNA II with modifications adapted from the test probe. The design was kept the same as fgRNA I, with the exception of a phosphorothioate backbone, which was now added onto one of the strands of the short duplex and the addition of extra 2'-OMe sugars in the short duplex. The oligonucleotide sequence can be found in Appendix Table 4.

fgRNA II was designed with the test probe in mind. Almost no change was made in the overhanging duplex involved in RNase H activation. The crRNA was chemically synthesised with 3 X HEG units linking the short overhanging strand bearing the fluorophore (Figure 2.17). The fluorophore was again incorporated by NHS ester labelling of an amine moiety in the RNA. 2'-OMe modifications were introduced in the same positions as the test probe, leaving only 4 unmodified DNA nucleotides for RNase H cleavage. Phosphorothioate linkages were included all the way from the HEG linkers to the end of the overhanging strand. The complementary strand of the overhanging duplex was synthesised separately and consisted of unmodified RNA nucleotides throughout the strand and 2'-OMe modifications in the 4 nucleotides preceding the BHQ-2, to protect the quencher against nucleases. The tracrRNA was unmodified and was made by *in vitro* transcription (discussed in section 1.9.1). The probe was assembled by mixing all three components in a 1:1:1 ratio.

2.7.2 RNase H mediated *in vitro* Fluorescence Studies

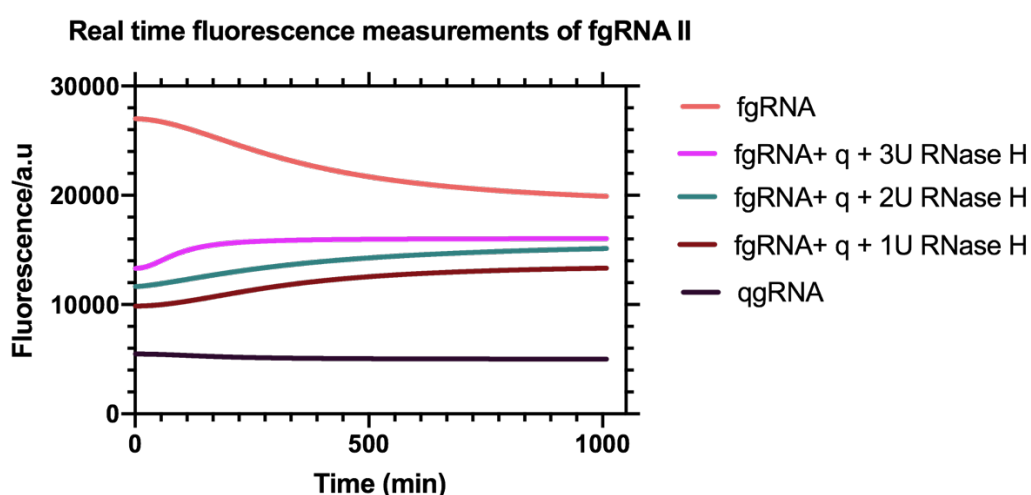


Figure 2.18 | Real time fluorescence measurements of the fgRNA II in the presence and absence of RNase H. FgRNA II was labelled with Alexa Fluor 594 and BHQ-2 and resuspended in buffer containing 50 mM Tris-HCl, 7.5 mM KCl, 3 mM MgCl₂ and 10 mM DTT. RNase H (1–3 units) was added to the mixture and the rise in fluorescence due to RNase H cleavage of the RNA-DNA heteroduplex was measured over time. The oligonucleotide sequences can be found in Appendix Table 4.

The assembled probe was tested *in vitro* for rise in fluorescence in the presence and absence of RNase H. As a control, the probe was also assembled without a quencher to record the maximum fluorescence that the probe can potentially reach. The maximum fluorescence of the probe was desirably high, averaging around 22,500 a.u (Figure 2.18). The fluorescence of the quenched probe in the absence of RNase H was around 5000 a.u, which was unusually high and meant that the probe was not efficiently quenched. Although the addition of RNase H gave a good rise in fluorescence of about 3-fold (when 3 units of RNase H was used), the issue of inefficient quenching needed to be addressed as it would lead to background fluorescence in cells.

As small impurities can give significant fluorescence, the purity of the modified oligonucleotides was examined. As shown in Figure 2.19, the HPLC and the MS traces of both the fluorescent and the quenched oligonucleotide displayed good purity.

In an attempt to improve quenching, all the components of the probe were mixed and then annealed together. The mixture is heated to 95 °C to denature all the base pairs including any secondary structures between the oligos. The mixture is then slowly cooled down, allowing hydrogen bonds between the complementary bases to form. If the short complement with the quencher was unable to hybridise or was hybridising to other parts of the construct, it would lead to inefficient quenching as the BHQ-2 and the fluorophore would be too far away from each other. In this case, the annealing process should help the complement hybridise to the correct strand. Unfortunately, when fluorescence measurements were carried out, there was no improvement in

quenching and heating the construct to 95 °C lowered the overall intensity of the fluorophore (Figure 2.20a).

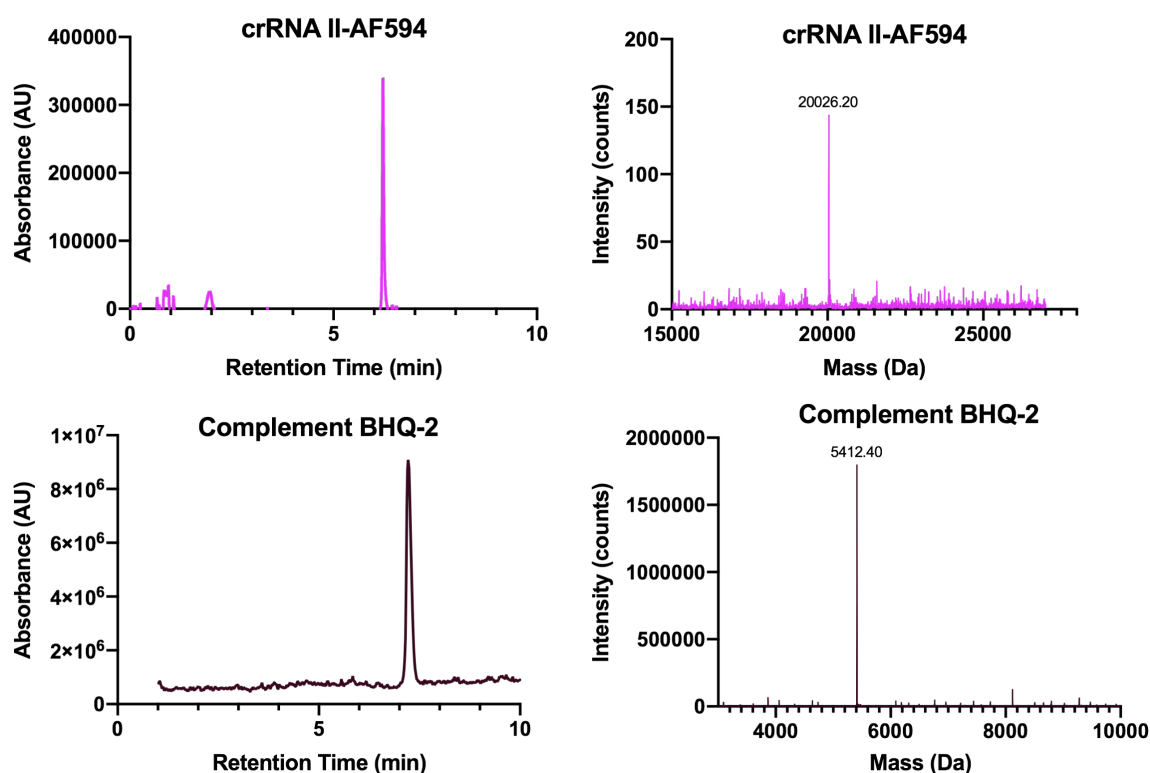
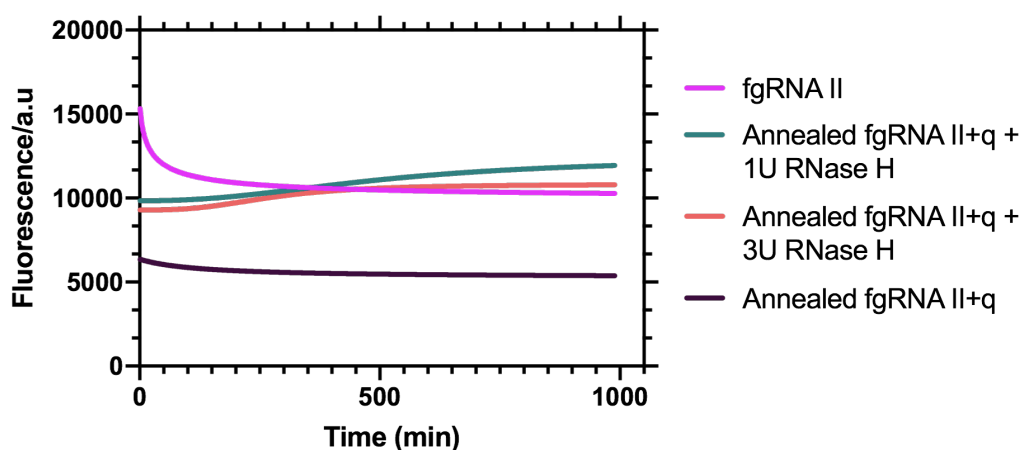


Figure 2.19 | HPLC and MS traces to show purity of the modified RNAs used.

Increasing the ratio of the quenched strand to the ratio of the fluorescence strand could be another solution to the problem. Inadequate quenching could be because the fluorescent strand was in slight excess. Inaccuracy during pipetting and measuring the optical density (OD) could have led to this. Therefore, the probe was assembled using an excess of the quencher strand. 2x excess of the quencher strand was used but unfortunately, no significant improvement in quenching was observed (Figure 2.20b).

A Real time fluorescence measurements of Annealed fgRNA II



B Real time fluorescence with excess quencher

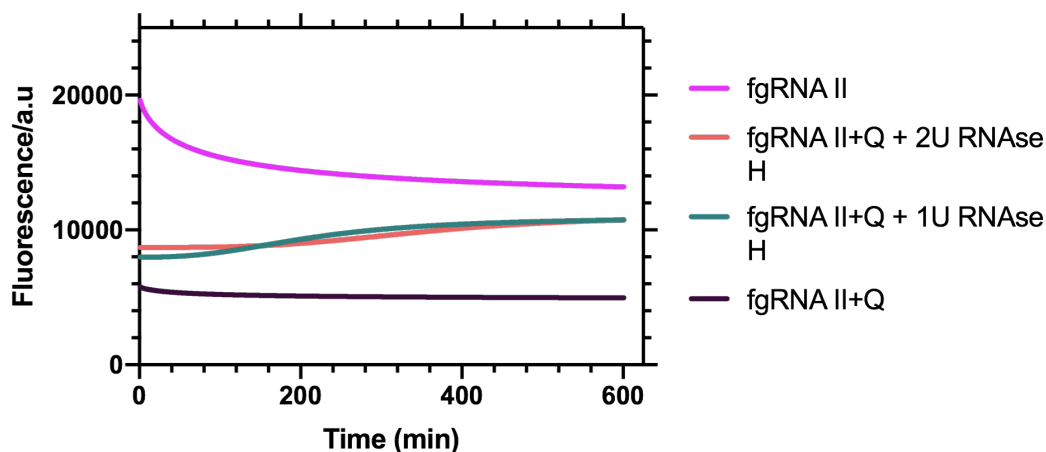


Figure 2.20 | Fluorescence measurements of fgRNA II after annealing or using an excess of quencher. (A) shows that annealing the probe did not improve the quenching efficiency of fgRNA II. (B) shows the fluorescence reading when 2x excess of the quencher strand was used and unfortunately did not show an improvement in quenching efficiency.

The main difference between the short duplex in the test probe and the fgRNA probe is that the two strands are held together by the HEG linkers in the test probe. In fgRNA II, the two strands are simply hybridised together and held by hydrogen bonding. One hypothesis could therefore be that at 37 °C, the two strands were dissociating from each other, increasing the distance between the fluorophore and the quencher and consequently leading to poor quenching. Fluorescence melting was therefore carried out to compare the T_m of the two probes. This will be the temperature at which 50% of

the fluorescent strand has been dissociated from the quenched strand. Figure 2.21 shows the fluorescence T_m of the two probes and the T_m of the test probe is significantly higher than that of fgRNA II. Although the T_m of fgRNA II is around 70 °C (higher than 37 °C), the lower T_m still means that a higher proportion of the double stranded duplex is dissociated in fgRNA II at 37 °C compared to the test probe. The design fgRNA II was consequently further optimised to ensure that the overhanging duplex were anchored together.

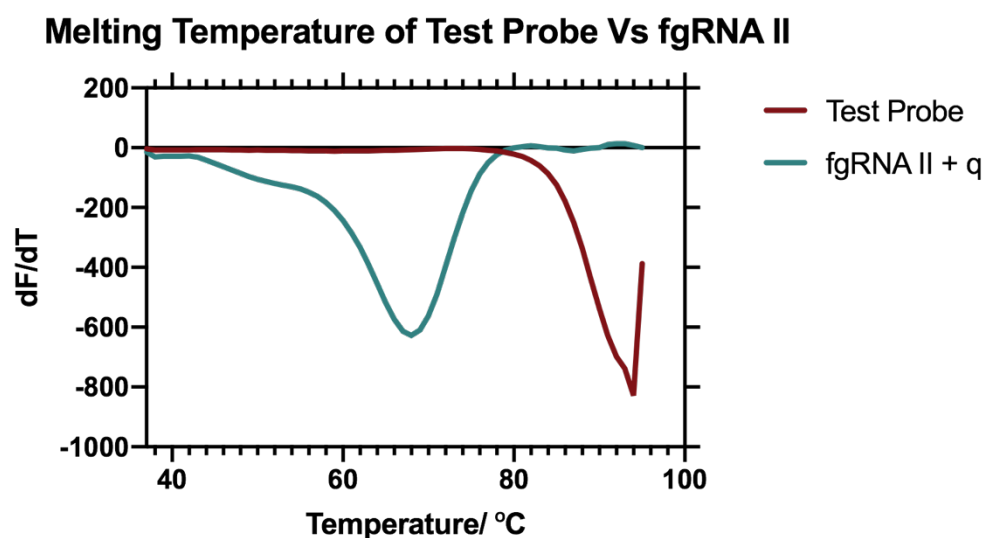


Figure 2.21 | Melting temperature comparison between the test probe and fgRNA II + q. The test probe and fgRNA II + q were resuspended in buffer containing 50 mM Tris-HCl, 7.5 mM KCl, 3 mM MgCl₂ and 10 mM DTT. The rise in fluorescence due to the melting of the strand bearing the quencher was measured over increasing temperatures. The oligonucleotide sequences can be found in Appendix Table 4.

2.8 Crosslinked Version of fgRNA Design II

2.8.1 Design Strategy

Instead of redesigning the whole probe to anchor the short duplex together, a simpler and more cost-effective alternative was to redesign only the quencher complement. The short complement was designed again to incorporate a 3-Cyanovinylcarbazole

(^{CNV}K) monomer at position 10 of the strand (Figure 2.22). ^{CNV}K undergoes very rapid photo cross-linking with the complementary strand when irradiated at 366 nm.¹⁷⁶ For efficient photo cross-linking, it is essential that the complementary strand contains a thymine or a cytosine adjacent to the ^{CNV}K. Since the crRNA already had a thymine base in that position, there was no need to redesign or resynthesise the latter. The structure and mechanism of cross-linking of ^{CNV}K is further discussed in Chapter 3.

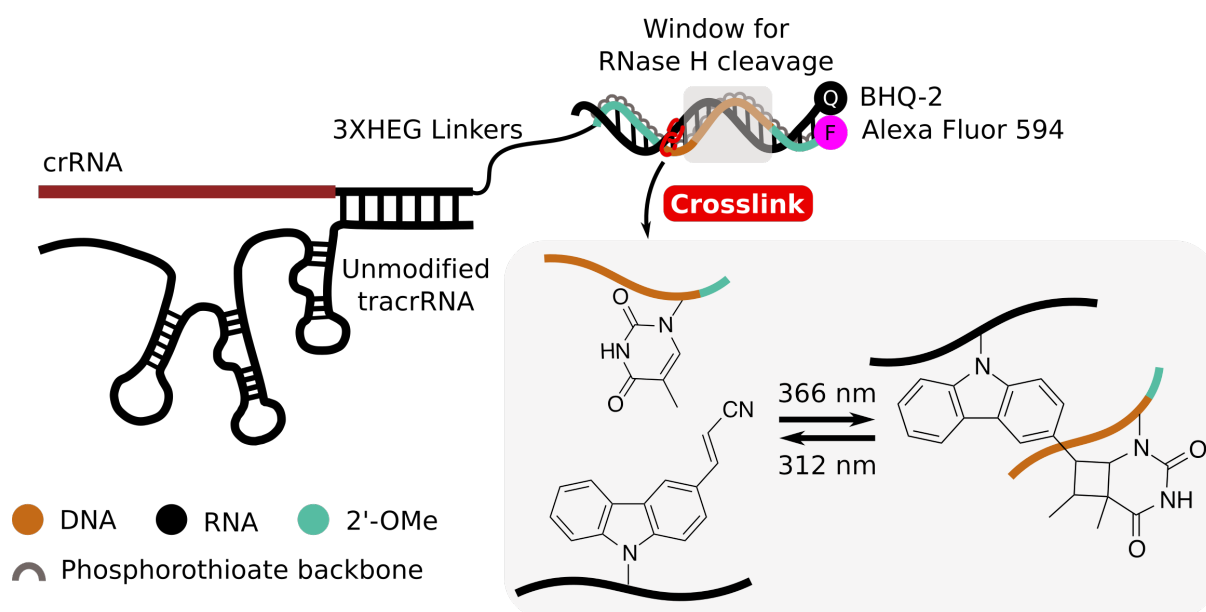


Figure 2.22 | Design of fgRNA III where the two strands in the short duplex are cross-linked. A ^{CNV}K modification was included in the RNA strand of the short duplex and upon UV irradiation, it cross-links with a thymine base on the complementary strand. The oligonucleotide sequence can be found in Appendix Table 4.

Cross-linking the two strands will create a strong bond, relatively close to the fluorophore and quencher, preventing them from distancing from each other. It will also increase the frequency of collision between them, increasing the efficiency of dynamic quenching. The cross-linked bond can only be broken when irradiated at 312 nm. After cross-linking the two strands, purification by denaturing gel electrophoresis was carried out. Gel electrophoresis revealed that two major products were formed (Figure 2.23). Unfortunately, mass spectrometry could not help determining the nature of the products due to the large size of the product, leading to inaccurate

deconvolution. When fluorescence analysis was carried out, both products seemed to contain the quencher. It was speculated that a combination of non-specific cross-linking and secondary structures gave rise to the second product. By comparing the product length and migration, the lower band was expected to be the right product. Since the two strands would be crosslinked together and would remain a duplex, an extreme migration on 8% PAGE was not expected. However, both bands were extracted and tested to ensure that the correct and best performing product was chosen.

The two products were separately assembled with the tracrRNA in a 1:1 ratio. The cross-linked bond was formed in the RNase H cleavage window and therefore it was imperative to carry out RNase H cleavage and fluorescence assay to make sure that the enzyme was still able to activate the probe.

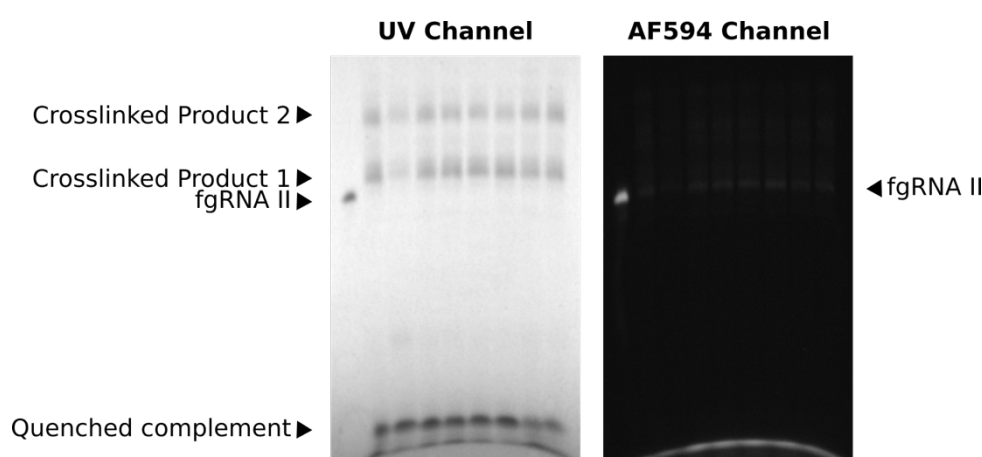


Figure 2.23 | Cross-linking of fgRNA III. The RNA strand with the ^{CNV}K and BHQ-2 modification was mixed with crRNA-AF594 in a 4:1 ratio in the presence of 100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl₂ and 100 µg/ml BSA (pH 7.9). The mixture was then exposed to 360 nm UV for 15 min. The reaction was loaded on an 8% denaturing PAGE gel and the efficiency of the reaction was determined by visualising under the UV channel and AF594 channel. The oligonucleotide sequence can be found in Appendix Table 4.

2.8.2 RNase H mediated *in vitro* Fluorescence Studies

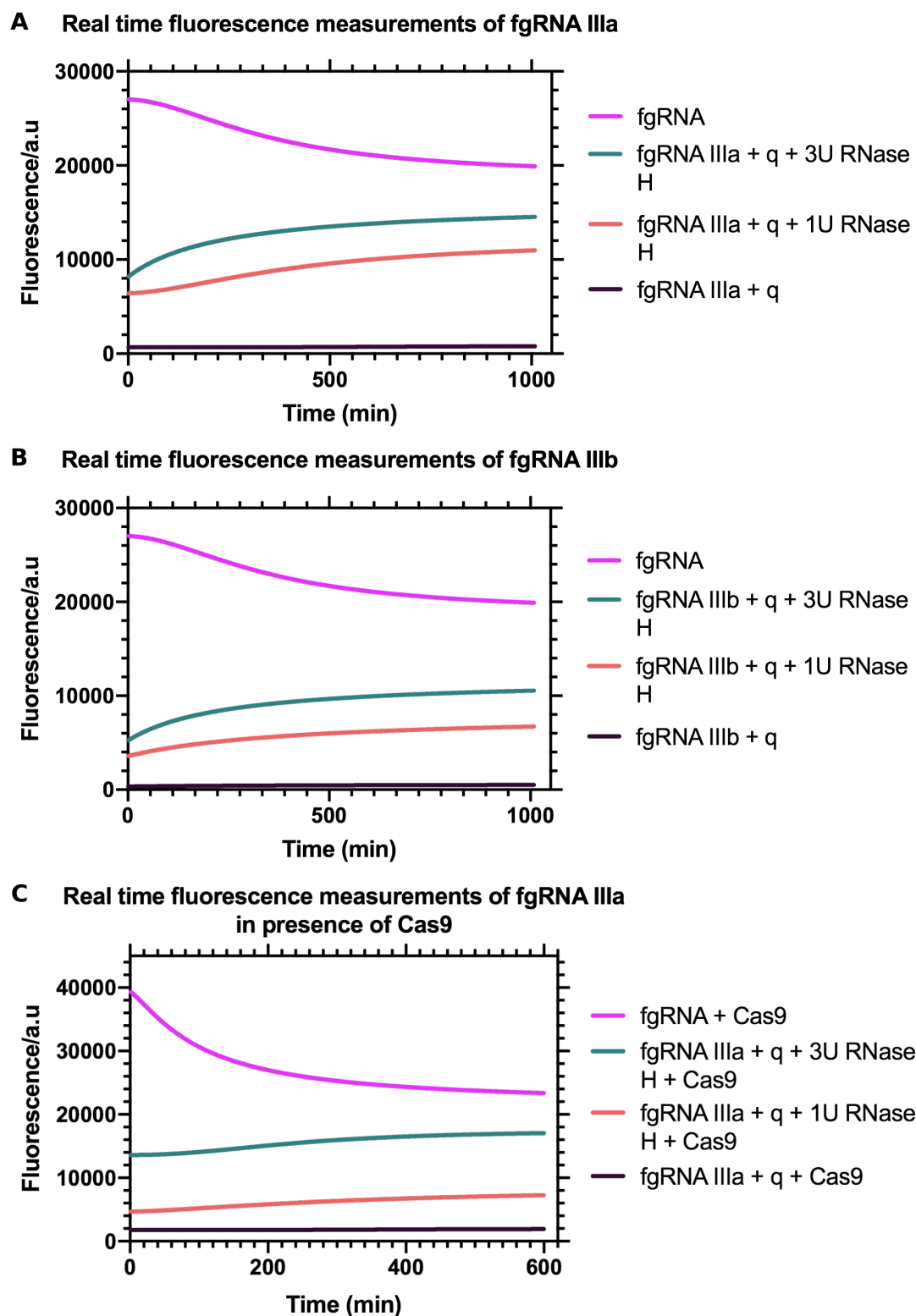


Figure 2.24 | Real time fluorescence measurements of the fgRNA IIIa compared to fgRNA IIIb in the presence and absence of RNase H. FgRNA IIIa/b were labelled with Alexa Fluor 594 and BHQ-2 and resuspended in buffer containing 50 mM Tris-HCl, 7.5 mM KCl, 3 mM MgCl₂ and 10 mM DTT. Cas9 was added to the mixture followed by RNase H (1 and 3 units) was added to the mixture and the rise in fluorescence due to RNase H cleavage of the RNA-DNA heteroduplex was measured over time. (A)

shows the rise in fluorescence of fgRNA IIIa + q. The probe displayed good quenching efficiency and good rise in fluorescence after adding RNase H. (B) shows the fluorescence measurements of fgRNA IIIb + q, which displayed good quenching efficiency but slightly lower rise in fluorescence after adding RNase H compared to fgRNA IIIa + q. (C) shows addition of Cas9 had negligible effect on fluorescence measurements. The oligonucleotide sequences can be found in Appendix Table 4.

The gRNA assembled with the product from the lowerband is referred to as fgRNA IIIa and the gRNA assembled from the upperband is referred to as fgRNA IIIb. There was a significant improvement in quenching efficiency compared to fgRNA II and both fgRNA IIIa and fgRNA IIIb had similar quenching efficiencies of around 600 a.u (Figure 2.24a-b). When the probes were treated with RNase H, both fgRNA IIIa + q and fgRNA IIIb + q performed equally well. They both gave an approximately 20-fold increase in fluorescent upon addition of three units of RNase H. This presumably means that they are both the same products and secondary structures are causing the difference in migration. Moreover, addition of Cas9 to the reaction mixture had negligible effect on RNase H activity (Figure 2.24c), which meant that even though Cas9 was in complex with the gRNA, RNase H could still effectively bind to the gRNA and cleave the target. FgRNA IIIa was extracted in a considerably higher yield and it was therefore tested in cells to check if the results are reproducible *in cellulo*.

2.8.3 Live Cell Imaging

fgRNA IIIa was transfected into cells using lipofectamine reagents. As shown in Figure 2.25a there was a significant improvement in background fluorescence in the presence of a quencher. In fact, background fluorescence was almost negligible when the quenched probe was transfected. fgRNA IIIa–q showed no telomeric labelling in the AlexaFluor 594 channel but instead showed labelling in the GFP channel. Moreover, there was considerable accumulation of the AlexaFluor 594 dye in the cytoplasm, suggesting degradation of the probe in proximity to the fluorophore (Figure

2.25b). The results are very similar as to when fgRNA I was transfected in cells. Interestingly, we successfully observed telomeric labelling with fgRNA IIIa+q inside the nucleus in both the GFP and the AlexaFluor 594 channel as shown in Figure 2.25c. There was also no significant accumulation of dye in the cytoplasm. This suggests that the cross-linked complement was providing additional protection against nucleases, allowing the probe to enter the nucleus. In comparison to labelling with GFP, AlexaFluor 594 emitted significantly less nuclear background fluorescence. In addition, we do not have the problem of accumulation of fluorophore in the nucleolus when the chemical fluorophore is used.

The main issue with probe fgRNA IIIa was the binding efficiency. When the telomeres are targeted, we expect to see multiple puncta, representing the numerous repetitive sequences present in cells. Nonetheless, in the case of fgRNA IIIa+q we only saw a few labelling representing the telomeres. This suggests that the probe does not bind well to the dCas9 or to the target. The HEG modifications along with the short duplex might not be well tolerated. The system needs to be redesigned to find an alternative way to incorporate the short duplex to improve binding. Another reason why such a low number of labelling was observed might be due to poor transfection rates or poor nuclear localisation. FgRNA IIIa+q was subjected to purification by gel electrophoresis after cross linking. This technique involves extracting the oligonucleotide from 7 M of urea, and the resulting oligo can be very salty, despite numerous desalting techniques. The presence of salts can interfere with the formation of lipid-oligo complex during transfection and can lead to poor transfection rates. To overcome this problem, the oligonucleotide can be precipitated after extraction. However, when this method was attempted, the yield obtained was too small to carry out any analysis.

In the few cases where we did observe multiple telomeric labellings in the GFP channel, not all of them were colocalised in the AlexaFluor 594 channel (Figure 2.25c). This suggests that some of the probe were still getting degraded and while the stability has been greatly improved, there is still improvement to be made to avoid probe degradation.

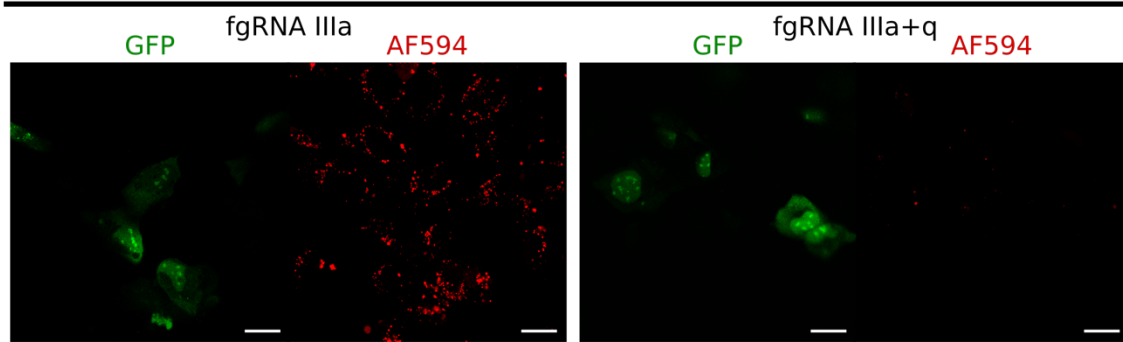
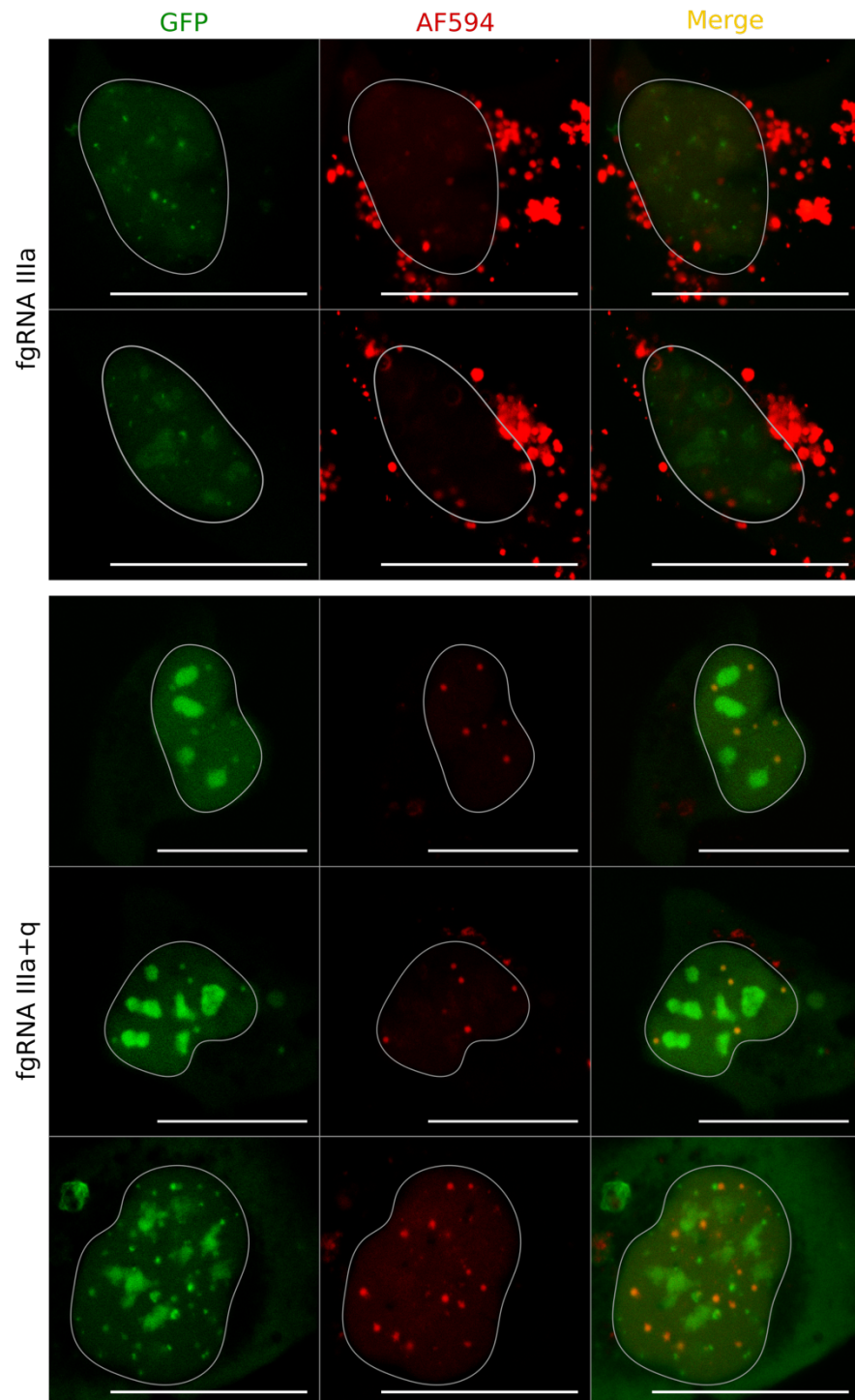
A**Background Signal****B**

Figure 2.25 | Live cell imaging using fgRNA IIIa. (A) shows a significant reduction in background fluorescence when BHQ-2 quencher is used. It also shows a significant reduction in nuclear background when GFP is used compared to Alexa-Fluor 594. (The same excitation and emission settings were used). (B) shows telomeric labelling when fgRNA IIIa that contains no quencher was used and accumulation of fluorescence in the cytoplasm can be observed (C) shows telomeric labelling when fgRNA IIIa with BHQ-2 quencher was used and there is significantly less accumulation of fluorescence in the cytoplasm. Transfection was performed in exactly the same way as the previous fgRNA constructs using Lipofectamine reagents. After 12–24 h, live-cell imaging was performed with different lasers used to visualise GFP (green) and Alexa Fluor 594 (red). Scale bar = 25 μ m. The oligonucleotide sequences can be found in Appendix Table 4.

2.9 Conclusions and Future Work

A novel method of live cell genomic imaging using CRISPR-Cas9 was explored. The main objective was to successfully image specific genomes, using bright, photostable small molecule fluorophores as well as improving the signal to background ratio. An RNase H mediated, fluorescence activation approach was discussed in this chapter and it has been shown that a key factor that contributes to the successful nuclear localisation was a stable fgRNA, that was protected against exonucleases. Modifications such as 2'-OMe sugars and phosphorothioate backbone were used to improve stability. A quencher was employed in the design of the probes to reduce background fluorescence and to eliminate granular cytoplasmic background due to the endosomally trapped fgRNA. It has also been shown that good strand affinity of the duplex with the fluorophore and quencher was essential for efficient quenching of the fluorophore. Cross-linking of the duplex with the fluorophore and the quencher was explored to increase strand affinity. Transfection of fgRNA IIIa+q showed negligible background fluorescence and minimal fluorescent endosomal entrapment. It also showed successful labelling of the telomere, confirmed by co-localisation with GFP. However, only a few puncta corresponding to the labelled telomeres were observed. We expected to see numerous labellings (>10) as the telomeres are highly repetitive.

Future work will involve incorporating the short stable duplex with the fluorophore and the quencher onto the sgRNA *via* a HEG-linker (Figure 2.26). The sgRNA binds considerably better to its target and as a result labelling efficiency is very good in comparison to a gRNA. Conjugating the short duplex onto a sgRNA should resolve the problem of observing only a few fluorescent puncta. By using an alkyne modified nucleotide incorporated in the tetraloop and an azido-modified nucleotide on the HEG-linker attached to the small duplex, we will be able to click the two components. This should result in a probe that stably binds to its target and that gets efficiently activated when located in the nucleus.

