

Bioconjugation Strategies for Oligonucleotide Functionalisation



Gerald Chik

St John's College, University of Oxford

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

Declaration

The author of this thesis declares that the thesis titled “Bioconjugation Strategies for Oligonucleotide Functionalisation” and the corresponding research presented in this thesis was performed and written by the author under the supervision of Professor Tom Brown at the University of Oxford. All experiments, except oligonucleotide synthesis which was conducted by Prof. Afaf El-Sagheer, were conducted by the author. The research is original and has not been submitted in full or in part for any other degree at this or at any other university. The copyright of this thesis rests with the author. Figures generated and licensed by third-party provider are acknowledged in relevant sections.

Table of Contents

Declaration.....	2
Table of Contents.....	3
Acknowledgements.....	7
List of Abbreviations.....	8
Abstract	11
Chapter 1 – Introduction.....	12
1.1 Nucleic Acids Structures and Nomenclature.....	13
1.1.1 Primary Structures of Nucleic Acids.....	13
1.1.2 Secondary Structures of Nucleic Acids	16
1.2 Biological Functions of Nucleic Acids	20
1.2.1 Central Dogma of Biology.....	20
1.2.2 Flow of Genetic information and Its Implications	21
1.2.3 Applications of Nucleic Acids in Medicine and Biotechnology.....	26
1.3 Chemical Synthesis of Oligonucleotides	31
1.3.1 Solid-Phase Synthesis.....	31
1.3.2 Protecting Groups and Deprotection.....	35
1.3.3 Chemical Modifications	37
1.3.4 Synthetic Strategies for Modified Oligonucleotides.....	40
1.3.5 Enzymatic Synthesis of Ribonucleic Acids (RNA).....	45
1.3.6 RNA Modifications.....	47
1.4 Aims and Objectives.....	51
Chapter 2 – Synthesis of Nucleic Acid Dendrons by Click Chemistry	52
2.1 Background	53
2.1.1 Nucleic Acid Dendrons and Dendritic Structures	53
2.1.2 Challenges in the Synthesis of Nucleic Acid Dendrons	55
2.1.3 Click Chemistry on Solid Support.....	56
2.2 Aims and Objectives.....	58
2.3 Solid-Phase Synthesis of Click Dendrons.....	59
2.3.1 Azide Conversion of 5'-hydroxyl in Oligonucleotide SPS.....	59
2.3.2 Synthesis of Multivalent Linkers.....	62
2.3.3 Optimisation of CuAAC Crosslinking.....	64
2.3.4 Syntheses of DNA and RNA Dendrons.....	68

2.3.5	Construction of Nonameric Oligonucleotide Dendrons.....	72
2.4	Properties of Click Dendrons	77
2.4.1	Duplex Hybridisation Study	77
2.4.2	Enzymatic Stability Assay	80
2.4.3	RNase H Activation	82
2.5	Conclusions	84
2.6	Future Outlook	85
Chapter 3 – Hepatocyte-Specific Delivery of CRISPR/Cas9 system		88
3.1	Background	89
3.1.1	The Origin of CRISPR.....	89
3.1.2	Mechanism and Functions of CRISPR in Prokaryotes	90
3.1.3	CRISPR-Cas9 in Mammalian Genome Editing	91
3.1.4	Challenges in Intracellular Delivery of CRISPR.....	94
3.1.5	ASGPr-Mediated Delivery to Hepatocytes.....	96
3.1.6	T7E1 Mismatch Detection Assay	98
3.2	Aims and Objectives.....	100
3.3	Synthesis of GalNAc-gRNA Conjugates via Click Chemistry.....	101
3.3.1	Design and Synthesis of Triantennary GalNAc Cluster	101
3.3.2	Conjugation Strategies for GalNAc-gRNA Constructs via SPAAC	103
3.4	Synthesis of GalNAc-gRNA Conjugates via Reductive Amination	107
3.4.1	3'-Azide Labelling on RNA by Reductive Amination	107
3.4.2	Synthesis of Triantennary GalNAc-ethylenediamine Conjugate	110
3.4.3	GalNAc-conjugation on RNA by Reductive Amination	110
3.5	Biological Activities of GalNAc-gRNA Conjugates.....	113
3.5.1	In vitro Cleavage Assay	113
3.5.2	Delivery of RNP with gRNA-GalNAc Conjugate into Hepatocytes.....	115
3.6	Dual Conjugation of gRNA for ASGPr-Mediated Uptake in Hepatocytes	117
3.6.1	Dual Conjugation with Endosomal Escape Domain.....	117
3.6.2	Biological Activities of gRNA-GalNAc-EED Conjugate.....	120
3.6.3	Imaging of RNP with gRNA-GalNAc-Cy5 Conjugate in HEPG2 Culture	122
3.7	Conclusions	126
3.8	Future Outlook	127
Chapter 4 – Bioconjugation of mRNA Fragments		129
4.1	Background	130
4.1.1	Eukaryotic pre-mRNA processing.....	130

4.1.2	The Essential Roles and Properties of Poly(A) Tail	133
4.1.3	3'-Untranslated Region (3'-UTR) and Non-canonical Poly(A) Tail.....	135
4.2	Aims and Objectives	137
4.3	Ligation Strategies and Oligonucleotide Designs	138
4.3.1	Synthesis of Ethylenediamine (EDA) Phosphoramidite	138
4.3.2	Splinted Ligation	139
4.3.3	Self-Templated Ligation	146
4.4	Conclusions	150
4.5	Future Outlook	151
Chapter 5 – Materials and Methods		153
5.1	Experimental Procedures	154
5.1.1	UV Melting Studies	154
5.1.2	Foetal Bovine Serum Assay	154
5.1.3	RNase H Assay	154
5.1.4	In Vitro Cleavage Assay	155
5.1.5	Cell Culture	155
5.1.6	Naked Transfection to Hepatocytes and T7E1 Assay.....	156
5.1.7	Cell Preparation for Microscopy	156
5.1.8	Live Cell Imaging	157
5.2	Synthesis, Purification and Characterisation of Oligonucleotides	157
5.2.1	Synthesis of Oligonucleotides	157
5.2.2	Deprotection of 2'-OTBDMS Protecting Groups on RNA.....	158
5.2.3	Oligonucleotide 5'-Azide Conversion on Solid Support	158
5.2.4	Copper-Catalysed Click-Crosslinking of 5'-Azido Oligonucleotides	159
5.2.5	NHS Ester Labelling	159
5.2.6	Azide Labelling by PAP Enzyme	160
5.2.7	Purification by HPLC	160
5.2.8	Purification by Gel Electrophoresis	161
5.2.9	Characterisation by Mass Spectrometry	161
5.2.10	Sampling of Oligonucleotide Samples from CPG	161
5.3	Organic Synthesis	162
5.3.1	General Procedures	162
Chapter 6 – Appendix		193
6.1	List of Oligonucleotides.....	194
6.2	NMR Spectra	197

6.3	High-Resolution Mass Spectra	209
6.4	Supplementary Experimental Data	213
References		220

Acknowledgements

I would like to thank Prof. Tom Brown for offering me a fantastic opportunity to conduct research in this exciting field of nucleic acid chemistry. Your guidance and intellectual input have been instrumental opened up a lot of possibilities in my research. I would also like to extend my gratitude to Prof. Afaf El-Sagheer for her support in bringing in new perspectives as well as managing the lab to its best possible conditions. For the work conducted in the cell culture facility, I would thank Prof. Angela Russell for providing me access and it has been a great time working with her lab members.

Working in this lab cannot be as pleasant without all Brownies. I would like to thank Dr. Ysobel Baker, Dr. Nicolò Zuin Fantoni and Dr. Cheng Jin for sharing valuable knowledge and insights in the oligo chemistry space. Dr. Lapatrada Taemaitree (Pui) has been supportive in my biological experiments and instructing me on techniques which I truly appreciate. I would also give credit to Nicolò and Pui for providing me emotional support during some down times in my DPhil study and always being encouraging. I am happy to have worked with everyone in the Brown group who has been really supportive and making the lab a fun place: Sven, Cam, Thanishta, Brendan, Phoebe, Siqi, Lillian, Ben and Alice – thank you very much!

I shall extend my gratitude to Chunsen, Sebastian and Diallo who have been great companions to share laughter and joy with. My DPhil experience would not be as enjoyable without any of you. I would specifically thank Sebastian for sharing a lot of his knowledge and experience which are invaluable for my research as well as being a great collaborator on multiple projects.

Finally, I would like to thank my family and friends who have supported me emotionally over my DPhil journey.

List of Abbreviations

1,2-DCE	1,2-dichloroethane
2'-F	2'-fluoro
2'-MOE	2'-O methoxyethyl
2'-OMe	2'-O methyl
A	Adenine
ASGPr	Asialoglycoprotein receptors
A-site	Aminoacyl site
ASO	Antisense oligonucleotides
ATP	Adenosine triphosphate
Boc	<i>tert</i> -butyloxycarbonyl
C	Cytosine
Cas	CRISPR-associated protein
^{CNV} K	3-cyanovinylcarbazole
CPG	Controlled-pore glass
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
CuAAC	Copper-catalysed Azide-Alkyne Cycloaddition
Cy3	Cyanine 3
Cy5	Cyanine 5
DBCO	Dibenzocyclooctyne
DCM	Dichloromethane
DDR	DNA damage response
DIPEA	N,N-Diisopropylethylamine
DMF	N,N-dimethylformamide
DMSO	Dimethyl sulfoxide
DMT	Dimethoxytrityl
DNA	Deoxyribonucleic acid
DPPA	Diphenylphosphoryl azide
ds	Double-stranded
DSB	Double strand breaks
EDA	Ethylenediamine
EDTA	Ethylenediaminetetraacetic acid
EED	endosomal escape domain
E-site	Exit site
FAM	Fluorescein
FBS	Foetal bovine serum
Fmoc	Fluorenylmethyloxycarbonyl
G	Guanine
GalNAc (GN)	N-acetylgalactosamine
GFP	Green fluorescent protein
gRNA	guide-RNA

H-bond	Hydrogen bond
HBTU	Hexafluorophosphate Benzotriazole Tetramethyl Uronium
HDR	Homology-directed repair
HPLC	High-performance liquid chromatography
IEX	Anion-exchange
IVT	In vitro transcription
K _D	Dissociation constant
LNA	Locked nucleic acid
LNP	Lipid-based nanoparticles
m ⁷ G	7-methylguanosine
MeCN	Acetonitrile
MeOH	Methanol
MMT	Monomethoxytrityl
mRNA	Messenger RNA
MS	Mass spectrometry
MTPPI	Methyltriphenoxyphosphonium iodide
NEB	New England Biolabs
NHEJ	Non-homologous end joining
NHS ester	N-hydroxysuccinyl ester
NMI	N-methylimidazole
NMR	Nuclear magnetic resonance
nt	Nucleotides
NTP	Nucleoside triphosphates
Osu	O-succinimide
PABPC	cytoplasmic poly(A)-binding protein
PAGE	Polyacrylamide gel electrophoresis
PAM	Protospacer-adjacent motif
PAP	Poly(A) polymerase
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PFP	Pentafluorophenyl
pre-mRNA	Pre-messenger RNA
PS	Phosphorothioate
P-site	Peptidyl site
r.t.	room temperature
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNAi	RNA interference
RNAP	RNA polymerase
RNP	Ribonucleoprotein
ROS	Reactive oxygen species

RP	Reverse-phase
RRM	RNA recognition motif
rSAP	shrimp alkaline phosphatase
RT	Reverse transcriptase
R _T	Retention time
siRNA	small-interfering RNA
SPAAC	Strain-promoted azide-alkyne cycloaddition
SPS	Solid-phase synthesis
ss	Single-stranded
SSB	Single strand breaks
T	Thymine
T7E1	T7 endonuclease I
TBDMS	tert-butyldimethylsilyl
t-BuOH	<i>tert</i> -butyl alcohol
TC	thiomorpholine-4-carbothioate
TCA	Trichloroacetic acid
TEAA	Triethylammonium acetate
TF	Transcription factors
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
THPTA	Trishydroxypropyltriazolylmethylamine
TLC	Thin layer chromatography
T _m	Melting temperature
Tr	Triazole
tracrRNA	trans-activating crRNA
tRNA	Transfer RNA
U	Uracil
UPLC	Ultra-performance liquid chromatography
UTR	untranslated region

Abstract

Chemical modification of oligonucleotides has gained growing attention in research and therapeutic areas as these modalities, such as ribose modifications or fluorescent tags, have shown great promise to improve therapeutic properties and enable real-time analysis of biodistribution in antisense and RNA-based therapeutics. Recent development of modified nucleoside phosphoramidites and triphosphates enables the incorporation of modified functionalities and small-molecule conjugates onto synthetic oligos by solid-phase or enzymatic syntheses. Such approaches, however, may prove inadequate in certain applications where modifications on long oligos or specific inter-nucleosidic linkages are desired. In this thesis, we explored various solid-phase and solution-phase chemical approaches to access oligonucleotide bioconjugates that are challenging to access by conventional methods. These modalities include oligonucleotide dendrons and 3'-conjugates of long RNA. In addition, we examined their physical properties and biological functions, and discussed their potential applications.

Chapter 1 – Introduction

1.1 Nucleic Acids Structures and Nomenclature

1.1.1 Primary Structures of Nucleic Acids

Nucleic acids are biopolymers which carry genetic information in all known forms of life. Most organisms employ deoxyribonucleic acid (DNA) to store and propagate genetic instructions essential to assemble and maintain a living system. Each DNA molecule is composed of a long polynucleotide chain formed by four types of nucleotides as monomeric units. In its polymer form, the chemical structure of a nucleotide is categorised into three units: (i) a five-carbon sugar ring (ribose); (ii) a nitrogen-containing heterocyclic base attached to the anomeric carbon and (iii) phosphate groups that connect two ribose rings. The basic unit that encodes genetic information is simply known as a nucleoside, which contains a ribose and a nucleobase but without a phosphate. As a structural analogue to DNA, ribonucleic acid (RNA) possesses a hydroxy group at the 2'-position of the ribose instead of a hydrogen (Figure 1.1). It principally serves as a conduit between DNA genome and protein-synthesising machinery in cellular organisms, but is also capable of encoding genomes in viruses.

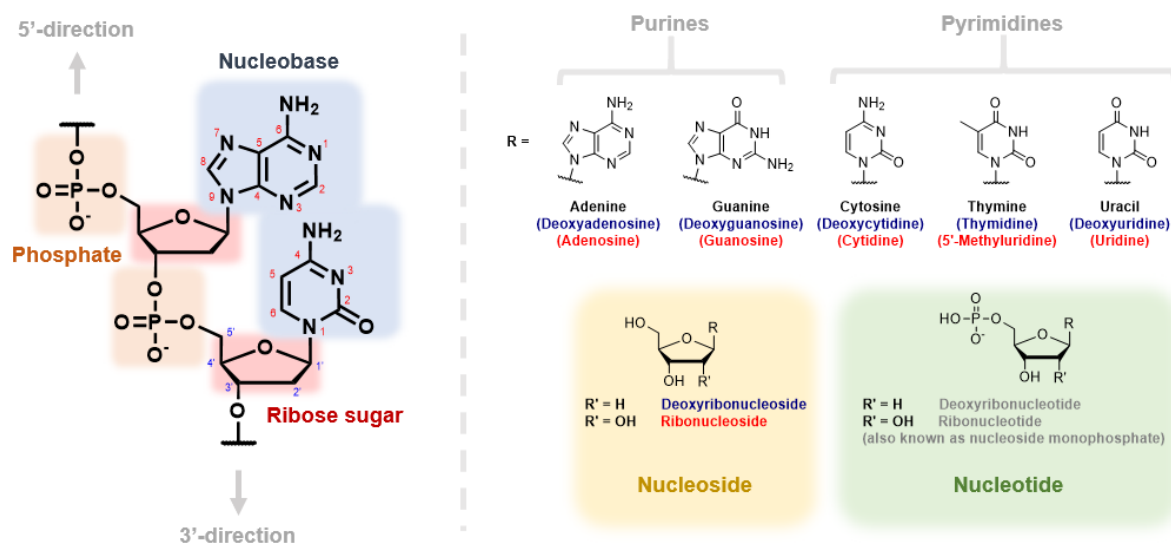


Figure 1.1 | Chemical structures and nomenclature of nucleic acid constituents

Left: chemical structure of a segment of DNA; highlighted are the structural motifs: nucleobase, phosphate and ribose sugar. Right: chemical structures of canonical nucleobases, generic forms of a nucleoside and a nucleotide.

The encoding of genetic information by nucleic acids is governed by the sequence of the polynucleotide chain, which stretches from, by convention, the 5'-end (where the terminal 5'-hydroxy group is situated) to the 3'-end. The sequence is based on the combination of four distinct nucleobases: adenine, guanine, cytosine and thymine (replaced by uracil in RNA) and offer a probabilistically unique value to identify a certain gene. These heterocyclic bases are connected to the sugar at the anomeric carbon (1'- position) via a flexible glycosidic linkage. Consequently, all nucleosides can adopt two different conformations, namely *syn*- (where the nucleobase projects towards the ribose) and *anti*- (where the nucleobase projects against the ribose) at two energy extrema. In addition, the anomeric carbon gives rise to two diastereoisomers – α -anomer (where the nucleobase points at the opposing direction from the 5'-carbon) and β -anomer (where the nucleobase points at the same direction as the 5'-carbon) (Figure 1.2). In naturally occurring nucleic acids, the *anti*-conformer and β -anomer which favour intermolecular hydrogen-bonding between nucleobases predominate.¹

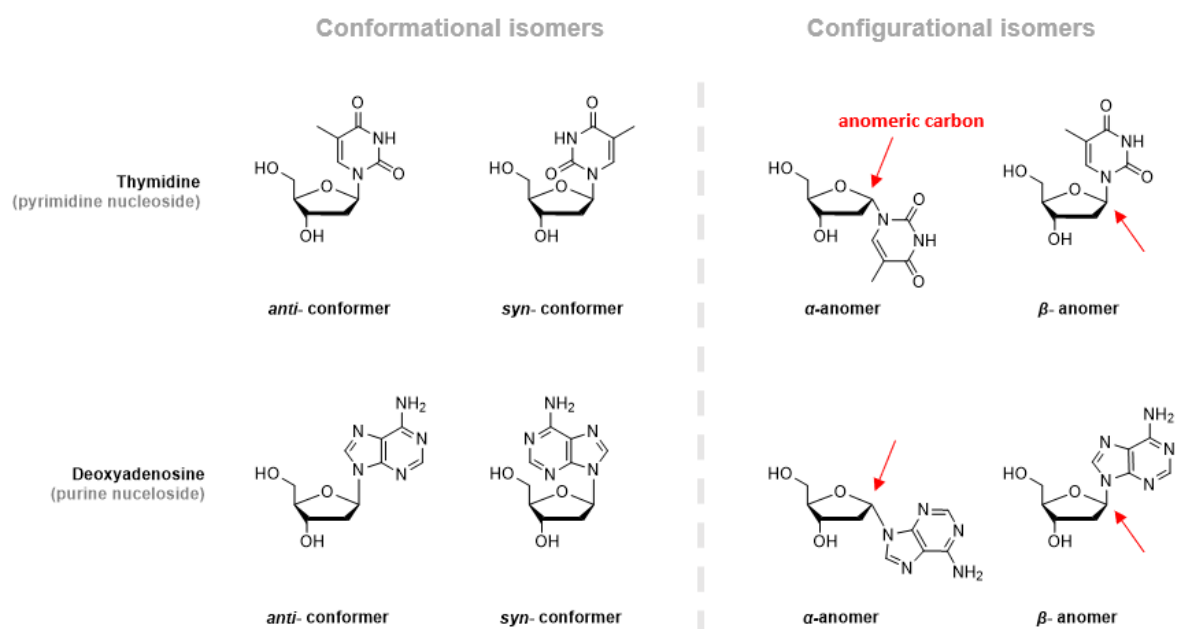


Figure 1.2 | Structures of the conformational and configurational isomers of nucleosides

In addition to the relative position of the nucleobase, the (2'-deoxy)ribose ring can adopt various conformations which engender structural nuances on nucleosides and different duplex properties in double-stranded nucleic acids.^{2,3} Generally, a planar conformation where all carbons are on the same plane results in high torsional and angle strains and is therefore energetically disfavoured. To relieve this steric crowding, the sugar ring deviates from a planar conformation and projects one or two carbons off the plane, resulting in either an envelop (E) or a twisted (T) conformation, respectively. This process is described as 'sugar puckering'. The spatial property of the carbon atom(s) deviated from the plane towards the relative orientation of the 5'-carbon is characterised as 'endo-', and of those opposite to it as 'exo-' (Figure 1.3). Although the sugar conformers are, in NMR time-scale, interchangeable due to a low energy barrier of transition in monomeric nucleosides, specific conformations are favoured in polymeric nucleic acids.³ For instance, cellular DNA adopts a C2'-endo conformation that is a geometry attribute of the predominant B-form double helix.⁴ Meanwhile, RNA generally prefers a C3'-endo conformation as it places the additional 2'-hydroxy group further away from the 3'-hydroxy group and nucleobase, and thus reduces overall energy. Of note, all conformations are in equilibria under different environments such as solvent or salt content.

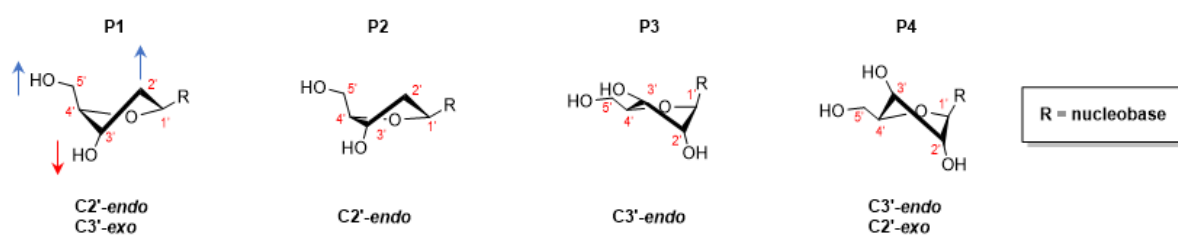


Figure 1.3 | Non-exhaustive representations of sugar puckers in nucleosides

Representations of 2'-deoxyribonucleoside in C2'-endo, C3'-exo (P1) and C2'-endo (P2) conformations, and of ribonucleoside in C3'-endo (P3) and C3'-endo, C2'-exo (P4) conformations. In the example of P1, blue arrows indicate the alignment in orientation of the 3'-carbon relative to the 5'-carbon; red arrows indicate an opposite one.

1.1.2 Secondary Structures of Nucleic Acids

All known organisms and certain classes of virus rely on DNA as a storage of genetic information. To date, we understand that naturally occurring DNA exists not as a single molecule but a duplex of two molecules that adopt a helical structure. The structural elucidation of DNA dates back to the early 1950s, when Erwin Chargaff first reported that the molar ratios between adenine (A) to thymine (T) and guanine (G) to cytosine (C) are close to unity across different sea urchin species.⁵ This regularity prompted a hypothesis that AT and GC bases are in some ways paired. In 1953, Rosalind Franklin discovered through X-ray fibre diffraction study that crystalline DNA is highly hygroscopic and it switches between two types of ordered structures depending on moisture levels, in which the interactions between the phosphates and the environment vary. Strikingly, these discoveries support that the phosphates must be interacting with the solvent so that it is positioned at the exterior, and neighbouring DNA chains are being held together by H-bonding between nucleobases.^{6,7} Concurrently, Watson and Crick proposed that DNA adopt a double helical structure in which the phosphate groups face outwards, the pentoses tilt to accommodate a roughly perpendicular angle of the nucleobases which face inwards and pair with another base⁸ (the discovery and verification of which earned them a Nobel Prize Physiology or Medicine in 1962). These seminal papers have provided the theoretical basis which has been verified and developed into our understanding on DNA structure and the molecular mechanism of heredity hitherto.

Currently, we understand the principles behind the formation of a stable DNA duplex (Figure 1.4), which is energetically driven by three physical properties, in aqueous environment. First, DNA carries both negative charges on its phosphate backbone and hydrophobic moieties on the nucleobases, thus it is an amphiphilic molecule which possesses affinity toward two different environments, such as those with distinct polarities. The presence of hydrophobic surface on nucleobases disrupts the highly organised hydrogen-bonding (H-bond) dynamics of

water and lowers the entropy of the system so that the hydrophobic part of the molecule tends to aggregate, or self-organise, to minimise contact with water.⁹ Consequently, a double helical structure with nucleobases buried inside the negatively charged phosphate backbone is energetically favoured as the non-polar part has minimised contact with the aqueous environment. This effect is an entropically driven event which is recognised as ‘*hydrophobic effect*’.

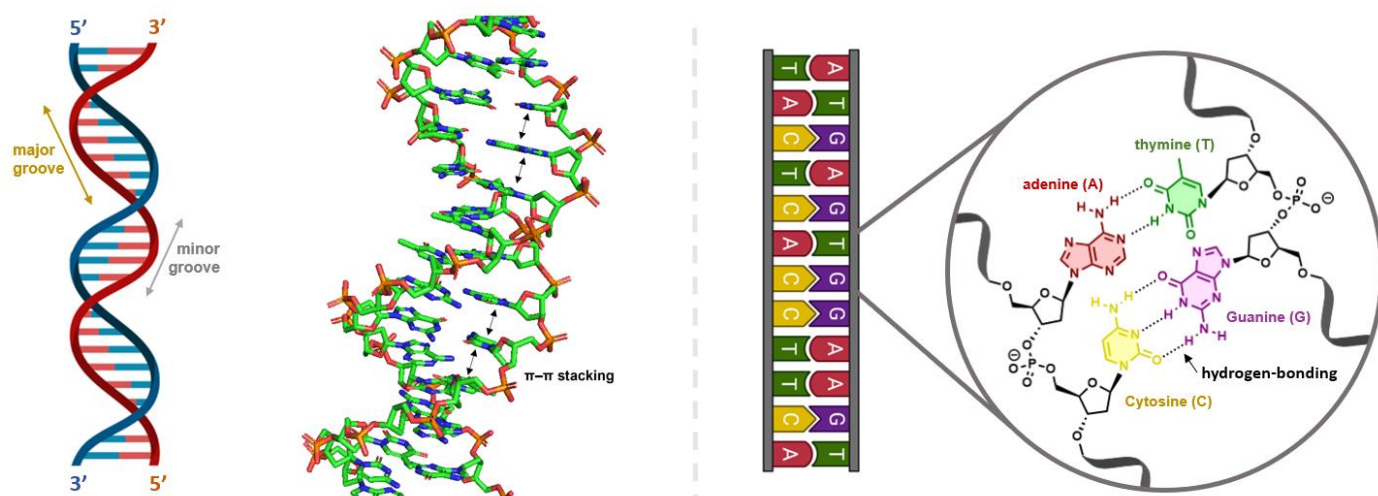


Figure 1.4 | Features of nucleic acid secondary structures

Left: graphical representation and computer simulation (PyMol) of a DNA duplex (double helix) formed by two DNA strands in anti-parallel orientation. Highlighted by the yellow arrow is the major groove, which is a wide, more accessible site of the helix; the grey is the minor groove, which is more narrow and less accessible to exogenous environment. π - π stacking is indicated between nucleobases in consecutive nucleotides. Right: scheme of hydrogen-bonding interactions between nucleobase pairs – adenine (A): thymine (T); and cytosine (C): guanine (G) on two antiparallel DNA strands in a duplex.

Second, DNA strands tend to adopt a vertical structure because of the energy contribution from ‘*nucleobase stacking*’, or more commonly known ‘ π - π stacking’’. Essentially, nucleobases on a DNA strand maintain a staggered stacking, in which two aromatic surfaces have energetically favourable electronic interactions between the individual dipoles and π systems¹⁰ (e.g., C-H dipole and π electrons) and orbital interactions between the aromatic rings¹¹, and support a vertical structure of the entire strand. On top of the geometry of a DNA

single strand, base stacking interactions contribute also to the stability of a duplex to an extent comparable to the H-bonding between base pairs (*vide infra*). For instance, Karabiyik and co-workers (2014) reported that the stacked A//T and G//C dimers contribute 11.6 and 16.9 kcal mol⁻¹ to the stabilisation energy of a DNA duplex, meanwhile the H-bonding between A:T and G:C base pairs contributed to 15.4 and 28.8 kcal mol⁻¹.¹⁰ The result indicates that such stacking interaction along the DNA strand is one of the major energy contributors for the formation of a double helical structure.

Finally, nucleobases can form extensive H-bonding with one another that brings two DNA molecules together to form a duplex structure, known as hybridisation. In agreement to the findings from Chargaff, Watson and Crick in 1953 proposed that each adenine (A) can form H-bonds with thymine (T) and guanine (G) with cytosine (C), given the roughly perpendicular orientation of nucleobases along the phosphoribosyl backbone.⁸ In fact, the A:T pair is able to form two hydrogen bonds whereas the G:C pair can form three as illustrated in Figure 1.4. These pairing patterns are coined as ‘*Watson-Crick base pairing*’ and considered as canonical. Despite such inter-strand H-bonding formed by nucleobases being relatively weak due to competing H-bonding interactions with water, structural pre-organisation of the phosphate-containing backbone and an additive effect of nucleobase stacking along a long polynucleotide strand favour the formation of a duplex between two antiparallel strands, not least if they contain high degree of complementarity with respect to the sequence of base pairs that maximises H-bond interactions. In addition to its structural implications, this canonical base pairing pattern leads to an important consequence: the stability of nucleic acid duplex hinges on the sequence complementarity between the two strands. It implies that the hybridisation of two DNA strands does not occur randomly, but it depends on their sequence specificities and the length of the polynucleotide.

Unlike DNA, most RNA molecules found in nature do not form duplexes but rather fold itself into secondary structures such as stems, loops and bulges.¹² Nonetheless, they are capable of hybridising with a complementary poly- or oligonucleotide to form a RNA:RNA or DNA:RNA duplex, which can be observed in certain classes of viral RNA genomes and antisense oligonucleotide therapeutics, respectively. These duplexes, however, possess contrasting structures and biophysical properties to a DNA:DNA duplex. As mentioned above, ribonucleotides adopt a C3'-endo sugar pucker and assume a more rigid structure than deoxynucleotides due to the presence of a 2'-hydroxy group. As a result, the hybridisation of two RNA strands produces a more rigid tertiary structure, namely A-RNA, with a deep but narrow major groove, shallow but wide minor groove and more compact spacing of base pairs.¹³ This structure is different to naturally occurring DNA:DNA duplexes that adopt a B-DNA conformation, in which the major and minor grooves are of similar depth and base pairs are less compact.¹ In general, RNA:RNA duplexes in A-form exhibit higher duplex stability than DNA:DNA ones in B-form; on the other hand, DNA:RNA hybrids can be more or less stable than DNA:DNA but generally less stable than RNA:RNA.¹⁴ Experimentally, the duplex stability of any oligonucleotide hybrid is indicated by a value called melting temperature (T_m), which is the temperature at which half of the duplexes are denatured under a controlled, gradual increase in temperature. This technique will be employed in relevant chapters within this thesis.

1.2 Biological Functions of Nucleic Acids

1.2.1 Central Dogma of Biology

In the late 1950s, the growing understanding in nucleic acids (as genetic materials) and proteins (as molecular machinery) has sparked the curiosity of molecular biologists on an important question: how proteins are synthesised. Indeed, the most intriguing puzzle at the time concerned the simplicity and regularity of protein synthesis from hereditary information, as the process seemed uniform regardless of peptide sequence or species. To answer such question, Crick is among the first to propose the idea of a '*central dogma*' of biology to indicate the overall framework of how genetic information is passed from nucleic acids to proteins.¹⁵

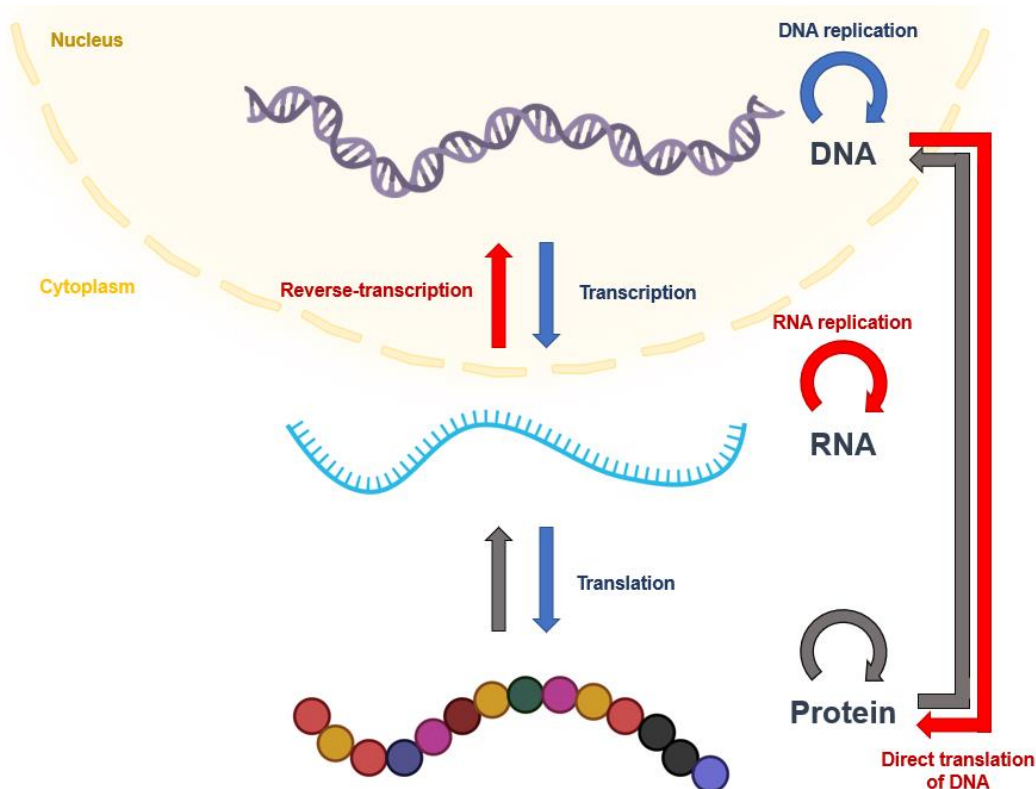


Figure 1.5 | Central dogma of biology indicating the flow of information between nucleic acids and proteins

This scheme highlights the transfer of genetic information among DNA, RNA and proteins by arrows. Blue arrows indicate a general transfer, which can occur in all cell types; red arrows indicate a special transfer, which does not occur in most cell types but only under specific circumstances; grey arrows indicate an unknown transfer, which has not been observed and is postulated to be impossible. Generally, cells can freely replicate its genetic information (DNA replication), copy the information into RNA (transcription) and synthesise proteins by reading through the genetic instructions on RNA (translation).

The formulation of an early central dogma by Crick and co-workers encompasses a few general principles which were supported by experimental observations at the time: (i) the specificity of a nucleic acid molecule is manifested solely by the sequence of its nucleobases, and this sequence is linearly translated to the amino acid sequence of a protein; (ii) the flow of information is only possible between nucleic acids and nucleic acids, or from nucleic acids to proteins, but not from proteins to nucleic acids, or between proteins and proteins; (iii) the amino acid is carried to the RNA template during protein synthesis via an adaptor molecule which is likely an RNA molecule that could fit into the template; and (iv) there exists a set of triplet codes (of nucleobases) that correspond to specific amino acids.¹⁵ The following decade has seen some minor modifications on the original concepts, but the core idea remains unchanged as Crick established it: once the residue-by-residue, sequential information is passed from nucleic acids to proteins, it cannot be passed back to nucleic acids again (Figure 1.5).¹⁶ This theory has set a paradigm for chemists and biologists to solve more complex problems such as understanding disease progression and relevant treatments.

1.2.2 Flow of Genetic information and Its Implications

To date, it is widely understood that genetic instructions governing all biological processes are passed from DNA to RNA, and finally to proteins which carry out a vast array of functions in cellular organisms (Figure 1.6). The process of copying such instructions from DNA to RNA is called '*transcription*' and the process of decoding the RNA transcript to generate a specific sequence of peptide is called '*translation*', also known as 'gene expression' (Figure 1.6). In broad sense, a gene is a DNA sequence in the genome that can be transcribed into a functional RNA, which either encodes for proteins or serves regulatory functions.

During transcription in eukaryotic systems, an RNA polymerase (RNAP), together with a set of transcription factors (TF), binds to the promoter region of a gene and melts the double-

stranded DNA (dsDNA) structure to form an accessible region called ‘transcription bubble’. Next, the RNAP brings a nucleotide to its active site and uses one strand of the genomic DNA as a template to guide nucleotides with correct base-pairing to polymerise in exact sequence. After a regulatory process called ‘abortive initiation’ where short RNA transcripts (< 10 nucleotides, nt) are synthesised and released,^{17,18} the nascent RNA transcript longer than 10 nt enters the exit channel of the polymerase and the RNAP moves along the DNA strand from its promoter to continue polymerisation of the other end of the transcript. Once the RNAP reaches the terminator region, it releases the full-length pre-messenger RNA (pre-mRNA) and dissociates from the dsDNA. This pre-mRNA will then undergo a series of post-transcriptional modifications, including capping of the 5’-end, RNA splicing and polyadenylation etc., and the resulting mRNA will be exported to the cytoplasm for templating protein synthesis.

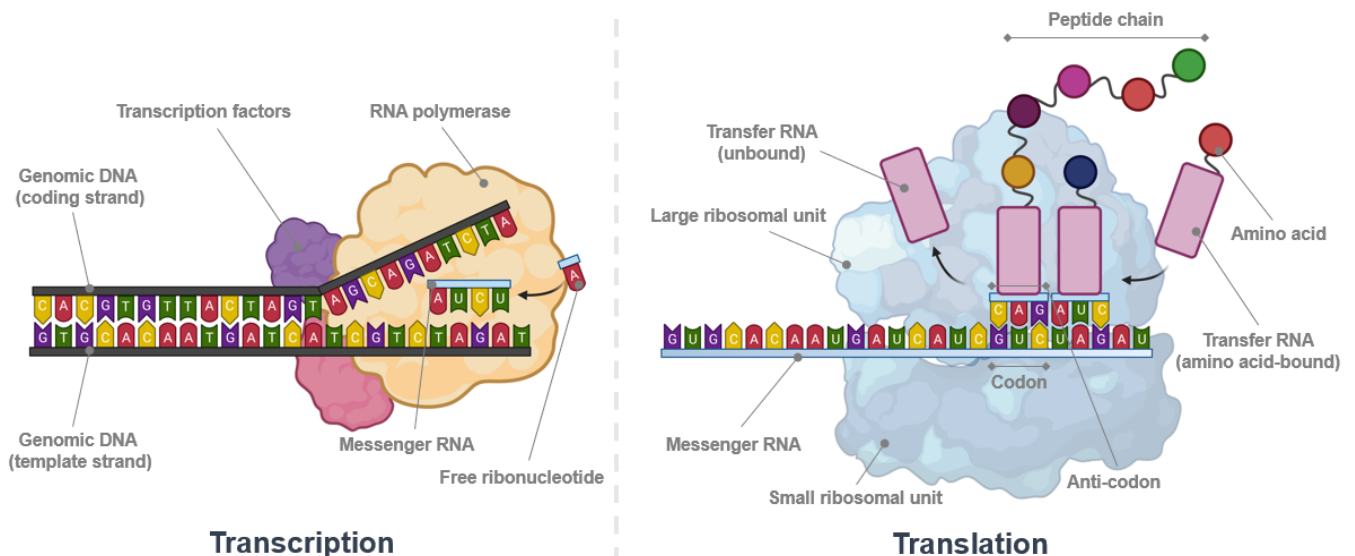


Figure 1.6 | Schematic overview of two information transfer processes - transcription and translation – in cellular organisms

Left: graphical representation of a *transcription* event. RNA polymerase bound to transcription factors unwinds a segment of genomic DNA and catalyses the incorporation of RNA nucleotides into a newly synthesised mRNA by complementary base pairing. Right: graphical representation of a *translation* event. The mRNA reaches the ribosome which decodes the transcript by inducing the binding of tRNA carrying a complementary trinucleotide (anti-codon) and corresponding amino acid according to the triplet code on the mRNA (codon) along the strand. The ribosome catalyses amide bond formation between the growing peptide chain and the incoming amino acid. Figure created under BioRender’s license.

Translation begins when the small ribosomal subunit, with a set of initiation factors, recognises and binds to the approaching mRNA then moves along the strand until it finds the start codon, which is a trinucleotide sequence of AUG. At this point, it recruits the large ribosomal unit and first transfer RNA (tRNA), which carries an amino acid via an aminoacyl bond at its 3'-end specific to a trinucleotide anti-codon complementary to the codon on the mRNA. As the tRNA sits on the start codon of the mRNA at the peptidyl site (P-site), the ribosome accommodates another tRNA with anticodon matching the next codon downstream at its aminoacyl site (A-site) and catalyses peptide bond formation between the amino acid (growing peptide) at the P site and the incoming amino acid at the A-site. Finally, the ribosome moves to the next codon and the first tRNA breaks the aminoacyl bond and enters the exit site (E-site) while the tRNA at the A-site moves to the P-site to free up space for the next tRNA for peptide elongation. This process is referred to as 'translocation'. When the stop codon (UAA, UAG, or UGA) is reached at the termination step of translation, it is recognised by release factor proteins which prompts the dissociation of the ribosome and mRNA and release of the peptide from the ribosomal peptidyl transferase centre.¹⁹ Notably, anti-codons made up of three nucleotides (each possessing four possible identities) provide a set of $4^3 = 64$ permutations, which encode for a smaller set of 20 canonical amino acids. It was discovered that most amino acids are encoded by more than one codon (except methionine and tryptophan), giving rise to the '*codon degeneracy*' property which allows codons with similar compositions to encode the same amino acid and enhances the tolerance of translation machinery toward point mutations.²⁰

During translation, the growing polypeptide chain starts folding into three dimensional (3D) structures and form proteins whose structural and catalytic properties are governed by their structures. This process is primarily driven by the non-covalent interactions among amino acid side chains (i.e., hydrophobic interactions, Van der Waals forces, H-bonding and salt bridges etc.) and, in some cases, covalent linkages such as disulfide between cysteine

residues.^{21–23} Therefore, the sequence of amino acids linearly translated from DNA sequence via mRNA and tRNA intermediaries on a residue-by-residue basis ultimately impacts the properties and functions of every protein.

The fundamental process of carrying out genetic instructions is essential to life as most parts of its functioning involve proteins that are synthesised in faithful sequences and folded in correct structures. However, nucleic acids are somehow prone to chemical alterations due to endogenous and exogenous factors. For example, reactive oxygen species (ROS) generated from metabolism can oxidise nucleobases and break chemical bonds²⁴; UV radiation can cause intra- or inter-strand crosslinking and DNA scission²⁵; alkylating agents can chemically transform nucleobase and induce scission events.²⁶ As a result, cellular organisms have developed a family of repair proteins to remove damaged nucleotides and repair scissions to prevent harmful mutations from being accumulated or passed on (Figure 1.7).

In reality, not all of these repair mechanisms are equally productive. Mutations in base pairs could sometimes evade cellular surveillance and *DNA damage response* (DDR) which is a cascade of signalling to recruit repair proteins. They could also introduce undesired changes to the gene and lead to genomic instability and progression of diseases such as cancers.²⁷ Among various known repair mechanisms, *direct reversal repair* is the simplest and most error-free pathway to treat base lesions. It involves enzymatic reactions that ‘reverse’ DNA damages on individual bases such as methylation (by alkyltransferase) or photo-crosslinking (by photolyase). There are more sophisticated lesions such as base mismatches which involve addition, deletion or alteration of nucleobases due to replication errors or chemical reactions, and single strand breaks (SSB) which is the bond breakage of one DNA strand due to ionising radiation or oxidative stress.²⁸ Such lesions are generally structurally anomalous and can easily trigger DDR; they often require excision of a few nucleotides around the damaged site by endonucleases and are repaired by DNA polymerases and ligases using the intact strand as a

template. These mechanisms, namely *nucleotide excision repair* and *mismatch repair*, are orchestrated by multiple sensing proteins and enzymes. Despite the presence of a template, such delicate processes can sometimes be prone to failures such as abortive enzymatic activities and lead to undesirable base mutations.²⁸ Double strand breaks (DSB), in which both DNA strands are cleaved in close proximity, are the most deleterious form of DNA damage and are linked to genomic rearrangements and carcinogenesis. During cell cycle S/G2 phase when sister chromatid exists, DSB can be repaired via *homology-directed repair* (HDR) pathway in which the two ends of the dsDNA are processed to form two hanging single strands (DNA end resection) and they will be elongated by polymerases using the homologous DNA as a template. This results in a conservative and error-free product of the repair. In other cell cycle phases, however, cells are programmed to repair DSB via *non-homologous end joining* (NHEJ) to prevent prolonged presence of the lesion. From this mechanism, DNA binding proteins will first bind to the break ends and recruit other enzymes to remove damaged or mismatched nucleotides (by nucleases), synthesise a new strand (by polymerases) and connect it to the other end (by ligases). This process can be accurate if the two ends harbour compatible overhangs as they can act as template for the repair process; otherwise, NHEJ tends to introduce random insertions and deletions which can threaten genomic integrity. The HDR and NHEJ repair mechanisms will be further discussed in Chapter 3 about CRISPR-mediated genome editing.

When any nucleic acid sequence within a genome or extrachromosomal DNA is altered without faithful repair, a mutation is formed. In some instances, mutations are a normal biological process and they do not change the phenotype, which is the observable characteristics, of an organism. Some mutations, however, lie within the protein-encoding gene and could result in changes of the amino acid sequence, namely *amino acid dysmorphism*, and functions of the protein product. This could be harmful to an organism and its offspring and lead to disease progression.

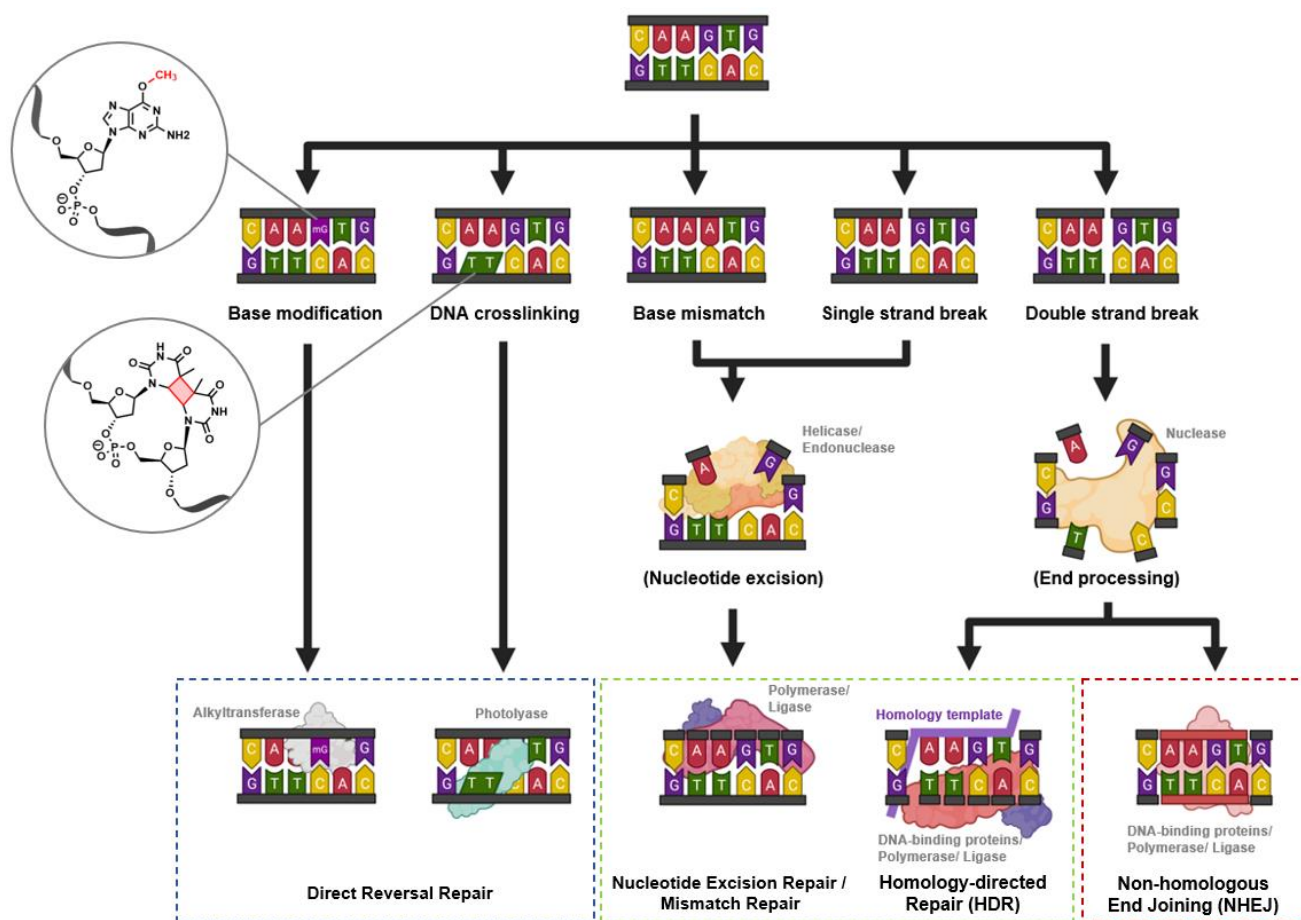


Figure 1.7 | Overview of DNA damages and possible repair mechanisms²⁷

Schematic representations of various types of DNA damages and possible repair mechanisms in cellular organisms. Figure created under BioRender's license.

1.2.3 Applications of Nucleic Acids in Medicine and Biotechnology

Naturally, nucleic acids are known for their roles as the storage and carrier of genetic information. With an increasing understanding of the principles of base-pairing and interactions with proteins, the utility of oligonucleotides in medicine and biotechnology has been widely expanded. Most oligonucleotide-based drug platforms interact with their cognate targets via Watson-Crick base-pairing. This property offers a few key advantages over conventional pharmaceutical modalities: (i) sequence design is modular and straight-forward; (ii) the range of potential drug targets is wider than protein-targeting small-molecule drugs; and (iii) the process of drug screening, optimisation and engineering is expedient due to

accessibility to automation.^{29,30} In this section, the growing popularity and utility of oligonucleotide-based drug platforms and gene-editing methodologies will be discussed.

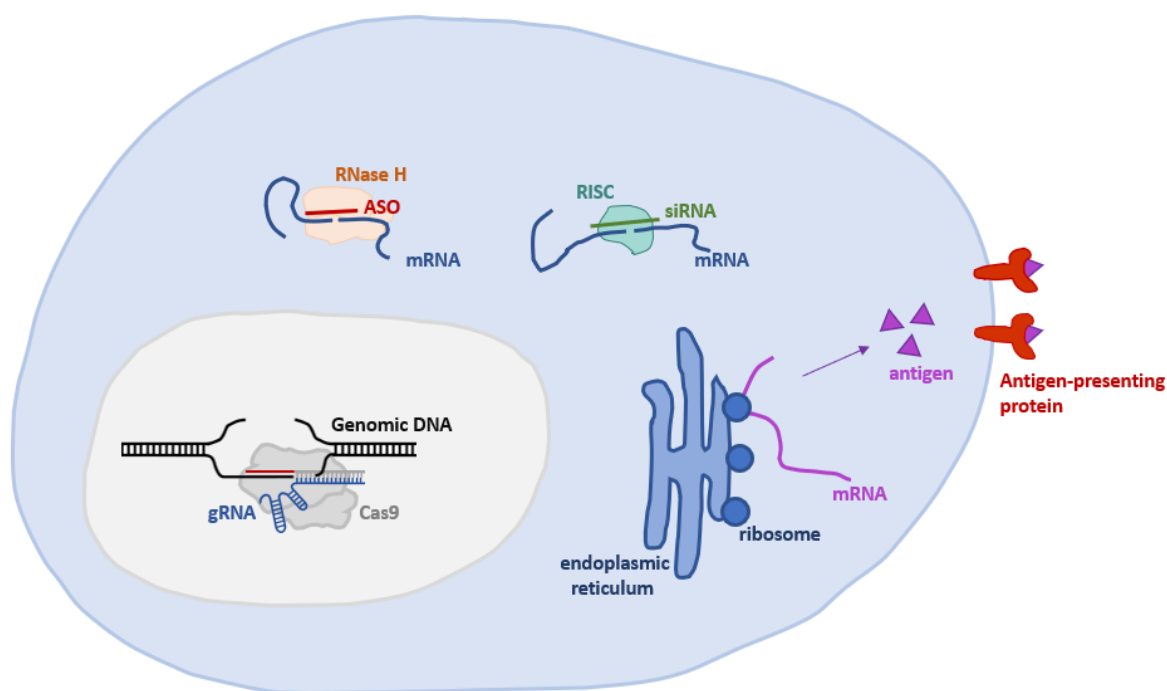


Figure 1.8 | Applications of artificially generated oligonucleotides in therapeutics and biotechnology

Selected examples of oligonucleotide-based therapeutic platforms and gene-editing technologies in cellular organisms are illustrated. Antisense oligonucleotides (ASO) and small-interfering RNA (siRNA) can induce cleavage at target mRNA by interacting with RNase H and RNA-induced silencing complex (RISC), respectively, resulting in the arrest of protein synthesis. CRISPR-mediated gene-editing system comprises Cas9 endonuclease and guide RNA (gRNA) which directs Cas9 to introduce breaks at the double-stranded genomic DNA target. This cleaved DNA can be repaired according to a donor template which contains the desired gene edits. The mRNA component in vaccine formulae is translated into peptide fragments of a pathogen which will be presented on cell surface and trigger recognition by the host's adaptive immunity.

Antisense Oligonucleotides (ASOs). Translation of mRNA is an essential process in gene expression but it is susceptible to interference by exogenous, short fragments of oligonucleotides bearing sequence complementarity. This mechanism is harnessed to silence or modulate the expression of a faulty gene that bears undesired mutations. Antisense oligonucleotides are a class of synthetic oligonucleotides, usually at a length of 18-30

nucleotide, which effect gene silencing and modulation when they bind to their cognate mRNA targets via RNase H activation or steric blocking. RNase H is an endonuclease residing in cell nuclei (and mitochondria) and is capable of hydrolysing RNA residues within a DNA/RNA duplex,³¹ leading to direct destruction of mRNA and thereby silencing of a gene.³² ASOs typically follow a ‘gapmer’ design in which the central part of a sequence is made with DNA to engage RNase H in cleavage activity whereas the flanking parts are composed by RNA-based, modified nucleotides to improve target binding as well as stability.³³ Alternatively, ASOs can adopt a ‘mixmer’ sequence in which a mixed construction of nucleotide chemistries is utilised to sterically block the target transcript from interacting with proteins such as spliceosomes and alternate the splicing decision to correct the translational reading frame.^{34,35}

RNA Interference (RNAi). RNAi is another gene silencing mechanism discovered in 1998 by Andrew Fire and Craig Mello, who were jointly awarded The Nobel Prize in Physiology or Medicine for their contributions in 2006. The RNAi pathway begins when double-stranded RNA (dsRNA) molecules are recognised and cleaved by Dicer, an endoribonuclease, which produces small dsRNA fragments known as small interfering RNAs (siRNAs). These siRNAs are then processed into single strands and one of which, namely the guide strand, is incorporated into a multi-protein complex known as the RNA-induced silencing complex (RISC). The resultant ribonucleoprotein then binds to and degrades its mRNA target via a complete complementarity.²⁹ RNAi-mediated silencing has originally been observed in plants as a natural immune response against viral infection in which dsRNA generated from the replication of virus is loaded onto RISC to effect gene silencing to the viral transcript.^{36,37} In 2001, the first example of synthetic siRNA effecting gene silencing in mammalian cells was reported.³⁸ Two years later, the utility of siRNA in disease prevention was demonstrated by Lieberman and co-workers who reported the utility of synthetic siRNA in preventing liver injury in mice by silencing the *Fas* gene, which mediates hepatocyte apoptosis.³⁹ These seminal

studies have prompted a rapid increase in research volume and commercial investment on siRNA as a therapeutic tool, with major directions to identify gene targets associated with disease progression and improve the pharmacological properties of siRNA drugs with chemical modifications.

CRISPR-Mediated Gene Editing. CRISPR stands for Clustered Regularly Interspaced Short Palindromic Repeats which refer to the pattern of DNA fragments first identified in the genome of *Escherichia coli*.⁴⁰ Unlike other tandem repeats, these repeat clusters are spaced by non-repeating DNA sequences which were found to be originated from viruses or exogenous genetic elements^{41–43} and involved in bacterial immune system to degrade foreign genetic materials from invading species such as bacteriophages.^{44,45} The function of CRISPR involves two major components: (i) the CRISPR-associated (Cas) protein, which is a family of DNA endonucleases encoded by *Cas* genes adjacent to the repeat sequences and (ii) a guide RNA (gRNA), which is transcribed from the CRISPR array and directs Cas enzymes to carry out site-specific scission at the specific location in the genome that bears complementarity to itself. After the initial findings and structural elucidation, numerous reports have demonstrated that CRISPR components could be re-engineered and expressed in human cells to effect site-specific gene editing.^{46–48} Such efforts have realised the potential of CRISPR as a preferred gene targeting tool unrivalled by other conventional gene-editing platforms, largely thanks to its superior editing efficiency and modularity to target various genes by simply altering the sequence of the gRNA instead of a number of endonuclease modalities.⁴⁹ To date, scientists are able to transfect cells with synthetic gRNA and Cas proteins, or CRISPR-encoding plasmids to not only edit genomes precisely, but also regulate gene expression by fusing deactivated Cas protein (without cleavage ability) with transcription factors⁵⁰ and image genetic loci by appending fluorescent molecules to CRISPR modalities.⁵¹ The simplicity to employ laboratory-made gRNA in various gene-targeting applications has proven immensely

valuable across scientific fields. In Chapter 3, the detailed mechanisms and relevant chemistries of CRISPR-based systems will be further discussed.

mRNA Vaccines. The COVID-19 pandemic in 2020 has catalysed recent advancements in mRNA-based vaccine platforms.⁵² This new class of vaccines comprises synthetic mRNA which is translated into pathogenic antigens, which are fragments of a pathogen that can generate an immune response, using the host's cellular machinery when delivered into cells. This platform offers major key advantages over conventional vaccine components such as attenuated virus as it does not pose risks of genomic integration, or antigenic protein that requires slow and costly biosynthesis in living organisms during its manufacturing process. Currently, mRNA vaccine formulae are often packaged in lipid nanoparticles that are composed of cationic lipids (to complex with mRNA), PEGylated lipids and cholesterol (to modulate the size and structure of the particle) and carry the mRNA cargo into cells across the cell membrane. These vehicles offer essential protection to mRNA against nuclease during systemic circulation as the RNA cannot be easily modified with nuclease-resistance motifs such as phosphorothioate and 2'-MOE by *in vitro* transcription. Another notable aspect of mRNA in vaccine applications is that its nucleobases are often heavily modified. This is because (1) the human immune system recognises unmodified ssRNA as foreign genetic materials and triggers the production of interferons that block RNA translation; and (2) certain types of nucleobase modification can regulate translation and control the level of protein expression.⁵³ These nuances have led to a growing interest in the discovery of chemical or biological capabilities to modify mRNA, which is often considered a difficult target due to its length (often >800 nt) and fragility (intrinsic susceptibility to basic hydrolysis and metal-catalysed degradation), for further development. Some of the available strategies will be discussed in Chapter 4 in more detail.

1.3 Chemical Synthesis of Oligonucleotides

1.3.1 Solid-Phase Synthesis

With an increasing demand for oligonucleotides in research and therapeutic development, an efficient preparative workflow for scalable production has become essential. Compared to the synthesis of small molecules, there are a few key differences in the process of oligonucleotide synthesis: (i) most oligonucleotides are large molecules (> 3,000 Dalton) composed of a limited number of structurally identical modules (nucleoside subunits); (ii) the sequence of these modules determines the identity and properties of the product; and (iii) the large number of steps involved precludes the possibility to remove by-product at individual step. These factors limit the utility of conventional solution-phase chemistry and have led to the adoption of solid-phase chemistry, which was first introduced by Prof. Robert Bruce Merrifield to improve the synthetic efficiency of polypeptides,⁵⁴ for the synthesis of oligonucleotides with great length and molecular complexity. In this method, a starting nucleoside (and the growing strand thereafter) is first mobilised on a solid support, such as controlled pore glass (CPG) or polystyrene (PS) resins, typically through a 3'-O ester linkage. Then, a solution of an activated nucleoside is passed through to provide a medium for the immobilised species to react and form an adduct via a P-O bond. Finally, the unreacted reagents are washed away and the addition product remains on the solid support. After successive cycles of couplings, the bond between the oligo and its support is then cleaved to yield the full-length product. This strategy has proven highly suitable to oligonucleotide synthesis as it offers a number of merits: (i) the coupling efficiency of each cycle is greatly enhanced and deletion events can be minimised as a much higher loading of coupling reagents can be used; (ii) removal of excess reagents by solvent washing is simple and does not incur product loss; (iii) the processes involved are all mechanical and amenable to automation. For such reasons, solid-

phase synthesis (SPS) has been the most popular method for oligonucleotide production in both laboratory and manufacturing scales.⁵⁵

The overall yield of oligonucleotide synthesis is a multiplicative product of the coupling efficiency of each iterative addition step, meaning that the latter value has a significant cumulative effect on the yield for full-length oligonucleotides. Over the past decades, the chemistry involved in nucleoside coupling has been progressing to drive stepwise coupling towards quantitative yield. In the early 1950s, benzyl phosphorochloridate was first demonstrated to activate a 5'-protected nucleoside which couples with the 5'-hydroxy group of another 3'-protected DNA nucleoside to afford a P-O linkage in solution.^{56,57} In the following decades, scientists have reported various methods to activate the phosphate oxygen on a nucleotide to facilitate coupling with another monomer with varying degree of success.^{58–}
⁶⁰ It was not until 1980s was the coupling protocol of full-length oligo on solid-support reported. Two groups have independently reported the use of H-phosphonate monoester, activated by pivaloyl chloride or 2,4,6-triisopropylbenzenesulfonyl chloride (TPS-Cl) and coupled on solid support, to synthesise oligomer up to 107 nt in length.^{61,62} To date, the most widely adopted and efficient coupling reagent is nucleoside phosphoramidite⁶³ – a 5'-dimethoxytrityl (DMT) protected nucleoside connected to a 3'-phosphite diester amide which is a stable P(III) compound and can be activated *in situ* before coupling to oligos on a solid support.