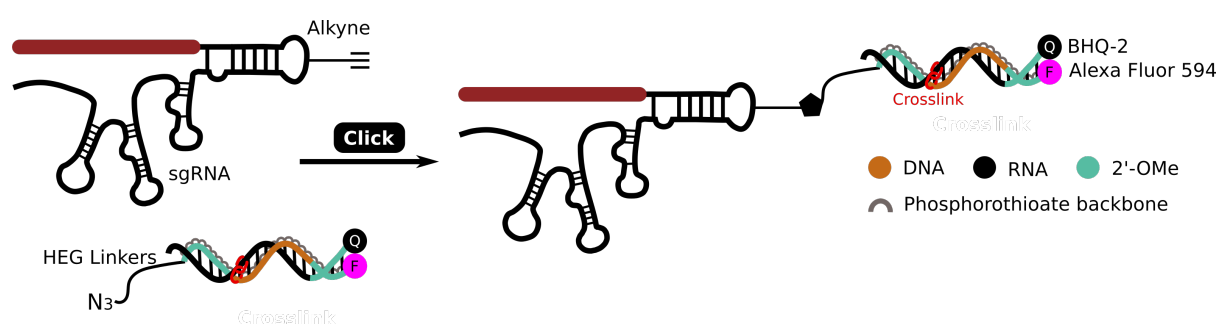


Figure 2.26 | Design of qqRNA for future work. The design consists of a sgRNA to get optimum target binding, with a metastable fluorescent duplex conjugated *via* click chemistry.

Chapter 3

3'-Modified sgRNAs

using crosslinking to improve synthesis

3.1 Background

The use of modified gRNAs has become incredibly popular as scientists continue to focus their efforts into improving CRISPR-Cas9 mediated live cell imaging, gene editing or finding better ways to deliver CRISPR systems into cells. With the exception of conventional gene editing, most applications of CRISPR-Cas9 require either a modified gRNA or a modified Cas9. It is considerably costly to make modified long RNAs, due to the cost of the monomers and due to the low coupling efficiency of RNA, that gets worst with the growing RNA chain. In this chapter, we aim to provide a novel solution to this problem by designing a two-part sgRNA system in which the parts can be crosslinked to each other. Psoralen and 3-Cyanovinylcarbazole (^{CNV}K) have widely

been used to crosslink oligonucleotides. Nonetheless, they are yet to be applied in the CRISPR field.

3.1.1 Modifications to induce photo-crosslinking

3.1.1.1 Psoralen Modifications

The psoralen aromatic system is a bi-functional photoactive molecule, which can be used to probe nucleic acid structure and function (Figure 3.1a).¹⁷⁷ It intercalates within dsDNA. Upon UV irradiation at 320–400 nm it either forms a pyrimidine-psoralen mono-adduct or a bi-adduct with adjacently stacked pyrimidine bases (Figure 3.1b). Crosslinking occurs through a [2+2] cycloaddition reaction between one of the two reactive double bonds of psoralen and the reactive double bond of thymine.

The photo-chemical properties of psoralen have been exploited in many ways and applied to various field. One of the earliest applications of psoralen derivatives was for the treatment of various skin diseases such as psoriasis and leukoderma.¹⁷⁸⁻¹⁸⁰ The treatment is called (Psoralen + UVA) therapy and involves the topical or oral application of psoralens and irradiation with UVA light (320–400 nm).¹⁸¹ [2+2] cycloaddition between psoralen derivatives and epidermal DNA inhibits DNA replication and causes cell cycle arrest. As a result, there is a drop in the expression of cytokines and cytokine receptors, reducing inflammation.¹⁸²⁻¹⁸³ However, psoralen derivatives have been reported to cause erythema and have been identified and carcinogens and mutagens when used in conjunction with UV.¹⁸⁴

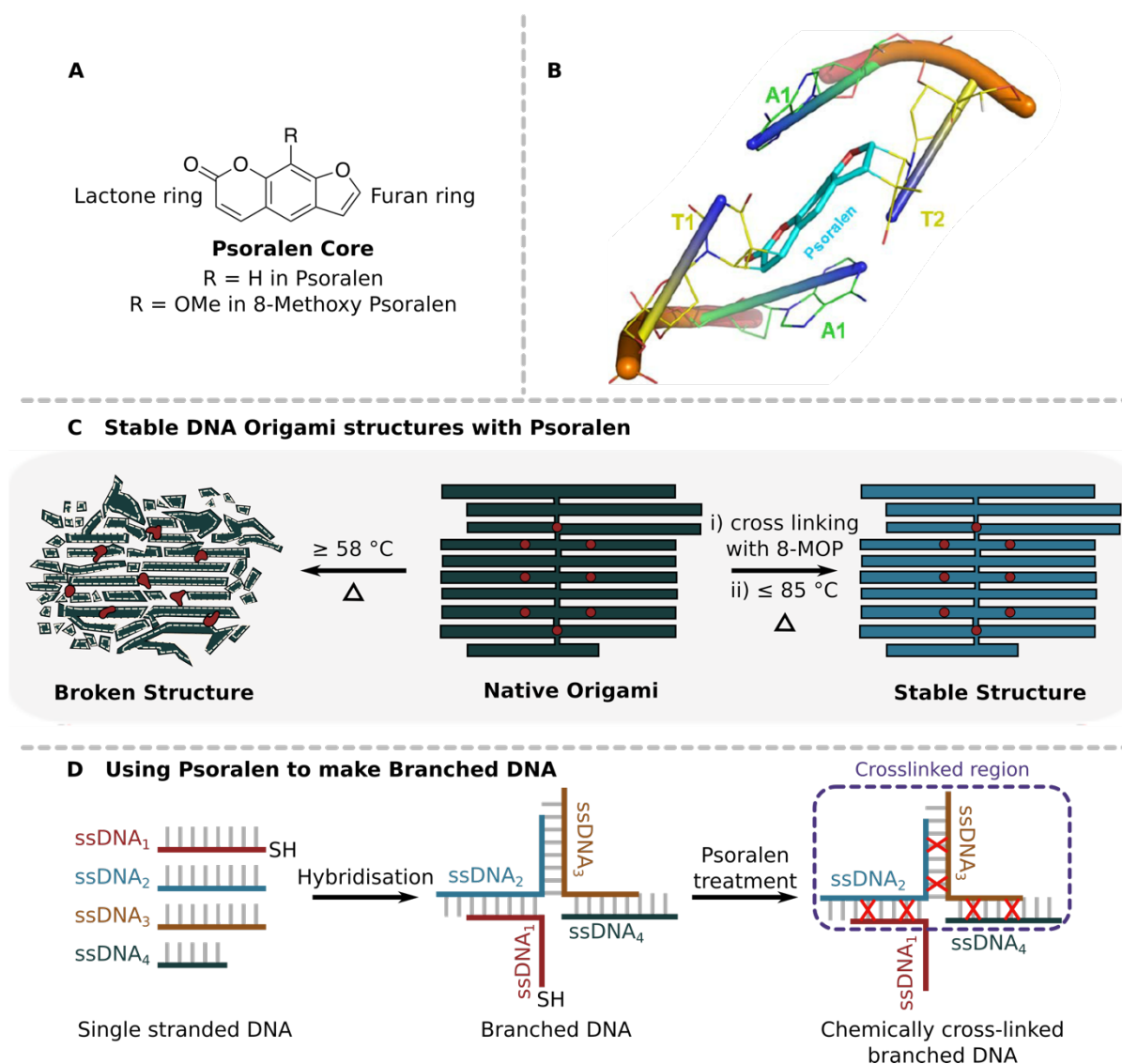


Figure 3.1 | Psoralen structure and its applications. A) shows a psoralen moiety consisting of three main parts: A psoralen core, a lactone ring and a furan ring. B) represents a model for the crosslinking of psoralen and thymine. Psoralen is interacted between two pairs of A-T bases. C) schematic drawing showing how crosslinking with 8-MOP helps to stabilise a DNA origami structure and helps it retain its shape after heating. Native origami loses its shape and denatures after heating. D) the synthetic steps to making branched DNA, comprising of annealing four ssDNA followed by psoralen treatment. This enabled the chemically crosslinked branched DNA to remain intact even after being exposed to denaturing conditions.

More recently, psoralen has been utilised to stabilise extended DNA origami structures.¹⁸⁵⁻¹⁸⁶ Scaffolded DNA origami has a plethora of applications such as: being used as a platform for the nanopatterning of a variety of functional molecules, templates for the growth of nanowires, as nanochips in solution to monitor biochemical

reactions and conformational changes, and for single molecular analysis of various functionalities.¹⁸⁷⁻¹⁹⁰ The origami structures can also be confined within a lithographically patterned surface for nanoscale electronic and photonic applications.¹⁹¹ Nonetheless, these applications are limited due to the low melting temperature of DNA, typically around 55 °C.¹⁹² In order to make more practical devices using DNA origami, this problem should be resolved. In 2011, Rajendran *et al.* reported the photo-crosslinking of DNA origami using 8-methoxypsoralen (8-MOP).¹⁸⁶ The crosslinked DNA origami had an improved thermal stability (+30 °C), which widened the scope of the DNA nanostructures (Figure 3.1c).

In a study in 2002, Patolsky *et al.* demonstrated the generation of gold nanoparticle wires onto DNA templates.¹⁹³ The DNA-gold nanoparticle wires were made by attaching a psoralen moiety onto the surface of 1.5 nm gold nanoparticles and mixing them with a double stranded poly A/poly T duplex. The resulting mixture was irradiated with 360 nm UV light to form the DNA-gold nanoparticle wires. A similar strategy was used by Li *et al.* to make a branched DNA nanostructure, with the aim of creating a recyclable biosensing platform for highly specific and ultra-sensitive DNA detection.¹⁹⁴ The system was made up of a DNA probe, consisting of a high density of thymine bases in the double-stranded DNA, to crosslink with psoralen molecules (Figure 3.1d). The DNA probe recognises a target DNA and is coupled with DNA-functionalised iron oxide (Fe₃O₄) nanoparticles for signal amplification. The presence of multiple covalent bonds in this biosensing platform enables it to endure denaturation and high temperature conditions, allowing regeneration of the interface. The device could also be recycled without loss of detection sensitivity.

While psoralen and its analogues have had a remarkable success, their applications were limited by the choice of DNA target as an efficient cross-linking reaction requires a TpA (intercalation of psoralen between 5'-T and 3'-A) step. In addition, reversing the reaction between psoralen and thymine base occurs at 254 nm, which is highly damaging to normal DNA as it can result in the formation of pyrimidine dimers. To widen the applications of photo-crosslinking and to overcome the challenges of sequence dependence and DNA damage, researchers have focussed on finding alternative photo-crosslinkers with high reactivity.

3.1.1.2 3-Cyanovinylcarbazole modifications

In 2008, Yoshimura and Fujimoto reported a novel ultrafast interstrand photo-crosslinking reaction via ^{CNV}K in duplex DNA. An oligonucleotide containing the ^{CNV}K can be rapidly and cleanly crosslinked to an adjacent pyrimidine base via UV irradiation at 366 nm in a [2+2] cycloaddition reaction. This crosslink can be efficiently reversed by irradiating at 312 nm, without causing damage to the parent OGN (oligonucleotide) (Figure 3.2).¹⁷⁶ Photo-crosslinking has been used in a myriad of applications.

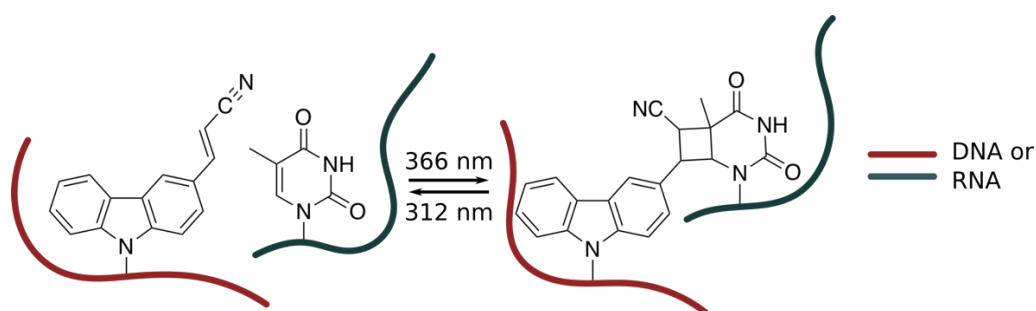
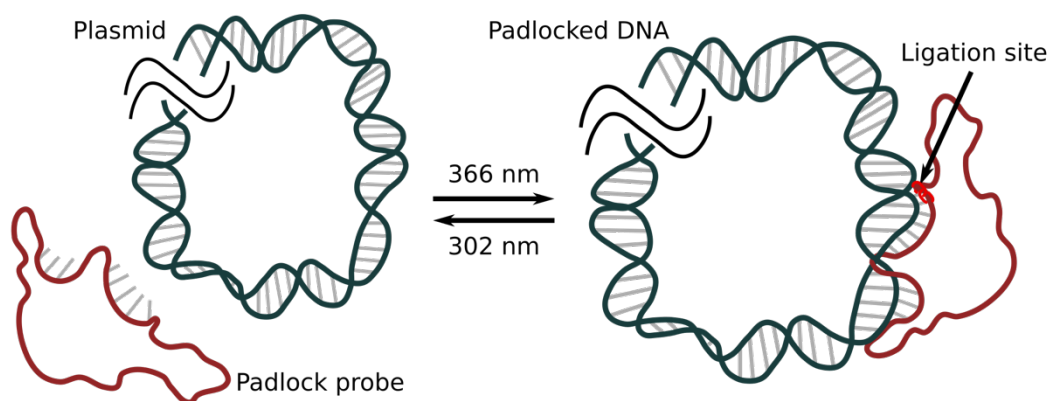


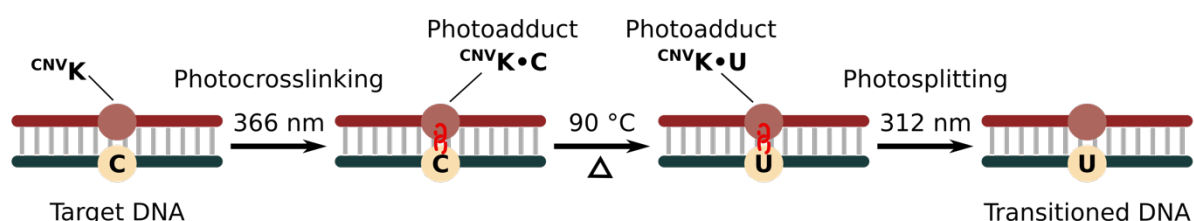
Figure 3.2 | Schematic representation of ^{CNV}K crosslinking. ^{CNV}K undergoes a [2+2] photocycloaddition with a pyrimidine base, thymine in this case, by photoirradiation at 365nm. The reaction is reversible and photo splitting occurs at 312 nm.

In addition to the development of cyanovinylcarbazole, Fujimoto *et al.* also explored the use of 5-carboxyvinyl-2'-deoxyuridine (^{cv}U), which is now a commercially available monomer.¹⁷⁷ They used the monomer to develop a highly efficient method for the formation of padlock probes, which are catenated plasmids. A modified OGN containing a 5'-^{cv}U was reversibly photo-ligated around the target sequence of the double-stranded plasmid DNA resulting in formation of a padlock probe (Figure 3.3a).¹⁹⁵ Irradiation at 302 nm resulted in the photo-ligated nucleic acids reverting to their original form. These padlock probes can act as inhibitors to block the DNA–protein interaction and hence, hinder transcriptional initiation at sites where the padlock probes bind to the plasmid DNA site-specifically.¹⁹⁶ They can also discriminate gene sequences with single base mutations due to the high sequence specificity of the padlock and serve as templates for rolling circle replication.¹⁹⁷

A Generation of Photo-ligated Padlock Probes



B Artificial Site-Specific Conversion of Cytosine into Uracil



C Using CNVK as an Antisense Strategy

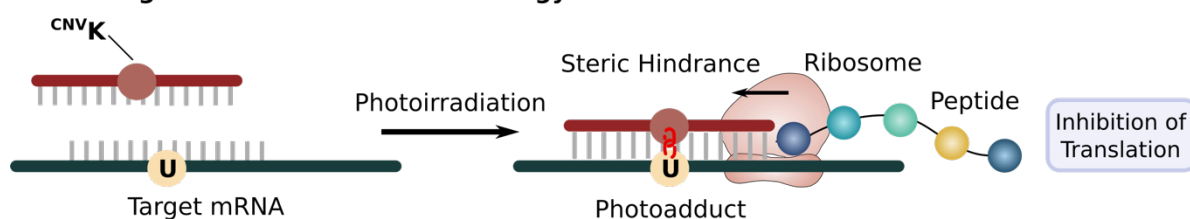


Figure 3.3 | Applications of CNVK -modified oligonucleotides. A) is a schematic representation of reversible photopadlocking, whereby a circular OGN forms a triplex with a double plasmid and is locked in place via photo-crosslinking. B) shows an overview of photo-crosslink based cytosine deamination involving CNVK crosslinking to a cytosine base, followed by heating at 90 °C to convert cytosine into uracil and lastly photo splitting at 312 nm. C) shows a schematic drawing of gene regulation using an AS-OGN containing CNVK , which acts as a physical obstacle at the ribosome after crosslinking.

They also explored the crosslinking of an RNA strand to a modified OGN containing artificial DNA bases and have successfully reported a method for the reversible RNA photo-crosslinking reaction *via* an artificial CNVK DNA base in an RNA-DNA heteroduplex.¹⁹⁸ They proved that the heteroduplex reverts back to the original RNA after irradiation at 312 nm and that the photo-crosslinking reaction is highly specific. The ultra-fast reactivity (1s), selectivity and reversibility of RNA-DNA photo-crosslinking means that it can be used for inhibiting transcription and eventually

translation; it can be used for the pinpoint fluorescent labeling of target RNA or the quantification of a multitude of non-coding RNAs.

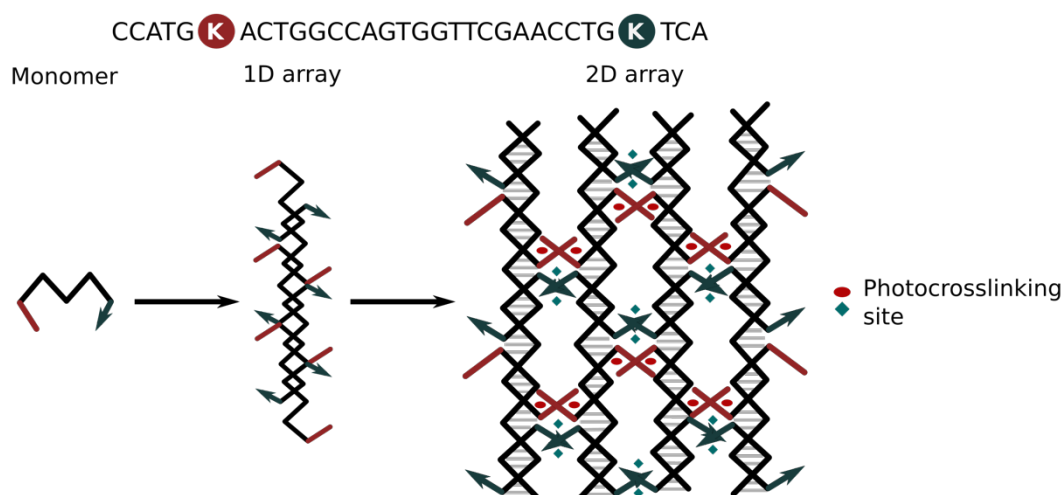
Further research in the area led to the development of a template-directed procedure for the artificial site-specific conversion of cytosine into uracil.¹⁹⁹ Initial attempts involved a photoligation reaction *via* the [2+2] cycloaddition between 5-carboxy-vinyldeoxyuridine (^{CV}U) and cytosine and the deamination of the photoligated dimer on heat treatment.²⁰⁰⁻²⁰² The [2+2] cycloaddition induces a change in the local conformation of the duplex and also in the hydrophobicity surrounding the cytosine base. Consequently, this results in a change in the solvent exposure of the cytosine base and this effect accelerates the deamination of cytosine.²⁰⁰ However, the use of ^{CV}U required photoirradiation on a timescale of hours. The same concept was repeated using ^{CNV}K instead of ^{CV}U, which undergoes [2+2] cycloaddition with cytosine in seconds. A modified DNA template containing a ^{CNV}K was reversibly photoligated to a cytosine on the complementary strand followed by heating at 90 °C, to convert the cytosine to uracil (Figure 3.3b). Photolysis at 312 nm released the DNA template. This procedure can precisely pinpoint the target cytosine for conversion into uracil, which is challenging when other methods based on small chemical reagents, such as bisulfite, are attempted.²⁰³⁻²⁰⁵

In 2012, the use of ^{CNV}K as part of an antisense strategy was also established.²⁰⁶ An antisense OGN containing ^{CNV}K was designed to photocrosslink to the complementary point mutated K-ras mRNA, which causes irregular cell growth with the perturbation of the EGFR/Ras signaling pathway (Figure 3.3c).²⁰⁷ RT-PCR experiments revealed ^{CNV}K-antisense OGN clearly photo-crosslinks to its target mRNA in a sequence

selective manner and then the function of mRNA as the substrate of reverse transcriptase is clearly regulated with only 1 s photoirradiation. In addition, the translation activity of target mRNA is clearly inhibited by photoirradiation, suggesting that ^{CNV}K-antisense OGNs have the potential to be effective photodynamic antisense drugs that act at desired locations around the body with a few seconds of photoirradiation.²⁰⁶

^{CNV}K has also been used in the construction and stabilisation of nanostructured DNA.²⁰⁸ There are various applications for DNA nanostructures such as nanopatterning, studying conformational changes, enzymatic reactions or computer systems.^{189-190, 209} Recent years have also seen the attempts to make a molecular robot, using DNA, RNA and a protein.²¹⁰ The main challenge to develop DNA nanostructures is the high thermal stability required. DNA melts at around 55 °C in solution, which limits the applications of DNA nanostructures.¹⁹² Thermal stability has been improved in DNA duplex by enzymatic ligation method and using artificial nucleotides such as PNA and LNA.²¹¹⁻²¹³ However, those solutions are still imperfect as enzymatic conditions require the use of buffers and modified nucleotides increase the thermal stability by only a few degrees. Therefore, the use of ^{CNV}K was explored, to make a 2D DNA array (Figure 3.4a). The resulting ultrafast photo-crosslinked DNA array had a very stable structure and was not broken under heating or denaturing conditions. Moreover, the position and the number of ^{CNV}K allows control over the size and conformation of the DNA array.²⁰⁸ The construction of the heat resistant DNA array is useful for making practical DNA nanodevices.

A Use of ^{CNV}K to make a 2D DNA array



B Ligation of DNA-gold nanoparticle assemblies

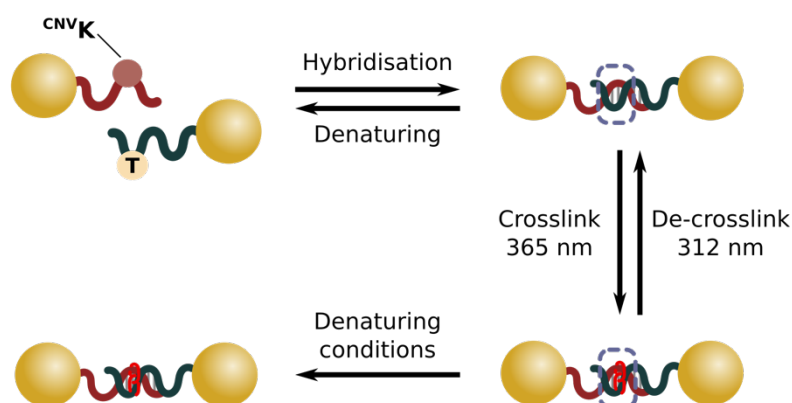


Figure 3.4 | Applications of ^{CNV}K-modified oligonucleotides (continued). A) displays a 2D DNA array, made by annealing and crosslinking an OGN containing two ^{CNV}K modifications, resulting in a heat resistant array. B) illustrates reversible photoligation of DNA-nanoparticle dimers. ^{CNV}K is incorporated in one of the DNA sequences and is crosslinked to a thymine base on the opposite strand such that the resulting dimers remain intact even under DNA denaturing conditions.

Similar to psoralen derivatives, ^{CNV}K has also been used to ligate DNA-gold nanoparticle assemblies.²¹⁴ Two batches of 15 nm gold nanoparticles were conjugated to complementary oligonucleotides, respectively. A ^{CNV}K nucleoside was incorporated in one of the DNA strands and the complementary strand was designed such that a thymine base would be present diagonally opposite the ^{CNV}K after hybridisation. Irradiation at 365 nm crosslinked the two strands together to form the DNA-gold

nanoparticles assemblies. Irradiation at 312 nm reversed the reaction and demonstrated a new approach to reversibly ligating nanoparticles without the need for special pH conditions or the use of additional molecules or ions (Figure 3.4b). This non-invasive approach unravels a new way to manipulate DNA-nanoparticle assemblies and can be applied to several fields such as catalysis, sensing, and photonics.²¹⁵⁻²¹⁶

As depicted by the above examples, ^{CNV}K has been successfully applied to numerous fields and has contributed to solving several scientific problems. Nevertheless, ^{CNV}K modifications are yet to be applied in the CRISPR field. In this chapter, we discuss how ^{CNV}K can help find a solution to the growing costs of modified gRNAs. By incorporating ^{CNV}K into our design, the only long RNA required will be an unmodified sgRNA, which can be easily and cost effectively made by *in vitro* transcription.

3.1.2 Modified CRISPR Cas9 systems

3.1.2.1 Modifications on Cas9

As elaborated in Chapter 2, Cas9 has been extensively modified for genomic imaging. Throughout the years, Cas9 has been subjected to several other modifications, some that goes beyond gene editing. One such application is the use of modified dCas9 as gene regulators. Conjugating certain DNA elements to dCas9 can repress transcription by sterically hindering the RNA polymerase machinery, a term coined 'CRISPR interference' (CRISPRi).¹¹² dCas9 was tethered to transcription repressor domains such as that of KRAB and has successfully been used to silence single genes and non-coding RNAs by targeting promoter regions, proximal and distal enhancer elements and 5' untranslated regions (Figure 3.5a).^{69, 114, 217-218}

Likewise, dCas9 can also be tethered to activator effectors, for gene activation, a term called 'CRISPR activation' (CRISPRa) (Figure 3.5b).⁶⁹ dCas9 has been fused to transcription activator domains such as the nuclear factor-κB transactivating subunit (p65) or to VP64.^{113, 219-220} gRNA modified with these synthetic transcription factors have successfully mediated synergistic gene activation. Different activator domains can also be combined, and multiplexed activation of endogenous genes can be used for cellular reprogramming.^{150, 221-223}

Epigenome editing such as acetylation and methylation of histones and methylation of DNA can also be carried out using modified CRISPR-Cas9 systems.²²⁴ By fusing dCas9 to the catalytic core of histone acetyltransferase p300, robust gene activation was successfully achieved *via* the acetylation of histone H3 Lys27.¹¹⁵ DNA demethylation was also successfully carried out by fusing dCas9 to the catalytic domain of methylcytosine dioxygenase TET1 and targeting the *BRCA1* promoter, resulting in transcription upregulation (Figure 3.5c).²²⁵

Modified dCas9 has also been used to study chromatin interactions. dCas9 fused with tag proteins such as FLAG tag can be used to identify proteins interacting with specific genomic loci by immunoprecipitating the chromatin targeted by the dCas9-tag protein-gRNA, with an antibody. The method is called 'engineered DNA-binding molecule-mediated chromatin immunoprecipitation' (enChIP).¹¹⁷ To identify the locus-associated proteins, mass spectrometry is carried out post-precipitation. Alternatively, dCas9 can also be fused to APEX2, which is an engineered peroxidase that promiscuously labels nearby proteins with biotin. The biotinylated proteins are then 'pulled-down' and identified by mass spectrometry (Figure 3.5d).²²⁶⁻²²⁷

To aid the delivery of CRISPR-Cas9 systems, Cas9 has also been modified to include ligands such as asialoglycoprotein (ASGPr) (Figure 3.5e). The reasoning behind the work carried out by Doudna *et al.* was that by harbouring ASGPr ligands on Cas9, the selective uptake of the CRISPR cargo can be expected in cells expressing ASGPr on their surface (eg. HEPG2).²²⁸ By assembling an RNP with two ASGPr ligands on Cas9 and sgRNA, internalisation of the RNP was successfully shown *via* receptor-mediated endocytosis. Selective gene editing was also successful through the use of an endosomolytic agent to release the RNP from the endosomes.

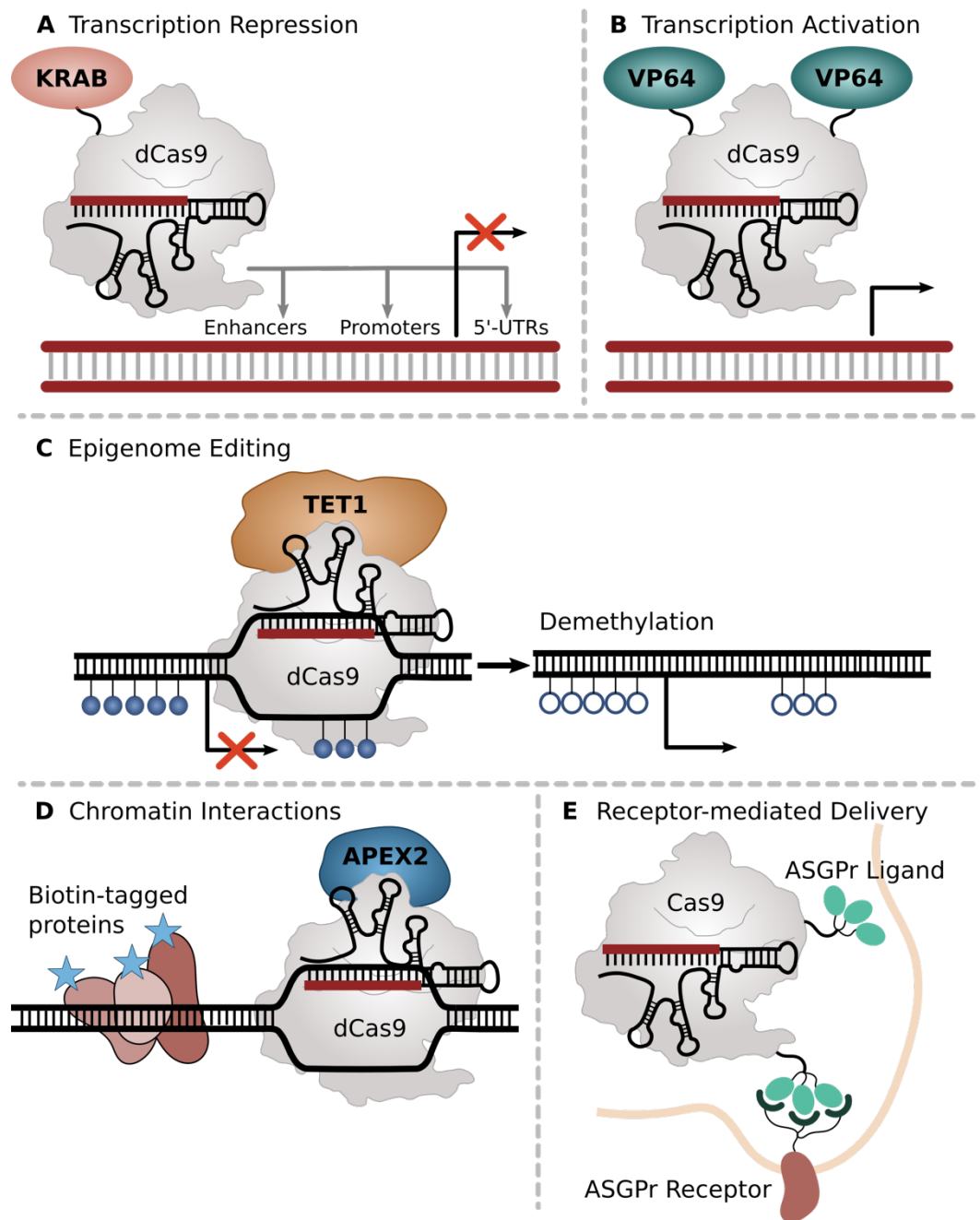


Figure 3.5 | Modified Cas9 and their respective functions. (A) shows dCas9 fused to KRAB to function as transcription repressor, targeting promoters, 5'-untranslated regions (UTRs) or enhancers. (B) illustrates transcription activation by fusing dCas9 to VP64 and resulting in the expression of transcription factors. (C) shows dCas9 fused to methylcytosine dioxygenase TET1 to reverse hypermethylation and silence genes. (D) is dCas9 fused to peroxidase APEX2 that can be used to biotinylate proteins in the vicinity of a targeted genomic locus; the proteins are then identified by mass spectrometry. (E) represents Cas9 conjugated to ASGPr ligand to aid receptor-mediated delivery specifically in cells with ASGPr receptors.

3.1.2.2 Modifications on the gRNA

Chapter 2 Section 2.1.2.2 explores the multiple ways fluorescently tagged gRNAs in combination with dCas9 can be used to specifically target genes of interest, to study real-time dynamics of genome editing, chromosome translocation, and transcription. gRNAs have also had additional functional RNA structures appended on to them such as aptazymes at the 5' end.²²⁹ Aptazymes are ligand-activated self-cleaving ribozymes that contain integrated aptamer domains.²³⁰ After binding to the correct ligands, aptazymes undergo structural changes that activate an associated ribozyme domain, triggering RNA cleavage. Aptzyme-embedded guide RNAs (agRNAs) allows a small molecule to control the structure of the gRNA and consequently genome engineering activity. The agRNA is designed with a 'blocking sequence', such that the agRNA is inactive. It is only in the presence of specific small molecules (eg. theophylline or guanine) that gene editing can resume, making the agRNA a controllable genome-editing agent (Figure 3.6a).

Similar to modified Cas9, gRNAs can also be used for controlling gene expression, by fusing viral RNA scaffolds at the 3'-end to recruit a variety of transcriptional activators, suppressors, or protein modifiers (Figure 3.6b). Well-characterised viral RNA sequences such as MS2, PP7 and com were integrated into the sgRNA, to which RNA-binding proteins MCP, PCP, and Com are recruited.²³¹ CRISPRa was successfully achieved in both yeast and human cells by fusing VP64 to each of the corresponding RNA-binding proteins. Likewise, CRISPRi was achieved by fusing KRAB to RNA-binding domains. Moreover, simultaneous ON/OFF gene regulation was also tested in human cells by targeting the endogenous gene *CXCR4* for activation with MCP-VP64 as well as targeting an additional endogenous gene

B4GALNT1 for repression with COM-KRAB. Simultaneous activation of *CXCR4* and repression of *B4GALNT1* was observed and the level of gene expression were similar to when single genes were targeted. The study showed that a single modified sgRNA could mediate both activation and repression.

In 2017, Lee *et al.* also investigated conjugating a donor DNA at the 5'-end of the gRNA as a single molecule (gDonor) to improve HDR (Figure 3.6c).¹²⁷ Using the gDonor ensures that a donor DNA is in close proximity to the site of cleavage for efficient HDR.²³²⁻²³³ Moreover, using the gDonor allows the simultaneous delivery of the gRNA and the donor DNA by polycations, which has proved challenging in the past due to the donor DNA and the Cas9 RNP have different charge densities, and it is therefore unlikely that a single nanoparticle will contain both Cas9 RNP and donor DNA.¹²⁷ As a result, it is challenging to deliver Cas9 RNP and donor DNA simultaneously into the same cell and efficiently induce HDR.

Modified gRNAs can be applied to various applications to solve crucial problems; from controlling gene expression to genomic imaging and potentially easing the process of delivering CRISPR-Cas9 RNPs by conjugating ligands onto the sgRNA. However, the cost of making these modified gRNAs are a massive obstacle, preventing the testing of these hypotheses. In this chapter, we discuss how a combination of crosslinking and chemical modifications can help to reduce the cost of modified sgRNA synthesis and help to find solutions for fundamental biological problems.

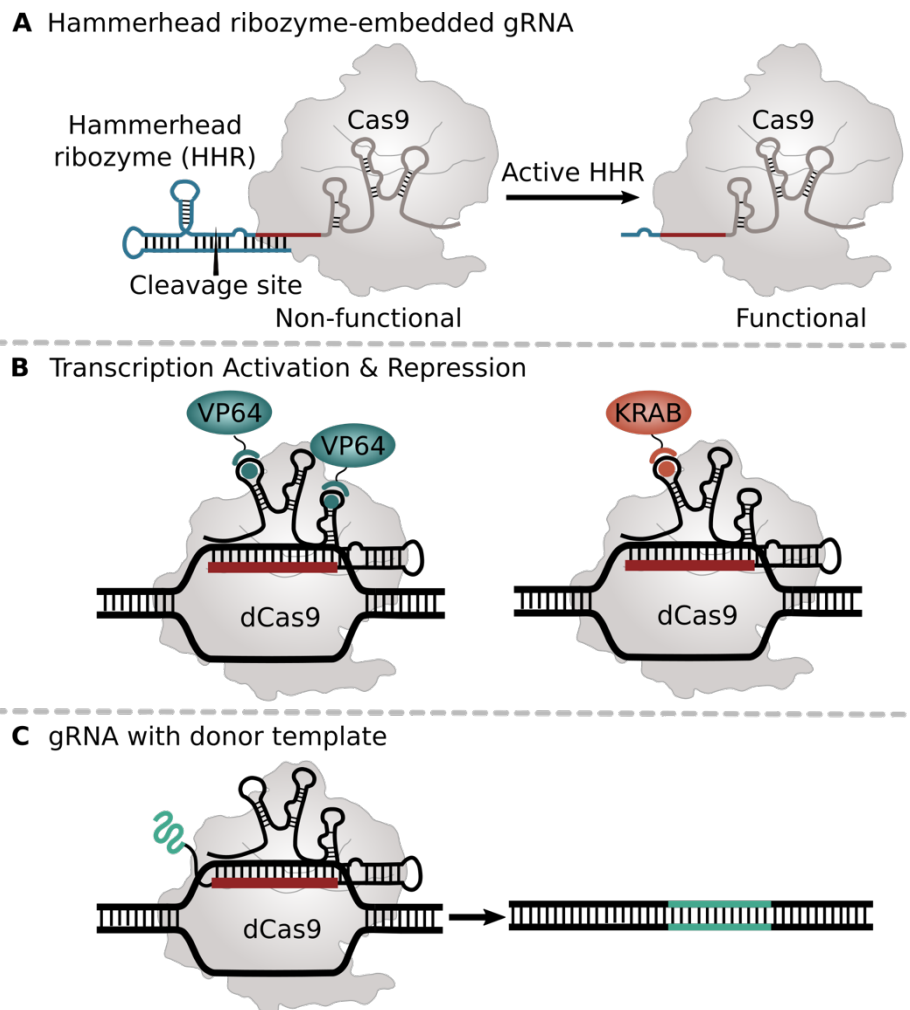


Figure 3.6 | Modifications on the sgRNA to expand the functions of CRISPR. (A) is a non-functional hammerhead ribozyme-embedded gRNA, that gets activated after being exposed to specific ligands such as guanine. (B) shows gRNAs modified with scaffolds designed to recruit transcription repressors or activators to control gene expression. (C) is a modified gRNA, with a donor DNA strand conjugated to improve the probability of HDR over the error prone NHEJ.