Phosphoramidite coupling in solid phase starts typically with a 5'-DMT protected nucleoside bound to the solid support via 3'-O linkage (Figure 1.9 and scheme 1.1, **I**), from which the strand grows in the direction of 3' → 5'. Before entering coupling cycles, the acid-labile DMT group is removed by treating with trichloroacetic acid (TCA) to liberate the 5'-hydroxy group on the nucleoside (**II**). This step is called '*detritylation*'. Complete detritylation releases DMT cations (**A**) in equimolar ratio to the nucleoside and hence the absorbance of DMT cations (~495 nm) offers a quantitative measure of the level of DMT-containing

nucleoside incorporated in the previous cycle. Next, the phosphoramidite (**B**) is activated *in situ* by mixing with an activator, typically 1H-tetrazole, 5-Ethylthio-1H-Tetrazole (ETT) or 5-Benzylthio-1H-Tetrazole (BTT), which is mildly acidic and replaces the diisopropylamine to form a reactive tetrazolide species (**C**). The P(III) core is now susceptible to nucleophilic attack by the 5'-hydroxy to readily release the tetrazole leaving group. This step is referred as 'coupling'. Then, the oligo is treated with acetic anhydride/N-methyl imidazole ($\text{Ac}_2\text{O}/\text{NMI}$) mixture which generates a 1-acetyl-3-methylimidazolium active intermediate (**D**) and transfers

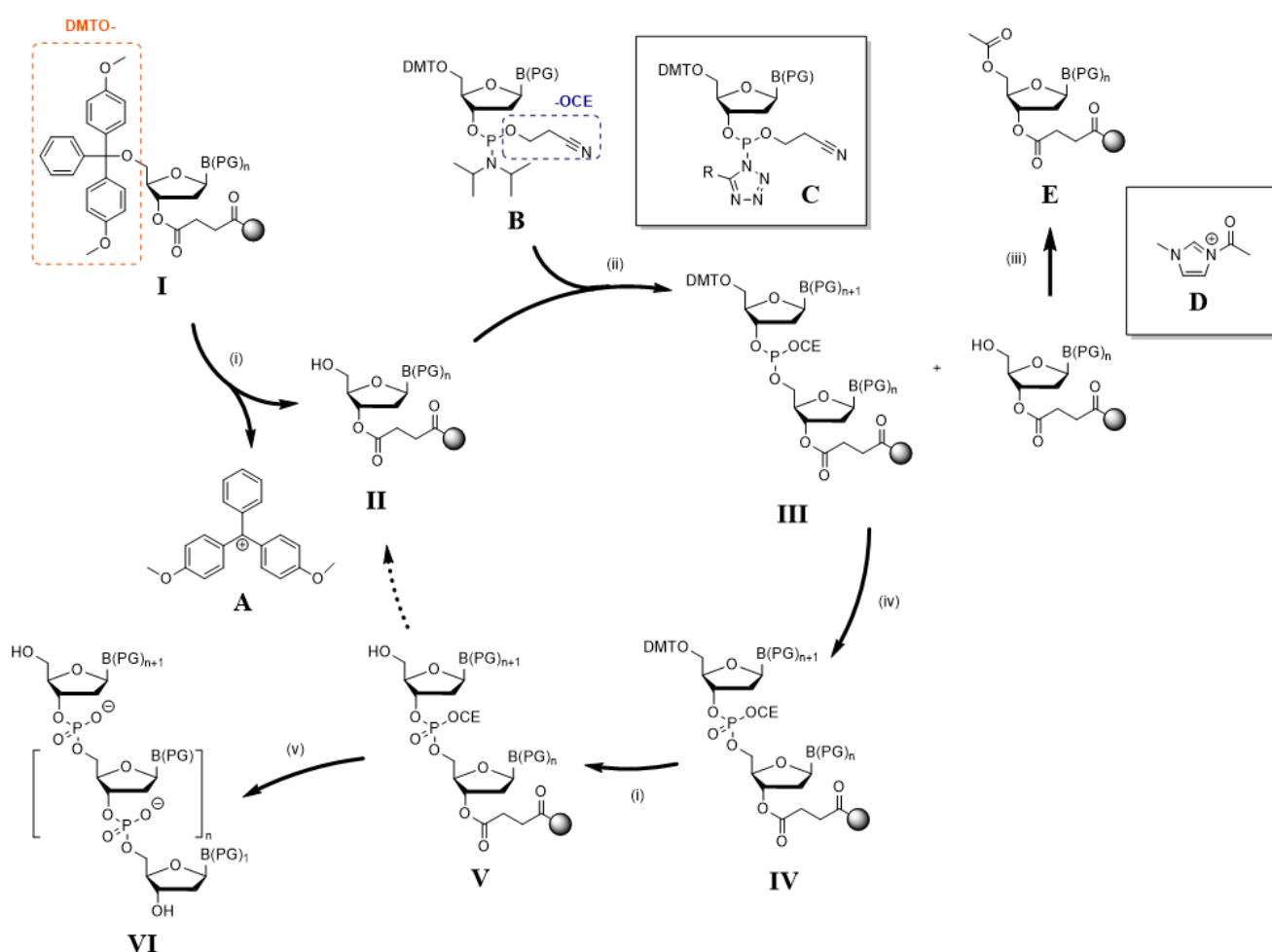
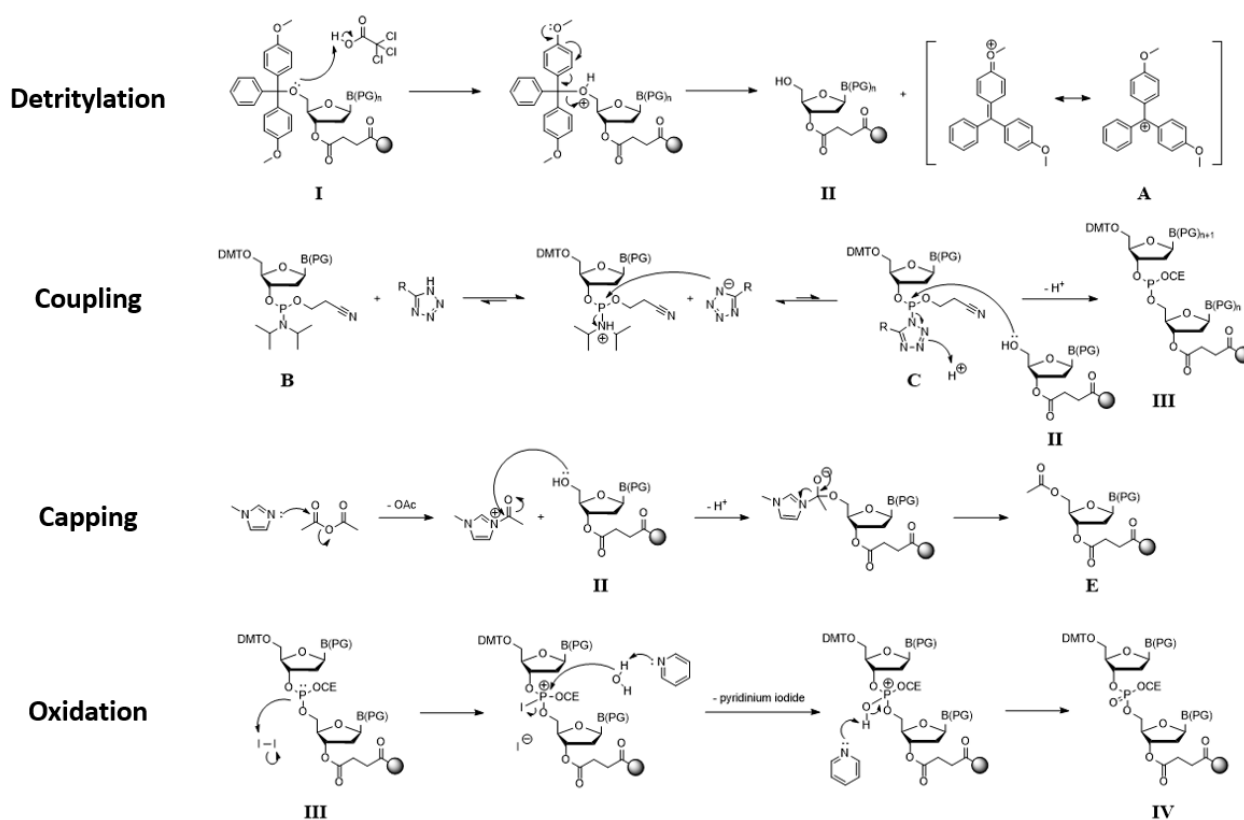


Figure 1.9 | Schematic overview of DNA synthesis by phosphoramidite method

This scheme describes the synthesis of a DNA oligonucleotide via standard phosphoramidite coupling cycles. Reagents and conditions: (i) 3% TCA in CH_2Cl_2 , 50 seconds; (ii) 0.1 M phosphoramidite (in MeCN), 0.5 M tetrazole (in MeCN), 30s [R = H]; (iii) Ac_2O /pyridine/THF (1/1/8 v/v/v), 17.6% NMI (in MeCN), 30 seconds; (iv) 0.015 M iodine in water/pyridine/THF (2/20/78 v/v/v), 45 seconds; (v) conc. NH_4OH , 55°C , 5 hours. Between each step of coupling cycles, washing with anhydrous MeCN (30s) and drying with argon (10s) is performed. Abbreviations: B = nucleobase, PG = protecting groups, DMT = dimethoxytrityl, CE = 2-cyanoethyl. Black ball represents solid support. Figure adapted from literature.⁶⁴

the acetyl group to the 5'-hydroxy group. This '*capping*' step can mask the reactivity of any unreacted 5'-hydroxy groups (**E**) so that they do not engage in further cycles of coupling. It ensures the events of internal deletions of oligonucleotides, which are difficult to remove by chromatography, are minimised. At this stage, a P(III)-O bond has formed between the nucleoside phosphoramidite and the resin-bound oligo (or nucleoside). Since this P(III) core is reactive towards acid in subsequent coupling steps, an '*oxidation*' step is performed with an iodine solution in a pyridine/water/tetrahydrofuran mixture to yield a stable P(V) species (**IV**). Then, the oligonucleotide intermediate is available for subsequent cycles of synthesis until the desired sequence is generated. Finally, the protecting groups on the oligo are removed and the free oligonucleotide product is released from the solid support by treating with aqueous ammonia (37%).



Scheme 1.1 | Mechanistic overview of each step in DNA synthesis cycle

This scheme describes the reaction mechanisms of each individual step in DNA synthesis.⁶⁴

1.3.2 Protecting Groups and Deprotection

During synthesis, some functional groups on an oligonucleotide may participate in side-reactions and generate undesired by-products. For example, the aromatic amines on cytosine, guanosine and adenosine are nucleophilic and able to attack activated P(III) species. Hence, these nucleoside phosphoramidites contain protecting groups on such sites in order to mitigate side-reactions during synthesis cycles. Common protecting groups include benzoyl (on A and C) and isobutyl (on G) (Figure 1.10), which are labile to basic hydrolysis and can be quantitatively removed by standard treatment with concentrated ammonium hydroxide at 55°C for 5 hours.⁶⁵ Another universal protecting group is placed on the non-nucleosidic phosphate oxygen as a 2-cyanoethyl phosphotriester which masks the negatively charged oxygen during oligo synthesis. This group is also susceptible to basic treatment via a β -elimination process to release acrylonitrile as a neutral leaving group.

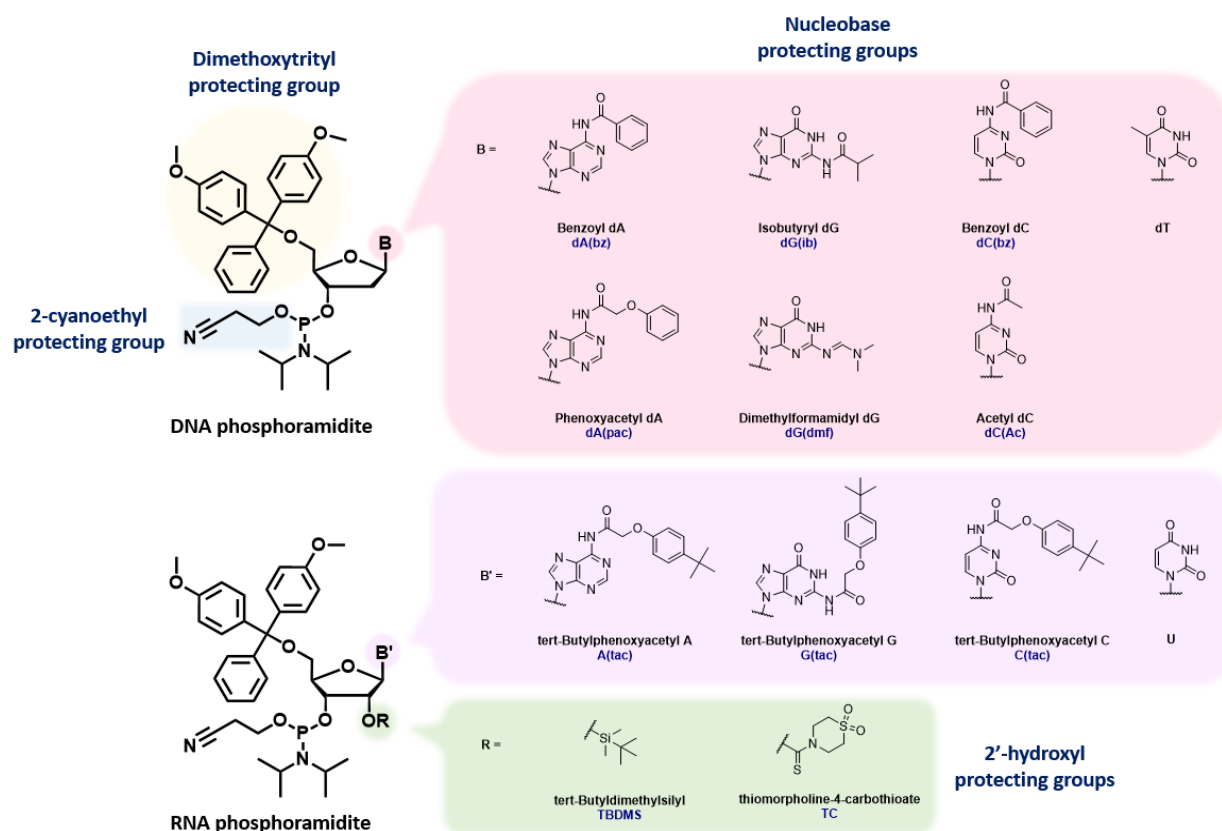
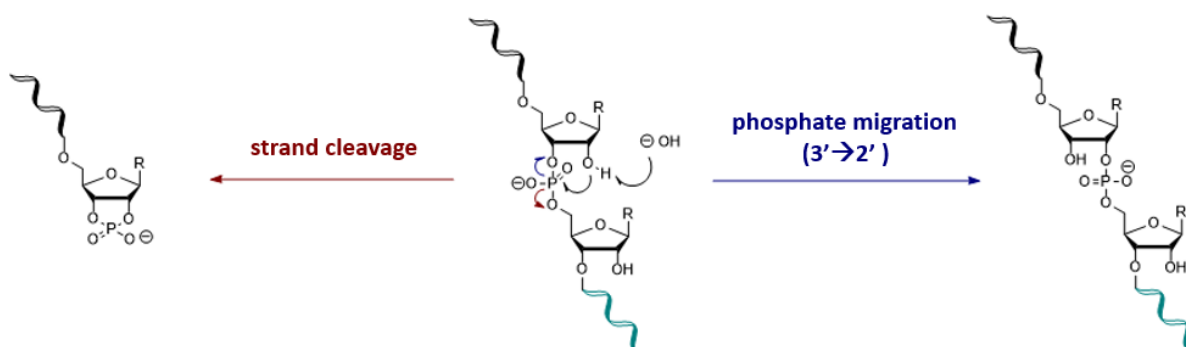


Figure 1.10 | Chemical structures of phosphoramidites with various protecting groups

Some oligonucleotides bear chemical modifications which are susceptible to highly basic environment. In such cases, more labile protecting group strategies on nucleobase is employed. For instance, phenoxyacetyl (on A), dimethylformamidyl (on G) and acetyl (on C) are common on commercially available phosphoramidites which require shorter exposure in concentrated ammonium hydroxide or milder conditions such as methanolic potassium carbonate for complete deprotection. For RNA synthesis, the presence of 2'-hydroxy groups requires additional consideration for protection because they can attack incoming phosphoramidites and promote intramolecular nucleophilic attack onto the neighbouring phosphate, causing strand cleavage or 3'→2' phosphate migration (Scheme 1.2), during ammonia deprotection. Thus, the need to mask the reactivity of 2'-hydroxyl is essential throughout coupling cycles until other protecting groups are removed. Popular protecting groups for this purpose are *tert*-butyldimethylsilyl (TBDMS) and thiomorpholine-4-carbothioate (TC), which are removed by triethylamine trihydrofluoride (TEA·3HF) in DMSO and neat ethylenediamine (EDA) respectively (Figure 1.10). These 2'-protecting groups are moderately stable under basic conditions but not with prolonged heating. As a result, the robust protecting groups on nucleobases, such as benzoyl and isobutyryl, are replaced by more labile *tert*-butylphenoxyacetyl groups which can be quantitatively removed without heating.⁶⁵



Scheme 1.2 | Side reactions of 2'-hydroxy group under basic conditions

This scheme illustrates the possible side reactions of 2'-unprotected RNA molecule during standard basic treatment, which generates faulty products including cleaved strands and phosphate-migrated oligonucleotides.

1.3.3 Chemical Modifications

Scalable synthesis of DNA and RNA has enabled the rapid development of nucleic acids-based therapeutics and biotechnology as discussed in Chapter 1.2.3. However, the utility of unmodified oligonucleotides in biological systems is limited by multiple factors: (i) DNA and RNA are naturally occurring molecules which are susceptible to enzymatic degradation during metabolic processes; (ii) they are large and densely packed with phosphate oxygen anions along their backbones so that they are repelled by the negatively charged phospholipid bilayer at the cell membrane; (iii) the low concentration of oligonucleotides available inside cells, due to inefficient delivery or rational design to avoid toxicity, hinders the binding with their cognate targets; and (iv) on some occasions, exogenous DNA or RNA can trigger undesirable immune response of the host.^{29,66–68}

To address these limitations, various strategies to modify oligonucleotides have been developed to enhance their physical and chemical properties for biological applications, including *in vivo* stability, delivery and targeting-binding efficacies. In a broad sense, chemical modifications can be categorised by the position at which they are introduced on an oligonucleotide, which includes: (i) the sugar and (ii) backbone, which contribute to its structure and involve in substrate recognition by nucleases; (iii) the nucleobase, which engages in base-pairing with a complementary strand; and (iv) the 5'- and 3'-termini, which are susceptible to exonuclease digestion (Figure 1.11). Many reported modifications aim to confer desirable properties to these sites.^{29,66,69–72} For example, the 2'-hydroxy group of RNA can be substituted by an O-methyl (2'-OMe),^{73–75} O-methoxyethyl (2'-MOE)⁷⁶ or fluoro (2'-F)⁷⁷ group which provide higher stability against nucleases and improve target binding affinity ($\Delta T_m = +0.9 - 2.5^\circ\text{C}$ per modification) due to changes in the structure and electronic properties of the sugar, respectively. An increasingly popular strategy to modify the sugar as locked nucleic acid (LNA), which contains a carbon bridge between the 2'-OH and 4'-carbon, is

justified by its tremendous stabilising properties ($\Delta T_m = + 4.0 - 8.0^\circ\text{C}$ per modification) on RNA-target binding,^{78–80} which renders it desirable in situations where high-affinity binding is warranted, such as in therapeutic ASOs.^{81,82} Apart from the ribose moiety, the native phosphodiester backbone is involved in most enzymatic degradation pathways as well as exhibiting a polyanionic character on oligonucleotides. These attributes have prompted research interest in artificial substitutes which can reduce its susceptibility to enzymes and enhances its permeability to cells. A well-developed modification is a phosphorothioate (PS) linkage, in which a non-bridging oxygen atom of the internucleoside phosphate is replaced by sulfur. This substitution has been shown to not only inhibit enzymatic degradation,⁸³ but also engage in binding with proteins or cell surface due to its reduced charged density (making it more lipophilic) and capability to form covalent disulfide bonds with cysteine that in turn promote biodistribution and internalisation to cells.^{84,85} Hence, it is commonly employed to either fully (in ASOs) or partially (in siRNAs) modify therapeutic oligos in order to improve its metabolic stability *in vivo* as well as delivery efficacy to target tissues.²⁹

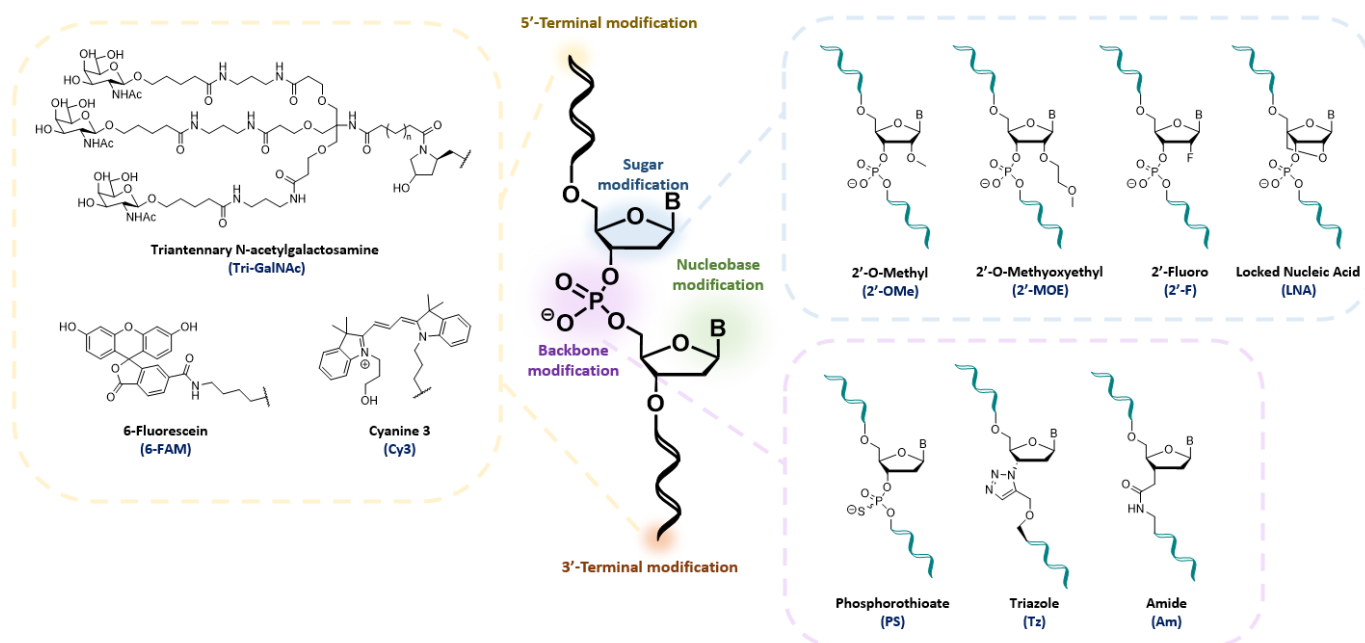


Figure 1.11 | Schematic overview of possible modification sites on oligonucleotides and selected examples of modifications

Another class of modifications has been developed to replace the anionic phosphate with charge-neutral internucleoside linkages, such as triazole and amide, which are structurally akin to phosphodiester but resistant to nuclease digestion.^{86,87} Early examples have been reported more than two decades ago to provide synthetic accessibility to these molecules,⁸⁸⁻⁹⁰ which have since been gathering attention as more evidence emerged that they can be biologically active alternative backbone for therapeutic applications. For example, our group has investigated in the properties of amide-substituted oligonucleotides in antisense applications, in which they have shown enhanced stability as well as potency in RNase H-mediated RNA digestion and exon-skipping actions.^{82,91} These attributes hold promising potential for developing future oligonucleotide therapeutics.

Besides structural modifications which can improve stability and binding properties, it is also possible to introduce small-molecule or biomolecule conjugates that confers a wide scope of functionalities to oligonucleotides. These modifications can be placed internally or, more commonly, at the termini. For instance, Givosiran® - an FDA-approved siRNA drug - contains a tri-N-acetylgalactosamine (GalNAc) conjugate at the 3'-end of its passenger strand that serves to target asialoglycoprotein receptors (ASGPr) expressed on hepatocytes and enhance internalisation to liver tissues.²⁹ Another platform for oligo delivery is antibody conjugates, which can bind specific antigens or biomarkers of certain cell types and target specific tissues from circulation. This strategy has been investigated on enhancing delivery efficacy of siRNA to tissues such as skeletal muscle by appending a fragment antigen-binding (Fab) region to its 3'-end.⁹² Fluorophores are also commonly incorporated to oligos with a short carbon linker. Such moieties can generate fluorescent read-outs under an optical detector or confocal microscope and therefore enable tracking of oligonucleotides by quantitative^{93,94} or qualitative means.⁵¹

1.3.4 Synthetic Strategies for Modified Oligonucleotides

(I) Modified Phosphoramidites (pre-coupling)

Precise control on the location of chemical modifications within an oligo is essential for therapeutic applications. For short oligonucleotides (i.e., <30 nt) such as ASOs and siRNAs, modifications are predominantly introduced on individual nucleoside before solid-phase synthesis. There are a vast number of commercially available and synthetically viable options to access modified phosphoramidites or resin-bound nucleoside with desired structural motifs which are compatible to standard coupling in a synthesiser (Figure 1.12). Ribose modifications such as 2'-F or LNA or nucleobase modifications such as C8-alkyne T are installed on the phosphoramidite and routinely used in oligo synthesis. Chemical handles such as amino group or functional molecules such as GalNAc or cyanine-3 can also be phosphitylated (i.e., a hydroxy group on the molecule undergoing addition reaction to a phosphite) or immobilised on solid support to act as building blocks during phosphoramidite coupling cycles.^{95,96}

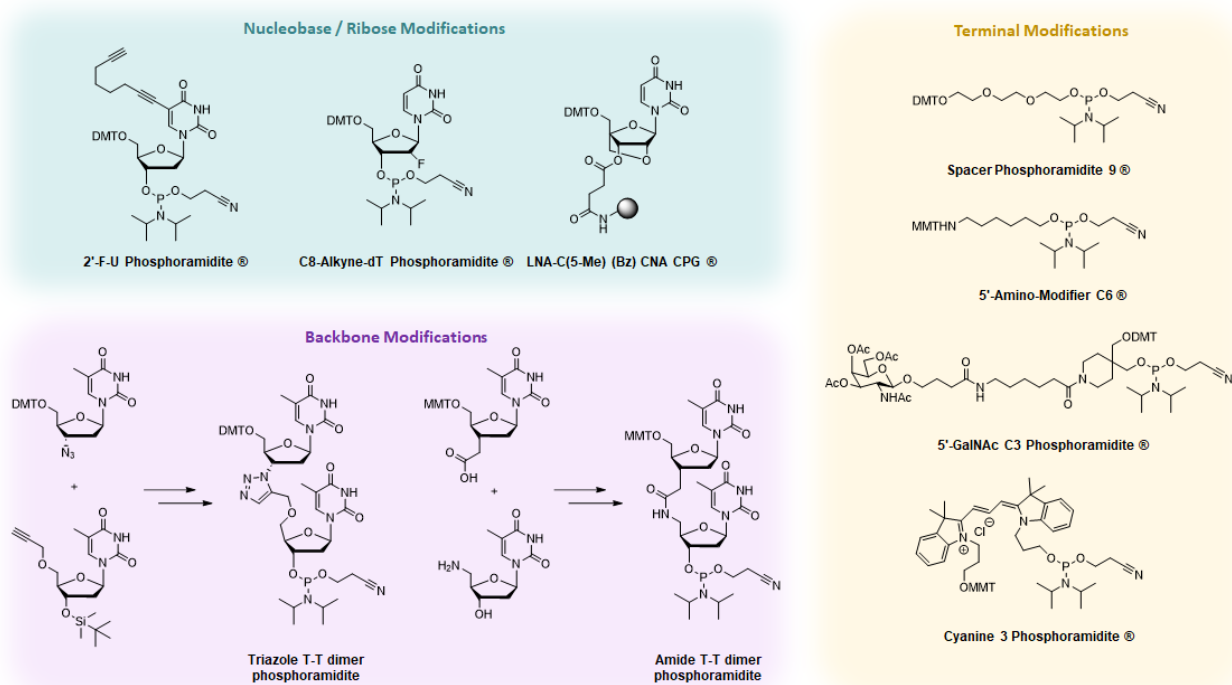


Figure 1.12 | Selected examples of phosphoramidites with various modifications

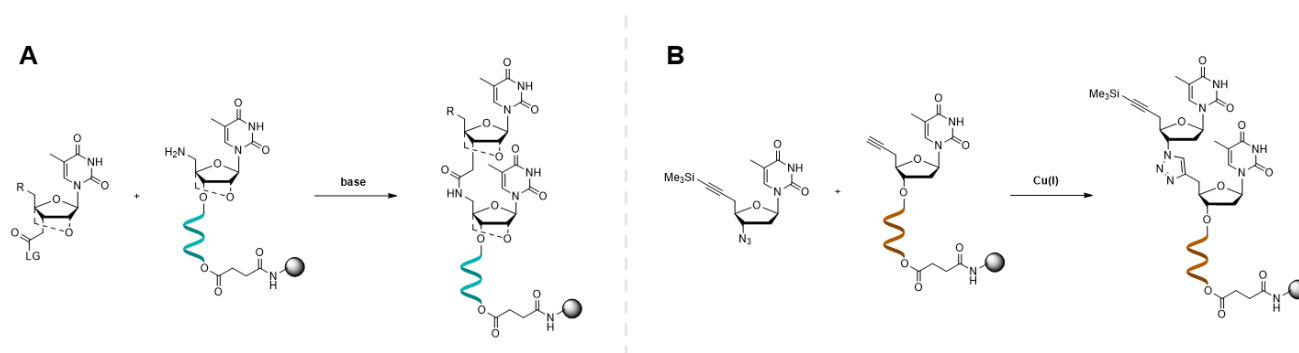
This figure illustrates the chemical structures of modified phosphoramidites categorised as nucleobase/ribose, terminal and backbone modifications. The first two categories are common from commercial sources^{95,96} while the third is synthetically accessible by reported methods.^{91,97}

Backbone modifications, however, involve a change in the bond formation chemistry and are not easily afforded by employing modified phosphoramidite monomers during coupling. Phosphorothioate moiety, for example, is often introduced non-stereospecifically by replacing iodine by sulfurising reagents such as tetraethylthiuram disulfide (disulfiram) during the oxidation step which will generate a P(V)-S bond in lieu of phosphate. Of note, recent work by Baran and co-workers reported a chiral P(V)-containing nucleoside which can directly couple with growing oligonucleotides to yield stereospecific PS linkage, but its usage remains niche.^{98,99} Other non-phosphorus backbone modalities such as amide and triazole are not accessible via standard phosphorus chemistry. To incorporate these unnatural backbone modifications, researchers have developed a ‘dimer approach’, by which two nucleosides are first covalently linked by the desired backbone modification and then phosphitylated as a single unit for standard coupling in a synthesiser.^{90,91,97} Although this approach offers a feasible way to site-specifically introduce backbone modifications, its application is limited by a few factors: (i) the dimer synthesis requires a large number of steps; (ii) a complete set of building blocks entails $4 \times 4 = 16$ dimers which can be synthetically challenging to access; and (iii) it is not possible to extend the unnatural linkage consecutively beyond two nucleosides from dimer phosphoramidites. As a result, it has been of ongoing interest to develop a streamlined synthetic strategy to access modified backbones.

(II) Modified Coupling Chemistries

The scope of solid-phase coupling chemistry is by no means limited at phosphoramidite addition. As discussed in the previous session, its early development was targeted at forming amide bonds for the preparation of polypeptides.⁵⁴ Based on this knowledge and subsequent advances in amide coupling, Von Matt and colleagues reported a 5'-amino-3'-carboxymethyl dT analogue which efficiently couples with 5'-amino oligonucleotide (introduced by 5'-amino phosphoramidite monomer) on solid support and affords amide linkages that span

consecutively over 15 nucleosides.¹⁰⁰ Recently, our group reported another strategy to access more challenging amide linkages between LNA nucleosides containing 5'-amine and 3'-active ester in solid phase and generated oligonucleotides with up to 4 LNA/amide modifications.⁸² This approach offers additional flexibility to introduce ribose/nucleobase modifications on each monomer and hence opens a new avenue to explore therapeutic properties of mixed modifications (Scheme 1.3A).



Scheme 1.3 | Generalised reaction scheme for the synthesis of amide and triazole linkages on solid support

A. Generalised reaction scheme of the formation of amide linkage from a resin-bound 5'-amino nucleoside and free 3'-carboxyl nucleoside.^{82,100} R = DMT-O- or MMT-NH-, LG = pentafluorophenoxy
 B. Formation of triazole linkage from a resin-bound 5'-alkynyl nucleoside and free 3'-azido nucleoside.¹⁰¹ Black ball represents solid support.

As another charge-neutral alternative like amide, the use of triazole linkage to conjugate biomolecules has been popularised since the inception of Copper-catalysed Azide-Alkyne Cycloaddition (CuAAC), also known as ‘click chemistry’ for its robustness and chemospecificity in joining an azide and alkyne moiety with high yield.¹⁰² Although click chemistry is versatile towards solution- or solid-phase reactions, this section will focus on its use over solid-phase coupling with oligonucleotides, particularly on forming internucleoside triazole linkages and labelling oligonucleotides with small-molecules.^{103,104} An early example is the work by Isobe and colleagues on the assembly of a 15-mer oligonucleotide with all internucleoside linkages fully substituted by 1,4-triazole via sequential CuAAC between 5'-alkynyl (silyl-protected) nucleosides and their 3'-azido counterparts (Scheme 1.3B).¹⁰¹ Apart

from nucleoside addition, contemporaneous work by Morvan and co-workers demonstrated the solid-phase assembly of glycoclusters on oligonucleotides via CuAAC-based coupling between alkyne-containing oligomer and azido carbohydrate derivatives.^{105–107} By this solid-phase click method, the resultant click product can be cleaved and deprotected with standard ammonia treatment to afford a carbohydrate-oligonucleotide chimera. These examples indicate the versatility of solid-phase synthesis which can support a wide scope of coupling chemistries for modified oligo synthesis. In Chapter 2, the mechanism and novel applications of CuAAC in oligonucleotide chemistry will be further discussed.

(III) Post-Synthetic Ligation and Labelling

Despite the popularity of solid-phase synthesis in the manufacturing of modified oligonucleotides for therapeutic purposes, thanks to its high efficiency and scalability, the physical and chemical limitations on stepwise coupling efficiency (<100%) inherently restrict the length of oligonucleotides accessible (200-mer for DNA and 120-mer for RNA).¹⁰⁸ Alternatively, enzymatic synthesis poses little constraints on length but the incorporation of modified nucleosides into oligos, especially RNA, remains challenging.¹⁰⁹ To circumvent these limitations, a solution-phase ligation strategy can be employed to covalently link two or more short, synthetic oligo fragments with desired modifications to form a single oligonucleotide via various chemistries. Ligation reactions between two oligos are usually performed in micromolar concentrations and thus this strategy generally has two features: (i) the reaction between the two oligos must be chemo-selective and high-yielding, otherwise side-reactions could drastically reduce its yield; (ii) a DNA splint which bridges two reacting oligos by complementary base-pairing is usually needed to increase the effective concentration of the two reacting groups. In the past two decades, a number of ligation chemistries have been developed to probe sequence mismatches¹¹⁰ as well as generate ultra-long (>300 nt) circular DNA compatible with rolling-circle amplification (RCA) (Figure 1.13).^{111–113} In general, these

examples demonstrated the utility of chemical ligation on preparing lengthy, and circumstantially, modified oligonucleotides beyond the capabilities of solid-phase synthesis.

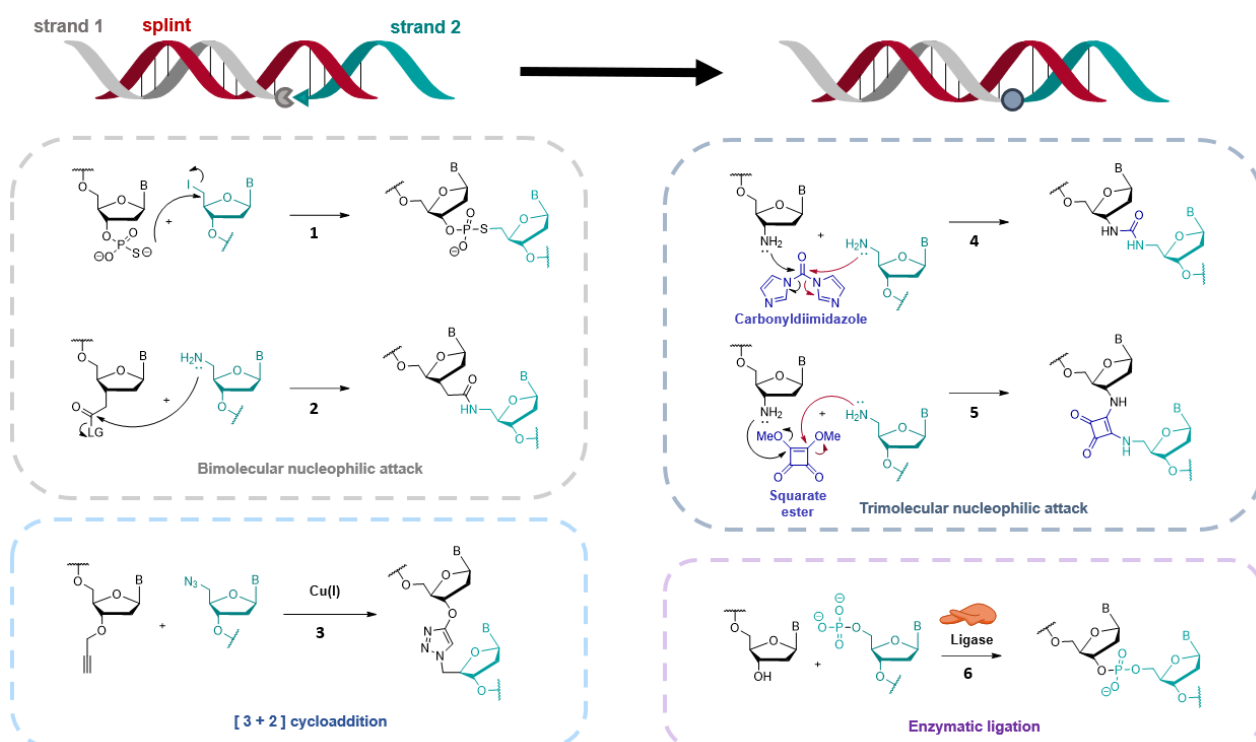


Figure 1.13 | Schematic overview of various post-synthetic ligation strategies

Top: an illustration of the general ligation strategy to covalently join two oligonucleotides via different reactive groups (partial circle and arrow) templated by a splint (red strand). Bottom: reaction schemes of various reported ligation strategies: (1) thiophosphate-iodide S_N2 reaction,¹¹⁰ (2) amide coupling,¹¹¹ (3) CuAAC,¹¹² (4) CDI-mediated urea coupling,¹¹³ (5) squaramide coupling¹¹³ and (6) enzymatic ligation.¹¹⁴ LG represents a leaving group.

On the other hand, enzymatic ligation is an alternative technique to ligate oligonucleotides by natural phosphodiester linkages. For example, T4 DNA ligase can join a 5'-phosphoryl DNA with a 3'-hydroxy DNA in the presence of a splint which brings the two into close proximity, forming a phosphate bond.^{115,116} Such activity is also compatible with RNA/DNA substrates, forming RNA-DNA hybrids, albeit at a lower efficiency. However, more extensive structural changes such as backbone modification or fluorescent labels is likely to impede reaction efficiency.¹¹⁰ In situations where heavily modified oligonucleotides are required, this method may not offer the best outcome.

Apart from ligating two oligo strands, small molecules can also be incorporated onto a strand by chemical or enzymatic means. Typically, a chemical handle (i.e., a reactive group) is installed on the oligo via coupling with modifier phosphoramidites or via enzymatic incorporation of modified building blocks. These functional groups can then react with its cognate partner to generate a conjugated product.

Commercial vendors offer a wide selection of modifiers on phosphoramidites or CPG formulae such as 5'-amino and 3'-alkynyl motifs for oligo SPS. These groups can be universally used to conjugate any molecule with a corresponding reactive group (e.g., amine with active ester, alkyne with azide) after the oligo is synthesised and thus offer great versatility with various reaction partners. Besides chemical methods, the incorporation of these functional motifs can be achieved by enzymatic reactions. For example, yeast poly(A) polymerase (PAP) was shown to append modified nucleoside triphosphates (NTPs), such as 2'-azido-2'-dNTP, onto the 3'-end of RNA molecules.¹¹⁷ This reaction incorporates an azide group onto the RNA which can then react with an alkyne-bearing molecule by CuAAC or copper-free strain-promoted azide-alkyne cycloaddition (SPAAC) to yield a bioconjugate. Alternatively, it is also possible to directly incorporate modified nucleotides such as non-canonical fluorescent bases into DNA by polymerases.¹¹⁸ This strategy allows the synthesis of fluorescent DNA amplicons in polymerase chain reaction (PCR) at a low cost, but it is limited by enzyme constraints and poor positional control of the modified base.

1.3.5 Enzymatic Synthesis of Ribonucleic Acids (RNA)

In the past century, RNA molecules were largely seen as a transient carrier of genetic information or effector of niche cellular functions without intriguing applications. But with the development of siRNA therapeutics, CRISPR-related technologies and mRNA vaccines, scientific interest in these molecules has grown significantly.¹¹⁹ This drives up the demand for

the synthetic capabilities of RNA ranging from short (< 30 nt) to ultra-long lengths (> 1000 nt). As previously discussed, chemical synthesis of long RNA is a challenging task as the coupling efficiency of RNA phosphoramidites is generally lower. To satisfy demands for long RNA such as CRISPR gRNA or mRNA, some commercial vendors offer services to synthesise RNA sequence for tailored purposes.^{120,121} On a manufacturing scale, where kilogram-quantities of RNA are required, *in vitro* transcription (IVT) is commonly employed in the production process.¹²² Compared to natural synthesis of RNA in cells, IVT can take place in a reaction vessel with all essential components supplied. These include a dsDNA template which bears a promoter sequence and the sequence of interest, an RNAP which recognises the promoter and reads through the downstream sequence to polymerise RNA strand, and nucleoside triphosphates (NTPs) which are building blocks of the polymerisation reaction (Figure 1.14). The benefits of IVT over SPS on RNA production are multifold: (i) it enables the synthesis of long RNA unreachable by SPS, which is currently limited at ~100 nt¹²³; (ii) enzymatic synthesis is template-dependent and highly accurate that generates products with high purity and requires simple purification procedures; and (iii) the up-scaling of solution-phase IVT is more facile than SPS which requires costly investment in specific equipment such as DNA synthesisers.

In the production of mRNA, the 5'-cap is a crucial molecular motif that interacts with the eukaryotic translation initiation factor (*eIF4E*) to initiate translation and thus has to be incorporated to the transcript.¹²⁴ During IVT, it is possible to introduce modified caps at the 5'-end of the transcript by supplying the cap reagents, such as an N7-methyl guanosine linked to the first transcribed nucleotide via a 5'-5' triphosphate bridge (m7GpppX, where X = nucleoside). At the end stage, a homopolymer of adenosine, also referred as poly(A) tail, is appended to the transcript by poly(A) polymerase. The properties of this motif, such as length or modifications, can impact the stability and translation efficacy of mRNA.¹²⁵

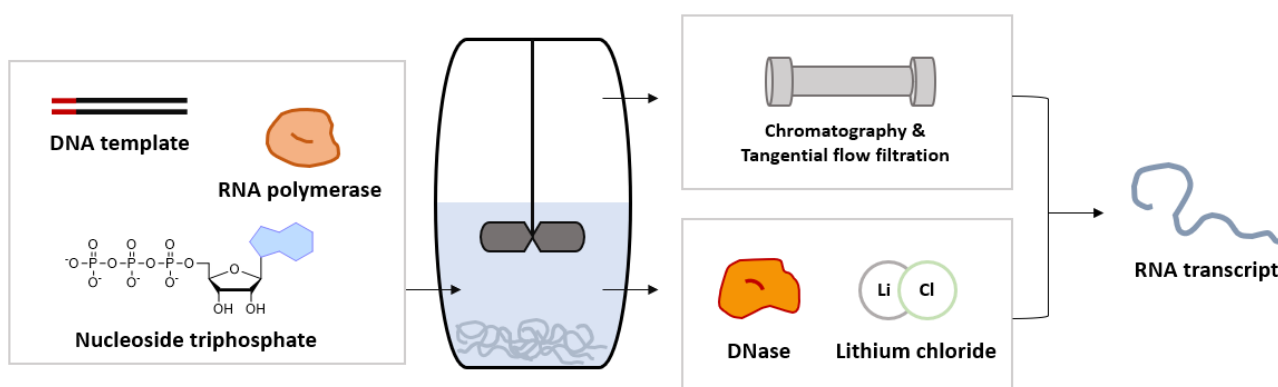


Figure 1.14 | Schematic representation of workflow in enzymatic RNA synthesis

This figure illustrates the production of RNA transcript in lab/industrial scales. In a reaction vessel, DNA template with the desired sequence and nucleoside triphosphates (NTPs) are incubated with RNA polymerase which synthesises RNA transcripts by reading through the template. The resultant mixture can be subjected to DNase treatment to remove the dsDNA template and precipitated by lithium chloride (LiCl) salt to yield the pure RNA product. Alternative, large-scale purification is obtained by established chromatographic strategies (e.g., anion exchange and ion-pair reverse-phase) coupled with tangential flow filtration.

1.3.6 RNA Modifications

As RNA emerged as a potent therapeutic agent, the need for site-specific modifications will only become more prominent. Currently, bacterial RNAP, which are commonly used in RNA manufacturing, have been shown to tolerate NTPs with nucleobase modification, albeit to a limited extent. For instance, RNAP can incorporate N7-substituted purine NTPs into RNA transcript with good efficiencies.^{126,127} Nonetheless, it is much more challenging to introduce bulkier nucleobase substituents, backbone modifications and modified ribose structures as they have less structural resemblance to the natural substrates of RNAP. Also, the insertion of these modifications appeared to be stochastic so that control over the site-specificity is lacking.

In areas where short, modified RNAs (i.e., < 100 nt) are required, SPS remains the most common synthetic strategy. For example, the gRNA component in CRISPR systems can be synthesised by phosphoramidite coupling with modified nucleoside monomers (e.g., 2'-OMe, 2'-F RNA and LNA), which enable researchers to probe the locations of ribose modifications

tolerated by Cas9 endonuclease.^{128–130} More recently, our group reported a method to ligate two synthetic RNA strands to produce a functional sgRNA via click chemistry.¹³¹ This strategy involves the installation of 5'- or 3'-amino-modifier on the synthetic RNA and subsequent labelling with the click-reactive moieties (i.e., azide or dibenzocyclooctyne). Such post-synthetic labelling method not only presents a means to produce lengthy RNA with shorter modified fragments but also enables one-step coupling with small molecules with corresponding reactive groups to generate RNA bioconjugates. Despite its usefulness, chemical labelling has its limitations. As RNA is more prone to degradation, the conditions employed in labelling chemistry can attract particular concerns. For example, amide coupling in basic medium can lead to hydrolytic degradation while the presence of copper in CuAAC can generate reactive radicals that damage the integrity of the RNA strand.¹⁰⁴ The presence of truncated RNA or oxidised nucleobase may affect its biological functions and downstream analysis. In view of this, mild reaction conditions are usually preferred for RNA labelling. One example is strain-promoted azide-alkyne cycloaddition (SPAAC) which requires no copper catalyst but a strained alkyne that is energetically activated to react with an azide species. This will be discussed in Chapter 3 for the preparation of CRISPR gRNA conjugates.

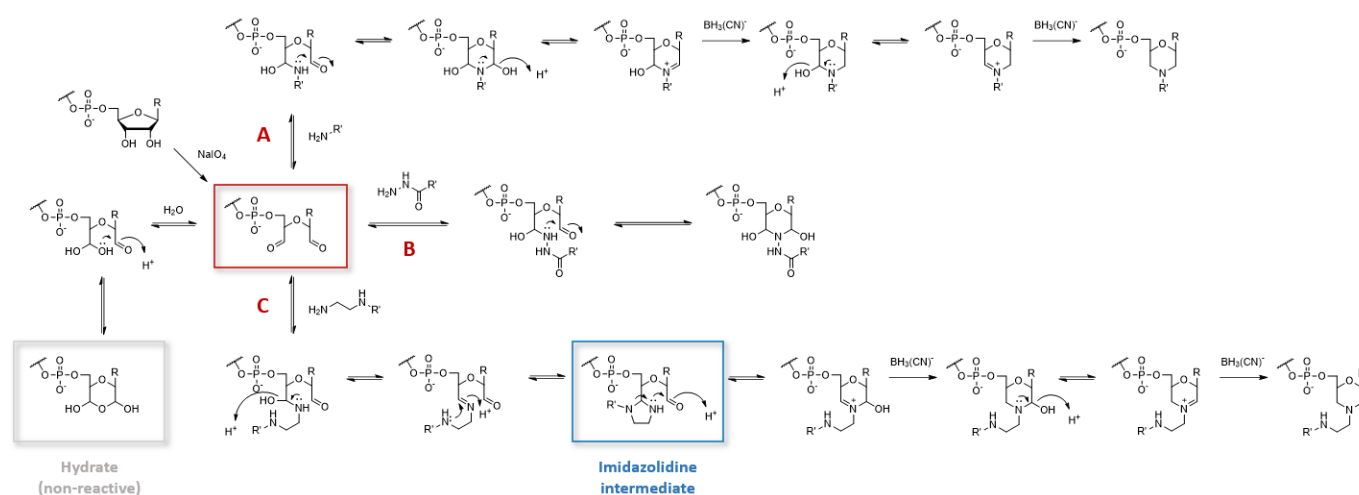
Besides solution-phase conjugation chemistry, click chemistry also demonstrates the potential in generating artificial backbone on RNA with good biocompatibility. Fujino and co-workers have reported a method to form trimeric RNA by CuAAC coupling between 5'-alkynyl and 3'-azido ribonucleoside monomers, with relevant protecting groups.¹³² This triazole-linked RNA fragment was then phosphitylated and incorporated onto a 13-mer 'mRNA' which could be read through and translated into simple peptides by bacterial ribosome *in vitro*. The study reveals the possibility of combining SPS and click chemistry to access mRNA analogues with backbone modifications, which may provide tools for researchers to modulate their properties *in vivo*.

To incorporate modifications onto long RNAs (i.e., > 300 nt), however, the limitations on oligo length and nucleoside substrate scope restrict the utility of SPS or enzymatic synthesis. Precise spatial control on internal modifications on nucleosides (e.g., backbone, ribose or nucleobase) hence remains particularly challenging. Alternatively, researchers have investigated different methods to modify RNA post-synthetically to generate bioconjugates by the combination of enzymatic reactions and solution-phase chemistry. To this end, some strategies to enable terminal modifications on RNA with mild conditions have been reported.

In search of compatible chemistries to transform RNA, a two-step reaction involving oxidation and reductive amination has attracted some attention over the past two decades. This type of reaction commonly utilises a mild oxidant, such as periodates, to open the ribose ring via 3'-diol oxidation; the resultant 3'-dial species subsequently undergoes reductive amination with an amine in the presence of a mild reductant, such as sodium cyanoborohydride, to afford a morpholino-like linkage that connects the amino derivative to the RNA (Scheme 1.4). For example, propargylamine was appended onto a 21-mer RNA via this strategy at an isolated yield of c.a. 50% and subsequently labelled with azido reagents via CuAAC.¹³³ On the other hand, direct amine addition (without reduction) can be achieved by using hydrazide nucleophiles to introduce Cy3 fluorophore at the 3'-end of RNA via 3,5-dihydroxymorpholino transformation.¹³⁴ This method, however, furnished a lower yield at 37%.¹³⁴

A recent study by Mamot and co-workers has identified a robust method to conjugate small molecules on RNA with length ranging from 1 to 2000 nt. This approach relies on an ethylenediamine (EDA) motif, in replacement of an amine, that is postulated to accelerate the second amine addition via the formation of an imidazolidine intermediate (Scheme 1.4). It was demonstrated that a wide scope of substrates, including propargyl, azidoethyl, biotinylated and fluorophore-conjugated EDA, can be labelled to a trimeric RNA fragment with an HPLC yield of >86%. Furthermore, Cy3-EDA was shown to participate in oxidation-reductive amination

to afford 3'-Cy3 mRNA (997 nt) at 70 – 85% HPLC yield, providing access to fluorescent mRNA for localisation studies in HeLa cells and zebrafish embryos.



Scheme 1.4 | Various conjugation strategies via 3'-dial intermediate of oxidised RNA

This scheme illustrates the general reaction mechanisms of various reported methods to conjugate small molecules on RNA via an oxidised intermediate with a reactive 3'-dialdehyde (red box). Amino derivatives can attack both electrophilic carbonyl groups sequentially to afford an iminium intermediate which is reduced by the reductant, $\text{BH}_3(\text{CN})^-$, in acidic buffer (pathway A).¹³³ Hydrazide derivatives can be used to generate stable 3,5-dihydroxymorpholino linkage without further reduction (pathway B).¹³⁴ Ethylenediamine derivatives have been shown to catalyse iminium formation (second amine addition) via an imidazolidine intermediate which provides an efficient and robust pathway to afford a morpholino-like linkage to connect small molecules with RNA (pathway C).¹³⁵ Notably, the 3'-dial can also be attacked by water to generate a hydrate species, which is unreactive and does not participate in reactions with amino nucleophiles. R = nucleobase; R' = any conjugation motifs.

As our understanding on RNA chemistry and bioconjugation progresses, it has become feasible to generate long RNA bioconjugates for *in vivo* studies. Such development proves particularly valuable when researchers started investigating and improving the utility of RNA-based therapeutic systems such as gRNA-directed CRISPR gene editing or engineered protein synthesis via artificial mRNA, which are lengthy and challenging to modify by conventional IVT or SPS methods. In Chapter 3 and 4, we will explore these bioconjugation strategies and discuss the potential applications of such conjugates to provide insights on the functionality of modified RNA in biological systems.

1.4 Aims and Objectives

The study of chemical modifications has been of immense interest in oligonucleotide research as they provide tools to improve therapeutic properties (e.g., nuclease resistance, organ-specificity) and enable real-time spatial analysis (e.g., fluorescent labelling) of antisense and RNA-based oligonucleotides *in vivo*. Thanks to a large volume of research and commercial availability of modified nucleoside phosphoramidites and triphosphates, these synthons are popular for the incorporation of modified functionalities and small-molecule conjugates onto synthetic oligos by solid-phase or enzymatic syntheses. Such approaches, however, may prove inadequate in certain applications where modifications on long oligos or specific inter-nucleosidic linkages are desired. In this thesis, we aim to explore various solid-phase and solution-phase chemical approaches to access oligonucleotide bioconjugates that are challenging to synthesise by conventional methods, and examine their physical properties and biological functions, if relevant. Finally, the methods reported should provide a helpful toolkit for researchers who are interested in accessing oligonucleotide conjugates discussed in this work.

The specific aims of this work are to:

Chapter 2: develop an efficient synthetic strategy to produce dendritic oligonucleotides in solid phase and study their properties in antisense applications.

Chapter 3: explore bioconjugation chemistries on RNA and prepare gRNA-ASGPr-ligand-conjugate to study receptor-mediated delivery of CRISPR RNP into hepatocytes.

Chapter 4: investigate methods to conjugate mRNA fragments with modified poly(A) tail and ASGPr-ligands which can potentially improve functionality of mRNA *in vivo*.

Chapter 2 – Synthesis of Nucleic Acid Dendrons by Click Chemistry

2.1 Background

2.1.1 Nucleic Acid Dendrons and Dendritic Structures

Dendrimers are an interesting class of highly branched macromolecules with well-defined molecular constitution and distinctive supramolecular properties. They are composed of polymeric molecules branching outwards symmetrically from a single core.¹³⁶ The constitutional unit of a dendrimer can be regarded as a dendron, which is a part of the branching structure with one free valence that can chemically react further (Figure 2.1).

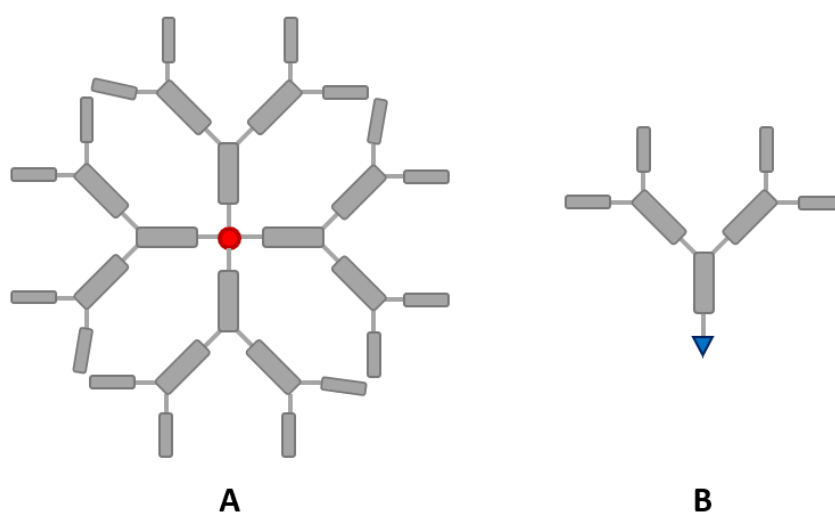


Figure 2.1 | Schematic representation of molecular dendrimers and dendrons

A represents a dendrimer, or dendritic structure, which comprises polymeric repetitive units (grey box) arranged in a branching geometry from its core (red circle); and B represents a dendron, which is a part of the branches from a dendrimer and bears a free valence (blue triangle).

Since the inception of dendrimers such as polyamidoamines (PAMAMs) in laboratory, chemists have been intrigued by the physical properties imparted by the supramolecular structure of these macromolecules. For example, the branched structures possess interior cavities which can encapsulate molecules by hydrophobic or electrostatic interactions and form guest-host complexes.^{137,138} Moreover, the solubility of three-dimensional dendrimers can be modulated by adjusting their terminal groups which are in contact with the solvent. These properties have been harnessed to solubilise hydrophobic drugs, such as steroids, in aqueous environment.¹³⁹ In the early 1990s, the curiosity over the role of branched RNA intermediates

during RNA splicing has prompted Hudson and Damha to investigate the synthesis of branching oligonucleotides, as an early example of nucleic acid dendrons.¹⁴⁰ They have reported a novel method to couple the 5'-hydroxyl of resin-bound DNA oligonucleotide with a protected adenosine 2',3'-O-bis(phosphoramidite) reagent to afford a branching junction with both 3'-5' and 2'-5' linkages between two oligonucleotides. This junction could be further elongated by addition of standard phosphoramidite and connected through this bis(phosphoramidite) reagent resulting in a branching pattern from 5' to 3'. Although this finding showed that branched oligonucleotides are synthetically accessible, the yields are badly limited as the tri-molecular coupling between two oligos and the bis(phosphoramidite) is inefficient and subjected to variables such as oligo length and inter-strand distance. To expand the utility of dendrimeric oligonucleotides, the Southern group from Oxford has reported two branched synthons – 1,3-bis-[5-(4,4'-dimethoxytrityloxy)pentylamido]propyl-2-[(2-cyanoethyl)-(N,N-diisopropyl)] (*Doubler*) and tris-2,2,2-[3-(4,4'-dimethoxytrityloxy)propyloxymethyl]ethyl-N,N-diisopropylaminocyanethoxy (*Trebler*) phosphoramidites – to introduce branching junctions of two and three valencies in dendritic oligos.^{141,142} Each addition of these branching points provides more than one hydroxy group for phosphoramidite coupling and thus strand propagation. As a result, a reiterative process of branching and chain elongation can be performed in a synthesiser to afford oligonucleotide dendrimers branching from 3' to 5'. Since the formation of dendritic cluster requires much fewer coupling steps as each step is done simultaneously on multiple 5'-hydroxy groups, it has been used to impart a large number of negative charges on a PCR primer to shift the electrophoretic mobility of the amplicons from those of compatible length. The addition of this dendrimer-tagged has been shown to have no effect on polymerase activity.¹⁴¹

As the commercialisation of branching phosphoramidite synthons brings down synthetic costs, investigation on the properties and applications of nucleic acid dendrons has

accelerated in recent years. In 2019, a trident siRNA duplex which contains a trebler-based passenger strand (trimer) to carry the guide strands was reported. This construct has been shown to maintain RNAi activity and improve the stability of the guide strand against exonucleases and thus is a promising motif for therapeutic use.¹⁴³ In the subsequent year, another group reported the incorporation of dendritic oligonucleotides on DNA brick nanostructures to improve their nuclease resistance in biological environment.¹⁴⁴ Soon later, Mirkin's group demonstrated the utility of dendritic poly(T) oligos to promote cellular uptake of peptide cargos via scavenger receptor A-mediated endocytosis, which is triggered by the interaction with polyanionic clusters.¹⁴⁵ The densely-packed nature of DNA dendrons also raises the local concentration of oligonucleotides and improves their interactions with certain proteins. Such property has been harnessed to enhance immunostimulatory effects of CpG-containing oligonucleotides in dendritic constructs, which elicit higher level of cytokine release in macrophage-like cells.¹⁴⁶ These favourable characteristics of dendritic oligos have demonstrated great potential in the development of new therapeutic reagents or materials.

2.1.2 Challenges in the Synthesis of Nucleic Acid Dendrons

Notably, all of the reported constructs from the previous session were synthesised from Trebler-based phosphoramidites which afford dendrons with up to 9 valencies (i.e., number of oligonucleotide strands within one molecule) and isolatable yield (< 1%). Although the yield remains quite poor, these constructs exhibit promising potential in biological applications; it highlights the importance of a more efficient synthetic strategy to access nucleic acid dendrons.

In addition to yields, this 'top-down' approach of dendron assembly which starts from the branching units requires efficient coupling of phosphoramidites on multiple hydroxy termini on one molecule. This could prove challenging for bulky monomers such as TBDMS-RNA and any failure in coupling can result in a large amount of *n*-1 and *n*-2 products which

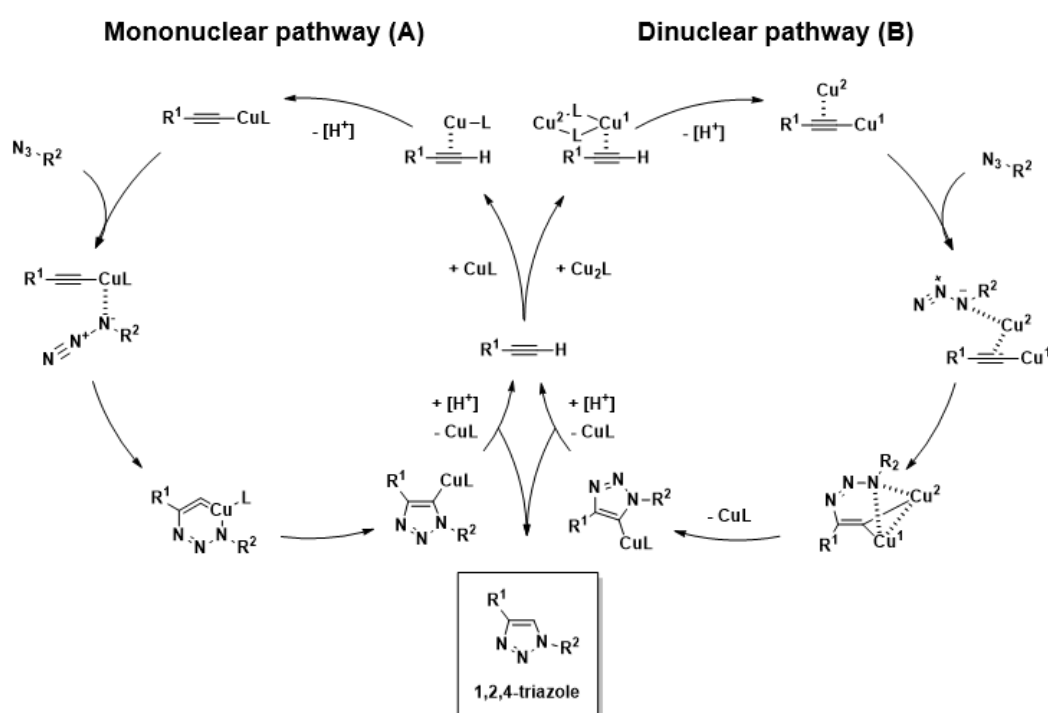
are difficult to remove by chromatography. Moreover, the coupling with multiple phosphoramidites simultaneously requires a high loading of the reagents and extended coupling times which could be wasteful in large scale syntheses. Hence, this chapter will introduce a novel ‘bottom-up’ method to generate nucleic acid dendrons by click chemistry that is fully compatible with standard mono-strand phosphoramidite coupling protocol.