3.2 Hypotheses and Aims

We aim to make access to modified sgRNAs considerably easier by using an extended sgRNA with a ‘tail’. A short oligonucleotide (OGN), complementary to the tail, will be designed with a ^{CNV}K and a 5'-amino group (or any other modifications) such that it can be crosslinked onto the extended sgRNA. The short OGN can be post-synthetically modified with various functionalities such as fluorophores or ligands and the extended sgRNA can be made efficiently and cost-effectively by *in vitro*

transcription. In addition to be a high-yielding method, this process will also result in purer modified sgRNA as shorter OGNs are easily purified by HPLC and the crosslinking reaction is very clean.

We will explore the conjugation of four different moieties in this chapter: Cholesterol, Folic Acid and ASGPrL for transfection-free (gymnotic) delivery of CRISPR-Cas9 RNP into cells and Cy3 modification for live cell imaging of specific genes. ASGPrL modified sgRNA will provide specific uptake in cells expressing ASGPr such as HEPG2 cells and folic acid modified sgRNA will be specially taken up by cells with folate receptors such as HeLa cells and cholesterol modified sgRNA can more generally be taken up by most cell lines. This study will allow us to develop a new way of incorporating modifications in sgRNAs as well as testing the functionalities of those modifications.

3.3 Design and Synthesis of the Modified sgRNAs

The DNA template for *in vitro* transcription of the extended sgRNA was made by chemical synthesis. The short OGN (20-mer) complementary to the tail was also made by chemical synthesis, with a ^{CNV}K monomer at position 10 and a 5'- amino-hexyl. The short OGN was first post-synthetically modified by NHS ester labelling, before being crosslinked with the extended sgRNA (Figure 3.7).

For the folic acid-modified sgRNA, folic acid NHS ester was synthesised according to the methods described in section 6.19.2 and then conjugated with the amino-modified OGN. For cholesterol and ASGPr ligands-modified sgRNA (ASGPrL-sgRNA), 6-azido-hexanoic NHS ester was first synthesised according to the methods described in section 6.19.1 and then conjugated with the amino-modified OGN. The commercially

available alkyne-cholesterol was then clicked onto the OGN. DBCO-ASGPrL was kindly gifted by Gerald Chik and was also clicked onto the OGN. Finally, to make the Cy3-modified sgRNA, the amino-modified OGN was coupled with the commercially available Cy3-NHS ester.

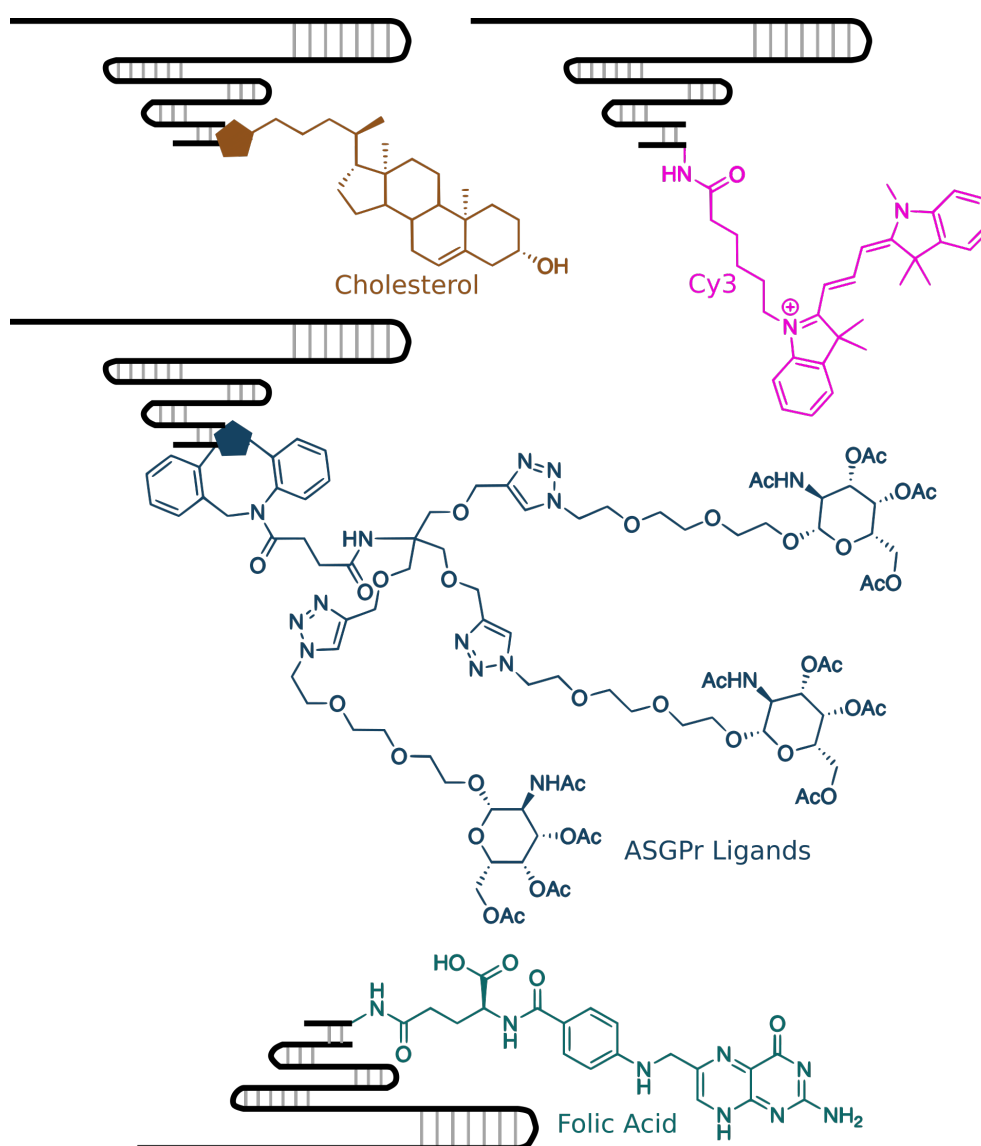


Figure 3.7 | Structures of the modifications. Cholesterol, Cy3, ASGPr ligands and folic acid conjugated to the sgRNA to add functionality.

The modified OGNs were mixed with the extended sgRNA and irradiated with 360 nm UV light for 15 min. The irradiation time was longer than previously reported protocols due to the long sgRNA forming multiple secondary structures and taking longer to

react. Therefore, the longer reaction ensured more efficient crosslinking and higher yield. Moreover, the prolonged irradiation time caused no damage to sgRNA when examined under gel-electrophoresis. The crosslinked-sgRNAs were purified on an 8% denatured gel-electrophoresis and extracting the bands corresponding to the desired product (figure 3.8a). The crosslinked products were confirmed by mass spectrometry (figure 3.8b).

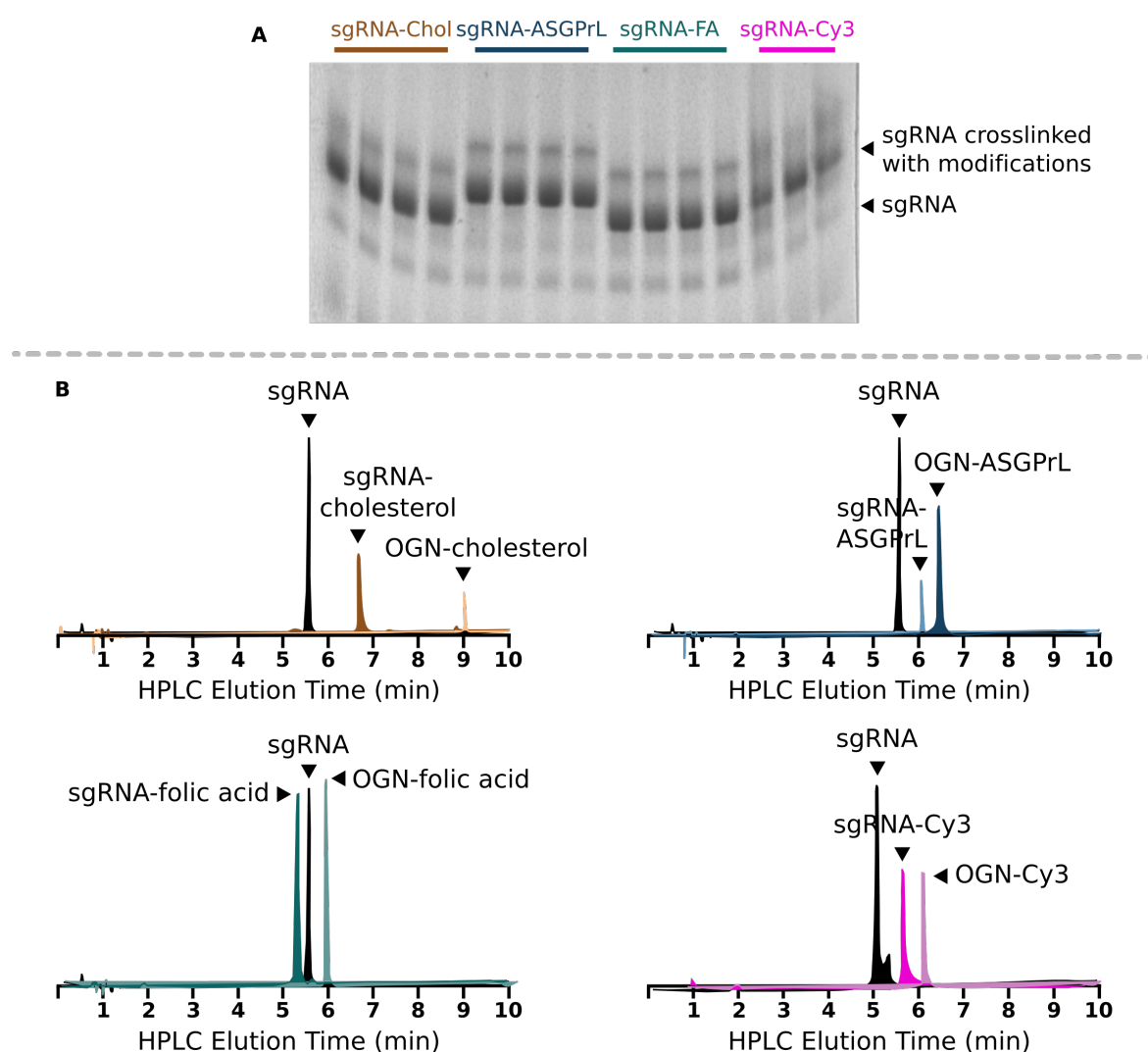


Figure 3.8 | Efficiency of crosslinking the modified OGNs onto the sgRNA. (A) The reaction was very efficient, regardless of the 5'-modification on the OGN. The crosslinking reaction proceeded for 15 min and were loaded onto an 8% denaturing PAGE gel. (B) Super-imposed mass spectrum of the sgRNA, the modified OGNs and the crosslinked OGNs, showing the difference in HPLC elution times. The oligonucleotide sequences can be found in Appendix 5.

3.5 *In Vitro* Cleavage Assay

The folic acid, cholesterol and ASGPrL modified sgRNAs will be used for gene editing after binding to Cas9. It is therefore crucial to test whether the modified sgRNA can still bind to the Cas9 enzyme and if they are still able to efficiently cleave the target DNA *in vitro*, despite having a crosslink in the middle of the structure and a bulky modification at the 3'-end. To evaluate whether the modifications were tolerated, *in vitro* DNA cleavage was performed using Cas9. The unmodified extended sgRNA produced prior to crosslinking was used as control. The reaction contained the sgRNA, Cas9 and a DNA template (plasmid) with a ratio of 50:10:1. As Cas9 is a single-turnover protein, the sgRNAs were used in greater excess to ensure that every Cas9 protein molecule forms a complex with the sgRNA and cleaves a target DNA.⁸⁵

It was found that all three modified sgRNAs were able to mediate DNA cleavage as efficiently as the unmodified sgRNA (Figure 3.9). Complete cleavage of the target was observed in 3 h. After establishing that the modifications and the crosslinking do not affect the efficiency of Cas9 target cleavage, the sgRNAs were tested in cells for their ability to enter the cells without the need for transfection reagents and for their ability to efficiently mediate gene editing.

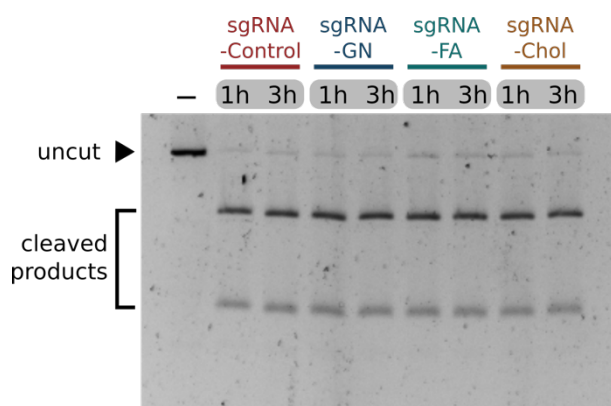


Figure 3.9 | In vitro cleavage of EMX1 target to test the effect of the modifications on the sgRNA. All targets were efficiently cut, regardless of the modifications present and the rates were comparable to the control (unmodified sgRNA). The DNA cleavage reaction contained 30 nM Cas9, 150 nM sgRNA and 3 nM linearised plasmid DNA, and proceeded for 1 h or 3 h at 37 °C. Samples were loaded onto a 1% SYBR gold-stained agarose gel. The oligonucleotide sequences can be found in Appendix 5.

3.5 In Cellulo Testing

3.5.1 Effect of the Constructs on Cells

The ability for the modified sgRNAs (with the exception of Cy3-sgRNA) to enter the cells were tested. Folic acid sgRNA was tested on HeLa cells, which have folate receptors, Cholesterol sgRNA was tested on U2OS cells and finally ASGPrL-sgRNA was tested on HEP-G2 cells, which express ASGPr. The RNP with Cas9 was first formed by mixing the modified sgRNAs with the protein. The higher the concentration of the RNP, the higher are the chances for the complex to go into cells. However, very high concentrations of RNP can be toxic to cells. Therefore, the optimum amount required had to be found. In the study reported by Doudna *et al.*, where a ASGPrL modified Cas9 was employed, the issue of the RNP being endocytosed was reported.²²⁸ The RNP was unable to escape the endosome and eventually, degraded *via* the lysosomal pathway. The solution to this problem was using the endosomolytic protein ppTG21, which has an isoelectric point of around 7 and is close to neutrally

charged at physiological pH to avoid cell penetration and thereby jeopardizing selectivity. The effect of ppTG21 on cells was also investigated in this study.

HeLa cells, U2OS cells and HEP-G2 cells were plated onto 24-well plates and grown until confluent. To the confluent cells, 250 pmols or 500 pmols of the modified sgRNAs were tested, in the presence and absence of ppTG21. Since the experiment can be conducted at 90-100% confluence of cells, the 'empty' space can be used as direct correlation to the amount of cell death. As shown in figure 3.10, both 250 pmols and 500 pmols of cholesterol modified RNP were well tolerated. The addition of ppTG21 resulted in some cell death. In the case of folic acid modified RNP, both 250 pmols and 500 pmols of the construct were well tolerated. The addition of ppTG21 did not result in any significant cell death in this case. Unfortunately, the results for ASGPrL modified RNP are not shown here as complete cell death was observed in all cases. This might be due to low viability of the batch of cells. Unfortunately, we did not have access to an alternate batch of HEP-G2 cells during the course of this study and further investigation could not have been carried out.

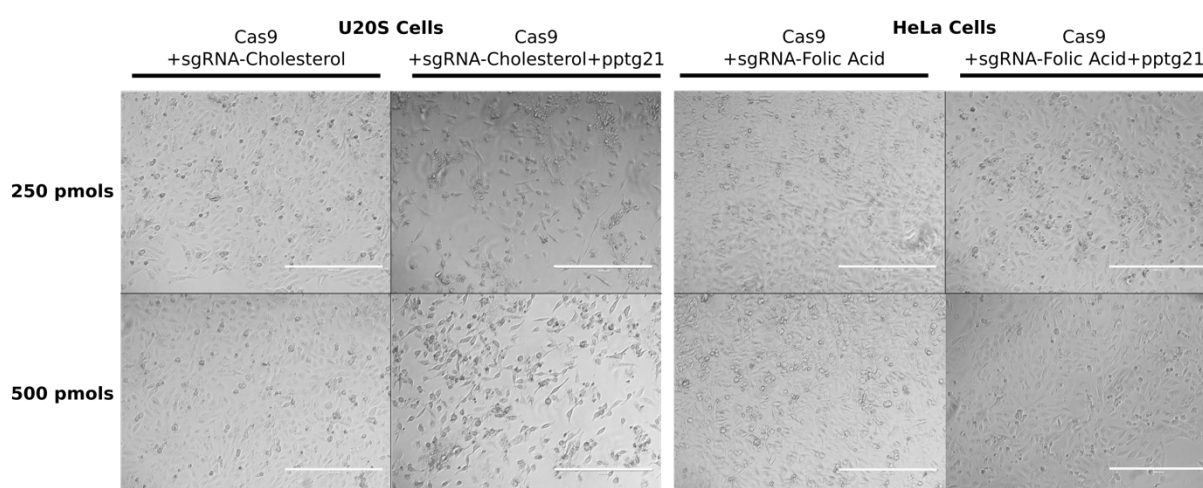


Figure 3.10 | Effect of 250 pmols or 500 pmols of modified sgRNA on cells in the presence or absence of ppTG21. U2OS cells tolerated both concentrations of sgRNA-Cholesterol well but addition of ppTG21 resulted in cell death. HeLa cells tolerated both concentrations of sgRNA-Folic Acid as well

and the addition of ppTG21 did not have a significant effect on cells. Cells were exposed to the modified sgRNAs for 16 h. Scale bar = 400 μ m.

3.5.2 *In Cellulo* Gene Editing Activity

3.5.2.1 T7E1 Assay

T7E1 assay will be used to monitor DSB and subsequent NHEJ-mediated indel mutations within the *EMX1* gene caused by the sgRNA-Cas9 complex. PCR amplification of the edited region generates amplicons with mixed length. Re-hybridisation of these amplicons yields imperfectly hybridised duplexes, which are recognised and cleaved by T7E1 endonuclease, a mismatch endonuclease. The more DNA cleavage, the greater the editing efficiency represented as % indels (where indels = $100 \times (1 - (1 - (b + c)/(a + b + c))^{0.5})$, where a is the integrated intensity of the uncleaved PCR amplicon, and b and c are the integrated intensities of each digestion product).¹⁰⁸

3.5.2.2 Gymnotic Delivery using the modified sgRNAs

The modified sgRNAs (with the exception of Cy3-sgRNA) were exposed to their respective cell lines. 250 pmols and 500 pmols of Cholesterol-sgRNA was mixed into the media of U2OS cells, in the presence and absence of ppTG21. The cells were incubated for 16 h, after which the T7E1 assay was carried out. The unmodified sgRNA was subjected to the same conditions as a control. Analysis of the results (Figure 3.11) revealed that there was no significant indel occurring in the cells. This means that the cholesterol-sgRNA was unable to enter the cells or it was unable escape the endosome once in the cell.

The same experiment was carried out for the folic acid-sgRNA and ASGPrL-sgRNA with HeLa cells and HEP-G2 cells respectively. There was no significant indels present in the culture exposed to the folic acid sgRNA either. Unfortunately, as explained in section 3.5.1, exposure to the ASGPrL-sgRNA resulted in cell death and data could not be acquired. Whether gymnotic delivery is possible with the modified sgRNAs is still under investigation. Future work will consist adding a tag onto the modified sgRNAs, to track them and find out if they are getting into cells. If the modified sgRNAs are not entering the cells, other variables need to be considered such as salt concentration in the sgRNA mixture, pH or whether the ligand modifications are being 'shielded' by Cas9 and are unable to interact with the cell membrane.

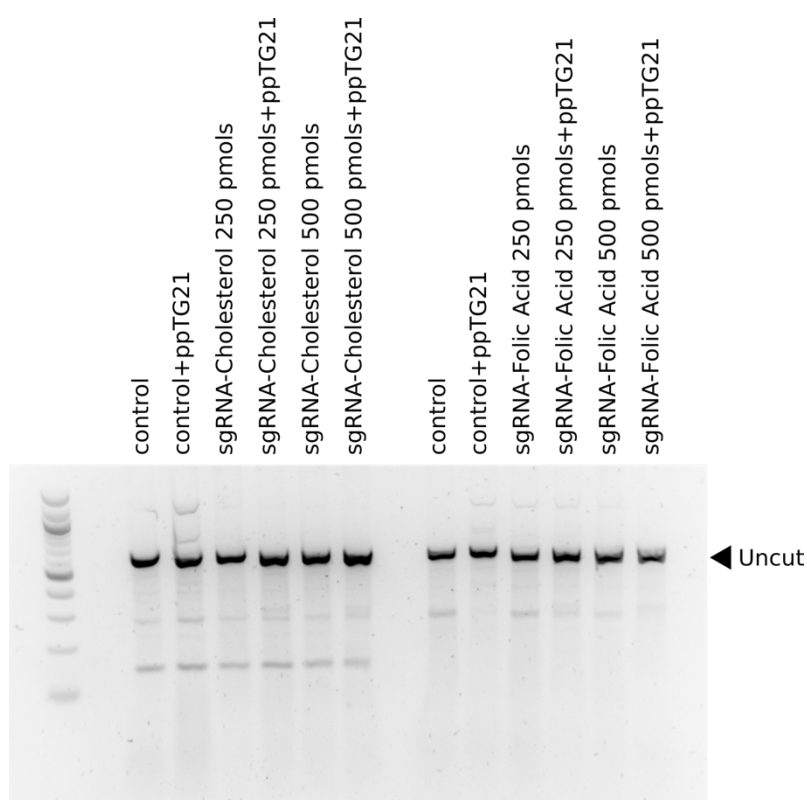


Figure 3.11 | *In cellulo* DNA cleavage activities of modified sgRNAs. The modified sgRNAs (and the controls) were added into cell media and incubated for 16 h and their ability to mediate indel formation was determined by T7E1 assay.

3.5.3 Live Cell Imaging

Cy3-sgRNA was tested in U2OS cells for its ability to label the telomeres. Cy3-sgRNA was transfected using lipofectamine reagents into cells expressing dCas9-3XGFP. Co-localisation between Cy3 and GFP will then be used to confirm positive target binding. As shown in Figure 3.12, the telomeres were successfully labelled by GFP by not by Cy3. The presence of granular cytoplasmic background is also prevalent under the Cy3 channel. This is a similar issue to that encountered in chapter 2, which suggests that part of the sgRNA where the Cy3 dye is present is getting degraded before it enters the nucleus. The degraded sgRNA is lacking Cy3 but is still functional and able to bind to dCas9-3xGFP and its target. To solve this issue, a similar approach to Chapter 2 should be considered where the short cy3-OGN is also modified with phosphorothioates.

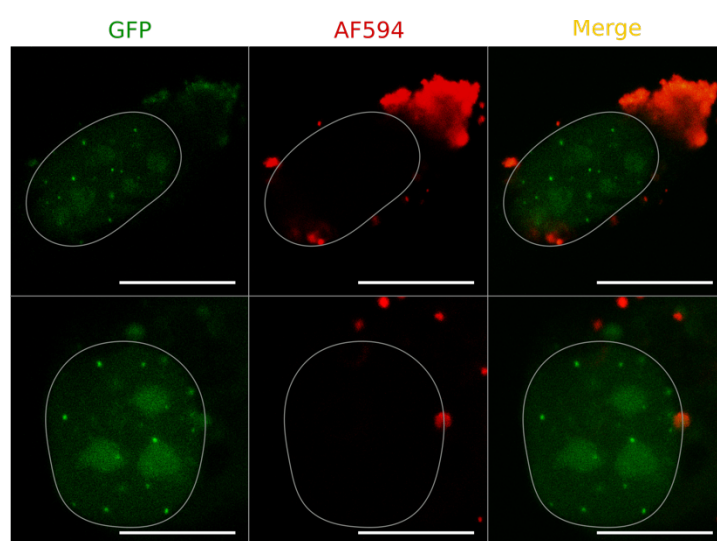


Figure 3.12 | Live cell imaging using sgRNA-Cy3. sgRNA-Cy3 was transfected into dCas9-3xGFP expressing U2OS cells using Lipofectamine reagents. After 16 h, live cells were imaged by confocal fluorescence microscopy using different lasers to visualise GFP (green) and Cy3 (red). Telomeric labellings were observed in the GFP channel but not in the Cy3 channel. Scale bar = 10 μ m.

3.6 Conclusion and Future Work

The main objective of this work was to devise a strategy to make the synthesis of modified sgRNAs simpler and more cost effective. The method described in this chapter consists of crosslinking an extended sgRNA made by IVT and a short OGN that bears all the modifications. Post-synthetic labelling of a short OGN is much more effective due to minimal secondary structures of the OGN and HPLC of short modified-OGN is relatively easier. Through this work, we have successfully developed a novel method for incorporating modifications onto the gRNA. Crosslinking of the OGN with the sgRNA was very efficient and high yielding. Both the crosslinked CNVK modifications and the 3'-modifications were well tolerated, as revealed by the *In vitro* Cas9 cutting assay.

The second part of the project consisted of testing the functions of the modifications in cells. Unfortunately, gymnotic delivery using the modified sgRNA has not yet been proven in this study. Future work will consist of fluorescently tagging the ligands-modified sgRNAs to check whether they are able to enter the cells. In case the sgRNA are unable to enter the cells, the modified sgRNA will be precipitated out after gel extraction to remove most of the salts (urea) present, which can affect the delivery. The amount of the modified sgRNA-Cas9 complex in the media can also be increased to higher than 500 pmols to increase the chances of the RNP going into the cells and inducing a DSB.

To improve genomic imaging using the crosslinked Cy3-sgRNA, the OGN bearing the Cy3 will be modified with a phosphorothioate backbone to increase cell stability and then crosslinked onto the sgRNA. This should make the probe stable enough to reach

the cells without degradation. Another solution that will be attempted will be using electroporation to transfect the Cy3-sgRNA-Cas9 RNP instead of lipofectamine reagents. The Cas9 will be able to protect the sgRNA against degradation and the RNP does not enter the cells through endosomes and therefore decreases the chances of the RNP being trapped in the cytoplasm.

Chapter 4

R-loop Structure Exploitation to introduce modifications

4.1 Background

4.1.1 Study of the R-loop structure

A defining feature of type I and type II CRISPR-Cas systems is the R-loop formation, which as detailed in Section 1.6.3, is a three stranded intermediate structure that consists of a DNA-RNA heteroduplex and a displaced DNA strand. Although the structures of Cas9 in complex with the target DNA and the sgRNA is well resolved, the mechanistic details on the formation of the Cas9-mediated R-loop structure remains elusive.²³⁴⁻²³⁵ In this chapter, we aim to exploit the DNA bubble formed during

R-loop propagation and prove that it can be key into finding a cost effective and efficient way to expand the functionalities of CRISPR-Cas9.

During the R-loop formation process, the spacer part of the crRNA invades double stranded DNA to form an RNA-DNA heteroduplex with the target DNA, while displacing the non-targeting strand, without requiring any external energy.^{92, 236-238} In type II CRISPR-Cas systems, the Cas9 endonuclease together with crRNA and a tracrRNA or an engineered sgRNA, is sufficient to form this R-loop.⁴⁹ A unidirectional model has been proposed whereby after the Cas9-gRNA complex recognises the PAM sequence, priming of the R-loop adjacent to the PAM leads to a unidirectional expansion towards the distal end of the protospacer (Figure 4.1).^{92, 239-240} Mutated PAM sequences or mismatches between the protospacer adjacent to the PAM and the crRNA had drastic effects on DNA cleavage, as opposed to distal mismatches, which further supported the model.⁸⁰ The model assumes that mutations within or adjacent to the PAM hinders R-loop formation. In other words, it suggests that the R-loop still forms with significant efficiency but displays a reduced lifetime, the magnitude of which depends on the position of the mutation.⁹²

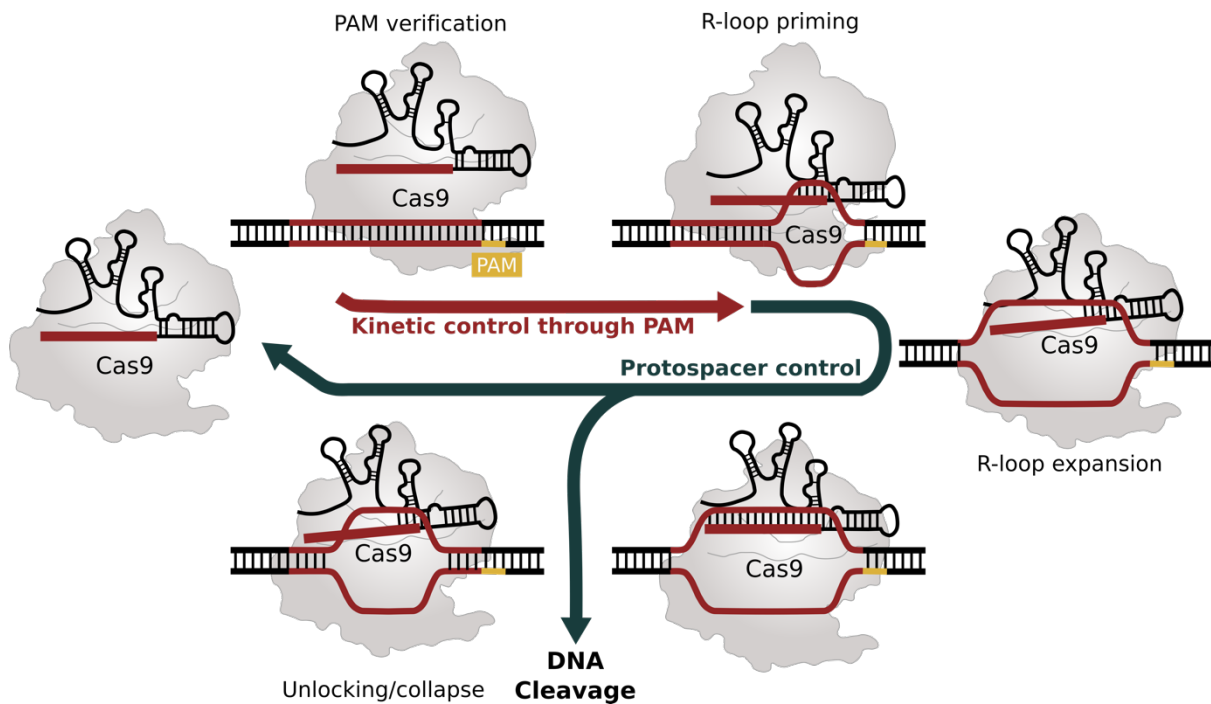


Figure 4.1 | Unified model for the formation and dissociation of the R-loop by Cas9-sgRNA RNP, without requiring ‘external energy’. The RNP carries out a homology search, guided by the PAM binding, which provides kinetic control over the R-loop formation. Complementarity (over 11 bases) between the protospacer and the target leads to DNA cleavage, otherwise the RNP leaves and the R-loop collapses.

In 2016, the use of cryo-electron microscopy (cryo-EM) shed further light on how Cas9, an individual non-polymerase enzyme was able to catalyse R-loop formation. The 3.4 Å resolution crystal structure of wild-type *Streptococcus pyogenes* Cas9 bound to sgRNA and a 30-bp target dsDNA containing a canonical 5'-TGG-3' PAM was solved, in order to determine the structural basis for R-loop formation. The bi-lobed architecture of the Cas9 enzyme has been mentioned in Chapter 1, whereby the bound target DNA resides within the central channel between the REC and the NUC lobes. Double-stranded DNA binding induces additional changes in the Cas9 structure that explain both HNH domain repositioning and R-loop stabilization. Upon binding of the dsDNA, the two hinge regions connecting the NHN domain with the RuvC domain, L1 (residues 765 to 780) and L2 (residues 906 to 918), are switched.⁸⁰ This rearrangement results in local changes in Cas9's secondary structure L1, which in

previous structures was completely distorted, and now refolds into a well-ordered loop. The refolding of the L1 is critical for the stabilisation of the R-loop, due to multiple interactions between L1 and the RNA-DNA hybrid minor groove and also the distal region of the non-target DNA strand (Figure 4.2).

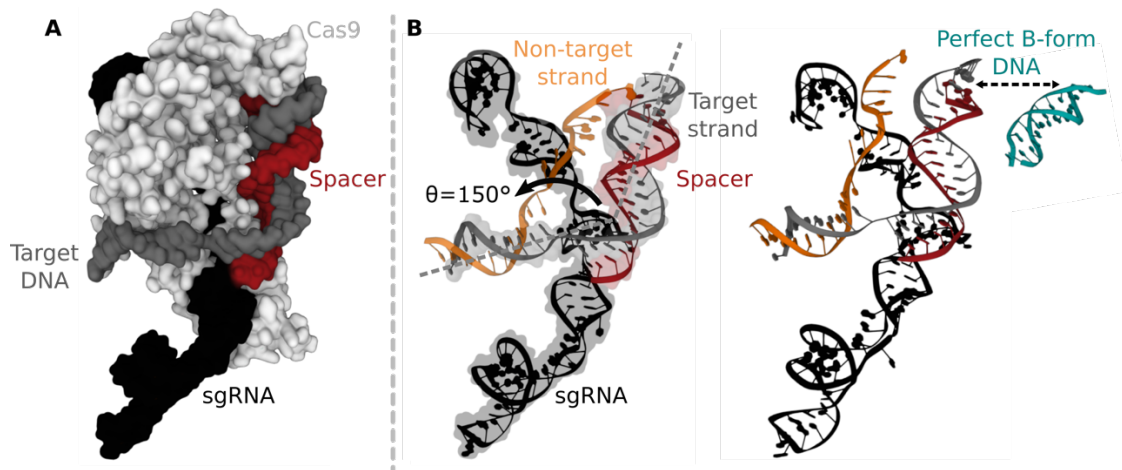


Figure 4.2 | R-loop surface representation. (A) shows the sgRNA (red and black), the target DNA strand (dark grey) and Cas9 with the HNH domain removed for clarity. (B) represents the pseudo-atomic model, with local bending of the dsDNA by $\sim 30^\circ$, creating an angle of 150° between ends (arrow) and the deviation from a linear helical duplex extending from the PAM-proximal end.

In spite of these advances, the structural dynamics of Cas9 and the molecular details of the R-loop formation were still not fully understood. In 2018, Zeng *et al.* carried out a combination of biochemical, single-molecule fluorescence energy transfer (smFRET) and molecular dynamics assays to further elucidate the structure of the R-loop.²³⁴ They found that the nucleotides at positions 3-7 from the PAM-proximal region in the target strand were the most sensitive to mismatches. In addition, the nucleotides of the sgRNA RNA in this region were shown to be immobilised by a highly conserved arginine-rich bridge helix in the high-resolution structures of the Cas9-sgRNA RNP in complex with and without the target DNA. Docking of a standard B-form dsDNA onto the Cas9-sgRNA RNP reveals a clear interference of DNA by the bridge helix. It was

therefore speculated that the initiation of R-loop structure requires the disruption of DNA duplex by this bridge helix and its associated sgRNA sequences.

They also demonstrated that the R-loop has a critical size for efficient dsDNA cleavage. Prior biochemical and single molecule experiments have shown that Cas9-sgRNA RNP can efficiently bind to its target with only 8 nt base pairing between the sgRNA and the PAM proximal region of the DNA target strand.²⁴¹⁻²⁴² However, efficient nicking of the DNA strand required 12 nt base pairing between the sgRNA and the target DNA and efficient DSB requires 14-17 nt base pairing.^{7, 243} The basis behind this R-loop propagation process were not well understood until the cleavage activities of Cas9 and its mutants were studied on DNA target sequences with 2–10 bp mismatches at the PAM distal end. The results revealed that only a 10-nt base pairing between the sgRNA and the target DNA could induce nicking of either the target or non-target strand. This implies that nicking only occurs once the Cas9-sgRNA stably binds to the target DNA after seed pairing and it can be speculated that the process of cleavage is highly coordinated. Cas9 adopts a 'partial active conformation' when base pairing is less than 12-nt, rather than an 'inactive conformation', that was previously reported.²⁴³ DSB occurs only once base pairing between the sgRNA and the target DNA reached 12 nt, suggesting that a minimum size of the R-loop structure is required for efficient target cleavage by Cas9 complex.

To study the dynamics of the R-loop structure in more details, smFRET was used to probe the R-loop propagation process, where the sgRNA is binding to the target DNA. The experiments consisted of variable guide sequences for the same target. The DNA was labelled with Cy3 on the target strand and Cy5 on the non-target strand such that

there is a decrease in FRET upon stable Cas9 R-loop formation. The experiments indicated that there exists an intermediate state before the fully stable R-loop complex. Kinetic analysis of this new intermediate state indicated that the lifetime of this state increases when the base-pairing length of guide-DNA duplex increases and reaches the maximum at the size of 18 bp. This was consistent with a previously carried out *in vitro* plasmid cleavage assay using Cas9 whereby the target cleavage reaction reaches highest efficiency at the R-loop size of 18-bp.

A different smFRET study on the dynamics motions of Cas9 HNH domain during target binding also reported that Cas9 adopted an intermediate state between target DNA binding and cleavage.²⁴⁴ In this study, it was also reported that any sequence mismatches between the DNA target and the sgRNA at PAM distal region prevents the transition from the intermediate state to the active confirmation. Zeng *et al.* therefore speculated that the structural intermediate of the R-loop corresponds to a Cas9 checkpoint intermediate before accessing its active conformation and initiating a DSB. Based on these observations, a new model for the formation of the R-loop was proposed (Figure 4.3).

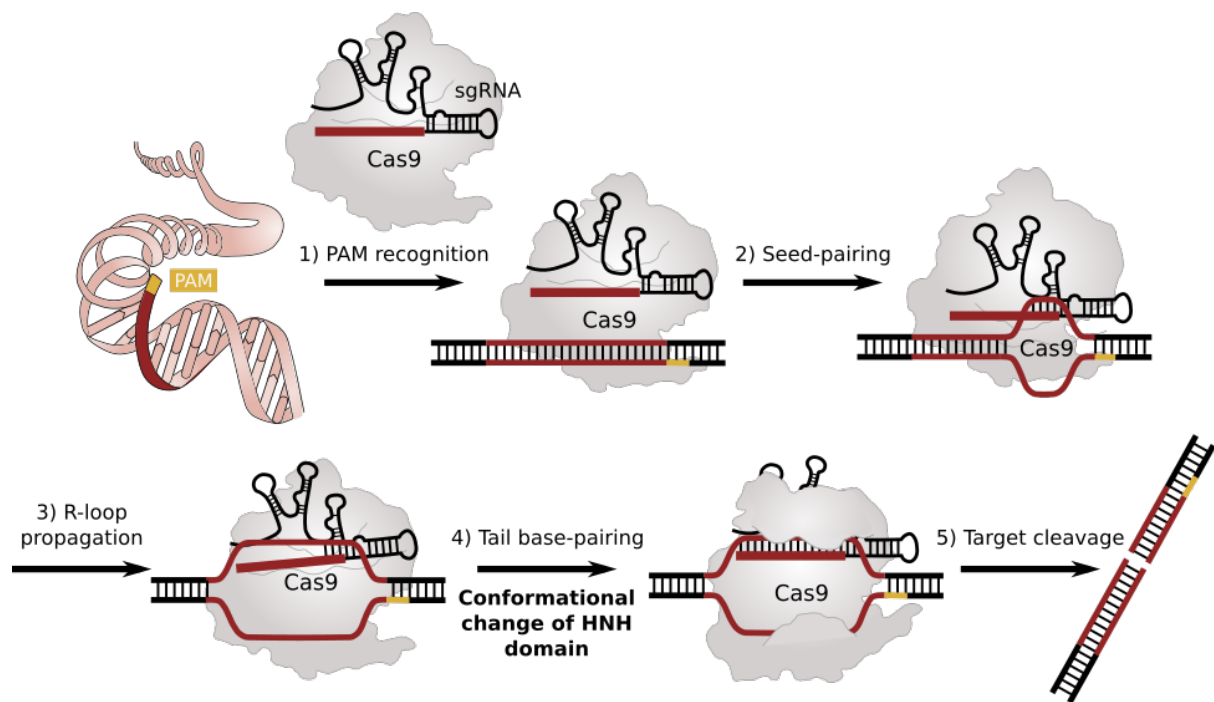


Figure 4.3 | Updated model for R-loop formation in CRISPR-Cas9 system. The first step consists of binding of the Cas9-sgRNA complex to the target DNA through transient PAM recognition. R-loop structure is then initiated by both PAM recognition motif and Arginine-rich bridge helix in Cas9 protein. The HNH domain adopts a 'partial active conformation' after seed pairing between the sgRNA and the target DNA. It is only after the R-loop structure reaches a critical size that the HNH domain adopts an active conformation and induces DSB.

While the studies described above set a foundation on the overall structure and dynamics of the R-loop and how it can be of potential applications in the future, the displaced ssDNA part of the R-loop is often overlooked. In this chapter, we aim to provide a proof of concept that the displaced ssDNA can be used to incorporate modifications (such as fluorophores) into the CRISPR-Cas9 system, instead of modifying the gRNA. There are a few examples where the DNA bubble played an important role in gene editing using CRISPR, which provides a good basis for our studies.

4.1.2 Role of the R-loop in 3'–3' Dual gRNAs

The R-loop structure played a key role in the work carried out by Dr Lapatrada Taemaitree, where a 3'–3' Dual gRNA was used to make 'staggered' cuts at the target site, in an attempt to improve the chances of HDR (Figure 4.4a). While CRISPR-Cas9 allows very efficient gene disruption, precise gene insertions remain a challenge due to the 'blunt' double stranded break induced by Cas9, which favours the error-prone repair, NHEJ over the more accurate HDR (Chapter 1, Section 1.3.2).

CRISPR-Cas12a system generates 5'-overhangs using only a single PAM site. However, the 5'-overhangs generated are complementary to each other, therefore the likelihood of ligation of the nicks to yield the original genomic sequence is high. Given that strand invasion by Cas12a at the target site displaces dsDNA and forms an R-loop structure, it was speculated that a second Cas12a–gRNA complex would not require a PAM site, if it could be colocalised with the first target site. In order to ensure that the second complex is present at the target site, two gRNAs were covalently linked by means of click chemistry (Figure 4.4b). The bi-functional guide RNA only requires a single PAM site and the unnatural 3'–3' linkage facilitates base-pairing to the second cleavage site. The system has the added advantage of having higher specificity by interrogating two recognition sites simultaneously.

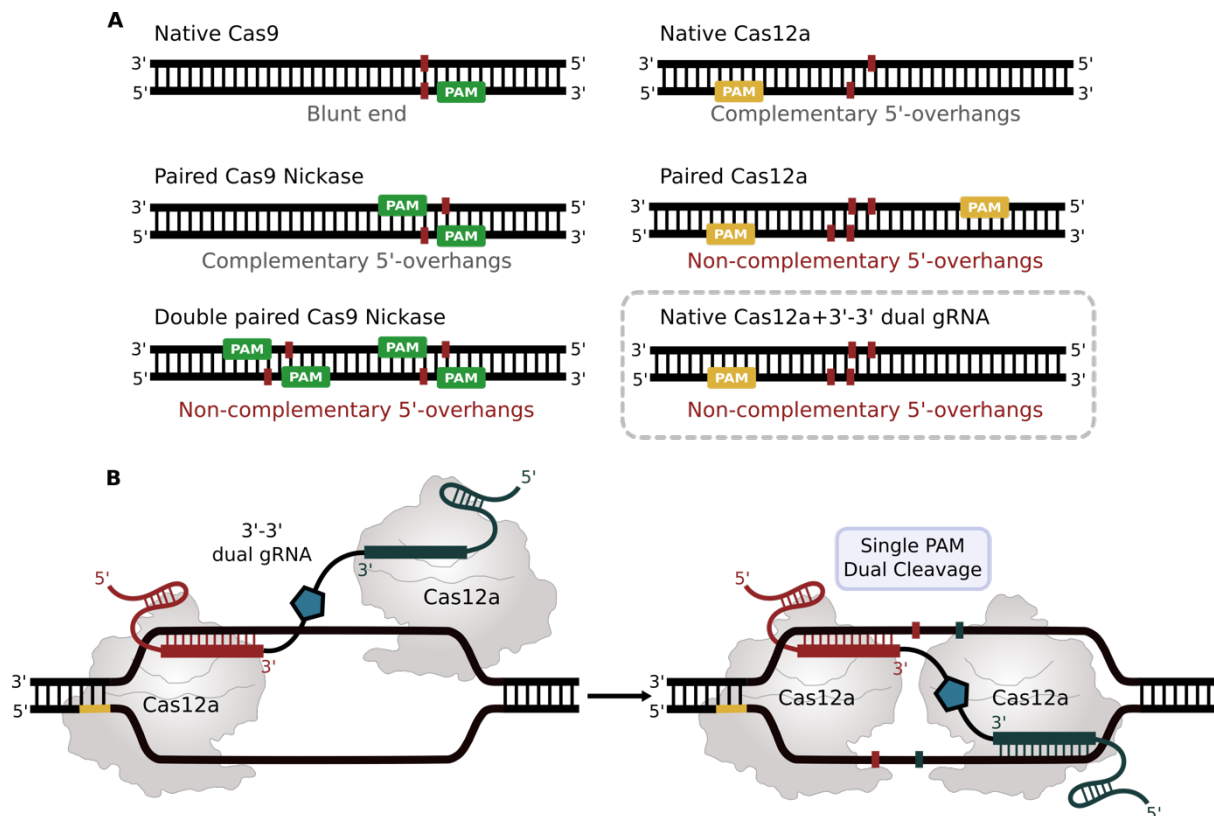


Figure 4.4 | Overview of Cas9 and Cas12a cleavage patterns and schematic of the DNA target cleavage by 3'-3' dual gRNA. (A) highlights the growing PAM motif dependence for generating non-complementary 5'-overhangs that favours DNA repair by HDR (B) Clicked guide RNA (crRNA) for Cas12a that reduces PAM dependence and facilitates the removal of short dsDNA to generate non-complementary 'sticky' ends.

Unfortunately, due to the non-specific ssDNase nature of Cas12a, it was challenging to prove the concept. Cas12a displayed a characteristic '*trans*-cleavage' (i.e. cleavage of non-hybridised DNA) as opposed to the usual '*cis*-cleavage' (i.e. specific on-target cleavage) normally observed when Cas binds to dsDNA or ssDNA. Misfolding of the RNA secondary structure due to close proximity of the second crRNA further prevented progress in the area.

Nonetheless, this concept shows that the R-loop structure is an attractive target, not just for Cas9 applications but also other variants of Cas. It is an indication that further

work needs to be carried out to find out more information on the R-loop structure such as its size, stability and other characteristics.

4.1.3 Role of the R-loop in Prime Editing

In 2019, a new method of gene editing known as prime editing was described. Prime editing is a versatile and precise genome editing tool that uses catalytically impaired Cas9 fused to an engineered reverse transcriptase, programmed with a prime editing guide RNA (pegRNA). The pegRNA directs Cas9 to the target site as well as contains new genetic information as a template to replace the target DNA nucleotides.²⁴⁵ In comparison to CRISPR-Cas9 mediated HDR, prime editing shows similar or higher editing efficiency and it produces fewer by-products. It also induces lower off-target editing compared to CRISPR-Cas9.

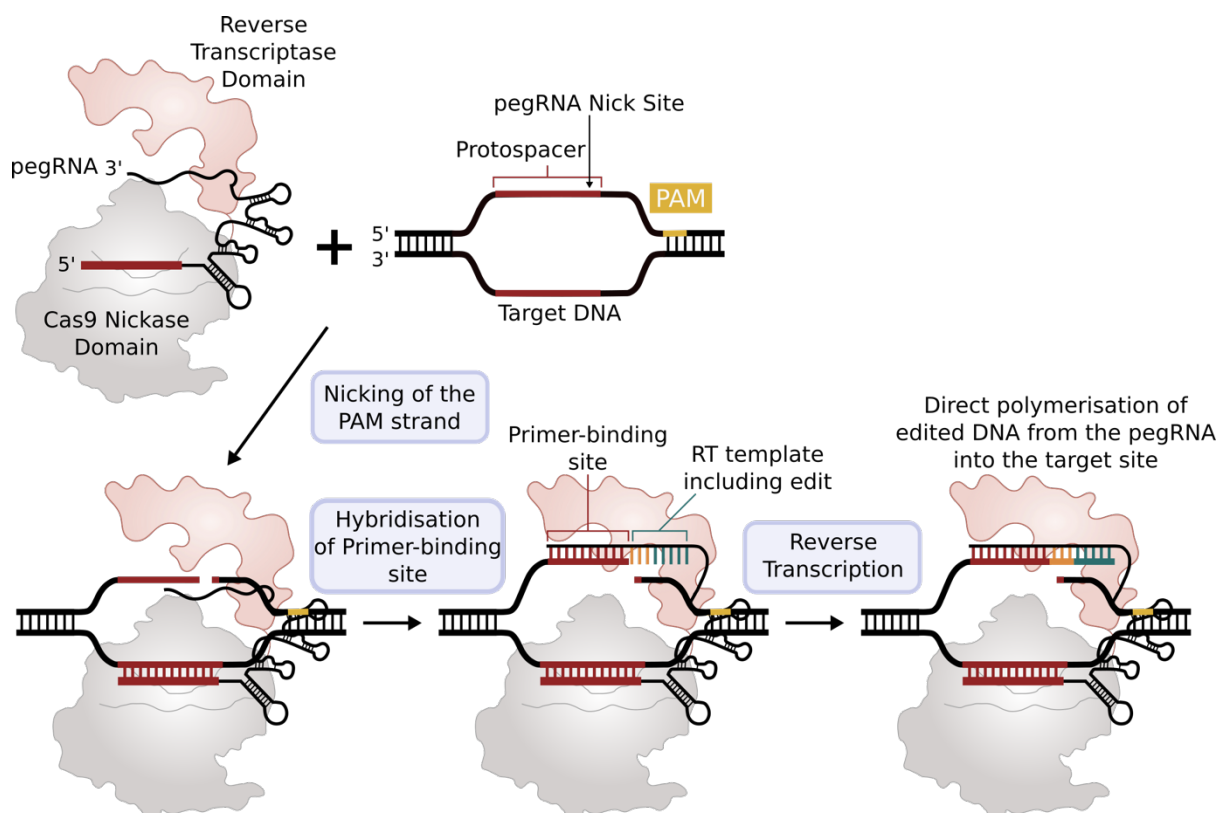


Figure 4.5 | Overview of Prime Editing by PE protein and pegRNA. The PE protein consists of a Cas9 nickase fused to a reverse transcriptase domain. The protein is complexed with a pegRNA. The

complex binds to its target DNA and nicks the PAM-containing strand. The 3'-end of the nicked strand hybridises to the primer-binding site of the pegRNA. The reverse transcriptase domain binds to the primer-binding site and starts the process of reverse transcription of new DNA using pegRNA as template. Equilibration between the edited 3' flap and the unedited 5' flap, cellular 5' flap cleavage and ligation, and DNA repair results in stably edited DNA

The prime editing complex consists of a prime-editing (PE) protein, which contains a Cas9 nickase domain as well as a reverse transcription (RT) domain, complexed with a pegRNA. The complex is guided to the target DNA by the pegRNA where dsDNA is unwound, and the R-loop structure is formed. The PE protein nicks the displaced strand in the R-loop, exposing a 3'-hydroxyl group (Figure 4.5). Hybridisation of the pegRNA onto the nicked strand primes the reverse transcription process, resulting in the incorporation of the desired edit, encoded within the pegRNA, directly into the target site.

The resulting intermediate is two redundant single-stranded DNA flaps: a 5' flap that contains the unedited DNA sequence and a 3'-flap that contains the edited sequence copied from the pegRNA. The unedited 5'-flap can then either be removed by endonucleases such as FEN1 or by 5' exonucleases such as EXO1.²⁴⁶⁻²⁴⁷ Prime editing has been successfully used to induce more than 175 edits in human cells, including targeted insertions, deletions and point mutations. The results provide a promising basis that the R-loop structure can be exploited and functionalised.

4.1.4 5'-extended sgRNAs

The most commonly employed sgRNAs are about 100 bp in length. The spacer sequence defines the genomic target of the sgRNA and is usually 20-nt long at the 5'-end. Studies involving longer spacer sequences in the guide RNA could provide an

estimate of the size of the DNA bubble. However, studies where this has been done is very limited and most of the time, truncated spacers such as 19–17-nt long spacers are studied.^{99, 248-249} In 2013, Zhang *et al.* designed a sgRNA with a 30-nt long spacer in an attempt to improve Cas9 targeting specificity. While the extended spacer did not improve specificity, it was still function and was able to induce indel.¹⁰⁸ Similarly, an independent study consisting of extending the sgRNA with a 30-nt, 40-nt and 53-nt long spacer sequence also showed that the extended sgRNAs were still able to form functional RNPs and cleave DNA targets.²⁵⁰ While these studies provide evidence that extended sgRNAs are still functional, they do not prove that the DNA bubble can accommodate more than 20 bp or whether it would still unravel enough to fit an independent OGN.

4.2 Hypotheses and Aims

In this chapter, we aim to develop a novel way of introducing modifications into the CRISPR-Cas9 system. The applications of modified CRISPR-Cas9 systems has been elaborated in Chapter 3 and more specifically, the importance of fluorescently labelled CRISPR-Cas9 systems has been discussed in Chapter 2. We are interested in proving that there is enough space in the R-loop structure, next to the bound sgRNA-Cas9 complex, where a short single stranded external oligomer (eOGN) can bind (Figure 4.6). This eOGN can bear all the modifications such as fluorophores to image specific genes or ^{CNV}K to introduce crosslinks at a specific target.

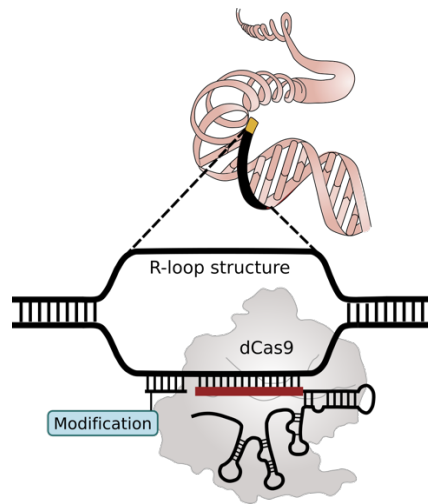


Figure 4.6 | Introducing modified oligos via CRISPR-mediated R-loop. Proposed method for the incorporation of modifications into CRISPR-Cas9 system. The method consists of using dCas9 in complex with an unmodified sgRNA to form the R-loop. A short eOGN, containing the desired modifications can then bind proximal to the sgRNA binding site.

The system will make use of an unmodified sgRNA, which can be made by *in vitro* transcription or it can be transcribed directly in cells. *In vitro* transcription is a cheap and easy way to make sgRNA in a large scale (Chapter 3, Section 3.1.2). However, making modified RNAs by *in vitro* transcription is not straightforward and the process is not high yielding. In this case, all the modifications will be incorporated into a short eOGN, which will be made by chemical synthesis. As opposed to long RNAs, short oligonucleotides can be made in high yields and large scale using chemical synthesis due to the excellent coupling efficiencies in shorter oligo strands. In addition, modified phosphoramidites can be incorporated at any desired position during the coupling. Shorter OGNs have the added advantage of being more stable and less prone to degradation by exo- or endonucleases.

Once the Cas9 (or dCas9) and sgRNA complex is formed in cells, it will be directed to its target and form the R-loop structure. Based on the work carried out in by Lui *et al* for Prime Editing, we are speculating that there will be enough space upstream of the

sgRNA binding site for a 12-mer OGN to stably bind. The 12-mer eOGN will be in large excess and will be co-transfected with the sgRNA. The displaced strand was not chosen as the target to avoid self-complementarity issues between the eOGN and the spacer sequence of the sgRNA, which will reduce the efficiency of binding of the sgRNA to the target, especially when targeting repeating sequences. However, the best position for the eOGN to bind in close proximity to the Cas9-sgRNA complex will need to be optimised.

4.3 Proof of Concept Using Live Cell Imaging

In line with our expertise and equipment available, live cell imaging was used to attempt to demonstrate that there is enough space in the R-loop structure for both the Cas9-sgRNA complex and the eOGN to bind.

4.3.1 Design Strategy

The modified eOGN was made by chemical synthesis, with an amino-c6 modification at the 3'-end. The oligo was made using 2'-OMe ribonucleotides for enhanced stability and to provide protection against nucleases in cells.¹⁶⁷ The eOGN was then post synthetically labelled *via* NHS ester coupling using Cy3 NHS ester. Four eOGN were designed, each 12 monomers long, and each binding one nucleotide further upstream of the dCas9-sgRNA complex: eOGN+1 was designed to bind three nucleotides upstream of the RNP complex, eOGN+2 was designed to bind four nucleotides upstream and so on (Figure 4.7). This will help to determine the size of the 'space' available next to the bound sgRNA and which eOGN binds more efficiently. The sgRNA targeting the telomeres was made by *in vitro* transcription and similar to

Chapter 2, dCas9-3xGFP plasmid will be transfected and expressed in cells before the co-transfection of the sgRNA and the eOGN. Co-localisation of GFP and Cy3 will provide proof that the eOGN is able to bind to its target specifically.

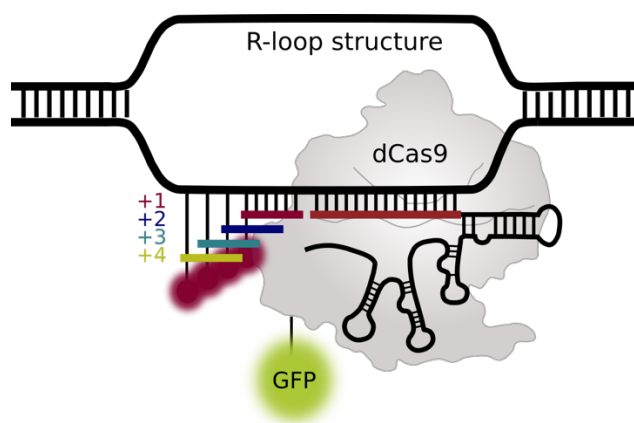


Figure 4.7 | Proof of concept of eOGN binding site using fluorescently tagged eOGNs and live cell imaging. The figure shows four Cy3-labelled eOGN, each binding 1-nt further from each other proximal to the sgRNA. This will help determine which eOGN binds more efficiently and thus how much 'extra space' is present on the R-loop and can also provide an alternate method to carry out genomic imaging.

4.3.2 Electrophoretic Mobility Shift Assay (EMSA)

Before attempting live cell imaging, the *in vitro* binding of the eOGN to the target DNA was investigated using the EMSA assay. EMSA is a rapid and sensitive method to detect protein–nucleic acid interactions.^{251–253} It is based on the observation that the electrophoretic mobility of a protein–nucleic acid complex is typically less than that of the free nucleic acid. Here, a non-denaturing PAGE was used to detect the ribonucleoprotein complex formation. The assay was used to find out whether the eOGN will bind to the target DNA after the dCas9-sgRNA complex unwinds dsDNA and binds to its target. For this purpose, the eOGN was labelled with a fluorophore, Cy3, so that the bands corresponding to the target DNA-eOGN-dCas9-sgRNA complex and the free eOGN are fluorescent and can be directly compared. Since the

complex is larger than the pore of the gel, the bound eOGN will remain in the well whereas the unbound species will migrate into the gel.

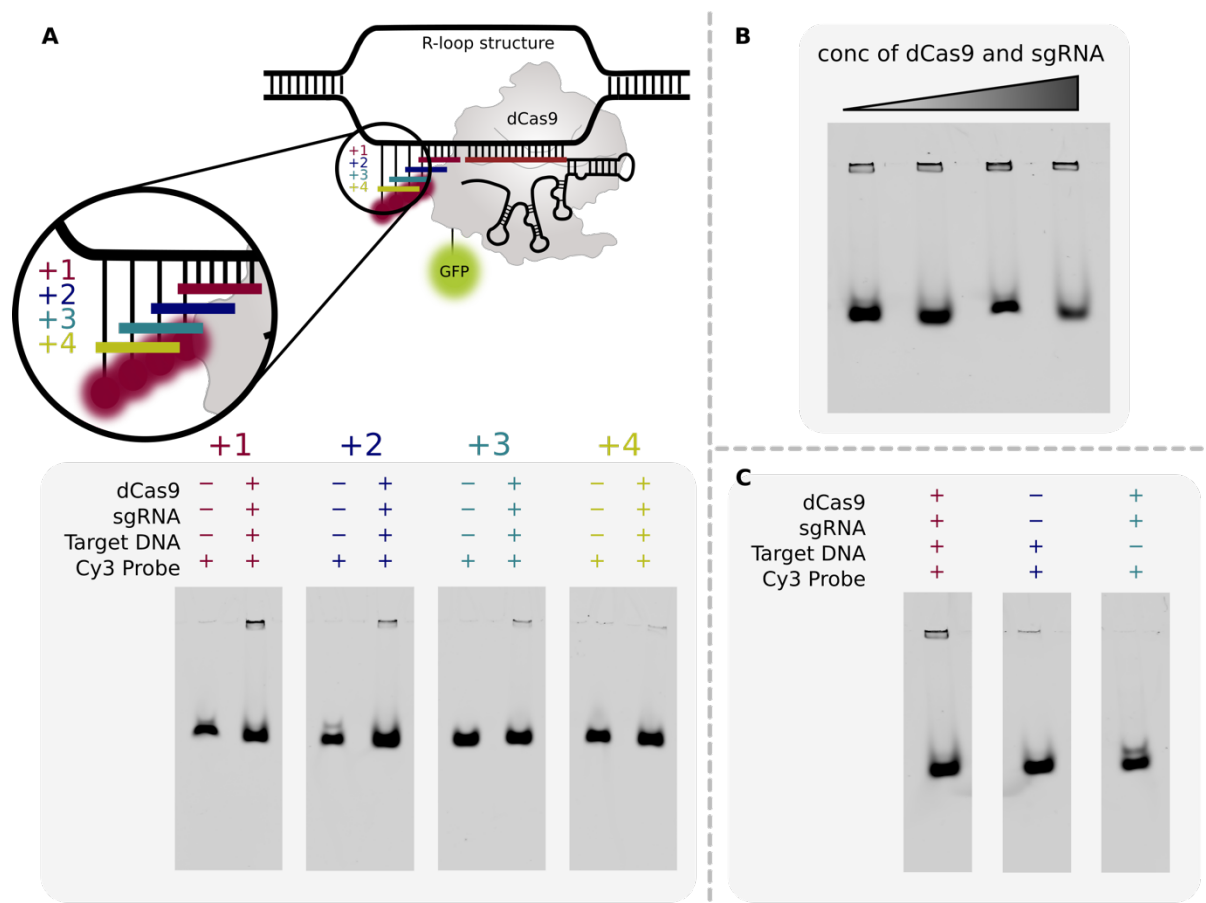


Figure 4.8 | EMSA assay to determine binding efficiency and specificity of the eOGNs. (A) shows the binding of eOGN+1–4 to the target DNA-dCas9-sgRNA complex. eOGN+1 shows the strongest binding efficiency. (B) shows an increase in binding efficiency on eOGN+1 upon increasing concentration of dCas9 and sgRNA (stronger top band and weaker lower band). (C) shows the binding specificity of eOGN+1 whereby the probe does not bind to the target DNA in the absence of Cas9-sgRNA (faint band present in image as artefact) and whereby the probe does not bind to Cas9-sgRNA in the absence of the target. The reaction was loaded on an 8% native PAGE gel and the efficiency of the reaction was determined by visualising under the Cy3 channel. The oligonucleotide sequence can be found in Appendix Table 7.

The binding of the four labelled eOGN was investigated and from figure 4.8a it is apparent that eOGN+1 and eOGN+2 bind the best *in vitro*. Next, the binding of the best performing eOGN, eOGN+1, was assessed by increasing the concentration of dCas9 and sgRNA. Since eOGN+1 and the target DNA was present in a large excess,

increasing the concentration of dCas9 and sgRNA should increase the binding. There was a slight increase in binding upon increasing dCas9 and the sgRNA concentration (Figure 4.8b). This shows that while binding to the target does occur, it is not very efficient.

Lastly, the specificity of the binding was assessed. The eOGN should only bind to the target in the presence of the dCas9-sgRNA complex, responsible for the formation of the R-loop structure. This is the case for eOGN+1, as shown in figure 4.8c. The binding was also assessed in the absence of the target DNA to rule out the possibility that eOGN+1 is binding non-specifically to the sgRNA or to dCas9. There was no band in the well in the absence of the target DNA, which concludes that eOGN+1 binds specifically to its target after dCas9-sgRNA forms the R-loop.

4.3.3 Live Cell Imaging

Since the EMSA assay showed a relatively good binding, especially with eOGN+1 and eOGN+2 and they displayed good specificity *in vitro*, the binding of the eOGN *in cellulo* was investigated. dCas9-3xGFP plasmid was first transfected using lipofectamine and expressed in U2OS cells. After 6 h, the sgRNA and each of the four Cy3-labelled eOGN (10 eq.) was co-transfected, also with lipofectamine. After 12-16 h, the transfected cells were visualised under a confocal microscope.

The puncta in the nucleus that can be seen under the GFP channel represent telomeric labelling by dCas9-3XGFP-sgRNA complex (Figure 4.9). The puncta in the Cy3 channel represent binding of the eOGN. Co-localisation of the two puncta indicates eOGN is binding to its target, the telomers. While successful telomeric

labelling was observed in the GFP channel and some red signals were observed in the Cy3 channel, there was no co-localisation between the signals from the two channels. The two puncta were close to each other but not perfectly aligned. This indicates that the eOGNs were not binding specifically to the telomeres. It was speculated that this was due to poor nuclear localisation of the eOGN, which meant that the eOGN accumulated in the cytoplasm where it was degraded. The smaller degraded eOGN entered the nucleus more easily and was binding to a different target, causing non-specific binding. If this was the case, the nuclear localisation and cellular stability of the eOGN needed to be improved and a common way of doing so is using phosphorothioate linkages.

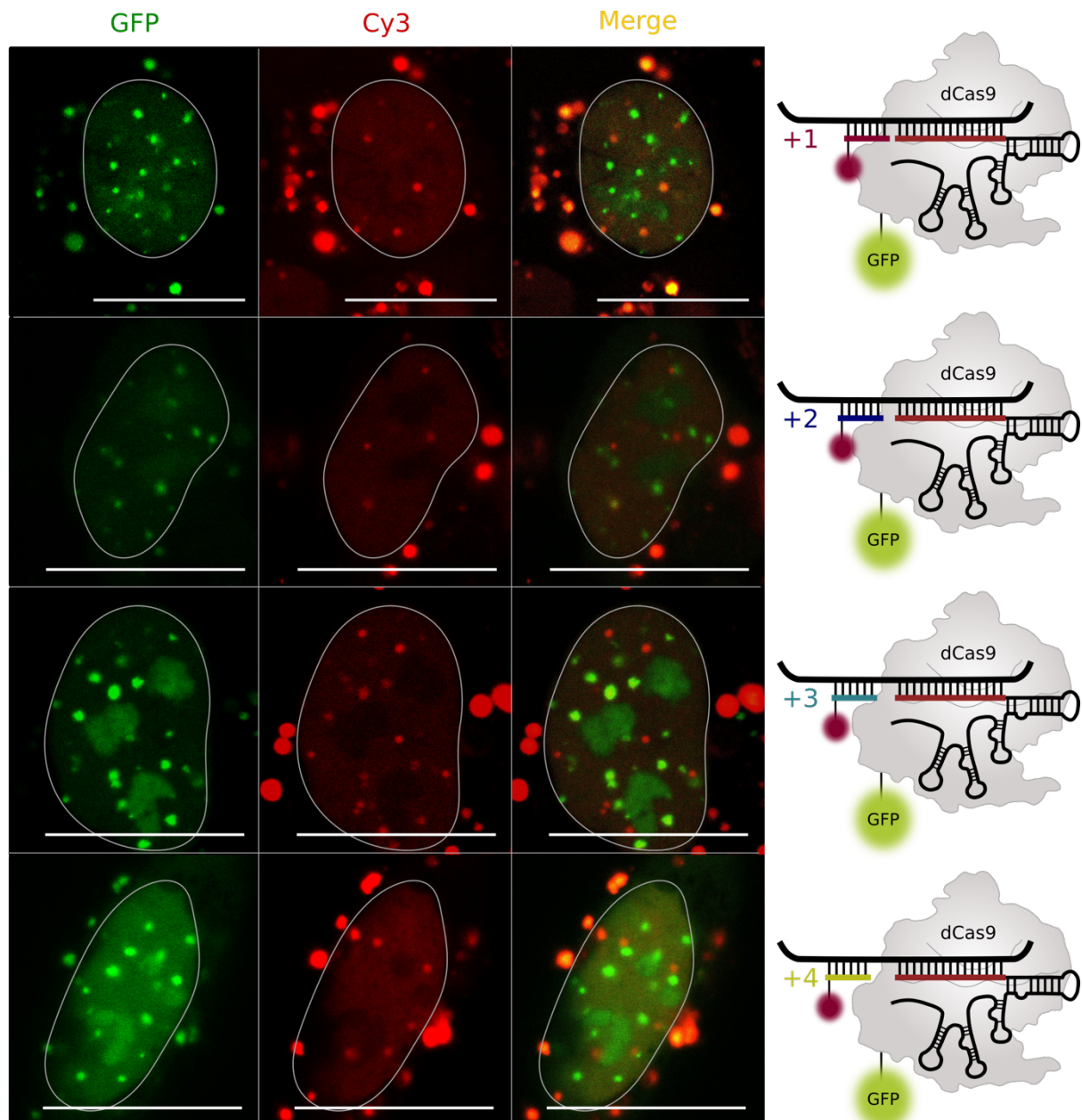


Figure 4.9 | Live cell imaging using eOGN+1–4 as probes. Successful telomeric labelling was observed in the GFP channel. However, the puncta corresponding to the telomeres were not colocalised in the Cy3 channel presumably due to degradation of the Cy3 labelled probes. Moreover, there were significantly less puncta observed in the Cy3 channel in comparison to the GFP channel which could either be due to degradation of the probes or due to inefficient binding of the probes. sgRNA and eOGN+1–4 were co-transfected into dCas9-3XGFP expressing U2OS cells using Lipofectamine reagents. After 12–24 h, live-cell imaging was performed with different lasers used to visualise GFP (green) and Cy3 (red). Scale bar = 25 μ m. The oligonucleotide sequences can be found in Appendix Table 8.

4.4 Phosphorothioate linkages to improve nuclear localisation

4.4.1 Design Strategy & Synthesis

Phosphorothioate (PS) linkages are one of the most widely used modification to enhance the metabolic stability of nucleic acid drugs.²⁵⁴ It is a backbone modification where one of the nonbridging oxygens of the phosphodiester linkage is replaced with sulfur (Figure 4.10a). In comparison to OGNs with phosphodiester (PO) backbone, OGNs with PS modifications show enhanced affinity for plasma,²⁵⁵ cell-surface,²⁵⁶ and intracellular proteins,²⁵⁷ which facilitates distribution and cellular uptake, even in the absence of delivery reagents (Figure 4.10b).²⁵⁸

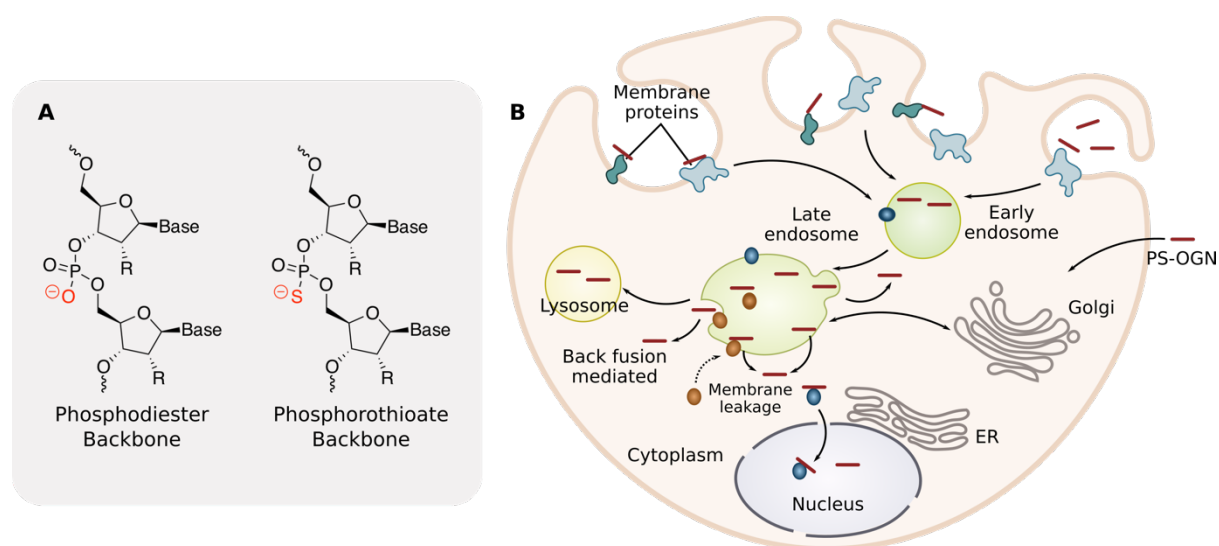


Figure 4.10 | Structure of phosphorothioate backbone and its interaction with cells. (A) compares the phosphodiester backbone to the phosphorothioate backbone, where one of the O is replaced with S. (B) shows the mechanism of cellular uptake and release of a PS-OGN. The PS-OGN interacts with cell surface proteins to enter the cell, via endocytic pathways. The PS-OGN either leaks out of the endosome or is escorted out by interacting with proteins. Interaction of PS-OGN with cellular proteins also facilitates entry into the nucleus.