2.1.3 Click Chemistry on Solid Support

Click chemistry describes the chemistry that joins two molecular building blocks in a simple, chemo-selective, high-yield reaction under mild conditions with few benign or no by-products.^{147,148} Among all reported click reactions, copper-catalysed azide-alkyne cycloaddition (CuAAC) has been a popular option to form covalent linkages between biomolecules as it is biocompatible and fits most criteria of click chemistry: (i) the azide and alkyne motifs are uncommon in naturally occurring molecules so they only react selectively and are inert to most chemical environments; (ii) the reaction is highly efficient (with near-quantitative yield) in the presence of copper;¹⁰³ (iii) the resultant triazole linkage has been shown to be non-toxic and non-reactive against hydrolysis or oxidation but retain considerable polarity to interact with water and hydroxy groups.^{104,149} These properties render triazole linkage an excellent mimetics for the peptide/phosphate linkages in natural biopolymers such as proteins and oligonucleotides as exemplified in Chapter 1. In this chapter, the terminology ‘click reaction’ is used interchangeably with CuAAC.

Since the discovery of CuAAC, various reaction mechanisms have been proposed. Density-functional theory (DFT) calculations and kinetic studies identified two plausible competing pathways in triazole formation: (i) mononuclear pathway (slower) and (ii) dinuclear pathway (faster). In both pathways, the initial step involves the replacement of a Cu(I) ligand donor with an organic alkyne, resulting in the formation of a π -complex. This coordination

significantly lowers the pKa of the alkyne and favours its deprotonation to form the Cu-acetylide adduct in aqueous environment. At this point, the organic azide coordinates to the copper core to displace its ligand and situate itself in close proximity to the acetylide. This step reduces the activation barrier considerably and promotes the formation of a six-membered ring via the attack of the distal nitrogen on the azide to the C-2 of the alkyne. In the final step, the ring contracts to generate the triazolyl-copper derivative, which then undergoes protolysis to release the 1,4-disubstituted triazole product and complete the catalytic cycle.^{150,151}



Scheme 2.1 | Reaction pathways of copper-catalysed azide-alkyne cycloaddition¹⁰⁴

Pathway A represents the azide-alkyne cycloaddition process catalysed by a mononuclear $Cu(I)$ species, which proceeds slower. Pathway B represents a competitive catalytic process catalysed by a dinuclear $Cu(I)$ core. The latter process is more kinetically favourable and hence proceeds faster.^{150,151}

In an early study by Meldal's group, the compatibility of CuAAC with SPS has been demonstrated in the couplings between resin-bound peptide with terminal alkyne group and various azide-bearing molecules with 67% to 95% yield.¹⁵² The chemistry has soon been applied to oligonucleotide SPS to synthesise a 15-mer DNA in which the backbone is substituted fully by triazole, from 5'-alkyne and 3'-azide-bearing building blocks.¹⁰¹ In both cases, the benefits of solid-phase synthesis to enable high loading of reactants and easy removal

of excess reagents are coupled with the superior reactivity of CuAAC to afford excellent coupling efficiency, which is essential for the assembly of biomolecules with polymeric characteristics. From this perspective, the robustness of solid-phase CuAAC coupling inspired a click-enabled synthetic method to access nucleic acid dendrons that is compatible with standard oligonucleotide SPS. This chapter will discuss this novel application and the physical properties of the dendron products generated by copper click chemistry (referred as ‘click dendrons’ in later sections).

2.2 Aims and Objectives

The new synthetic route proposed in this chapter is a ‘bottom-up’ approach in which the oligonucleotides are first assembled on solid support via standard phosphoramidite coupling followed by a 5’-azide conversion step and crosslinking with a multivalent alkynyl linker. It is envisaged that the new synthetic strategy is more efficient than conventional doubler/trebler-based coupling as the formation of failure sequences is much less prominent than the latter. Also, this protocol requires only standard coupling which means no excess reagent or coupling duration is required, making it compatible with most existing programmes on DNA synthesisers. In this chapter, the concept and synthesis of multivalent linkers for oligonucleotide crosslinking are described. Next, the synthetic procedures, purification and characterisation of DNA dendrons are discussed. This part also examines the reaction optimisation and remarkable observation on isolating the dendritic product to act as a reference for readers interested in accessing this type of molecules. Finally, this chapter reports also the physical properties (e.g., duplex stability) and interactions between click dendrons and enzymes relevant to antisense applications and hence provides insight into their therapeutic potential.

To assess the conversion efficiency of 5'-hydroxy-to-azide conversion, we prepared a T₁₂ model oligo on solid support (1 µmol/column) synthesised by standard coupling cycles and detritylated at the last step, with purity > 99%. The identity of the material was assessed by detachment and deprotection of an analytical amount of resin-bound oligo (c.a. 0.1 mg) in concentrated ammonia hydroxide then analysed by UPLC-MS and deconvoluted on MassLynx™. This analytical method is applied to all oligonucleotides in this chapter and details can be found in Chapter 5 and 6. The resin was then subjected to treatment under standard azide conversion conditions. The resultant oligonucleotide product was identified as a major peak with m/z corresponding to 5'-azido-T₁₂ (calc. mass: 3613) with 98% purity (with the remaining 2% as unreacted T₁₂) (Figure 2.2).

Next, we further studied the possibility to convert non-nucleosidic alcohol into azide which supports click reactions when the 5'-terminal moiety is not a nucleoside. To this end, we synthesised a 5'-propyl alcohol T₁₂ and subjected the resin to the same treatment with MTPPI and sodium azide. On UPLC-MS, the majority of the product mixture was obtained as 5'-propyl azide T₁₂ as a major peak with found mass 3753 (Figure 2.3). This experiment demonstrated that terminal hydroxy groups can be efficiently converted into azide groups on resin-bound oligonucleotides. This set of conditions has been used for all azide conversion in this chapter.

Oligonucleotide	Sequence (5'-to-3')
ON2.1	HO-TTTTTTTTTTTT
ON2.2	N ₃ -TTTTTTTTTTTT
ON2.3	HO (CH ₂) ₃ -TTTTTTTTTTTT
ON2.4	N ₃ (CH ₂) ₃ -TTTTTTTTTTTT

Table 2.1 | List of oligonucleotide sequences in 5'-hydroxy-to-azide conversion studies

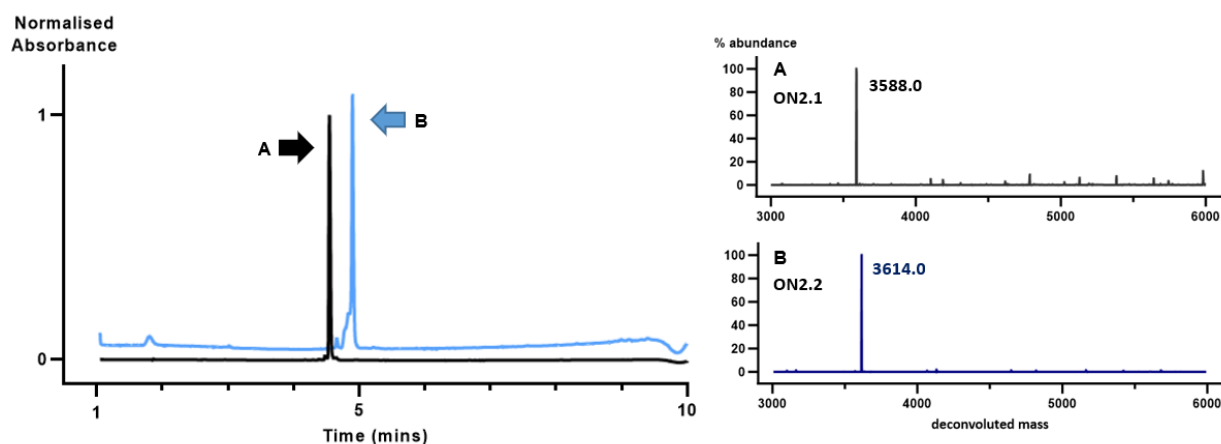


Figure 2.2 | UPLC-MS results of 5'-hydroxy-to-azide conversion on T₁₂

Overlaid UPL chromatograms are shown on the left. Black line indicates the retention profile of ON2.1 (starting material HO-T₁₂, calc. mass = 3588.4). Blue line indicates that of ON2.2 (conversion product N₃-T₁₂, calc. mass = 3613.4) under standard azide conversion conditions: (i) 0.5 M MTPPI/DMF, r.t., 1 h; (ii) wash with dry DMF and MeCN then dry under Ar; (iii) sat. NaN₃/DMF, 55°C, 5 hours on UPLC-MS. Deconvoluted masses of individual peaks are shown on the right.

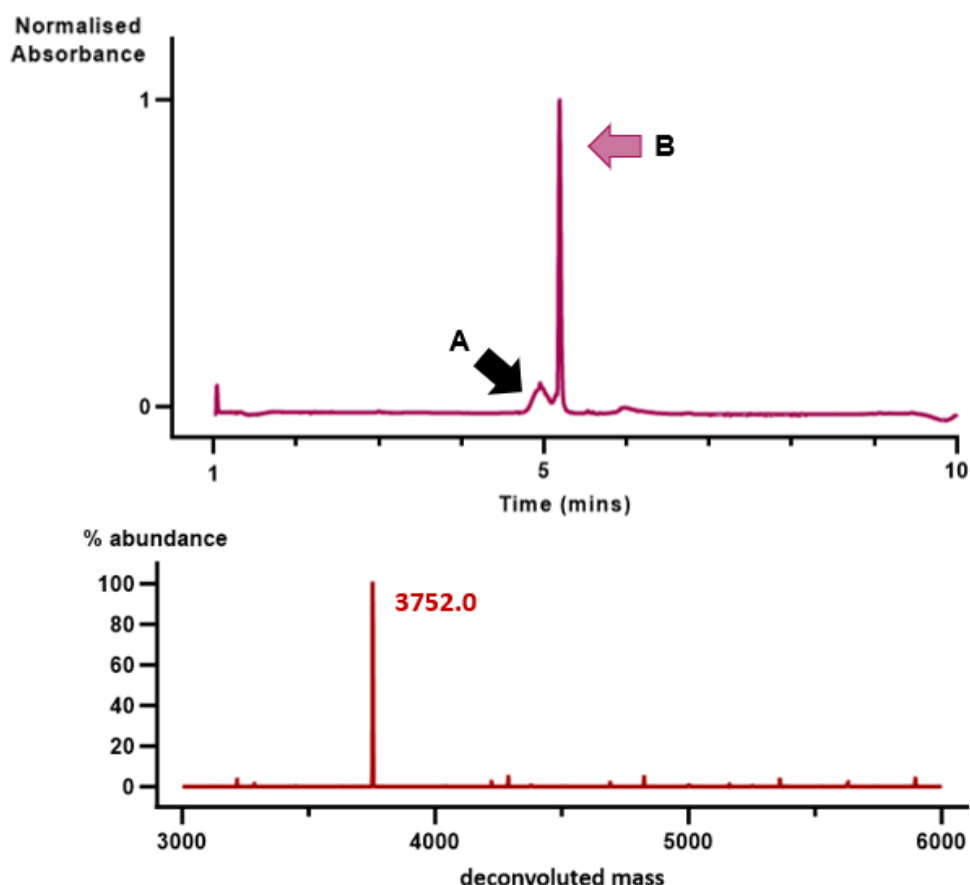


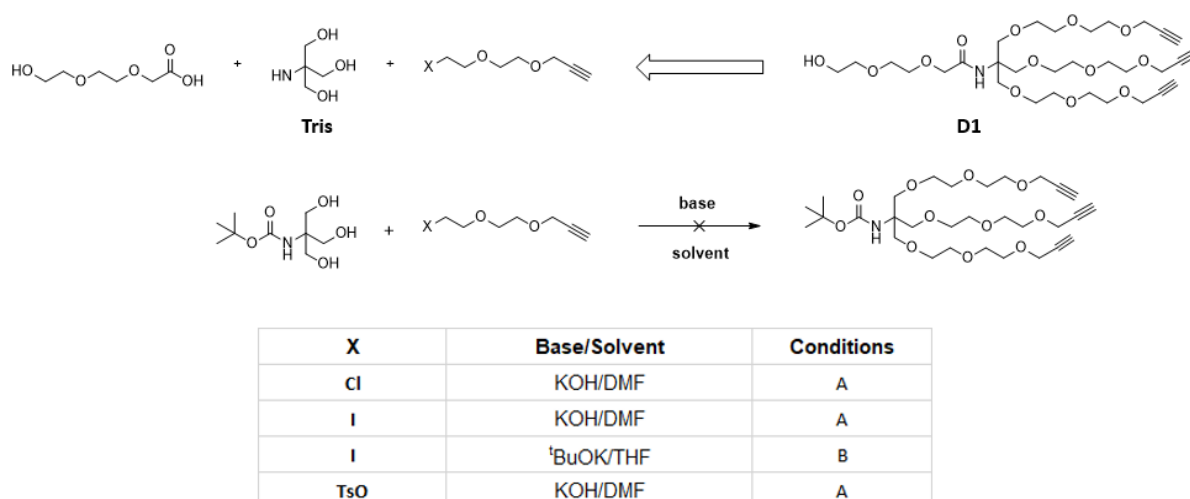
Figure 2.3 | UPLC-MS result of 5'-hydroxy-to-azide conversion on hydroxypropyl T₁₂

Retention profile of the crude product from azide conversion of starting material ON2.3 (calc. mass = 3725.5). Both peaks A and B have a mass corresponding to ON2.4 (calc. mass = 3750.5).

2.3.2 Synthesis of Multivalent Linkers

With a robust method to access 5'-azido oligonucleotides on solid support, we envisaged that a crosslinking reaction with a polyalkynyl linker via CuAAC be plausible. When having more than one alkyne group (i.e., valence), this linker should crosslink neighbouring oligonucleotides up to the number of its valence by forming triazole linkages, resulting in a dendritic oligonucleotide covalently bound at the 5'-end. At a starting point, we aimed to demonstrate the feasibility of simultaneous click reactions and thus designed the linker prototype as a branching molecule with three terminal alkynes and one hydroxy end.

There are a number of considerations in this design: (i) the linker should efficiently react with the 5'-azido oligonucleotides without side reactions in CuAAC conditions; (ii) it should be stable under standard deprotection conditions (i.e., strongly basic media); (iii) it should be hydrophilic in nature to avoid reduction in aqueous solubility or hydrophobic aggregation of the oligo construct; (iv) the reactive alkynyl ends should be spaced in an optimal distance by flexible linkers to react with nearby azides; and (v) it should be achiral to avoid the formation of stereoisomers which are difficult to isolate by chromatography. Based on these rules, we first designed a synthetic route starting from a Boc-protected Tris base which contains three hydroxy groups that can be symmetrically elongated with an alkynyl polyethylene glycol (PEG) linker via nucleophilic substitution; the deprotected amino group can then undergo condensation with a hydroxy acid to afford prototype **D1**. However, the attempts to drive ether formation between Tris and various alkynyl-PEG-halides did not result in satisfactory yield, with the isolation of only monosubstituted product and unidentified species (Scheme 2.3). This is attributed to the fact that aliphatic alcohol is difficult to deprotonate to form alkoxide ion which is more potent nucleophile and three substitution reactions in one-pot do not proceed efficiently. We avoided using NaH as a base for fear that the carbamate nitrogen will be deprotonated, resulting in undesired substitution products.



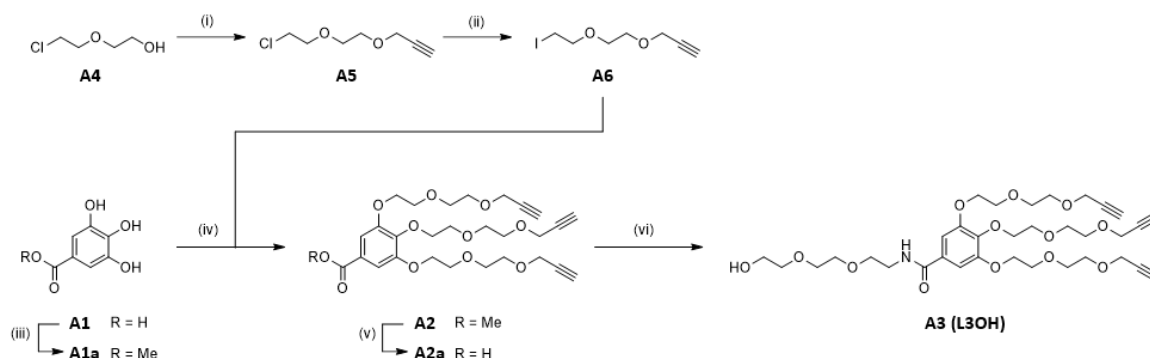
Scheme 2.3 | Reaction design for trialkynyl linker from Tris core

This scheme shows the initial reaction design to access the trialkynyl linker, D1 from a Tris-based core. Various reaction conditions are indicated in the table below. Conditions A: 0°C to r.t., 16 hours and B: -78°C to r.t., 16h. Negligible quantity of product was isolated from all four conditions.

From this observation, we decided to approach the desired alkynyl linker via a polyphenolic core because phenols are readily deprotonated under basic conditions and phenoxide ions can act as good nucleophiles. In this route, gallic acid, **A1** was chosen as it bears three adjacent phenols at the 3,4,5-positions and a carboxylic acid which enables orthogonal condensation chemistry in the next step. First, the acid was protected as methyl ester (**A1a**) by condensation in acidic methanol to mask its nucleophilicity. An electrophilic alkynyl PEG spacer, **A6** was obtained by propargylation of 2-(2-chloroethoxy)ethanol, **A4** followed by halide exchange with sodium iodide. This transhalogenation step is essential because chloro spacer **A5** has much lower reactivity than its iodo counterpart **A6** when acting as an electrophile in the following S_N2 reaction (8% yield), due to chloride being a poor leaving group. The next step involved Williamson ether reaction between **A1a** and **A6** with Na₂CO₃ as a base in DMF to furnish 75% yield of the trialkynyl intermediate **A2**. After quantitative removal of the methyl group, **A2a** was activated by HBTU and DIPEA to form a hydroxybenzotriazole (HOBt) active ester, which undergoes aminolysis with an incoming amino-PEG-alcohol to form the product **A3** in 85% yield (Scheme 2.4). Alternative coupling

reagents such as N,N'-dicyclohexylcarbodiimide (DCC) and 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) were also tested but they provided lower yield at around 30%. This might be due to carbodiimide being a poorer leaving group than HOBT whose anionic form is resonance-stabilised by the aromatic ring.

We expected that this novel trivalent linker (**A3**, referred also as **L3OH**) is suitable for oligonucleotide dendron formation because it contains flexible PEG spacers which separate the alkyne groups within inter-oligonucleotide distance (25 Å) at standard loading density (47 µmol/g) on resin. This means that all alkyne groups possess sufficient degree of freedom to react with 5'-azides but not too much which disfavours the reaction entropically. Moreover, **L3OH** is connected through mostly PEG and amide moieties which are polar and provide excellent aqueous solubility.



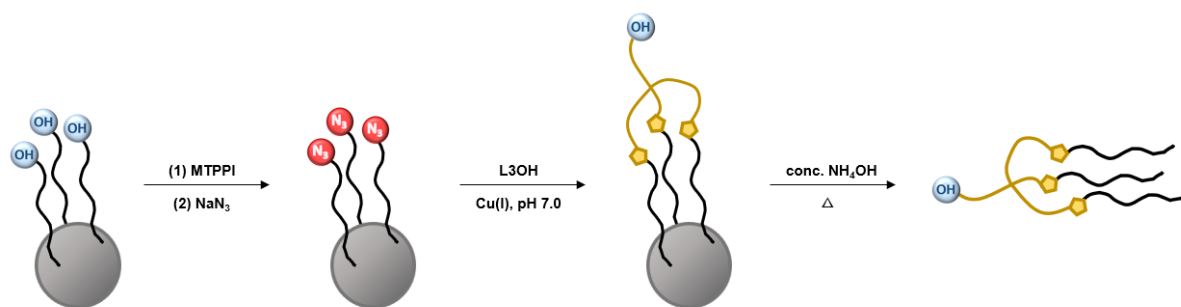
Scheme 2.4 | Reaction design for trialkynyl linker from polyphenolic core

Reaction conditions are as follows: (i) NaH, dry THF, -78°C to reflux, 4h, 75%; (ii) NaI, acetone, reflux, 24h, 98%; (iii) conc. H₂SO₄, methanol, reflux, 18h, 87%; (iv) K₂CO₃, DMF, 60°C, 18 hours, 75%; (v) aq. NaOH, dioxane, 50°C, 4 hours, quant.; (vi) 2-[2-(2-aminoethoxy)ethoxy]ethanol, HBTU, DIPEA, DMF, r.t., 24h, 85%.

2.3.3 Optimisation of CuAAC Crosslinking

After obtaining the linker molecule, we proceeded to study the efficiency of CuAAC crosslinking with oligonucleotides on solid support. The workflow of this protocol starts from 5'-hydroxy-to-azide conversion followed by copper-catalysed click reaction with **L3OH**

(Scheme 2.5). Assuming any azide group exposed at the 5'-end of oligonucleotides can react quantitatively with all available alkyne groups, the loading of alkynyl linker in the reaction was stoichiometrically controlled. In order to minimise the formation of mono- and bi-conjugated products, the loading of L3OH was limited in the initial trial. For CuAAC conditions, the copper source was chosen to be CuSO_4 which is aqueous soluble and can be reduced to Cu(I) *in situ* by sodium ascorbate. Copper-chelating ligand, trishydroxypropyltriazolylmethylamine (THPTA), was added to accelerate the rate of CuAAC and protect the Cu(I) core from oxidation.¹⁵⁵ The reaction medium is a mixture of water and DMSO, which solubilises all organic components, in triethylammonium acetate (TEAA) buffer (pH 7.0). We first attempted to aliquot 0.2 μmol (6 mg) equivalent of T_{12} oligo on resin (34 $\mu\text{mol/g}$) in a vial and subject it to the reaction mixture with various equivalents of **L3OH**: 10 nmol (0.15 eq.), 20 nmol (0.30 eq.) and 30 nmol (0.45 eq.). The conversion profile is visualised by UPLC-MS in Figure 2.4.



Scheme 2.5 | Cartoon representation of CuAAC crosslinking protocol

This general scheme shows the synthesis of oligonucleotide dendrons in solid phase. Grey ball represents CPG and black curve represents oligonucleotide chain.

Within 0.45 equivalents (alkyne to azide), it was observed that the formation of tri-conjugated (trimer) T_{12} oligo dendron, $\text{L3OH-(T}_{12})_3$ (ON2.5) increases with **L3OH** loading. The conversion rate from $\text{N}_3\text{-T}_{12}$ to $\text{L3OH-(T}_{12})_3$, calculated as the ratio between areas of the two peaks, follows an increasing trend from 9% at 0.15 eq. to 41% at 0.3 eq. to 58% at 0.45 eq.

(Note that this value is an estimation based on 200 nmole of oligonucleotide loading, which could be overestimated due to loss in coupling/weighing errors). Interestingly, only an insignificant quantity of mono- and di-conjugated oligo was observed. The absence of such by-products indicates that subsequent CuAAC reaction(s) after the first triazole formation proceed rapidly. This phenomenon can be rationalised by two factors: (i) when the linker reacts with the first azide and enters the solid-phase from the solvent, the local concentration of resin-bound azide is high relatively to the solution phase; and (ii) the linker is then covalently constrained on the solid surface with low conformational freedom which reduces the entropic cost of reacting with neighbouring azide which is also solid-bound. The acceleration of reaction rate beyond the first CuAAC is essential in a preparative context because any by-products with free valence (i.e., not fully crosslinked) have similar retention properties with the desired product and thus are difficult to remove by chromatography.

From the knowledge that tri-conjugated dendrons are the most favourable product from solid-phase click reactions, we next attempted to drive the crosslinking reaction to completion by increasing the loading of **L3OH** in the reaction mixture. If **L3OH** is present in abundance that overcrowds the surface of the resin-bound oligos, the 5'-azide may react rapidly with the free linker instead of the covalently bound (already reacted) ones. Hence, in this experiment, the concentration of **L3OH** was controlled at 50 and 100 μM but the total reaction volume was increased to 1.0 mL. We also tested running the reaction for shorter durations (4 hours & 8 hours) and elevated temperature (55°C). Lastly, this experiment was conducted with T₁₂ oligo on commercially available 0.2 μmole CPG column directly by passing through click reaction mixture via syringes. This is a more facile approach to remove solvent and wash the resin. The results were summarised in Figure 2.5.

When the reaction was carried out at room temperature with 50 μM (0.75 eq.) of **L3OH** for 4 hours, the conversion was far from complete but similarly low level of by-products was

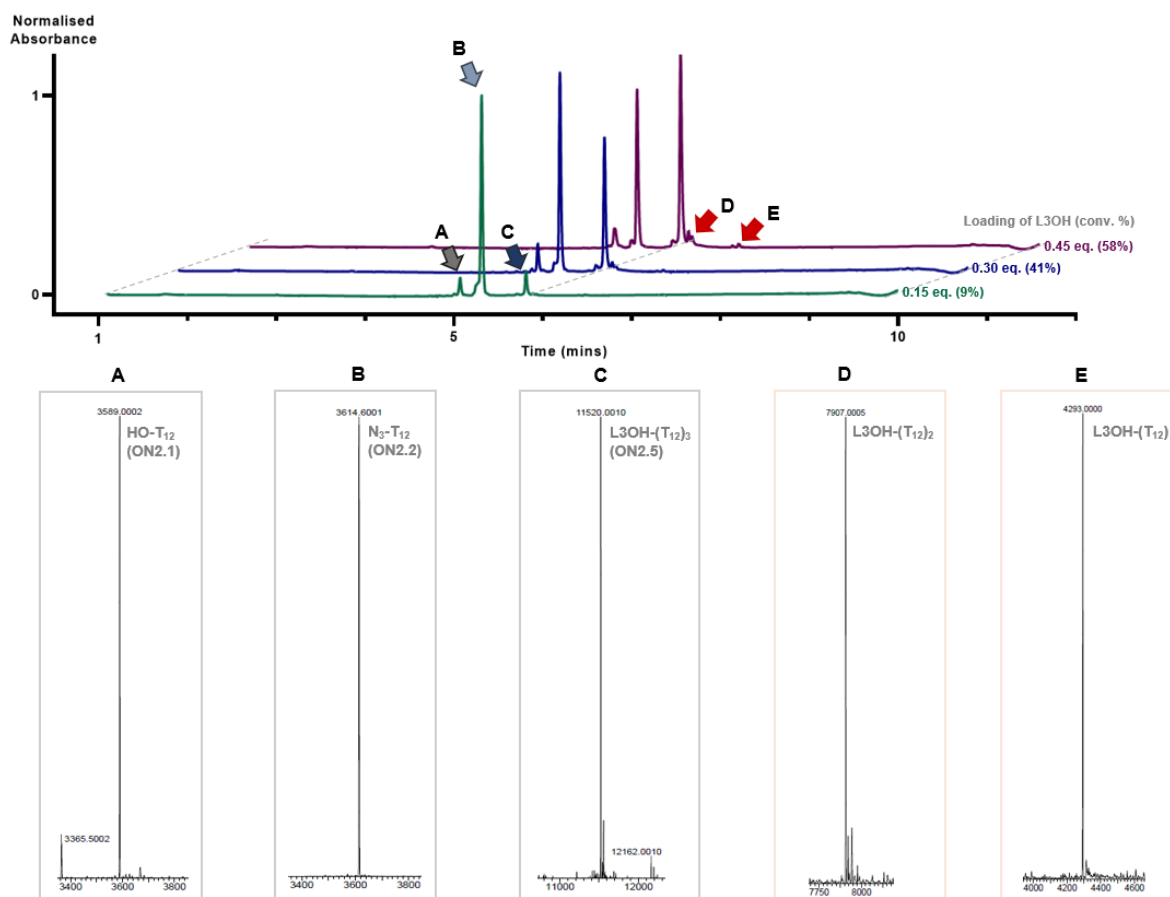


Figure 2.4 | UPLC-MS results from CuAAC crosslinking with various eq. of L3OH

Stacked UPL chromatograms are shown on the top. General CuAAC condition: 0.5 mM CuSO₄, 1.0 mM THPTA, 1.0 mM sodium ascorbate in 100 mM TEAA buffer (20% DMSO, total volume = 200 µL), incubated at r.t. in an Eppendorf Thermomixer®, 24 hours. The three sets of experiment are loaded with various equivalents of L3OH (to a final conc. of 50, 100 and 150 µM). The retention profiles are indicated as green, blue and violet lines, respectively. Deconvoluted masses of the major peaks are shown at the bottom, assigned to a corresponding species: (A) HO-T₁₂, calc. mass = 3588.4; (B) N₃-T₁₂, calc. mass = 3613.4; (C) L3OH-(T₁₂)₃, calc. mass = 11519.5; (D) L3OH-(T₁₂)₂, calc. mass = 7906.1; and (E) L3OH-(T₁₂), calc. mass = 4292.7. Conversion % was calculated as the ratio between area under the product peak and that of starting material, product and by-product.

observed. By increasing the reaction time from 4 hours to 8 hours, the conversion of click reaction has recorded a significant increase to 54%. Since no significant amount of by-product formed, we increased the concentration of **L3OH** to 100 µM to drive the reaction further. The resulting conversion was good, at 80% but the level of bi-conjugated (dimer) product also increased to 14%. Finally, when the resin was incubated with 100 µM of **L3OH** at an elevated

temperature (at 55°C) for 4 hours, the reaction also proceeded at 80% conversion. The results suggest that increasing the linker concentration from 50 μM to 100 μM improve the conversion of crosslinking reaction but also slightly compromise selectivity on the tri-conjugated product; also, the reaction may proceed faster at elevated temperature without decomposition or generating additional by-products.

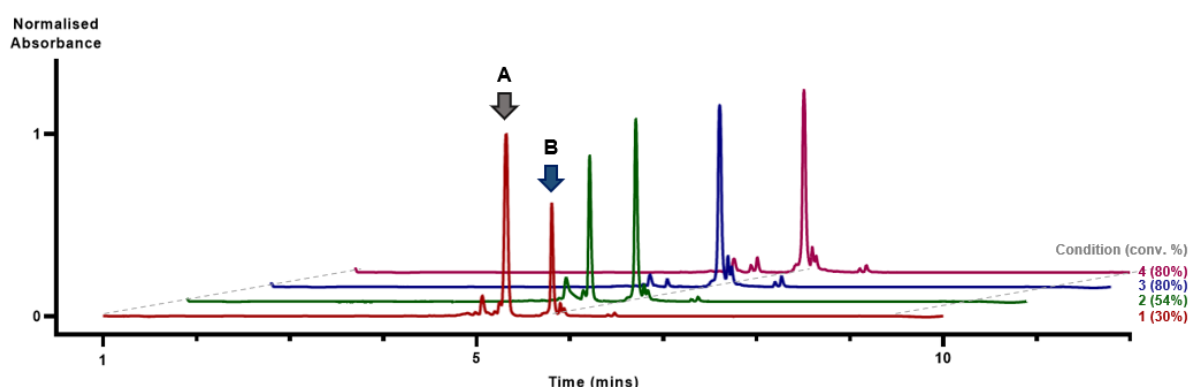


Figure 2.5 | UPLC-MS results from CuAAC crosslinking in various conditions

Stacked UPL chromatograms show the retention profiles of product mixture under various reaction conditions. A indicates the peak corresponding to starting material $\text{N}_3\text{-T}_{12}$ and B indicates the product $\text{L3OH-(T}_{12})_3$. General CuAAC condition: 0.5 mM CuSO_4 , 1.0 mM THPTA, 1.0 mM sodium ascorbate in 100 mM TEAA buffer (20% DMSO, total volume = 1000 μL). Condition (1): 50 μM L3OH, r.t., 4 hours; (2) 50 μM L3OH, r.t., 8 hours; (3) 100 μM L3OH, r.t., 8 hours; and (4) 100 μM L3OH, 55°C, 4 hours. Conversion % was calculated as the ratio between area under the product peak and that of starting material, product and by-product.

2.3.4 Syntheses of DNA and RNA Dendrons

The optimisation results from T_{12} model oligo show that CuAAC crosslinking from single strand to trimer could reach satisfactory conversion at elevated linker concentration within 8 hours. We next sought to test this reaction on a scrambled, mixed sequence to assess the possibility of assembling dendrons with functional sequences (i.e., capable of target binding or antisense applications). This is deemed more challenging because mixed sequences contain A, C and G nucleosides which have bulky nucleobase protecting groups which might change the solubility profile and affect the efficiency of crosslinking. Also, the effect of crosslinking

on nucleobase deprotection kinetics is unknown. Hence, we introduced some variations to the reaction conditions for the mixed sequence.

First, we designed a fully random sequence, TCACTTCGCTTCAAC (ON2.6) and synthesised it on a 1 μ mol CPG column. The UPLC-MS results of the crude product indicated 88% yield of the full-length product (Figure 2.6). Then, the resin was subjected treatment with MTPPI and sodium azide to afford the 5'-azido intermediate (ON2.7) with quantitative conversion. In the next click reaction, the concentration of L3OH was raised to 500 μ M correspondingly to the increase in synthetic scale of oligo from 0.2 to 1 μ mol. The solvent composition was modified to 50% DMSO in order to improve solubility of organic components. To shorten the reaction time, the temperature was set to 45°C and the concentration of copper catalyst was raised to 1 mM. We applied these conditions to the crosslinking reaction of ON2.7. After subjecting the resin to hot ammonium hydroxide treatment (55°C) for 5 hours, the analysis from UPLC-MS showed that the conversion to the trimer product (ON2.8) was 80% but 19% of dimer by-product was also found. The remaining 1% was unreacted oligo.

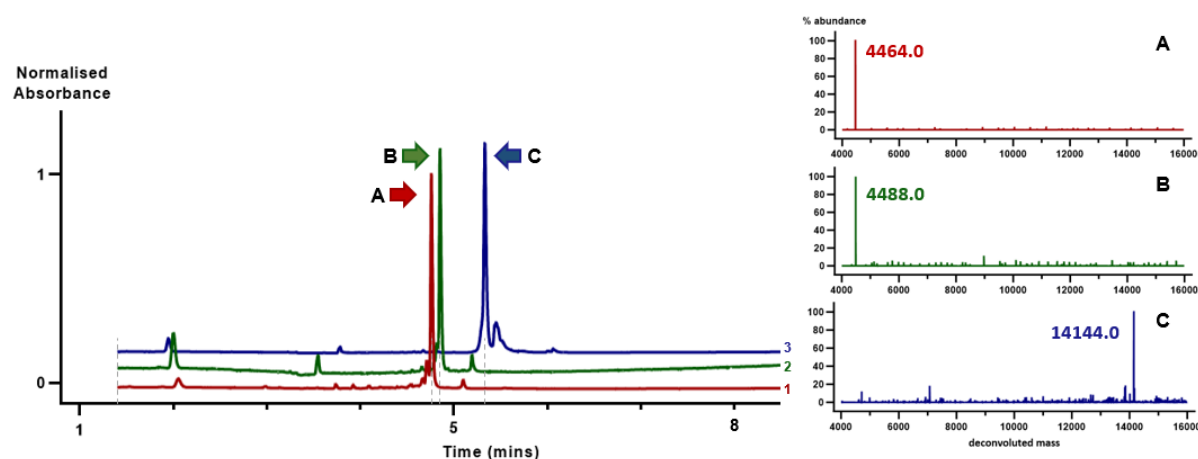


Figure 2.6 | Crude UPLC-MS results from DNA dendron synthesis

Overlaid UPL chromatograms show the retention profiles of the crude sample of (1) ON2.6, starting material (calc. mass = 4462.9); (2) ON2.7, 5'-azido intermediate (calc. mass = 4487.9); and (3) ON2.8, tri-conjugated product (calc. mass = 14143). Deconvoluted mass of individual peaks was shown on the right. CuAAC crosslinking condition: 1.0 mM CuSO₄, 2.0 mM THPTA, 5.0 mM sodium ascorbate, 500 μ M L3OH in 100 mM TEAA buffer (50% DMSO, total volume = 1000 μ L), 45°C, 4 hours.

Encouraged by the positive results from a scrambled DNA sequence, it is intriguing to explore the possibility of crosslinking ribonucleosides to form RNA dendrons. The additional 2'-protecting group present on RNA monomers could further complicate the synthetic conditions because (i) it has to be stable throughout azide conversion and click reaction; and (ii) the linker has to be stable under the deprotection conditions for the 2'-protecting group. Report from Dr. Lapatrada Taemaitree (previous group member) showed that 2'-thiocarbamate (TC) readily decomposes under iodination conditions. Hence, we commenced the study with ON2.9, a 2'-TBDMS protected RNA whose nucleobase sequence is identical to ON2.6 except T is replaced by U (Table 2.2).

Oligonucleotide	Sequence (5'-to-3')
ON2.5	L3OH- (Tr-TTTTTTTTTTTT) ₃
ON2.6	HO-TCACTTCGCTTCAAC
ON2.7	N ₃ -TCACTTCGCTTCAAC
ON2.8	L3OH- (Tr-TCACTTCGCTTCAAC) ₃
ON2.9	HO-UCACUUCGCUUCAAC
ON2.10	N ₃ -UCACUUCGCUUCAAC
ON2.11	L3OH- (Tr-UCACUUCGCUUCAAC) ₃

Table 2.2 | List of oligonucleotide sequences and modifications in crosslinking studies
L3OH represents the linker motif. Tr represents triazole group.

When ON2.9 was subjected to MTPPI/sodium azide treatment, the conversion was 80% on UPLC retention profile. The analytical sample was obtained by treating the resin in hot ammonium hydroxide/ethanol and then in triethylamine trihydrofluoride (3HF·Et₃N)/DMSO before precipitation in sodium acetate buffer/n-butanol mixture. There was no degradation observed on the 5'-azido oligo after HF treatment but the formation of 5'-azido full-length oligo was much less efficient than in DNA, with the shoulder of the major peak identified as unreacted hydroxy-oligo or elimination product from the iodo-intermediate. Moving forward, we treated the remaining resin in the column with CuAAC conditions and analysed the product

with the same deprotection protocol. From the chromatogram, the major peak in the product mixture was identified as the trimer product, ON2.11 by mass. The conversion from the starting material ON2.10 to product ON2.11 was found to be 72% and the formation of dimer by-product was at similarly low level at 11% (Figure 2.7). Likewise, no degraded product was observed after the HF treatment. Although the crosslinking is less efficient in RNA compared to DNA, possibly due to the steric hindrance or rigidity imparted by the additional 2'-TBDMS group, the results indicated the feasibility of generating RNA trimers with the same reaction conditions. It is envisaged that this protocol would be compatible with other types of 2'-modifications such as 2'-fluoro (2'-F) or 2'-O methyl (2'-OMe).

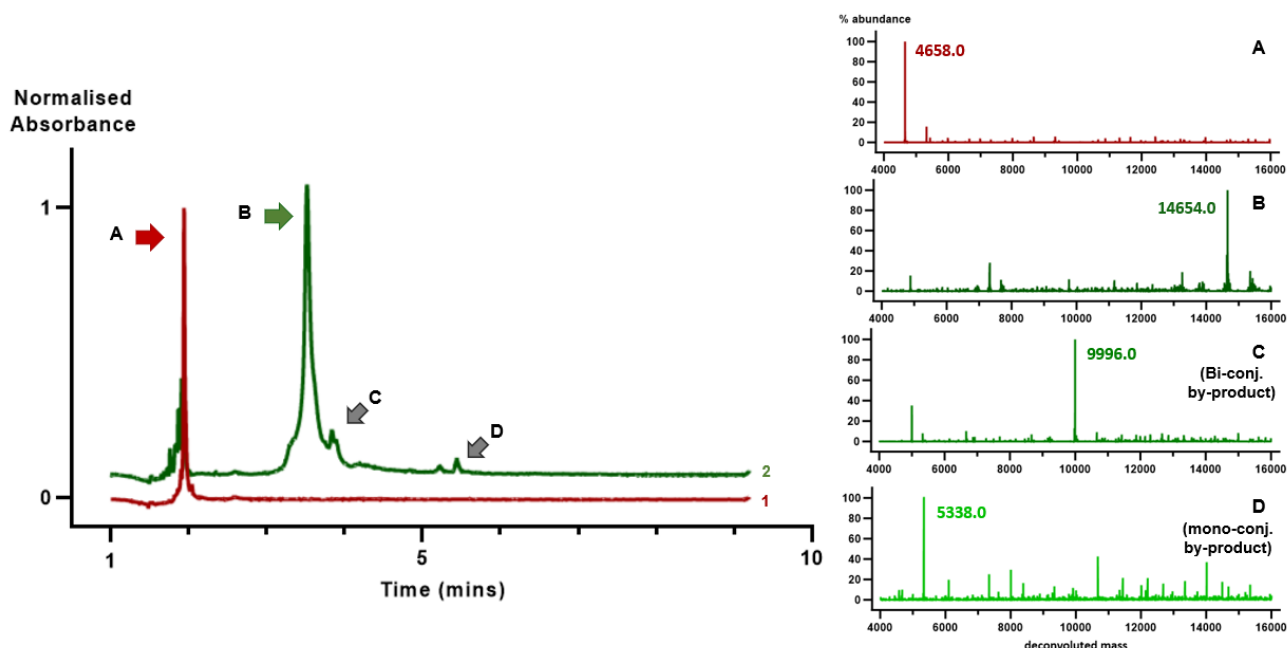
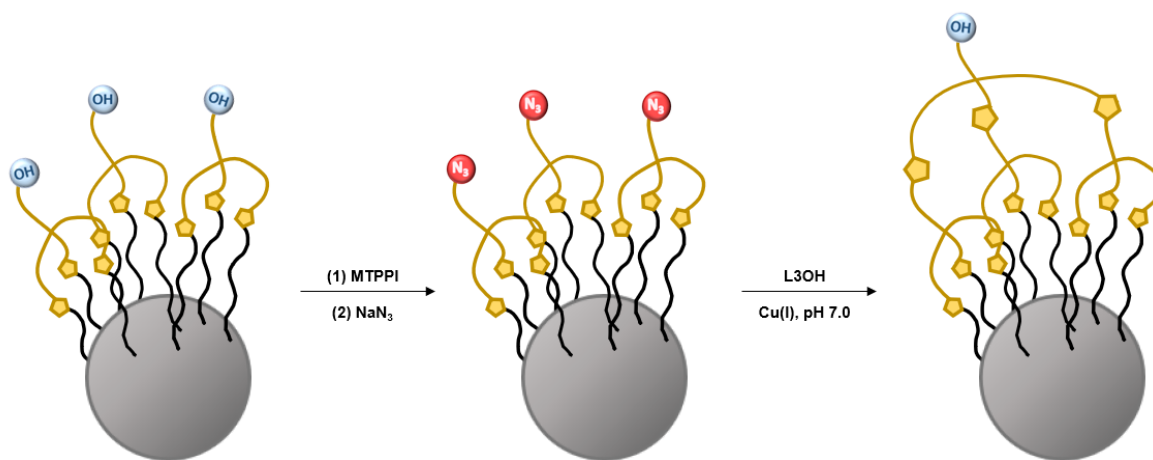


Figure 2.7 | Crude UPLC-MS results from RNA dendron synthesis

Overlaid UPL chromatograms show the retention profiles of the crude sample of (1) ON2.10, 5'-azido intermediate (calc. mass = 4657.8); (2) ON2.11, tri-conjugated product (calc. mass = 14652.7). Deconvoluted mass of individual peaks was shown on the right. CuAAC crosslinking condition: 1.0 mM CuSO₄, 2.0 mM THPTA, 5.0 mM sodium ascorbate, 500 μM L3OH in 100 mM TEAA buffer (50% DMSO, total volume = 1000 μL), 45°C, 4 hours.

2.3.5 Construction of Nonameric Oligonucleotide Dendrons

As the efficiency of azide conversion and CuAAC crosslinking was shown to be good, the majority of oligos on resin should have reacted to form a dendron branching at the 5'-end with a hydroxy group exposed. The following part of this study explores the possibility of forming a nonameric dendrons with two iterative crosslinking steps. First, the 5'-hydroxy groups on the oligo are converted to azide and then react with **L3OH** to form trimers; then, the hydroxy end of the linker is further converted to azide and crosslink again with **L3OH** and two other trimers to form a nonamer (Scheme 2.6). The utility of this chemistry can be demonstrated in multiple applications: (i) when a dense display of oligos is required such as in nanomaterials to improve stability,¹⁴⁴ or (ii) when a large molecular size is desirable such as in antisense therapeutics to avoid renal clearance.¹⁵⁶



Scheme 2.6 | Cartoon representation of crosslinking oligonucleotide trimers

This general scheme shows the synthesis of a nonamer oligo dendron in solid phase. Grey ball represents CPG and black curve represents oligonucleotide chain. The hydroxy ends on oligo trimers are first converted to azide and undergo another iterative step of CuAAC crosslinking.

We first applied the same reaction conditions to 1 μmol T₁₂ oligo on resin to generate trimer T₁₂ (ON2.5). The UPLC profile of the crude product showed an 85% conversion to the desired trimer species while a small amount (6%) of dimer was observed. With a majority of

oligos crosslinked to L3OH, we attempted to convert the hydroxy group on oligo trimers with MTPPI/sodium azide to azido group. The azide conversion proceeded quantitatively with only 3% of starting material (hydroxy trimer) remaining; the dimer by-product was also substituted with azide from its hydroxy group. Hereafter, we refer this linker motif as **L3N₃** (Table 2.3). At this step, the available azide groups for click reaction are theoretically only one-thirds (1/3 μ moles) of the starting scale as three oligos crosslink with one **L3OH** molecule which is further substituted with an azide. Hence, we lowered the concentration of **L3OH** to 250 μ M in order to drive the formation of the nonamer product. The sample from each step was analysed after standard detachment and deprotection.

Oligonucleotide	Sequence (5'-to-3')
ON2.12	L3N₃ - (Tr-TTTTTTTTTTTT)₃
ON2.13	L3OH - [Tr-L3 - (Tr-TTTTTTTTTTTT)₃]₃
ON2.14	L3N₃ - (Tr-TCACTTCGCTTCAAC)₃
ON2.15	L3OH - [Tr-L3 - (Tr-TCACTTCGCTTCAAC)₃]₃

Table 2.3 | List of oligonucleotide sequences and modifications in nonamer construction
L3OH and L3N₃ represent the linker motif with a hydroxy and azido end, respectively. Tr represents triazole group.

In the chromatogram, a broad peak was observed at a high retention time than any monomeric or trimeric species (Figure 2.8). Mass analysis indicated the presence of a mixture of dendritic oligo nonamer (calc. mass = 35312.8; found = 35314) and hexamer (calc. mass = 23768.4; found = 23767) in that peak. In order to resolve the identity of the mixture, the crude mixtures of ON2.5 and ON2.13 were analysed also by polyacrylamide gel electrophoresis (PAGE) on an 8% gel (Figure 2.8). From the first click reaction, we observed an intense band which corresponds to the major peak in UPLC as ON2.5 and a faint band with higher electrophoretic mobility which is inferred as the bi-conjugated dimer by-product. By comparison, the second click reaction generated a mixture of dendritic oligonucleotides with

various valences; the presence of eight bands suggests the formation of dendrons with 2 to 9 valences that include the nonameric product and any by-products from incomplete click reaction. Attempts to separate each species in the mixture via reverse-phase (RP) and anion-exchange (IEX) HPLC were unsuccessful, as dendrons with different valences possess similar retention profiles even in different mobile phases and at elevated temperatures. This phenomenon is more pronounced in RP systems because the affinity of the dendron toward the stationary phase is heavily governed by the hydrophobic linker motif instead of the number of oligonucleotides. From this observation, we decided to isolate the nonameric dendron from preparative PAGE in order to obtain mass characterisation data as well as study its physical properties.

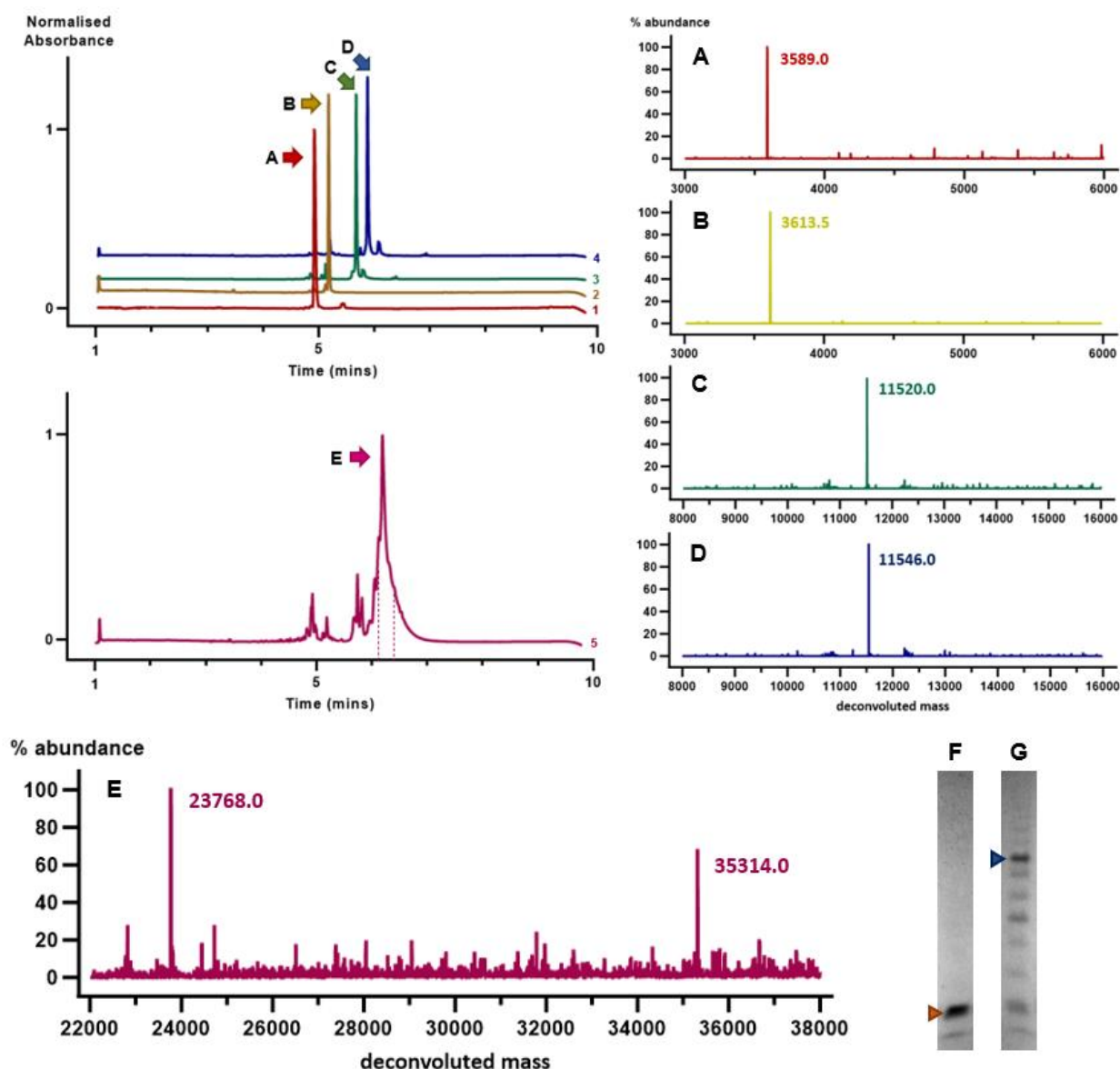


Figure 2.8 | Crude UPLC-MS results from nonameric oligo dendron synthesis

Overlaid UPL chromatograms show the retention profiles of the crude sample of (1) ON2.1, 5'-hydroxy- T_{12} (calc. mass = 3588.4); (2) ON2.2, 5'-azido T_{12} (calc. mass = 3613.4); (3) ON2.5, L3OH- $(T_{12})_3$ (calc. mass = 11519.5); (4) ON2.12, L3N $_3$ - $(T_{12})_3$ (calc. mass = 11544.5). Bottom left: retention profile the crude mixture of (5) ON2.13, nonameric T_{12} product (calc. mass = 35312.8). Deconvoluted masses of peaks A-E are visualised. The results of 8% PAGE analysis of the mixture in (3) and (5) are excerpted from the same gel at identical height, as shown as F and G, respectively. Orange triangle indicates the major band as ON2.5 and blue indicates that as ON2.13. The 1st CuAAC crosslinking condition: 1.0 mM CuSO $_4$, 2.0 mM THPTA, 5.0 mM sodium ascorbate, 500 μ M L3OH in 100 mM TEAA buffer (50% DMSO, total volume = 1000 μ L), 45°C, 4 hours. The 2nd crosslinking conditions: 250 μ M L3OH and other parameters unchanged.

In the following part, we attempted to construct a nonameric dendron with the mixed sequence, ON2.6 which is a better model system for target-binding studies. We applied the same reaction conditions (with 500 μM **L3OH**) for the formation of the trimer dendron, ON2.8 and lowered the concentration of **L3OH** to 200 μM in the second crosslinking (with other conditions unchanged). Then, the crude product mixture was obtained by standard detachment and deprotection. Purification of the nonamer, ON2.15 was performed by denaturing the oligo in formamide/water mixture (1/1 v/v) at 95°C before loading on an 8% denaturing PAGE gel (7M urea) which was connected to a 20W power supply. After RP-HPLC purification and MS characterisation, ON2.8 was loaded on the same gel to indicate the electrophoretic mobility of the trimer dendron (Figure 2.9). The top band with intense signal was excised and incubated in water overnight to re-dissolve the oligo into solution. Analysis from UPLC-MS indicated a single species with a broad peak with a corresponding mass (43189.0) to ON2.15 (calc. mass = 43183.4). The broad shape of the peak can be attributed to its large molecular size because the molecule might not migrate along the porous stationary phase efficiently. This demonstrates the utility of preparative PAGE over HPLC in isolating oligo dendrons.

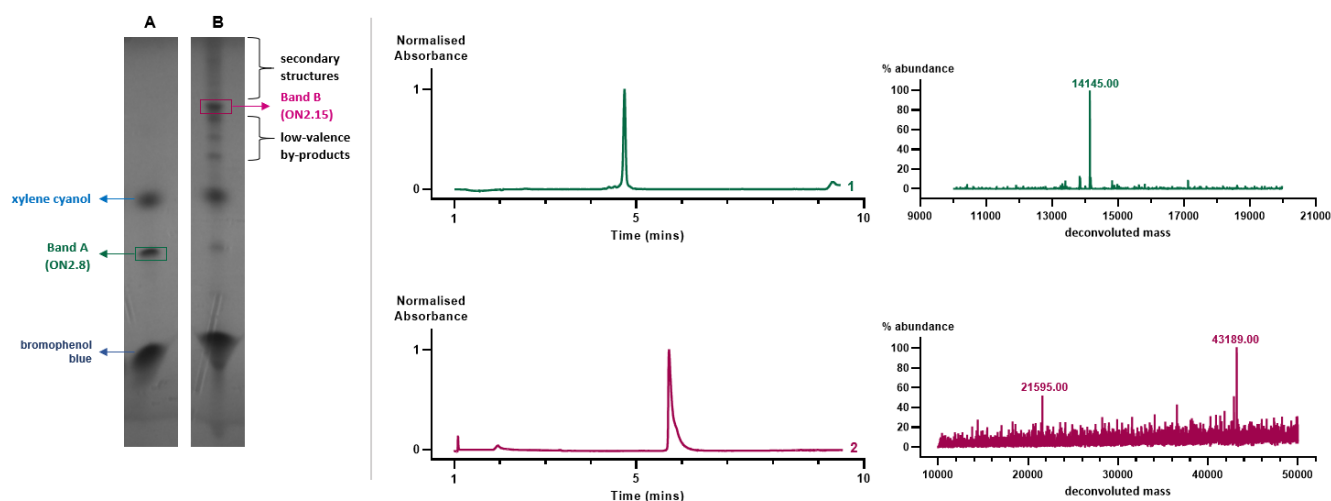


Figure 2.9 | UPLC-MS analysis of purified oligonucleotide dendrons

Left: the image of 8% PAGE analysis of (A) HPLC-purified ON2.8 and (B) crude mixture from the second click reaction (ON2.15) - excerpted from the same gel at identical height. Right: UPLC chromatograms and deconvoluted masses of the purified sample of (1) ON2.8 (calc. mass = 14143.0) and (2) ON2.15 (calc. mass = 43183.4).

Although the synthesis and isolation of nonameric oligo dendrons is successful, the gel image shows that the reaction generates a mixture of products which are ‘clicked’ to varying degrees. Interestingly, inefficiency in the second click reaction might not be sufficient to rationalise the formation of by-products as electrophoretic analysis indicates the presence of octamer, heptamer etc. which must be constituted by smaller, dimer or monomer (single strand) dendritic units. That means such by-products, with one or two free alkyne groups on the linker motif, can engage in CuAAC crosslinking in subsequent click reactions. Hence, it is envisaged that an efficient protocol to generate nonameric, or larger, oligo dendrons on solid support via CuAAC should eliminate the presence of any by-products from the first iterative step so that all available azide groups are connected to a trimer unit for propagation. In this study, we would like to move beyond the optimisation step to explore the properties on nucleic acid dendrons synthesised by CuAAC crosslinking on solid support, thus we will discuss the physical and biological properties of the dendrons obtained in the following section.

2.4 Properties of Click Dendrons

2.4.1 Duplex Hybridisation Study

The binding affinity of oligonucleotides to their complementary target sequence has important implications in various applications such as in self-assembling nanostructures or in antisense therapeutics.^{144,157} Conventional method to measure hybridisation strength of a duplex is UV-thermal melting, by which the duplex is gradually heated up from a low to high temperature that forces the hydrogen bonds between the duplex to break and divides the two strands. Since the magnitude of UV absorbance of two oligonucleotide strands is about 40% higher than that of their duplex, this value increases with the process of melting. Hence, a plot of absorbance vs. temperature can indicate the melting profile of an oligonucleotide duplex. The standard value to indicate the hybridisation stability is coined ‘melting temperature’ (T_m) which describes the temperature when half of duplex in solution have melted into their

constituent single stranded oligonucleotides, the point when the first derivative of the melting function reaches its maximum. The following section reports the UV-thermal melting data and compares the hybridisation strength of single-stranded DNA, trimer DNA dendron and nonamer DNA dendron against DNA and RNA targets (Figure 2.10 and Table 2.4). The concentration of oligonucleotide strands in each duplex sample is constant at 3 μ M regardless of the concentration of dendron molecules.

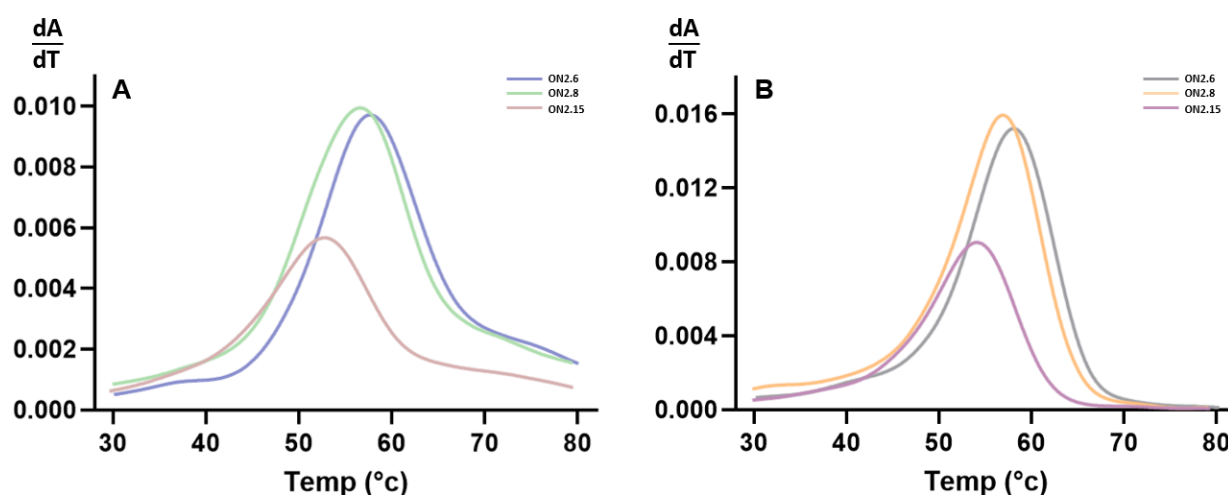


Figure 2.10 | First derivative of UV-thermal melting curves of oligonucleotide dendrons

The two plots describe the first derivative T_m curves of oligonucleotide dendrons against DNA (A) and RNA (B) targets. Measurement was taken with 3 μ M of oligonucleotides with equimolar DNA or RNA target in 10 mM phosphate, 200 mM NaCl buffer (pH 7.0) and from 30–80°C. T_m values were determined from an average of six ramps from a smoothed plot of dA/dT .

Entry	Sequence (5'-to-3')	DNA Target		RNA Target	
		T_m	ΔT_m	T_m	ΔT_m
ON2.6	HO-TCACTTCGCTTCAAC	57.5	---	57.8	---
ON2.8	L3OH-(Tr-TCACTTCGCTTCAAC) ₃	56.8	-0.7	57.1	-0.7
ON2.15	L3OH-[Tr-L3-(Tr-TCACTTCGCTTCAAC) ₃] ₃	53.0	-4.5	54.4	-3.4

Table 2.4 | Melting temperatures of oligonucleotide dendrons against DNA/RNA targets

L3OH represents the linker motif. Tr represents triazole group. T_m values were obtained from the maxima of dA_{260} versus temperature using 3 μ M of each oligonucleotide in a 10 mM phosphate, 200 mM NaCl buffer (pH 7.0) from 30–80°C. ΔT_m is relative to the unmodified control ON2.6.

Previously, a trimeric RNA dendron synthesised from a pentaerythritol-based phosphoramidite (at 3'-end) was shown to possess similar hybridisation stability ($\Delta T_m < 0.4^\circ\text{C}$) versus RNA target compared to a single-stranded RNA of the same sequence.¹⁴³ This similarity in thermodynamic stability can be attributed to the sufficient spacing between oligonucleotides and from the branching, quaternary carbon core (at three PEG units) that contributes little impact by steric crowding or promoting intramolecular binding. In other words, each strand within the trimeric dendron can bind with its target independently with small thermodynamic cost. This phenomenon was analogously observed in the UV-thermal melting data of trimer ON2.8 which shows minor decrease in duplex stability ($\Delta T_m = -0.7^\circ\text{C}$) when binding to either DNA or RNA target, compared to its ssDNA counterpart. The optimality in spacer distance possibly plays a similar role here.

When the number of oligonucleotides is increased to nine in the nonameric dendron ON2.15, the duplex stability drops more significantly against both DNA and RNA targets, with a decrease in melting temperature by 4.5°C and 3.4°C , respectively. This is possibly due to increasing steric hindrance which impacts the accessibility of the target strand as the oligonucleotide cluster grows. Moreover, with a higher local concentration of oligonucleotides bound on one molecule, the propensity of forming intramolecular secondary structure increases. This intramolecular binding can compete with intermolecular binding with the target especially during the initial phase of hybridisation where H-bonds are formed between the first few nucleobase pairs. Finally, the relative instability of the fully hybridised ON2.15 can be entropically driven. Compared to one and three released in the smaller dendrons, nine molecules are freed in the melting of ON2.15, rendering it more thermodynamically favourable.

Interestingly, the local maxima of dA/dT in the melting of ON2.15 are almost 40% lower in magnitude than those in ON2.6 and ON2.8. This indicates the melting curve is less steep and the melting of ON2.15 happens within a broader range of temperature. One plausible

explanation is that the melting of ON2.15 entails multiple dissociation processes of oligos from the nonameric dendron and the first few are thermodynamically favoured due to electrostatic repulsion of negative charges when the number of duplexes increases within close proximity. In general, it can be concluded that target binding in trimeric dendrons is not significantly influenced whereas that in nonameric dendrons is destabilised by the dendritic structure.