Cellular uptake of PS-OGNs are usually endocytic.^{257, 259} Importantly, PS-OGNs make no direct interactions with the lipid components of the membrane during gymnotic cell uptake.²⁶⁰⁻²⁶¹ The enhanced cellular uptake is solely due to key interactions with cellular proteins. Although in this chapter, the PS-eOGNs are transfected using

lipofectamine reagents, this is an important point as the fate of any PS-OGNs depend on proteins.²⁶² This means if a PS-OGN is present in a biological site, a protein or protein complex is responsible for its location. For instance, once the PS-OGN enters the cell in an endosome, proteins such as TCP1 and VARS, are recruited to facilitate release of the PS-OGN (Figure 4.10b).²⁶³ This is important as inefficient release of OGNs from endosomes lead to lysosomes fusion and degradation of the trapped OGN. PS-OGNs have the advantage of an “easy” endosomal escape *via* proteins. Once released from the endosomes, PS-ASOs can interact with cellular proteins and enter the nucleus more easily. In fact, it has been shown that nuclear localisation was predominant shortly after microinjection of PS-modified OGNs into the cytoplasm.²⁶⁴⁻²⁶⁶ Lastly, PS-modified OGNs also confer increased resistance to the nucleases that degrade OGNs, thereby extending their half-lives in cells.²⁶²

Increased nuclease resistance, faster endosomal escape and enhanced nuclear localisation make PS an attractive modification to incorporate in the probes. Therefore, four new eOGN were synthesised using the same sequence and using 2'-OMe ribonucleotides, but with PS-linkages throughout the OGN. The new PS-eOGNs were also synthesised with a 3'-amino c6 modification and were later on post-synthetically labelled with Cy3 NHS ester. The four PS-eOGNs were then tested in cells.

4.4.2 Live Cell Imaging

The live cell imaging experiment was repeated, but this time the Cy3-labelled PS-eOGN+1–4 were used. Confocal microscopy revealed perfect colocalization between the genomic labelling observed under the GFP channel and those observed under the Cy3 channel (Figure 4.11). This means all four PS-eOGN were binding specifically to

their target. The binding for all the four probes were very good as all the telomeric spots that were observed in the GFP channel were also seen in the Cy3 channel. This means the probes were binding at every R-loop formed by the dCas9-3xGFP-sgRNA complex. These results are different from what was observed *in vitro*; however, this might be due to the different physiological conditions in cells (temperature, pH), which cannot be exactly replicated *in vitro*. For instance, the *in vitro* assay involves loading the samples on a gel and thus exposing them to harsh conditions (high voltage and temperature) which might cause the probes to dissociate. Handling of the samples such as pipette mixing could also have caused the probes to dissociate *in vitro*.

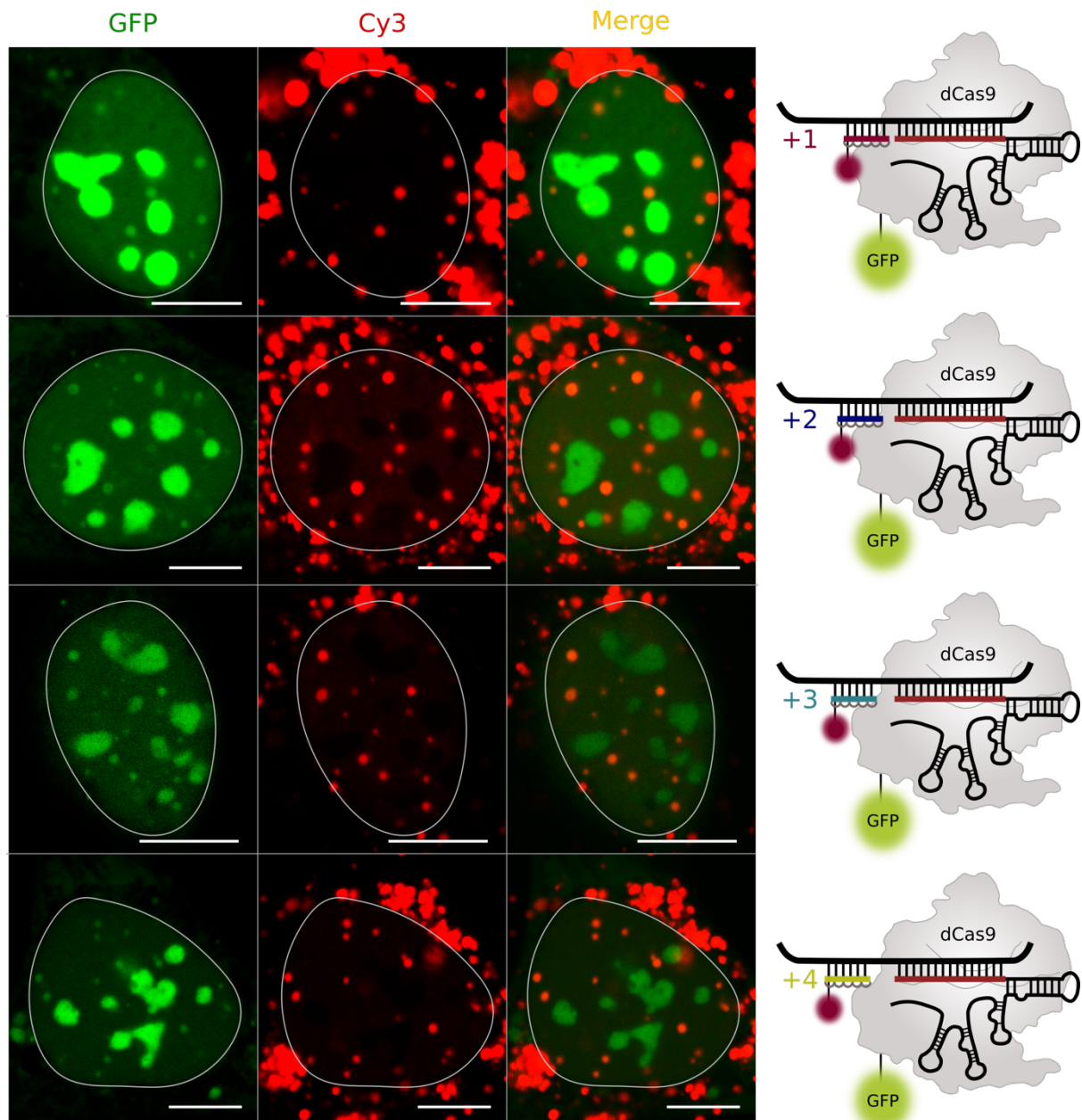


Figure 4.11 | Live cell imaging using eOGN+1–4 with phosphorothioate backbones. Successful telomeric labellings were observed in both the GFP and the Cy3 channel. The puncta observed in both channels perfectly colocalised, which implies that the probes are binding to the right target. Nonetheless, a substantial amount of the Cy3-labelled probes accumulated in the cytoplasm, possibly due to toxicity of phosphorothioates. sgRNA and PS-eOGN+1–4 were co-transfected into dCas9-3XGFP expressing U2OS cells using Lipofectamine reagents. After 12–24 h, live cell imaging was performed with different lasers used to visualise GFP (green) and Cy3 (red). Scale bar = 10 μ m. The oligonucleotide sequences can be found in Appendix Table 9.

The only issue with using the PS-eOGNs was the increased accumulation of the fluorescent probe in the cytoplasm. There are two main reasons for this: PS-modified

oligos are known to interact non-specifically with proteins. Those interactions could lead to clustering of the fluorescent probes. Another reason to account for the accumulation could be due to the toxicity of phosphorothioates, causing a rise in lysosome fusion to vesicles containing the PS-eOGNs, resulting in degradation of the PS-eOGNs and eventually they will be removed from the cells by exocytosis. The accumulation of fluorophore in the cytoplasm interferes with the desired signals in the nucleus and is not ideal for live cell imaging applications. Moreover, the toxicity of the probe could pose serious problems if this concept were to be applied to other areas such as gene editing.

4.4.3 Toxicity of Phosphorothioates

The sulfur atom in PS-OGNs forms strong electrostatic interactions with the positively charged guanidino moiety of arginines and ϵ - ammonium of lysines. However, it also has the ability to form Van Der Waals interactions with the hydrophobic alkyl side chains of amino acids. Hence, the ability of PS to form both electrostatic and hydrophobic interactions forms the basis for the enhanced protein binding and this characteristic has profound biological impacts on toxicity.²⁵⁸

For example, on a global perspective it has been shown that the protein binding for 'safe' OGNs was significantly less relative to toxic OGNs.²⁶⁷⁻²⁶⁸ The discovery of the molecular mechanism underlying the observed toxicity of PS took almost 30 years. However, the mechanism behind most toxic PS-OGNs is now well understood, and it is defined by several key interactions with paraspeckle proteins and RNase H1. Toxic PS-OGNs bind more proteins more tightly. The tight interactions between PS-OGNs and proteins can cause paraspeckle protein mislocalisation to the nucleolus, in a

RNase H1-dependent manner, and can affect pre-ribosomal RNA (rRNA) synthesis, causing nucleolar stress and apoptotic cell death (Figure 4.12).²⁶⁷⁻²⁶⁸ PS-OGNs have also been shown to form various protein aggregates at different concentrations of the PS-OGNs in cells.²⁵⁷ Some of these aggregates are RNPs that are normally found in cells and others are circumstantially formed, only in the presence of PS-OGNs (Table 4.1).²⁶³

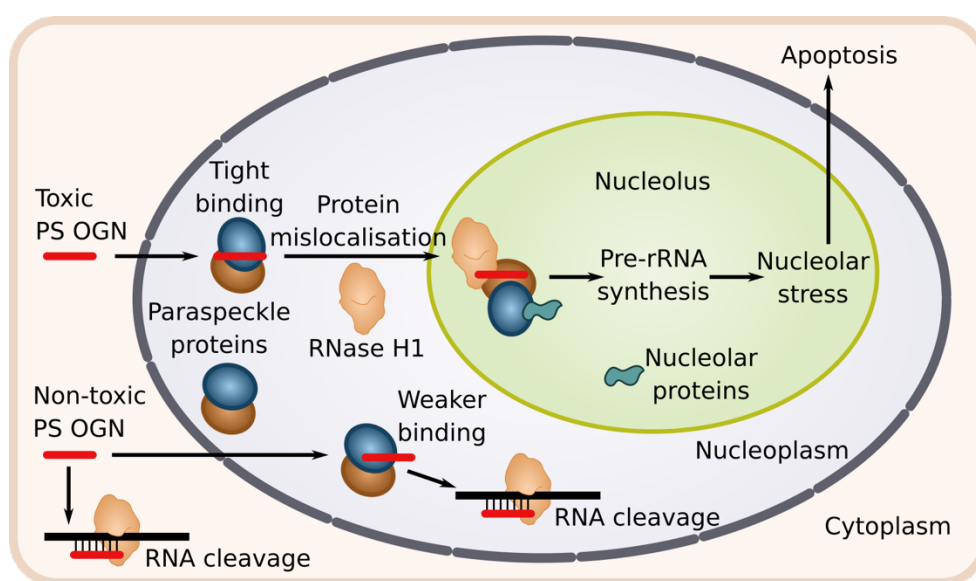


Figure 4.12 | Major molecular toxicity mechanism of PS-OGN. Toxic PS-OGN forms very strong interactions with paraspeckle proteins, leading to protein mislocalisation and eventually apoptosis. Non-toxic PS-OGN on the other hand forms weaker interactions and is able to carry to its normal functions such as RNA cleavage (as an ASO).

The increase in RNP aggregation was in accordance with what was observed in the live cell imaging experiments. 2'-OMe modifications have been shown to mitigate toxicity and reduce protein binding properties of PS-OGNs.²⁶⁷ However, since the PS-eOGNs were already modified with 2'-OMe and were still showing strong presence of protein aggregation, the concentration of the eOGNs transfected was reduced instead. As summarised in Table 4.1, the concentration of PS-OGN transfected has an effect on toxicity. The PS-OGN is only toxic once the amount goes beyond a specific concentration. It was presumed that reducing the excess of the PS-eOGNs would

reduce the chance of protein aggregation in the cytoplasm. However, an optimisation was required to find the correct concentration that reduces aggregation but still binds efficiently to its target.

Table 4.1 | PS-OGN induced protein and RNP complexes. Summary of the different aggregates formed upon transfection of a PS-OGN. The aggregates are concentration dependent and are toxic at varying lengths of time post transfection, depending on the type of aggregates formed.²⁶²

Aggregates	Localisation	Minimum transfected OGN conc. (nM)	Notes
PS bodies	Nucleoplasm	~20	
PS body-like structure	Cytoplasm	~20	can form after RAN reduction
Paraspeckle-like	Nucleoplasm	~10	
Nuclear filaments	Nucleoplasm	~30-50	
Cloudy nucleolar structures	Nucleolus	~15	toxic OGN induced > 1 h after transfection
Perinucleolar cap	Perinucleolar region	~40	toxic OGN induced > 6 h after transfection
Nucleolus-like structure	Nucleolus	~20	toxic OGN induced > 16-18 h after transfection
Stress granules	Cytoplasm	~20	
Mutant protein granule	Cytoplasm	~50	
P-bodies	Cytoplasm	~1	

4.4.4 Effect of Phosphorothioate Concentration on Aggregation

In the previous experiment, 500 pmols (10 eq.) of the PS-eOGNs was used. Consequently, a concentration screen was carried out by transfecting 50 pmols, 100 pmols, 250 pmols and 500 pmols (as a control) of PS-eOGN+4, to find out the optimum concentration required to reduce aggregation but keep telomeric labelling efficient. After carrying out the live cell imaging experiment, a clear reduction in aggregation was observed, as the amount of probe was reduced from 500 pmols to 50 pmols (Figure 4.13). The least amount of cytoplasmic accumulation of the probe was observed when 50 pmols and 100 pmols of probe was used. Reduction in the concentration of PS-eOGN+4 did not have a negative effect on genomic labelling, even when 50 pmols was transfected (1 eq.). However, to ensure that the probe was

present in excess and hence to ensure efficient genomic labelling, it was concluded that 100 pmols (2 eq.) was the optimum amount of PS-eOGN+4 to be transfected in future experiments.

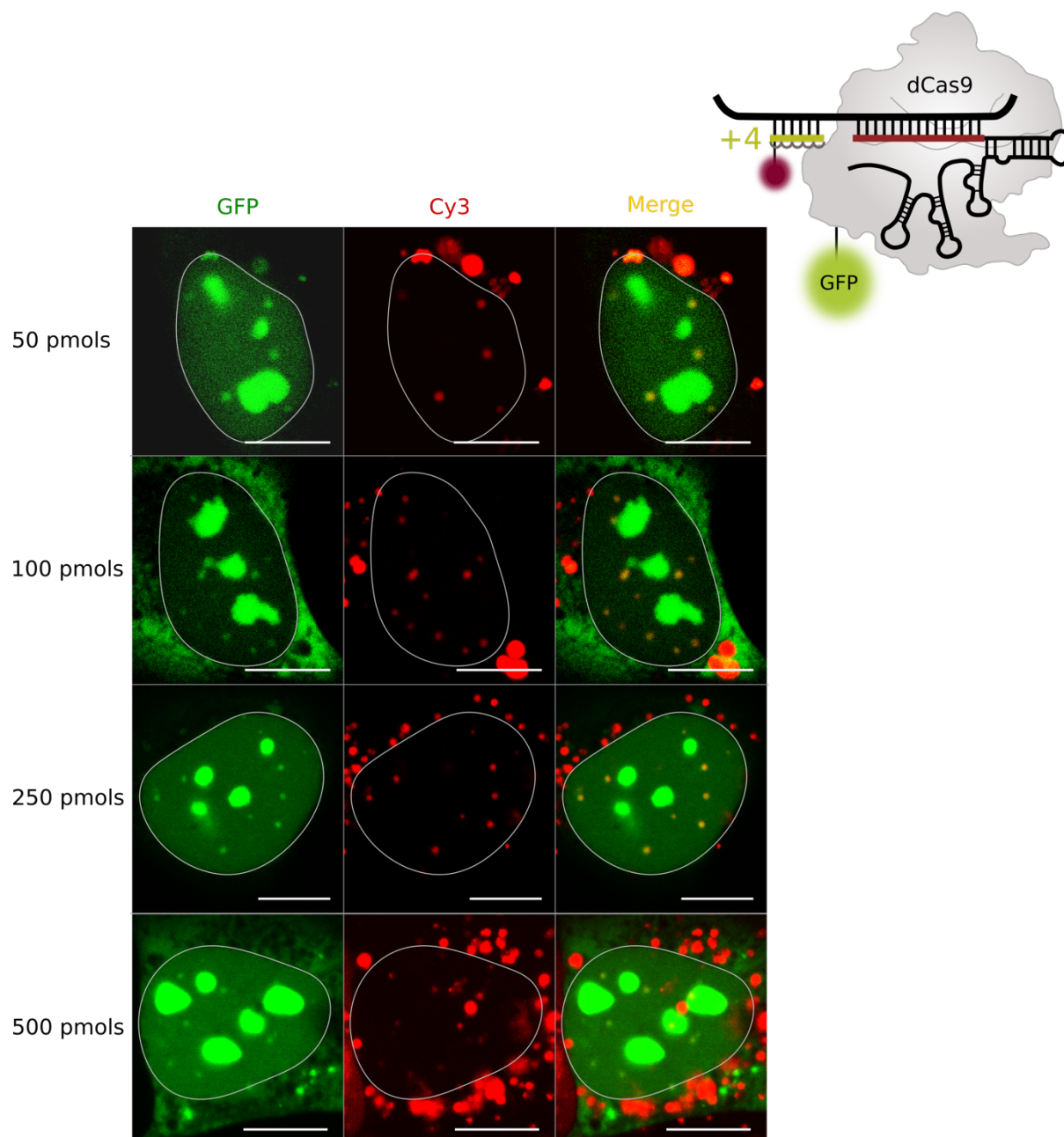


Figure 4.13 | Effect of concentration on toxicity of PS-modified eOGNs. A clear increase in cytoplasmic accumulation can be observed when the amount of eOGN+4 transfected increases from 50 pmols to 500 pmols. The efficiency of genomic labelling was not affected despite transfecting amounts as low as 50 pmols of the eOGN+4. sgRNA and PS-eOGN+4 were co-transfected into dCas9-3XGFP expressing U2OS cells using Lipofectamine reagents. After 12–24 h, live-cell imaging was performed with different lasers used to visualise GFP (green) and Cy3 (red). Scale bar = 10 μ m. The oligonucleotide sequences can be found in Appendix Table 9.

While aggregation was significantly reduced by transfecting only 100 pmols of probe, it still was not completely eliminated. Some accumulation in the cytoplasm was still observed and if this concept were to be used as a tool for live cell imaging, it was imperative to eliminate those interfering signals. Previous studies have shown that the hydrophobic properties of Cy3 can lead to non-specific interactions with proteins. In order to eliminate the possibilities of this scenario, an alternate fluorescent dye was employed.

4.5 Optimisation of the fluorescent dye

4.5.1 Hydrophobicity of Cy3

The cyanine dye Cy3 is widely used as chemical labels in various fluorescent studies. It is composed of two indole rings and three methine groups (Figure 4.14).²⁶⁹ It can be easily conjugated to nucleic acids and proteins and has the advantage of having sharp absorption band, high extinction coefficient, less autofluorescence, great photostability and low pH sensitivity.²⁷⁰ However, studies have found that Cy3-labelled siRNA produced strong fluorescence signals on the surface of cells when electroporated, even though the transfection efficiency was very low. Similarly, when transfected using lipofectamine, Cy3-labelled RNA accumulated around the cell surface.²⁷¹⁻²⁷²

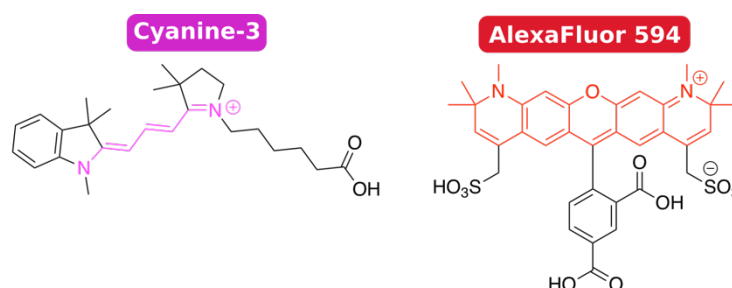


Figure 4.14 | Structure of Cy3 and AlexaFluor 594 dye. The structure differences between the two dyes greatly affect the physical properties. For instance, the linker moiety of Cy3 contributes to the dye's high hydrophobicity.

The interaction that is responsible for most of the non-specific binding is the hydrophobic interactions between the cyanine and linker moieties of the fluorophore and the lipid headgroup/tail, which help to stabilise the fluorophore binding to the lipid bilayer.²⁷³⁻²⁷⁴ In order to estimate the hydrophobicity of a fluorophore, the LogD value of several commonly used fluorophores was measured.²⁷⁵ The LogD values are calculations of the equilibrium partition coefficient of the dye molecule between octanol and water at pH 7.4 and the larger the LogD, the stronger the hydrophobicity.²⁷³ Table 4.2 summarises the LogD of some fluorophores, including Cy3 and it can be noted that the LogD of Cy3 is 4.69, which is considerably high in comparison to the other fluorophores. This means that Cy3-labelled probes can result in artefacts and can lead to errors during interpretation of fluorescent images. Therefore PS-eOGN+1–4 were labelled with the dye AlexaFluor 594 instead, which has a much lower LogD.

Table 4.2 | LogD values of commonly used fluorescent dyes.²⁷³ Cy3 has a considerably higher logD value compared to the other fluorophores, which can result in unspecific hydrophobic interactions in cells. SE is succinimidyl ester reactive group and M is maleimides reactive group. LogD_{hyd} is the calculated logD using the molecular structure of the dye with a hydrolysed reactive group (maleimides and succinimidyl esters). LogD_{unhyd} is the calculated logD using the structure of the unhydrolysed reactive group. QY is the quantum yield.

Dye	logD _{hyd}	logD _{unhyd}	λ_{max} (ex)	λ_{max} (em)	QY
Alexa 488 SE	−11.09	−8.02	494	517	0.55
Alexa 594 M	−7.1	−3.66	588	612	0.66
Cy3 SE	4.69	3.29	555	570	0.31
Sulfo-Cy3 M	−2.72	1.12	550	570	0.15
Sulfo-Cy5 M	−2.19	1.65	649	670	0.28
Texas Red M	−3.19	2.87	595	616	-
Atto 488 SE	−7.6	−4.67	501	523	0.8

4.5.2 Live Cell Imaging using AlexaFluor 594-labelled probes

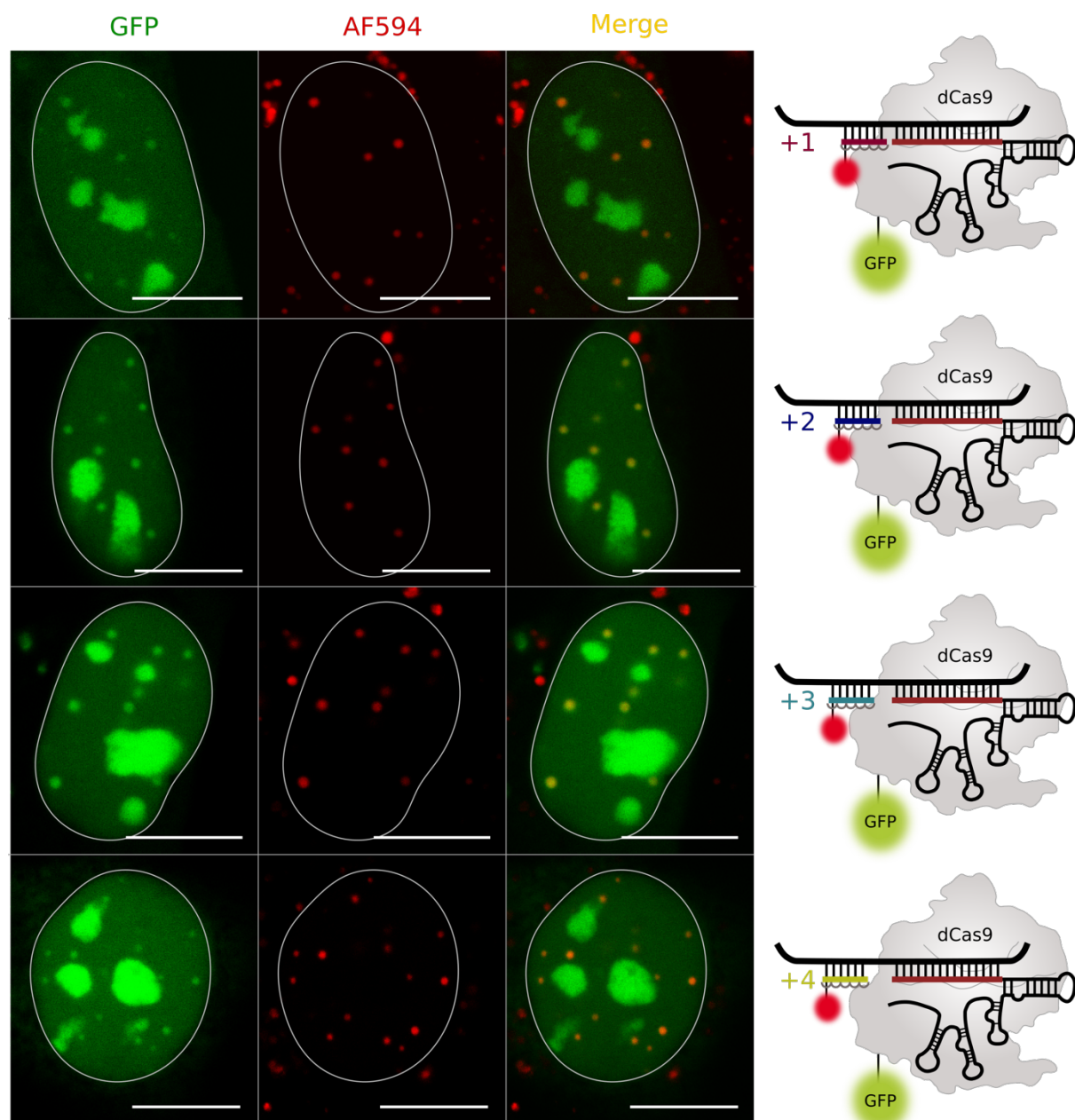


Figure 4.15 | Live cell imaging using AlexaFluor 594 labelled eOGNs. Telomeric labellings can be observed in both GFP and AlexaFluor 594 channel. All the spots were perfectly colocalised and there was negligible signal in the cytoplasm from the AlexaFluor 594 labelled eOGNs. sgRNA and PS-eOGN+1–4 were co-transfected into dCas9-3XGFP expressing U2OS cells using Lipofectamine reagents. After 12–24 h, live-cell imaging was performed with different lasers used to visualise GFP (green) and Alexa Fluor 594 (red). Scale bar = 10 μ m. The oligonucleotide sequences can be found in Appendix Table 9.

PS-eOGN+1–4 were post synthetically labelled using AlexaFluor 594 NHS ester. They were then co-transfected with the sgRNA into U2OS cells expressing dCas9-3xGFP. There was a significant drop in cytoplasmic fluorescence when AlexaFluor 594 was used instead of Cy3. The telomeric labelling remained unaffected and was still very efficient (Figure 4.15). All four probes performed equally well as all the puncta observed under the GFP channel were also replicated under the AlexaFluor 594 channel for all the four probes. There was only negligible amount of fluorescence observed in the cytoplasm. These results were very promising and confirmed that the eOGNs can indeed bind to their target after R-loop formation and that using both phosphorothioate linkages and the correct fluorophore can help to optimise the probe so that it is better suited for its application. Moreover, the puncta observed from AlexaFluor 594 dye were brighter and displayed no interfering signals in the nucleus. GFP, on the other hand, accumulated in the nucleolus to form bright green globules.

4.6 Specificity of PS-eOGN probes

4.6.1 Scrambled probe design

While the results described in section 4.5.2 are very promising, the specificity of the probes need to be tested. Only the probe with the correct specific sequence should bind to the R-loop. A random OGN should in theory should not bind at all and should not form any specific spots in the nucleus or colocalise with the puncta from GFP. Therefore, a scrambled PS-eOGN was designed. The number of As, Ts, Cs and Gs were kept the same as PS-eOGN+4 but the sequence order was scrambled (Figure 4.16). The scrambled oligo was labelled with AlexaFluor 594 and was tested in cells.

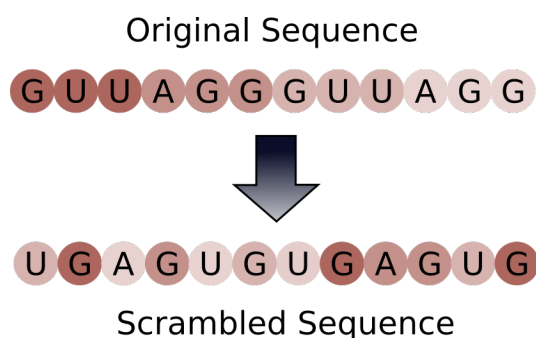


Figure 4.16 | Design of scrambled eOGN. Scrambled OGN designed to test specificity of the eOGN in cells. The length and nucleotides content were kept the same. The positions of the nucleotides were changed such that the OGN would no longer be complementary to the target DNA.

4.6.2 Live cell using scrambled PS-eOGN

The same procedure for live cell imaging was repeated but this time the scrambled oligo was transfected. Confocal microscopy of the transfected cells showed telomeric labelling in the GFP channel, as expected. Most of the fluorescent PS-eOGN accumulated in the cytoplasm of the cells, as shown in Figure 4.17. There were only a few faint spots in the nucleus. However, they did not colocalise with the puncta in the GFP channel, which means they were not binding to the telomeres. It was speculated that the faint spots inside the nucleus was due to non-specific interactions with proteins. These interactions could be minimised by reducing the number of PS-linkages in the probe, such that only the wing regions of the probe are modified with PS to protect against exonucleases and to aid nuclear localisation.

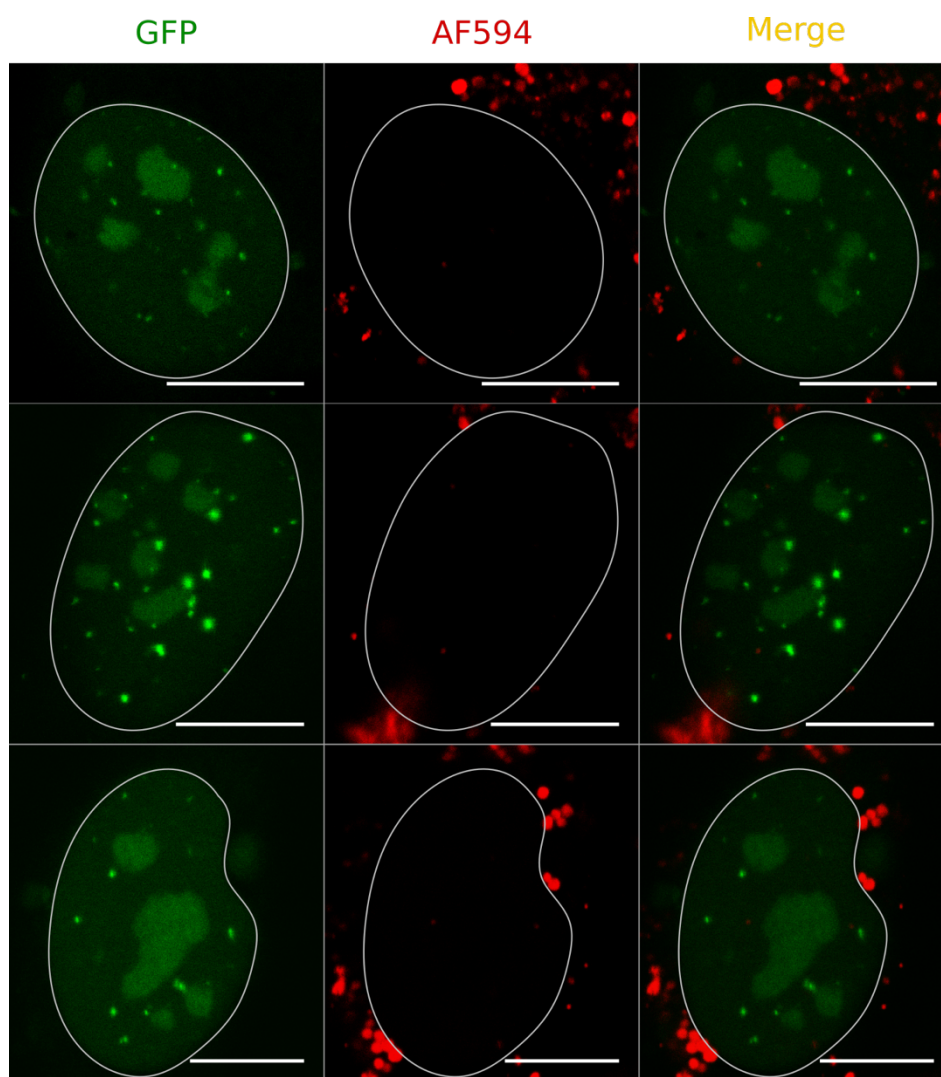


Figure 4.17 | Live cell imaging using scrambled version of PS-eOGN+4. Successful labelling of the telomere can be observed in the GFP channel due to the sgRNA-dCas9-3XGFP RNP. No labelling is observed in the AlexaFluor 594 channel as expected. Instead, the AlexaFluor 594 labelled probe accumulated in the nucleus. sgRNA and the scrambled oligo were co-transfected into dCas9-3XGFP expressing U2OS cells using Lipofectamine reagents. After 12–24 h, live-cell imaging was performed with different lasers used to visualise GFP (green) and Alexa Fluor 594 (red). Scale bar = 10 μ m. The oligonucleotide sequences can be found in Appendix Table 10.

4.7 Conclusions and Future Work

The aim of this work was to prove that after the formation of the R-loop structure, there was enough space for a short 12-mer e-OGN to bind, which could have several important applications in the future. Live cell imaging was used to successfully prove the concept and the work presented a novel method for live cell imaging. The method for live cell imaging summarised in this chapter has the advantage of having negligible

background fluorescence in the cytoplasm but instead forms very bright spot in the nucleus, after specifically binding to its target. It is a cheaper way of doing live cell imaging experiments as the unmodified sgRNA can be made in large scale by *in vitro* transcription and the short eOGN can be effectively made by chemical synthesis and post-synthetically labelled. The sgRNA provided the added advantage of having a strong on-target binding while the eOGN probe emits a strong signal. PS-linkages and 2'-OMe modifications were important incorporations into the eOGN, to prevent degradation and to improve nuclear localisation.

There was a few non-specific labelling observed when a scrambled sequence of PS-eOGN was used. This was most likely due to PS-linkages interacting with proteins and could be minimised by using less PS-modifications. The live cell imaging has only been tested on repetitive target sequences (telomeres). Future work will focus on using this method and a library of sgRNAs and fluorescent eOGNs to efficiently label non-repetitive targets as they have proven to be challenging to image using previously reported methods. The applications will also be expanded beyond live cell imaging by incorporating a ^{CNV}K modification in the eOGN. This will allow us to specifically form crosslinks in proximity to faulty genes, which will be presumably repaired by the cell's repair mechanism and would be a novel method of gene editing using CRISPR-Cas9.

Chapter 5

Conclusions

The aim of this study was to chemically engineer gRNAs to improve their ability to edit genes or to use them beyond gene editing.

In Chapter 2, a metastable fluorescent gRNA was designed such it is only active once it reaches the nucleus, where the target is present. Despite facing several hurdles in the early stages, the design and synthesis of test probes mimicking the gRNA helped considerably to understand the fate of the fluorescent probe in cells and how the probe can be improved. A fluorescent gRNA combining a quencher was successfully designed, that was stable enough to reach the nucleus without significant degradation and cleaved by RNase H once in the nucleus to activate the probe. The fluorescent

gRNA successfully labelled its target, the telomeres. This work enabled us to deliver the fluorescent probe using the user-friendly method, transfection by lipofectamine, without facing the issue of granular cytoplasmic background fluorescence. This was made possible by designing a gRNA that was modified with both a fluorophore and a quencher and by making the gRNA more resistant to exo- and endonucleases.

Chapter 3 consisted of incorporating 3'-modifications to the sgRNA to aid delivery of sgRNAs into cells or to functionalise the sgRNA. A novel way of incorporating modifications was developed, consisting of using an extended sgRNA, made by *in vitro* transcription, and crosslinking a short OGN onto that extension such that the short OGN bears all the modifications. This method was successfully used to add ASGPr-ligands, cholesterol, folic acid and Cy3 to the sgRNA. Unfortunately, gymnotic delivery of sgRNAs modified with ligands has not yet been proved and further work is required to establish the applications of the modified sgRNAs.

In Chapter 4, the R-loop structure, formed when the sgRNA-Cas9 complex binds to its target, was exploited by transfecting another short oligonucleotide (OGN) that bears all the modifications. This novel method of functionalising the CRISPR-Cas9 system was successfully proven using live cell imaging. After carrying out optimisations on the short OGN to make it stable to nucleases and prevent background fluorescence, the fluorescent short OGN not only labelled the telomeres very efficiently, but also produced negligible granular cytoplasmic background fluorescence when transfected by lipofectamine. The work carried out for this chapter presented a new way to introduce modifications, without having to modify the sgRNA, which can be costly and inefficient.