2.4.2 Enzymatic Stability Assay

In biological environment, there are various digestive enzymes such as exonucleases and endonucleases, which can cleave phosphodiester bonds between nucleosides.¹⁵⁸ Unmodified oligonucleotides as natural substrate for these enzymes degrade rapidly. Terminal modifications on nucleic acid dendrons with a crosslinking moiety introduce structural changes to either terminus and thus may impart better stability against exonucleases that recognise the end of an oligo and cleave one nucleotide at a time. Moreover, a high density of negative charges in oligonucleotide clusters attract more charge-balancing counter-ions, which have been shown to retard nuclease activity.¹⁵⁹ As a result, we next sought to examine and compare the nuclease stability of the click dendrons ON2.8 (trimer) and ON2.15 (nonamer) with their single-stranded counterpart ON2.6. All oligonucleotides (50 μ M) were incubated in 50% v/v foetal bovine serum (FBS) and phosphate-buffered saline (PBS), a commonly used buffer which mimics a serum-rich environment, for designated durations (0, 1, 2, 4 and 8 hours) at which point the enzymatic reaction was aliquoted, quenched with equal volume of formamide and snap-frozen in liquid nitrogen until analysis. The samples prepared were loaded on PAGE gel at different percentages to visualise the digestion pattern over time (Figure 2.11).

When incubated with FBS, all three oligonucleotides showed a similar degradation pattern. Within 1 hour of incubation, both dendrons ON2.8 and ON2.15 migrated faster than the sample at 0 hour, indicating some nucleotides are cleaved from the strands. While a small

ratio of oligo remained undegraded in ON2.6, the band of intact oligonucleotide completely disappeared in ON2.8 and ON2.15 after 2 hours. After 8 hours of incubation, all unmodified single-stranded oligonucleotide ON2.6 was degraded and only a faint smear was observed from the dendron samples. The results demonstrate that the effects of 5'-end protection and high density of negative charges do not prevent oligonucleotides from degradation, possibly due to the open end at the 3'-termini of the oligonucleotide cluster which are still accessible to exonucleases. Also, unlike the dense packing of DNA on nanoparticles, the distribution of oligonucleotides on ON2.8 and ON2.15 might be spaced out by long distance via the flexible linker, which diminishes the protective effect of steric hindrance and counter-ion crowding. Recent report has shown that 3-branched and 9-branched unmodified DNA dendrons assembled by the 'top-down' strategy via Trebler phosphoramidite have more than double the half-life to their single-stranded counterpart in 20% FBS/PBS environment.¹⁴⁶ Such difference in observation could be due to the serum content, where the protection from the dendritic structure is not sufficient in a higher FBS (50%) environment.

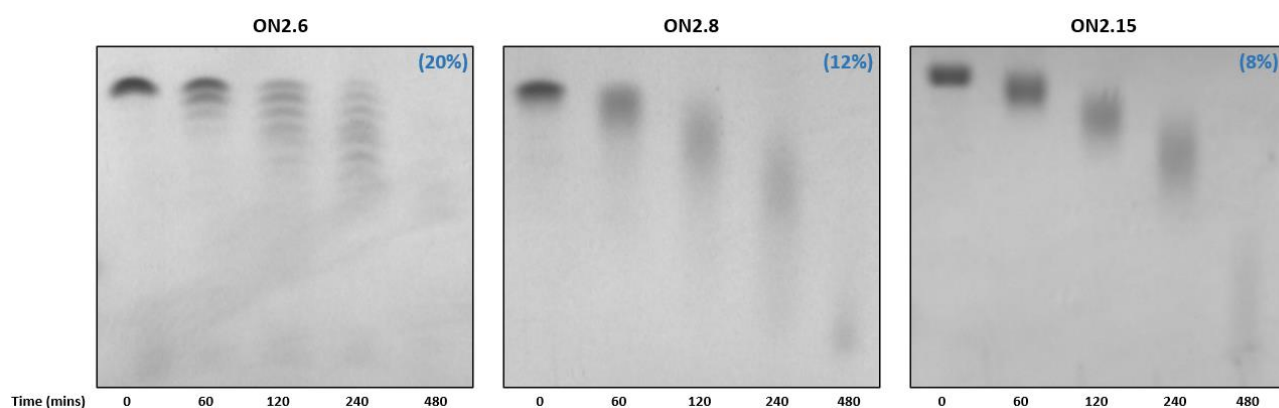


Figure 2.11 | Results of PAGE analysis from FBS stability assay

The three images show the electrophoretic migration of the aliquots of samples ON2.6, ON2.8 and ON2.15 incubated in 50% v/v FBS/PBS solution (pH 7.4) over 0, 60, 120, 240 and 480 minutes at 37°C. The degradation pattern of each oligonucleotide species is shown. The % of the gel is indicated on the top-left corner in each image.

Even though modification via 5'-crosslinking does not offer apparent advantage in preventing oligonucleotides from degradation in serum-rich milieu, it is reasonable to expect stability enhancement by combining with backbone or 3'-terminal modifications, given the versatility of the 'bottom-up' strategy which allows incorporation of modified monomers/dimers as in standard coupling.

2.4.3 RNase H Activation

RNase H-induced cleavage is the principal mechanism in various ASO drugs to silence mRNA and arrest downstream translation of faulty proteins (See Chapter 1). Although various modifications have been adopted to improve the nuclease resistance of ASOs, they can also reduce RNase H competence (i.e., the ability to be recognised by and activate RNase H by form a DNA:RNA heteroduplex) to different extent.²⁹ For example, 2'-modifications such as 2'-OMe and 2'-MOE are not recognised by RNase H whereas backbone modifications such as phosphorothioates only slow down its digestion. The effect of clustering oligonucleotides as a dendritic structure on RNase H activation, however, has not been widely studied. The following section examines the compatibility of DNA dendrons ON2.8 and ON2.15 with RNase H-mediated cleavage.

An RNA target ON2.16 was designed as a complementary sequence to ON2.6 that bears a fluorescein (FAM) modification at the 5'-end for visualisation on PAGE analysis. The RNA target (2.0 μ M, 5 equiv.) was incubated with ssDNA ON2.6 and DNA dendrons ON2.8 and ON2.15 (0.4 μ M of oligonucleotides) for designated durations (0, 1, 5, 15, 30 and 60 minutes) in the presence of *E. Coli* RNase H (NEB, 20 U/mL). At each time point, an aliquot of sample was taken, quenched with Na-EDTA solution (0.5 M) and formamide to stop enzymatic activity and snap-frozen in liquid nitrogen until the point of analysis. The samples prepared were loaded on 20% PAGE gel to visualise RNA digestion over time (Figure 2.12).

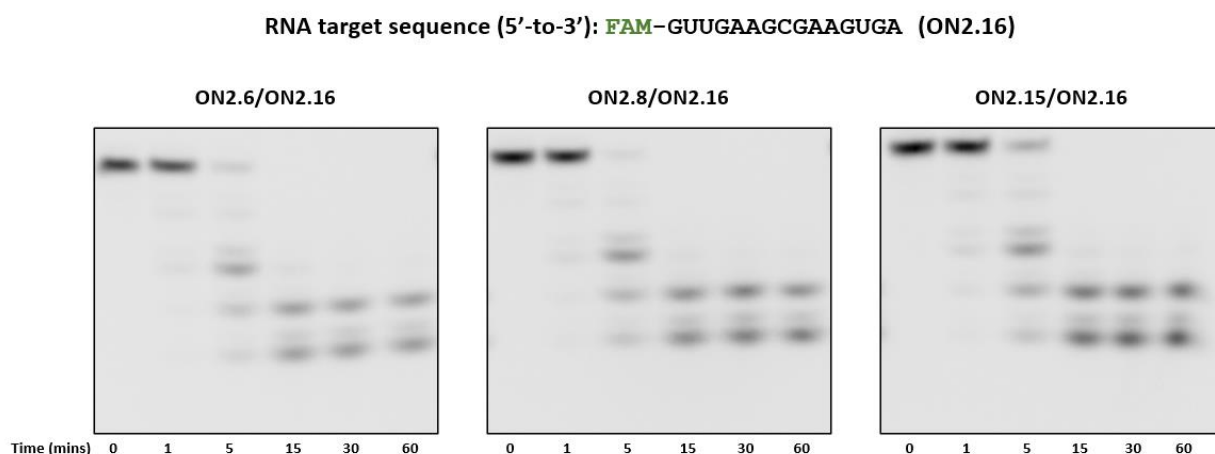


Figure 2.12 | Results of PAGE analysis from RNase H activation assay

RNA target sequence (5'-to-3') (ON2.16) is shown on the top. The three images show the electrophoretic migration of the aliquots of RNA sample ON2.16 (2.0 μ M) when incubated with ON2.6, ON2.8 and ON2.15 (0.4 μ M) together with *E. Coli* RNase H (20 U/mL) in RNase H reaction buffer® (50 mM Tris-HCl, 75 mM KCl, 3 mM MgCl₂, 10 mM DTT) over 0, 1, 5, 15, 30 and 60 mins at 37°C. The digestion pattern of ON2.16 is shown as fluorescent band (inverted) on the gel.

Single-stranded ON2.6 induced significant level of RNA cleavage after 5 minutes of incubation with RNase H and complete cleavage was observed in 15 minutes. Notably, RNase H prefers to cleave the phosphodiester bond downstream of a G-rich sequence,¹⁶⁰ which rationalises the discrete, shorter bands which correspond to the cleavage product when the dinucleotide sequence is cleaved downstream of guanosine. Compared to ssDNA, both trimeric (ON2.8) and nonameric (ON2.15) dendrons show a similar cleavage profile, with major RNA cleavage activity observed after 5 minutes and complete cleavage after 15 minutes. The results demonstrated that the crosslinking motif at the 5'-terminus does not inhibit the interaction between the DNA:RNA duplex and the recognition site of RNase H.

While terminal modification has been applied to ASO therapeutics in form of GalNAc conjugates that demonstrated RNase H competence in clinical models,¹⁶¹ it has not been reported that oligonucleotide dendrons can be used in the context of ASO therapeutics at the time of writing. The results confirmed that 5'-crosslinking does not interfere with RNase H-mediated RNA digestion and hence can potentially be used in ASO applications.

2.5 Conclusions

The use of dendritic oligonucleotides in material and therapeutic contexts is gaining popularity over the past few years. Although these molecules demonstrated desirable properties over their single-stranded counterparts, their syntheses are not efficient and wasteful in reagent input. In this chapter, we reported a new synthetic strategy to assemble oligonucleotide dendrons via a ‘bottom-up’ approach, by which full-length oligonucleotides are first synthesised by standard phosphoramidite coupling and then converted to 5'-azido intermediate for CuAAC crosslinking by a novel linker molecule. By this method, we are able to access trimeric DNA dendrons with up to 80% conversion and RNA at 72%. These products can be isolated by RP-HPLC and characterised on UPLC-MS. Furthermore, nonameric DNA dendrons are also accessible by repeating the iterative crosslinking process beyond trimer construction. Nevertheless, the product cannot be purified via HPLC systems possibly due to the presence of extensive secondary structures or its large molecular size that impacts its affinity to the stationary phase; it can be analysed and isolated via PAGE.

Previous reports have demonstrated the utility of dendritic oligos in various applications, such as an efficient siRNA loading platform via modified siRNA dendrons (duplex) and nonameric T₁₂ conjugate that elicits scavenger receptor A-mediated delivery of peptides.^{143,145} In this study, we investigated the properties of unmodified, mixed-sequence DNA dendrons with 3 and 9 valences. In terms of duplex stability, it was found that a trimeric dendron poses little effect on melting temperature against DNA and RNA targets whereas the nonamer has reduced duplex stability ($\Delta T_m = -4.5$ for DNA and -3.4 for RNA). This suggests the negative effect of higher valence on target binding which is likely due to steric crowding and charge repulsion; such destabilising effect should be taken into account when applying dendritic construct on antisense applications. Next, we found that both dendritic structures have similar stability profile in serum compared with single-stranded oligo, with degradation

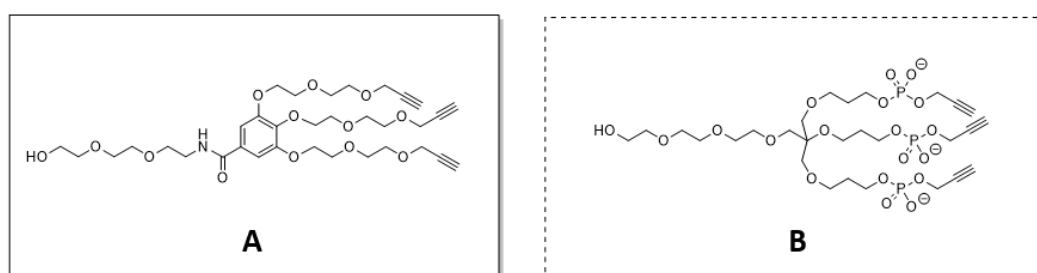
observed after 1 hour of incubation. Unlike previously reported RNA dendrons which are hybridised with terminally modified siRNA and protected from exonuclease action, the DNA dendrons used in this study have an accessible 3'-end which renders it susceptible to exonuclease digestion, in addition to endonucleases which can cleave the strand around the internal region. Finally, we demonstrated that DNA in dendritic structures can elicit RNase H-mediated cleavage when bound to its RNA target with no observable difference compared to ssDNA. It is possible to extend the use of dendritic construct with modified oligonucleotides for antisense applications.

2.6 Future Outlook

This study introduces a new methodology in accessing DNA and RNA dendrons. In the interest of time, we attempted to identify the best performing conditions and turned the focus to investigate the properties of the new molecules. From there on, a few aspects can be further optimised to improve their synthetic efficiency and physical properties. First, the efficiency of CuAAC crosslinking can be improved further in order to improve yield and access high-valence dendrons. The current protocol of click reaction converts at best 80% of oligonucleotides into trimeric dendrons on solid support, with the remaining as dimeric and linker-conjugated by-products. Such level of impurity is acceptable in trimer synthesis as the impurities can be removed quantitatively by HPLC. To carry out higher-order crosslinking, however, the presence of by-products with free alkyne can compete with free linkers for azido groups and thus drastically reduce the yield of subsequent click reactions and produce by-products that cannot be removed by chromatography. We envisage a few changes of the linker design and the reaction protocol might help overcome these issues.

First, it is possible to titrate the minimum loading of the linker that provides quantitative conversion of all oligonucleotides on resin. This is to reduce the formation of by-product by decreasing the availability of linker in the solution phase. Second, a longer reaction time with

lower concentration of linker may reduce the tendency of by-product formation because the resin-bound oligos can only react slowly in solution phase but are driven to react with covalently bound linker by proximity. Third, since the linker contains a relatively hydrophobic aromatic core, the retention time for any conjugated oligonucleotides is shifted right significantly in RP-HPLC systems. This makes the separation between impurities and products difficult. It is possible to improve its hydrophilic properties by replacing the phenolic core with pentaerythritol-based ones and introducing phosphate groups on the spacer units (Scheme 2.7). Such modifications may also alter the properties of the dendron as the distance between and packing of oligonucleotides change. In which case, a hydrophobic core tends to aggregate in aqueous solution while negative-charged spacers can separate strands further by charge-repulsion. This enhanced spacing might compensate for the loss in duplex stability due to steric crowding.



Scheme 2.7 | Molecular design of a hydrophilic linker

A: L3OH, the linker molecule used in this study. B: design of a hydrophilic linker with a pentaerythritol core and phosphate spacers

In addition to synthetic optimisation, improvements in dendron properties are worth further investigation. In the current construct, protection of oligonucleotides from nucleases has not been as pronounced as we originally believed. We anticipate the incorporation of backbone modifications (such as triazole, amide) at the 3'-end of the oligos can improve their nuclease resistance significantly as they are not substrate for exonucleases and hence can inhibit their activity. Thanks to their large molecular size (> 50 kDa), oligo dendrons are not

readily cleared by the renal system. An improvement in serum stability can increase their utility in therapeutic or drug delivery platform when systemically administered. However, their use in antisense applications might be limited by the fact that their target-binding is weaker than single-stranded oligos. To address this, a cleavable linker can be introduced between the 5'-nucleoside and the crosslinking unit for intracellular release of oligos. In this case, the therapeutic oligos can be protected in circulation and be released for target binding after the dendritic cluster has been internalised. For example, a recently reported enzyme-sensitive dipeptide phosphoramidite by our group can be incorporated to the 5'-end of an oligonucleotide via standard coupling and then converted to azide at its hydroxy end for CuAAC crosslinking.¹⁶² This linkage can be recognised and cleaved by *Cathepsin B* in lysosomal compartments and hence release the oligos inside cells.

Chapter 3 – Hepatocyte-Specific Delivery of CRISPR/Cas9 system

3.1 Background

3.1.1 The Origin of CRISPR

In 1987, the sequencing data published by Ishino and co-workers revealed the presence of an unusual pattern of repeats, which are not arranged consecutively but interspaced by non-repeating sequences (named ‘spacers’), adjacent to the isozyme-converting alkaline phosphatase (*iap*) gene of *E. Coli*.⁴⁰ Not until early 2000s have more research papers published over the origin of these interspaced repeat sequences, which were linked to foreign genetic elements from extrachromosomal entities and bacteriophages, in bacterial and archaeal genomes.^{41–43} This unique genomic feature in microorganisms has been coined ‘Clustered Regularly Interspaced Short Palindromic Repeats’ (CRISPR) and has since sparked research interest in the fields of microbiology and genomics alike. Added to the puzzle was the finding of an array of CRISPR-associated (*cas*) genes which encode a family of helicase, exonuclease and other enzymes and are found exclusively adjacent to the CRISPR locus.^{163–165} Such evidence hinted at the idea that the entire CRISPR-*cas* locus might function in DNA metabolism and spacer acquisition.

Although various hypotheses arose at the time, an increasing volume of evidence emerged to attribute the functions of CRISPR to bacterial immunity. Apart from the sequencing data which indicates CRISPR spacers often originate from plasmid or phage-DNA, it was discovered that bacteria are less sensitive to phage infection as the number of spacers in their genome increases.⁴² Soon after, the process of spacer acquisition following phage challenge in *S. thermophilus* has been reported independently by three groups.^{166–168} Barrangou and co-workers have provided direct evidence on conferring phage-resistance to a bacterial strain by inserting a spacer with sequence derived from the challenging phage, and phage-sensitivity by deletion of the corresponding spacer.⁴⁵ This study established that CRISPR is functionally involved in bacterial immunity against phage infection.

3.1.2 Mechanism and Functions of CRISPR in Prokaryotes

The identification of CRISPR as an acquired immune system in prokaryotes has led scientists to further elucidate the mechanism of its action. To date, two classes and various subtypes of CRISPR-Cas systems have been identified.^{169,170} In this section, the class II type 2 CRISPR-Cas9 system which is relevant to this chapter is discussed. In the event of an infection, the invading viral DNA is first recognised by Cas1-Cas2 complex which processes and incorporates a piece of the sequence into the host's CRISPR array.^{171,172} The selection of the sequence, namely 'protospacer', requires a 'protospacer-adjacent motif' (PAM) which serves to distinguish between the acquired spacer (that does not have a PAM) and the foreign genetic material. The protospacer sequence is then integrated into the CRISPR array to yield a new spacer that is flanked by 20 - 50 bp direct repeats.^{163,173,174} This process is also referred to as 'adaptation', whereby the host can memorise a piece of genetic material from the invading species.

In the next step, various effectors of CRISPR-mediated immunity are encoded and processed. First, the CRISPR array is transcribed to form a non-coding transcript – namely 'precursor CRISPR RNA' (pre-crRNA) – and then processed by Cas effector proteins and host factors such as RNase III to generate mature crRNAs that contain a single spacer and at least one repeat.^{44,175,176} Meanwhile, trans-activating crRNA (tracrRNA), which carries an anti-repeat sequence that is complementary to the crRNA repeats, is transcribed in the vicinity of the *cas* operon. The two complementary sequences then bind to form a dsRNA structure that is bound and stabilised by Cas9 endonuclease and furthered processed by RNase III, resulting in the formation of a crRNA:tracrRNA:Cas9 ribonucleoprotein (RNP) complex.^{175,177} After the 5'-trimming of crRNA by an exonuclease, the surveillance RNP is now primed to recognise its cognate target adjacent to PAM by complementary base-pairing from its crRNA and cleave the target by its endonuclease activity (Figure 3.1).

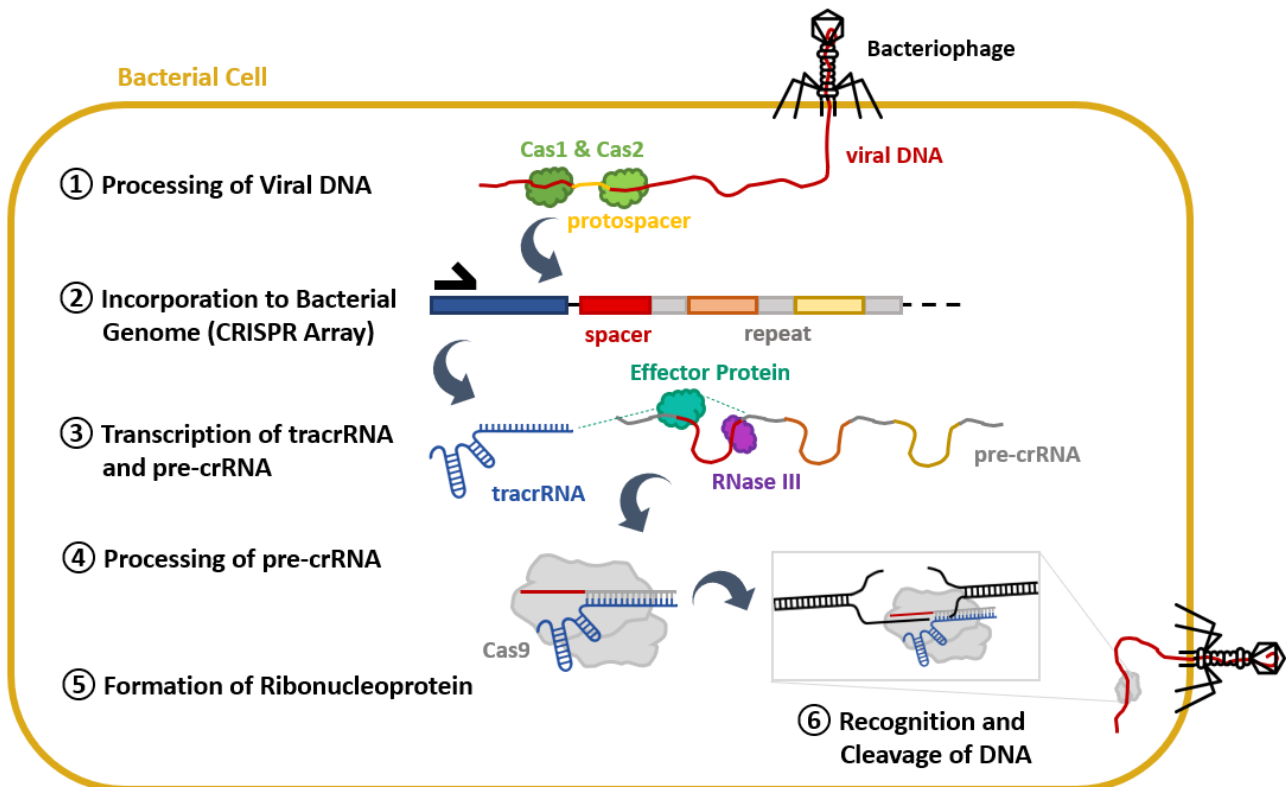


Figure 3.1 | Overview of CRISPR–Cas9 (Class II) adaptive immunity and interference

In a bacterial cell, (1) Viral DNA is recognised by Cas1–Cas2 complex which identifies the protospacer adjacent to a PAM sequence and processes it; (2) the complex then integrates a portion of the invading gene into the host's CRISPR array as an immunological memory; (3) transcription of the CRISPR–*cas* array produces pre-crRNA and tracrRNA which has a complementary anti-repeat to the repeat sequence in crRNA; (4) the former is further processed by specific Cas proteins or host factors into mature crRNA containing a unique spacer with identical sequence to the invading genetical elements; (5) assembly of interference Cas9 protein, encoded from *cas* operon, with crRNA and tracrRNA forms a ribonucleoprotein (RNP) surveillance complex; and (6) the complex scans invading nucleic acids and targets foreign complementary sequences to initiate strand degradation.

3.1.3 CRISPR-Cas9 in Mammalian Genome Editing

The precision and efficiency of CRISPR-mediated DNA cleavage has sparked immense interest in its utility on gene targeting and editing beyond prokaryotic systems. In 2012, Professors Jennifer Doudna and Emmanuelle Charpentier jointly published a study showing the possibility of targeted DNA cleavage by employing an artificially expressed Cas9 endonuclease and transcribed crRNA–tracrRNA (or a chimeric single-guide RNA constituted by fusing the 3'-end of crRNA to 5'-end of tracrRNA) *in vitro*.¹⁷⁸ The ground-breaking findings have demonstrated the potential of re-programming CRISPR RNA components to direct

cleavage on any dsDNA of interest, which has ushered in a new research field in CRISPR-mediated genome editing and has awarded the two Nobel Prize in Chemistry in 2020.

Indeed, the year succeeding the initial publication has seen a number of studies reporting CRISPR-mediated genome-editing in human cells. For instance, heterologous expression of human codon-optimized Cas9 and the required guide RNA (gRNA), which refers to the RNA component in DNA- or RNA-targeting ribonucleoproteins, in human kidney cancer HEK293T cells was shown to direct cleavage at almost any loci of interest in the vicinity of the PAM sequence by altering the 20-nt guide sequence within the gRNA.^{46,48,179} More remarkable is the possibility to cleave multiple targets simultaneously in *EMX1* locus reported by Zhang's group.⁴⁷ These cleavage events generated DSB in the genomic DNA target at the position of the -3 phosphate (the 5'-phosphate of the third nucleotide upstream of the PAM)¹⁶⁹ and are detected by the formation of insertions or deletions (indels) arising from the error-prone NHEJ repair pathway in the PCR amplicons (Figure 3.2).

In subsequent years, researchers have endeavoured to further control the editing output by promoting a template-mediated repair mechanism after DNA cleavage in order to precisely re-write genetic information in human cells. One notable example of which is 'prime editing' introduced by Liu's group.¹⁸⁰ In its design, a short RNA extension bearing the desired editing output was fused to the 3'-end of the gRNA to form a prime editing guide RNA (pegRNA). This sequence was used as a template for the reverse transcriptase (RT) fused to a re-engineered Cas9 nickase which introduces only a single strand break (SSB) on the target sequence. After the cleavage event, the cut strand binds to the primer-binding site on the pegRNA and acts as a primer for the RT subunit using the programmed RNA sequence template to synthesise a desired DNA extension. When reverse transcription completes, the flap (i.e., extended segment of the target strand) is processed and the break is repaired by ligase to incorporate the newly synthesised edit into the genome (Figure 3.2). Not only has this idea been proven highly

accurate and efficient (with up to 55% editing rate), it is also applicable in various human cell lines. This shows a great promise in correcting up to about 66,800 known pathogenic human genetic variants.

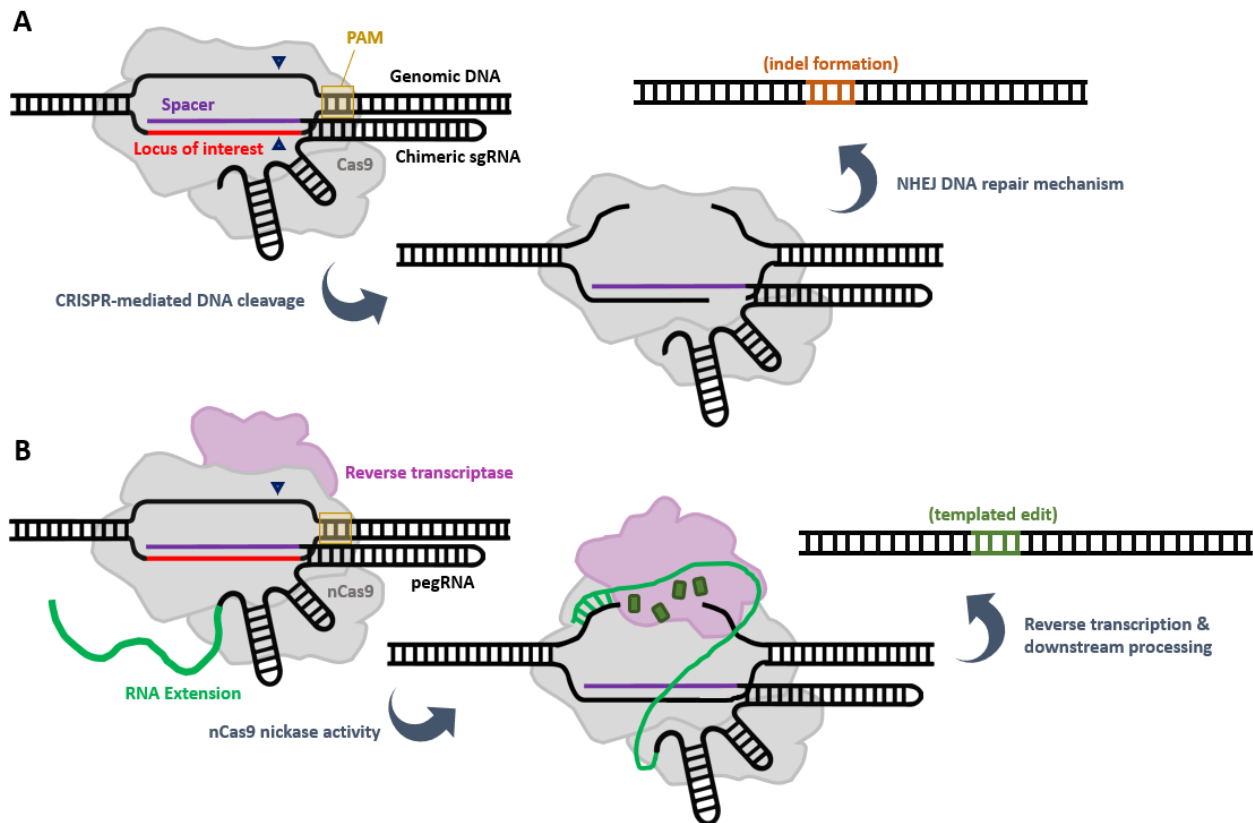


Figure 3.2 | Schematic representation of CRISPR-mediated genome editing mechanisms

A: Mechanism of CRISPR-mediated genome-editing at target locus. The genomic target is initially interrogated by Cas9 which binds to PAM and rapidly unwinds the upstream DNA duplex for target validation. The spacer sequence on the chimeric sgRNA then binds to adjacent DNA strand, if it is a *bona fide* target, and initiates conformational changes on Cas9 which becomes catalytically active and produces a DSB at the -3 phosphate (blue triangles).¹⁶⁹ The induced DNA damage then triggers NHEJ repair mechanism which involves a cascade of protein recruitment, DNA polymerisation and ligation, at a higher propensity than the templated HDR pathway, resulting in indel formation at the target site.¹⁸¹

B: Mechanism of prime editing. The Cas9 protein subunit is sequence-modified to replace the nuclease domains by a nickase domain (nCas9), and fused to a reverse transcriptase subunit. The sgRNA is extended by 'primer binding' and 'RT template' sequences (pegRNA). After target recognition, the nickase domain creates a SSB on the target's complement whereby the cut strand binds to the primer binding site and acts as a primer for templated RT activity. After downstream processing, the editing output bears an identical sequence to the programmed template.¹⁸⁰

3.1.4 Challenges in Intracellular Delivery of CRISPR

Although CRISPR technologies hold exciting promise in tackling genetic diseases and cancers, it remains a challenging task for *in vivo* research and clinical use due to the lack of methods for cell- and tissue-selective delivery.¹⁸² Current protocols for intracellular delivery can be efficient and target-specific in certain contexts but constraints and side-effects abound. This section provides an overview of existing methodologies and discusses their limitations.

In general, reprogrammed CRISPR components can be delivered to cells in form of a (1) plasmid which encodes Cas9 and gRNA; (2) mRNA which encodes Cas9 with a separate gRNA; or (3) pre-organised Cas9:gRNA ribonucleoprotein complex. Even though all three types of cargoes have been shown to effect CRISPR-mediated gene editing in human cells, plasmids are generally not considered in clinical applications as the amount of Cas9 expressed cannot be precisely controlled inside cells and overexpression could lead to off-target binding and DNA damage at undesired sites.¹⁸³ A few commonly employed methods can achieve high delivery efficiency to cultured cells *in vitro*. Microinjection, for example, is a technique to inject plasmids, mRNA or ribonucleoprotein to single cells via a fine needle and is regarded as a ‘gold standard’ of introducing CRISPR components to cells with up to 100% efficiency.^{184,185} To introduce these cargoes to multiple cells simultaneously, electroporation has also been commonly employed. By this method, pulsed high-voltage electrical currents are transiently applied to cells suspended in buffer, forcing open nanometer-sized pores on the cell membrane to allow passage of CRISPR RNP¹⁸⁶ and expression vectors¹⁸⁷ into the cells. Despite their high efficiencies, microinjection and electroporation are not suitable for *in vivo* applications as it is nigh on impossible to conduct injections to individual body cells or apply electrical current in biological medium inside an animal. Thus, they have been used exclusively in cell line models.

Achieving cell-specific delivery via systemic injection is a desirable objective. To this end, researchers have developed various viral vectors which can envelop CRISPR-expressing

vectors and deliver them to targeted cells. Adeno-associated virus (AAV), for example, is an excellent vehicle to delivery genetic materials because it is not known to cause or relate to any diseases in humans and it can target a multitude of human cell types.¹⁸⁸ However, its loading capacity is limited to 4.5–5 kb of genetic material, meaning that the packaging of Cas9-encoding genes is difficult. Apart from AAV, adenovirus (AV) and lentivirus (LV) with much higher loading capacity are also used in delivering vectors expressing CRISPR in human cells. But they are also unsuitable in clinical applications due to their propensity to elicit strong immunological response and integrate into the host genome.¹⁸³

Lipid-based nanoparticles (LNPs) present another attractive option to deliver genes or proteins into cells. These particles can form a complex with anionic species via its cationic lipid composition and form an envelope to protect the negatively charged nucleic acids or ribonucleoproteins *ex vivo* or *in vivo*. The neutralisation of charge promotes uptake of cargoes as large as the CRISPR RNP across the anionic cell membrane by endocytosis.¹⁸⁹ Such trafficking vehicles are superior to their virus counterparts in animal or human models as they tend not to trigger immunological response or undesired gene integration. However, the internalisation pathway via endocytosis generally brings about a lower transfection efficiency because the majority of cargoes, without an escape mechanism, are trapped inside endosomes and eventually degraded in lysosomes. Moreover, most LNP formulations are cytotoxic as their cationic and hydrophobic lipid motifs can interact with negatively charged organelles and disrupt cell membrane, respectively.¹⁹⁰ It has been shown that commercial LNP transfecting agents exhibit levels of cytotoxicity with positive correlation to their efficacies.¹⁹¹ In addition, LNPs are also not target-specific without modifications of recognition motifs. Hence, they are rarely used in systemic administration to target tissues and organs. An alternative methodology which combines endocytotic internalisation and endosomal release without membrane-disrupting reagents should mitigate such issues and promotes efficient uptake of CRISPR.

3.1.5 ASGPr-Mediated Delivery to Hepatocytes

Apart from exogenous vehicles which can package genes or proteins and translocate them across cell membrane, specific ligands can also initiate endocytosis of cargo molecules by the recognition of cell surface receptors. One example is asialoglycoprotein receptor (ASGPr) which is abundantly expressed on hepatocytes in mammals and primarily responsible for clearing glycoproteins from circulation.¹⁹² ASGPr has a high affinity towards terminal galactose (Gal) or N-acetylgalactosamine (GalNAc) residues, which can naturally be found on desialylated glycoproteins and trigger the formation of clathrin-coated vesicles on cell membrane to endocytose target molecules for degradation.¹⁹³ This mechanism is highly specific to ASGPr-expressing cells and thus has been harnessed as a method of intracellular delivery of GalNAc-conjugated antibodies¹⁹⁴ and therapeutic oligonucleotides²⁹ to hepatocytes.

In particular, tri-antennary GalNAc display is a widely used moiety in therapeutic formulations (Figure 3.3). ASGPr has trimeric sugar binding sites which are 15–25 Å apart from each other in a triangular spatial geometry. This allows for high binding affinities with trivalent ligands spaced by optimal distance.¹⁹³ Commercially available CPG enables installation of a tri-antennary GalNAc group at the 3'-end of oligonucleotides. The effect of such modification has been found to increase the potencies in ASO and siRNA designs in gene silencing of mouse hepatocytes *in vivo*, compared to unconjugated or bi-antennary GalNAc-conjugated constructs.^{195,196} In fact, there are already a few oligonucleotide drugs employing GalNAc conjugation as a delivery platform under development or already approved by the FDA, such as Givosiran and Inclisiran, highlighting the utility of such strategy.²⁹

Although ASGPr-mediated endocytosis presents an efficient pathway to deliver large biomolecules (i.e., nucleic acids and proteins alike) into hepatocytes, these molecules are prone to endosomal entrapment as their size and charge prevent passive diffusion across the endosomal membrane.¹⁹⁵ Eventually, maturing endosomes will fuse with lysosomal bodies

which contain digestive enzymes to degrade their content. Small antisense oligonucleotides often bear heavy modifications which render them resistant to enzymatic as well as higher membrane permeability via neutral backbone modifications. Hence, endosomal entrapment might not influence the effective uptake of these oligos into the cytoplasm for mRNA silencing while the slow release from endosomes can be desirable for steady silencing effects. Unmodified biomolecules such as CRISPR RNP, however, are prone to degradation during prolonged entrapment in endosome/lysosome compartments. Hence, efficient endosomal escape mechanism is essential for intracellular delivery of CRISPR via endocytotic pathways.

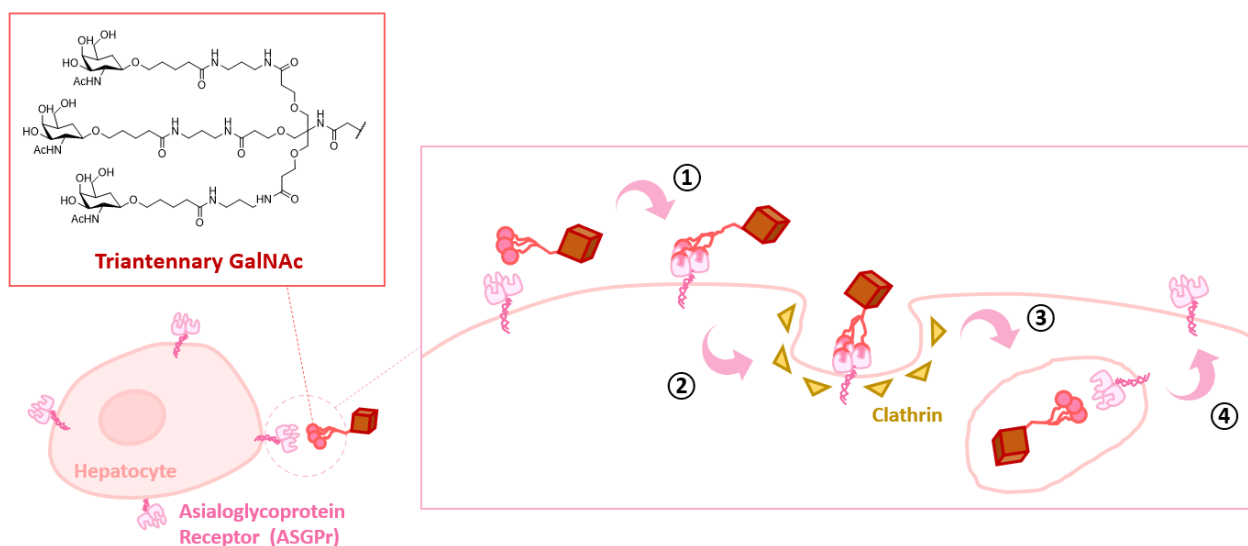


Figure 3.3 | Mechanism of ASGPr-mediated endocytosis

Cartoon representation of ASGPr-mediated endocytotic pathway of a GalNAc-conjugate on a hepatocyte. The cargo is indicated as a red cube and ASGPr-ligand is exemplified by commonly used triantennary GalNAc as red circles (structure shown on top left). (1) ASGPr recognises and binds to its ligands on the GalNAc-cargo conjugate; (2) the recognition triggers a cascade of downstream signalling and promotes clathrin nucleation around the receptor to form a clathrin-coated pit; (3) the pit buds into the cytoplasm to form a clathrin-coated vesicle which rapidly disassembles and leaves the cargo in an endosome. At acidic pH, the ligands dissociate from ASGPr and allows the release of cargo inside the endosome; (4) ASGPr is recycled to the cell surface.

Recently, Doudna and co-workers have reported a method to deliver CRISPR RNP into hepatocytes by conjugating Cas9 to novel ASGPr-ligands which are artificial sugar motifs with high binding affinity to ASGPr ($K_D \leq 100$ pM).¹⁸² The ligands were covalently linked to a

point-mutated Cas9 protein via its solvent-exposed cysteine residues by disulfide exchange chemistry. This conjugation was shown to induce receptor-mediated uptake of CRISPR RNP by HEPG2 cells (liver carcinoma) without any transfecting agents ('naked transfection') by confocal microscopy. However, T7E1 assays which analyse indel formation indicated only nominal gene editing activity after internalisation. Such discrepancy between uptake and function has been attributed to endosomal entrapment which can be particularly challenging for the release of the RNP-conjugate with a molecular weight > 170 kDa. To induce endosomal release, the group co-incubated the RNP with an endosomolytic peptide, ppTG21 which becomes cationic via its histidine residues¹⁹⁷ and membrane-disruptive at low pH (late endosomal environment) in HEPG2 cultures. This peptide has shown negligible impact on promoting the activity of unconjugated RNP but a significant increase in indel formation (4.8%) with ASGPr-ligands-conjugate which can trigger receptor-mediated endocytosis. Such results suggest endosomolytic agents be necessary in intracellular trafficking of endocytosed RNP into the cytoplasm (and eventually the nucleus) to effect its function.

3.1.6 T7E1 Mismatch Detection Assay

To evaluate CRISPR gene-editing activity in cells, T7E1 mismatch detection assay is one of the most commonly used techniques. This section introduces the mechanism of this assay and its expected readout, which will be used to assess the gene-editing efficiencies of the CRISPR constructs discussed in this chapter.

T7 endonuclease I (T7E1) recognises and cleaves dsDNA at a site where base-pair mismatch exists. In order to identify CRISPR-induced indel formation at the target locus, cells transfected with CRISPR components are first lysed and the genomic DNA is extracted. Then, the locus of interest is amplified by PCR to generate a large number of copies. If there is gene editing activity at the target site, indels will be introduced by NHEJ repair mechanism after

Cas9 cleavage activity and this edit will be amplified altogether. Theoretically, the number of copies containing indels corresponds to the number of cells that undergo successful gene-editing events and the former should be a quantitative representation of the latter within the population via an unbiased PCR amplification. The resulting mixture of original and edited amplicons are then melted and re-annealed at different temperatures to generate DNA heteroduplexes (i.e., duplexes formed by ssDNA amplified from different origins) which contain base-pair mismatches if the original anneals with the edited strand. This re-annealed mixture is then subjected to T7E1 digestion which cleaves the dsDNA at the mismatched site and results in the formation of shorter fragments. The quantity of these fragments relative to the intact amplicons is hence indicative of CRISPR gene-editing efficiency and can be visualised on gel (Figure 3.4).

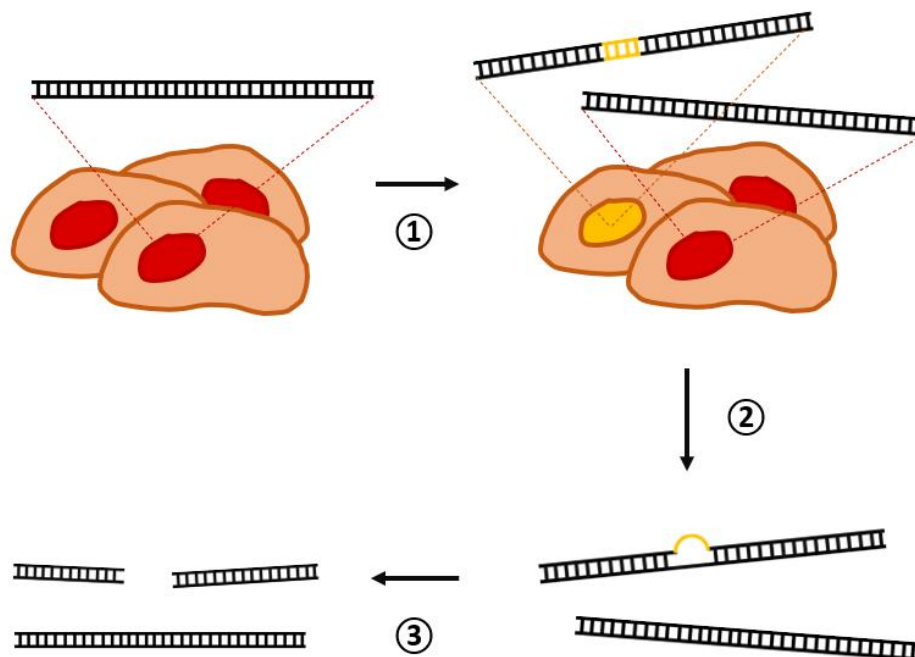


Figure 3.4 | Schematic representation of T7E1 mismatch detection assay

Three steps of T7E1 mismatch detection assay to evaluate indel formation: (1) cells are subjected to CRISPR transfection and successful cleavage at the locus of interest will result in indel formation (yellow segment); (2) genomic DNA of the population is extracted and amplified by PCR. Amplicons are denatured and re-annealed to form DNA heteroduplexes, which contain base-pair mismatch (yellow bulge) if it is hybridised with both edited and original strands; (3) PCR amplicons are incubated with T7E1 enzyme which induces cleavage at the site of mismatch, resulting in shorter fragments relative to the parental amplicon. The digestion pattern can be visualised and quantified by gel electrophoresis.

Notably, although T7E1 assay is a convenient tool to assess CRISPR activity based on indel frequencies, the results can be error-prone.¹⁹⁸ For instance, single nucleotide dimorphism is poorly recognised by the enzyme so that indel of one nucleotide cannot be accurately identified from the digestion pattern. Also, at low indel frequencies, the parental (i.e., uncut) band intensity can be equal to the negative control and cut products may be barely visible above background, resulting in biased comparison. In general, T7E1 assays tend to underestimate CRISPR activity when indel frequencies are only nominal or exceeding 30% and hence it could only serve as a semi-quantitative indicator.

3.2 Aims and Objectives

ASGPr-mediated uptake is an efficient and specific platform for intracellular delivery of CRISPR RNP to hepatocytes without the use of cytotoxic transfecting agents. The strategy presents great promise for liver-specific *in vivo* delivery of CRISPR reagents. While Cas9-ASGPr-ligands conjugates have exhibited good delivery efficacy, the modification and purification of protein conjugates can be laborious. In this study, we attempt to focus on gRNA modification and aim to develop a synthetic strategy to prepare a gRNA-ASGPr-ligands conjugate that exhibits good delivery efficacy of CRISPR RNP to hepatocytes. Covalent linkage with ASGPr-ligands is expected to direct ASGPr-mediated endocytosis as the gRNA complexes with the effector Cas9 protein. The chemistry involved should be simple and efficient whereas the product should be easy to purify and apply in downstream applications. Then, we aim to evaluate the efficacy of this delivery strategy by assessing gene-editing efficiency of CRISPR-Cas9 at the target locus via ‘naked transfection’ of the gRNA-ASGPr-ligands:Cas9 RNP complex, compared to commercially available LNP transfection formulations.

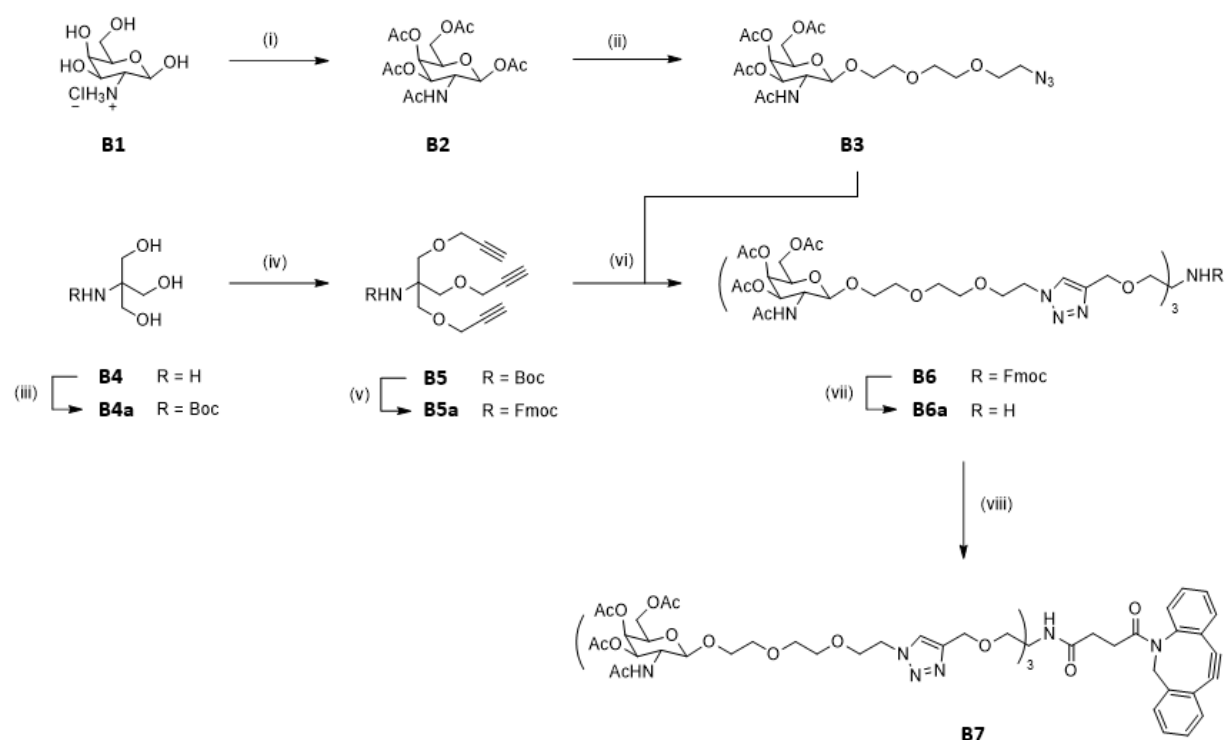
3.3 Synthesis of GalNAc-gRNA Conjugates via Click Chemistry

3.3.1 Design and Synthesis of Triantennary GalNAc Cluster

As a natural ligand to ASGPr, GalNAc residues have been extensively used in decorating biomolecules for hepatocyte-specific delivery. We commenced the study by investigating the possibility to incorporate GalNAc in its triantennary form, which has the highest binding affinity to ASGPr, to gRNA. Conjugation chemistry of RNA is more intricate than with DNA since it is susceptible to degradation in basic conditions as well as in the presence of Cu(II) which mediates oxidative damage. Hence, we attempted to access GalNAc-gRNA conjugates through strain-promoted azide-alkyne cycloaddition (SPAAC, a.k.a. copper-free click chemistry), which relies on an alkyne species that is heavily strained and energetically activated to react with an azide species without the need of a catalyst. For its incorporation, the GalNAc cluster should therefore contain dibenzocyclooctyne (DBCO), a commonly used strained alkynyl moiety, that is reactive toward an azido-modified gRNA. As for the molecular structure of this sugar cluster, we slightly modified the reported strategy of ‘clicking’ a tris-based core with acetylated GalNAc on PEG linker via CuAAC.¹⁸² In our design, the PEG spacer was shortened by one unit that reduces the distance between the sugar residue and the branching point below 19 Å as evidence suggested that longer linkers (> 19 Å) space out the sugars by too much and hinder their multivalent binding with ASGPr.¹⁹³

The synthesis of this triantennary GalNAc-DBCO derivative, **B7** is described in Scheme 3.1. We started from incorporating three PEG units at the functional end of the protected sugar with an azido end group which can couple with a Tris-based alkynyl core for trivalent display of GalNAc ligands. First, the hydroxy and amino groups on the precursor D-galactosamine hydrochloride, **B1** were globally protected by acetic anhydride in pyridine to yield the peracetylated intermediate **B2**. The reaction between **B2** and 2-(2-(2-azidoethoxy)ethoxy)ethanol in the presence of scandium (III) triflate as a Lewis acid catalyst

afforded **B3** with an azido-PEG linker incorporated via a β -glycosidic linkage. Notably, a recent study has indicated the anomeric carbon configuration does not impact ASGPr-binding significantly and in this study we employed the predominantly used β -configuration in pharmaceutical formulations.¹⁹⁹



Scheme 3.1 | Synthesis of triantennary GalNAc-DBCO

Reaction conditions are as follows: (i) Ac_2O , pyridine, r.t., 18h, 89%; (ii) $\text{HO}(\text{CH}_2)_2\text{O}(\text{CH}_2)_2\text{O}(\text{CH}_2)_2\text{N}_3$, $\text{Sc}(\text{OTf})_3$, anhydrous 1,2-DCE, reflux, 24 hours, 62%; (iii) Boc_2O , MeOH/t-BuOH, r.t., 18 hours, 85%; (iv) propargyl bromide, KOH, DMF, 40°C, 24 hours, 68%; (v) TFA/DCM, r.t., 2 hours, then Fmoc-OSu, aq. NaHCO_3 /THF, r.t., 18h, 75% over two steps; (vi) CuSO_4 , THPTA, sodium ascorbate, water/MeCN, 40°C, 48 hours, 52%; (vii) morpholine, DMF, r.t., 3.5 hours, quant.; (viii) DBCO-NHS ester, DIPEA, MeCN, r.t., 18 hours, 75%.

For the trivalent alkynyl motif, the precursor **B4** was first protected by tert-butyloxycarbonyl (Boc) protecting group by di-tert-butyl dicarbonate (Boc_2O) that makes the reactivity of the amino group. To the intermediate **B4a** were then inserted three alkynyl groups on its hydroxy ends via nucleophilic substitution with propargyl bromide as an electrophile and potassium hydroxide as a base to afford **B5** in 68% yield. Since Boc requires harsh deprotection

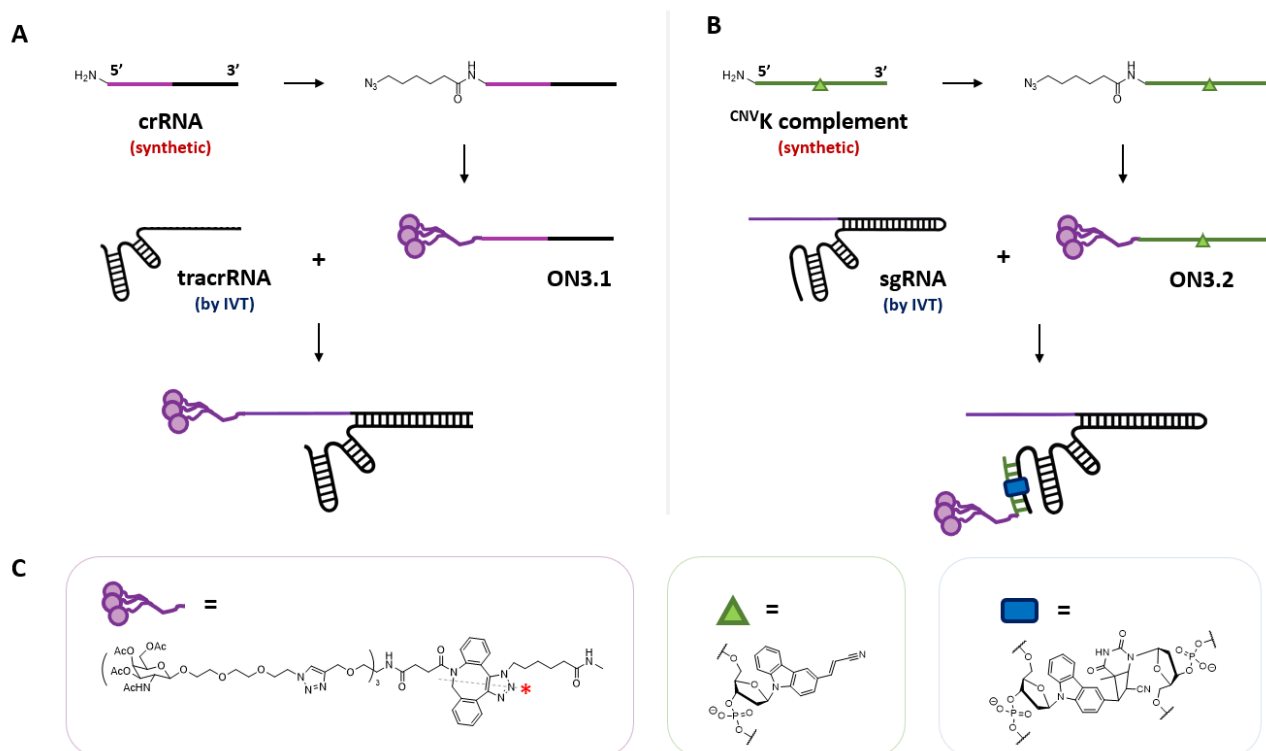
conditions (i.e., trifluoro acetic acid, TFA) which were found to cleave acetyl groups on the sugar, the next step involved the replacement of Boc protecting group on **B5** to fluorenylmethoxycarbonyl (Fmoc) group which can be deprotected under mild basic conditions. The resultant **B5a** precursor was reacted with **B3** via CuAAC in the presence of copper (I) ions and THPTA ligand to afford the tri-conjugated product **B6** in 52% yield. Then, Fmoc was quantitatively removed by the treatment with morpholine and the liberated amino group on **B6a** was allowed to react with DBCO-N-hydroxysuccinimidyl ester (DBCO-NHS ester) to generate the SPAAC-capable product **B7**. This product was further purified by HPLC before reactions with azido oligonucleotides.

3.3.2 Conjugation Strategies for GalNAc-gRNA Constructs via SPAAC

After obtaining **B7**, we proceeded to investigate possibilities to incorporate the triantennary GalNAc motif to gRNA. Although it is feasible to introduce an azide group on synthetic RNA via 5'-azide conversion or post-synthetic labelling on amino modifiers, the synthesis of modified sgRNA, whose length is often > 100 nt, is challenging and costly.²⁰⁰ Hence, we conceived two approaches to incorporate this motif to shorter synthetic oligos and then construct the full-length GalNAc-gRNA conjugate. In this study, we chose a spacer sequence which targets human genomic locus EMX1, a model to evaluate CRISPR activity.

The first strategy is a direct SPAAC coupling between a synthetic 3'-azido crRNA and **B7** and subsequent hybridisation with tracrRNA to form a functional gRNA conjugate which can be delivered to cells as a trimolecular complex with Cas9 (Scheme 3.2A). The second involves photo-crosslinking chemistry to incorporate a **B7**-conjugated oligonucleotide to the 3'-end of gRNA. In this case, a synthetic 2'-OMe oligo complementary to the 3'-end of gRNA modified with an internal 3-cyanovinylcarbazole (^{CNV}K) and 5'-azide is first conjugated to **B7** to impart the GalNAc functionality; then, the carbazole group is allowed to covalently link to

gRNA under UV irradiation (Scheme 3.2B).²⁰¹ These two approaches are synthetically viable as they require only short synthetic fragments which react with the GalNAc moiety and unmodified RNA components (i.e., tracrRNA and gRNA) which can be generated via *in vitro*



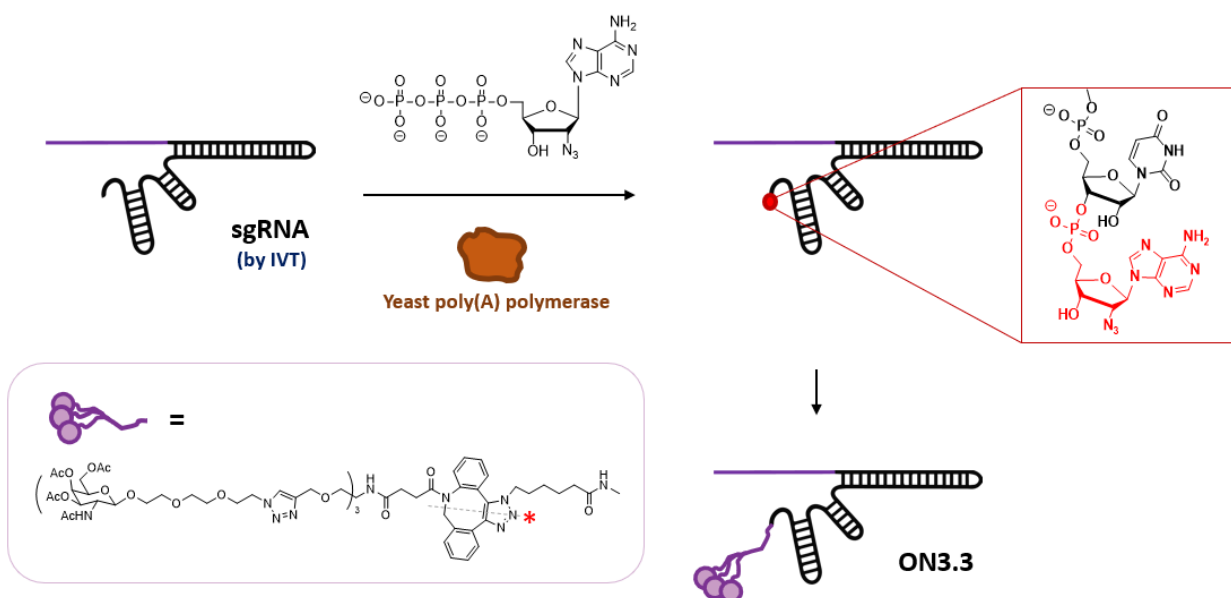
transcription by T7 polymerase.

Scheme 3.2 | Conjugation strategies to incorporate GalNAc on gRNA

(A) Synthetic 5'-amino-modified crRNA is first labelled with 6-azido-hexanoic NHS ester to form a 5'-azido intermediate which further reacts with B7 to yield GalNAc-conjugated ON3.1. This conjugated can bind with *in vitro* transcribed tracrRNA to form a function gRNA. (B) 5'-amino-2'-OMe oligo fragment bearing internal ^{CNV}K modifier undergoes the same azide labelling and conjugation with B7 to form ON3.2. It then binds to transcribed sgRNA and photo-crosslinks with the thymine base at -1 position on the complementary gRNA. (C) various modifications introduced to the oligonucleotides. Regio-isomers are indicated with the plane marked with asterisk by SPAAC coupling.

Preliminary UPLC-MS results indicated the feasibility of these conjugation strategies in forming GalNAc-conjugates (Figure 3.5). However, we realised two issues with these two approaches. First, various labelling and purification steps are time-consuming and they engendered significant product loss; and second, the synthesis of modified RNA is costly and

the coupling efficiency is relatively poor. Although we partially used RNA transcripts generated by enzymatic synthesis, it is overall challenging and cost-ineffective to scale up the synthesis for biological studies. Hence, we designed the third strategy to simplify the process by using enzymatic methods. In this approach, the full-length sgRNA is synthesised via *in vitro* transcription by T7 RNA polymerase. Then, the RNA transcript is labelled with an azide functionality at the 3'-end via the addition of 2'-azido-2'-deoxyadenosine-5'-triphosphate (2'-azido-2'-dATP) catalysed by yeast poly(A) polymerase (PAP). The introduction of a 2'-azido group at the 3'-terminus of the sgRNA enables SPAAC coupling with **B7** to yield a 3'-GalNAc-conjugated construct (Scheme 3.3).



Scheme 3.3 | Conjugation of GalNAc on sgRNA via enzymatic azide labelling

The sgRNA prepared by *in vitro* transcription is incubated with 2'-azido-2'-dATP in the presence of yeast PAP. The crude mixture containing sgRNA with 2'-azido group at the 3'-end is treated with B7 to afford 3'-GalNAc conjugate ON3.3. Regio-isomers are indicated with the plane marked with asterisk by SPAAC coupling.