In conclusion, the work carried out in this thesis shows that chemically engineered sgRNAs can have a great impact in improving the applications of CRISPR-Cas9 and established several new ways to make chemically engineered sgRNAs more easily accessible.

Chapter 6

Experimental

6.1 Oligonucleotide synthesis and Purification

6.1.1 DNA synthesis

DNA was synthesised using an Applied Biosystems 394 automated DNA/RNA synthesizer using a standard phosphoramidite cycle of detritylation, coupling, capping and oxidation using TCA (3% in dichloromethane), tetrazole (0.45 M in acetonitrile), Cap A (10% acetic anhydride in 10% lutidine and 80% tetrahydrofuran (THF))/Cap B (16% N-methylimidazole (NMI) in THF) and iodine (0.02 M in THF, pyridine and water) respectively on a 0.2 or 1.0 μmol scale. The relevant solid support (described in Section 6.1.3) was weighed in a twist column, ready for synthesis. Immediately prior

to use, β -cyanoethyl protected phosphoramidites (dA-bz, dG-ib, dC-ac, dT, Sigma-Aldrich) were dissolved in anhydrous acetonitrile (0.1 M). Unmodified monomers had a coupling time of 1 min while modified monomers had a coupling time of 10 min. Stepwise coupling efficiencies were >98% in all cases as determined by automated trityl cation conductivity monitoring.

6.1.2 RNA synthesis

RNA was synthesised using an Applied Biosystems 394 automated DNA/RNA synthesizer using a standard phosphoramidite cycle of detritylation, coupling, capping and oxidation on a 1.0 μ mol scale. The relevant solid support (described in Section 6.1.3) was weighed in a twist column, ready for synthesis. RNA was either made using 2'-O-TBDMS protected monomers or 2'-O-TC protected monomers.

For RNA synthesis employing 2'-O-TBDMS protected phosphoramidites (A-tac, C-tac, G-tac, U, Link Technologies), the monomers were dissolved in anhydrous acetonitrile (0.1 M) immediately before use. Benzylthiotetrazole (BTT) (0.3 M in acetonitrile, Link Technologies), fast deprotection Cap A (5% *t*butylphenoxyacetyl acetic anhydride in THF)/Cap B (16% *N*-methylimidazole (NMI) in THF) and iodine (0.1 M in tetrahydrofuran (THF), pyridine and water) were used as the coupling, capping and oxidation reagents respectively. The monomers had a coupling time of 6 mins and stepwise coupling efficiencies were >97% in all cases as determined by automated trityl cation conductivity monitoring.

For RNA synthesis employing 2'-O-TC protected phosphoramidites (A-bz, C-ac, G-ib, U, Sigma-Aldrich), the monomers were dissolved in anhydrous toluene (0.1 M) immediately before use. Ethylthiotetrazole (ETT) (0.25 M in acetonitrile, Link

Technologies) was used as the coupling reagent when capping and oxidation reagents were the same as those used in DNA synthesis. The monomers had a coupling time of 3 mins and stepwise coupling efficiencies were >97% in all cases as determined by automated trityl cation conductivity monitoring.

6.1.3 Monomers and solid supports

Unmodified DNA phosphoramidites, RNA phosphoramidites and solid supports were purchased from Link Technologies and Applied Biosystems Ltd. Modified monomers and solid support are summarised in Table 6.3.

Table 6.3 | Summary of the modified monomers and the solid supports used.

	Monomer	Supplier	Cat. No.	Diluent
1	Amino-Modifier C ₆ -dT-CE Phosphoramidite	Biosearch Technologies	2135-F100	Anhydrous acetonitrile
2	BHQ-2-dT-CE Phosphoramidite	Biosearch Technologies	2157-F100	Anhydrous acetonitrile
3	Spacer CE Phosphoramidite 18	Biosearch Technologies	LK2129-F100	Anhydrous acetonitrile
4	3-Cyanovinylcarbazole Phosphoramidite (CNVK)	Cambio	10-4960-90	Anhydrous acetonitrile
	Solid Support	Supplier	Cat. No.	
1	3'-Amino-Modifier C ₇ CPG 1000	Biosearch Technologies	2350-P008	

6.2 Oligonucleotide Deprotection

6.2.1 Selective β -cyanoethyl removal

Prior to any deprotection or purification, OGNs containing a primary amine were treated with 20% diethylamine solution in MeCN (20 min at rt) to prevent side reactions between cyanoethyl and the amine during the deprotection step.

6.2.2 DNA deprotection

DNA was cleaved from the solid support and deprotected by heating in a sealed vial at 55 °C for 5 h in concentrated aqueous ammonia. After evaporating the ammonia in vacuo, oligonucleotides were dissolved in water and purified further by High-Performance Liquid Chromatography (HPLC).

6.2.3 2'-O-TBDMS RNA deprotection

RNA was cleaved from the solid support and the nucleobases deprotected by heating in a solution of concentrated aqueous ammonia: ethanol (3:1 v/v) in a sealed vial for 2.5 h at 55 °C. After cooling to room temperature, the ammonia was evaporated in vacuo until the solution becomes turbid. The suspension was then freeze-dried before re-dissolving in a 1:1 mixture of dry DMSO (300 μ L) and trimethylamine trihydrofluoride (300 μ L). The solution was heated for 2.5 h at 65 °C. After cooling to room temperature, sodium acetate (3 M pH 5.2, 50 μ L) and butanol (3 mL) were added and the RNA was cooled to –80 °C for 30 min. The RNA was then pelleted by centrifugation (8500 rpm, 30 min, 4 °C). The supernatant was discarded, and the pellet

washed twice with 70% ethanol in water (750 μ L). The pellet was then freeze-dried, dissolved in water and further purified.

6.2.4 2'-O-TC RNA deprotection

The solid support was treated with dry ethylenediamine: toluene (1:1 v/v) such that the support was completely exposed. The deprotection was carried out for 6 h at room temperature, after which the solid support was washed with toluene (3 $\times$ 1 mL) and dried using argon. The cleaved RNA was eluted with water and further purified.

6.3 Oligonucleotide purification

6.3.1 RP-HPLC

Oligonucleotides were purified using a Gilson HPLC system with ACE[®] C8 column (10 mm $\times$ 250 mm, pore size 100 Å, particle size 10 μ m) with a gradient of buffer A (0.1 M TEAB, pH 7.5) to buffer B (0.1 M TEAB, pH 7.5 containing 50% v/v acetonitrile) and flow rate of 4 mL/min. For unmodified DNA and short RNA, the gradient was 10–40% buffer B over 20 min and for longer RNA (50-mer and higher), the gradient was 20–30% over 30 min. The gradient was adjusted accordingly for other oligonucleotides.

6.3.2 Desalting

NAPTM-5, NAPTM-10, and NAPTM-25 columns (G.E. Healthcare, cat. no. GE17-0854-02) were used according to the manufacturer's instructions.

The Amicon Ultra-0.5 mL 3K and Amicon Ultra-0.5 mL 10K centrifugal filters (Merck Millipore, cat. no. UFC500308) were used according to the manufacturer's instructions.

6.3.3 Denaturing PAGE

A denaturing 8–20% polyacrylamide gel (1 × tris-borate-EDTA (TBE)) buffer containing 7 M urea, percentage and reagent quantities listed in Table 6.4, W × D × H = 18 × 0.2 × 24.4 cm) was prepared and casted. Oligonucleotides mixed with formamide (50% v/v) was loaded onto the gel and separated at 20 W for 1.5–3 h. Bands were visualised under UV or using the Syngene™ G:Box Chemi 16 bio imaging system. The desired bands were excised, crushed, extracted in water (for DNA, ~15 mL) or buffer (for RNA, 50 mM Tris-HCl pH 7.5 containing 25 mM NaCl, ~15 mL) overnight at 37 °C with 900 rpm shaking.

The gel was filtered and discarded, and the oligonucleotide solution was desalted using either a NAP™-10 column followed by a NAP™-25 column (G.E. Healthcare, cat. no. GE17-0852-01) or using Amicon Ultra-0.5 mL centrifugal filters.

Table 6.4 | Example of the reagents and quantities used for preparing denaturing gel.

*Note that APS and TEMED were added to acrylamide mixture before casting.

Reagent	Gel Percentage	
	8%	20%
Urea	29.4 g	29.4 g
Milli-Q H ₂ O	26 mL	5 mL
40% acrylamide:bis-acrylamide (29:1)	14 mL	35 mL
10x TBE (1 M Tris-base, 1 M boric acid, 20 mM EDTA, pH 8.3)	7 mL	7 mL
10% ammonium persulfate (APS)*	560 µL	560 µL
TEMED*	56 µL	56 µL

6.4 Oligonucleotide mass spectrometry

Negative ion electrospray was recorded using a UPLC-MS Xevo G2-QTOF (Waters) mass spectrometer and an Acquity UPLC system with a BEH C18 column (130Å, 1.7 µm, 2.1 mm x 50 mm) and was used to characterise all the oligonucleotides. The gradient used was 0–70% buffer B in buffer A over 7.5 min or 15–30% over 12.5 min for normal oligonucleotides and 50–100% over 7.5 min for very hydrophobic oligonucleotides, with a flow rate of 0.2 mL/min (buffer A: 8.6 mM TEA, 200 mM HFIP in 5% methanol/water (v/v); buffer B: 20% v/v buffer A in methanol). Raw data obtained was processed using the deconvolution software MassLynx v4.1.

6.5 Native PAGE

Native PAGE gels were prepared using the reagents and quantities described below:

Table 6.5 | Example of how native gels were prepared.

*Note that APS and TEMED were added to acrylamide mixture before casting.

Reagent	Gel Percentage	
	5%	8%
Milli-Q H ₂ O	8.2 mL	7.45 mL
40% acrylamide:bis-acrylamide (29:1)	1.25 mL	2 mL
10x TBE (1 M Tris-base, 1 M boric acid, 20 mM EDTA, pH 8.3)	0.5 mL	0.5 mL
10% ammonium persulfate (APS)*	80 µL	80 µL
TEMED*	8 µL	8 µL

Native gels (W × D × H = 80 mm × mm × 75 mm) were run in 1× TBE buffer and samples were mixed with 10% glycerol.

6.6 Agarose Gel Electrophoresis

1% w/v of agarose was dissolved in 1× TBE buffer (100 mM Tris-base, 100 mM boric acid, 2 mM EDTA, pH 8.3) by heating in a microwave. After adding 0.5x SYBR[™] Gold Stain (Thermo Fisher Scientific, cat. no. S11494, 10,000x), the agarose was allowed to set at room temperature. Samples were mixed with Blue gel loading buffer (6×, NEB, cat. no. B7021S) and loaded onto the gel. The gel was run at 126 V for 60–90 min in 1× TBE running buffer. Bands were visualised using a Syngene G:Box Imager.

6.7 *In vitro* Transcription

A dsDNA template mixture was first formed by mixing the ssDNA template (142 μ M, 2.1 μ L) with a short ssDNA complementary to the T7 promoter sequence (300 μ M, 1.1 μ L) and milli-Q water (56.9 μ L) such that the DNA template has a double stranded region that is recognizable by RNA polymerase. DNA was transcribed using the MEGAScript™ T7 *in vitro* transcription kit (ThermoFisher Scientific, AM1334). The transcription reaction was assembled by mixing thoroughly milli-Q water (14 μ L), ATP (75 mM, 4 μ L), CTP (75 mM, 4 μ L), GTP (75 mM, 4 μ L), UTP (75 mM, 4 μ L), enzyme mix (4 μ L) and dsDNA template (5 μ M, 2 μ L). The reaction mixture was incubated at 37 °C for 16 h, after which it was treated with TURBO DNase and incubated at 37 °C for 1 h. The mixture was purified using MEGAClear™ transcription Clean-up kit (Thermo Fisher Scientific, AM1908). Elution Solution (80 μ L), Binding Solution (350 μ L) and ethanol (100%, 250 μ L) were added to the reaction mixture in a filter cartridge. The solution was centrifuged, the flow discarded and the RNA on the cartridge was washed twice with a Wash Solution (500 μ L). Lastly, pre-heated Elution solution (50 μ L, 95 °C) was applied to the cartridge and the *in vitro* transcribed RNA was collected.

6.8 Post-synthetic oligonucleotide labelling

6.8.1 *N*-Hydroxysuccinimide (NHS) labelling

Freeze-dried oligonucleotide was dissolved in NaHCO₃ buffer (0.5 M, pH 8.5, 18 μ L) and mixed with the NHS ester (20 eq. unless stated otherwise) dissolved in DMSO (12 μ L). The reaction mixture was shaken (950 rpm) at 20 °C for 4 h. The final solution was desalted and excess NHS ester reagent removed using a NAP™-10 column (G.E. Healthcare, cat. no. GE17-0854-02) prior to RP-HPLC purification.

6.8.2 Click chemistry labelling (CuAAC)

The volumes of the reagents required was calculated according as described below:

Table 6.4 | Volume of reagents used to prepare stock solutions and in the reaction mixture for click labelling.

Reagent	Final concentration in mixture	Stock solution concentration
Oligonucleotide, azido-modified	Varies (20–200 μ M)	Varies
Alkyne	1.5 x oligonucleotide concentration	10 mM in DMSO
DMSO	50 vol %	–
Ascorbic Acid	0.5 mM	5 mM in water
Cu-TBTA Complex	0.5 mM	10 mM in 55 vol % DMSO

The azido-modified oligonucleotide was dissolved in water in a pressure-tight vial. 2M triethylammonium acetate buffer, pH 7.0 was added to a final concentration of 0.2 M. DMSO was added and the solution was mixed by vortexing. The alkyne stock solution (10 mM in DMSO) was added followed by vortex again. 5mM ascorbic acid stock solution was added and mixed briefly followed by degassing by bubbling helium into the solution for 1 min. 10 mM copper (II)-TBTA stock in 55% DMSO was added, the vial was flushed with helium and the cap was immediately closed. The solution was thoroughly mixed and if precipitation was observed, the mixture was heated at 80 °C for 3 mins. The reaction was left overnight at room temperature after which the mixture was thoroughly desalted using the Amicon Ultra-0.5 mL 3K columns.

6.9 Real-time fluorescence measurements

6.9.1 Real-time fluorescence with stoichiometric amount of quencher

RNase H (1 μ L, NEB, cat. no. M0297S), diluted to the desired concentration, was added to the RNase H activated probes (0.44 μ M) in RNase H buffer (10 $\times$, 1 μ L, NEB). The total volume was made up to 10 μ L by adding milli-Q water. Where necessary, diluted Cas9 (10 μ M, 1 μ L, NEB, cat. no. M0386T) was added to the suspended probe prior to the addition of RNase H and was incubated at room temperature for 2 min. The mixture was incubated at 37 °C and the fluorescence was monitored using a CFX96™ Real-Time PCR Detection System (Bio-Rad) every minute for 150 min. The controls were modified accordingly.

6.9.2 Real-time fluorescence with increasing amount of quencher

To the fluorophore bearing strand (0.44 μ M) in RNase H buffer (10 $\times$, 1 μ L, NEB), increasing equivalents of the quencher bearing strand was added (1 eq.–2 eq.). RNase H (1 μ L, NEB, cat. no. M0297S), diluted to the desired concentration, was added. The total volume was made up to 10 μ L by adding milli-Q water. The mixture was incubated at 37 °C and the fluorescence was monitored using a CFX96™ Real-Time PCR Detection System (Bio-Rad) every minute for 150 min.

6.9.3 Melting temperature studies

RNase H activated probes (0.44 μ M) was suspended in RNase H buffer (10 $\times$, 1 μ L, NEB). The total volume was made up to 10 μ L by adding milli-Q water. The melting fluorescence change was monitored using a CFX96™ Real-Time PCR Detection

System (Bio-Rad). The mixture was first heated from 37 °C to 95 °C, with an increment of 1 °C every 6 s. The mixture was held at 95 °C for 2 min and data points were taken every minute for 15 min. The mixture was then cooled from 95 °C to 10 °C, with a decrement of 1 °C every 6 s. The mixture was annealed again by heating from 10 °C to 95 °C, with an increment of 0.5 °C and data point taken every 30 s. The mixture was held at 95 °C and then melted again by cooling to 10 °C, with a decrement of 0.5 °C and data point taken every 30 s.

6.10 Time course reaction with RNase H

RNase H activated probes (200 pmols) were suspended in RNase H buffer (10x, 5 µL, NEB) and 43 µL milli-Q water. RNase H (2 µL, NEB, cat. no. M0297S) was added and the reaction was incubated at 37 °C for 30 s, 1 min, 5 min, 15 min, 30 min and 1 h. The reaction was terminated by incubating the reaction mixture at 65 °C for 20 min. The samples were loaded on an 8% denaturing PAGE for analysis.

6.11 Crosslinking ligation of oligonucleotides

250 pmols of RNA was suspended into NEBuffer™ r3.1 (10x, 5 µL, NEB). The ^{CNV}K-containing oligonucleotide (500 pmols) was added and the total volume was made up to 50 µL by adding milli-Q water. Photoirradiation was carried out with a UV-LED (366 nm, at a distance of 1.5 cm) at 0 °C on an aluminium block incubator for 15 min. The reaction mixture desalted using Amicon Ultra-0.5 mL 10K columns and loaded onto a denaturing PAGE for purification.

6.12 Electrophoretic Mobility Shift Assay

The RNP complex was formed by mixing EnGen Spy dCas9 (20 μ M, 0.5 μ L, NEB, M0652T), IVT sgRNA (10 μ M, 1 μ L), NEBuffer™ r3.1 (10x, 1 μ L, NEB) and 6.3 μ L milli-Q water. The complex was allowed to form for 15 mins at room temperature. dsDNA (10 μ M, 1 μ L) and Cy3-labelled eOGN (10 μ M, 0.2 μ L) was then added to the mixture and the RNP was allowed to bind for 30 min at 37 °C. The mixture was mixed with Blue gel loading buffer (6x, 2 μ L, NEB cat. no. B7703S) loaded onto an 8% native PAGE gel and visualised using Typhoon™ laser-scanner platform.

6.13 *In vitro* Cas9 Cleavage Assays

gRNA (300 nM, 5 μ L), Cas9 (1 μ M, 0.3 μ L, NEB, M0386S), NEBuffer™ r3.1 (10x, 1 μ L, NEB), target DNA (20 nM, 1.5 μ L) and H₂O (2.2 μ L) were mixed together and incubated for 3 h at 37 °C. The RNA was degraded by adding RNase A (20 mg/mL, 5 μ L) and incubating for 1 h at 37 °C. The enzymes were digested by adding proteinase K (20 mg/mL, 0.5 μ L) for incubating for another 1 h at 37 °C. The reaction was mixed with Blue gel loading buffer (6x, 2 μ L, NEB, B7703S) and loaded on a 1% agarose stained with 0.5x SYBR gold (1x TBE running buffer, 126 V).

6.14 Cell Culture

Human osteosarcoma cell line U2OS and human adenocarcinoma cell line HeLa were cultured in DMEM (high glucose, HEPES-buffered, no phenol red, with glutamine, ThermoFisher Scientific, cat. no. 21063029) supplemented with 10% (v/v) FBS (ThermoFisher Scientific, cat. no. 10270098) in a humidified incubator at 37 °C with 5% CO₂.

Human hepatocellular HEPG2 cells were cultured in Minimum Essential Medium Eagle (EMEM) (with Earle's salts, L-glutamine and sodium bicarbonate, Merck, cat. no. M4655), supplemented with 10% (v/v) FBS (Thermo Fisher Scientific, cat. no. 10270098) in a humidified incubator at 37 °C with 5% CO₂.

6.15 Lipid Transfection for Live Cell Imaging

U2OS cells were plated in 8-well Nunc™ Lab-tek™ chambered plates (ThermoFisher Scientific, cat. no. 155409PK) at a density of 8×10^4 – 1×10^5 cells per well. After 24 h, cells were washed with DPBS (ThermoFisher Scientific, cat. no. 14190250) and the growth media was replaced with colourless Opti-MEM™ reduced serum media. pHAGE-TO-dCas9-3xGFP (1 µg/µL, 0.5 µL, Addgene, plasmid number 64107, gifted by Dr Taemitree) in Opti-MEM™ was gently mixed with Lipofectamine™ 2000 (1 µL, ThermoFisher Scientific, cat. no. 11668030) diluted in Opti-MEM™, in a 1:1 ratio and allowed to incubate for 5 min at rt. The transfection mixture was added to the cells. After 6 h, the cells were washed with DPBS and the media was replaced with fresh colourless Opti-MEM™.

For gRNA transfection, 10–30 pmols of gRNA was diluted in Opti-MEM™ (25 µL) and gently mixed with Lipofectamine™ RNAiMAX Reagent (1.5 µL, ThermoFisher Scientific, cat. no. 13778030) diluted in Opti-MEM™ medium (25 µL), and incubated for 5 min at rt. The gRNA-lipofectamine mixture was added to the cells and after 16h, the media was replaced with fresh DMEM+10% FBS media and confocal microscopy was carried out.

6.16 Confocal Microscopy

Live cell imaging was performed 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 was detected as follows: For dCas9-3xGFP, GFP was excited at 488 nm and the emission recorded from 493–542 nm. For gRNA labelled with AF594, the dye was excited at 594 nm and recorded from 548–604 nm. For probes labelled with Cy3, the dye was excited at 543 and recorded from 548–604 nm. All raw images were processed using the software FIJI.

6.17 Gymnotic Delivery

The RNP complex was prepared by mixing EnGen Spy Cas9 NLS (20 µM, 12.5 µL, NEB, cat. no. M0646M) with the modified sgRNA (50 µM, 5 µL), NEBuffer™ r3.1 (5 µL) and the total volume was made up to 50 µL with sterilised water. The solution was incubated at 37 °C for 15 min to allow the complex to form.

To prepare RNP complex with 500 pmols, the volume of Cas9 and sgRNA was doubled while the concentrations and total volume was kept constant.

8 x 10⁴ cells (U2OS, HeLa and HEP G2) cells were plated in 24-well gelatin-coated tissue culture plates. The RNP-complex solution (50 µL) was mixed with 50 µL RNP buffer [20 mM HEPES-NaOH pH 7.45, 150 mM NaCl, 10% (v/v) glycerol and 1 mM MgCl₂] and added to 400 µL of OptiMEM. 75 µL 100 µM ppTG21 salt (7.5 nmol) and 75 µL of sterilised water was then added to the mixture and incubated at room temperature for 5 min. For controls without ppTG21, 150 µL water was used instead. The OptiMEM mixture (650 µL) was added to the plated cells and incubated for 48 h. After incubation, the cells were washed, and genomic extraction was carried out.

6.18 Genomic DNA Extraction

Cells were washed with DPBS, trypsinised and pelleted (200 rpm, 5 min). PureLink® genomic DNA mini kit (Invitrogen, cat. no. K1820-00) was used to extract genomic DNA from the cell pellet. The cells were resuspended in DPBS and (200 µL) and proteinase K (20 µL) and RNase A (20 µL) was added and incubated at room temperature for 2 min. Purelink® genomic Lysis/Binding buffer (200 µL) was then added and the sample was heated for 10 min at 55 °C before the addition of ethanol (200 µL). The sample was applied to a purification column and centrifuged at 10,000 rpm for 1 minute at room temperature. The sample in the column was then washed with Wash I (500 µL), centrifuged (10,000 rpm, 1 min), washed with Wash II (500 µL), and centrifuged (10,000 rpm, 1.5 min). Genomic DNA was then eluted using Elution buffer (30 µL) by centrifuging (13,300 rpm, 1min).

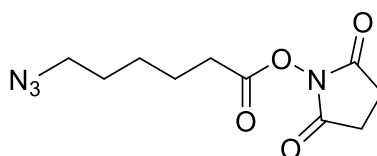
6.19 T7E1 Assay

PCR amplicons from genomic DNA were generated using forward primer (10 µM, 2.5 µL), reverse primer (10 µM, 2.5 µL), genomic DNA (100 ng, 5 µL) dNTPs (10 mM, 1 µL), Phusion Hot Start Flex polymerase (2 U/µL, 0.5 µL, NEB, cat. no. M0535S) and Phusion HF buffer (5x, 10 µL, NEB) and water (28.5 µL). The PCR cycle consisted of a denaturation step (30 sec, 98 °C), an amplification step (34 cycles of 5 sec, 98 °C; 20 sec, 60 °C and 20 sec, 72 °C) and a final extension step (5 min, 72 °C). The resulting PCR products were purified using AMPure XP beads (90 µL, 1.8x beads to DNA ratio, Beckman Coulter, cat. no. A63880) or using the Monarch PCR & DNA Cleanup Kit (5 µg) (NEB, cat. no. T1030S) according to the manufacturer's instructions.

The purified PCR product (200 ng) was mixed with NEBuffer™ 2 (10x, 2 μ L, NEB) and water (total volume = 20 μ L) and annealed by heating for 5 min at 95 °C and cooling to 85 °C (ramp rate = -2 °C/sec) and then 25 °C (ramp rate = -0.1 °C/sec). T7 endonuclease I (10 U/ μ L, 1 μ L, NEB, cat. no. M0302S) was added to the annealed mixture and incubated at 37 °C for 15 min. The reaction was terminated by adding proteinase K (20 mg/mL, 1 μ L) for 1 h at 37°C. The mixture was analysed by mixing it with Blue gel loading buffer (6x, 4 μ L, NEB cat. no. B7703S) and by loading it on a 2 % agarose gel stained with 0.5x SYBR gold (1x TBE running buffer, 126 V).

6.20 Chemical Synthesis

6.20.1 Synthesis of 6-Azido hexanoic acid NHS ester



To a solution of 6-bromohexanoic acid (1.00 g, 5.13 mmol) in dry DMF (4 mL) was added sodium azide (0.67 g, 10.3 mmol) and the mixture was allowed to stir at 85 °C for 3 h. The solvent was then evaporated under high vacuum. The crude product was dissolved in ethyl acetate (50 mL) and washed with 0.1 M HCl (50 mL). The aqueous layer was extracted twice with ethyl acetate (2x 50 mL). The organic layers were combined, dried over magnesium sulfate and concentrated in vacuo to afford crude 6-azidohexanoic acid (1.89 g, quantitative).

To the crude product (1.6 g, 10.2 mmol) was added *N*-Hydroxysuccinimide (261 mg, 2.27 mmol) and EDC•HCl (435 mg, 2.27 mmol). The reaction was kept under an argon atmosphere and 3:1 CHCl₃/DMF (1.5 mL:0.5 mL) was added. The reaction was stirred

under argon overnight. The solvents were evaporated under high vacuum and the crude product was dissolved in ethyl acetate (50 mL) and successively washed with 1 M HCl (50 mL), dilute NaHCO₃ (50 mL) and brine (50 mL). The organic layer was then dried over magnesium sulfate and concentrated *in vacuo*. The crude product was dissolved in CH₂Cl₂ (2 mL), loaded onto a 25 g SNAP Ultra cartridge and purified by column chromatography [SiO₂, CH₂Cl₂: EtOAc (100:0→70:30)]. The product-containing fractions were combined, and the solvent removed under reduced pressure to afford 6-azido hexanoic acid NHS ester (936 mg, 72%) as a colourless oil.

R_f (5% MeOH in CH₂Cl₂): 0.79

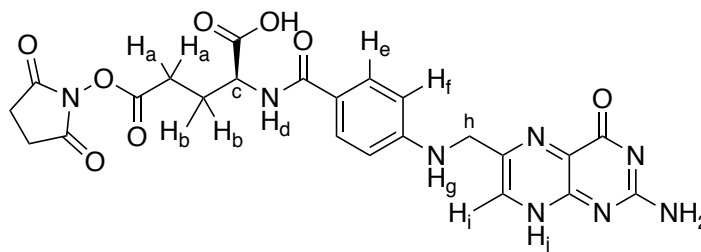
LRMS (ESI-TOF) m/z: [M+K]⁺ 285.1

¹H NMR (400 MHz, Chloroform-*d*) δ 3.25 (t, *J* = 6.8 Hz, 2H, CH₂-C(O)-O), 2.79 (s, 4H, CO-CH₂-CH₂-CO), 2.59 (t, *J* = 7.4 Hz, 2H, CH₂-N₃), 1.78 – 1.70 (m, 2H, CH₂-CH₂-N₃), 1.65 – 1.54 (m, 2H, CH₂-CH₂-CO), 1.50 – 1.42 (m, 2H, CH₂-CH₂-CH₂).

¹³C NMR (101 MHz, Chloroform-*d*) δ 169.3 (2x CO-N), 168.42 (C(O)-O), 51.1 (CH₂-C(O)-O), 30.8 (CH₂-N₃), 28.4 (CH₂-CH₂-CO), 25.9 (CH₂-CH₂-CH₂), 25.6 (CO-CH₂-CH₂-CO), 24.1 (CH₂-CH₂-N₃).

Data agrees with literature.²⁷⁶

6.20.2 Synthesis of Folic Acid NHS Ester



Folate (1.0 g, 2.26 mmol) and 1,3-dicyclohexylcarbodiimide (DCC) coupling agent (1.4 g, 6.78 mmol) were dissolved in anhydrous dimethyl sulfoxide (DMSO) and stirred for 2 h at room temperature. *N*-hydroxysuccinimide (0.78 g, 6.78 mmol) and 4-dimethylaminopyridine (0.14 g, 1.14 mmol) were then added to the solution and the reaction was stirred overnight at 40 °C under a nitrogen atmosphere.

The reaction mixture was filtered to remove the precipitated dicyclohexylurea and the filtrate was concentrated under high vacuum. The concentrated crude product was added dropwise to a cold solution of acetone/diethyl ether (30:70 v/v ratio) and the resulting precipitate was collection by filtration. The precipitate was washed with 100 mL of acetone/diethyl ether (30:70 v/v ratio) solution and diethyl ether (3x 100 mL) and was dried under high vacuum to yield folic acid NHS ester (652 mg, 54%) as an orange solid.

LRMS (ESI-TOF) m/z : $[M+Na]^+$ 561.1

^1H NMR (400 MHz, $\text{DMSO-}d_6$) δ 11.52 (b, 1H, COOH), 8.66 (s, 1H, H_i), 8.15 (d, J = 7.3 Hz, 1H, H_d), 7.66 (d, J = 8.4 Hz, 2H, H_e), 6.96 (t, J = 5.3 Hz, 1H, H_g), 6.64 (d, J = 8.4 Hz, 2H, H_f), 5.01 (m, 1H, H_j), 4.94 (m, 2H, NH_2), 4.61 (m, 2H, H_h), 4.47 (dd, J = 7.6, 12.3 Hz, 1H, H_c), 2.79 (m, 4H, NHS ester), 2.39 (m, 2H, H_a), 2.07 (m, 2H, H_b).

^{13}C NMR (101 MHz, DMSO- d_6) δ 174.4, 173.9, 165.1, 152.3, 151.0, 148.3, 128.9, 127.8, 121.5, 111.7, 52.1, 46.2, 30.8, 169.1, 25.5.

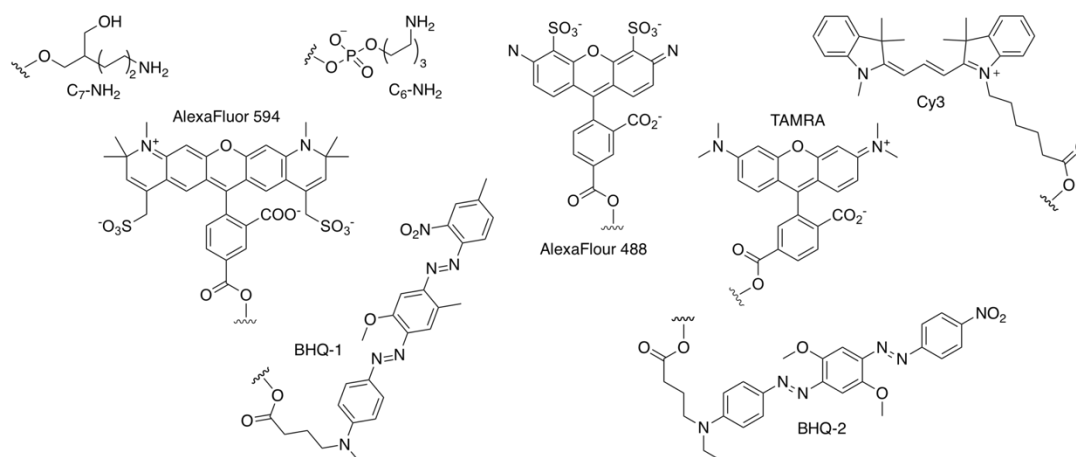
Data agrees with literature.²⁷⁷⁻²⁷⁹

Appendix

Appendix Table 1 | Oligonucleotides used in optimisation of fgRNA (Probes 1–3).

Uppercase bases are RNA nucleotides and lowercase bases are DNA nucleotides. Italicised bases are 2'-OMe modified nucleotides. The structures of modifications are shown below the table. The oligonucleotides are given a unique identifier ('res') and a code name.

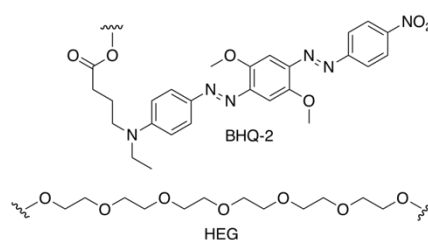
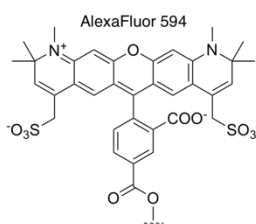
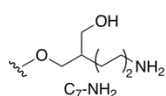
res	Code	Sequence (5'-3')	Mol. weight (g/mol)	
			Expected	Found
8229	Probe 1	gucagautcatcacc-C ₇ -NH ₂	4693	4694
8230	Probe 2	<i>GUCAGAU</i> tcacc-C ₇ -NH ₂	4902	4903
8231	Probe 3	<i>GUCAGAUU</i> Cacc-C ₇ -NH ₂	4948	4950
8193	Complement	NH ₂ -C ₆ -GGUGAUGAAUCUGAC	5050	5051
8196	Complement-BHQ2	BHQ2-GGUGAUGAAUCUGAC	5411	5412
8193-BHQ1	Complement-BHQ1	BHQ1-(O)C-NH-GGUGAUGAAUCUGAC	5537	5538
8229-Cy3	Probe 1-Cy3	gucagautcatcacc-C ₇ -NH-CO-Cy3	-	5280
8229-AF594	Probe 1-AF594	gucagautcatcacc-C ₇ -NH-CO-AF594	-	5608
8229-TAMRA	Probe 1-TAMRA	gucagautcatcacc-C ₇ -NH-CO-TAMRA	-	5253
8229-AF488	Probe 1-AF488	gucagautcatcacc-C ₇ -NH-CO-AF488	-	5419
8230-AF594	Probe 2-AF594	<i>GUCAGAU</i> tcacc-C ₇ -NH-CO-AF594	-	5652
8231-AF594	Probe 3-AF594	<i>GUCAGAUU</i> Cacc-C ₇ -NH-CO-AF594	-	5700



Appendix Table 2 | crRNA and tracrRNA used to assemble fgRNA I.

Uppercase bases are RNA nucleotides and lowercase bases are DNA nucleotides. Italicised bases are 2'-OMe modified nucleotides. **X** is a HEG linker. The structures of modifications are shown below the table. The oligonucleotides are given a unique identifier ('res') and a code name.

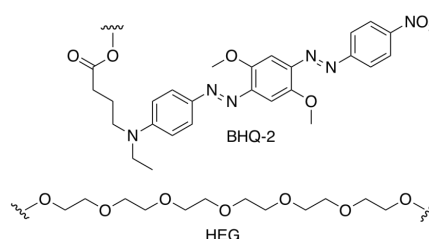
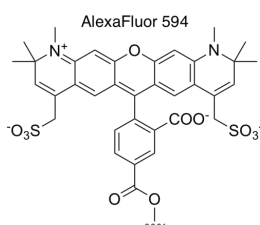
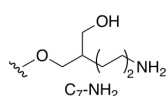
res	Code	Sequence (5'-3')	Mol. weight (g/mol)	
			Expected	Found
6491	IVT telomeric sgRNA template	aaaagcaccgactcgggtgccactttttcaagttgataacggact agccttattttaacttgctatttctagctctaaaactaacctaacc ctaaccctaccctatagtgagtcgtattaga	37319	37324
8553	crRNA I	GGGUUAGGGUUAGGGUUAGUUUAAGAGCUA UGCUGUUUUG XXX GUCAGAUtcatCACC-C ₇ -NH ₂	19044	19044
8553-AF594	crRNA I-AF594	GGGUUAGGGUUAGGGUUAGUUUAAGAGCUAUGCU GUUUUUG XXX GUCAGAUtcatCACC-C ₇ -NH-CO-AF594	19749	19750
8196	Complement-BHQ2	BHQ2-GGUGAUGAAUCUGAC	5411	5412
8558	IVT tracrRNA template	aaaagcaccgactcgggtgccactttttcaagttgataacggact agccttatttaaacttgctatgctgtttgcctatagtgagtcgtattaga	29541	29545



Appendix Table 3 | Oligonucleotide design for the stability test probe.

Uppercase bases are RNA nucleotides and lowercase bases are DNA nucleotides. Italicised bases are 2'-OMe modified nucleotides. Underlined bases are modified with phosphorothioate linkages. **X** is a HEG linker. The structures of modifications are shown below the table. The oligonucleotides are given a unique identifier ('res') and a code name.