After performing the labelling experiment, it was found that PAP activity with 2'-azido-dATP is inefficient and non-specific in the number of addition (Figure 3.5). The crude UPLC-MS chromatogram indicated that only 51% of the desired mono-conjugate was formed, with

20% of unreacted RNA ($R_T = 5.22$ min) and the remainder as bis- and tri-conjugated by-products ($R_T = 7.36$ and 7.88 min, respectively). Since the analogous 3'-azido-2',3'-ddATP is not a substrate for known poly(A) polymerases,¹¹⁷ the use of 2'-azido-2'-dATP enables multiple labelling as the 3'-hydroxy group remains available for further ATP addition via Poly(A) polymerisation, resulting in a mixture of SPAAC products and difficulty in purification. Also, the need to remove PAP enzyme after the reaction requires purification by affinity chromatography, on top of HPLC purification to isolate the product. Hence, this is also not an ideal method to prepare GalNAc-gRNA conjugates.

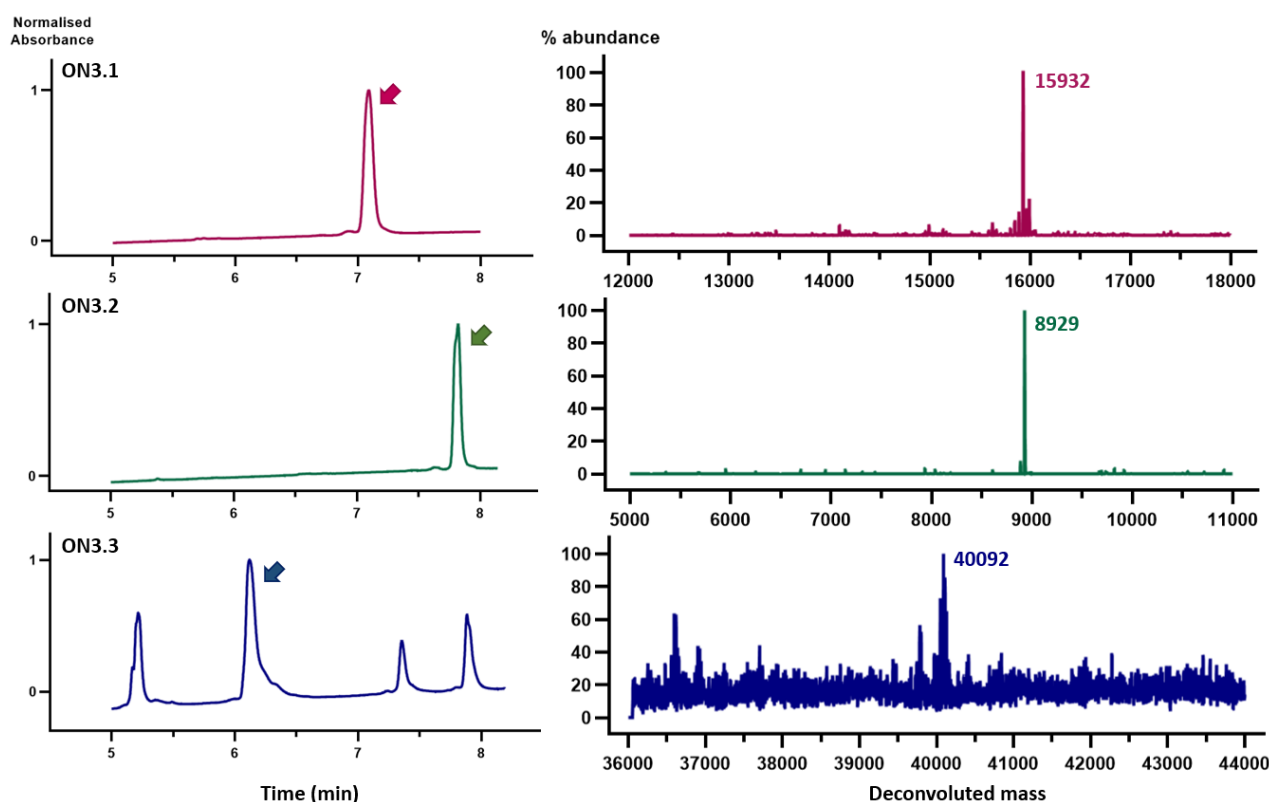


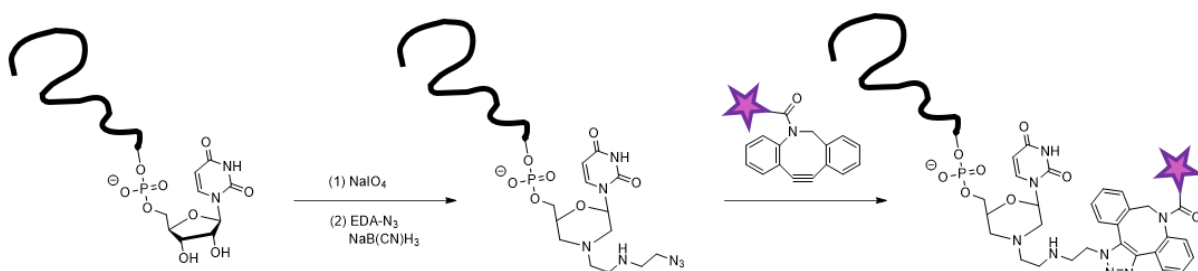
Figure 3.5 | UPLC-MS analysis of triantennary GalNAc-conjugated oligonucleotides

Left: UPLC-MS retention profiles of HPLC-purified ligation products ON3.1, ON3.2 and of crude mixture from PAP-catalysed azido-dATP addition on sgRNA (ON3.3 as major product). Major peak with mass corresponding to the conjugated product is indicated with an arrow. Right: deconvoluted mass from the major peak of ON3.1 (calc. mass = 15929.4), ON3.2 (calc. mass = 8927.1); ON3.3 (calc. mass = 40096.2, bis-Mg²⁺ adduct). Conditions for PAP-catalysed 2'-azido-2'-dATP addition: Yeast PAP (4400 U, MCLab, PAPY-30), 2'-azido-2'-dATP (1.04 mM, 65 eq.) in 1x reaction buffer. Conditions for GalNAc labelling: B7 (1.13 mM, 20 eq.), TEAA (100 mM, pH 7.0), 50% v/v DMSO.

3.4 Synthesis of GalNAc-gRNA Conjugates via Reductive Amination

3.4.1 3'-Azide Labelling on RNA by Reductive Amination

The installation of **B7** on synthetic gRNA fragments and enzymatically generated azido-sgRNA has been shown to be a feasible route to access GalNAc-conjugated gRNA. However, the synthetic routes demand multiple steps and do not guarantee good yields. We thus proposed a new reductive amination approach to introduce an azide functionality for SPAAC with the GalNAc-DBCO motif. Recent work by Mamot demonstrated the possibility to introduce 3'-modification to RNA molecules of length from 3 – 1000 nt via reductive amination with ethylenediamine (EDA) derivatives (See Chapter 1).¹³⁵ Essentially, the 2',3'-diol of an RNA is first oxidised with sodium periodate, a mild oxidant, and then subjected to reductive amination in the presence of the EDA motif and sodium cyanoborohydride, a mild reductant. After that, dehydration removes the hydroxy groups at the 2'- and 3'-positions and results in a morpholino-like linkage that connects the functional motif on EDA with the RNA at its 3'-terminus. The reported conversion of this two-step reaction with EDA-Cy3 derivative is 85% for RNA up to 276 nt long. With this encouragingly high coupling efficiency, we decided to apply this chemistry to incorporate terminal EDA-N₃ to our gRNA construct for click-enabled 3'-functionalisation (Scheme 3.4).



Scheme 3.4 | Schematic representation of RNA 3'-functionalisation via reductive amination and strain-promoted azide-alkyne cycloaddition

Chemical transformation of the 3'-end of an RNA molecule (black curve) is illustrated. The 2',3'-diol is first oxidised into 2',3'-dial which then undergoes reductive amination in the presence of EDA-N₃ derivative and sodium cyanoborohydride. The conjugation of azido handle enables SPAAC labelling of a functional molecule (violet star) which carries a DBCO-motif.

To commence, we prepared N-(2-azidoethyl)ethylenediamine (EDA-N₃) according to literature.¹³⁵ In order to assess the efficiency of this approach, a short, mixed-sequence RNA ON3.4 was prepared and subjected to NaIO₄ treatment. After 30 mins of oxidation, the RNA-dial intermediate was incubated with EDA-N₃ and NaB(CN)H₃ in KH₂PO₄ buffer (pH 6.0) for 2 hours. UPLC-MS analysis of the crude mixture indicated 82% of the starting material has been transformed into EDA-N₃ conjugate via a morpholino linkage (calc. mass = + 95.1) after this one-pot reaction and without work-up (Figure 3.6). The efficiency and simplicity of this reaction have led us to investigate the possibility to conduct EDA-N₃ ligation with longer, structurally complex sgRNA. We first prepared a sgRNA ON3.5 which targets human genomic locus *EMX1*, a model target for evaluating CRISPR activity. The RNA transcript after IVT retains 5'-triphosphate and hence we subjected it to shrimp alkaline phosphatase (rSAP) to remove these groups for mass analysis. Then, we carried out oxidation-reductive amination to conjugate an EDA-N₃ and subsequently coupled it with **B7** to generate the gRNA-GalNAc construct. From UPLC-MS analysis, we were not able to determine the conversion of EDA conjugation as ON3.5-EDA-N₃ has a retention time close to the parental ON3.5. Thus, the crude mixture was incubated with an excess of **B7** (20 eq.) to drive the formation of the SPAAC product which indicates the amount of azido derivative in the mixture. With identical reaction conditions for ON3.4, stepwise EDA-N₃ conjugation and SPAAC with **B7** afforded the desired GalNAc-conjugated gRNA (R_T = 6.17 min) at 57% conversion (Figure 3.7). The reasons for such lower conversion could be two-fold: (i) the reaction kinetics with the 3'-end of the long and structured gRNA is lower due to steric hindrance; and (ii) the SPAAC reaction is incomplete. Indeed, this strategy also requires a two-step reaction which complicates the workflow in addition to being inefficient. Therefore, we decided to modify the strategy and synthesise a triantennary GalNAc-EDA derivative which can be directly coupled on gRNA in order to simplify the procedures and minimise product loss in the following part of this study.

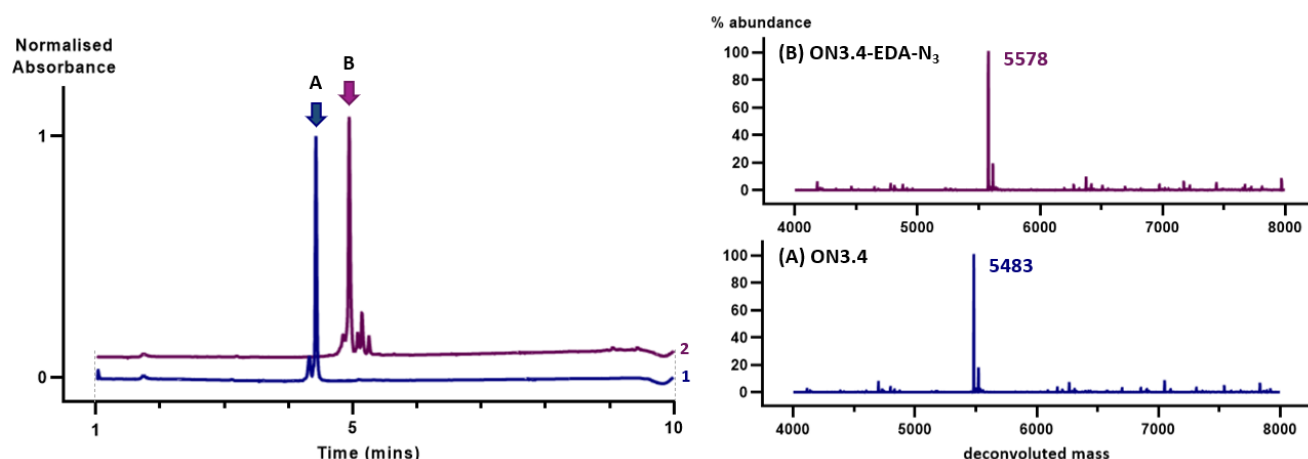


Figure 3.6 | UPLC-MS analysis of EDA-N₃-conjugated short RNA

Left: overlaid UPL chromatograms show the retention profiles of (1) short RNA sequence ON3.4; and (2) crude sample of ON3.4 after oxidation-reductive amination with EDA-N₃. Right: deconvoluted mass of the major peaks from corresponding chromatograms: (A) ON3.4 (calc. mass = 5481.3); (B) ON3.4-EDA-N₃ conjugate (calc. mass = 5576.4). Reaction condition: (1) NaIO₄ (1.5 mM), r.t., 30 mins, in darkness; (2) EDA-N₃ (1.0 mM, 10 eq.), NaB(CN)H₃ (20 mM), KH₂PO₄ (100 mM, pH 6.0), r.t., 2 h, in darkness.

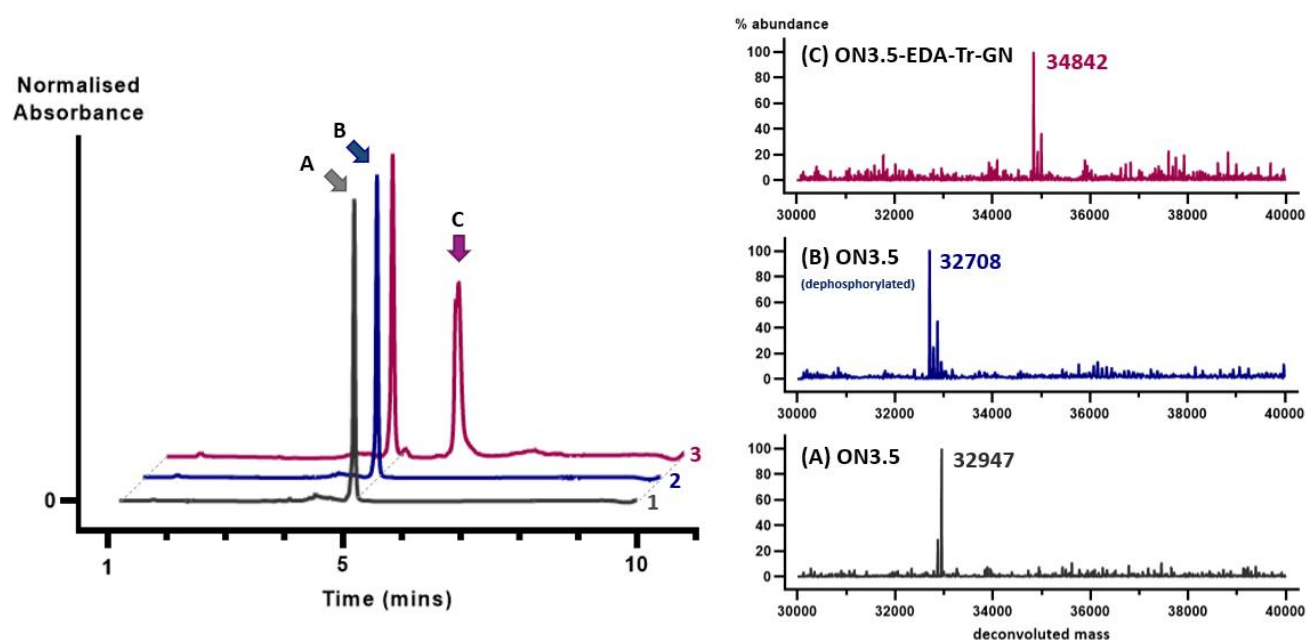


Figure 3.7 | UPLC-MS analysis of EDA-N₃ and DBCO-conjugation on gRNA

Left: Stacked UPL chromatograms show the retention profiles of (1) sgRNA ON3.5 after IVT; (2) crude sample of ON3.5 treated with rSAP; and (3) crude sample of dephosphorylated ON3.5 after oxidation-reductive amination with EDA-N₃ and SPAAC with B7. Right: deconvoluted mass of the major peaks from corresponding chromatograms: (A) ON3.5 (calc. mass = 32942.4); (B) 5'-dephosphorylated ON3.5 (calc. mass = 32706.5); (C) ON3.5-EDA-Tr-GN (calc. mass = 34837.7). Reaction condition: (1) NaIO₄ (1.5 mM), r.t., 30 minutes, in darkness; (2) EDA-N₃ (1.0 mM, 10.3 eq.), NaB(CN)H₃ (20 mM), KH₂PO₄ (100 mM, pH 6.0), r.t., 2 hours, in darkness; (3) B7 (3.2 mM, 20 eq.), TEAA (100 mM, pH 7.0), 50% v/v/ DMSO. Tr denotes triazole and GN denotes the GalNAc-DBCO (B7) motif.

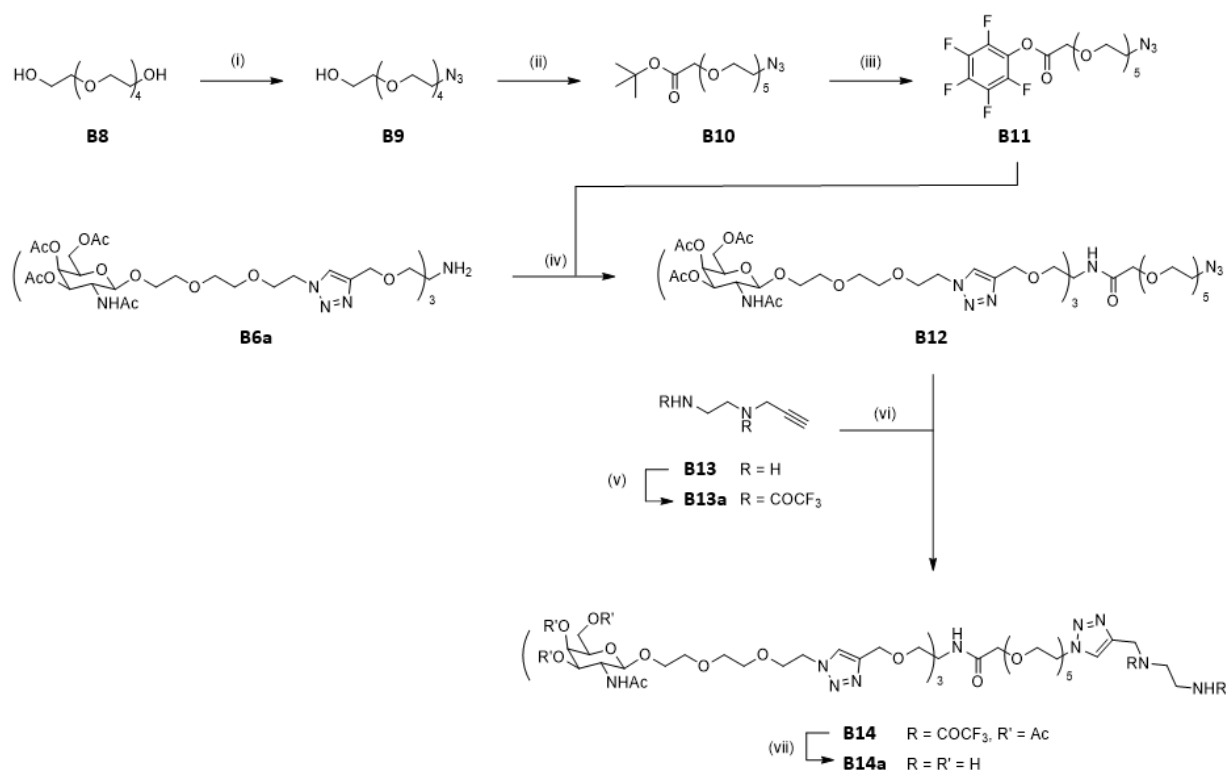
3.4.2 Synthesis of Triantennary GalNAc-ethylenediamine Conjugate

To functionalise the triantennary GalNAc cluster with an EDA moiety, the trimeric sugar motif was first modified with hexa-ethylene glycol spacer (PEG₆) appended with an azido end group. The azido cluster was then connected to an alkynyl-EDA adaptor via CuAAC (Scheme 3.5). The synthesis of the PEG spacer started from an azide conversion of pentaethylene glycol **B8** by diphenylphosphoryl azide (DPPA) and sodium azide, and then the addition of an acetate unit via S_N2 with tert-butyl bromoacetate and sodium hydride to afford a tert-butyl ester derivative **B10**. Deprotection by trifluoroacetic acid and esterification with pentafluorophenyl (PFP) trifluoroacetate yielded active ester **B11** which is reactive toward the amino GalNAc cluster **B6a**. The coupling between **B6a** and **B11** with DIPEA as a base furnished 56% yield of the azido-GalNAc intermediate **B12**. The alkynyl adaptor **B13**¹³⁵ was first protected by trifluoroacetic anhydride to mask the reactivity of the ethylenediamine motif, and then coupled with **B12** via CuAAC using CuSO₄ as copper source which was reduced *in situ* by sodium ascorbate. The resultant click product **B14** was subjected to treatment in methanolic ammonia (7.0 N) to globally remove all trifluoroacetate and acetate protecting groups and provide the final product **B14a** (also referred as EDA-GalNAc) in 87% yield after HPLC purification.

3.4.3 GalNAc-conjugation on RNA by Reductive Amination

The coupling reagent **B14a** provides a reactive ethylenediamine moiety to engage in reductive amination with sgRNA transcript directly, after *in vitro* transcription. After obtaining the compound, we attempted to test this new conjugation strategy with ON3.5 by replacing EDA-N₃ to EDA-GalNAc. Given the previous coupling results from EDA-N₃ coupling, we decided to double the concentration of EDA-GalNAc (20 eq.) in the hope of driving the reaction to completion, meanwhile other parameters were kept identical.

At a loading of 5 nmol, UPLC-MS analysis indicated that 76% of ON3.5 in the crude mixture was converted to ON3.5-EDA-GalNAc (i.e., the conjugated product with **B14a**), with only minor amount of by-products remained (Figure 3.8). The increment in reaction efficiency highlights the merits of a one-pot conjugation strategy which avoids product loss over multiple reactions, clean-up and transfer steps, especially due to the vulnerability of RNA towards degradation. It also demonstrates the simplicity of the workflow as compared to fully synthetic approach. In this case, long RNA (> 100 nt) can be enzymatically generated in large quantities and subjected to EDA conjugation in one step. Such outcome is difficult and costly to achieve by conventional SPS and chemical labelling which can consume excess reagents and require numerous purification steps.



Scheme 3.5 | Synthesis of triantennary GalNAc-EDA conjugate

Reaction conditions are as follows: (i) PO(OPh)₂N₃, DBU, NaN₃, THF, reflux, 18h, 43%; (ii) tert-butyl bromoacetate, NaH, dry THF, 0°C to reflux, 4 hours, 25%; (iii) TFA, dry DCM, r.t., 2 hours, then pentafluorophenyl trifluoroacetate, DIPEA, dry DCM, 2 hours, 46% over two steps; (iv) DIPEA, DMF, r.t., 24 hours, 56%; (v) trifluoroacetic anhydride, TEA, DCM, 0°C to r.t., 12 hours, 97%; (vi) CuSO₄, THPTA, sodium ascorbate, water/DMSO, r.t., 2 hours, 47%; (vii) ammonia/MeOH (7.0 N), r.t., 12 hours, 87%.

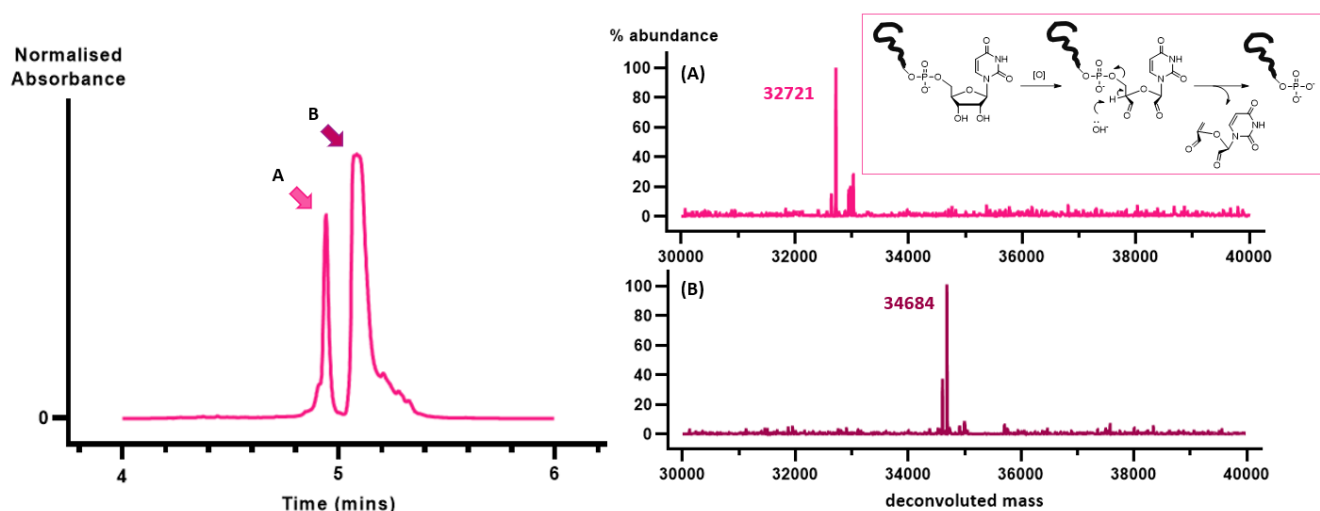


Figure 3.8 | UPLC-MS analysis of EDA-GalNAc conjugation on gRNA

Left: UPL chromatograms show the retention profile of sgRNA ON3.5 subjected to oxidation and reductive amination with B14a. Right: deconvoluted mass of the major peaks from the chromatogram: (A) by-product from β -elimination at the 3'-nucleoside of ON3.5 (calc. mass = 32717.8); (B) ON3.5-EDA-GalNAc (calc. mass = 34680.9). The pink box indicated a plausible mechanism for the formation of elimination product after initial oxidation. Reaction condition: (1) NaIO_4 (1.5 mM), r.t., 30 minutes, in darkness; (2) EDA-GalNAc (2.0 mM, 20 eq.), $\text{NaB}(\text{CN})\text{H}_3$ (20 mM), KH_2PO_4 (100 mM, pH 6.0), r.t., 2 hours, in darkness.

Notably, the crude trace did not indicate a significant presence of the starting material ON3.5 but that of a major by-product from the parental RNA. From mass analysis, we determined the corresponding species to be the 3'-phosphate (n-1) product formed via β -elimination. Due to the electron-withdrawing nature of the adjacent aldehyde group, the C4' proton becomes more acidic and is more readily removed by a base after initial oxidation to generate the elimination product (Figure 3.7). This decomposition pathway can compete with the reductive amination pathway in an aqueous environment where base is abundant and generate the 3'-phosphate by-product which is no longer reactive towards oxidation. Prolonging reaction duration is hence not conducive in driving the reaction to completion.

The desired ON3.5-EDA-GalNAc conjugate from this reaction was purified by RP-HPLC and ion-exchanged with NaCl solution on Amicon® centrifugal columns to remove triethylammonium counter-ions, and utilised in the next-step biological studies.

3.5 Biological Activities of GalNAc-gRNA Conjugates

3.5.1 In vitro Cleavage Assay

Although extensive studies have been carried out to identify possible internal modifications on gRNA, such as nucleobase or backbone substitutes, which can enhance gene editing efficiency and intracellular stability without compromising CRISPR activity, the effects of terminal modifications remained less well-known.^{202,203} Previous work has indicated that the introduction of oligonucleotide extensions such as poly(A) tails²⁰⁴ or G-quadruplexes²⁰⁵ at the 3'-end on gRNA do not impair CRISPR activity but even enhance gene editing efficiency in cells. It hints that steric hindrance might not play a significant role in the activity of RNP surveillance and cleavage. We were therefore intrigued to understand if 3'-conjugation is compatible with gRNA-directed RNP cleavage.

We decided to compare the cleavage efficiency of unmodified gRNA, ON3.5 and GalNAc conjugate, ON3.5-EDA-GalNAc against EMX1 target in vitro. In this assay, the two gRNA samples were separately incubated with Cas9 protein (Thermo Fisher, A36496) to form an RNP which was then added to a solution containing the target DNA *EMX1* produced from PCR. The resultant DNA fragments from the mixture were then analysed on Agilent Bioanalyzer 2100® which performs capillary electrophoresis on microfluidic chips to quantify DNA of various lengths by UV absorbance. In the reaction mixture, an excess (5 eq.) of Cas9:gRNA RNP was used against the DNA target (1 eq.) as Cas9 is a single turnover enzyme and this is to ensure DNA target can be recognised and cleaved by catalytically active RNP. It was found that the cleavage mediated by Cas9:ON3.5-EDA-GalNAc effected 96% cleavage on the target DNA within 1 minute of incubation, compared to 79% by Cas9:ON3.5 (unmodified). The results indicated not only that the 3'-conjugation of triantennary GalNAc does not compromise RNP nuclease activity, but also the performance is superior to unmodified gRNA prepared by *in vitro* transcription. A plausible explanation can be that Cas9

protein is sensitive to the residual salt from IVT and hence the RNP exhibited lower activity when incubated with the gRNA transcript. This property is less relevant after the RNP is internalised into cells and the salt content reaches an equilibrium with the cellular environment. From this experiment, we identified that RNP formed by ON3.5-EDA-GalNAc exhibited almost complete cleavage of target DNA (at 20% mol eq.) within 1 minute and the cleavage efficiency was better than gRNA transcript ON3.5. Such results are consistent with reported gRNA constructs with 3'-extensions which showed no negative impact on gene editing in cells.^{204,205}

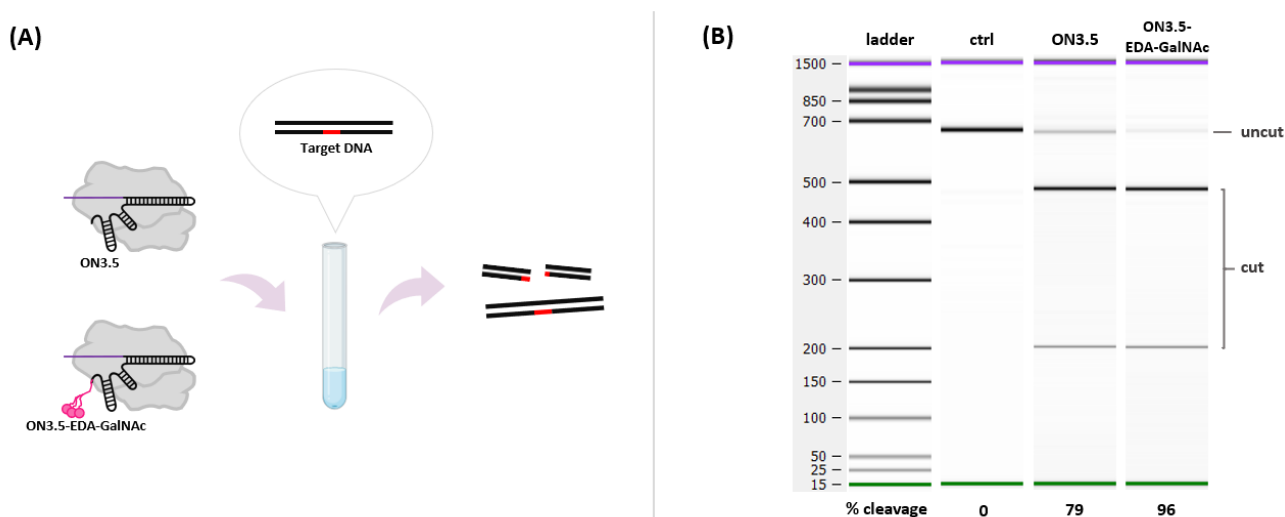


Figure 3.9 | In vitro cleavage assay of RNP with gRNA-GalNAc conjugate

(A) Schematic representation of *in vitro* cleavage assay. The RNP complexes formed by Cas9 protein and gRNA constructs (ON3.5 and ON3.5-EDA-GalNAc) are incubated with target DNA *EMX1* in reaction buffer. The resultant cleavage fragments are analysed by gel electrophoresis to determine the % cleavage efficiency. (B) Cleavage patterns of individual RNP constructs is visualised on the electropherogram and cleavage efficiency is quantified by the equation $f_{\text{cut}}/f_{\text{total}} \times 100$, where f stands for fraction, using Agilent Bioanalyzer 2100®. Results were calculated from three independent experiments ($n = 3$).

3.5.2 Delivery of RNP with gRNA-GalNAc Conjugate into Hepatocytes

After confirming the compatibility of ON3.5-EDA-GalNAc, we proceeded to investigate the efficiency of transfecting the GalNAc-bearing RNP into hepatocytes without using transfecting agents. It has been reported that endosomolytic peptide is essential for releasing RNP from the endosomes via receptor-mediated endocytosis. The peptide ppTG21 was found to be an effective agent in inducing endosomal release at low pH when co-incubated with RNP cargos.¹⁸² Hence, we included the use of this peptide in our cell study.

In this experiment, ASGPr-expressing HEPG2 cells were seeded in 24-well plates and transfection was conducted on the next day when the cells reached confluency. We transfected each well in triplicates with various concentrations of Cas9:ON3.5-EDA-GalNAc RNP together with ppTG21. Results were compared with cells incubated in only transfection medium (negative control) and with Cas9:ON3.5 and lipofectamine CRISPRMAX™ (positive control). In reported protocols, the liver cells were incubated with ASGPr-ligands-decorated RNP at 360 nM and ppTG21 at 10.4 μ M to effect observable level of indel formation (4.8%). We pre-incubated our GalNAc-gRNA construct with Cas9 protein in HEPES-NaOH buffer and transfected the cells with the RNP at concentrations ranging from 50 nM to 1.25 μ M and ppTG21 at 10 μ M. After 48 hours, the cells were harvested and lysed and the genomic DNA extracted. The target *EMX1* locus was amplified by PCR and subjected to T7E1 enzyme digestion to identify indel formation.

It was found that at all transfection concentrations no indel was observed in the cells incubated with Cas9:ON3.5-EDA-GalNAc, whereas lipofectamine-based transfection of unmodified RNP resulted in 17% of indel formation (Figure 3.10). The result indicated that even at 1.25 μ M, which is about 3.47 times higher than reported protocol with Cas9-ASGPr-ligands transfection, the RNP was not efficiently delivered to the nucleus to induce DNA cleavage at the target site. This condition was not mitigated by co-incubation with ppTG21

which has endosome-disruptive properties. This contrasts with conventional lipofectamine formula which promotes internalisation of RNP and CRISPR activity inside the nucleus. We surmised that it might be due to poor endosomal release of the RNP cargo after ASGPr-mediated endocytosis and the resulting degradation in lysosomes that abrogated CRISPR activity. Therefore, we conceived an alternative strategy to improve endosomal escape.

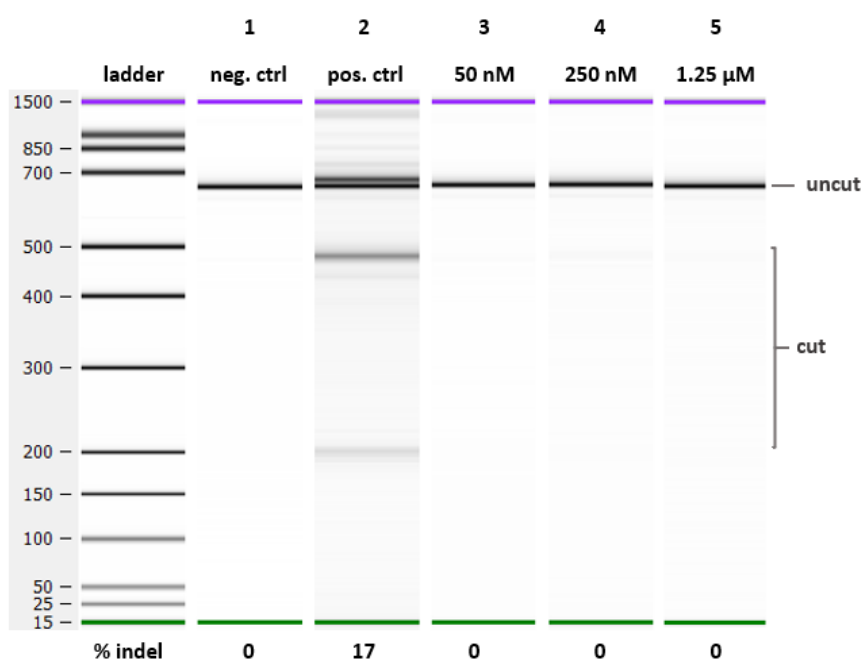


Figure 3.10 | Indel frequencies by GalNAc-conjugated RNP by naked transfection

Electropherogram of T7E1 digestion of EMX1 amplicon originated from HEPG2 cells transfected by various concentrations of RNP-GalNAc (Cas9:ON3.5-EDA-GalNAc). Lane 1: without RNP transfection (negative control); lane 2: RNP (Cas9:ON3.5) transfected by lipofectamine CRISPRMAX™ (positive control); lane 3-5: naked transfection of 50, 250 and 1250 nM of RNP-GalNAc. Indel frequency is quantified by the equation $(1 - (1 - f_{\text{cut}}/f_{\text{total}})^{0.5}) \times 100$, where f stands for fraction, using Agilent Bioanalyzer 2100®. Results were calculated from three independent experiments (n = 3).

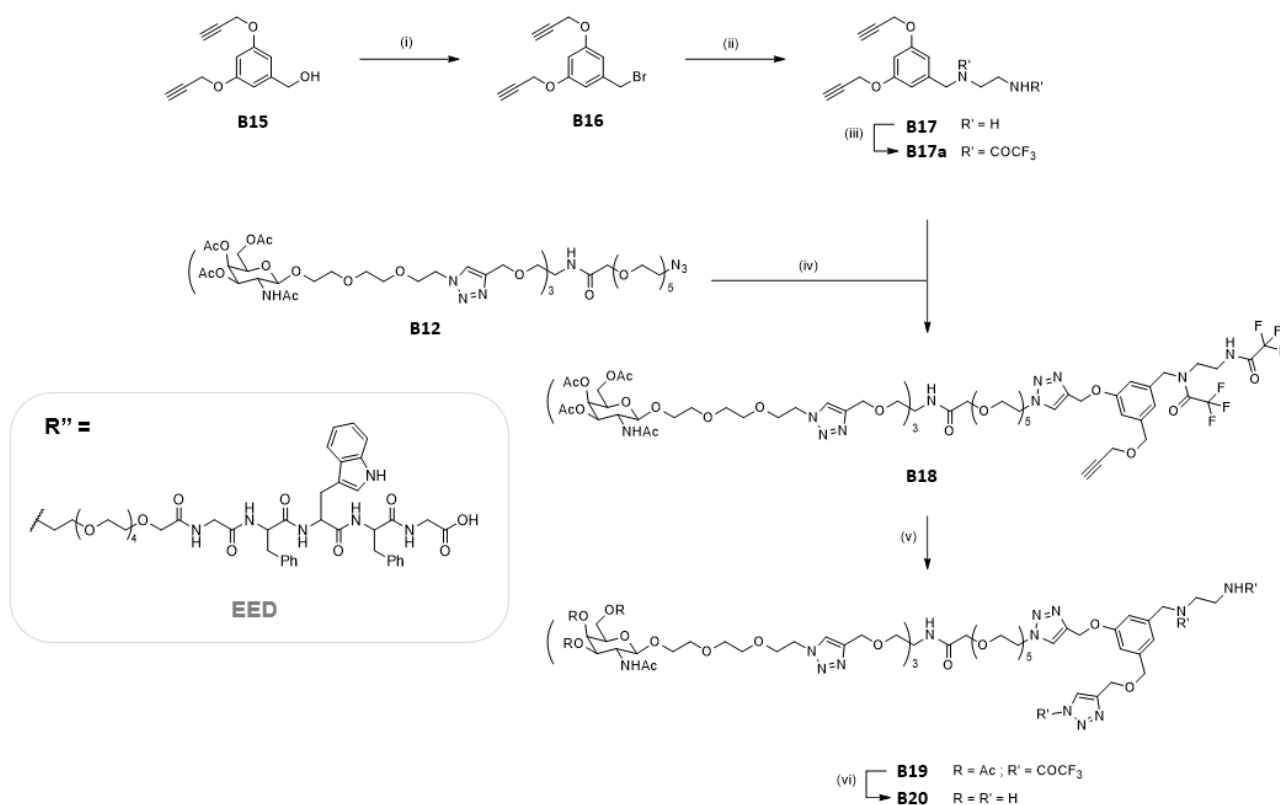
3.6 Dual Conjugation of gRNA for ASGPr-Mediated Uptake in Hepatocytes

3.6.1 Dual Conjugation with Endosomal Escape Domain

Co-incubating endosomolytic peptides with RNP for receptor-mediated endocytosis has been proven an effective strategy for endosomal release. However, the uptake of such peptides is possibly through pinocytosis when the ASGPr-ligands trigger vesicle formation and the endosome-disrupting effect can be subject to factors such as the efficiency of peptide uptake as well as charge interactions with the cargo. We decided to attempt another strategy to covalently link a short endosomal disruptive peptide to GalNAc-conjugated gRNA to induce endosomal escape. In the literature, this endosomal escape domain (EED) is a synthetic peptide fragment that has a sequence -GWFWG with mostly hydrophobic residues that induced multiple-fold higher endosomal release efficiency of covalently bound protein cargos internalised via receptor-mediated endocytosis.²⁰⁶ It was believed that the hydrophobic patch linked to the delivery domain inserts into the endosomal membrane after internalisation to disrupt the structure and promotes cargo release. Hence, we designed a synthetic scheme to incorporate this EED as a secondary functionality to the GalNAc-EDA molecule and attempted dual conjugation with our gRNA construct.

Appending two functional molecules on RNA can be challenging and might lead to significant degradation over the clean-up and purification steps. Hence, we looked upstream to modify the synthesis the EDA-motif which incorporates the two functional moieties (i.e., triantennary GalNAc and EED) and enables one-step conjugation with gRNA (Scheme 3.6). The synthesis began with bis-alkynyl precursor **B15** which underwent Appel reaction to yield bromo-substituted derivative **B16** that further reacted with ethylenediamine in large excess to afford bis-alkynyl EDA adaptor **B17**. This EDA derivative was protected by trifluoroacetic anhydride and then reacted with **B12** via CuAAC to provide mono-conjugated GalNAc-EDA

derivative **B18**, under stoichiometric control where 10 eq. of **B17a** was used. The available alkyne enables a second click reaction with EED-N₃ (Biomatik) to provide bifunctional intermediate **B19** which was then globally deprotected to afford the EDA-GalNAc-EED product, **B20** after HPLC purification.



Scheme 3.6 | Synthesis of EDA-GalNAc-EED conjugate

Reaction conditions are as follows: (i) CBr₄, PPh₃, DCM, 0°C to r.t., 1 hours, 97%; (ii) ethylenediamine, Cs₂CO₃, dry DCM, r.t., 24 hours, 61%; (iii) trifluoroacetic anhydride, DIPEA, dry DCM, 0°C to r.t., 4 hours, 89%; (iv) CuSO₄, THPTA, sodium ascorbate, water/DMSO, r.t., 2 hours, 80%; (v) R''-N₃, CuSO₄, THPTA, sodium ascorbate, water/DMSO, r.t., 2 hours, 66%; (vi) ammonia/MeOH (3.5 N), r.t., 18 hours, 98%. R'' = PEG₅-GFWGF (EED) where G = glycine, F = phenylalanine and W = tryptophan in the order from N- to C-termini.

Next, we attempted to perform reductive amination on gRNA ON3.5 with **B20** which was expected to be a less reactive substrate compared to EDA-N₃ or EDA-GalNAc due to steric hindrance. UPLC-MS analysis indicated that 69% of the starting material was converted to GalNAc-EED dual conjugate ($R_T = 5.79$ & 5.86 min, with identical mass) whereas 22% of oxidised RNA ($R_T = 5.09$ min) remained and the rest was degraded product which cannot be

identified by mass (Figure 3.11). In this experiment, we observed relatively low formation of the (*n*-1) RNA by-product (calc. mass + $[\text{Mg}]^{2+} = 32742$) from mass analysis at c.a. 40% abundance, compared to the conjugation with **B14-H** (EDA-GalNAc) (Figure 3.8A). We rationalised it as a trivial variable because the parental peak of ON3.5 after the oxidation reaction should in theory result in a mixture of oxidised (dial) and elimination (*n*-1) products, and the overall abundance of these species are much lower than that of the EDA-conjugated product that makes quantitative mass analysis less accurate. In general, the reagent **B20** provided low conversion than its EDA-GalNAc counterpart **B14-H** but it remains an efficient strategy to install two functionalities on long RNA simultaneously. We purified the gRNA-EDA-GalNAc-EED construct via RP-HPLC (MeCN/water, TEAA pH 7.0) and proceeded to biological study in the next part.

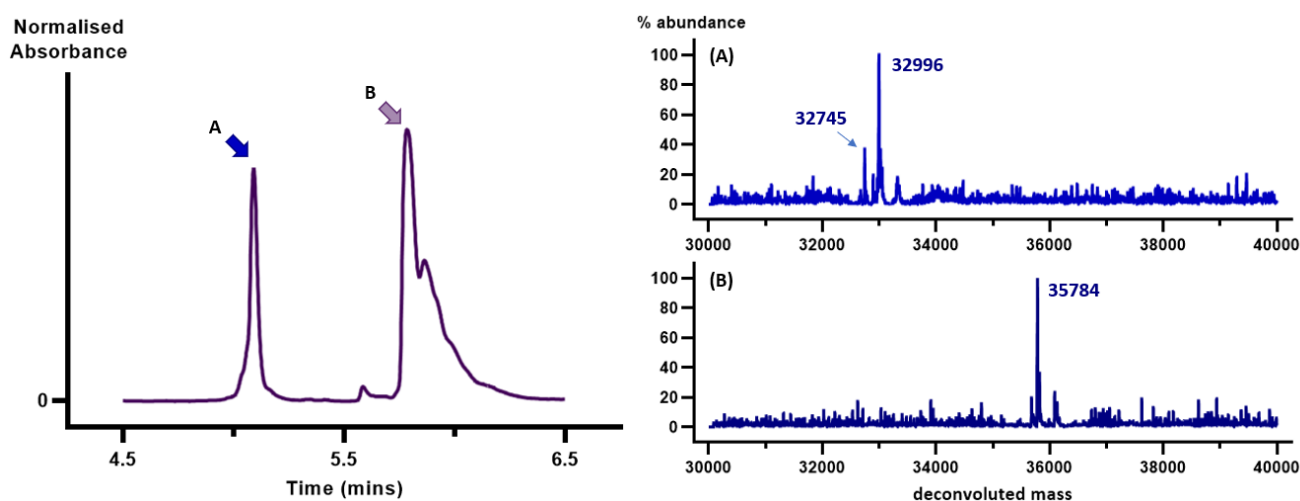


Figure 3.11 | UPLC-MS analysis of EDA-GalNAc-EED conjugation on gRNA

Left: UPL chromatograms show the retention profile of sgRNA ON3.5 subjected to oxidation and reductive amination with B20. Right: deconvoluted mass of the major peaks from the chromatogram: (A) oxidised ON3.5-dial (calc. mass + $2[\text{Mg}]^{2+} = 32992$); (B) ON3.5-EDA-GalNAc-EED (calc. mass + $[\text{Mg}]^{2+} = 35781$). Reaction condition: (1) NaIO_4 (1.5 mM), r.t., 30 minutes, in darkness; (2) EDA-GalNAc-EED (2.0 mM, 26.6 eq.), $\text{NaB}(\text{CN})\text{H}_3$ (20 mM), KH_2PO_4 (100 mM, pH 6.0), r.t., 2 hours, in darkness.

3.6.2 Biological Activities of gRNA-GalNAc-EED Conjugate

We first performed *in vitro* cleavage assay to assess the effect of a bulkier GalNAc-EED dual conjugate on the RNP cleavage activity. Compared to unmodified gRNA, the conjugate ON3.5-EDA-GalNAc-EED exhibited higher cleavage activity at 93% than unmodified ON3.5 at the same incubation conditions (Figure 3.12). The results are similar to the cleavage of ON3.5-EDA-GalNAc and indicated the 3'-modification exerts little impact on the RNP activity. Hence, we next attempted to transfect the EED-conjugated construct to HEPG2 cells to evaluate the delivery efficiency by T7E1 assay.

The hepatocytes were incubated with ON3.5-EDA-GalNAc-EED at 50, 250 and 1250 nM without external endosomolytic agents and indel formation as an indicator of RNP delivery was assessed by T7E1 digestion, compared to a control with conventional lipofectamine transfection. Results from gel electrophoresis indicated that the absence of indel event in every concentration of transfection (lanes 3-5) as only the parental amplicon was visualised, compared to the positive control (lane 2) where an indel frequency of 27% was identified (Figure 3.13). The results did not satisfy our initial expectation that GalNAc-conjugation could promote the internalisation and subsequent release of RNP with endosomolytic agents as no cleavage event was detected. There could be several reasons for this finding: (i) the RNP may not be internalised into intracellular vesicle or endosome via receptor-mediated endocytosis due to a lack of binding affinity; (ii) the RNP may be trapped in the endosome regardless of the addition of endosomolytic agents and subsequently undergo decomposition; (iii) the functionality of the RNP may be altered or abrogated. In order to better understand the limiting factor of this delivery approach, we next carried out an imaging study to investigate the localisation of GalNAc-conjugated RNP in HEPG2 culture by confocal microscopy.

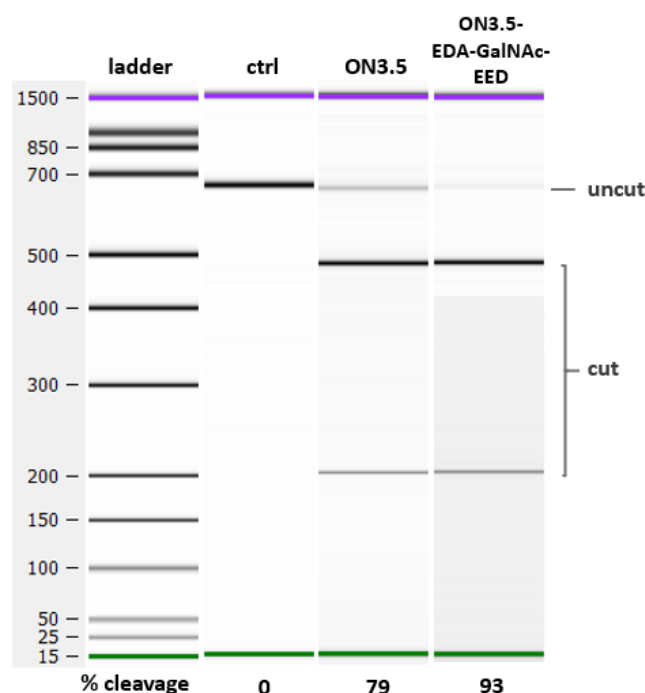


Figure 3.12 | In vitro cleavage assay of RNP with gRNA-GalNAc-EED conjugate

The cleavage pattern of RNP constructs with gRNA-GalNAc-EED modification is visualised and compared with unmodified gRNA on the electropherogram and cleavage efficiency is quantified by the equation $f_{\text{cut}}/f_{\text{total}} \times 100$, where f stands for fraction, using Agilent Bioanalyzer 2100®. Results were calculated from three independent experiments (n = 3).

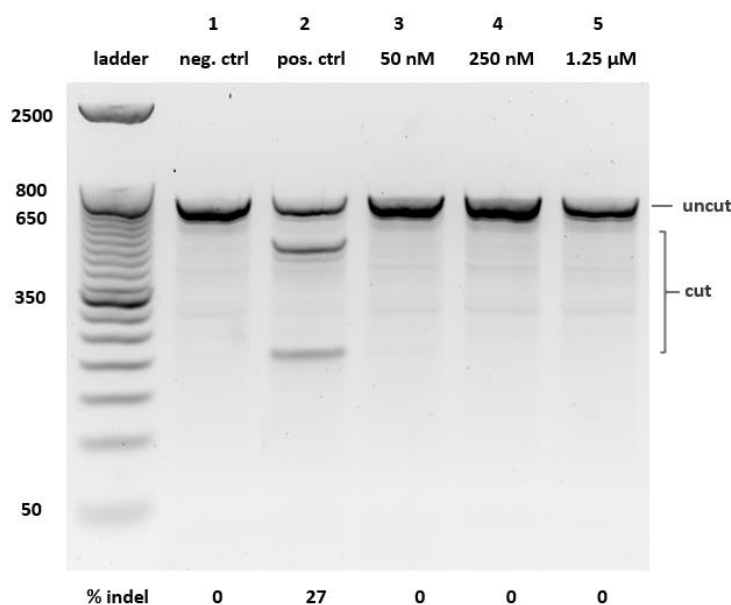


Figure 3.13 | Indel frequencies by GalNAc-conjugated RNP by naked transfection

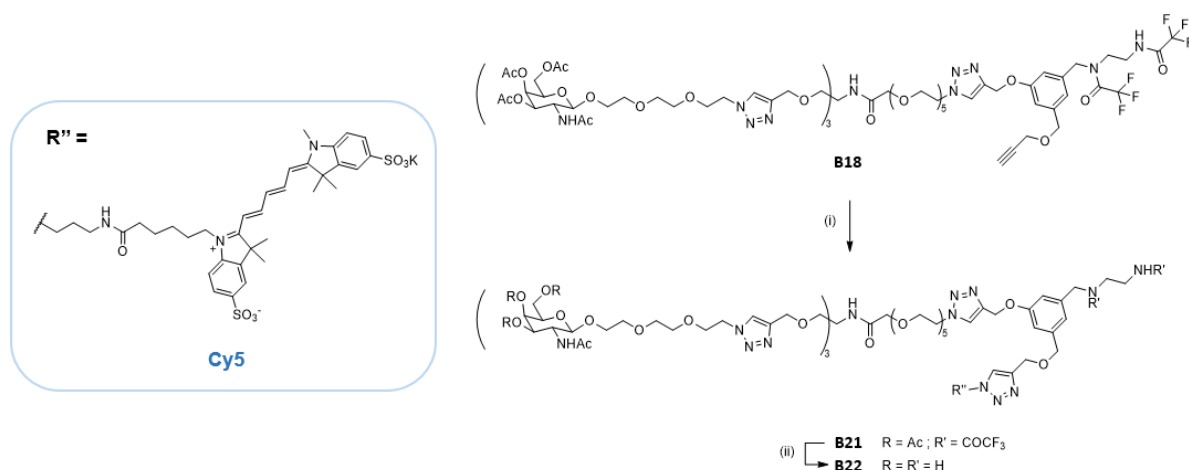
Results of agarose gel electrophoresis from T7E1 assay of RNP transfection to HEPG2 cells at various concentrations of RNP-GalNAc-EED (Cas9:ON3.5-EDA-GalNAc-EED). Electrophoretic migration of T7E1-digested EMX1 amplicon is visualised. Lane 1: without RNP transfection (negative control); lane 2: RNP (Cas9:ON3.5) transfected by lipofectamine CRISPRMAX™ (positive control); lane 3-5: naked transfection of 50, 250 and 1250 nM of RNP-GalNAc-EED. Indel frequency is quantified by the equation $(1 - (1 - f_{\text{cut}}/f_{\text{total}})^{0.5}) \times 100$, where f stands for fraction, using Agilent Bioanalyzer 2100®.

3.6.3 Imaging of RNP with gRNA-GalNAc-Cy5 Conjugate in HEPG2 Culture

The previous two transfection experiments indicated the absence of RNP activity from undetectable indel formation, which is a downstream readout, but did not provide sufficient information about the efficiency of the internalisation process. Hence, we designed another bi-functional conjugate which incorporates both GalNAc and a fluorescent tag, Cy5, which could fluoresce and indicate the localisation pattern of the conjugated gRNA in real time under confocal microscope.

The synthesis proceeded from precursor **B18**, which already contains the triantennary GalNAc and EDA moieties in their protected forms (Scheme 3.7). The free alkynyl group was coupled with commercially available sulfo-cyanine 5 azide (Lumiprobe) via CuAAC and all protecting groups were globally removed by treatment with methanolic ammonia and purified on RP-HPLC to afford EDA-GalNAc-Cy5 reagent **B22**. We then subjected gRNA ON3.5 to reductive amination with **B22** that provided 82% conversion, a value in line with previous trials, to the desired fluorescent conjugate ON3.5-EDA-GalNAc-Cy5 (Figure 3.14). The product was then isolated by RP-HPLC (MeCN/water, TEAA pH 7.0) and subsequently used in naked transfection to HEPG2 cells.

In the transfection protocol, we employed recombinant eGFP-Cas9 protein (Sigma-Aldrich), which possesses an enhanced GFP fusion that fluoresces when excited at 488 nm and a nuclear localisation sequence (NLS) that induces translocation across the nuclear membrane, to form RNP with ON3.5-EDA-GalNAc-Cy5 to cross-validate its localisation profile in HEPG2 culture. To study the endosomolytic effect of ppTG21, we also attempted to incubate the cells in the presence of ppTG21 at 10 μ M. A control set with untreated cells was also analysed in comparison to RNP transfected cells (Figure 3.15A). It was found that incubation with the RNP resulted in co-localisation of both eGFP and Cy5 signals on the cell surface after 2 hours (Figure 3.15B), indicating that the RNP with GalNAc conjugation interacts with the



Scheme 3.7 | Synthesis of EDA-GalNAc-Cy5 conjugate

Reaction conditions are as follows: (i) CuSO₄, THPTA, sodium ascorbate, water/DMSO, r.t., 2h, 51%; (ii) ammonia/MeOH (3.5 N), r.t., 18h, 82%.

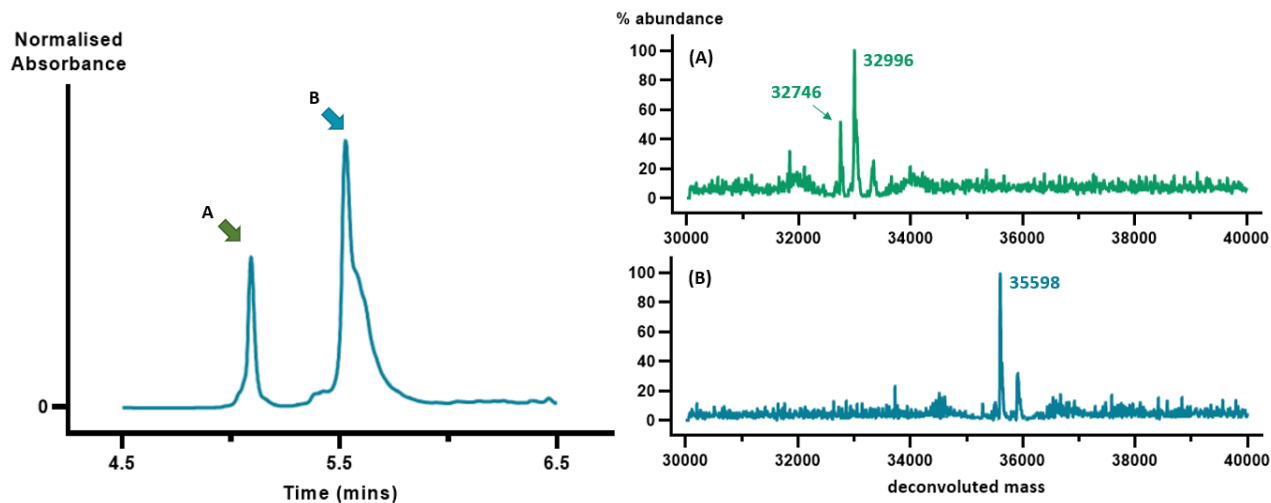


Figure 3.14 | UPLC-MS analysis of EDA-GalNAc-Cy5 conjugation on gRNA

Left: UPL chromatograms show the retention profile of sgRNA ON3.5 subjected to oxidation and reductive amination with B22. Right: deconvoluted mass of the major peaks from the chromatogram: (A) oxidised ON3.5-dial (calc. mass + 2[Mg]²⁺ = 32992); (B) ON3.5-EDA-GalNAc-Cy5 (calc. mass + [Mg]²⁺ = 35589). Reaction condition: (1) NaIO₄ (1.5 mM), r.t., 30 minutes, in darkness; (2) EDA-GalNAc-Cy5 (2.0 mM, 26.6 eq.), NaB(CN)H₃ (20 mM), KH₂PO₄ (100 mM, pH 6.0), r.t., 2 hours, in darkness.

cell membrane of HEPG2 cells and plausibly with ASGPr according to literature findings.¹⁸² The signals remained after 4 hours of incubation but exhibited only nominal intensity in the cytoplasm, indicating that minimal internalisation took place. Overnight incubation (18 hours) resulted in the disappearance of fluorescence on cell surface but minute puncta of Cy5 signals in the cytoplasm. These observations provide evidence that the GalNAc-conjugated RNP could interact with ASGPr on hepatocytes but endocytosis could not proceed efficiently to internalise the RNP into endosomes; and after longer periods of incubation, small amounts of GalNAc-bearing gRNA might dissociate from Cas9 and enter the cell via endocytosis. In such scenario, no CRISPR-mediated gene editing activity can be observed.

When ppTG21 was co-incubated with the RNP and HEPG2 cells for 2 hours, an aggregation of Cy5 signal was observed beyond the cell membrane, with little intensity found within the membrane. The fluorescence from eGFP-Cas9 co-localised with the smearing shade of Cy5 (See appendix). It suggests no RNP had been internalised and it was present outside of the cell as an aggregate possibly due to the interaction with the cationic ppTG21 (charge = +1.3 at pH 7.4). After overnight incubation, the intensity of Cy5 signals from gRNA dropped significantly and an insignificant level of Cas9 signal remained as aggregates.

The microscopic images provided mechanistic indications to rationalise the results from the CRISPR activity assays (T7E1). We suspected that the incorporation of a triantennary GalNAc moiety on gRNA does not confer sufficiently strong binding between the large, non-covalently bound RNP complex and the ASGPr on HEPG2 cell surface. This binding affinity might be inadequate in nature and further compromised by the presence of a bulky protein unit (> 192 kDa). Previous studies on ASGPr-mediated delivery of RNP substantiated this hypothesis as they discussed the need of using two trivalent clusters of a stronger-binding ASGPr-ligand (with an affinity more than 5 times higher) conjugated to Cas9 to initiate endocytosis of the RNP.^{182,207} Such discussion reflected that it might be necessary to induce

endocytosis of very large molecules, such as RNP, with ASGPr-ligands of high affinities or valences. On the other hand, Cas9 was chosen to be the conjugation platform for ASGPr-ligands in the study. The reason we believe is that RNP, unlike ASO or siRNA, has a large molecular weight and is formed from weak non-covalent interactions; these properties render it requiring a much larger activated clathrin-coated pit to engulf and more susceptible to dissociation. This might highlight the importance of inducing endocytosis of the protein instead of the gRNA subunit within RNP to effect efficient internalisation.

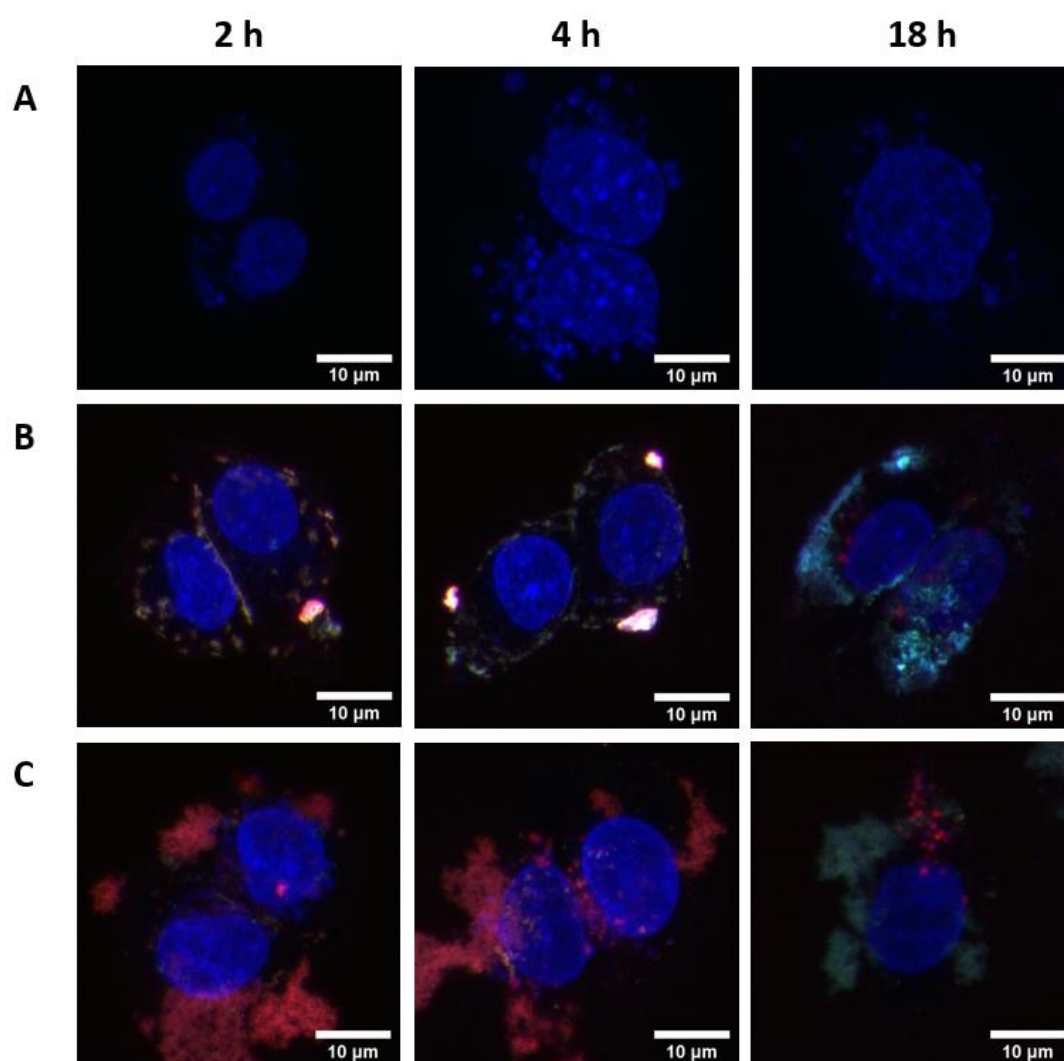


Figure 3.15 | Images of RNP localisation in HEPG2 cells by confocal microscopy

Internalization in HEPG2 cells of RNP via naked transfection at designated time points (T = 2, 4 and 18 hours). (A) untreated control; (B) eGFP-Cas9:ON3.5-EDA-GalNAc-Cy5 (50 nM); and (C) eGFP-Cas9:ON3.5-EDA-GalNAc-Cy5 (50 nM) with ppTG21 (10 μM). Blue: Hoechst stain of cell nuclei. Green: fluorescent signal from eGFP. Red: fluorescent signal from Cy5 on gRNA. Images not adjusted. Images in split channels are illustrated in Figure S4 – S6 in Appendix.

3.7 Conclusions

Recent years have witnessed a surge in the adoption of GalNAc-conjugated oligonucleotides as a drug delivery platform to liver tissues. The idea of ASGPr-mediated intracellular delivery has prompted us to investigate the possibility of delivering CRISPR RNP to hepatocytes by conjugating gRNA with GalNAc moieties. In this study, we explored methods of incorporating GalNAc to long, structured gRNA which is conventionally difficult to achieve by solid-phase oligonucleotide synthesis (SPOS) due to constraints in coupling efficiencies and instability of RNA in basic or nucleophilic environments. The synthetic strategies include post-synthetic labelling of shorter RNA fragments, enzymatic labelling and photo-crosslinking of GalNAc-bearing short oligos. We identified the simplest and most economical approach is to employ reductive amination to modify the 3'-end of gRNA with triantennary GalNAc reagents via a morpholino linkage transformation where the conversion can reach 69 – 82 % by using an excess of GalNAc-EDA derivatives.

We also established synthetic routes to access bi-functional EDA reagents which provide additional functionalities (e.g., endosomal escape domain and fluorescent dye) on the GalNAc conjugate, when coupled to gRNA. The one-pot syntheses of various gRNA constructs with these novel EDA derivatives proved the feasibility and efficiency of achieving dual conjugation on long RNA without excessive work-up and purification steps. However, after transfecting RNP formed by GalNAc-conjugated gRNA and Cas9 to cells, in the presence of endosomolytic peptides or EED, we observed no CRISPR activity at genomic level. The unexpected findings led us to investigate the localisation of RNP with GalNAc conjugation. By installing fluorescent Cy5 on the gRNA construct, we observed that the RNP constructs were able to interact with hepatocytes on cell surface but nominal internalisation was effected within the transfection timeframe. It is postulated that the binding of GalNAc with ASGPr is not sufficiently strong and/or the activation of clathrin-mediated endocytosis is not capable of

internalising RNP with a large molecular size. As a result, dissociation is favoured over internalisation when RNP interacts with the receptors on the cell surface. In conclusion, we determined that the conjugation of a single triantennary GalNAc cluster on gRNA does not trigger efficient delivery of CRISPR RNP into HEPG2 cells, to the contrary to the clinically proven delivery efficacy of shorter, less charged oligonucleotide drugs such as ASO and siRNA.

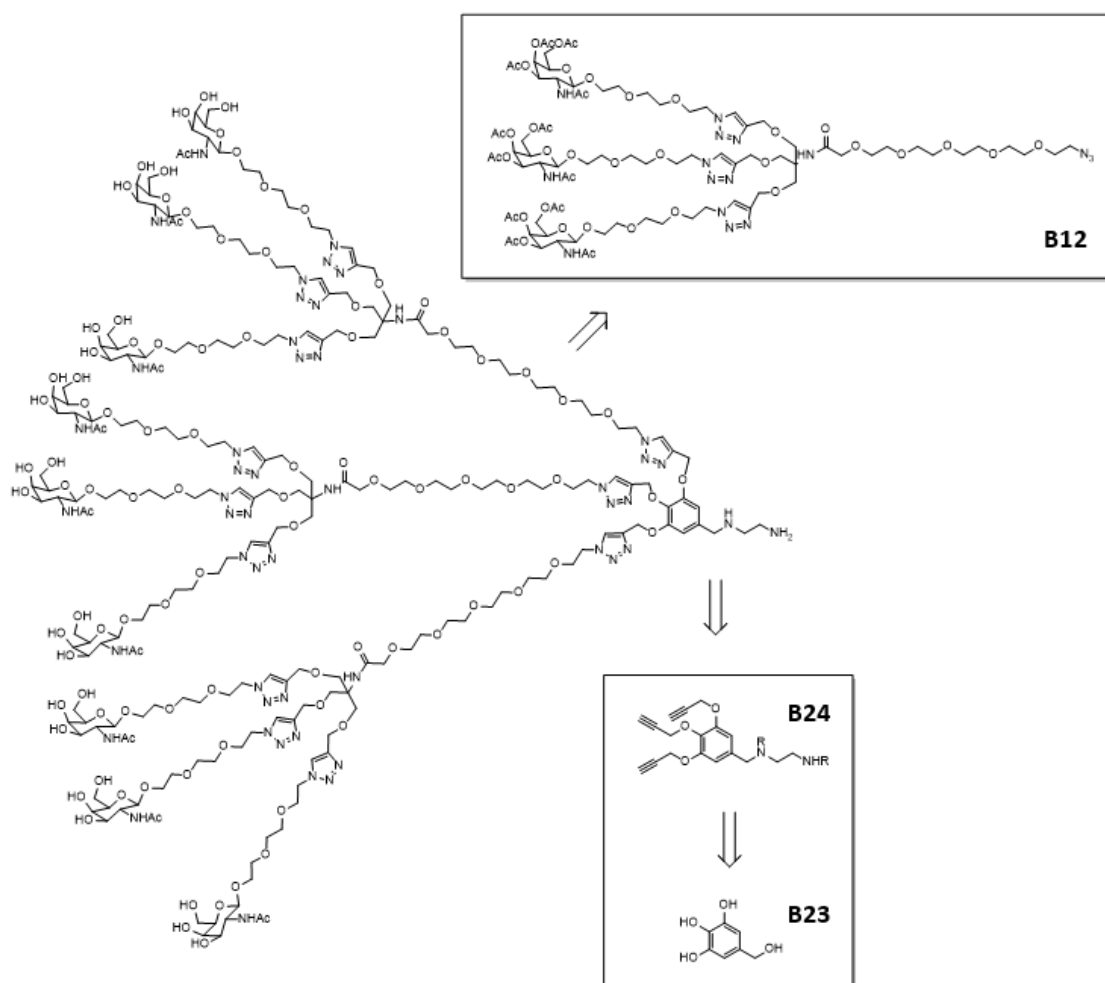
3.8 Future Outlook

Although the results did not meet our initial expectations, delivering CRISPR RNP to hepatocytes via ASGPr-ligand conjugation on gRNA remains an attractive prospect due to its relative biological compatibility and target specificity over lipid- or virus-based platforms, and the simplicity in the manufacturing process in regard to protein modifications in reported transfection protocols. For the continuation of this effort, some aspects are worth further investigation and improvement.

First, it is beneficial to understand the binding affinity between GalNAc-conjugated gRNA and ASGPr. Although existing studies provide approximations on the dissociation constant (K_D) between GalNAc (and other ASGPr-ligands) and immobilised ASGPr via surface plasmon resonance measurement, little is known about the effect on binding affinity when the ligands are covalently linked to gRNA, which is highly charged, or when complexed with Cas9, which has large molecular weight. Because of such shortfall, it is difficult to determine if the interaction between GalNAc-conjugated gRNA and ASGPr is sufficient to maintain strong binding and trigger subsequent internalisation of the entire RNP complex.

If poor binding strength is a valid reason for the limitation internalisation efficiency, possible solutions are to switch to ligands with stronger binding affinities such as 1,5-bridged GalNAc derivative described in literature,²⁰⁷ or the incorporation of multivalent ligand clusters (i.e. more than one sugar trimer) on gRNA to promote binding with multiple ASGPr units.

These approaches will be chemically accessible by the synthesis of the corresponding EDA derivatives with ASGPr-ligand clusters and subsequent reductive amination with gRNA. An example of the synthesis of such cluster with polyvalent display of GalNAc (or alternative ligands) from triantennary GalNAc azide, **B12** and commercially available **B23** is illustrated in Scheme 3.8.



Scheme 3.8 | Molecular design of polyvalent GalNAc-EDA derivative

An EDA conjugate capable of coupling to RNA via reductive amination can be synthesised from synthon B12 reported in this study and commercially available polyphenol B23 via CuAAC.

Chapter 4 – Bioconjugation of mRNA Fragments

4.1 Background

4.1.1 Eukaryotic pre-mRNA processing

Proteins are an indispensable class of molecules for any living organisms as they serve an array of essential functions including metabolism, signalling, and structural support. In normal cases, the biosynthesis of these molecules is tightly controlled by genes. As discussed in Chapter 1.2, modern biology describes proteins as the product of *translation*, a process where a transcript of genetic information is decoded to form peptides with a unique amino acid sequence that folds into a specific three-dimensional structure that defines their properties to bind to molecules and hence their functions. In this process, messenger RNA (mRNA) is a crucial actor to relay genetic information between genes and cellular machinery for protein synthesis. Following *transcription*, where a gene is copied from the genomic DNA to a primary transcript mRNA (also known as pre-mRNA), a number of processes are carried out by various enzymes to convert pre-mRNA into mature mRNA which is then transported to the cytoplasm for translation. Such modification processes significantly alter the structure of the mRNA molecule and are essential to regulate its stability and ability of being translated into proteins. Such structural alterations are regarded as ‘post-transcriptional modifications’ (Figure 4.1).

There are three major types of biochemical processes for these modifications. The first one is *5'-capping*, which involves the addition of 7-methylguanosine (m^7G) to the 5' end of the pre-mRNA via a 5'-5' triphosphate linkage. To accomplish this, one phosphate group is first removed from the 5'-triphosphate by RNA triphosphatase (RNA TPase), followed by the addition of one guanosine monophosphate (GMP) unit by guanylyl transferases (GTase) on the remaining 5'-diphosphate on the pre-mRNA in a 5'-5' fashion. After the linkage is formed, the covalently bound guanosine (G) cap is methylated by guanine-N7 methyltransferase (N7 MTase) at the N7 position by the methyl donor, S-adenosylmethionine (SAM). The 2'-O on the +1 ribose will further be methylated by 2'-O methyltransferase (2'-O MTase) that protects

the primary transcript from digestion by ribonuclease.²⁰⁸ This cap, namely m⁷Gppp, not only provides resistance against nucleases but also a recognition site for the binding of eukaryotic translation initiation factor 4E (eIF4E) which triggers a cascade of additional protein interactions with eIF4G and eIF4A, activates the recruitment of small ribosomal unit (40S) and helicase activity of eIF4A to enable the initiation of the translation process.^{125,209}

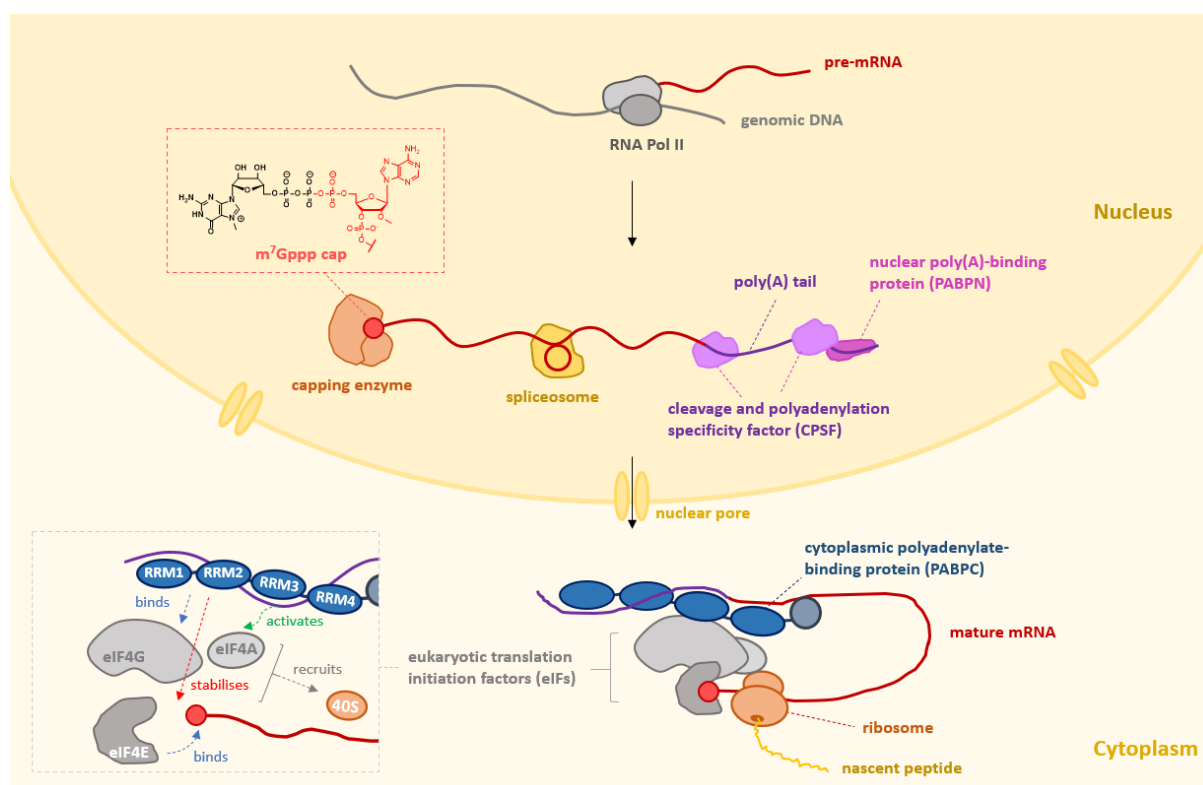


Figure 4.1 | Schematic overview of post-transcriptional modifications in eukaryotes

After transcription by RNA polymerase (RNA Pol II) in the nucleus, pre-mRNAs are processed by 5' capping, splicing, and 3' cleavage by various enzymes and polyadenylation, which is regulated by nuclear poly(A)-binding protein (PABPN). Mature mRNAs bearing 3'-poly(A) sequence are exported to the cytoplasm. Cytoplasmic poly(A)-binding protein (PABPC) binds to poly(A) tails and promotes translation initiation via multiple interactions with eukaryotic translation initiation factors (eIFs).

In addition to 5'-processing, the second modification process at the 3'-end is imperative to confer mRNA stability and activate translation machinery. This step is called *3'-cleavage and polyadenylation*. Following the synthesis of pre-mRNA, cleavage and polyadenylation specificity factors (CPSF) bind to its 3'-proximal region that usually contains a polyadenylation

signal (PAS) sequence, AAUAAA, and recruit other cleavage factors to catalyse cleavage of the strand at 10 – 30 nt downstream of this sequence.^{210,211} Meanwhile, CPSF also interact with poly(A) polymerase (PAP) which incorporates adenosine monophosphate units from adenosine triphosphate to the cleavage site at the 3'-end of RNA, resulting in the assembly of a poly(A) tail which often contains 50 – 100 adenosines.^{125,212} This poly(A) tail, together with its upstream segment at the 3'-proximal end regarded as 3'-untranslated region (3'-UTR) that does not code for proteins, serves significant regulatory functions on gene expression. For instance, cytoplasmic poly(A)-binding protein (PABPC) binds to poly(A) tail on mature mRNA in the cytoplasm and protects it from degradation; it also interacts with a number of translation initiation factors at the 5'-end that act in tandem to recruit the small ribosomal unit (40S) for translation initiation.¹²⁵ These synergistic interactions between the 5'-cap, 3'-poly(A) tail and protein factors have been found to strongly enhance translation efficiency compared with mRNAs with either or no such modification in eukaryotic cells.²¹³

Finally, RNA splicing process takes place on the pre-mRNA to remove any non-coding region (introns) and splicing back the coding parts (exons) by spliceosomes, a ribonucleoprotein residing in the nucleus. Some introns of the mRNA can fold into a complex secondary structure and catalyse self-splicing while tRNA can also involve in splicing in certain species such as yeast *Saccharomyces cerevisiae*.²¹⁴ This activity happens in all eukaryotes whereas the mechanism of which can differ in prokaryotes. Although introns are normally not translated into protein products, the presence of introns provides a reservoir of long non-coding RNA (lncRNA), which is involved in regulating gene expression, as well as potential new genes by allowing non-coding genes to be expressed through evolutionary processes.²¹⁵ Hence, such features are retained in most protein-coding genes but are subject to extensive splicing processes upon mRNA maturation.

4.1.2 The Essential Roles and Properties of Poly(A) Tail

As discussed in the previous section, the poly(A) tail acts as an important structural motif of mRNA to provide stability against nucleases and activate a cascade of protein interactions for translation initiation. An early study on mRNA stability demonstrated that the samples (of rabbit globin mRNA) in *Xenopus* oocytes with 32 – 122 nt long poly(A) tail were translated as efficiently as the native mRNA with 150 nt poly(A), whereas those with fewer than 21 adenosines were translated poorly or not translated at all.²¹⁶ Echoing this finding, deadenylation (removal of adenosine residues) was found to promote mRNA decay following the dissociation of PABPC.²¹⁷ These results shone light on the interactions of poly(A) with protein factors and its stabilising effects on mRNA that are functionally important for eukaryotic translation.

In recent years, more mechanistic insights over the roles of poly(A) in controlling mRNA stability and translation efficiency have been revealed. Importantly, the binding modes and functions of the key effector protein, PABPC, have been determined, offering researchers greater understanding on how translation is regulated via the mRNA 3'-region. Poly(A) RNA segment adopts an A-form-like single-stranded structure.²¹⁸ Structural elucidation of eukaryotic PABPC by crystallography indicated the presence of four highly conserved N-terminal RNA recognition motif (RRM) domains which possess high binding affinities to poly(A) to different extents.²¹⁹ The RRM domains in total covers around 30 nucleotides but only about 12 are needed for high-affinity binding by RRM1 and RRM2, which possess higher specificity and binding affinity to the poly(A) sequence than RRM3 and RRM4. A long stretch of poly(A) tail (> 60 nt) may be bound to multiple PABPCs which interact with each other and promote cooperative binding, but the amount of associated PABPCs bound to the tail is limited by its concentration available in the cell.^{220,221}

Apart from binding to the poly(A) tail at the 3'-region, PABPC also interacts with eIF4G, the protein factor that associates with the 5'-m⁷G cap. It in turn promotes the formation of the translation initiation factor complex by stabilising the interactions between eIF4E and 5'-cap as well as stimulates the ATPase and helicase activities of eIF4A, resulting in a closed-loop structure where the 5'-cap, eIF4E, eIF4G, eIF4A, PABPC and 3'-poly(A) are closely bound together in a configuration that can trigger the recruitment of small ribosomal unit (40S) and opening of the RNA strand for translation initiation (Figure 4.1).¹²⁵ It has been understood that the closed-loop structure serves a quality-control function to ensure partially degraded mRNA, with shortened or no poly(A) tail, is not translated.

Finally, PABPC also serves an important function to confer stability to the mRNA via binding to the poly(A) tail. From a mechanistic perspective, deadenylation or poly(A) shortening is considered as the first and rate-determining step of mRNA decay, followed sequentially by the dissociation of PABPC and 5'-decapping which eventually provide access for exoribonucleases to digest the mRNA from either the 5'- or 3'-end.²²² The binding of PABPC thus offers an immediate steric effect to protect the poly(A) from being accessed by deadenylases and prevents the dissociation of the eIF4E-cap complex that allows decapping activities, resulting in a slower degradation rate. The process of mRNA decay follows an ordered progression from poly(A) deadenylation which is a controlled process governed by its binding dynamics with PABPC and deadenylation enzymes. These properties of the poly(A) tail render it an important control of translation efficiency and metabolism of the mRNA transcript.

4.1.3 3'-Untranslated Region (3'-UTR) and Non-canonical Poly(A) Tail

Other than introns, mRNAs usually contain two untranslated sequences upstream and downstream of its protein-coding regions, namely 5'- and 3'-untranslated regions (UTRs). Although scant evidence has suggested that 5'-UTR participates in protein-mRNA interactions, there are multiple studies indicating that 3'-UTR is involved in translation regulation via its interactions with RNA-binding proteins.²⁰⁹ One example of which is demonstrated by its inhibitory effect on translation initiation. A widely understood model illustrates the binding of 3'-UTR to a sequence-specific binding protein (protein X), which forms an alternative closed-loop structure with the 5'-cap via a bridging protein (protein Y) and a cap-binding protein (protein Z). The identities of these proteins can vary in different cell types but they generally possess inhibitory properties to translation as such alternative loop structure precludes access of eIF4E, eIF4G and eIF4A to the 5'-end, preventing the formation of the translation initiation complex necessary for ribosome recruitment.^{223,224} In addition to that, it was reported that the 3'-UTR-5'-cap interactions can promote mRNA oligomerisation and aggregation, leading to translation repression.²²⁵

Interestingly, 3'-UTR has also the potential to up-regulate gene expression by prolonging mRNA half-life. For instance, the incorporation of a consecutive U sequence within the 3'-UTR can introduce secondary structures at the poly(A) tail via Watson-Crick base-pairing. Such structure was found to attenuate deadenylation and increase the half-life of mRNA, possibly by altering the affinity of deadenylases or PABPC at the 3'-region.^{226,227} As more experimental evidence emerges, the role of the 3'-region of mRNA in translation regulation has become more apparent.

Another factor affecting mRNA stability has been identified at the poly(A) tails in certain cell models. Sequencing studies have shown that non-adenosine residues are sometimes found in the poly(A) tails in cells, and guanosine residues, in particular, can confer stability to

mRNA, possibly by altering the helical structure of poly(A) and reducing its binding affinity to deadenylases.^{228–230} A recent study has identified the properties of human cytomegalovirus (HCMV) and hepatitis B virus (HBV) to hijack cellular post-transcriptional modification machinery to incorporate mixed nucleosides (C, G and U) into its poly(A) tail, resulting in enhanced protection of the viral mRNA against the host's deadenylation mechanisms.²³¹ These sequence anomalies naturally existing in biological systems reveal that the regulatory roles of 3'-poly(A) in mRNA stability and translation efficiency can be multi-fold. The findings have intrigued researchers to investigate possible means to modify this important regulatory motif in mRNA and the possibility to enhance translation by such modifications.

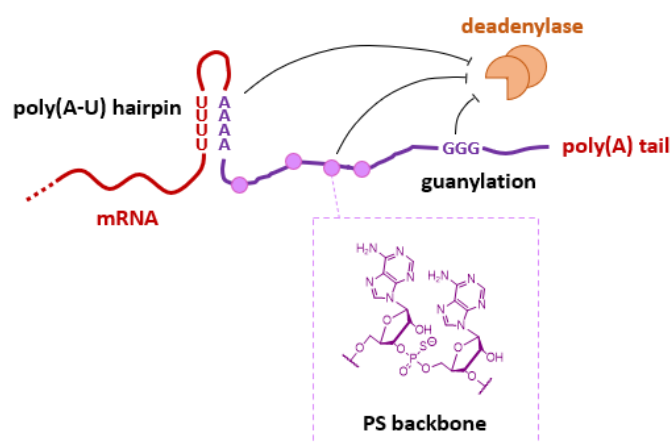


Figure 4.2 | Examples of 3'-UTR and poly(A) tail modifications

Various non-canonical motifs can be introduced to the 3'-UTR and poly(A) tail artificially or naturally. Poly(U) sequence can be transcribed from a synthetic DNA template and incorporated at the 3'-UTR to form secondary structure with the poly(A) tail, resulting in the inhibition of deadenylase activity. Phosphorothioate backbone, which exhibits greater stability against CNOT7 deadenylase, can be incorporated into poly(A) tail non-specifically with adenosine-5'-O-(1-Thiotriphosphate) (ATP α S) by PAP enzyme. Some cells and viruses can introduce non-A residues, such as guanosine, on the poly(A) tail that can slow down deadenylase digestion.

Recently, it has been shown that mRNA poly(A) with non-specific insertion of phosphorothioate (PS) and boranophosphate (PBH₃) modifications can be prepared by incubating IVT mRNA with modified ATP analogues, namely ATP α S and ATP α BH₃, in the presence of poly(A) polymerase (PAP) *in vitro*.²³² Treating these constructs with human

CNOT7 deadenylase, the mRNAs with PS-modified poly(A) tail have shown corresponding increase in stability against deadenylation with increasing concentration of ATP α S supplied in IVT; those with PBH₃-modifications are conversely more susceptible to deadenylation. Although the assessment on translation efficiency in cells is lacking, it offers a reasonable hypothesis that PS-modifications on mRNA poly(A) could slow down deadenylation and hence prolong mRNA half-life and promote translation.

These examples have indicated that the 3'-proximal regions of mRNA can be subject to sequence or structural modifications which improve mRNA stability and, in some cases, translation efficiency without losing recognition of cellular translation machinery. The precise control of such modifications thus proves great potential to enhance the *in vivo* stability and functionality of future mRNA-based therapeutics.

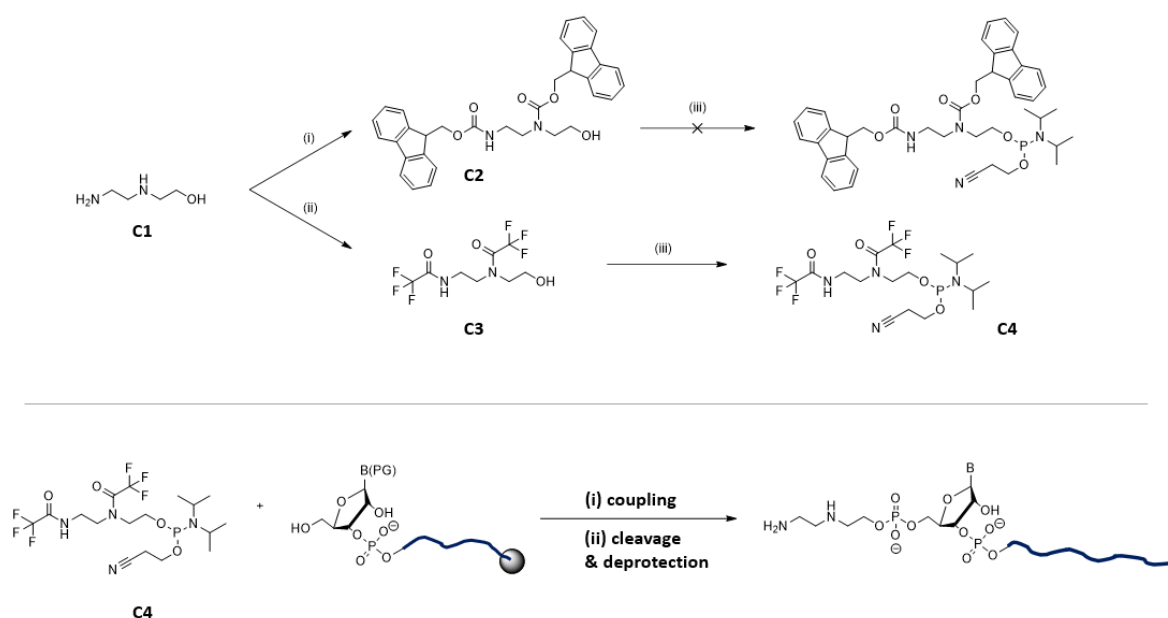
4.2 Aims and Objectives

Although modifications on the poly(A) tail in mRNA have shown great promise in promoting translation *in vivo*, they are difficult to be translated into therapeutic candidates as methods to precisely control modifications on long RNA remain lacking. In this chapter, we aim to develop a method of introducing sequence and structural modifications in the poly(A) region via chemical ligation. A 5'-EDA-conjugated synthetic oligonucleotide, which may bear any modifications or sequence alterations, will be ligated to the 3'-end of an RNA by reductive amination in a templated manner. Such strategy enables the incorporation of modified nucleosides (XNA) in any desired quantities and positions within synthetic limits. As demonstrated in Chapter 3, we envisage the chemistry between the RNA 3'-end and the 5'-EDA moiety to be robust and efficient in forming the chimeric RNA/XNA hybrid. The investigation of the ligation reaction under various conditions will be discussed. Our final goal is to provide a new tool to site-specifically modify the 3'-region of mRNA by chemical means.

4.3 Ligation Strategies and Oligonucleotide Designs

4.3.1 Synthesis of Ethylenediamine (EDA) Phosphoramidite

To ligate an oligonucleotide to an the 3'-end of an oxidised RNA (dialdehyde) via reductive amination, a reactive amino modifier is first installed at its 5'-end. As discussed in Chapter 3, ethylenediamine (EDA) is a more reactive function group than any alkyl amine; thus, we first attempted to prepare a protected EDA derivative by phosphitylating its hydroxy end and incorporate the EDA-motif at the 5'-end of the attachment oligo with this phosphoramidite by SPS.



Scheme 4.1 | Synthesis of EDA phosphoramidite and incorporation of EDA-motif

Reaction conditions are as follows: (i) Fmoc-OSu, dioxane/water, r.t., 18h, 89%; (ii) trifluoroacetic anhydride, Et₃N, DCM, 0°C to r.t., 6 hours, 37%; (iii) 2-Cyanoethyl N,N,N',N'-tetraisopropylphosphoramidite, tetrazole, MeCN, r.t., 1 hour, 54 %. Coupling time of EDA-phosphoramidite is discussed in experimental section.

In our initial attempt, we chose the Fmoc group, a base-labile protecting group, to mask the reactivity of the amino groups. We first prepared a Fmoc-protected N-(2-hydroxyethyl)ethylenediamine (C2) which was treated with 2-cyanoethyl N,N,N',N'-tetraisopropylphosphoramidite, a phosphitylating reagent, and tetrazole to form the desired

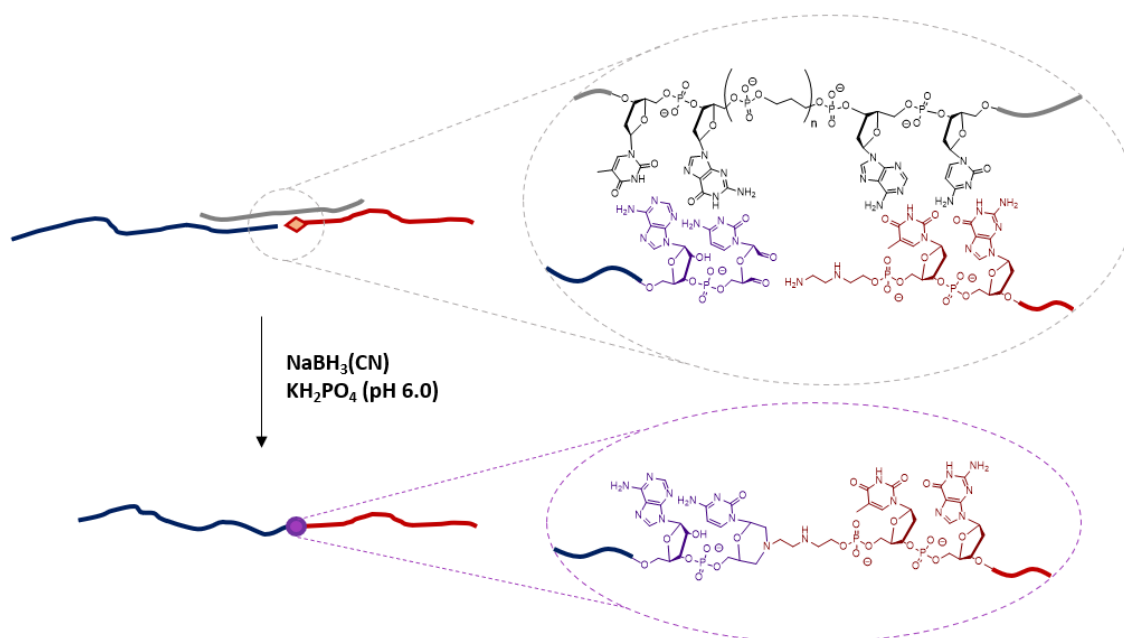
phosphoramidite. However, TLC analysis indicated the formation of multiple species after 1 hour of reaction, revealing that a number of side reactions or degradation might have taken place. As a result, we switched to trifluoroacetamide (TFA), a more stable protecting group, on the same precursor, N-(2-hydroxyethyl)ethylenediamine (C1). After the TFA groups were incorporated by treatment with trifluoroacetic anhydride to form C3, the free hydroxy group was phosphitylated with the same reagent in 1 hour to afford the product phosphoramidite C4 at 54% yield (Scheme 4.1). The phosphoramidite product C4 was aliquoted into 100 mg portion in oligonucleotide synthesis vial and kept at -20°C until being used for coupling in oligo synthesis. No loss of reactivity was observed after 3 months of storage.

4.3.2 Splinted Ligation

Compared to reported reductive amination protocols where the EDA-derivative is present in high concentrations (> 1 mM), the concentrations of the 3'-end of an RNA and the 5'-EDA modifier between two oligo strands are much lower (< 10 μ M) and the rate of reaction is also significantly reduced. Hence, it is essential to increase the effective concentrations of the two reactive groups. A common strategy for such purpose is templated ligation, by which a splint oligo is introduced to bind the two reacting oligos complementarily and bring them in close proximity to speed up the reaction. As a proof of concept, we designed a three-component system whereby a model RNA fragment (56-mer) is first oxidised by sodium periodate and incubated with a splint, and then a 5-EDA oligo fragment was introduced to the mixture under various conditions for reductive amination. Two splint oligonucleotides were designed with one having full complementarity to the reacting oligos and the other having a C3-spacer at the site of reaction (Scheme 4.2 and Table S1).

The aim of the initial screening is to identify if the 5'-EDA attachment on an oligo is reactive towards the 3'-dial of an oxidised RNA. For this purpose, we employed crude

oligonucleotides obtained from SPS without further purification. The RNA species remaining in the reaction was isolated by precipitation and subsequent elution on Monarch® RNA Cleanup Kit. After incubating with NaIO_4 without a splint, no conjugated product was identified as the RNA was subjected to reductive amination in the presence of EDA-oligo (10 eq.) and $\text{NaBH}_3(\text{CN})$ at pH 6.0. When a splint oligo without spacer (1 eq.) was introduced at the oxidation step, a mass peak corresponding to the conjugated product via a morpholino-type linkage after reductive amination was observed by UPLC-MS, but it is covered in the same elution peak with the splint (Figure S1 and S2). When the splint contains an internal C3 spacer, the product peak was observed as a single entity from chromatography, confirming the presence of the ligated product after reductive amination. Within the 2-hours duration, increase of temperature from 25°C to 37°C did not significantly alter the chromatographic pattern, indicating product formation might not be limited to kinetic factors. This study provides evidence that reductive amination can ligate RNA with synthetic oligo. In the next part, we used a quantitative methodology to study the optimal conditions for the conjugation reaction.



Scheme 4.2 | Splinted conjugation between 3'-oxidised RNA and 5'-EDA oligonucleotide by reductive amination

This scheme illustrates the trimolecular reaction between the RNA (red) and 5'-EDA oligo (blue) assisted by a splint (black) under reductive amination conditions. $n = 0$ or 1 .

Having confirmed from the previous experiment that the 5'-EDA motif on an oligonucleotide is reactive towards the 3'-end of oxidised RNA to form the desired linkage, we moved to investigate the effect of reduced loading of 5'-EDA oligonucleotide on product conversion. Theoretically, the splint should bring an equimolar quantity of two reacting oligos in vicinity and thus 1 equivalent of the EDA-oligo is sufficient for complete conversion. In this study, a new set of oligonucleotides were designed to improve the chromatographic separation between starting-material and product peaks for quantitative analysis. They include a 42-mer RNA, 33-mer splint and a 72-mer 5'-EDA oligo with 64 adenosines at the 3'-end as a poly(A) tail mimic (Table 4.1). This mimic also contains full 2'-OMe ribose and 3'-Cy3 modifications to prevent depurination and provides fluorescent signal for identification.

Using identical oxidation-reductive amination condition, we then analysed product conversion with loadings of the EDA-oligo at 1.0, 2.0 and 10.0 equivalents and splint at 1.0 equivalent by injecting the crude reaction mixture into UPLC-MS. The conversion is calculated by the ratio between the quantity of product (integrated area under the peak divided by its extinction coefficient) that of the RNA starting material and product (Figure 4.3). As shown before, in the absence of a splint oligo, no conjugation reaction was observed at 10 eq. loading of EDA-oligo (set 1). When the splint was introduced to the reaction with 10 eq. of EDA-oligo, the conversion to product was 18% by chromatography (set 2). As the loading was decreased to 2 eq., product conversion slightly improved to 23% (set 3) whereas further reduction to 1 eq. provided 23% conversion (set 4). We refrained from concluding that lowering EDA-oligo loading can improve conversion as this could be due to the attenuation of peak intensities from oversaturation of the EDA-oligo peak. However, this experiment provides evidence that 1 eq. loading of EDA-oligo furnishes comparable level of product to 10 eq., indicating that an equimolar reaction is possible.

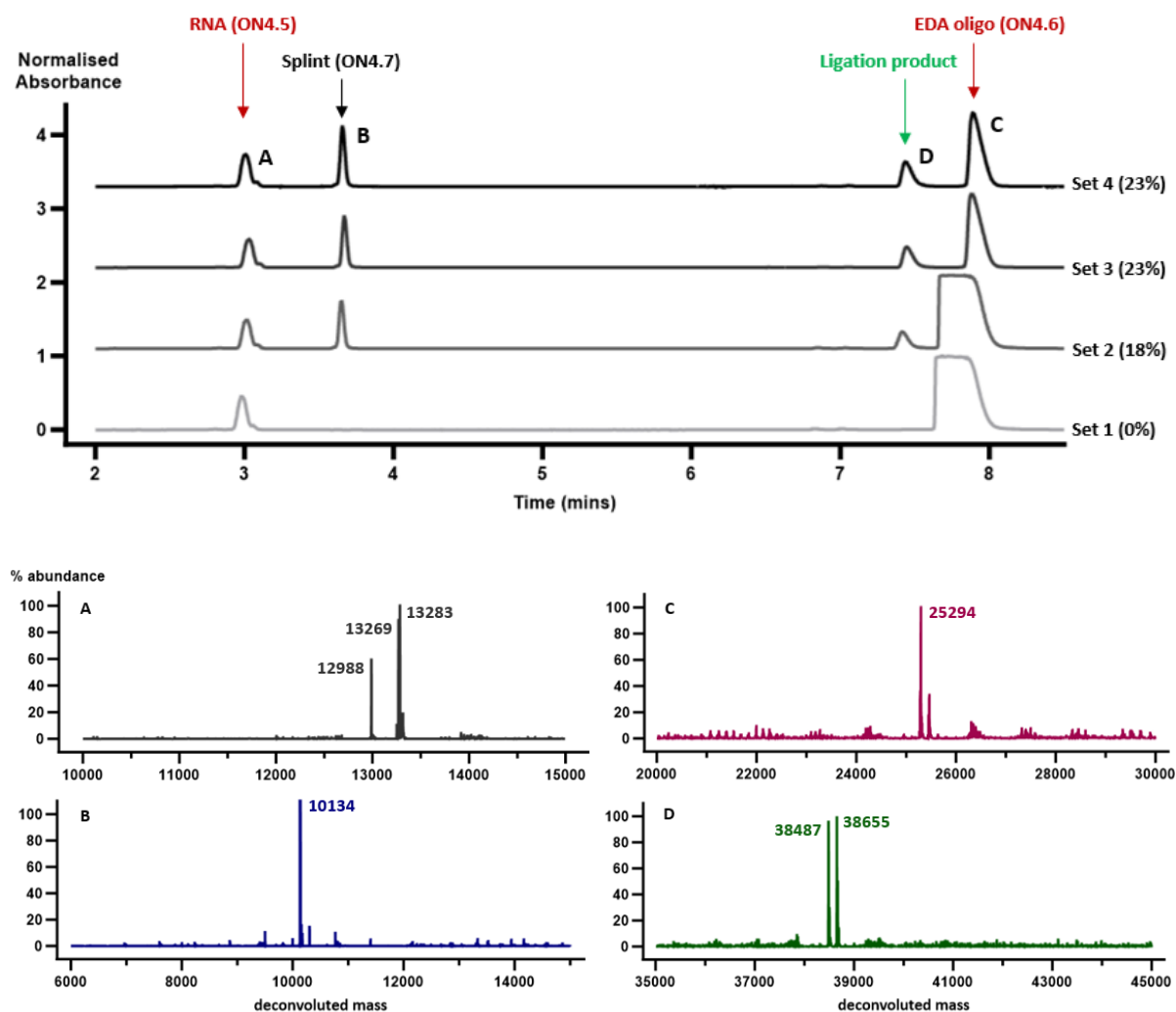


Figure 4.3 | UPLC-MS results of RNA 3'-conjugation at various EDA-oligo loadings

Top: Overlaid UPL chromatograms show the retention profiles of the crude reaction mixtures after oxidation-reductive amination with various loadings of EDA-oligo. A control experiment without splint was performed. Before injecting to UPLC-MS, to the mixture was added Et3N.EDTA (final conc. = 50 mM) to remove sodium adducts at early elution. Conversion % in parentheses was calculated as the ratio between the quantity of product (integrated area under the peak divided by its extinction coefficient) that of the RNA starting material and product with triplicate experiments ($n = 3$). General reaction condition: RNA (final conc. = 5 μ M) was first incubated with NaIO₄ (1.5 mM) with or without a splint (1 eq.) in darkness for 30 minutes, then to the mixture was added various components to a final conc.: KH₂PO₄ buffer (100 mM, pH 6.0), 5'-EDA oligo (varied) and NaBH₃CN (20 mM) for 2 hours at r.t. Set 1: no splint, 10 eq. of EDA-oligo; set 2: with splint, 10 eq. of EDA-oligo; set 3: with splint, 2 eq. of EDA-oligo; set 4: with splint, 1 eq. of EDA-oligo. Bottom: Mass determination of corresponding peaks from UPL chromatograms. (A) RNA (calc. mass = 13254; found mass = 12988 – 3'-(n-) phosphate, 13269 – aldehyde hydrate, 13283 – dihydrate). (B) splint (calc. mass = 10134; found mass = 10134). (C) EDA-oligo (calc. mass = 25281; found mass = 25294 – mobile phase adduct). (D) RNA-EDA-oligo conjugate (calc. mass = 38500; found mass = 38487, 38655 – mobile phase HFIP adduct).

We next attempted to increase the reaction time in order to push the conversion to completion. By incubating the oxidised RNA and splint with 1 eq. of EDA-oligo for 4 and 6 hours, it was found that the conversion did not improve significantly with an increase in incubation time (Figure 4.4). It was hypothesised that after two hours of reaction, the reactive RNA-dialdehyde intermediate has degraded or been inactivated. Indeed, mass analysis on the peak at $R_T = 2.88 - 2.94$ corresponding to the RNA starting material indicates only unreactive species such as 3'-phosphate (n-1) RNA, aldehyde-hydrate and dihydrate were present. This is in agreement to the observations found in Chapter 3 where a majority of oxidised RNA has undergone cleavage at the 3'-end nucleoside via β -elimination by deprotonation at the 4'-position. To improve product conversion in this conjugation strategy, we moved on to investigate means to increase the reaction kinetics over such competing degradation pathways.

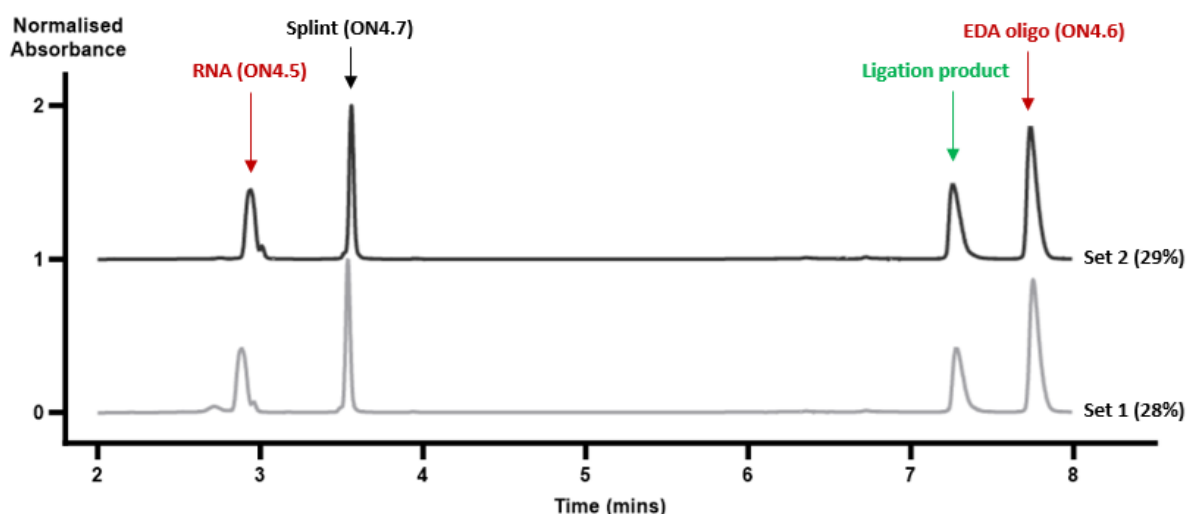


Figure 4.4 | UPLC results of RNA 3'-conjugation for extended reaction times

Overlaid UPL chromatograms show the retention profiles of the crude reaction mixtures after oxidation-reductive amination with 1 eq. loadings of EDA-oligo for 4 and 6 hours. Before injecting to UPLC-MS, to the mixture was added Et₃N.EDTA (final conc. = 50 mM) to remove sodium adducts at early elution. Conversion % in parentheses was calculated as the ratio between area under the product peak and that of the RNA starting material and product with triplicate experiments (n = 3). General reaction condition: RNA (final conc. = 10 μ M) was first incubated with NaIO₄ (1.5 mM) with a splint (1 eq.) in darkness for 30 minutes, then to the mixture was added various components to a final conc.: KH₂PO₄ buffer (100 mM, pH 6.0), 5'-EDA oligo (5 μ M, 1 eq.) and NaBH₃(CN) (20 mM) at r.t. Set 1: incubation time = 4 hours; set 2: incubation time = 6 hours.

Oligonucleotide	Sequence (5'-to-3')
ON4.5	UUCUGCCUAAUAAAAACAUUUAUUUUCAUUGCAUUUAAAUG
ON4.6	<u>XGTGCCTGC</u> AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA <u>3</u>
ON4.7	TTTTTGCAGGCACCATTAAATGCAATGAAAAAT
ON4.8	TTTTTTTTGCAGGCACCATTAAATGCAATGAAAAATAAATG
ON4.9	TTTTTTTTGCAGGCACTCATTAAATGCAATGAAAAATAAATG
ON4.10	<u>TTTTTTT</u> <u>TGCAGGCACCATTAAATGCAATGAAAA</u> <u>TAAATG</u>

Table 4.1 | List of oligonucleotide sequences in RNA-EDA-oligo conjugation studies

Modifications are indicated with special characters: X = N-ethyl ethylenediamine; 3 = Cy3; bold underlined = LNA; double underlined = 2'-OMe. ON4.5 is fully RNA.

As it was understood that the reductive amination reaction between the 3'-dial of RNA and 5'-EDA motif competes with hydration and β -elimination pathways, we postulated that the binding of the two reacting oligos with the splint could be a limiting factor to the reaction kinetics which controls product conversion. As a result, we performed a test with various annealing conditions to examine the effect of strand hybridisation on product formation. In this test, we lowered the RNA concentration to 5 μ M as literature suggested that high RNA concentration can increase the viscosity of the solution which has a negative effect on reaction kinetics. When the RNA oligo was first incubated with the splint with KH_2PO_4 buffer (high salt concentration) and annealed from 95°C to 25°C at -0.4°C/s, subsequent additions of the EDA-oligo and $\text{NaBH}_3(\text{CN})$ furnished a product conversion of 25%. Comparable levels of product conversion was observed when the RNA, splint and EDA-oligo were incubated altogether at the annealing step at -0.4°C /s (23%) and at -0.1°C /s (22%), respectively (Figure S3). From this experiment, it was concluded that the additional annealing step did not improve the reaction kinetics significantly as about half of the RNA degraded under these conditions.

We next moved to study the effect of enhancing binding between the splint and the two reacting oligos by modifying its design. It was hypothesised that the increase in binding affinity may kinetically favour the two reacting groups approaching each other and hence increase the reaction rate. To this end, three splint designs which include a longer binding arm were

synthesised and purified. They include a full DNA splint (ON4.8), the one with an internal T spacer (ON4.9) and the one with 6 LNA nucleosides incorporated (ON4.10). In the UPLC-MS analysis, we devised a smaller gradient to separate the starting materials. In the presence of the short DNA splint, annealing from 95°C to 25°C, followed by oxidation-reductive amination transformed 37% of the RNA into the desired conjugated product. Longer splints with and without an internal spacer provided a conversion at comparable 32% and 34%, respectively. In all these cases, we observed the formation of aldehyde-hydrate and (n-1) cleavage by-product which arose from the oxidation of RNA. However, when the reactants were incubated and annealed with the long splint with LNA incorporation, the conversion dropped to 5% and the majority of RNA species remained as un-oxidised 3'-diol. We rationalised this as the structurally constrained ribose of LNA decreases the flexibility of LNA-RNA duplex and hence reduces the accessibility of the 3'-diol to the oxidants, resulting in incomplete oxidation before $\text{NaBH}_3(\text{CN})$ was added to quench the reaction. Hence, the LNA-containing splint resulted in a much lower conversion.

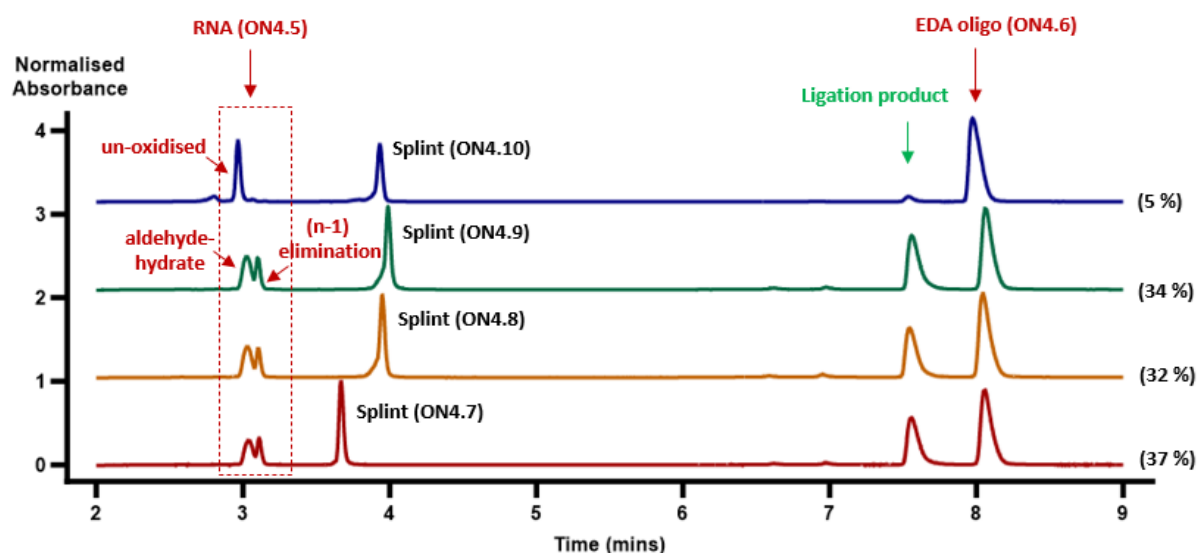


Figure 4.5 | UPLC results of RNA 3'-conjugation with various splint designs

Overlaid UPL chromatograms show the retention profiles of the crude reaction mixtures after oxidation-reductive amination with 1 eq. loading of different splints. General reaction condition: RNA (final conc. = 5 μM) was first incubated with KH_2PO_4 buffer (100 mM, pH 6.0), NaIO_4 (1.5 mM), the splint (1 eq.) and EDA-oligo (1 eq.) in darkness for 30 minutes, then to the mixture was added $\text{NaBH}_3(\text{CN})$ to a final conc. of 20 mM at r.t. for an incubation of 2 hours.

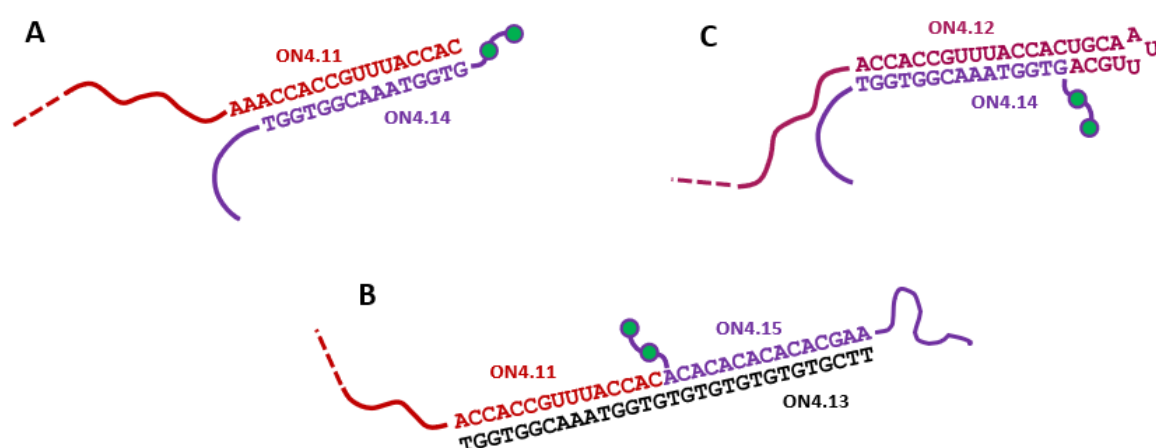
4.3.3 Self-Templated Ligation

Since there is no significant improvement in reaction kinetics when the splint is modified to improve binding with reactants, we devised a new strategy to bring the two reacting oligos in close proximity in a self-templated manner. To achieve this, two sequences on the RNA strand have been designed to include a 3'-region which is complementary to the 5'-region of the EDA-oligo. One of the sequences contains a trinucleotide loop and a hairpin structure (ON4.12) whereas the other does not (ON4.11) (Table 4.2). When incubated in sodium phosphate buffer, the two oligos will hybridise to form a hairpin structure and position the reactive EDA and dialdehyde groups in vicinity (Scheme 4.3).

Oligonucleotide	Sequence (5'-to-3')
ON4.11	AAAAACAUUUAUUUUCAUUGCAUUUAAAUGUUUAAACCACCGUUUACCAC
ON4.12	AAAAACAUUUAUUUUCAUUGCAUUUAAAUGUUUAAACCACCGUUUACCACUGCAAUUUGCA
ON4.13	TTCGTGTGTGTGTGTGTGGTAAACGGTGGT
ON4.14	XGTGGTAAACGGTGGTAAAAAAA
ON4.15	XACACACACACACGAAAAAAA

Table 4.2 | List of oligonucleotide sequences in RNA-EDA-oligo conjugation studies

Modifications are indicated with special characters: X = N-ethyl ethylenediamine. ON4.11 and ON4.12 are fully RNA.



Scheme 4.3 | Oligonucleotide designs for self-templated and splinted conjugation

This scheme illustrates three reaction designs that include an EDA-oligo which hybridises with (A) the 3'-end of the RNA, (B) the splint oligo that brings the RNA in close proximity, and (C) the upstream region of the 3'-hairpin of the RNA. The green icon illustrates the position of the 5'-EDA modification where reductive amination takes place with the neighbouring dialdehyde.

To compare between self-templated and splint-templated systems, we also used the identical RNA (ON4.11) to perform reductive amination with an EDA-tail in the presence of a splint (ON4.13). For these experiments, the RNA and EDA-oligo are in 1:1 mole equivalent and the concentrations of other reagents remained unchanged from previous experiments. We incubated the mixture at various temperatures (i.e., 25°C, 37°C and 45°C) after the reductant was added to study the impact of temperature on reaction progress. It was found that, in a bimolecular system where the EDA-oligo is recruited via self-templating, the ligation efficiency increased from 27% to 42% when the incubation temperature rose from 25°C to 45°C (Figure 4.6). However, the efficiency was not superior to the trimolecular system where a splint oligo brings together two reacting groups by complementary base-pairing with oligos. The conversion in this splinted system was between 47 - 48% at all temperatures. The most efficient ligation took place when the RNA with an internal 3'-hairpin was incubated with the EDA-oligo. The reaction with this couple furnished a conversion to product from 64% (at 25°C) to 70% (at 37°C).

It is believed that in the first design (A) where the two reacting oligos hybridise and are aligned in parallel, the EDA moiety is not situated in a favourable position to react with the neighbouring dialdehyde and thus demands a higher entropic cost for reaction. This resulted in a slower reaction rate than the splinted design (B) where the two reactive groups are positioned heads-on. Finally, the hairpin design (C) harnesses benefits of both (A) and (B) by allowing a heads-on configuration between the EDA and dialdehyde groups in a bimolecular, self-templated hybridisation mode. This resulted in the highest reaction rate as bimolecular binding is expected to be faster and the reactants are pre-organised in a favourable configuration.

Lastly, we attempted to further improve the conversion by increasing the loading of the EDA-oligo. Previously, no improvement was observed when the loading of the EDA-oligo increased. It was rationalised that a high concentration of EDA-oligo might non-specifically

bind to the splint and inhibit its binding to the RNA. In a self-templated system, however, the increase in loading should not interfere with target binding. We attempted to test this with design (C) by adding 1.0, 2.0 and 5.0 eq. of EDA-oligo to the reaction mixture. After 2 hours of incubation at 25°C, no significant enhancement in conjugation efficiency was observed among three sets, at between 65 – 78% (Figure 4.7). We thus conclude that the improvement on conversion is marginal when higher equivalents of the EDA reactant are loaded.

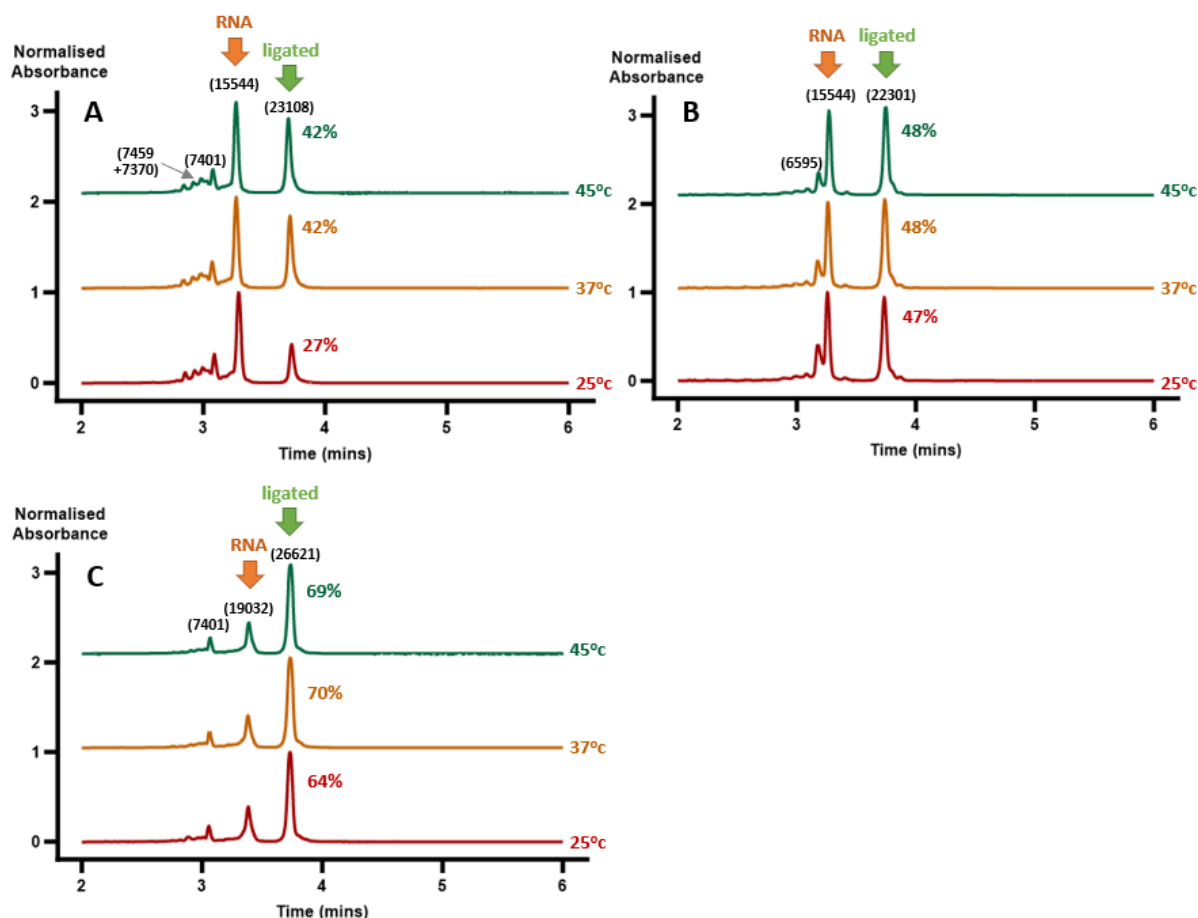


Figure 4.6 | UPLC results of RNA 3'-conjugation of self-templated and splinted designs

Overlaid UPL chromatograms show the retention profiles of the crude reaction mixtures after oxidation-reductive amination with various RNA and EDA-oligo designs at different temperatures. General reaction condition: RNA (final conc. = 5 μ M) was first incubated with KH_2PO_4 buffer (100 mM, pH 6.0), NaIO_4 (1.5 mM), the splint (1 eq. or none) and EDA-oligo (1 eq.) in darkness for 30 mins, then to the mixture was added $\text{NaBH}_3(\text{CN})$ to a final conc. of 20 mM at r.t. for an incubation of 2 hours at various temperatures. The chromatograms correspond to the oligonucleotide mixtures illustrated in Scheme 4.3. The deconvoluted mass is indicated in parentheses. Conversion is shown next to the product peak.