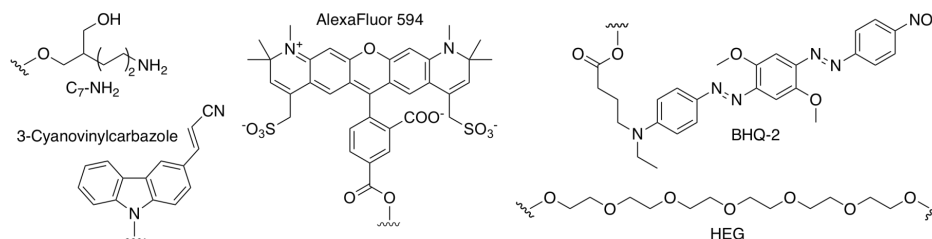
res	Code	Sequence (5'-3')	Mol. weight (g/mol)	
			Expected	Found
9171	Stability probe	BHQ2-GGUGAUGAAUCUGAC X <u>GUCAGAU</u> tcatCACC-C ₇ -NH ₂	11023	11025
9171-AF594	Stability probe-AF594	BHQ2-GGUGAUGAAUCUGAC X <u>GUCAGAU</u> tcatCACC-C ₇ -NH-CO-AF594	-	11723



Appendix Table 4 | crRNA and tracrRNA used to assemble fgRNA II and fgRNA III.

Uppercase bases are RNA nucleotides and lowercase bases are DNA nucleotides. Italicised bases are 2'-OMe modified nucleotides. Underlined bases are modified with phosphorothioate linkages. **X** is a HEG linker. **W** is 3-Cyanovinylcarbazole. The structures of modifications are shown below the table. The oligonucleotides are given a unique identifier ('res') and a code name.

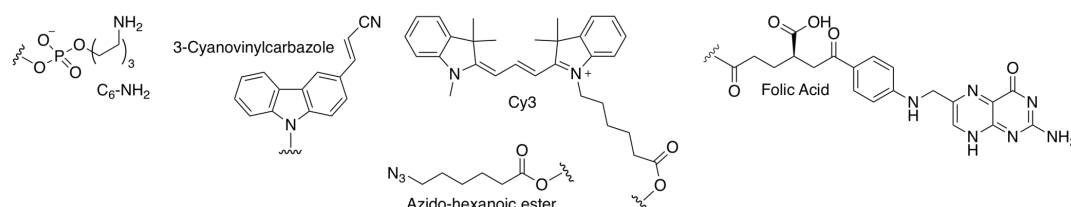
res	Code	Sequence (5'-3')	Mol. weight (g/mol)	
			Expected	Found
9258	crRNA II	GGGUUAGGGUUAGGGUUAGUUUAAGAGCUA UGCUGUUUUUG XXX GU CAGAU tc <u>atCACC</u> -C ₇ -NH ₂	19332	19332
9258-AF594	crRNA II-AF594	GGGUUAGGGUUAGGGUUAGUUUAAGAGCUA UGCUGUUUUUG XXX GU CAGAU tc <u>atCACC</u> -C ₇ -NH-CO-AF594	-	20026
8558	IVT tracrRNA template	aaaagcaccgactcggtgccactttttcaagttgataacggact agccttatttaaacttgctatgctgttttgcctatagtgagtcgtattaga	29541	29545
8196	Complement-BHQ2	BHQ2-GGUGAUGAAUCUGAC	5411	5412
9333	Complement-BHQ2-CNV _K	BHQ2-GGUGAUGAA W CUGAC	5444	5445



Appendix Table 5 | Oligonucleotides used to assemble modified extended sgRNAs.

Uppercase bases are RNA nucleotides and lowercase bases are DNA nucleotides. Italicised bases are 2'-OMe modified nucleotides. **W** is 3-Cyanovinylcarbazole. The structures of modifications are shown below the table. The oligonucleotides are given a unique identifier ('res') and a code name.

res	Code	Sequence (5'-3')	Mol. weight (g/mol)	
			Expected	Found
9018	Extended sgRNA <i>EMX1</i> template	caaatcttattctatattcaaaaagcaccgactcggtgccactttt caagttgataacggactagccttattttaacttgctatttctagctct aaaacttcttctctgctcggaactccctatagtgagtcgtattaga	43168	43171
8856	Extended sgRNA telomeric template	caaatcttattctatattcaaaaagcaccgactcggtgccactttt caagttgataacggactagccttattttaacttgctatttctagctct aaaactaaccctaaccctaaccctaccctatagtgagtcgtattaga	43405	43409
9020	Short OGN	NH ₂ -C ₆ -CAAUUCUUA W UCUAUAUUA	6878	6878
9020-N ₃	Short OGN-N ₃	N ₃ -C ₆ -C(O)-NH-C ₆ -CAAUUCUUA W UCUAUAUUA	-	6895
9020-Folic Acid	Short OGN-FA	Folic Acid-C(O)-NH-C ₆ -CAAUUCUUA W UCUAUAUUA	-	7179
9020-Cy3	Short OGN-Cy3	Cy3-C(O)-NH-C ₆ -CAAUUCUUA W UCUAUAUUA	-	7352



Appendix Table 6 | PCR primers for genomic DNA amplification of *EMX1* sequence.

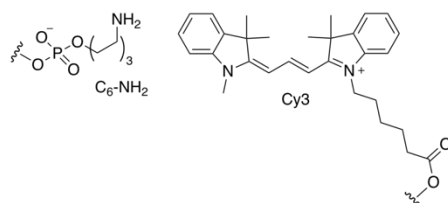
Lowercase bases are DNA nucleotides. Oligonucleotides have both a unique identifier ('res') and a code name.

res	Code	Sequence (5'-3')	Mol. weight (g/mol)	
			Expected	Found
7513	<i>EMX1</i> -Forward Primer	ggagcagctggtcagagggg	6288	6288
7414	<i>EMX1</i> -Reverse Primer	ccataggaagggggacactgg	6874	6874

Appendix Table 7 | Oligonucleotides used in the EMSA assay targeting the *MUC4-E3* gene.

Lowercase bases are DNA nucleotides. Italicised bases are 2'-OMe modified nucleotides. The structures of modifications are shown below the table. The oligonucleotides are given a unique identifier ('res') and a code name.

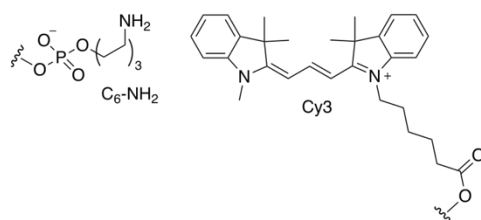
res	Code	Sequence (5'-3')	Mol. weight (g/mol)	
			Expected	Found
8766	Non-target strand <i>MUC4-E3</i>	tgaggaagtgtcggtgacaggaagaggggtg gcgtgacctgtggatgctgaggaagtgtcggt	19864	19868
8767	Target strand <i>MUC4-E3</i>	actccttcacagccactgtccttctcccacc gcactggacacctacgactccttcacagcca	18944	18946
8742	IVT <i>MUC4-E3</i> sgRNA template	aaaagcaccgactcgggtgccacttttcaagttgataac ggactagccttattttaacttgctatttctagctctaaacc agcatccacaggtcacgccctatagtgagtcgtattaga	36822	36831
8738	eOGN+1	NH ₂ -C ₆ -CAGGAAGAGGGG	4323	4323
8739	eOGN+2	NH ₂ -C ₆ -ACAGGAAGAGGG	4307	4307
8740	eOGN+3	NH ₂ -C ₆ -GACAGGAAGAGG	4307	4307
8741	eOGN+4	NH ₂ -C ₆ -UGACAGGAAGAG	4268	4268
8738-Cy3	eOGN+1-Cy3	Cy3-C(O)-NH-C ₆ -CAGGAAGAGGGG	4797	4797
8739-Cy3	eOGN+2-Cy3	Cy3-C(O)-NH-C ₆ -ACAGGAAGAGGG	4781	4781
8740-Cy3	eOGN+3-Cy3	Cy3-C(O)-NH-C ₆ -GACAGGAAGAGG	4781	4781
8741-Cy3	eOGN+4-Cy3	Cy3-C(O)-NH-C ₆ -UGACAGGAAGAG	4742	4742



Appendix Table 8 | Oligonucleotides used for live cell imaging *via* R-loop binding, targeting the telomeres.

Lowercase bases are DNA nucleotides. Italicised bases are 2'-OMe modified nucleotides. The structures of modifications are shown below the table. The oligonucleotides are given a unique identifier ('res') and a code name.

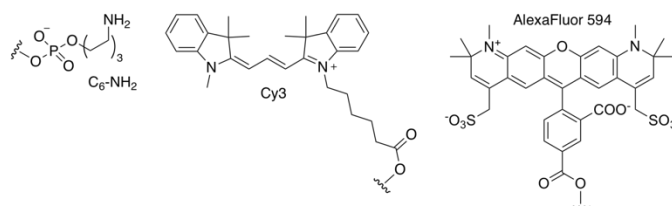
res	Code	Sequence (5'-3')	Mol. weight (g/mol)	
			Expected	Found
6491	IVT telomeric sgRNA template	aaaagcaccgactcgggtgccacttttcaagttgataacggact agccttattttaacttgctatttctagctctaaaactaacctaacc ctaaccctaccctatagtgagtcgtattaga	37319	37324
8696	eOGN+1	NH ₂ -C ₆ -UUAGGGUUAGGG	4238	4238
8697	eOGN+2	NH ₂ -C ₆ -GUUAGGGUUAGG	4238	4238
8698	eOGN+3	NH ₂ -C ₆ -GGUUAGGGUUAG	4238	4238
8699	eOGN+4	NH ₂ -C ₆ -GGGUUAGGGUUA	4238	4238
8696-Cy3	eOGN+1-Cy3	Cy3-C(O)-NH-C ₆ -UUAGGGUUAGGG	4679	4680
8697-Cy3	eOGN+2-Cy3	Cy3-C(O)-NH-C ₆ -GUUAGGGUUAGG	4679	4680
8698-Cy3	eOGN+3-Cy3	Cy3-C(O)-NH-C ₆ -GGUUAGGGUUAG	4679	4680
8699-Cy3	eOGN+4-Cy3	Cy3-C(O)-NH-C ₆ -GGGUUAGGGUUA	4679	4680



Appendix Table 9 | Phosphorothioate-modified oligonucleotides used for live cell imaging *via* R-loop binding, targeting the telomeres.

Lowercase bases are DNA nucleotides. Italicised bases are 2'-OMe modified nucleotides. Underlined bases are modified with phosphorothioate linkages. The structures of modifications are shown below the table. The oligonucleotides are given a unique identifier ('res') and a code name.

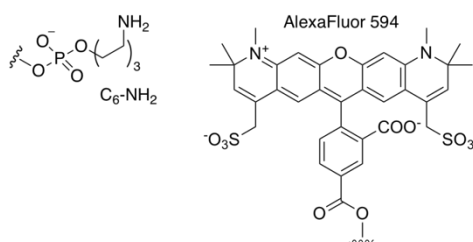
res	Code	Sequence (5'-3')	Mol. weight (g/mol)	
			Expected	Found
6491	IVT telomeric sgRNA template	aaaagcaccgactcgggtgccacttttcaagttgataacggact agccttattttaacttgctatttctagctctaaaactaacccctaac ctaaccctaccctatagtgagtcgtattaga	37319	37324
9201	PS-eOGN+1	NH ₂ -C ₆ - <u>UUAGGGUUAGGG</u>	4430	4431
9202	PS-eOGN+2	NH ₂ -C ₆ - <u>GUUAGGGUUAGG</u>	4430	4431
9203	PS-eOGN+3	NH ₂ -C ₆ - <u>GGUUAGGGUUAG</u>	4430	4431
9204	PS-eOGN+4	NH ₂ -C ₆ - <u>GGGUUAGGGUUA</u>	4430	4431
9201-Cy3	PS-eOGN+1-Cy3	Cy3-C(O)-NH-C ₆ - <u>UUAGGGUUAGGG</u>	4871	4871
9202-Cy3	PS-eOGN+2-Cy3	Cy3-C(O)-NH-C ₆ - <u>GUUAGGGUUAGG</u>	4871	4871
9203-Cy3	ePS-OGN+3-Cy3	Cy3-C(O)-NH-C ₆ - <u>GGUUAGGGUUAG</u>	4871	4871
9204-Cy3	PS-eOGN+4-Cy3	Cy3-C(O)-NH-C ₆ - <u>GGGUUAGGGUUA</u>	4871	4871
9201-AF594	PS-eOGN+1-AF594	AF594-C(O)-NH-C ₆ - <u>UUAGGGUUAGGG</u>	5135	5137
9202-AF594	PS-eOGN+2-AF594	AF594-C(O)-NH-C ₆ - <u>GUUAGGGUUAGG</u>	5135	5139
9203-AF594	ePS-OGN+3-AF594	AF594-C(O)-NH-C ₆ - <u>GGUUAGGGUUAG</u>	5135	5139
9204-AF594	PS-eOGN+4-AF594	AF594-C(O)-NH-C ₆ - <u>GGGUUAGGGUUA</u>	5135	5139



Appendix Table 10 | Scrambled oligonucleotides used to test specificity of R-loop binding, targeting the telomeres.

Lowercase bases are DNA nucleotides. Italicised bases are 2'-OMe modified nucleotides. Underlined bases are modified with phosphorothioate linkages. The structures of modifications are shown below the table. The oligonucleotides are given a unique identifier ('res') and a code name.

res	Code	Sequence (5'-3')	Mol. weight (g/mol)	
			Expected	Found
6491	IVT telomeric sgRNA template	aaaagcaccgactcggtgccacttttcaagttgataacggact agccttattttaacttgctatttctagctctaaaactaacctaacc ctaaccctaccctatagtgcgtgattaga	37319	37324
9581	Scrambled eOGN	NH ₂ -C ₆ - <u>UGAGUGUGAGUG</u>	4430	4431
9581-AF594	Scrambled eOGN+AF594	AF594-C(O)-NH-C ₆ - <u>UGAGUGUGAGUG</u>	5135	5139



References

1. Watson, J. D.; Crick, F. H., Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **1953**, *171* (4356), 737-8.
2. Erwin Chargaff, E. V., Ruth Doniger, Charlotte Green and Fernanda Misani, THE COMPOSITION OF THE DESOXYRIBOSE NUCLEIC ACIDS OF THYMUS AND SPLEEN. *J. Biol. Chem.* **1949**, *177*, 405-416.
3. Franklin, R. E.; Gosling, R. G., Molecular configuration in sodium thymonucleate. *Nature* **1953**, *171* (4356), 740-1.
4. Wilkins, M. H.; Seeds, W. E.; Stokes, A. R.; Wilson, H. R., Helical structure of crystalline deoxyribose nucleic acid. *Nature* **1953**, *172* (4382), 759-62.
5. Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E., Structure of a B-DNA dodecamer: conformation and dynamics. *Proc Natl Acad Sci U S A* **1981**, *78* (4), 2179-83.
6. 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 D Biol Crystallogr* **2014**, *70* (Pt 7), 1790-800.
7. Jiang, F.; Zhou, K.; Ma, L.; Gressel, S.; Doudna, J. A., STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **2015**, *348* (6242), 1477-81.
8. Nucleic Acids in Chemistry and Biology. In *Chapter 2 DNA and RNA Structure* [Online] The Royal Society of Chemistry: 2006; pp. 13–76.
9. Huang, M.; Giese, T. J.; Lee, T. S.; York, D. M., Improvement of DNA and RNA Sugar Pucker Profiles from Semiempirical Quantum Methods. *J Chem Theory Comput* **2014**, *10* (4), 1538-1545.

10. Leulliot, N.; Varani, G., Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry* **2001**, 40 (27), 7947-56.
11. Williamson, J. R., Induced fit in RNA-protein recognition. *Nat Struct Biol* **2000**, 7 (10), 834-7.
12. Aoki, K., Crystallographic and Structural Data II. Saenger, W., Ed. SpringerMaterials: 1989.
13. Rich, A.; Zhang, S., Timeline: Z-DNA: the long road to biological function. *Nat Rev Genet* **2003**, 4 (7), 566-72.
14. Olson, M. V., The human genome project. *Proc. Natl. Acad. Sci. USA* **1993**, 90, 4338-4344.
15. Alberts B, J. A., Lewis J, et al., In *Molecular Biology of the Cell*, 4th ed.; Garland Science: New York, 2002.
16. Crick, F., Central dogma of molecular biology. *Nature* **1970**, 227 (5258), 561-3.
17. Watson, J. D.; Crick, F. H., The structure of DNA. *Cold Spring Harb Symp Quant Biol* **1953**, 18, 123-31.
18. Okazaki, R.; Okazaki, T.; Sakabe, K.; Sugimoto, K.; Sugino, A., Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc Natl Acad Sci U S A* **1968**, 59 (2), 598-605.
19. Alberts, B., DNA replication and recombination. *Nature* **2003**, 421 (6921), 431-5.
20. Guo, J., Transcription: the epicenter of gene expression. *J Zhejiang Univ Sci B* **2014**, 15 (5), 409-11.
21. Nikolov, D. B.; Burley, S. K., RNA polymerase II transcription initiation: a structural view. *Proc Natl Acad Sci U S A* **1997**, 94 (1), 15-22.

22. Kapp, L. D.; Lorsch, J. R., The molecular mechanics of eukaryotic translation. *Annu Rev Biochem* **2004**, *73*, 657-704.
23. Jackson, R. J.; Hellen, C. U.; Pestova, T. V., The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **2010**, *11* (2), 113-27.
24. Carroll, D., Genome Editing: Past, Present, and Future. *Yale J. Biol. Med.* *90* (4), 653–659.
25. Rothstein, R. J., One-step gene disruption in yeast. *Methods Ezymol.* **1983**, *101*, 202-211.
26. Smithies, O.; Gregg, R. G.; Boggs, S. S.; Koralewski, M. A.; Kucherlapati, R. S., Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature* **1985**, *317* (6034), 230-4.
27. Thomas, K., High frequency targeting of genes to specific sites in the mammalian genome. *Cell* **1986**, *44* (3), 419-428.
28. Capecchi, M. R., Altering the genome by homologous recombination. *Science* **1989**, *244* (4910), 1288-92.
29. Mao, Z.; Bozzella, M.; Seluanov, A.; Gorbunova, V., DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle* **2008**, *7* (18), 2902-6.
30. Lin, F. L.; Sperle, K.; Sternberg, N., Recombination in mouse L cells between DNA introduced into cells and homologous chromosomal sequences. *Proc Natl Acad Sci U S A* **1985**, *82* (5), 1391-5.
31. Rouet, P.; Smih, F.; Jasin, M., Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol* **1994**, *14* (12), 8096-106.

32. Bassing, C. H.; Alt, F. W., The cellular response to general and programmed DNA double strand breaks. *DNA Repair (Amst)* **2004**, 3 (8-9), 781-96.
33. Lieber, M. R., The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* **2010**, 79, 181-211.
34. Liang, F.; Han, M.; Romanienko, P. J.; Jasin, M., Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc Natl Acad Sci U S A* **1998**, 95 (9), 5172-7.
35. Lewis, L. K.; Westmoreland, J. W.; Resnick, M. A., Repair of endonuclease-induced double-strand breaks in *Saccharomyces cerevisiae*: essential role for genes associated with nonhomologous end-joining. *Genetics* **1999**, 152 (4), 1513-29.
36. Lewis, L. K.; Kirchner, J. M.; Resnick, M. A., Requirement for end-joining and checkpoint functions, but not RAD52-mediated recombination, after EcoRI endonuclease cleavage of *Saccharomyces cerevisiae* DNA. *Mol Cell Biol* **1998**, 18 (4), 1891-902.
37. Jacquier, A.; Dujon, B., An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* **1985**, 41 (2), 383-394.
38. Seligman, L. M.; Chisholm, K. M.; Chevalier, B. S.; Chadsey, M. S.; Edwards, S. T.; Savage, J. H.; Veillet, A. L., Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Res* **2002**, 30 (17), 3870-9.
39. Sussman, D.; Chadsey, M.; Fauce, S.; Engel, A.; Bruett, A.; Monnat, R., Jr.; Stoddard, B. L.; Seligman, L. M., Isolation and characterization of new homing endonuclease specificities at individual target site positions. *J Mol Biol* **2004**, 342 (1), 31-41.
40. Rosen, L. E.; Morrison, H. A.; Masri, S.; Brown, M. J.; Springstubb, B.; Sussman, D.; Stoddard, B. L.; Seligman, L. M., Homing endonuclease I-Crel

derivatives with novel DNA target specificities. *Nucleic Acids Res* **2006**, 34 (17), 4791-800.

41. Porteus, M., Genome Editing: A New Approach to Human Therapeutics. *Annu Rev Pharmacol Toxicol* **2016**, 56, 163-90.

42. Urnov, F. D.; Rebar, E. J.; Holmes, M. C.; Zhang, H. S.; Gregory, P. D., Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* **2010**, 11 (9), 636-46.

43. Klug, A.; Rhodes, D., Zinc fingers: a novel protein fold for nucleic acid recognition. *Cold Spring Harb Symp Quant Biol* **1987**, 52, 473-82.

44. Liu, Q.; Segal, D. J.; Ghiara, J. B.; Barbas, C. F., 3rd, Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc Natl Acad Sci U S A* **1997**, 94 (11), 5525-30.

45. Bibikova, M.; Carroll, D.; Segal, D. J.; Trautman, J. K.; Smith, J.; Kim, Y. G.; Chandrasegaran, S., Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol* **2001**, 21 (1), 289-97.

46. Porteus, M. H.; Baltimore, D., Chimeric nucleases stimulate gene targeting in human cells. *Science* **2003**, 300 (5620), 763.

47. Joung, J. K.; Sander, J. D., TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* **2013**, 14 (1), 49-55.

48. Miller, J. C. *et al.*, A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **2011**, 29 (2), 143-8.

49. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **2012**, 337 (6096), 816-21.

50. Mali, P.; Yang, L.; Esvelt, K. M.; Aach, J.; Guell, M.; DiCarlo, J. E.; Norville, J. E.; Church, G. M., RNA-guided human genome engineering via Cas9. *Science* **2013**, 339 (6121), 823-6.
51. Wang, H.; Yang, H.; Shivalila, C. S.; Dawlaty, M. M.; Cheng, A. W.; Zhang, F.; Jaenisch, R., One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **2013**, 153 (4), 910-8.
52. Ishino, Y.; Shinagawa, H.; Makino, K.; Amemura, M.; Nakata, A., Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol* **1987**, 169 (12), 5429-33.
53. Mojica, F. J.; Diez-Villasenor, C.; Soria, E.; Juez, G., Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol* **2000**, 36 (1), 244-6.
54. Jansen, R.; Embden, J. D.; Gastra, W.; Schouls, L. M., Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **2002**, 43 (6), 1565-75.
55. Brouns, S. J.; Jore, M. M.; Lundgren, M.; Westra, E. R.; Slijkhuys, R. J.; Snijders, A. P.; Dickman, M. J.; Makarova, K. S.; Koonin, E. V.; van der Oost, J., Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **2008**, 321 (5891), 960-4.
56. Hale, C. R.; Zhao, P.; Olson, S.; Duff, M. O.; Graveley, B. R.; Wells, L.; Terns, R. M.; Terns, M. P., RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **2009**, 139 (5), 945-56.
57. Mojica, F. J.; Diez-Villasenor, C.; Garcia-Martinez, J.; Soria, E., Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **2005**, 60 (2), 174-82.

58. Pourcel, C.; Salvignol, G.; Vergnaud, G., CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **2005**, *151* (Pt 3), 653-663.
59. Bolotin, A.; Quinquis, B.; Sorokin, A.; Ehrlich, S. D., Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **2005**, *151* (Pt 8), 2551-2561.
60. Barrangou, R.; Fremaux, C.; Deveau, H.; Richards, M.; Boyaval, P.; Moineau, S.; Romero, D. A.; Horvath, P., CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **2007**, *315* (5819), 1709-12.
61. Deveau, H.; Barrangou, R.; Garneau, J. E.; Labonte, J.; Fremaux, C.; Boyaval, P.; Romero, D. A.; Horvath, P.; Moineau, S., Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* **2008**, *190* (4), 1390-400.
62. Deltcheva, E.; Chylinski, K.; Sharma, C. M.; Gonzales, K.; Chao, Y.; Pirzada, Z. A.; Eckert, M. R.; Vogel, J.; Charpentier, E., CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **2011**, *471* (7340), 602-7.
63. Garneau, J. E.; Dupuis, M. E.; Villion, M.; Romero, D. A.; Barrangou, R.; Boyaval, P.; Fremaux, C.; Horvath, P.; Magadan, A. H.; Moineau, S., The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **2010**, *468* (7320), 67-71.
64. Makarova, K. S. *et al.*, An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **2015**, *13* (11), 722-36.
65. Redding, S.; Sternberg, S. H.; Marshall, M.; Gibb, B.; Bhat, P.; Guegler, C. K.; Wiedenheft, B.; Doudna, J. A.; Greene, E. C., Surveillance and Processing of Foreign DNA by the *Escherichia coli* CRISPR-Cas System. *Cell* **2015**, *163* (4), 854-65.

66. Gasiunas, G.; Barrangou, R.; Horvath, P.; Siksnys, V., Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **2012**, *109* (39), E2579-86.
67. Mojica, F. J. M.; Diez-Villasenor, C.; Garcia-Martinez, J.; Almendros, C., Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **2009**, *155* (Pt 3), 733-740.
68. Stern, A.; Keren, L.; Wurtzel, O.; Amitai, G.; Sorek, R., Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet* **2010**, *26* (8), 335-40.
69. Gilbert, L. A.; Larson, M. H.; Morsut, L.; Liu, Z.; Brar, G. A.; Torres, S. E.; Stern-Ginossar, N.; Brandman, O.; Whitehead, E. H.; Doudna, J. A.; Lim, W. A.; Weissman, J. S.; Qi, L. S., CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **2013**, *154* (2), 442-51.
70. Westra, E. R.; Semenova, E.; Datsenko, K. A.; Jackson, R. N.; Wiedenheft, B.; Severinov, K.; Brouns, S. J., Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLoS Genet* **2013**, *9* (9), e1003742.
71. Goldberg, G. W.; Jiang, W.; Bikard, D.; Marraffini, L. A., Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature* **2014**, *514* (7524), 633-7.
72. Samai, P.; Pyenson, N.; Jiang, W.; Goldberg, G. W.; Hatoum-Aslan, A.; Marraffini, L. A., Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell* **2015**, *161* (5), 1164-1174.
73. Abudayyeh, O. O. *et al.*, C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **2016**, *353* (6299), aaf5573.

74. East-Seletsky, A.; O'Connell, M. R.; Knight, S. C.; Burstein, D.; Cate, J. H.; Tjian, R.; Doudna, J. A., Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* **2016**, 538 (7624), 270-273.
75. Sapranaukas, R.; Gasiunas, G.; Fremaux, C.; Barrangou, R.; Horvath, P.; Siksnys, V., The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res* **2011**, 39 (21), 9275-82.
76. Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; Zhang, F., Multiplex genome engineering using CRISPR/Cas systems. *Science* **2013**, 339 (6121), 819-23.
77. Nishimasu, H.; Ran, F. A.; Hsu, P. D.; Konermann, S.; Shehata, S. I.; Dohmae, N.; Ishitani, R.; Zhang, F.; Nureki, O., Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **2014**, 156 (5), 935-49.
78. Jinek, M. *et al.*, Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **2014**, 343 (6176), 1247997.
79. Anders, C.; Niewoehner, O.; Duerst, A.; Jinek, M., Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **2014**, 513 (7519), 569-73.
80. Jiang, F.; Taylor, D. W.; Chen, J. S.; Kornfeld, J. E.; Zhou, K.; Thompson, A. J.; Nogales, E.; Doudna, J. A., Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **2016**, 351 (6275), 867-71.
81. Hirano, S.; Nishimasu, H.; Ishitani, R.; Nureki, O., Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9. *Mol Cell* **2016**, 61 (6), 886-94.
82. Jiang, F.; Doudna, J. A., CRISPR-Cas9 Structures and Mechanisms. *Annu Rev Biophys* **2017**, 46, 505-529.

83. Chen, H.; Choi, J.; Bailey, S., Cut site selection by the two nuclease domains of the Cas9 RNA-guided endonuclease. *J Biol Chem* **2014**, 289 (19), 13284-94.
84. Heler, R.; Samai, P.; Modell, J. W.; Weiner, C.; Goldberg, G. W.; Bikard, D.; Marraffini, L. A., Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **2015**, 519 (7542), 199-202.
85. Sternberg, S. H.; Redding, S.; Jinek, M.; Greene, E. C.; Doudna, J. A., DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **2014**, 507 (7490), 62-7.
86. Jinek, M.; East, A.; Cheng, A.; Lin, S.; Ma, E.; Doudna, J., RNA-programmed genome editing in human cells. *Elife* **2013**, 2, e00471.
87. Mekler, V.; Minakhin, L.; Semenova, E.; Kuznedelov, K.; Severinov, K., Kinetics of the CRISPR-Cas9 effector complex assembly and the role of 3'-terminal segment of guide RNA. *Nucleic Acids Res* **2016**, 44 (6), 2837-45.
88. Wright, A. V.; Sternberg, S. H.; Taylor, D. W.; Staahl, B. T.; Bardales, J. A.; Kornfeld, J. E.; Doudna, J. A., Rational design of a split-Cas9 enzyme complex. *Proc Natl Acad Sci U S A* **2015**, 112 (10), 2984-9.
89. Marraffini, L. A.; Sontheimer, E. J., Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **2010**, 463 (7280), 568-71.
90. Knight, S. C.; Xie, L.; Deng, W.; Guglielmi, B.; Witkowsky, L. B.; Bosanac, L.; Zhang, E. T.; El Beheiry, M.; Masson, J. B.; Dahan, M.; Liu, Z.; Doudna, J. A.; Tjian, R., Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science* **2015**, 350 (6262), 823-6.
91. Ma, H.; Tu, L. C.; Naseri, A.; Huisman, M.; Zhang, S.; Grunwald, D.; Pederson, T., CRISPR-Cas9 nuclear dynamics and target recognition in living cells. *J Cell Biol* **2016**, 214 (5), 529-37.

92. Szczelkun, M. D.; Tikhomirova, M. S.; Sinkunas, T.; Gasiunas, G.; Karvelis, T.; Pschera, P.; Siksnys, V.; Seidel, R., Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc Natl Acad Sci U S A* **2014**, *111* (27), 9798-803.
93. Josephs, E. A.; Kocak, D. D.; Fitzgibbon, C. J.; McMenemy, J.; Gersbach, C. A.; Marszalek, P. E., Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Res* **2015**, *43* (18), 8924-41.
94. Siksnys, V.; Gasiunas, G., Rewiring Cas9 to Target New PAM Sequences. *Mol Cell* **2016**, *61* (6), 793-4.
95. Kleinstiver, B. P.; Prew, M. S.; Tsai, S. Q.; Topkar, V. V.; Nguyen, N. T.; Zheng, Z.; Gonzales, A. P.; Li, Z.; Peterson, R. T.; Yeh, J. R.; Aryee, M. J.; Joung, J. K., Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **2015**, *523* (7561), 481-5.
96. Hu, J. H.; Miller, S. M.; Geurts, M. H.; Tang, W.; Chen, L.; Sun, N.; Zeina, C. M.; Gao, X.; Rees, H. A.; Lin, Z.; Liu, D. R., Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **2018**, *556* (7699), 57-63.
97. Ma, D.; Xu, Z.; Zhang, Z.; Chen, X.; Zeng, X.; Zhang, Y.; Deng, T.; Ren, M.; Sun, Z.; Jiang, R.; Xie, Z., Engineer chimeric Cas9 to expand PAM recognition based on evolutionary information. *Nat Commun* **2019**, *10* (1), 560.
98. Chatterjee, P.; Jakimo, N.; Jacobson, J. M., Minimal PAM specificity of a highly similar SpCas9 ortholog. *Sci Adv* **2018**, *4* (10), eaau0766.
99. Fu, Y.; Sander, J. D.; Reyon, D.; Cascio, V. M.; Joung, J. K., Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* **2014**, *32* (3), 279-284.

100. Kleinstiver, B. P.; Pattanayak, V.; Prew, M. S.; Tsai, S. Q.; Nguyen, N. T.; Zheng, Z.; Joung, J. K., High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **2016**, 529 (7587), 490-5.
101. Hultquist, J. F.; Schumann, K.; Woo, J. M.; Manganaro, L.; McGregor, M. J.; Doudna, J.; Simon, V.; Krogan, N. J.; Marson, A., A Cas9 Ribonucleoprotein Platform for Functional Genetic Studies of HIV-Host Interactions in Primary Human T Cells. *Cell Rep* **2016**, 17 (5), 1438-1452.
102. Schumann, K.; Lin, S.; Boyer, E.; Simeonov, D. R.; Subramaniam, M.; Gate, R. E.; Haliburton, G. E.; Ye, C. J.; Bluestone, J. A.; Doudna, J. A.; Marson, A., Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proc Natl Acad Sci U S A* **2015**, 112 (33), 10437-42.
103. Kim, S.; Kim, D.; Cho, S. W.; Kim, J.; Kim, J. S., Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res* **2014**, 24 (6), 1012-9.
104. Vakulskas, C. A. *et al.*, A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat Med* **2018**, 24 (8), 1216-1224.
105. Shen, B.; Zhang, W.; Zhang, J.; Zhou, J.; Wang, J.; Chen, L.; Wang, L.; Hodgkins, A.; Iyer, V.; Huang, X.; Skarnes, W. C., Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods* **2014**, 11 (4), 399-402.
106. Dianov, G. L.; Hubscher, U., Mammalian base excision repair: the forgotten archangel. *Nucleic Acids Res* **2013**, 41 (6), 3483-90.
107. Trevino, A. E.; Zhang, F., Genome editing using Cas9 nickases. *Methods Enzymol* **2014**, 546, 161-74.

108. Ran, F. A.; Hsu, P. D.; Lin, C. Y.; Gootenberg, J. S.; Konermann, S.; Trevino, A. E.; Scott, D. A.; Inoue, A.; Matoba, S.; Zhang, Y.; Zhang, F., Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **2013**, *154* (6), 1380-9.
109. Wang, H.; La Russa, M.; Qi, L. S., CRISPR/Cas9 in Genome Editing and Beyond. *Annu Rev Biochem* **2016**, *85*, 227-64.
110. Fernandes, L. G. V.; Guaman, L. P.; Vasconcellos, S. A.; Heinemann, M. B.; Picardeau, M.; Nascimento, A., Gene silencing based on RNA-guided catalytically inactive Cas9 (dCas9): a new tool for genetic engineering in *Leptospira*. *Sci Rep* **2019**, *9* (1), 1839.
111. Bikard, D.; Jiang, W.; Samai, P.; Hochschild, A.; Zhang, F.; Marraffini, L. A., Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* **2013**, *41* (15), 7429-37.
112. Qi, L. S.; Larson, M. H.; Gilbert, L. A.; Doudna, J. A.; Weissman, J. S.; Arkin, A. P.; Lim, W. A., Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **2013**, *152* (5), 1173-83.
113. Perez-Pinera, P.; Kocak, D. D.; Vockley, C. M.; Adler, A. F.; Kabadi, A. M.; Polstein, L. R.; Thakore, P. I.; Glass, K. A.; Ousterout, D. G.; Leong, K. W.; Guilak, F.; Crawford, G. E.; Reddy, T. E.; Gersbach, C. A., RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods* **2013**, *10* (10), 973-6.
114. Kearns, N. A.; Pham, H.; Tabak, B.; Genga, R. M.; Silverstein, N. J.; Garber, M.; Maehr, R., Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods* **2015**, *12* (5), 401-403.
115. Hilton, I. B.; D'Ippolito, A. M.; Vockley, C. M.; Thakore, P. I.; Crawford, G. E.; Reddy, T. E.; Gersbach, C. A., Epigenome editing by a CRISPR-Cas9-based

acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* **2015**, 33 (5), 510-7.

116. Fujita, T.; Fujii, H., Identification of proteins associated with an IFN γ -responsive promoter by a retroviral expression system for enChIP using CRISPR. *PLoS One* **2014**, 9 (7), e103084.

117. Fujita, T.; Fujii, H., Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochem Biophys Res Commun* **2013**, 439 (1), 132-6.

118. Mir, A.; Alterman, J. F.; Hassler, M. R.; Debacker, A. J.; Hudgens, E.; Echeverria, D.; Brodsky, M. H.; Khvorova, A.; Watts, J. K.; Sontheimer, E. J., Heavily and fully modified RNAs guide efficient SpyCas9-mediated genome editing. *Nat Commun* **2018**, 9 (1), 2641.

119. Yin, H. *et al.*, Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing. *Nat Biotechnol* **2017**, 35 (12), 1179-1187.

120. Rahdar, M.; McMahon, M. A.; Prakash, T. P.; Swayze, E. E.; Bennett, C. F.; Cleveland, D. W., Synthetic CRISPR RNA-Cas9-guided genome editing in human cells. *Proc Natl Acad Sci U S A* **2015**, 112 (51), E7110-7.

121. Ryan, D. E.; Taussig, D.; Steinfeld, I.; Phadnis, S. M.; Lunstad, B. D.; Singh, M.; Vuong, X.; Okochi, K. D.; McCaffrey, R.; Olesiak, M.; Roy, S.; Yung, C. W.; Curry, B.; Sampson, J. R.; Bruhn, L.; Dellinger, D. J., Improving CRISPR-Cas specificity with chemical modifications in single-guide RNAs. *Nucleic Acids Res* **2018**, 46 (2), 792-803.