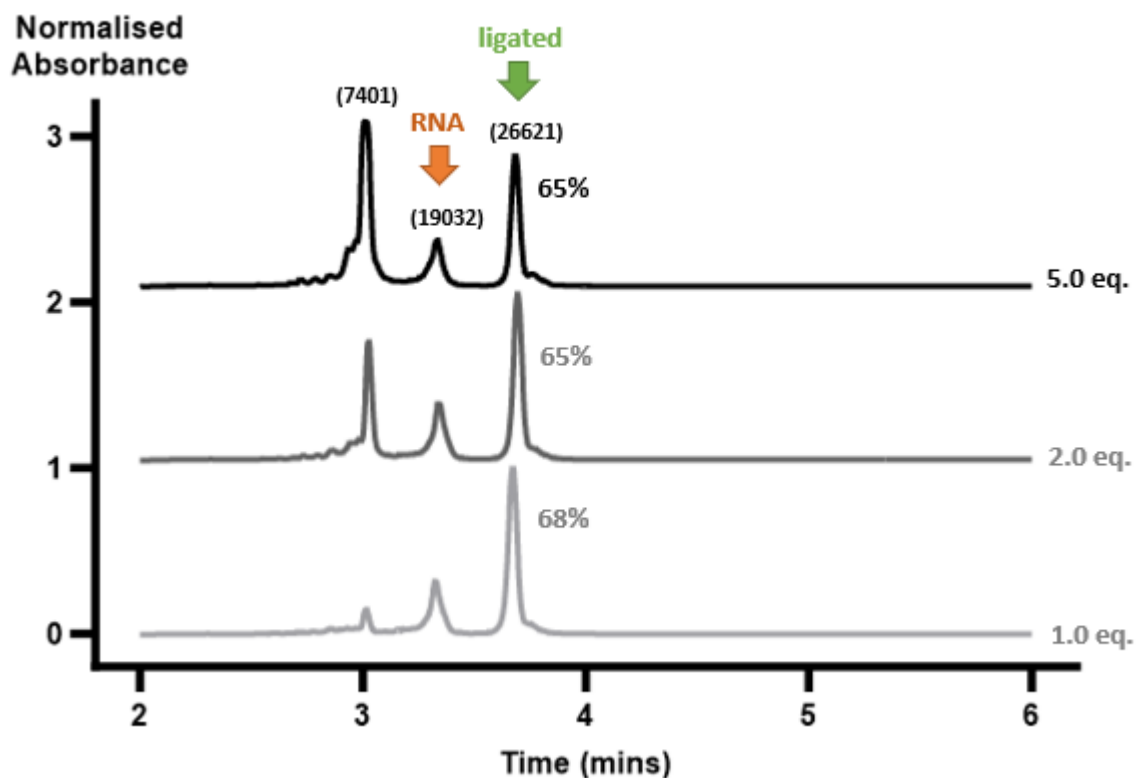


Figure 4.7 | UPLC results of self-templated RNA 3'-conjugation at various EDA-oligo loadings

Overlaid UPL chromatograms show the retention profiles of the crude reaction mixtures after oxidation-reductive amination with various loadings of EDA-oligo. General reaction condition: RNA (final conc. = 5 μ M) was first incubated with KH_2PO_4 buffer (100 mM, pH 6.0), NaIO_4 (1.5 mM), the splint (1 eq.) and EDA-oligo (1, 2 or 5 eq.) in darkness for 30 minutes, then to the mixture was added $\text{NaBH}_3(\text{CN})$ to a final conc. of 20 mM at r.t. for an incubation of 2 hours. The deconvoluted mass is indicated in parentheses. Conversion is shown next to the product peak. $n = 3$.

4.4 Conclusions

Site-specific introduction of modified nucleotides on RNA holds significant potential in promoting its pharmacological properties such as serum stability and tuning of metabolism and distribution. Long RNAs, however, have been proven challenging to modified due to inaccessibility by conventional SPS. Combining the merits of *in vitro* transcription and the simplicity of reductive amination, we successfully demonstrated the conjugation of a synthetic, modified oligo with an RNA species via chemical transformation with good efficiency.

To commence the study, we designed an ethylenediamine phosphoramidite with trifluoroacetamide protecting groups that can be modularly introduced on synthetic oligonucleotides by SPS. This novel amidite has shown good coupling efficiency to generate 5'-EDA-modified oligos to be deprotected by standard ammonia treatment and decent stability after 3 months of storage at -20°C. Model RNA fragments were designed to mimic the 3'-end of mRNA and prepared by TBDMS-protected amidites. Both oligos could be efficiently purified on RP-HPLC system.

This study has demonstrated that good conversion efficiency (>41%) can be afforded by incubating an oxidised RNA fragment with a 5'-EDA-modified oligonucleotide in the presence of a splint. By such means, it was shown that DNA and 2'-OMe-modified oligonucleotides, and those with 3'-Cy3 modification, can be conjugated to the 3'-end of RNA via a morpholino-like linkage. Due to the templated nature of reductive amination between the 3'-dialdehyde and 5'-EDA, it was found that increasing the equivalence of 5'-EDA oligo beyond equimolarity does not improve efficiency as all available reactive groups are positioned in close proximity by complementary base-pairing under the reaction conditions. However, analysis by mass spectroscopy indicated that a majority of RNA underwent degradation, such as hydration or strand cleavage, as competing reaction pathways to reductive amination. This

highlights the need to further increase the rate of reaction in order to improve conversion efficiency.

Through iterations of reaction design, we identified that 3'-hairpin-containing RNA performs better in reductive amination with 1 equivalent of 5'-EDA-oligo bearing complementarity with the 3'-region of the RNA. The high conversion efficiency (up to 70%) was attributed to the self-templated nature of the two reacting oligos which hybridise at a higher rate than a trimolecular system with the splint. At a higher temperature, the conversion was found to marginally improve but increasing the equivalence of 5'-EDA-oligos similarly saw no improvement in reaction efficiency. In general, the efficient ligation of oxidised RNA and 5'-EDA-oligos proves possible the concept of generating RNA conjugates with various modifications at the 3'-end. The modularity of incorporating the 5'-EDA moiety and simplicity of this protocol render it applicable to larger scale syntheses and screening applications.

4.5 Future Outlook

The idea of conjugating site-specifically modified oligos to RNA has potential to help scientists develop RNA-based therapeutics with better properties and also understand the molecular dynamics of RNA molecules *in vivo*, especially for mRNA. The development of such methodology is valuable to advance RNA research. Based on the findings from the current study, we conceived that further study may be carried out in the following directions.

First, this reaction protocol can be extended to modifying mRNA with synthetic 5'-EDA oligos. To achieve this, the 3'-end sequence of an mRNA sequence can be edited via recombinant DNA which is a DNA vector with the desired synthetic sequence inserted via enzymatic reaction. To remove random insertion of nucleotides or polyadenylation at the 3'-end, an additional 3'-homogenisation step can be introduced via sequence-specific cleavage by DNazyme to leave the 3'-end sequence as desired.²³³ Such work has been undertaken by our

group member Sebastian Golojuch who demonstrated 3'-homogenisation of IVT mRNA from recombinant DNA. The flexibility to introduce specific sequence at the 3'-end of the RNA renders it possible to form a short hairpin structure similar to the oligo design (C) in Scheme 4.3 and perform self-templated reductive amination with synthetic 5'-EDA-oligos.

Second, this methodology offers a useful platform to conjugate synthetic oligonucleotides with different modification patterns on RNA. This can be harnessed to screen a panel of chemical modifications with mRNA that can confer stability without compromising translation efficiency. As evidence emerged that non-specific incorporation of phosphorothioate backbone on the 3'-poly(A) tail produces hybrid mRNA construct which is tolerated by ribosomes as well as enhances translation efficiency, we envisage other types of modifications (i.e., ribose and backbone) such as 2'-OMe RNA, 2'-F and phosphoramidate backbone may effect similar or higher enhancement in mRNA stability and translation performance.²³² The methodology that we developed enables a streamlined process of screening oligo conjugates with various modifications as the 5'-EDA oligos can be modularly produced by SPS and they can be subjected to identical reaction conditions with the mRNA model. This improves throughput by avoiding multiple synthetic/procedural designs for various candidates.

Chapter 5 – Materials and Methods

5.1 Experimental Procedures

5.1.1 UV Melting Studies

UV DNA melting curves were recorded on a Cary 4000 Scan UV-Visible Spectrophotometer using complementary oligonucleotides (3 μ M each strand) in a 10 mM phosphate buffer containing 200 mM NaCl at pH 7.0. Samples were melted by heating to 85°C (10 °C/min) and then annealed by slowly cooling to 20 °C (1 °C/min). Six successive cycles (including heating and cooling) were then performed at a gradient of 0.5 °C/min and the change in UV absorbance at 260 nm was recorded. The melting temperature of each experiment was calculated from a smoothed plot of dA/dT (first derivative of the melting curve) from six ramps using in built software.

5.1.2 Foetal Bovine Serum Assay

Oligonucleotides (ON2.6, ON2.8 and ON2.15; 5 nmole of single-stranded oligo) were dissolved in 50 μ L of Dulbecco's PBS and 50 μ L of FBS (Gibco). This was then vortexed and 20 μ L were removed (t = 0 min; control), mixed with formamide 1/1 (v/v), snap-frozen and kept at –20 °C. The remaining solution was incubated at 37 °C and 20 μ L aliquots were removed at specified time points, mixed with formamide 1/1 (v/v), snap-frozen and kept at –20 °C. Samples were analysed by PAGE (300 V) at 8 – 20% depending on size and viewed under short wavelength UV light using a G:Box (Syngene).

5.1.3 RNase H Assay

Escherichia coli RNase H (NEB) was used for all RNase H studies. Oligonucleotides (0.4 μ M for ON2.6, ON2.8 and ON2.15) and corresponding FAM-labelled RNA targets (2.0 μ M) and *E. coli* RNase H (20 U/mL) were incubated at 37 °C or 20 °C in 1 $\times$ RNase H reaction buffer (50 mM Tris-HCl, 75 mM KCl, 3 mM MgCl₂, 10 mM dithiothreitol, pH 8.3). At each time point, 13 μ L aliquots were taken and quenched with aq EDTA-Na₂ (2 μ L, 500 mM) at a

concentration of 66.7 mM, denatured with formamide 1/1 (v/v), snap-frozen and kept at –20 °C in the dark. The aliquots were submitted to a 20% PAGE (400 V) and analysed using a G:Box (Syngene) with an excitation wavelength of $\lambda = 460$ nm and an emission filter of $\lambda = 516$ –600 nm for detection of the FAM label.

5.1.4 In Vitro Cleavage Assay

For DNA target cleavage, an RNP mixture (100 nM) was prepared by mixing water (30 μ L), NEBuffer™ r3.1 (100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl₂, 100 μ g/ml Recombinant Albumin, pH 7.9 at 25 °C, NEB) (10x, 10 μ L), gRNA (1 μ M, 5 μ L) and TrueCut™ Cas9 Protein v2 (Invitrogen™) (1 μ M, 5 μ L) and incubated at 25 °C for 5 minutes. The DNA target was prepared as a 20 nM stock solution by mixing *EMX1* amplicon (100 nM, 30 μ L) and water (120 μ L) and aliquoted into separate PCR vials (5 μ L each), to which the RNP mix (5 μ L) was added and the mixture was incubated at 37 °C in for the designated durations. At each timepoint, Cas9 was deactivated by heating the mixture to 65 °C for 5 minutes before cooling down to room temperature. The RNP was then digested by the addition of proteinase K (0.5 μ L, NEB) and RNase A (0.5 μ L, NEB) and incubation at room temperature for 5 minutes. The reaction (1 μ L) was analysed using an Agilent Bioanalyzer DNA 12000 kit (Agilent Technologies, 5067-1508 for DNA chips and 6067-1509 for reagents), where appropriate, according to the manufacturer's protocol.

5.1.5 Cell Culture

HEPG2 cells (regularly tested for mycoplasma contamination) were cultured in Dulbecco's modified eagle media (DMEM, high glucose) (Thermo Fisher Scientific) supplemented with 10% (v/v) FBS (Thermo Fisher Scientific) in a humidified incubator at 37 °C with 5% CO₂.

5.1.6 Naked Transfection to Hepatocytes and T7E1 Assay

HEPG2 cells were seeded on a 24-well plate at a density of 1.0×10^5 cells/well. After 24 h, the growth media (DMEM, high glucose) was replaced with Opti-MEM reduced serum media and transfected with 6.25 μ M RNP complex prepared from TrueCut™ Cas9 Protein v2 (Invitrogen™):gRNA in RNP buffer (HEPES-NaOH (20 mM, pH7.5), NaCl (0.15 M), MgCl₂ (1.0 mM), 10% glycerol) to designated RNP concentration of 50, 250 and 1250 nM in 200 μ L of Opti-MEM media. After 24 hours of incubation, the cells were harvested and genomic DNA was extracted by PureLink™ Genomic DNA Mini Kit (Invitrogen™) according to manufacturer's protocol. The genomic DNA was amplified with *EMX1* primers (See appendix). The *EMX1* amplicon products were aliquoted into c.a. 100 ng in 8 μ L, to which NEBuffer™ 2 (10x, 1 μ L, NEB) was added. The amplicon solution was re-annealed by heating to 95 °C in 5 minutes and cooling down to 85 °C (ramp rate = -2 °C /s) and then 25 °C (ramp rate = -0.1 °C /s) to form heteroduplexes which were subjected to digestion by T7 endonuclease I (10 U/ μ L, 1 μ L, NEB) for 15 minutes at 37 °C. The reaction was terminated by adding 1 μ L of 0.5 M Na-EDTA (pH 7.0) and 1 μ L was aliquoted for analysis using by an Agilent Bioanalyzer DNA 12000 kit (Agilent Technologies), where appropriate, according to the manufacturer's protocol. Alternatively, the reaction (3 μ L) was analysed by mixing with water (2 μ L) and blue gel loading buffer (6x, 1 μ L, NEB), loaded on a 2 % agarose gel stained with 0.5x SYBR gold and run in 1x TBE running buffer at 60 V for 1 hour.

5.1.7 Cell Preparation for Microscopy

HEPG2 cells were seeded on a Ibidi® 8-well chambered slide (Ibidi) at a density of 6.0×10^4 cells/well. After 24 hours, the growth media (DMEM, high glucose) was replaced with Opti-MEM reduced serum media and transfected with 5 μ M RNP complex prepared from eGFP-Cas9 (Sigma Aldrich):ON3.5-EDA-GalNAc-Cy5 in RNP buffer (HEPES-NaOH (20 mM, pH7.5), NaCl (0.15 M), MgCl₂ (1.0 mM), 10% glycerol) to a final RNP concentration of

50 nM in 200 μ L of Opti-MEM media. In the experiment where ppTG21 was used, 20 μ L of Opti-MEM media was replaced with 20 μ L of 100 μ M ppTG21 solution (in 5% DMSO in water) to a final concentration of 10 μ M. After 2 hours of incubation, the Opti-MEM media was replaced with growth media.

5.1.8 Live Cell Imaging

Live-cell imaging was performed using Zeiss LSM 780 confocal microscope system with a 63x 1.4 Numerical Aperture (NA) oil immersion objective at 37 °C with 5% CO₂ (Zeiss Ltd). The samples were excited by internal lasers and emission detected as described below. Different channels were imaged using sequential excitation and detection of emission. Images were then processed by ImageJ. Cell nuclei were stained with Hoechst 33342 stain (Thermo Fisher Scientific), which was excited at 352 nm and emission recorded from 455 nm. For eGFP-Cas9, GFP was excited at 488 nm and emission recorded from 493–542 nm. For Cy5-conjugated oligonucleotides, the dyes were excited at 633 nm and emission recorded from 664 nm.

5.2 Synthesis, Purification and Characterisation of Oligonucleotides

5.2.1 Synthesis of Oligonucleotides

Standard DNA and RNA phosphoramidites, solid supports and reagents were purchased from Sigma Aldrich, Link Technologies and Applied Biosystems. Modifier phosphoramidites were purchased from Link Technologies and Glen Research. Automated solid phase synthesis of oligonucleotides (trityl off) was performed on an Applied Biosystems 394 synthesiser. Synthesis was performed on a 0.2 or 1.0 μ mol scale involving cycles of acid-catalysed detritylation, coupling, capping, and iodine oxidation. Standard DNA phosphoramidites were coupled for 60 seconds, TBDMS-RNA and chemically modified phosphoramidites for 10 minutes. Coupling efficiencies and overall synthesis yields were

determined by automated trityl cation conductivity monitoring facility. For on-resin functionalisation, the oligonucleotide was detritylated on solid support after the final coupling cycle. To cleave the oligonucleotides from solid support with removal of nucleobase and backbone protecting groups, the resin was exposed to conc. NH_4OH for 60 minutes at r.t. followed by heating in a sealed glass vial at 55 °C for 5 hours, unless otherwise specified for modified oligonucleotides. The crude oligonucleotides in solution were then concentrated *in vacuo*, re-dissolved in water and desalted using NAPTM-10 column (G.E. Healthcare) prior to downstream lyophilisation or purification.

For oligonucleotides with 5'-EDA modification, EDA phosphoramidite C4 was coupled for 10 minutes. Prior to cleavage and deprotection, the resin was washed with 20% DEA in MeCN for 1 hour to remove cyanoethyl protecting groups and rinsed with MeCN (5 mL) three times.

5.2.2 Deprotection of 2'-OTBDMS Protecting Groups on RNA

TBDMS-protected RNA oligonucleotides were concentrated under reduced pressure until formation of a turbid solution and lyophilised overnight. The residue was then resuspended in a mixture of DMSO (300 μL) and $\text{Et}_3\text{N}\cdot 3\text{HF}$ (300 μL) and incubated at 65 °C for 2.5 hours. Next, sodium acetate (3 M, 50 μL) and n-butanol (3 mL) were added to the sample and the solution was stored at -80 °C for a minimum of 1 hour. After thawing the suspension was pelleted under centrifuge (13,000 rpm, 4°C, 10 minutes) and the supernatant decanted. The precipitate was then washed with 70% v/v ethanol ($2 \times 750 \mu\text{L}$), centrifuged at identical speed and then dried under vacuum.

5.2.3 Oligonucleotide 5'-Azide Conversion on Solid Support

After solid-phase synthesis, the resin-bound 5'-detritylated oligonucleotides were treated with methyltriphenoxyposphonium iodide (0.5 M) in anhydrous DMF (1 mL) for 1

hour at r.t. The solid support was then washed with dry DMF (3×1 mL) and dry MeCN (1 mL) and dried with argon. A saturated solution of sodium azide was then prepared by heating sodium azide (100 mg) in dry DMF (2 mL) for 5 minutes at 70 °C. After cooling to rt, the resin was treated with the azide solution for 5 hours at 55 °C. The resin was then washed with DMF (3×1 mL) and MeCN (1 mL) and dried with argon. The 5'-azido oligonucleotides were cleaved from solid support and deprotected as described previously.

5.2.4 Copper-Catalysed Click-Crosslinking of 5'-Azido Oligonucleotides

After 5'-azide conversion, the resin containing the 5'-azido oligonucleotides (1 μ mol) was washed with dry MeCN (1 mL) and dried thoroughly with argon. In the optimised click conditions, a click-reaction solution was prepared in a vial by mixing L3OH (5 mM, 100 μ L), TEAA buffer (1.0 M, pH 7.0, 100 μ L), Cu-THPTA solution (10 mM: 20 mM mixture, 100 μ L), DMSO (500 μ L) and water (150 μ L). The catalyst was activated by the addition of sodium ascorbate (100 mM, 50 μ L). Next, the mixture was immediately degassed under a gentle stream of argon for 30 seconds and transferred to a 1-mL syringe. An empty syringe was connected to one end of the synthetic column containing the resin, and the reaction mixture was then passed through the column by plunging the syringes back and forth for 10 times or until the resin was completely submerged (i.e., when no air bubbles were observed from the column). The reaction set-up was incubated in the oven at 45 °C for 4 hours and the resin was agitated by gentle plunging every 60 minutes. After the reaction, the mixture was cooled down to r.t., washed with DMF (3×1 mL) and MeCN (3×1 mL) and dried under argon. The crosslinked oligonucleotides were cleaved from solid support and deprotected as described previously.

5.2.5 NHS Ester Labelling

Freeze-dried oligonucleotide (20 nmol) was dissolved in NaHCO₃ buffer (0.5 M, pH 8.5, 150 μ L) and mixed with the corresponding NHS ester reagent (20 mM in DMSO, 100 μ L).

The reaction was then left for 4 hours at 25 °C with 750 rpm shaking. After dilution with water, the samples were desalted using a NAPTM-10 column (G.E. Healthcare) prior to RP-HPLC purification.

5.2.6 Azide Labelling by PAP Enzyme

A reaction mixture (20 µL) was prepared by diluting stock solutions of RNA oligonucleotide into a final concentration of 16 µM together with Yeast poly(A) polymerase (Jena Bioscience, 600 U/µL) (to 27.6 U/µL), Yeast PAP reaction buffer (Jena Bioscience, 5 x) (to 1 x), and 2'-azido-2'-dATP (Jena Bioscience, 20 mM) (to 1 mM). The mixture was incubated at 25 °C overnight. The RNA was isolated by precipitation on MEGAclearTM Transcription Clean-Up Kit (Thermofisher) and analysed by UPLC-MS.

5.2.7 Purification by HPLC

After complete deprotection, oligonucleotides were purified on Agilent HPLC system via RP-HPLC or IEX-HPLC methods. In RP method, Phenomenex Kinetex EVO C18 column (5.0 µm particle size, 10 mm × 250 mm) was used with a gradient of MeCN in aq. triethylammonium acetate (TEAA) buffers (Buffer A: 0.1 M TEAA, pH 7.0; buffer B: 0.1 M TEAA, 50% v/v MeCN, pH 7.0). Buffer B was increased from 20 - 40% or 20% - 60% or 20% - 80% over 30 min at a flow rate of 5 mL·min⁻¹ at 60 - 75°C, and elution was monitored by UV absorption at 260 nm with a 10.0 mm or 0.3 mm flow cell. In IEX method, Thermo Fisher DNAPac PA200 column (9.0 mm × 250 mm) was used with a gradient of aq. NaCl in Tris-HCl buffers (Buffer A: 20 mM Tris-HCl, pH 8.5; buffer B: 1.0 M NaCl, 20 mM Tris-HCl, pH 8.5). Buffer B was increased from 20–80% over 60 minutes at a flow rate of 1.2 mL·min⁻¹ at 75°C, and elution was monitored by UV absorption at 260 nm with a 10.0 mm or 0.3 mm flow cell.

5.2.8 Purification by Gel Electrophoresis

Oligonucleotide samples were first denatured in 1:1 v/v formamide/water at 95 °C for 5 minutes (for DNA) or 80 °C for 5 minutes (for RNA). Then, they were loaded on denaturing PAGE at specific % and run at 20 W (ca. 1 nmol of oligonucleotide per lane) and the gel was visualised under UV. Bands of interest were excised, crushed and incubated in a ThermoMixer (Eppendorf) at 37 °C, 950 rpm in water (ca. 1 mL per band) for 16 hours. The gel residue was then filtered off and the filtrate was concentrated under reduced pressure. Oligonucleotides were desalted using illustra™ NAP-10 columns (GE Healthcare) before MS analysis.

5.2.9 Characterisation by Mass Spectrometry

Oligonucleotide MS was recorded on a UPLC-MS Waters XEVO G2-QTOF (ESI⁻) using an ACQUITY UPLC ON BEH C18 column, 130 Å (1.7 µm particle size, 2.1 mm × 50 mm). Mobile phase was a gradient of triethylamine (Et₃N) and hexafluoroisopropanol (HFIP) in methanol/water buffers (Buffer A: 8.60 mM Et₃N, 200 mM HFIP in 5% v/v MeOH/H₂O; buffer B: 20% buffer A in MeOH). The gradient was formulated by increasing the content of buffer B from 0 – 70% over 8 minutes at a flow rate of 0.2 mL·min⁻¹ at 55 °C. A specific gradient was formulated for the analysis of RNA conjugates in Chapter 4 by increasing the content of buffer B from 15 – 35% over 8 minutes at a flow rate of 0.3 mL·min⁻¹ at 55 °C.

5.2.10 Sampling of Oligonucleotide Samples from CPG

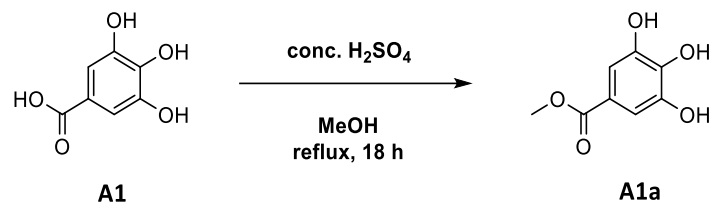
Analytical samples of oligonucleotide after solid-phase syntheses/reactions were obtained by aliquoting 0.8 ~ 1.0 mg of CPG resin from the synthetic column and subjecting it to treatment in conc. ammonium hydroxide at 55 °C in a heat block for 5 hours. After that, the solution was concentrated *in vacuo* and re-dissolved in 1 mL of water. The crude oligo solution was filtered through a PTFE syringe filter to remove residual resin particles and analysed by UPLC-MS.

5.3 Organic Synthesis

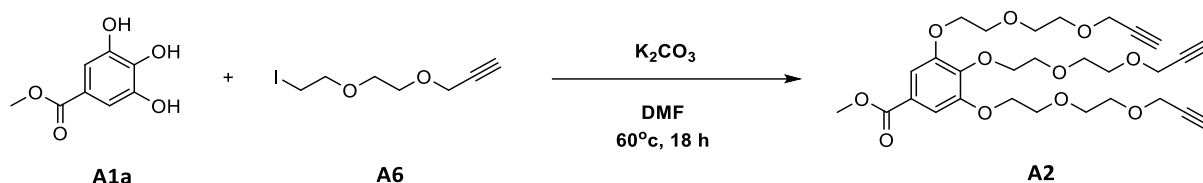
5.3.1 General Procedures

All reagents were purchased from Sigma-Aldrich, Acros Organics, Fisher Scientific, FluoroChem, Jena Bioscience or Alfa Aesar and used without further purification. Dry solvents (pyridine, Et₃N, CH₂Cl₂, THF, MeCN) were obtained using an MBraun SPS Bench Top solvent purification system (SPS). All air/moisture sensitive reactions were carried out under inert atmosphere (argon) in oven-dried glassware. Reactions were monitored by thin layer chromatography (TLC) using Merck Kieselgel 60 F24 silica gel plates (0.22 mm thickness, aluminium backed). The compounds were visualized by UV irradiation at 254/265 nm and by staining in p-anisaldehyde or ninhydrin solution followed by gentle heating. Flash column chromatography was performed using Merck Geduran® 60 Å (40–62 µm) or using a Biotage Sfär column. ¹H NMR (400 MHz, 500 MHz or 600 MHz), ¹³C NMR (101 MHz, 126 MHz or 151 MHz) and ³¹P NMR (162 MHz) spectra were measured on a Bruker AVIIIHD 400, a Bruker AVII 500 (with a ¹³C cryoprobe) or a Bruker AVIII 600 (with a prodigy N2 broadband cryoprobe) spectrometer. ¹H and ¹³C NMR spectra were referenced to the appropriate deuterated solvent signal, chemical shifts are given in ppm. Assignment of the NMR signals was aided by correlation spectroscopy (COSY; ¹H–¹H), heteronuclear single quantum correlation (HSQC; ¹H–¹³C) and heteronuclear multiple bond correlation (HMBC; ¹H–¹³C) experiments in specific experiments. Data was then processed using Mestrenova® software. High resolution mass spectrometry (HRMS) using electrospray ionisation (ESI+) was recorded on a Thermo Exactive High-Resolution Orbitrap FTMS. Data was then deconvoluted using MassLynx v4.1.

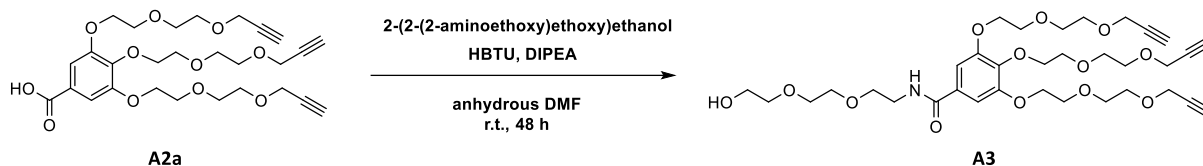
Methyl gallate (A1a)²³⁴



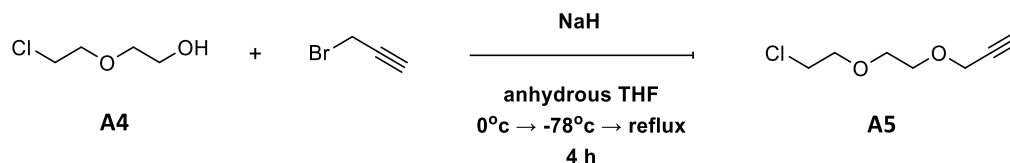
To a flask charged with gallic acid, **A1** (6.00 g, 35.3 mmol) dissolved in methanol (125 mL) was added concentrated sulfuric acid (5 mL, 93.3 mmol). The solution was stirred at 80 °C under reflux for 18 hours. At this point, the solvent was removed under reduced pressure and the crude product was re-dissolved in ethyl acetate (150 mL) and washed with saturated sodium bicarbonate (150 mL). The product was further extracted from the aqueous phase with ethyl acetate (150 mL × 2). The combined organic phase was dried over MgSO₄ and *in vacuo* to afford compound **A1a** (5.68 g, 87%) as an off-white solid. ¹H NMR (400 MHz, DMSO) δ 9.16 (s, 3H), 6.95 (s, 2H), 3.75 (s, 3H).

Methyl 3,4,5-tris(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethoxy)benzoate (A2)

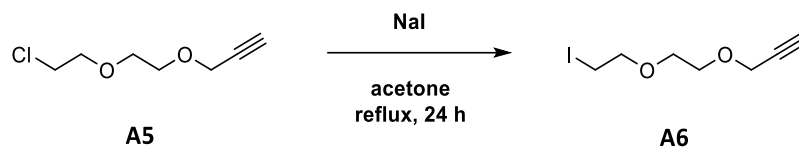
To a flask charged with compound **A1a** (460 mg, 2.5 mmol) and potassium carbonate (1.73 g, 12.5 mmol, 5 equiv.) was added N,N-dimethylformamide (12 mL). The suspension was stirred and heated up to $60^\circ C$. To the mixture was added 3-(2-(2-Iodoethoxy)ethoxy)prop-1-yne, compound **A6** (3.18 g, 12.5 mmol) dropwise over 45 minutes. The mixture was left stirring at $60^\circ C$ for 18 hours and then concentrated under reduced pressure and resuspended in water/brine mixture (1:1 v/v) (120 mL). The crude product was extracted with dichloromethane ($100\text{ mL} \times 3$), dried over $MgSO_4$, concentrated *in vacuo* and purified by column chromatography (SiO_2 , EtOAc/Petroleum ether 50% to 70%) to afford compound **A2** (1.05 g, 75%) as an oil. R_f (EtOAc/Petroleum ether 70%) = 0.50. ^1H NMR (600 MHz, DMSO) δ 7.25 (s, 2H), 4.17 – 4.12 (m, 12H), 3.84 (s, 3H), 3.79 – 3.75 (m, 4H), 3.71 – 3.68 (m, 2H), 3.64 (dd, $J = 5.9, 3.4$ Hz, 4H), 3.61 – 3.57 (m, 6H), 3.55 (dd, $J = 5.7, 3.3$ Hz, 2H), 3.39 (q, $J = 2.3$ Hz, 3H). ^{13}C NMR (151 MHz, DMSO) δ 166.26, 152.46, 142.40, 124.87, 108.65, 80.80, 80.78, 77.51, 77.47, 72.42, 70.37, 70.14, 69.94, 69.38, 69.04, 69.00, 58.00, 57.99, 52.63. HRMS (ESI-TOF): calculated m/z : 563.2487 ($[M+H]^+$, $C_{29}H_{39}O_{11}^+$); found m/z : 563.2486.

N-(2-(2-(2-hydroxyethoxy)ethoxy)ethyl)-3,4,5-tris(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethoxy)benzamide (A3)

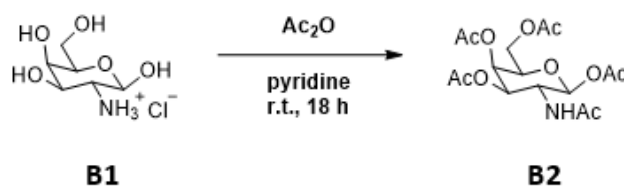
To a solution of compound **A2a** (291 mg, 0.53 mmol) in anhydrous N,N-dimethylformamide (10 mL) were added DIPEA (174 μ L, 3 mmol) and HBTU (212 mg, 0.58 mmol, 1.1 equiv.). The mixture was stirred at room temperature for 30 minutes under argon atmosphere with gradual deepening of colour. Then, 2-(2-(2-aminoethoxy)ethoxy)ethanol (87 mg, 0.58 mmol, 1.1 equiv.) in anhydrous N,N-dimethylformamide (2 mL) was added to the mixture dropwise over 5 minutes. The mixture was stirred at room temperature for 24 hours, at which point the solvent was removed under reduced pressure. The crude product was re-suspended in water (30 mL), extracted with dichloromethane (30 mL $\times$ 3), washed with brine (50 mL $\times$ 1), dried over MgSO_4 and concentrated *in vacuo*. Column chromatography (SiO_2 , MeOH/DCM 5% to 10%) afforded compound **A3** (606 mg, 85%) as an oil. R_f (MeOH/DCM 10%) = 0.40. ^1H NMR (500 MHz, DMSO) δ 8.46 (t, J = 5.7 Hz, 1H), 7.18 (s, 2H), 4.57 (s, 1H), 4.14 (dd, J = 4.9, 2.4 Hz, 10H), 4.08 – 4.05 (m, 2H), 3.78 – 3.75 (m, 4H), 3.69 – 3.66 (m, 2H), 3.65 – 3.62 (m, 4H), 3.61 – 3.50 (m, 15H), 3.47 (d, J = 5.7 Hz, 2H), 3.42 – 3.38 (m, 6H). ^{13}C NMR (126 MHz, DMSO) δ 166.23, 152.36, 141.17, 130.06, 107.78, 80.89, 80.87, 77.02, 76.98, 72.88, 72.47, 70.39, 70.31, 70.27, 70.24, 70.05, 69.60, 69.34, 69.21, 60.90, 58.11, 31.78. HRMS (ESI-TOF): calculated m/z : 680.3277 ($[\text{M}+\text{H}]^+$, $\text{C}_{34}\text{H}_{50}\text{NO}_{13}^+$); found m/z : 680.3262.

3-(2-(2-Chloroethoxy)ethoxy)prop-1-yne (A5)²³⁵

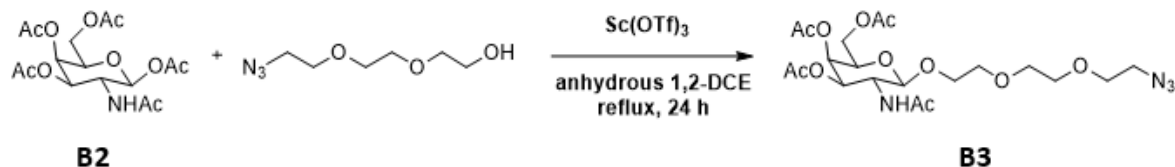
To an oven-dried flask was loaded sodium hydride (60% in mineral oil) (5.50 g, 138 mmol, 2 equiv.) under an argon atmosphere. While the flask was cooled down in an ice bath and kept at 0 °C, anhydrous tetrahydrofuran (140 mL) was transferred to the flask through a cannula. After 10 minutes of stirring at 0 °C, the ice bath was replaced by an acetone-dry ice bath and the mixture was cooled down to -78 °C. To the suspension was added 2-(2-chloroethoxy)ethanol, compound **A4** (8.59 g, 69 mmol) dropwise over 5 minutes. The effervescing mixture was stirred at -78 °C for 10 minutes, at which point propargyl bromide in toluene (80% w/w) (7.68 mL, 69 mmol, 1 equiv.) was added to the mixture slowly. Then, the mixture was heated up to 85 °C under reflux for 4 hours. After cooling down to room temperature, the reaction mixture was quenched with water (3 mL) and concentrated *in vacuo*. The crude product was re-suspended in water (100 mL) and extracted with dichloromethane (100 mL × 3). The combined organic phase was washed with brine (100 mL), dried over MgSO₄, concentrated *in vacuo* and purified by column chromatography (SiO₂, Et₂O/Petroleum ether 30% to 50%) to afford compound **A5** (8.45 g, 75%) as a yellow oil. *R*_f (Et₂O/Petroleum ether 30%) = 0.40. ¹H NMR (400 MHz, CDCl₃) δ 4.24 (d, *J* = 2.4 Hz, 2H), 3.79 (td, *J* = 5.9, 0.7 Hz, 2H), 3.73 (s, 4H), 3.66 (td, *J* = 5.9, 0.7 Hz, 2H), 2.46 (t, *J* = 2.4 Hz, 1H).

3-(2-(2-Iodoethoxy)ethoxy)prop-1-yne (A6)²³⁵

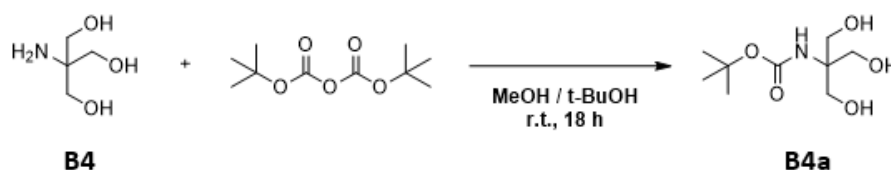
To a solution of compound **A5** (6.20 g, 38.1 mmol) in acetone (70 mL) was added sodium iodide (45.7 g, 300 mmol, 7.9 equiv.). The suspension was allowed to react at 85 °C under reflux for 24 hours. The orange suspension was diluted to 100 mL with acetone and filtered to remove the solids. The filtrate was concentrated under reduced pressure and the crude product was re-suspended in water (80 mL), extracted with ethyl acetate (80 mL $\times$ 3), dried over MgSO_4 and *in vacuo* to afford compound **A6** (9.51 g, 98%) as a brown oil. ^1H NMR (400 MHz, DMSO) δ 4.16 (d, J = 2.4 Hz, 2H), 3.67 (t, J = 6.4 Hz, 2H), 3.58 (q, J = 1.3 Hz, 4H), 3.43 (t, J = 2.4 Hz, 1H), 3.34 (d, J = 6.4 Hz, 2H).

β -D-Galactosamine pentaacetate (B2)²³⁶

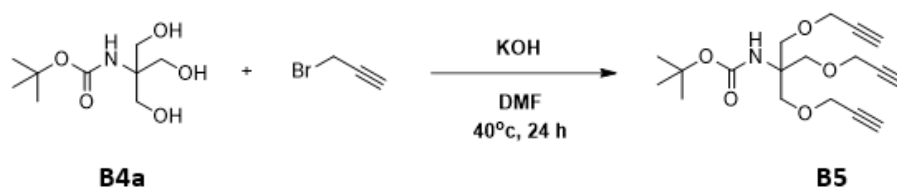
To a flask charged with D-(+)-Galactosamine hydrochloride, **B1** (3.95 g, 18.3 mmol) dissolved in pyridine (22 mL) was added acetic anhydride (20.76 mL, 219.6 mmol, 12 equiv.). The solution was stirred at room temperature for 18 hours. At this point, the reaction mixture was poured into a beaker filled with ice cold water (90 mL) and stirred for 45 minutes. The white precipitate formed was collected over a glass frit, washed with water (50 mL $\times$ 2) and diethyl ether (50 mL $\times$ 2) and dried under reduced pressure to afford compound **B2** (6.352 g, 89 %) as white powder. ^1H NMR (400 MHz, CDCl_3) δ 5.70 (d, J = 8.8 Hz, 1H), 5.41 – 5.34 (m, 2H), 5.08 (dd, J = 11.3, 3.3 Hz, 1H), 4.50 – 4.40 (m, 1H), 4.20 – 4.09 (m, 2H), 4.02 (t, J = 5.9 Hz, 1H), 2.17 (s, 3H), 2.13 (s, 3H), 2.05 (s, 3H), 2.02 (s, 3H), 1.94 (s, 3H).

2-(2-(2-Azidoethoxy)ethoxy)ethyl-2-acetamido-3,4,6-tri-O-acetyl-2-deoxy- β -D-galactopyranoside (B3)²³⁷

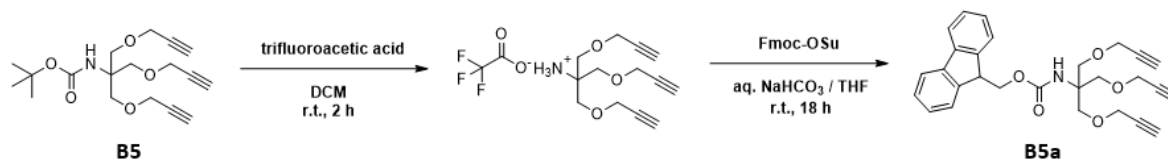
To an oven-dried flask were added **B2** (5.48 g, 14 mmol), 2-(2-(2-azidoethoxy)ethoxy)ethanol (2.96 g, 16.9 mmol, 1.2 equiv.) and scandium (III) triflate (345 mg, 0.7 mmol, 0.05 equiv.). The mixture was kept in an argon atmosphere and dissolved in anhydrous 1,2-dichloroethane (55 mL). The off-white suspension was stirred at 85 °C under reflux for 24 hours with a gradual colour change from yellow to dark brown. Then, the solvent was removed under reduced pressure and the crude product was re-dissolved in dichloromethane (50 mL), washed with saturated sodium bicarbonate (50 mL), dried over MgSO_4 and concentrated *in vacuo*. Purification by column chromatography (SiO_2 , MeOH/EtOAc 2% to 5%) afforded **B3** (4.36 g, 62%) as an orange oil. R_f (MeOH/EtOAc 5%) = 0.30; ^1H NMR (400 MHz, DMSO) δ 7.79 (d, J = 9.2 Hz, 1H), 5.22 (d, J = 3.4 Hz, 1H), 4.98 (dd, J = 11.2, 3.4 Hz, 1H), 4.57 (d, J = 8.5 Hz, 1H), 4.04 (d, J = 4.0 Hz, 3H), 3.93 – 3.85 (m, 1H), 3.82 – 3.75 (m, 1H), 3.64 – 3.59 (m, 3H), 3.57 – 3.51 (m, 6H), 3.41 – 3.38 (m, 2H), 2.11 (s, 3H), 2.00 (s, 3H), 1.90 (s, 3H), 1.78 (s, 3H).

N-(tert-Butyloxycarbonyl)tris(hydroxymethyl)aminomethane (B4a)²³⁸

To a flask charged with Tris, compound **B4** (6.0 g, 49.5 mmol) suspended in methanol/tert-butanol (90 mL, 1:1 v/v) was added a solution of di-tert-butyl carbonate (14.0 g, 64.4 mmol, 1.3 equiv.) in tert-butanol (60 mL). The mixture was stirred for 18 hours at which point the solvent was removed under reduced pressure and recrystallised in hot ethyl acetate (c.a. 500 mL) which was allowed to cool down to room temperature and stay still for 45 minutes. The white solids were collected, washed with cold ethyl acetate (30 mL $\times$ 2) and diethyl ether (30 mL $\times$ 3) and dried under reduced pressure to afford compound **B4a** (9.32 g, 85 %) as white solids. ¹H NMR (400 MHz, *d*-DMSO) δ 4.92 (s, 1H), 4.15 (d, *J* = 2.4 Hz, 6H), 3.79 (s, 6H), 2.42 (t, *J* = 2.4 Hz, 3H), 1.42 (s, 9H).

N-(tert-Butyloxycarbonyl)tris[(propargyloxy)methyl]aminomethane (B5)²³⁸

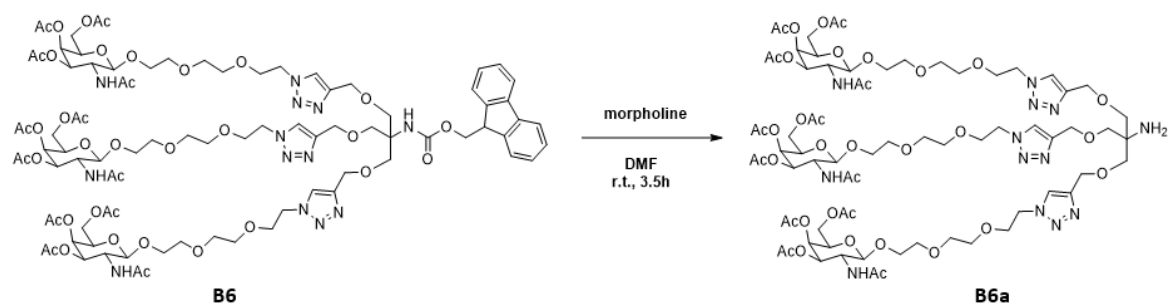
Compound **B4a** (3.0 g, 13.6 mmol) was dissolved in N,N-dimethylformamide (32 mL). To the solution was added propargyl bromide in toluene (80% w/w) (9.09 mL, 81.6 mmol, 6 equiv.) and the mixture was cooled down to 0 °C. To the yellow mixture was added KOH powder (4.58 g, 81.6 mmol, 6 equiv.) portion-wise over 15 minutes at which point it turned into a brown suspension. The suspension was stirred at 40 °C for 24 hours and then diluted to 100 mL with ethyl acetate. The mixture was washed with water (150 mL) and the product was further extracted with ethyl acetate (100 mL × 2), dried over MgSO₄, filtered off and concentrated *in vacuo*. Purification by column chromatography (SiO₂, EtOAc/Petroleum ether 5% to 10%) afforded compound **B5** (3.12 g, 68%) as a pale yellow oil. R_f (EtOAc/Petroleum ether 30%) = 0.58; ¹H NMR (400 MHz, CDCl₃) δ 4.92 (s, 1H), 4.15 (d, *J* = 2.4 Hz, 6H), 3.79 (s, 6H), 2.42 (t, *J* = 2.4 Hz, 3H), 1.42 (s, 9H).

N-(9-Fluorenylmethoxycarbonyl)tris[(propargyloxy)methyl]aminomethane (B5a)

Compound **B5** (2.68 g, 8.00 mmol) was dissolved in dichloromethane (30 mL) and cooled down to 0 °C. To which was added trifluoroacetic acid (10 mL) dropwise over 30 minutes. After stirring at r.t. for 3 hours, the reaction mixture was diluted with toluene to 50 mL and evaporated to dryness. The residual trifluoroacetic acid was removed by co-evaporation with toluene. The crude product was used in the next step without purification. The brown trifluoroacetate salt was dissolved in a mixture of tetrahydrofuran (20 mL) and 0.5M aqueous sodium bicarbonate (20 mL). To which N-(9-fluorenylmethoxycarbonyloxy)succinimide (2.97 g, 8.8 mmol, 1.1 equiv.) was added. The mixture was stirred at room temperature for 18 hours, at which point the solvent was removed under reduced pressure. The crude product was re-dissolved in dichloromethane (80 mL), washed with 0.1 M citric acid (40 mL) and brine (40 mL), dried over MgSO₄ and concentrated *in vacuo*. Purification by column chromatography (SiO₂, EtOAc/Petroleum ether 10% to 30%) afforded compound **B5a** (2.75 g, 75% over two steps) as an off-white oil. *R*_f (EtOAc/Petroleum ether 30%) = 0.30; ¹H NMR (500 MHz, DMSO) δ 7.89 (d, *J* = 7.5 Hz, 2H), 7.73 (d, *J* = 7.4 Hz, 2H), 7.44 – 7.41 (m, 2H), 7.35 (t, *J* = 7.4 Hz, 2H), 6.94 (s, 1H), 4.24 (s, 3H), 4.14 (s, 6H), 3.65 (s, 6H), 3.42 (d, *J* = 2.5 Hz, 3H). ¹³C NMR (126 MHz, DMSO) δ 155.17, 144.38, 141.21, 128.08, 127.56, 125.77, 120.55, 80.65, 77.69, 68.08, 65.67, 58.90, 58.57, 47.18. HRMS (ESI-TOF): calculated *m/z*: 458.1962 ([*M*+*H*]⁺, C₂₈H₂₈NO₅⁺); found *m/z*: 458.1959.

Compound B6

To a flask charged with a solution of compound **B5a** (147 mg, 0.32 mmol) in acetonitrile (1.4 mL) was added a solution of compound **B3** (810 mg, 1.6 mmol, 5 equiv.) in acetonitrile (4.2 mL). The mixture was purged with argon for 15 minutes at room temperature. The catalyst solution prepared by dissolving copper (II) sulfate pentahydrate (20 mg, 80 μmol), THPTA (34 mg, 80 μmol) and sodium ascorbate (31.7 mg, 160 μmol) in water (2.8 mL) and degassing with argon was added to the reaction mixture. After stirring for 48 hours at 40 $^\circ\text{C}$, the crude mixture was concentrated *in vacuo* and directly purified by column chromatography (SiO_2 , MeOH/DCM 5% to 10%). Compound **B6** (330 mg, 52%) was afforded as a white foam after drying *in vacuo* overnight. R_f (MeOH/DCM 10%) = 0.25; ^1H NMR (500 MHz, DMSO) δ 7.92 (s, 3H), 7.86 (d, J = 7.5 Hz, 2H), 7.69 (d, J = 7.5 Hz, 2H), 7.56 (d, J = 9.2 Hz, 3H), 7.41 (t, J = 7.4 Hz, 2H), 7.31 (td, J = 7.5, 1.2 Hz, 2H), 6.36 (s, 1H), 5.26 – 5.23 (m, 3H), 5.05 (dd, J = 11.2, 3.4 Hz, 3H), 4.62 (d, J = 8.5 Hz, 3H), 4.51 – 4.48 (m, 12H), 4.29 (d, J = 6.6 Hz, 2H), 4.21 (t, J = 6.7 Hz, 1H), 4.06 (d, J = 7.2 Hz, 6H), 3.99 (t, J = 6.4 Hz, 3H), 3.92 – 3.88 (m, 3H), 3.84 (t, J = 5.5 Hz, 6H), 3.78 (dd, J = 10.5, 5.7 Hz, 3H), 3.60 (q, J = 6.5 Hz, 9H), 3.54 – 3.49 (m, 18H), 2.10 (s, 9H), 1.99 (s, 9H), 1.90 (s, 9H), 1.78 (s, 9H). ^{13}C NMR (126 MHz, DMSO) δ 170.26, 170.20, 169.88, 169.72, 155.18, 144.48, 141.26, 128.00, 127.49, 125.59, 124.24, 120.41, 101.54, 71.09, 70.67, 70.23, 70.19, 69.99, 69.25, 68.95, 68.71, 67.49, 67.37, 65.88, 64.85, 62.03, 59.56, 50.29, 49.90, 47.40, 31.78, 23.17, 20.80, 20.75, 20.73. HRMS (ESI-TOF): calculated m/z : 1971.8196 ($[\text{M}+\text{H}]^+$, $\text{C}_{88}\text{H}_{124}\text{N}_{13}\text{O}_{38}^+$); found m/z : 1971.8140.

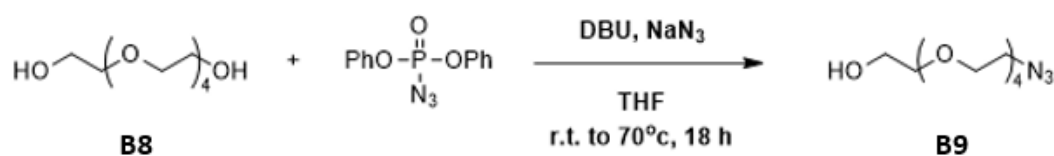
Compound B6a

To a flask charged with compound **B6** (500 mg, 254 μmol) in N,N-dimethylformamide (8.5 mL) was added morpholine (1.5 mL) dropwise over 5 minutes. The mixture was stirred at room temperature for 3.5 hours, at which point the solvent was removed under reduced pressure and the crude mixture was resuspended in water (35 mL). The insoluble precipitate was solubilised and removed by washing with diethyl ether (20 mL $\times$ 3). The aqueous phase remained cloudy at this point. The product, compound **B6a** (440 mg, quant.) was obtained after removing water by co-evaporation with acetonitrile (20 mL) and toluene (20 mL) and dried *in vacuo*. ^1H NMR (600 MHz, DMSO) δ 8.04 (s, 3H), 7.79 (d, J = 9.2 Hz, 3H), 5.22 (d, J = 3.4 Hz, 3H), 4.99 (dd, J = 11.2, 3.4 Hz, 3H), 4.56 (d, J = 8.5 Hz, 3H), 4.53 – 4.48 (m, 12H), 4.06 – 4.01 (m, 9H), 3.91 – 3.86 (m, 3H), 3.82 (t, J = 5.4 Hz, 6H), 3.77 (ddd, J = 11.1, 5.6, 3.8 Hz, 3H), 3.59 – 3.56 (m, 3H), 3.54 – 3.50 (m, 9H), 3.49 – 3.45 (m, 9H), 3.29 (s, 6H), 2.10 (s, 9H), 2.00 (s, 9H), 1.90 (s, 9H), 1.76 (s, 9H). ^{13}C NMR (151 MHz, DMSO) δ 169.97, 169.90, 169.60, 169.26, 160.96, 143.88, 124.08, 100.92, 79.15, 72.08, 70.45, 69.88, 69.65, 69.52, 69.37, 68.68, 68.29, 66.71, 65.73, 64.13, 61.44, 55.67, 49.31, 49.28, 45.07, 22.76, 20.49, 20.42, 20.40.

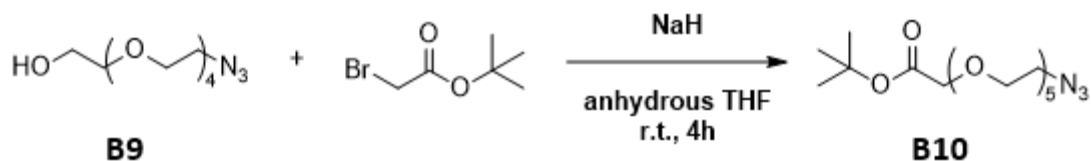
To a flask charged with compound **B6a** (68 mg, 39 μ mol) in acetonitrile (1 mL) was added a solution of DBCO-NHS ester (60 mg, 149 μ mol, 3.82 equiv.) in acetonitrile (1.5 mL). The mixture was stirred at room temperature for 1 minute and then DIPEA (10 μ L, 57.4 μ mol, 1.47 equiv.) was added dropwise. The mixture was stirred at room temperature for 18 hours, at which point the solvent was removed under reduced pressure and the crude mixture was re-dissolved in a mixture of water and acetonitrile (4 mL, 1:1 v/v). The crude mixture was purified by reverse-phase HPLC (mobile phase A: 0.1% TFA in water; mobile phase B: 0.1% TFA in acetonitrile; monitored at 260 nm) with a gradient of 0% - 100% B over 30 minutes and lyophilised to afford the product **B7** (59 mg, 75%). HRMS (ESI-TOF): calculated m/z : 2036.8462 ($[M+H]^+$, C₉₂H₁₂₇N₁₄O₃₈⁺); found m/z : 2036.8427.

Azido-PEG₅-alcohol (B9)

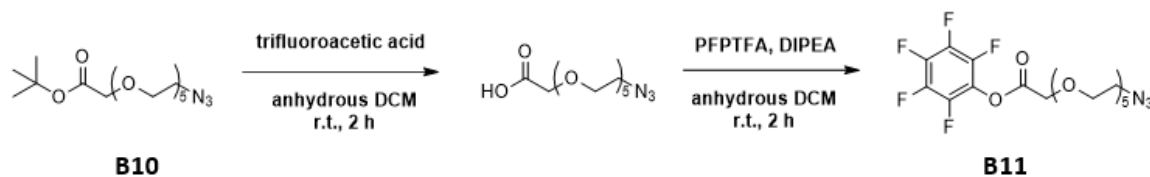
This procedure is modified from the literature.²³⁹



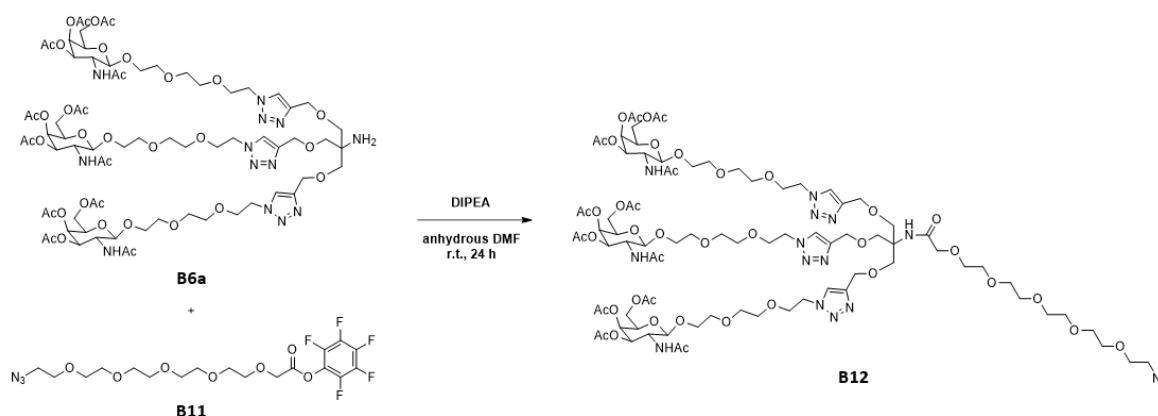
Pentaethylene glycol, **B8** (5.72 g, 24 mmol, 2 equiv.) was dissolved in tetrahydrofuran (50 mL) in a round-bottom flask. To the stirring solution was added 1,8-Diazabicyclo[5.4.0]undec-7-ene (DBU) (2.00 g, 13.2 mmol, 1.1 equiv.) and the mixture was cooled down to 0 °C. Diphenylphosphoryl azide (DPPA) (3.30 g, 12 mmol) was added to the mixture dropwise over 5 minutes at 0 °C. After stirring at room temperature for 1 hour, sodium azide (2.34 g, 36 mmol, 3 equiv.) was added to the turbid solution and the mixture was allowed to react at 70 °C under reflux for 18 hours. Then, the mixture was diluted to 100 mL with diethyl ether and filtered to remove the solids. The filtrate was concentrated *in vacuo* and purified by column chromatography (SiO₂, MeOH/DCM 2% to 5%) afforded compound **B9** (1.37 g, 43%) as a pale yellow oil. *R*_f (MeOH/DCM 5%) = 0.18; ¹H NMR (400 MHz, CDCl₃) δ 3.76 – 3.73 (m, 2H), 3.71 – 3.68 (m, 14H), 3.64 – 3.62 (m, 2H), 3.44 – 3.38 (m, 2H), 2.42 (s, 1H).

Azido-PEG₅-tert-butyl ester (B10)²⁴⁰

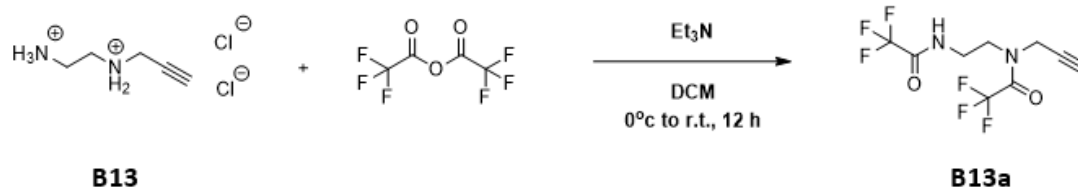
To an oven-dried flask charged with compound **B9** (1.36 g, 5.17 mmol) was added anhydrous tetrahydrofuran (15 mL). The solution was cooled down to 0 °C with stirring for 5 minutes before sodium hydride (60% in mineral oil) (414 mg, 10.3 mmol, 2 equiv.) was added in portionwise over 1 minute under an argon atmosphere. The mixture was allowed to react at 0°C for 15 minutes. Then, tert-Butyl bromoacetate (1.51 g, 7.76 mmol, 1.5 equiv.) dissolved in anhydrous tetrahydrofuran (5 mL) was added to the reaction flask dropwise over 5 minutes. After the mixture being stirred at room temperature for 3 hours, sodium hydride was quenched with water (1 mL) and the orange-coloured suspension was diluted with diethyl ether (80 mL) and filtered off. The filtrate was concentrated *in vacuo* and purified by column chromatography (SiO_2 , MeOH/EtOAc 2% to 5%) afforded compound **B10** (495 mg, 25%) as a yellow oil. R_f (MeOH/EtOAc 2%) = 0.55; $^1\text{H NMR}$ (400 MHz, CDCl_3) δ 4.04 (s, 2H), 3.75 – 3.67 (m, 20H), 3.43 – 3.39 (m, 2H), 1.49 (s, 9H).

Azido-PEG₅-pfp ester (B11)

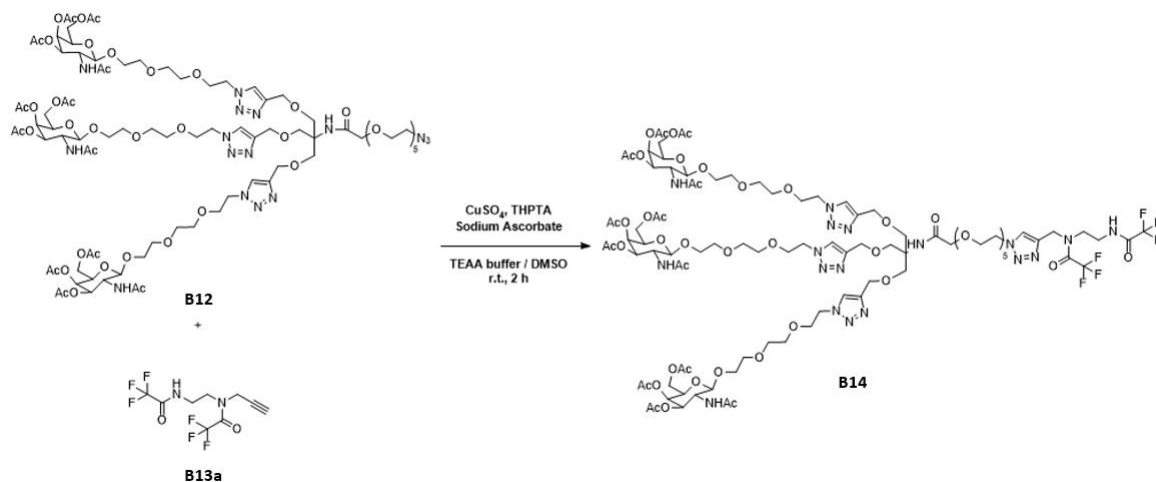
To a solution of compound **10** (490 mg, 1.30 mmol) in anhydrous dichloromethane (8.5 mL) was added trifluoroacetic acid (1.5 mL) dropwise over 10 minutes at 0 °C. The mixture was warmed up to room temperature and stirred for 2 hours. At this point the solution was diluted to 50 mL with toluene and concentrated under reduced pressure. The residual acid was removed by co-evaporating with toluene (50 mL × 2). The crude product was used in the next step without purification. The carboxylic acid product was re-dissolved in anhydrous dichloromethane (12 mL) and cooled down to 0 °C. To the stirring solution was gradually added DIPEA (680 µL, 3.90 mmol, 3 equiv.) over 5 minutes, followed by pentafluorophenyl trifluoroacetate (PFPTFA) (437 mg, 1.56 mmol, 1.2 equiv.) over 10 minutes. The mixture was stirred at room temperature for 2 hours. Then, the solvent was removed *in vacuo* and the crude product was re-dissolved in dichloromethane (30 mL), washed with water (30 mL), dried over MgSO₄ and concentrated for column chromatography (SiO₂, Et₂O/Petroleum ether 50% to 100%) which afforded compound **B11** (290 mg, 46%) as a yellow oil. *R_f* (MeOH/DCM 2%) = 0.10; ¹H NMR (400 MHz, CDCl₃) δ 4.57 (s, 2H), 3.87 – 3.83 (m, 2H), 3.77 – 3.74 (m, 2H), 3.69 (d, *J* = 2.9 Hz, 14H), 3.41 (t, *J* = 5.1 Hz, 2H).