122. Hendel, A.; Bak, R. O.; Clark, J. T.; Kennedy, A. B.; Ryan, D. E.; Roy, S.; Steinfeld, I.; Lunstad, B. D.; Kaiser, R. J.; Wilkens, A. B.; Bacchetta, R.; Tsalenko, A.;

Dellinger, D.; Bruhn, L.; Porteus, M. H., Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat Biotechnol* **2015**, 33 (9), 985-989.

123. Basila, M.; Kelley, M. L.; Smith, A. V. B., Minimal 2'-O-methyl phosphorothioate linkage modification pattern of synthetic guide RNAs for increased stability and efficient CRISPR-Cas9 gene editing avoiding cellular toxicity. *PLoS One* **2017**, 12 (11), e0188593.

124. Rueda, F. O.; Bista, M.; Newton, M. D.; Goeppert, A. U.; Cuomo, M. E.; Gordon, E.; Kroner, F.; Read, J. A.; Wrigley, J. D.; Rueda, D.; Taylor, B. J. M., Mapping the sugar dependency for rational generation of a DNA-RNA hybrid-guided Cas9 endonuclease. *Nat Commun* **2017**, 8 (1), 1610.

125. Yin, H.; Song, C. Q.; Suresh, S.; Kwan, S. Y.; Wu, Q.; Walsh, S.; Ding, J.; Bogorad, R. L.; Zhu, L. J.; Wolfe, S. A.; Koteliansky, V.; Xue, W.; Langer, R.; Anderson, D. G., Partial DNA-guided Cas9 enables genome editing with reduced off-target activity. *Nat Chem Biol* **2018**, 14 (3), 311-316.

126. Cromwell, C. R.; Sung, K.; Park, J.; Kryslar, A. R.; Jovel, J.; Kim, S. K.; Hubbard, B. P., Incorporation of bridged nucleic acids into CRISPR RNAs improves Cas9 endonuclease specificity. *Nat Commun* **2018**, 9 (1), 1448.

127. Lee, K.; Mackley, V. A.; Rao, A.; Chong, A. T.; Dewitt, M. A.; Corn, J. E.; Murthy, N., Synthetically modified guide RNA and donor DNA are a versatile platform for CRISPR-Cas9 engineering. *Elife* **2017**, 6.

128. Clow, P. A.; Du, M.; Jillette, N.; Taghbalout, A.; Zhu, J. J.; Cheng, A. W., CRISPR-mediated multiplexed live cell imaging of nonrepetitive genomic loci with one guide RNA per locus. *Nat Commun* **2022**, 13 (1), 1871.

129. Beckert, B.; Masquida, B., Synthesis of RNA by in vitro transcription. *Methods Mol Biol* **2011**, *703*, 29-41.
130. Milligan, J. F.; Uhlenbeck, O. C., Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol* **1989**, *180*, 51–62.
131. 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* **1987**, *15* (21), 8783-98.
132. Huang, F.; Yarus, M., 5'-RNA self-capping from guanosine diphosphate. *Biochemistry* **1997**, *36* (22), 6557-63.
133. Brakmann, S.; Grzeszik, An error-prone T7 RNA polymerase mutant generated by directed evolution. *Chembiochem* **2001**, *2*, 212–219.
134. Beaucage, S. L.; Caruthers, M. H., Deoxynucleoside phosphoramidites—A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Letters* **1981**, *22* (20), 1859-1862.
135. Gasparutto, D.; Livache, T.; Bazin, H.; Duplaa, A. M.; Guy, A.; Khorlin, A.; Molko, D.; Roget, A.; Teoule, R., Chemical synthesis of a biologically active natural tRNA with its minor bases. *Nucleic Acids Res* **1992**, *20* (19), 5159-66.
136. Ogilvie, K. K.; Usman, N.; Nicoghossian, K.; Cedergren, R. J., Total chemical synthesis of a 77-nucleotide-long RNA sequence having methionine-acceptance activity. *Proc Natl Acad Sci U S A* **1988**, *85* (16), 5764-8.
137. Dellinger, D. J.; Timar, Z.; Myerson, J.; Sierzchala, A. B.; Turner, J.; Ferreira, F.; Kupihar, Z.; Dellinger, G.; Hill, K. W.; Powell, J. A.; Sampson, J. R.; Caruthers, M. H., Streamlined process for the chemical synthesis of RNA using 2'-O-thionocarbamate-protected nucleoside phosphoramidites in the solid phase. *J Am Chem Soc* **2011**, *133* (30), 11540-56.

138. Misteli, T., Beyond the sequence: cellular organization of genome function. *Cell* **2007**, 128 (4), 787-800.
139. Zink, D.; Sadoni, N.; Stelzer, E., Visualizing chromatin and chromosomes in living cells. *Methods* **2003**, 29 (1), 42-50.
140. Martin, R. M.; Leonhardt, H.; Cardoso, M. C., DNA labeling in living cells. *Cytometry A* **2005**, 67 (1), 45-52.
141. Bauman, J. G. J.; Wiegant, J.; Borst, P.; van Duijn, P., A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochrome-labelled RNA. *Experimental Cell Research* **1980**, 128 (2), 485-490.
142. Singer, R. H.; Ward, D. C., Actin gene expression visualized in chicken muscle tissue culture by using in situ hybridization with a biotinated nucleotide analog. *Proc Natl Acad Sci U S A* **1982**, 79 (23), 7331-5.
143. Amano, T.; Sagai, T.; Tanabe, H.; Mizushina, Y.; Nakazawa, H.; Shiroishi, T., Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev Cell* **2009**, 16 (1), 47-57.
144. Speicher, M. R.; Gwyn Ballard, S.; Ward, D. C., Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nat Genet* **1996**, 12 (4), 368-75.
145. Schrock, E.; du Manoir, S.; Veldman, T.; Schoell, B.; Wienberg, J.; Ferguson-Smith, M. A.; Ning, Y.; Ledbetter, D. H.; Bar-Am, I.; Soenksen, D.; Garini, Y.; Ried, T., Multicolor spectral karyotyping of human chromosomes. *Science* **1996**, 273 (5274), 494-7.
146. Lindhout, B. I.; Fransz, P.; Tessadori, F.; Meckel, T.; Hooykaas, P. J.; van der Zaal, B. J., Live cell imaging of repetitive DNA sequences via GFP-tagged polydactyl zinc finger proteins. *Nucleic Acids Res* **2007**, 35 (16), e107.

147. Miyanari, Y.; Ziegler-Birling, C.; Torres-Padilla, M. E., Live visualization of chromatin dynamics with fluorescent TALEs. *Nat Struct Mol Biol* **2013**, *20* (11), 1321-4.
148. Chen, B.; Gilbert, L. A.; Cimini, B. A.; Schnitzbauer, J.; Zhang, W.; Li, G. W.; Park, J.; Blackburn, E. H.; Weissman, J. S.; Qi, L. S.; Huang, B., Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **2013**, *155* (7), 1479-91.
149. Ma, H.; Naseri, A.; Reyes-Gutierrez, P.; Wolfe, S. A.; Zhang, S.; Pederson, T., Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc Natl Acad Sci U S A* **2015**, *112* (10), 3002-7.
150. 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* **2014**, *159* (3), 635-46.
151. 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* **2018**, *19* (1), 39.
152. Zhao, M. X.; Zeng, E. Z., Application of functional quantum dot nanoparticles as fluorescence probes in cell labeling and tumor diagnostic imaging. *Nanoscale Res Lett* **2015**, *10*, 171.
153. Ma, Y.; Wang, M.; Li, W.; Zhang, Z.; Zhang, X.; Wu, G.; Tan, T.; Cui, Z.; Zhang, X. E., Live Visualization of HIV-1 Proviral DNA Using a Dual-Color-Labeled CRISPR System. *Anal Chem* **2017**, *89* (23), 12896-12901.
154. Tang, H.; Peng, J.; Peng, S.; Wang, Q.; Jiang, X.; Xue, X.; Tao, Y.; Xiang, L.; Ji, Q.; Liu, S. M.; Weng, X.; Zhou, X., Live-cell RNA imaging using the CRISPR-

dCas13 system with modified sgRNAs appended with fluorescent RNA aptamers. *Chem Sci* **2022**, 13 (47), 14032-14040.

155. Ma, H.; Tu, L. C.; Naseri, A.; Huisman, M.; Zhang, S.; Grunwald, D.; Pederson, T., Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nat Biotechnol* **2016**, 34 (5), 528-30.

156. Hur, J. K.; Kim, K.; Been, K. W.; Baek, G.; Ye, S.; Hur, J. W.; Ryu, S. M.; Lee, Y. S.; Kim, J. S., Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins. *Nat Biotechnol* **2016**, 34 (8), 807-8.

157. Anders, C.; Jinek, M., In vitro enzymology of Cas9. *Methods Enzymol* **2014**, 546, 1-20.

158. Zuris, J. A.; Thompson, D. B.; Shu, Y.; Guilinger, J. P.; Bessen, J. L.; Hu, J. H.; Maeder, M. L.; Joung, J. K.; Chen, Z. Y.; Liu, D. R., Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat Biotechnol* **2015**, 33 (1), 73-80.

159. McNaughton, B. R.; Cronican, J. J.; Thompson, D. B.; Liu, D. R., Mammalian cell penetration, siRNA transfection, and DNA transfection by supercharged proteins. *Proc Natl Acad Sci U S A* **2009**, 106 (15), 6111-6.

160. 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* **2019**, 17 (2), 119-128.

161. Wu, X.; Mao, S.; Yang, Y.; Rushdi, M. N.; Krueger, C. J.; Chen, A. K., A CRISPR/molecular beacon hybrid system for live-cell genomic imaging. *Nucleic Acids Res* **2018**, 46 (13), e80.

162. 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* **2019**, *47* (20), e131.
163. Wang, H.; Nakamura, M.; Abbott, T. R.; Zhao, D.; Luo, K.; Yu, C.; Nguyen, C. M.; Lo, A.; Daley, T. P.; La Russa, M.; Liu, Y.; Qi, L. S., CRISPR-mediated live imaging of genome editing and transcription. *Science* **2019**, *365* (6459), 1301-1305.
164. Li, S., Electroporation gene therapy: new developments in vivo and in vitro. *Curr Gene Ther* **2004**, *4* (3), 309-16.
165. 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* **2015**, *112* (38), 11870-5.
166. Manoharan, M., 2'-Carbohydrate modifications in antisense oligonucleotide therapy: importance of conformation, configuration and conjugation. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1999**, *1489* (1), 117-130.
167. Schneider, P. N.; Olthoff, J. T.; Matthews, A. J.; Houston, D. W., Use of fully modified 2'-O-methyl antisense oligos for loss-of-function studies in vertebrate embryos. *Genesis* **2011**, *49* (3), 117-23.
168. Kenski, D. M.; Butora, G.; Willingham, A. T.; Cooper, A. J.; Fu, W.; Qi, N.; Soriano, F.; Davies, I. W.; Flanagan, W. M., siRNA-optimized Modifications for Enhanced In Vivo Activity. *Mol Ther Nucleic Acids* **2012**, *1*, e5.
169. Chesnoy, S.; Huang, L., Structure and function of lipid-DNA complexes for gene delivery. *Annu Rev Biophys Biomol Struct* **2000**, *29*, 27-47.
170. Hirko, A.; Tang, F.; Hughes, J. A., Cationic lipid vectors for plasmid DNA delivery. *Curr Med Chem* **2003**, *10* (14), 1185-93.

171. Liu, D.; Ren, T.; Gao, X., Cationic transfection lipids. *Curr Med Chem* **2003**, *10* (14), 1307-15.
172. Xu, Y.; Szoka, F. C., Jr., Mechanism of DNA release from cationic liposome/DNA complexes used in cell transfection. *Biochemistry* **1996**, *35* (18), 5616-23.
173. Hafez, I. M.; Maurer, N.; Cullis, P. R., On the mechanism whereby cationic lipids promote intracellular delivery of polynucleic acids. *Gene Ther* **2001**, *8* (15), 1188-96.
174. Hoekstra, D.; Rejman, J.; Wasungu, L.; Shi, F.; Zuhorn, I., Gene delivery by cationic lipids: in and out of an endosome. *Biochem Soc Trans* **2007**, *35* (Pt 1), 68-71.
175. Duffy, K.; Arangundy-Franklin, S.; Holliger, P., Modified nucleic acids: replication, evolution, and next-generation therapeutics. *BMC Biol* **2020**, *18* (1), 112.
176. Yoshimura, Y.; Fujimoto, K., Ultrafast reversible photo-cross-linking reaction: toward in situ DNA manipulation. *Org Lett* **2008**, *10* (15), 3227-30.
177. De Fazio, A. F.; Misatziou, D.; Baker, Y. R.; Muskens, O. L.; Brown, T.; Kanaras, A. G., Chemically modified nucleic acids and DNA intercalators as tools for nanoparticle assembly. *Chem Soc Rev* **2021**, *50* (23), 13410-13440.
178. Parrish, J. A.; Fitzpatrick, T. B.; Tanenbaum, L.; Pathak, M. A., Photochemotherapy of psoriasis with oral methoxsalen and longwave ultraviolet light. *N Engl J Med* **1974**, *291* (23), 1207-11.
179. Dubertret, L.; Averbeck, D.; Zajdela, F.; Bisagni, E.; Moustacchi, E.; Touraine, R.; Latarjet, R., Photochemotherapy (PUVA) of psoriasis using 3-carbethoxypsoralen, a non-carcinogenic compound in mice. *Br J Dermatol* **1979**, *101* (4), 379-89.
180. Scott, B. R.; Pathak, M. A.; Mohn, G. R., Molecular and genetic basis of furocoumarin reactions. *Mutation Research/Reviews in Genetic Toxicology* **1976**, *39* (1), 29-74.

181. Wolff KW, F. T., Parrish JA, Gschnait F, Gilchrest B, Hönigsmann H, Pathak MA, Tannenbaum L., Photochemotherapy for psoriasis with orally administered methoxsalen. *Arch Dermatol.* **1976**, 7 (112), 943-50.
182. Bulat V, S. M., Dediol I, Ljubicić I, Bradić L. , The mechanisms of action of phototherapy in the treatment of the most common dermatoses. *Coll Antropol.* **2011**, (35), 147-51.
183. Wong, T.; Hsu, L.; Liao, W., Phototherapy in psoriasis: a review of mechanisms of action. *J Cutan Med Surg* **2013**, 17 (1), 6-12.
184. Seki, T.; Nozu, K.; Kondo, S., Differential causes of mutation and killing in *Escherichia coli* after psoralen plus light treatment: monoadducts and cross-links. *Photochem Photobiol* **1978**, 27 (1), 19-24.
185. Chandrasekaran, A. R.; Rusling, D. A., Triplex-forming oligonucleotides: a third strand for DNA nanotechnology. *Nucleic Acids Res* **2018**, 46 (3), 1021-1037.
186. Rajendran, A.; Endo, M.; Katsuda, Y.; Hidaka, K.; Sugiyama, H., Photo-cross-linking-assisted thermal stability of DNA origami structures and its application for higher-temperature self-assembly. *J Am Chem Soc* **2011**, 133 (37), 14488-91.
187. Maune, H. T.; Han, S. P.; Barish, R. D.; Bockrath, M.; Goddard, W. A., III; Rothmund, P. W.; Winfree, E., Self-assembly of carbon nanotubes into two-dimensional geometries using DNA origami templates. *Nat Nanotechnol* **2010**, 5 (1), 61-6.
188. Douglas, S. M.; Chou, J. J.; Shih, W. M., DNA-nanotube-induced alignment of membrane proteins for NMR structure determination. *Proc Natl Acad Sci U S A* **2007**, 104 (16), 6644-8.

189. Endo, M.; Katsuda, Y.; Hidaka, K.; Sugiyama, H., Regulation of DNA methylation using different tensions of double strands constructed in a defined DNA nanostructure. *J Am Chem Soc* **2010**, *132* (5), 1592-7.
190. Sannohe, Y.; Endo, M.; Katsuda, Y.; Hidaka, K.; Sugiyama, H., Visualization of dynamic conformational switching of the G-quadruplex in a DNA nanostructure. *J Am Chem Soc* **2010**, *132* (46), 16311-3.
191. Kershner, R. J.; Bozano, L. D.; Micheel, C. M.; Hung, A. M.; Fornof, A. R.; Cha, J. N.; Rettner, C. T.; Bersani, M.; Frommer, J.; Rothmund, P. W.; Wallraff, G. M., Placement and orientation of individual DNA shapes on lithographically patterned surfaces. *Nat Nanotechnol* **2009**, *4* (9), 557-61.
192. Castro, C. E.; Kilchherr, F.; Kim, D. N.; Shiao, E. L.; Wauer, T.; Wortmann, P.; Bathe, M.; Dietz, H., A primer to scaffolded DNA origami. *Nat Methods* **2011**, *8* (3), 221-9.
193. F. Patolsky, Y. W., O. Lioubashevski and I. Willner, Au-Nanoparticle Nanowires Based on DNA and Polylysine Templates. *Angew. Chem* **2002**, *41* (13), 2323-2327.
194. Li, F.; Dong, Y.; Zhang, Z.; Lv, M.; Wang, Z.; Ruan, X.; Yang, D., A recyclable biointerface based on cross-linked branched DNA nanostructures for ultrasensitive nucleic acid detection. *Biosens Bioelectron* **2018**, *117*, 562-566.
195. Fujimoto, K.; Matsuda, S.; Yoshimura, Y.; Ami, T.; Saito, I., Reversible photopadlocking on double-stranded DNA. *Chem Commun (Camb)* **2007**, (28), 2968-70.
196. Nilsson, M.; Krejci, K.; Koch, J.; Kwiatkowski, M.; Gustavsson, P.; Landegren, U., Padlock probes reveal single-nucleotide differences, parent of origin and in situ distribution of centromeric sequences in human chromosomes 13 and 21. *Nat Genet* **1997**, *16* (3), 252-5.

197. Fire, A.; Xu, S. Q., Rolling replication of short DNA circles. *Proc Natl Acad Sci U S A* **1995**, 92 (10), 4641-5.
198. Yoshinaga Yoshimura, T. O., Hajime Okada, Kenzo Fujimoto, A New Approach for Reversible RNA Photocrosslinking Reaction: Application to Sequence-Specific RNA Selection. *Chem bio chem* **2009**.
199. Fujimoto, K.; Konishi-Hiratsuka, K.; Sakamoto, T.; Yoshimura, Y., Site-specific cytosine to uracil transition by using reversible DNA photo-crosslinking. *Chembiochem* **2010**, 11 (12), 1661-4.
200. Fujimoto, K.; Matsuda, S.; Yoshimura, Y.; Matsumura, T.; Hayashi, M.; Saito, I., Site-specific transition of cytosine to uracil via reversible DNA photoligation. *Chem Commun (Camb)* **2006**, (30), 3223-5.
201. Ogasawara, S.; Yoshimura, Y.; Hayashi, M.; Saito, I.; Fujimoto, K., Highly Efficient Method for Constructing a Single-Stranded Comb-Like Oligonucleotide via Reversible Photocrosslinking. *Bulletin of the Chemical Society of Japan* **2007**, 80 (11), 2124-2130.
202. Matsumura, T.; Ogino, M.; Nagayoshi, K.; Fujimoto, K., Photochemical Site-specific Mutation of 5-Methylcytosine to Thymine. *Chemistry Letters* **2008**, 37 (1), 94-95.
203. Tanaka, K.; Okamoto, A., Degradation of DNA by bisulfite treatment. *Bioorg Med Chem Lett* **2007**, 17 (7), 1912-5.
204. Raizis, A. M.; Schmitt, F.; Jost, J. P., A bisulfite method of 5-methylcytosine mapping that minimizes template degradation. *Anal Biochem* **1995**, 226 (1), 161-6.
205. Frommer, M.; McDonald, L. E.; Millar, D. S.; Collis, C. M.; Watt, F.; Grigg, G. W.; Molloy, P. L.; Paul, C. L., A genomic sequencing protocol that yields a positive

display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A* **1992**, 89 (5), 1827-31.

206. Shigeno, A.; Sakamoto, T.; Yoshimura, Y.; Fujimoto, K., Quick regulation of mRNA functions by a few seconds of photoirradiation. *Org Biomol Chem* **2012**, 10 (38), 7820-5.

207. Watanabe, M.; Nobuta, A.; Tanaka, J.; Asaka, M., An effect of K-ras gene mutation on epidermal growth factor receptor signal transduction in PANC-1 pancreatic carcinoma cells. *International Journal of Cancer* **1996**, 67 (2), 264-268.

208. Nakamura, S.; Fujimoto, K., Creation of DNA array structure equipped with heat resistance by ultrafast photocrosslinking. *Journal of Chemical Technology & Biotechnology* **2014**, 89 (7), 1086-1090.

209. Chhabra, R.; Sharma, J.; Ke, Y.; Liu, Y.; Rinker, S.; Lindsay, S.; Yan, H., Spatially addressable multiprotein nanoarrays templated by aptamer-tagged DNA nanoarchitectures. *J Am Chem Soc* **2007**, 129 (34), 10304-5.

210. Murata, S.; Konagaya, A.; Kobayashi, S.; Saito, H.; Hagiya, M., Molecular Robotics: A New Paradigm for Artifacts. *New Generation Computing* **2013**, 31 (1), 27-45.

211. Erie, D.; Sinha, N.; Olson, W.; Jones, R.; Breslauer, K., A dumbbell-shaped, double-hairpin structure of DNA: a thermodynamic investigation. *Biochemistry* **1987**, 26 (22), 7150-9.

212. Giesen, U.; Kleider, W.; Berding, C.; Geiger, A.; Orum, H.; Nielsen, P. E., A formula for thermal stability (T_m) prediction of PNA/DNA duplexes. *Nucleic Acids Res* **1998**, 26 (21), 5004-6.

213. Christensen, U.; Jacobsen, N.; Rajwanshi, V. K.; Wengel, J.; Koch, T., Stopped-flow kinetics of locked nucleic acid (LNA)-oligonucleotide duplex formation: studies of LNA-DNA and DNA-DNA interactions. *Biochem J* **2001**, 354 (Pt 3), 481-4.
214. Harimech, P. K.; Gerrard, S. R.; El-Sagheer, A. H.; Brown, T.; Kanaras, A. G., Reversible Ligation of Programmed DNA-Gold Nanoparticle Assemblies. *J Am Chem Soc* **2015**, 137 (29), 9242-5.
215. Barrow, S. J.; Funston, A. M.; Wei, X.; Mulvaney, P., DNA-directed self-assembly and optical properties of discrete 1D, 2D and 3D plasmonic structures. *Nano Today* **2013**, 8 (2), 138-167.
216. Xu, L.; Yan, W.; Ma, W.; Kuang, H.; Wu, X.; Liu, L.; Zhao, Y.; Wang, L.; Xu, C., SERS encoded silver pyramids for attomolar detection of multiplexed disease biomarkers. *Adv Mater* **2015**, 27 (10), 1706-11.
217. Gao, X.; Tsang, J. C.; Gaba, F.; Wu, D.; Lu, L.; Liu, P., Comparison of TALE designer transcription factors and the CRISPR/dCas9 in regulation of gene expression by targeting enhancers. *Nucleic Acids Res* **2014**, 42 (20), e155.
218. Thakore, P. I.; D'Ippolito, A. M.; Song, L.; Safi, A.; Shivakumar, N. K.; Kabadi, A. M.; Reddy, T. E.; Crawford, G. E.; Gersbach, C. A., Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods* **2015**, 12 (12), 1143-9.
219. Maeder, M. L.; Linder, S. J.; Cascio, V. M.; Fu, Y.; Ho, Q. H.; Joung, J. K., CRISPR RNA-guided activation of endogenous human genes. *Nat Methods* **2013**, 10 (10), 977-9.
220. Farzadfard, F.; Perli, S. D.; Lu, T. K., Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS Synth Biol* **2013**, 2 (10), 604-13.

221. Chavez, A. *et al.*, Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* **2015**, 12 (4), 326-8.
222. Konermann, S.; Brigham, M. D.; Trevino, A. E.; Joung, J.; Abudayyeh, O. O.; Barcena, C.; Hsu, P. D.; Habib, N.; Gootenberg, J. S.; Nishimasu, H.; Nureki, O.; Zhang, F., Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **2015**, 517 (7536), 583-8.
223. Black, J. B.; Gersbach, C. A., Synthetic transcription factors for cell fate reprogramming. *Curr Opin Genet Dev* **2018**, 52, 13-21.
224. Thakore, P. I.; Black, J. B.; Hilton, I. B.; Gersbach, C. A., Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat Methods* **2016**, 13 (2), 127-37.
225. Choudhury, S. R.; Cui, Y.; Lubecka, K.; Stefanska, B.; Irudayaraj, J., CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. *Oncotarget* **2016**, 7 (29), 46545-46556.
226. Lam, S. S.; Martell, J. D.; Kamer, K. J.; Deerinck, T. J.; Ellisman, M. H.; Mootha, V. K.; Ting, A. Y., Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat Methods* **2015**, 12 (1), 51-4.
227. Myers, S. A.; Wright, J.; Peckner, R.; Kalish, B. T.; Zhang, F.; Carr, S. A., Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling. *Nat Methods* **2018**, 15 (6), 437-439.
228. Rouet, R. *et al.*, Receptor-Mediated Delivery of CRISPR-Cas9 Endonuclease for Cell-Type-Specific Gene Editing. *J Am Chem Soc* **2018**, 140 (21), 6596-6603.
229. Tang, W.; Hu, J. H.; Liu, D. R., Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nat Commun* **2017**, 8, 15939.

230. Robertson, M. P.; Ellington, A. D., In vitro selection of an allosteric ribozyme that transduces analytes to amplicons. *Nat Biotechnol* **1999**, *17* (1), 62-6.
231. Zalatan, J. G.; Lee, M. E.; Almeida, R.; Gilbert, L. A.; Whitehead, E. H.; La Russa, M.; Tsai, J. C.; Weissman, J. S.; Dueber, J. E.; Qi, L. S.; Lim, W. A., Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **2015**, *160* (1-2), 339-50.
232. Holloman, W. K., Unraveling the mechanism of BRCA2 in homologous recombination. *Nat Struct Mol Biol* **2011**, *18* (7), 748-54.
233. Durai, S.; Mani, M.; Kandavelou, K.; Wu, J.; Porteus, M. H.; Chandrasegaran, S., Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res* **2005**, *33* (18), 5978-90.
234. Zeng, Y.; Cui, Y.; Zhang, Y.; Zhang, Y.; Liang, M.; Chen, H.; Lan, J.; Song, G.; Lou, J., The initiation, propagation and dynamics of CRISPR-SpyCas9 R-loop complex. *Nucleic Acids Res* **2018**, *46* (1), 350-361.
235. Elskens, J.; Madder, A., Crosslinker-modified nucleic acid probes for improved target identification and biomarker detection. *RSC Chem Biol* **2021**, *2* (2), 410-422.
236. Jore, M. M. *et al.*, Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* **2011**, *18* (5), 529-36.
237. Rutkauskas, M.; Sinkunas, T.; Songailiene, I.; Tikhomirova, M. S.; Siksnys, V.; Seidel, R., Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. *Cell Rep* **2015**, *10* (9), 1534-1543.
238. Blosser, T. R.; Loeff, L.; Westra, E. R.; Vlot, M.; Kunne, T.; Sobota, M.; Dekker, C.; Brouns, S. J. J.; Joo, C., Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Mol Cell* **2015**, *58* (1), 60-70.

239. Semenova, E.; Jore, M. M.; Datsenko, K. A.; Semenova, A.; Westra, E. R.; Wanner, B.; van der Oost, J.; Brouns, S. J.; Severinov, K., Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* **2011**, *108* (25), 10098-103.
240. Wiedenheft, B.; van Duijn, E.; Bultema, J. B.; Waghmare, S. P.; Zhou, K.; Barendregt, A.; Westphal, W.; Heck, A. J.; Boekema, E. J.; Dickman, M. J.; Doudna, J. A., RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci U S A* **2011**, *108* (25), 10092-7.
241. Singh, D.; Sternberg, S. H.; Fei, J.; Doudna, J. A.; Ha, T., Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat Commun* **2016**, *7*, 12778.
242. Sternberg, S. H.; LaFrance, B.; Kaplan, M.; Doudna, J. A., Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **2015**, *527* (7576), 110-3.
243. Lim, Y.; Bak, S. Y.; Sung, K.; Jeong, E.; Lee, S. H.; Kim, J.-S.; Bae, S.; Kim, S. K., Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease. *Nature Communications* **2016**, *7* (1).
244. Dagdas, Y. S.; Chen, J. S.; Sternberg, S. H.; Doudna, J. A.; Yildiz, A., A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci Adv* **2017**, *3* (8), eaao0027.
245. Anzalone, A. V.; Randolph, P. B.; Davis, J. R.; Sousa, A. A.; Koblan, L. W.; Levy, J. M.; Chen, P. J.; Wilson, C.; Newby, G. A.; Raguram, A.; Liu, D. R., Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **2019**, *576* (7785), 149-157.

246. Liu, Y.; Kao, H. I.; Bambara, R. A., Flap endonuclease 1: a central component of DNA metabolism. *Annu Rev Biochem* **2004**, *73*, 589-615.
247. Keijzers, G.; Bohr, V. A.; Rasmussen, L. J., Human exonuclease 1 (EXO1) activity characterization and its function on flap structures. *Biosci Rep* **2015**, *35* (3).
248. Zhao, C.; Wang, Y.; Nie, X.; Han, X.; Liu, H.; Li, G.; Yang, G.; Ruan, J.; Ma, Y.; Li, X.; Cheng, H.; Zhao, S.; Fang, Y.; Xie, S., Evaluation of the effects of sequence length and microsatellite instability on single-guide RNA activity and specificity. *Int J Biol Sci* **2019**, *15* (12), 2641-2653.
249. Shojaei Baghini, S.; Gardanova, Z. R.; Zekiy, A. O.; Shomali, N.; Tosan, F.; Jarahian, M., Optimizing sgRNA to Improve CRISPR/Cas9 Knockout Efficiency: Special Focus on Human and Animal Cell. *Front Bioeng Biotechnol* **2021**, *9*, 775309.
250. Matson, A. W.; Hosny, N.; Swanson, Z. A.; Hering, B. J.; Burlak, C., Optimizing sgRNA length to improve target specificity and efficiency for the GGTA1 gene using the CRISPR/Cas9 gene editing system. *PLoS One* **2019**, *14* (12), e0226107.
251. Hellman, L. M.; Fried, M. G., Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc* **2007**, *2* (8), 1849-61.
252. Garner, M. M.; Revzin, A., The use of gel electrophoresis to detect and study nucleic acid— protein interactions. *Trends in Biochemical Sciences* **1986**, *11* (10), 395-396.
253. Fried, M. G., Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay. *Electrophoresis* **1989**, *10* (5-6), 366-76.
254. Eckstein, F., Phosphorothioates, essential components of therapeutic oligonucleotides. *Nucleic Acid Ther* **2014**, *24* (6), 374-87.
255. Gaus, H. J.; Gupta, R.; Chappell, A. E.; Ostergaard, M. E.; Swayze, E. E.; Seth, P. P., Characterization of the interactions of chemically-modified therapeutic nucleic

acids with plasma proteins using a fluorescence polarization assay. *Nucleic Acids Res* **2019**, 47 (3), 1110-1122.

256. Seth, P. P.; Tanowitz, M.; Bennett, C. F., Selective tissue targeting of synthetic nucleic acid drugs. *J Clin Invest* **2019**, 129 (3), 915-925.

257. Crooke, S. T.; Wang, S.; Vickers, T. A.; Shen, W.; Liang, X. H., Cellular uptake and trafficking of antisense oligonucleotides. *Nat Biotechnol* **2017**, 35 (3), 230-237.

258. Hyjek-Skladanowska, M.; Vickers, T. A.; Napiorkowska, A.; Anderson, B. A.; Tanowitz, M.; Crooke, S. T.; Liang, X. H.; Seth, P. P.; Nowotny, M., Origins of the Increased Affinity of Phosphorothioate-Modified Therapeutic Nucleic Acids for Proteins. *J Am Chem Soc* **2020**, 142 (16), 7456-7468.

259. Juliano, R. L.; Carver, K., Cellular uptake and intracellular trafficking of oligonucleotides. *Adv Drug Deliv Rev* **2015**, 87, 35-45.

260. Wang, S.; Allen, N.; Liang, X. H.; Crooke, S. T., Membrane Destabilization Induced by Lipid Species Increases Activity of Phosphorothioate-Antisense Oligonucleotides. *Mol Ther Nucleic Acids* **2018**, 13, 686-698.

261. Wang, S.; Allen, N.; Prakash, T. P.; Liang, X. H.; Crooke, S. T., Lipid Conjugates Enhance Endosomal Release of Antisense Oligonucleotides Into Cells. *Nucleic Acid Ther* **2019**, 29 (5), 245-255.

262. Crooke, S. T.; Vickers, T. A.; Liang, X. H., Phosphorothioate modified oligonucleotide-protein interactions. *Nucleic Acids Res* **2020**, 48 (10), 5235-5253.

263. Crooke, S. T.; Seth, P. P.; Vickers, T. A.; Liang, X. H., The Interaction of Phosphorothioate-Containing RNA Targeted Drugs with Proteins Is a Critical Determinant of the Therapeutic Effects of These Agents. *J Am Chem Soc* **2020**, 142 (35), 14754-14771.

264. Chin, T. K.; Perloff, J. K.; Williams, R. G.; Jue, K.; Mohrmann, R., Isolated noncompaction of left ventricular myocardium. A study of eight cases. *Circulation* **1990**, 82 (2), 507-13.
265. Amery, A.; Birkenhager, W.; Bulpitt, C. J.; Clement, D.; De Leeuw, P.; Dollery, C. T.; Fagard, R.; Fletcher, A.; Forette, F.; Leonetti, G.; et al., Syst-Eur. A multicentre trial on the treatment of isolated systolic hypertension in the elderly: objectives, protocol, and organization. *Aging (Milano)* **1991**, 3 (3), 287-302.
266. Fisher, W.; Piazza, C.; Cataldo, M.; Harrell, R.; Jefferson, G.; Conner, R., Functional communication training with and without extinction and punishment. *J Appl Behav Anal* **1993**, 26 (1), 23-36.
267. Shen, W. *et al.*, Chemical modification of PS-ASO therapeutics reduces cellular protein-binding and improves the therapeutic index. *Nat Biotechnol* **2019**, 37 (6), 640-650.
268. Migawa, M. T. *et al.*, Site-specific replacement of phosphorothioate with alkyl phosphonate linkages enhances the therapeutic profile of gapmer ASOs by modulating interactions with cellular proteins. *Nucleic Acids Res* **2019**, 47 (11), 5465-5479.
269. Lu, T.; Lin, Z.; Ren, J.; Yao, P.; Wang, X.; Wang, Z.; Zhang, Q., The Non-Specific Binding of Fluorescent-Labeled MiRNAs on Cell Surface by Hydrophobic Interaction. *PLoS One* **2016**, 11 (3), e0149751.
270. Moreira, B. G.; You, Y.; Owczarzy, R., Cy3 and Cy5 dyes attached to oligonucleotide terminus stabilize DNA duplexes: predictive thermodynamic model. *Biophys Chem* **2015**, 198, 36-44.
271. Dunne, J.; Drescher, B.; Riehle, H.; Hadwiger, P.; Young, B. D.; Krauter, J.; Heidenreich, O., The apparent uptake of fluorescently labeled siRNAs by

electroporated cells depends on the fluorochrome. *Oligonucleotides* **2003**, 13 (5), 375-80.

272. Barreau, C.; Dutertre, S.; Paillard, L.; Osborne, H. B., Liposome-mediated RNA transfection should be used with caution. *RNA* **2006**, 12 (10), 1790-3.

273. Hughes, L. D.; Rawle, R. J.; Boxer, S. G., Choose your label wisely: water-soluble fluorophores often interact with lipid bilayers. *PLoS One* **2014**, 9 (2), e87649.

274. Zhang, Z.; Yomo, D.; Gradinaru, C., Choosing the right fluorophore for single-molecule fluorescence studies in a lipid environment. *Biochim Biophys Acta Biomembr* **2017**, 1859 (7), 1242-1253.

275. Lam, K.; Tajkhorshid, E., Membrane Interactions of Cy3 and Cy5 Fluorophores and Their Effects on Membrane-Protein Dynamics. *Biophys J* **2020**, 119 (1), 24-34.

276. Jones, P. R.; Rothenberger, S. D., Nucleophilic aromatic substitution with elimination in a dinitrosalicylic lactone or ester via Meisenheimer intermediates. *The Journal of Organic Chemistry* **2002**, 51 (15), 3016-3023.

277. Atkinson, S. F.; Bettinger, T.; Seymour, L. W.; Behr, J. P.; Ward, C. M., Conjugation of folate via gelonin carbohydrate residues retains ribosomal-inactivating properties of the toxin and permits targeting to folate receptor positive cells. *J Biol Chem* **2001**, 276 (30), 27930-5.

278. Ak, G.; Sanlier, S. H., Synthesis of folate receptor-targeted and Doxorubicin-coupled chemotherapeutic nanoconjugate and research into its medical applications. *Prep Biochem Biotechnol* **2012**, 42 (6), 551-63.

279. Alexander, C. M.; Hamner, K. L.; Maye, M. M.; Dabrowiak, J. C., Multifunctional DNA-gold nanoparticles for targeted doxorubicin delivery. *Bioconjug Chem* **2014**, 25 (7), 1261-71.