Compound B12

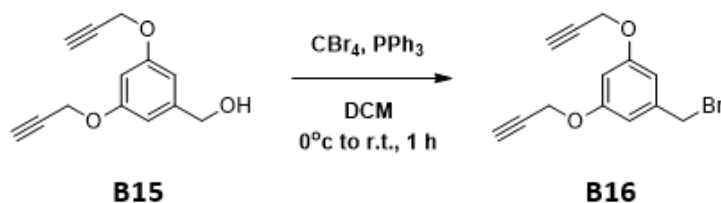
To a vial charged with compound **6a** (92 mg, 52.6 μmol) dried *in vacuo* overnight was added a solution of azido-PEG5-pfp ester, compound **B11** (103 mg, 210 μmol , 4 equiv.) in dry N,N-dimethylformamide (3 mL). To the mixture was added dry DIPEA (10.5 μL , 60 μmol) and the reaction was stirred at room temperature for 24 hours. The crude mixture was purified by reverse-phase HPLC (mobile phase A: 0.1% TFA in water; mobile phase B: 0.1% TFA in acetonitrile; monitored at 260 nm) with a gradient of 0% - 80% B over 30 minutes and lyophilised to afford compound **B12** (60 mg, 56 %) as an oily white solid. R_T = 20.3 min. ^1H NMR (600 MHz, DMSO) δ 8.02 (s, 3H), 7.79 (d, J = 9.2 Hz, 3H), 6.87 (s, 1H), 5.22 (d, J = 3.4 Hz, 3H), 4.98 (dd, J = 11.2, 3.4 Hz, 3H), 4.56 (d, J = 8.5 Hz, 3H), 4.52 (t, J = 5.3 Hz, 6H), 4.49 (s, 6H), 4.03 (h, J = 4.3 Hz, 9H), 3.89 (dt, J = 11.2, 8.9 Hz, 3H), 3.82 (t, J = 5.3 Hz, 6H), 3.79 (d, J = 3.3 Hz, 2H), 3.78 – 3.75 (m, 2H), 3.67 (s, 6H), 3.61 – 3.59 (m, 2H), 3.58 – 3.46 (m, 38H), 3.40 – 3.37 (m, 2H), 2.10 (s, 9H), 2.00 (s, 9H), 1.90 (s, 9H), 1.76 (s, 9H). ^{13}C NMR (151 MHz, DMSO) δ 170.46, 170.39, 170.10, 169.76, 169.35, 158.86, 158.61, 144.10, 124.63, 116.77, 114.85, 101.42, 70.95, 70.61, 70.38, 70.27, 70.25, 70.22, 70.20, 70.15, 70.02, 69.92, 69.86, 69.71, 69.19, 68.78, 68.49, 67.20, 64.52, 61.92, 59.60, 50.46, 49.80, 49.79, 31.77, 23.24, 20.96, 20.90, 20.89. HRMS (ESI-TOF): calculated m/z : 1026.9509 ($[\text{M}+2\text{H}]^{2+}$, $\text{C}_{85}\text{H}_{136}\text{N}_{16}\text{O}_{42}^{2+}$); found m/z : 1026.9470.

2,2,2-trifluoro-N-(prop-2-yn-1-yl)-N-(2-(2,2,2-trifluoroacetamido)ethyl)acetamide (B13a)

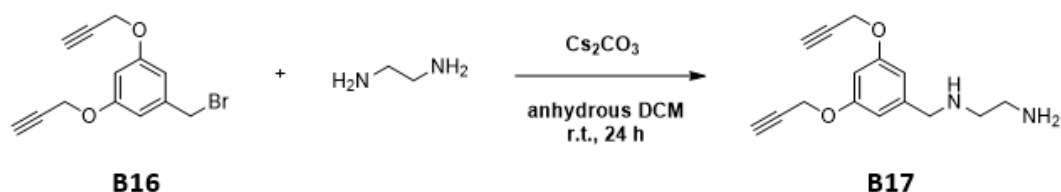
To a flask charged with compound **B13**¹³⁵ (171 mg, 1 mmol) in dichloromethane (5 mL) was added triethylamine (836 μ L, 6 mmol) and the mixture was cooled down to 0 °C. After 5 minutes of stirring, trifluoroacetic anhydride (630 mg, 3 mmol) was added to the mixture dropwise over 5 minutes and the reaction mixture was stirred at room temperature for 12 hours. Then, the mixture was diluted to 20 mL with dichloromethane and washed with saturated sodium bicarbonate solution (20 mL). The crude product was extracted once more from the aqueous phase by dichloromethane (20 mL). The combined organic phase was dried over MgSO₄ and concentrated *in vacuo*. Purification by column chromatography (SiO₂, EtOAc/Petroleum ether 10% to 30%) afforded compound **B14** (294 mg, quant.) as a brownish oil. R_f (EtOAc/Petroleum ether 30%) = 0.20. ¹H NMR (400 MHz, CDCl₃) δ 7.05 (s, 1H), 4.29 (dd, J = 21.3, 2.8 Hz, 2H), 3.80 (dd, J = 6.9, 4.9 Hz, 2H), 3.75 – 3.67 (m, 2H), 2.42 (dt, J = 4.8, 2.5 Hz, 1H).

Compound B14

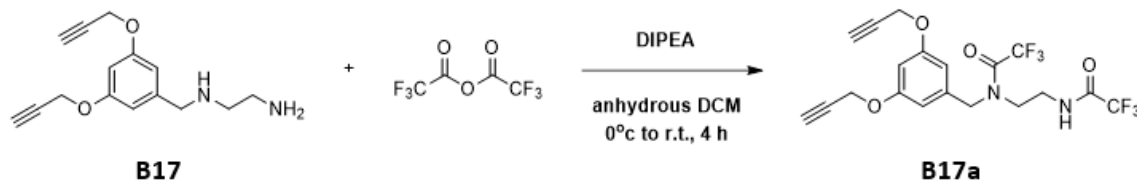
To a vial charged with compound **B12** (33 mg, 16 μmol) dissolved in dimethyl sulfoxide (500 μL) was added compound **B13a** (16 mg, 55 μmol) in equal volume of DMSO. The mixture was purged with argon for 5 minutes at room temperature. The catalyst solution prepared by mixing copper (II) sulfate solution (1.0 M, 10 μL , 10 μmol), THPTA (100 mM, 100 μL , 10 μmol) and sodium ascorbate (1.0 M, 100 μL , 100 μmol) in TEAA buffer (1.0 M, 150 μL , pH 7.0) and degassing with argon was added to the reaction mixture. After stirring for 2 hours at room temperature, the crude mixture was purified by reverse-phase HPLC (mobile phase A: 0.1% TFA in water; mobile phase B: 0.1% TFA in acetonitrile; monitored at 215 nm) with a gradient of 0% - 80% B over 30 minutes and lyophilised to afford compound **B14** (17.6 mg, 47 %). R_T = 22.0 min. HRMS (ESI-TOF): calculated m/z : 2341.9404 ($[\text{M}+\text{H}]^+$, $\text{C}_{94}\text{H}_{143}\text{F}_6\text{N}_{18}\text{O}_{44}^+$); found m/z : 2341.9519.

1-(bromomethyl)-3,5-bis(prop-2-yn-1-yloxy)benzene (B16)²⁴¹

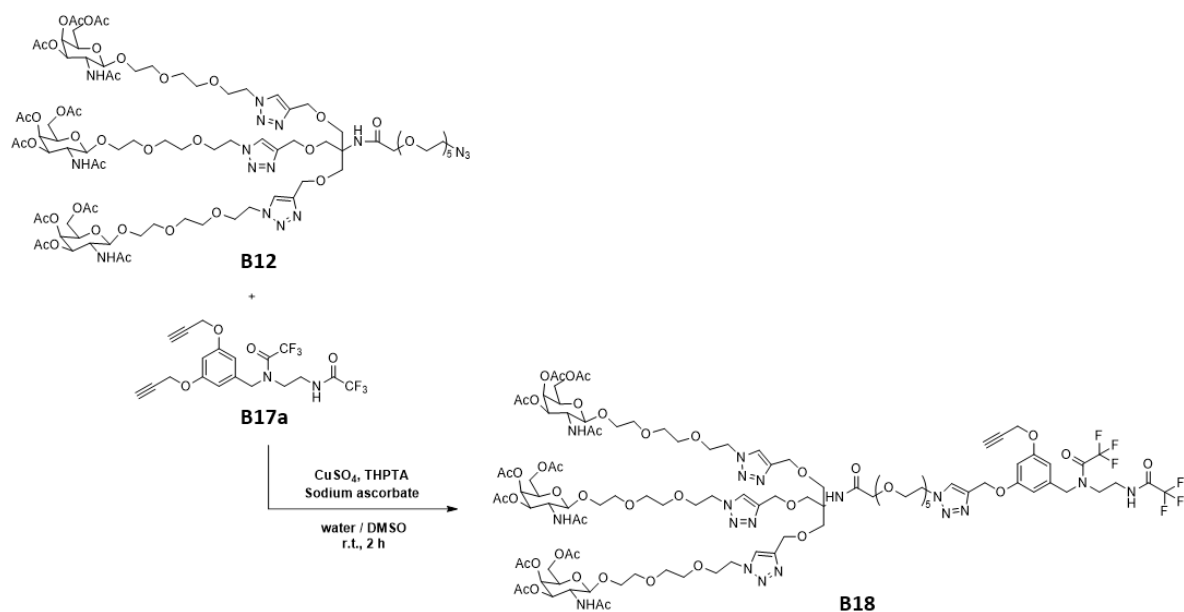
To a flask charged with (3,5-bis(prop-2-yn-1-yloxy)phenyl)methanol, compound **B15**²⁴² (840 mg, 3.88 mmol) in dichloromethane (12 mL) was added carbon tetrabromide (1.60 g, 4.85 mmol, 1.25 equiv.) at 0 °C. Under a stream of argon, triphenylphosphine (1.27 g, 4.85 mmol, 1.25 equiv.) was added and the mixture was stirred at room temperature for 1 hour. At this point, the solvent was removed and the crude product was re-dissolved in ethanol (10 mL) and mixed with anhydrous zinc (II) chloride (1.06 g, 7.77 mmol) to induce precipitation of triphenylphosphine oxide for 4 hours. The crude product was filtered and concentrated *in vacuo* and subjected to column chromatography (SiO₂, DCM/Petroleum ether 30% to 50%) which afforded compound **B16** (1.04 g, 97%) as an oil. *R_f* (DCM/Petroleum ether 50%) = 0.42. ¹H NMR (400 MHz, CDCl₃) δ 6.65 (d, *J* = 2.3 Hz, 2H), 6.56 (t, *J* = 2.3 Hz, 1H), 4.68 (d, *J* = 2.4 Hz, 4H), 4.42 (s, 2H), 2.54 (t, *J* = 2.4 Hz, 2H).

N¹-(3,5-bis(prop-2-yn-1-yloxy)benzyl)ethane-1,2-diamine (B17)

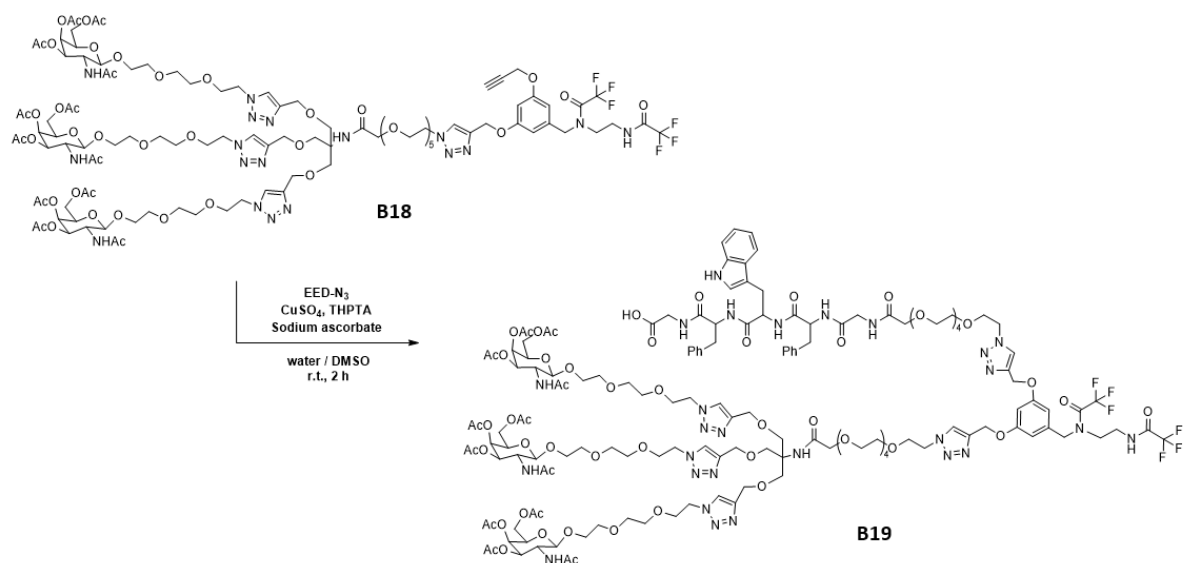
To an oven-dried flask charged with caesium carbonate (1.32 g, 4.06 mmol, 1.1 equiv.) was added dry dichloromethane (16 mL) and ethylenediamine (2.22 g, 37 mmol, 10 equiv.). A solution of 1-(bromomethyl)-3,5-bis(prop-2-yn-1-yloxy)benzene, compound **B16** (1.03 g, 3.69 mmol) in dry DCM (10 mL) dropwise over 40 minutes. The suspension was stirred at room temperature for 24 hours and then diluted to 50 mL with toluene. The excess amine was removed by co-evaporation with toluene (50 mL × 2). The concentrated crude product was re-suspended in water (50 mL) and extracted with dichloromethane (50 mL × 4), dried over MgSO₄ and concentrated *in vacuo*. Purification by column chromatography (SiO₂, MeOH/DCM 5% to 20% + 1% conc. ammonia) afforded compound **B17** (580 mg, 61%) as an oil. *R_f* (MeOH/DCM 10% + 1% conc. ammonia) = 0.07. ¹H NMR (400 MHz, DMSO) δ 6.60 (d, *J* = 2.3 Hz, 2H), 6.46 (t, *J* = 2.3 Hz, 1H), 4.76 (d, *J* = 2.4 Hz, 4H), 3.63 (s, 2H), 3.56 (t, *J* = 2.4 Hz, 2H), 3.35 (br, 1H), 2.59 (t, *J* = 6.2 Hz, 2H), 2.47 (t, *J* = 5.8 Hz, 2H), 1.63 (br, 2H). ¹³C NMR (101 MHz, DMSO) δ 158.64, 144.36, 107.51, 100.61, 79.75, 78.64, 55.90, 53.31, 52.44, 42.00. HRMS (ESI-TOF): calculated *m/z*: 259.1441 ([*M*+H]⁺, C₁₅H₁₉N₂O₂⁺); found *m/z*: 259.1437.

N-(3,5-bis(prop-2-yn-1-yloxy)benzyl)-2,2,2-trifluoro-N-(2-(2,2,2-trifluoroacetamido)ethyl)acetamide (B17a)

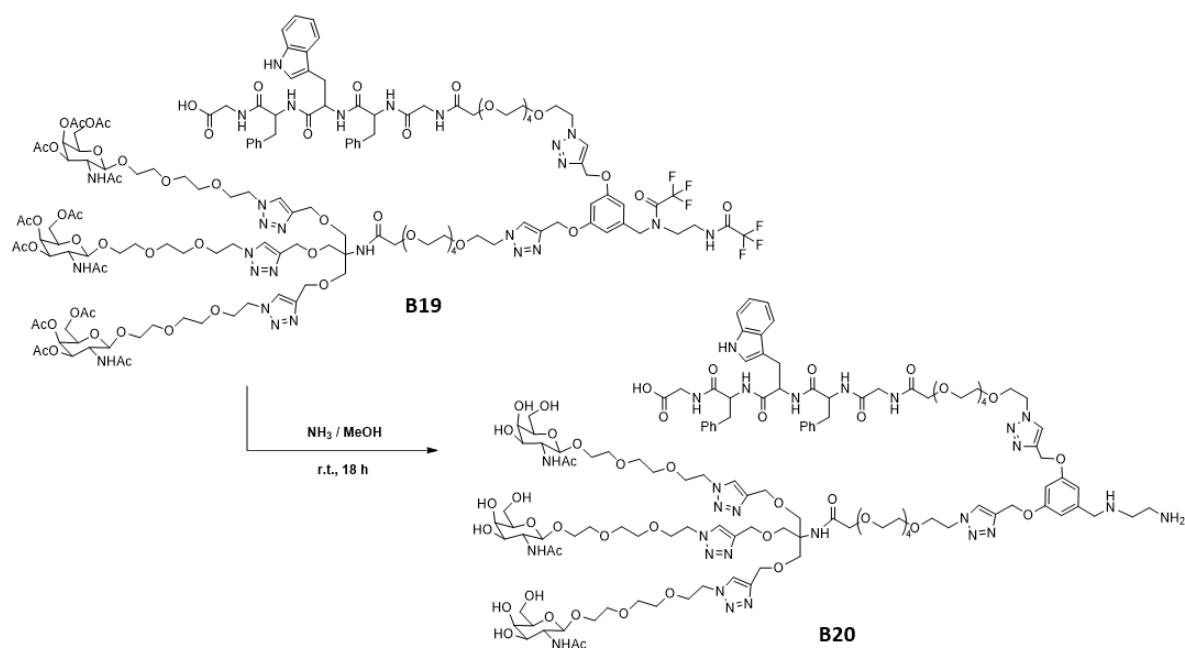
To a flask charged with a solution of compound **B17** (570 mg, 2.20 mmol) in dry dichloromethane (12 mL) was added DIPEA (2.30 mL, 13.2 mmol, 6 equiv.). After the mixture was cooled down to 0 °C, trifluoroacetic anhydride (1.39 g, 6.60 mmol, 3 equiv.) was added to it dropwise over 10 minutes. The mixture was stirred at room temperature for 4 hours, at which point the acid and solvent were removed by co-evaporation with toluene (30 mL × 3) under reduced pressure. The crude product was re-suspended in water (30 mL), extracted with dichloromethane (30 mL × 3), dried over MgSO₄ and concentrated *in vacuo*. Purification by column chromatography (SiO₂, EtOAc/Petroleum ether 30% to 50%) afforded compound **B17a** (878 mg, 89%) as an oil. *R_f* (EtOAc/Petroleum 50%) = 0.55. ¹H NMR (400 MHz, DMSO) δ 9.57 (s, 1H), 6.61 – 6.56 (m, 1H), 6.50 (t, *J* = 2.2 Hz, 2H), 4.79 (dd, *J* = 6.4, 2.4 Hz, 4H), 4.62 (d, *J* = 2.2 Hz, 2H), 3.58 (dt, *J* = 3.3, 2.4 Hz, 2H), 3.50 (s, 2H), 3.46 (q, *J* = 3.3 Hz, 2H). ¹⁹F NMR (376 MHz, DMSO) δ -67.28, -67.67, -74.65 (d, *J* = 4.3 Hz). HRMS (ESI-TOF): calculated *m/z*: 451.1087 ([M+H]⁺, C₁₉H₁₇F₆N₂O₄⁺); found *m/z*: 451.1084.

Compound 18

To a vial charged with compound **B12** (50.7 mg, 24.7 μmol) dissolved in dimethyl sulfoxide (750 μL) was added compound **B17a** (109 mg, 247 μmol , 10 equiv.) in equal volume of DMSO. The mixture was purged with argon for 5 minutes at room temperature. The catalyst solution prepared by mixing copper (II) sulfate solution (1.0 M, 10 μL , 10 μmol), THPTA (100 mM, 200 μL , 20 μmol) and sodium ascorbate (1.0 M, 100 μL , 100 μmol) in TEAA buffer (1.0 M, 240 μL , pH 7.0) and degassing with argon was added to the reaction mixture. After stirring for 2 hours at room temperature, the crude mixture was purified by reverse-phase HPLC (mobile phase A: 0.1% TFA in water; mobile phase B: 0.1% TFA in acetonitrile; monitored at 215 nm) with a gradient of 0% - 80% B over 30 minutes and lyophilised to afford compound **B18** (49.5 mg, 80 %). $R_T = 22.9$ min. This intermediate compound has only been characterised by LCMS and used in the next step after isolation.

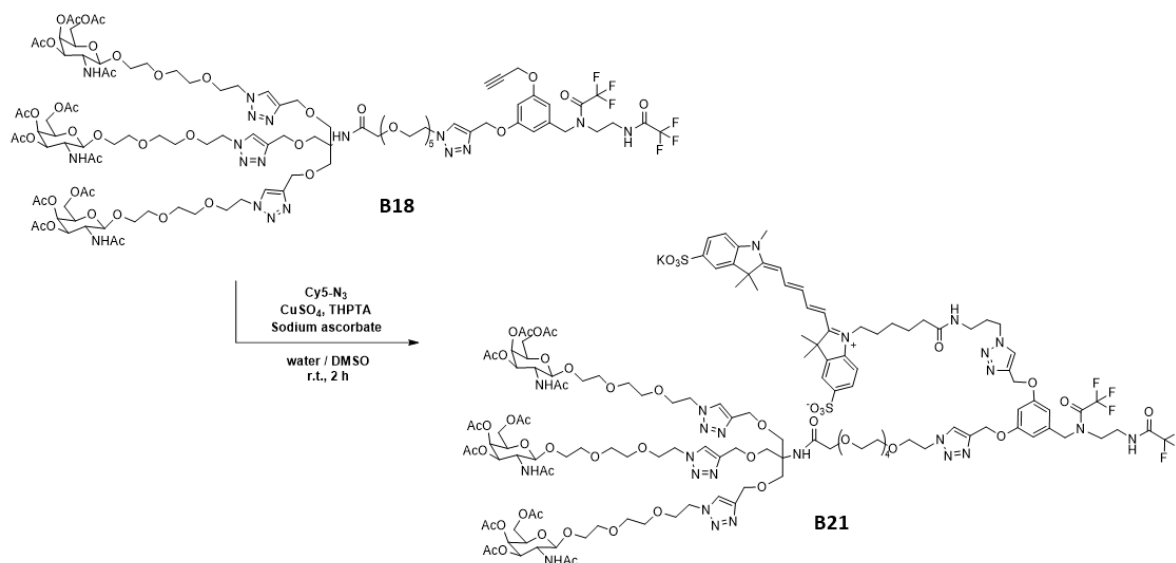
Compound B19

To a vial charged with compound **B18** (15 mg, 6.0 μmol) dissolved in dimethyl sulfoxide (500 μL) was added EED- N_3 (6.0 mg, 6.5 μmol , 1.08 equiv., Biomatik) in equal volume of DMSO. The mixture was purged with argon for 5 minutes at room temperature. The catalyst solution prepared by mixing copper (II) sulfate solution (50 mM, 100 μL , 5 μmol), THPTA (100 mM, 100 μL , 10 μmol) and sodium ascorbate (1.0 M, 50 μL , 50 μmol) in TEAA buffer (1.0 M, 150 μL , pH 7.0) and degassing with argon was added to the reaction mixture. After stirring for 2 hours at room temperature, the crude mixture was purified by reverse-phase HPLC (mobile phase A: 0.1% TFA in water; mobile phase B: 0.1% TFA in acetonitrile; monitored at 215 nm) with a gradient of 0% - 80% B over 30 minutes and lyophilised to afford compound **B19** (15 mg, 66 %). R_T = 20.1 min. This intermediate compound has only been characterised by LCMS and used in the next step after isolation.

Compound B20

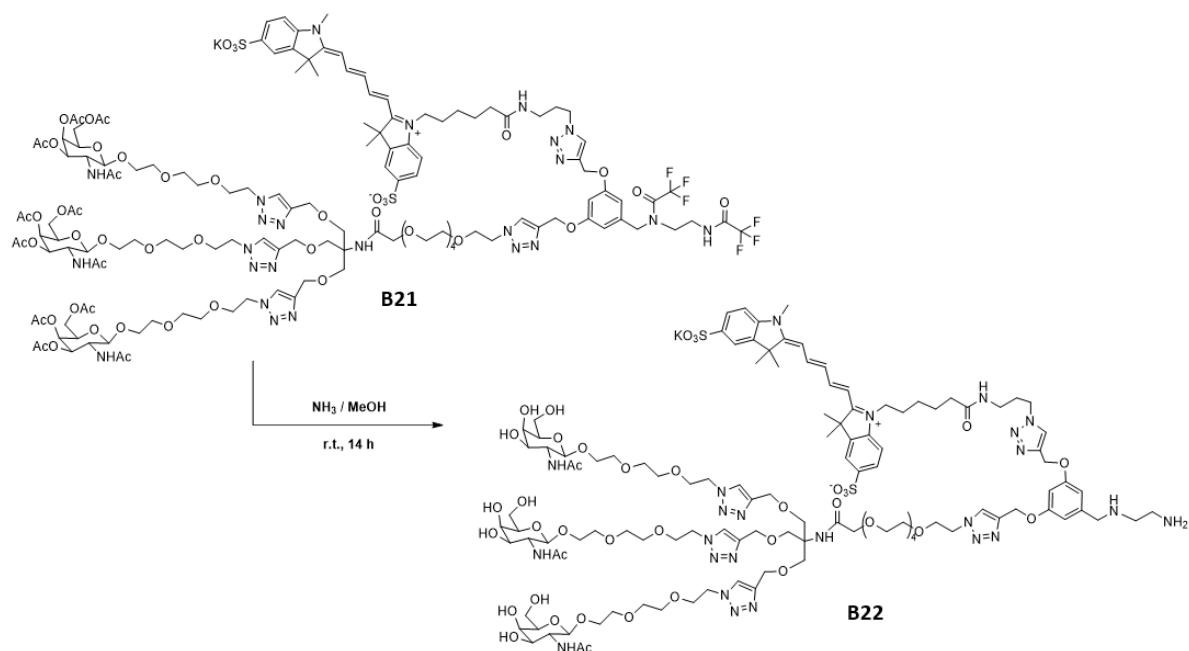
Compound **B19** (2.7 mg, 790 nmol) was dissolved in ammonia in methanol (0.9 mL, 3.5 N) and left at room temperature for 18 hours. After that, the solution was concentrated by evaporation under a gentle stream of argon and re-dissolved in 2 mL of water. Then, the crude mixture was purified by reverse-phase HPLC (mobile phase A: 0.1% TFA in water; mobile phase B: 0.1% TFA in acetonitrile; monitored at 215 nm) with a gradient of 0% - 80% B over 30 minutes and lyophilised to afford compound **B20** (2.2 mg, 98 %). $R_T = 17.3$ min. HRMS (ESI-TOF): calculated m/z : 2848.3488 ($[M+H]^+$, $C_{127}H_{192}N_{27}O_{47}^+$); found m/z : 2848.3372.

Compound 21

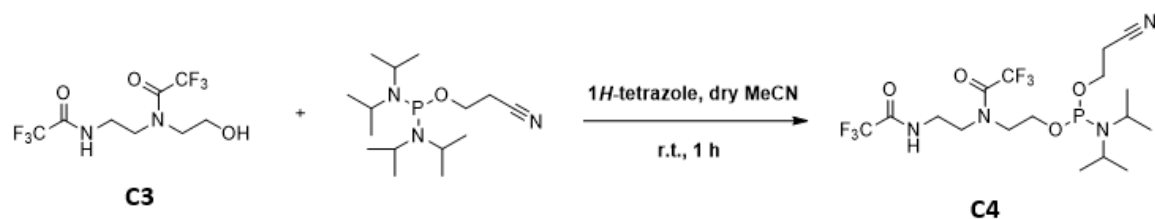


To a vial charged with compound **B18** (15 mg, 6.0 μmol) dissolved in dimethyl sulfoxide (500 μL) was added sulfo-cyanine5-N₃ (5.0 mg, 6.5 μmol , 1.08 equiv., Lumiprobe) in equal volume of DMSO. The mixture was purged with argon for 5 minutes at room temperature. The catalyst solution prepared by mixing copper (II) sulfate solution (50 mM, 100 μL , 5 μmol), THPTA (100 mM, 100 μL , 10 μmol) and sodium ascorbate (1.0 M, 50 μL , 50 μmol) in TEAA buffer (1.0 M, 150 μL , pH 7.0) and degassing with argon was added to the reaction mixture. After stirring for 2 hours at room temperature, the crude mixture was purified by reverse-phase HPLC (mobile phase A: 0.1% TFA in water; mobile phase B: 0.1% TFA in acetonitrile; monitored at 215 and 647 nm) with a gradient of 0% - 80% B over 30 minutes and lyophilised to afford compound **B21** (10 mg, 51 %). R_T = 20.7 min. HRMS (ESI-TOF): calculated m/z : 1612.1227 ($[\text{M}-\text{H}]^{2-}$, C₁₃₉H₁₉₂F₆N₂₄O₅₃S₂²⁻); found m/z : 1612.1268.

Compound B22



Compound **B21** (2.6 mg, 796 nmol) was dissolved in ammonia in methanol (0.9 mL, 3.5 N) and left at room temperature for 14 hours. After that, the solution was concentrated by evaporation under a gentle stream of argon and re-dissolved in 2 mL of water. Then, the crude mixture was purified by reverse-phase HPLC (mobile phase A: 0.1% TFA in water; mobile phase B: 0.1% TFA in acetonitrile; monitored at 215 and 647 nm) with a gradient of 0% - 80% B over 30 minutes and lyophilised to afford compound **B22** (1.73 mg, 82 %). R_T = 14.8 min. HRMS (ESI-TOF): calculated m/z : 2655.1928 ($[M]^-$, $C_{117}H_{177}N_{24}O_{42}S_2^-$); found m/z : 2655.1899.

Compound C4

To a round-bottom flask charged with compound **C3** (800 mg, 2.7 mmol) dissolved in dry acetonitrile (6 mL) was added 2-cyanoethyl N,N,N',N'-tetraisopropylphosphodiamidite (1.3 mL, 4.05 mmol, 1.50 equiv.) and 1H-tetrazole (0.45 M solution) (6.0 mL, 2.7 mmol, 1 equiv.). The mixture was stirred at room temperature for 1 hour. The crude mixture was filtered and the filtrate was concentrated under reduced pressure and purified by (SiO₂, DCM/Petroleum ether 50% + 1% Et₃N, isocratic) afforded compound **C4** (730 mg, 54%) as an oil. *R_f* (DCM/Petroleum 60% + 1% Et₃N) = 0.45. ¹H NMR (500 MHz, CDCl₃) δ 7.42 (dt, *J* = 116.4, 6.5 Hz, 1H), 3.98 – 3.45 (m, 12H), 2.74 – 2.50 (m, 2H), 1.17 (ddd, *J* = 10.1, 6.8, 3.6 Hz, 12H). ³¹P NMR (202 MHz, CDCl₃) δ 149.00, 148.18. ¹⁹F NMR (470 MHz, CDCl₃) δ -68.68 (d, *J* = 99.7 Hz), -76.03 (d, *J* = 83.2 Hz).

Chapter 6 – Appendix

6.1 List of Oligonucleotides

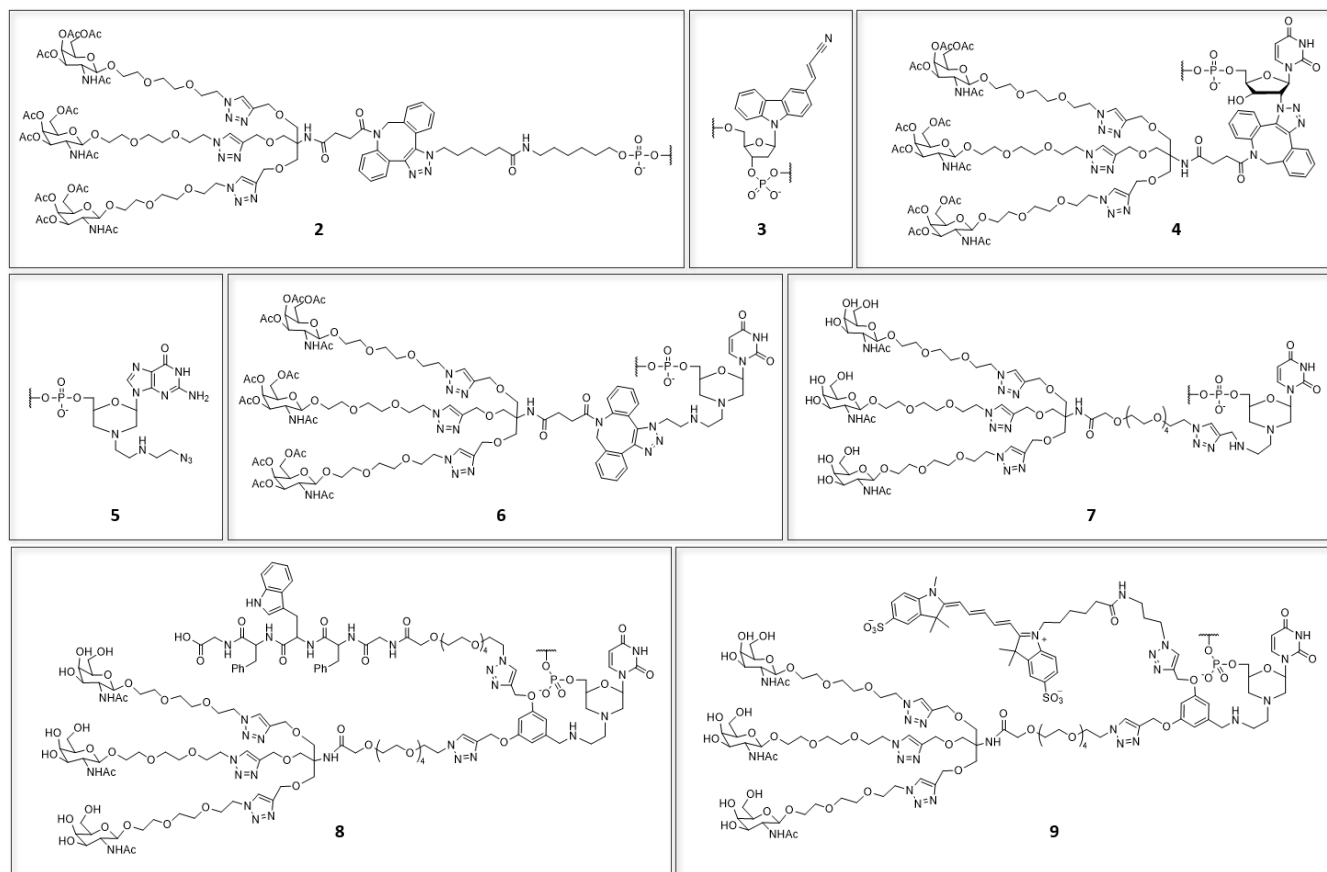
Oligonucleotides used in each chapter are shown in this section. Each oligonucleotide has an ON number corresponding to the experiments conducted in this study. The unique identification number provides reference to the raw analytical data.

ON	Sequence	Calc. mass (Da)	Found mass (Da)	Unique Identification #
ON2.1	TTTTTTTTTTTTT	3588.4	3588.0	Res9663
ON2.2	N ₃ -TTTTTTTTTTTTT	3613.4	3613.0	Res9663N ₃
ON2.3	HO (CH ₂) ₃ -TTTTTTTTTTTTT	3725.5	N/A	Res9836
ON2.4	N ₃ (CH ₂) ₃ -TTTTTTTTTTTTT	3750.5	3753.0	Res9836N ₃
ON2.5	L3OH- (Tr-TTTTTTTTTTTTT) ₃	11519.5	11520.0	Res9663L3OH
ON2.6	TCACTTCGCTTCAAC	4462.9	4464.0	Res10552
ON2.7	N ₃ -TCACTTCGCTTCAAC	4487.9	4489.0	Res10552N ₃
ON2.8	L3OH- (Tr-TCACTTCGCTTCAAC) ₃	14143	14145.5	Res10552L3OH
ON2.9	UCACUUCGCUUCAAC	4632.8	N/A	Res10651
ON2.10	N ₃ -UCACUUCGCUUCAAC	4657.8	4659.0	Res10651N ₃
ON2.11	L3OH- (Tr-UCACUUCGCUUCAAC) ₃	14652.7	14655.0	Res10651L3OH
ON2.12	L3N ₃ - (Tr-TTTTTTTTTTTTT) ₃	11544.5	11545.5	Res10011L3N ₃
ON2.13	L3OH- [Tr-L3- (Tr-TTTTTTTTTTTTT) ₃] ₃	35312.8	35314.0	Res10011L3L3
ON2.14	L3N ₃ - (Tr-TCACTTCGCTTCAAC) ₃	14168	N/A	Res10552L3N ₃
ON2.15	L3OH- [Tr-L3- (Tr-TCACTTCGCTTCAAC) ₃] ₃	43183.4	43189.0	Res10552L3L3

Figure 6.1 | List of oligonucleotides used in Chapter 2

L3OH and L3N₃ represent the linker motif with a hydroxy and azido end, respectively. Tr represents triazole group.

ON	Sequence	Calc. mass (Da)	Found mass (Da)	Unique Identification #
ON3.1	2GAGUCCGAGCAGAAGAAGAAGUUUUAGAGCUAUGCUGUUUUG	15929.4	15932	Res9058_GN
ON3.2	2CAAAUCUUA3UCUAUAUUA	8927.1	8929	Res9020_GN
ON3.3	PPFGGAGUCCGAGCAGAAGAAGAAGUUUUAGAGCUAG AAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU UCAUGUGUAUU4	40096.2 [M+2Mg] ⁴⁺	40092	Res9037_GN
ON3.4	GAGGAGCUGUUCACCGG	5481.3	5483	Res8843
ON3.4-EDA-N ₃	GAGGAGCUGUUCACCG5	5576.4	5578	Res8843N ₃
ON3.5	PPFGGAGUCCGAGCAGAAGAAGAAGUUUUAGAGCUAG AAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU	32942.4	32947	Res9072
ON3.5-EDA-Tr-GalNAc	PPFGGAGUCCGAGCAGAAGAAGAAGUUUUAGAGCUAG AAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUUU6	34837.7	34842	Res9072_GN_1
ON3.5-EDA-GalNAc	PPFGGAGUCCGAGCAGAAGAAGAAGUUUUAGAGCUAG AAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUUU7	34680.9	34684	Res9072ox5_GN
ON3.5-EDA-GalNAc-EED	PPFGGAGUCCGAGCAGAAGAAGAAGUUUUAGAGCUAG AAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUUU8	35781 [M+Mg] ²⁺	35784	Res9072_GNEED
ON3.5-EDA-GalNAc-Cy5	PPFGGAGUCCGAGCAGAAGAAGAAGUUUUAGAGCUAG AAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUUU9	35589 [M+Mg] ²⁺	35598	Res9072_GNCy5

**Figure 6.2 | List of oligonucleotides used in Chapter 3**

Each nucleoside modification is represented by a number. The structures of the modifications are illustrated.

ON	Sequence	Calc. mass (Da)	Found mass (Da)	Unique Identification #
ON4.1	CCACCACCACCACCACCACUAAUAGUGAAGGCGGC CGCAGCUCGCUUUGUUUAAAC	17797	17799	Res9972
ON4.2	XTGATTTCATTCTAATAGCAGC	6256	6257	Res10420
ON4.3	GCTGCTATTAGAATGAATCAGTTTAAACAAAGCGA GCTGC	12352	12351	Res10421
ON4.4	GCTGCTATTAGAATGAATCA ~~Y~~ GTTTAAACAAAGCG AGCTGC	12488	12490	Res10422
ON4.1-4.2 conjugate	CCACCACCACCACCACCACUAAUAGUGAAGGCGGC CGCAGCUCGCUUUGUUUAAACMTGATTTCATTCTAA TAGCAGC	24019	24022	N/A
ON4.5	UUCUGCCUAAUAAAAACAUUUUAUUUCAUUGCAU UUAAAUUG	13254	13253	Res10454
ON4.6	~~XGTGCCTGCAAAAAAAAAAAAAAAAAAAAAAAAAA~~ ~~AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA~~ ~~AAA3~~	25280	25281	Res10447
ON4.7	TTTTTGCAGGCACCATTTAAATGCAATGAAAAT	10134	10134	Res10445
ON4.8	TTTTTTTTGCAGGCACCATTTAAATGCAATGAAAA TAAATG	12620	12619	Res10494
ON4.9	TTTTTTTTGCAGGCACTCATTTAAATGCAATGAAA ATAAATG	12924	12923	Res10495
ON4.10	TTTTTTTTGCAGGCACCATTTAAATGCAATGAAAA TAAATG	12800	12801	Res10502
ON4.5-4.6 conjugate	UUCUGCCUAAUAAAAACAUUUUAUUUCAUUGCAU UUAAAUUGMTGCCTGCAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAA3	38500	38487	N/A
ON4.11	AAAAACAUUUUAUUUCAUUGCAUUUAAAUGUUUAA ACCACCGUUUACCAC	15766	15768	Res10528
ON4.12	AAAAACAUUUUAUUUCAUUGCAUUUAAAUGUUUAA ACCACCGUUUACCACUGCAAUUUGCA	19278	19281	Res10529
ON4.13	TTCGTGTGTGTGTGTGTGGTAAACGGTGGT	9386	9386	Res10532
ON4.14	XGTGGTAAACGGTGGTAAAAAAA	7359	7359	Res10539
ON4.15	XACACACACACGAAAAAAA	6552	6552	Res10540
ON4.11-4.14 conjugate	AAAAACAUUUUAUUUCAUUGCAUUUAAAUGUUUAA ACCACCGUUUACCACMTGGTAAACGGTGGTAAAA AAAA	23091	23108	N/A
ON4.12-4.14 conjugate	AAAAACAUUUUAUUUCAUUGCAUUUAAAUGUUUAA ACCACCGUUUACCACUGCAAUUUGCA ~~MTGGTAAA~~ CGGTGGTAAAAAAA	26603	26621	N/A
ON4.11-4.15 conjugate	AAAAACAUUUUAUUUCAUUGCAUUUAAAUGUUUAA ACCACCGUUUACCACMACACACACACGAAAAAAA AA	22284	22301	N/A

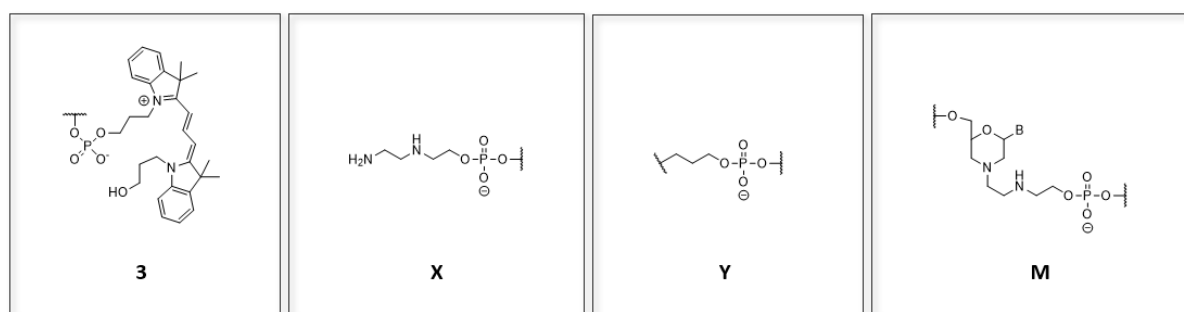
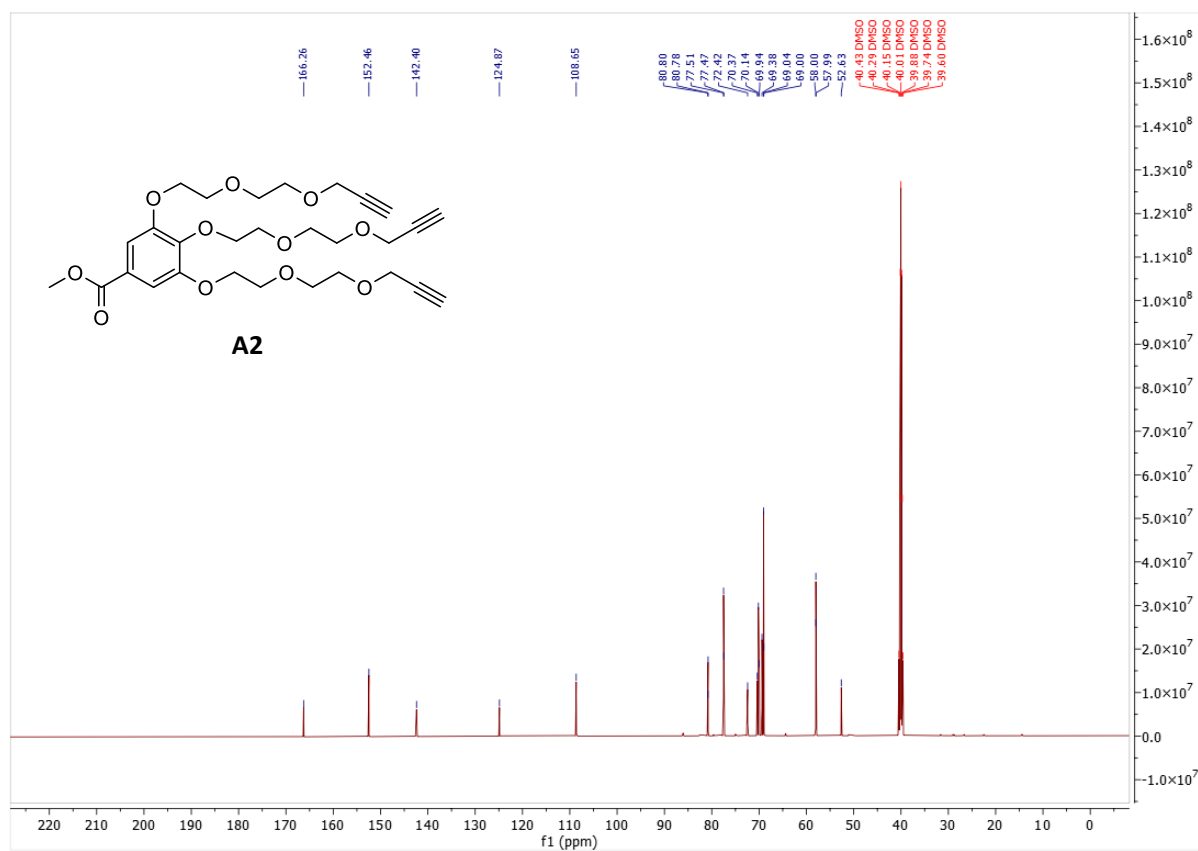
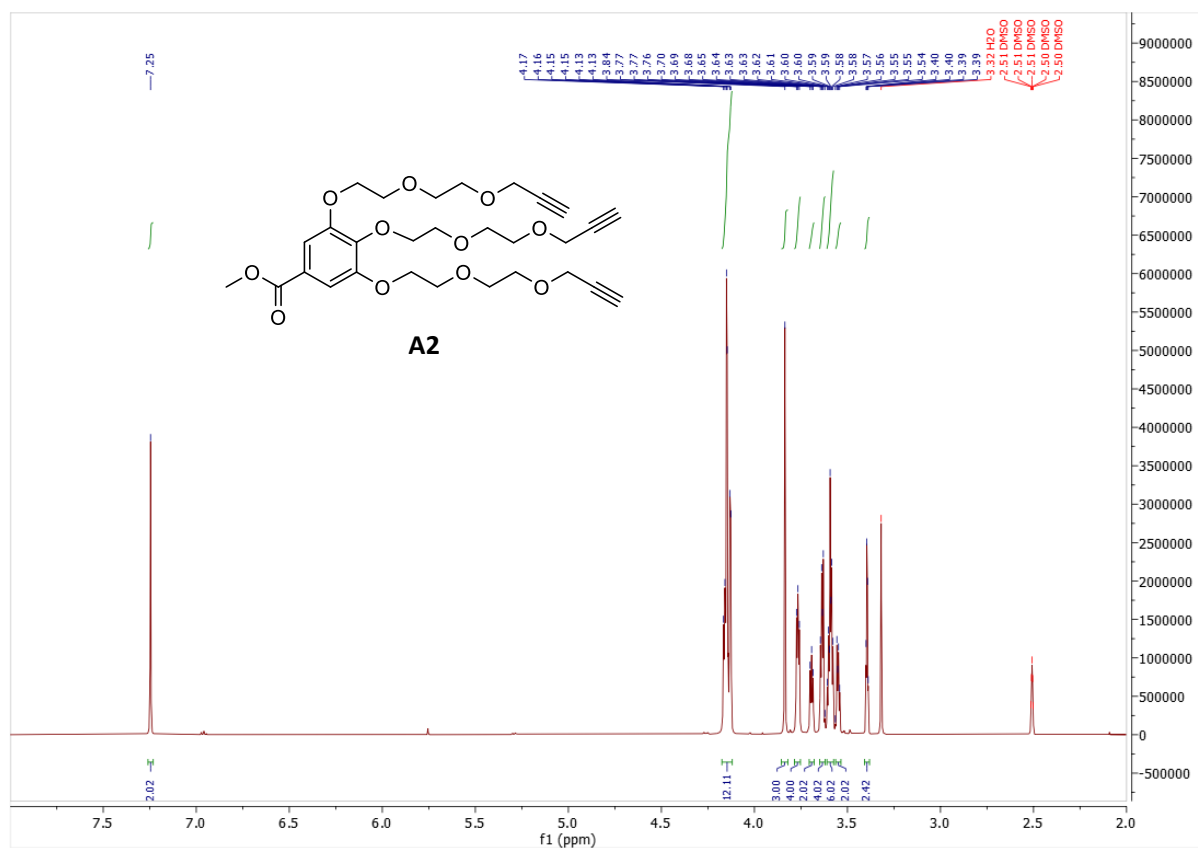


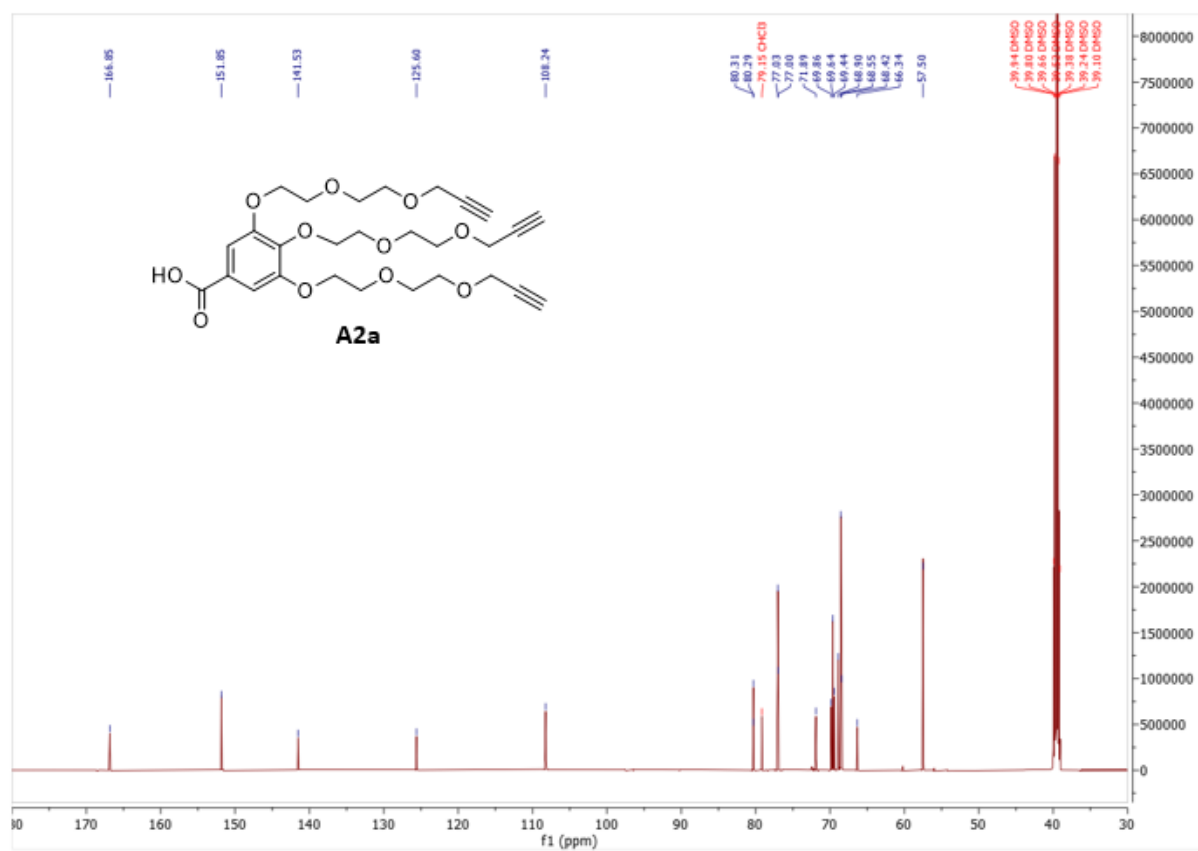
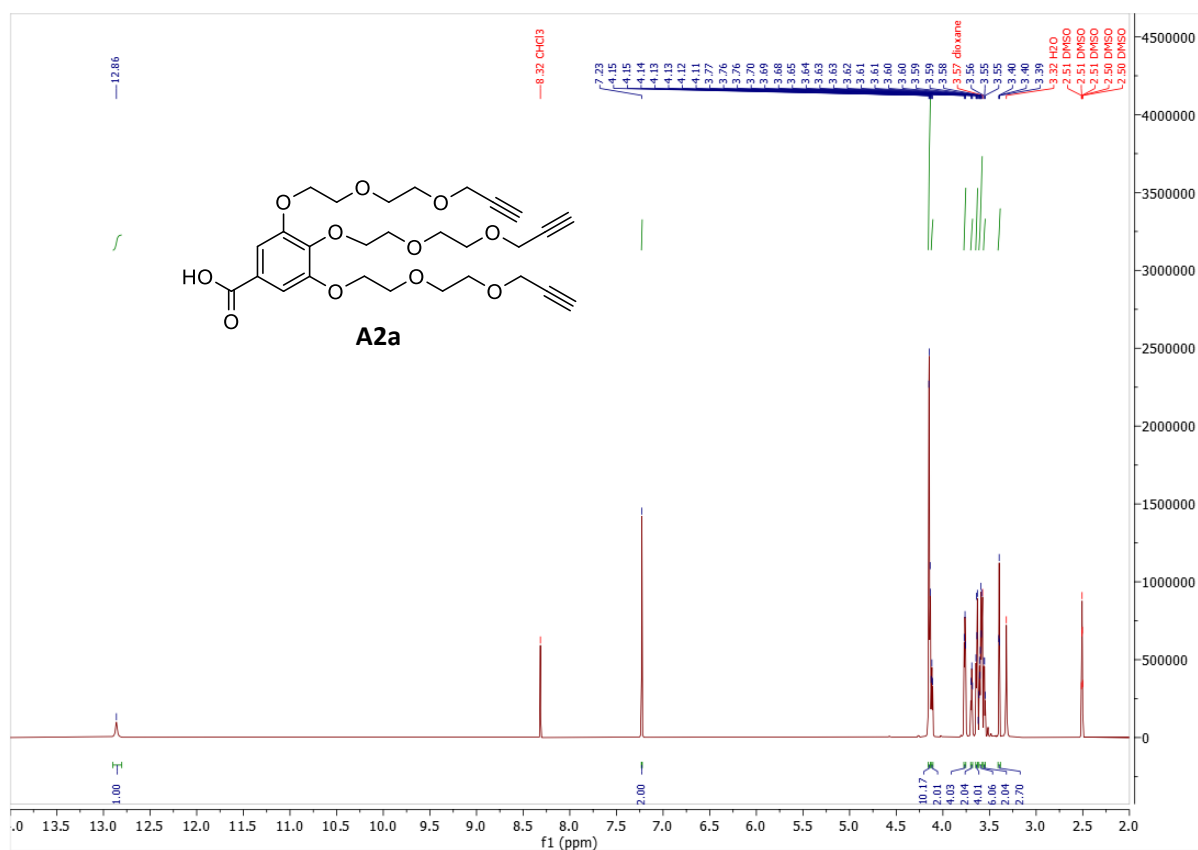
Figure 6.3 | List of oligonucleotides used in Chapter 4

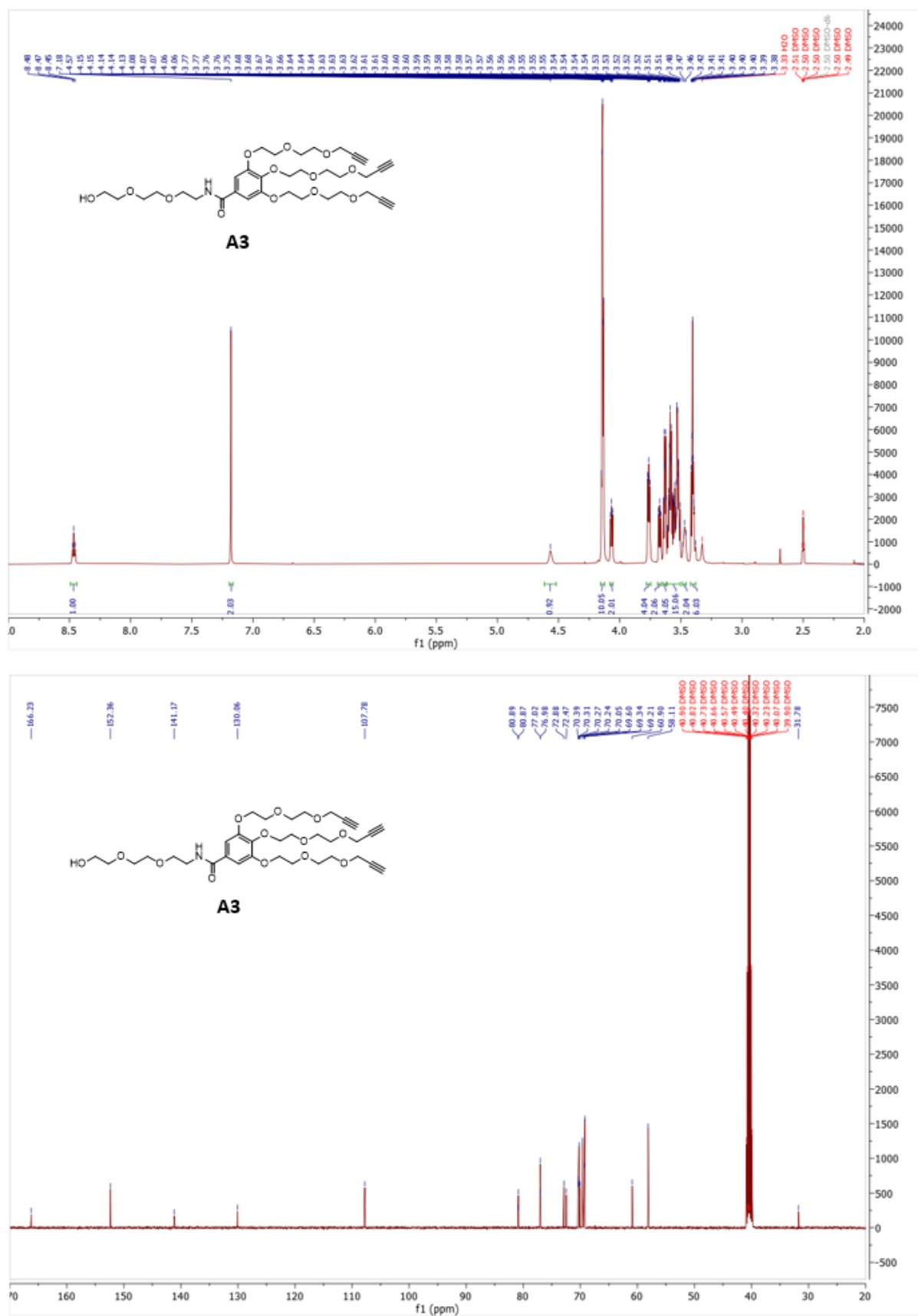
Each nucleoside modification is represented by a number. The structures of the modifications are illustrated. Underlined = 2'-OMe. Bold/underlined = LNA. Blue letters = RNA. M indicates a special modifier as a morpholino linkage between the 5'- and 3'- nucleotides.

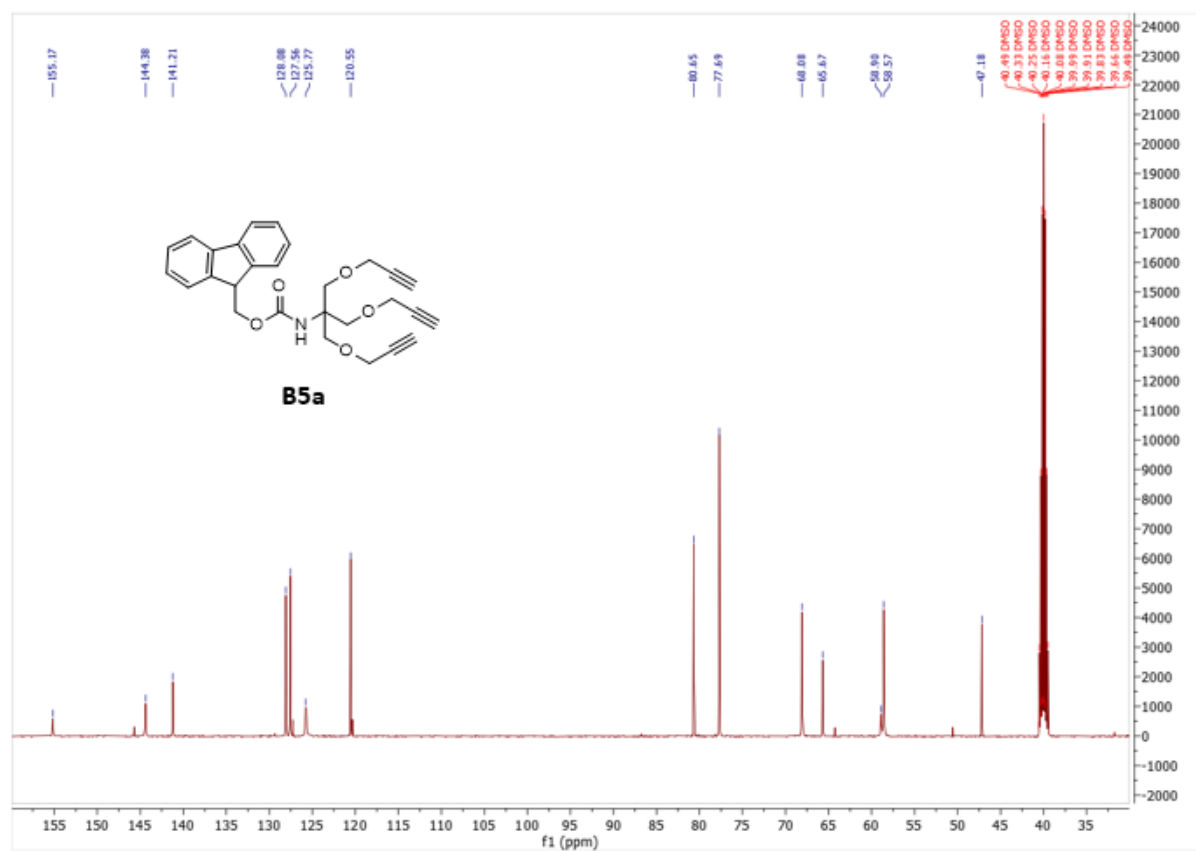
6.2 NMR Spectra

NMR spectra of novel compounds are shown in this section. Reagents for reactions with oligonucleotides are characterised by high resolution mass spectroscopy (HRMS) illustrated in the next section.

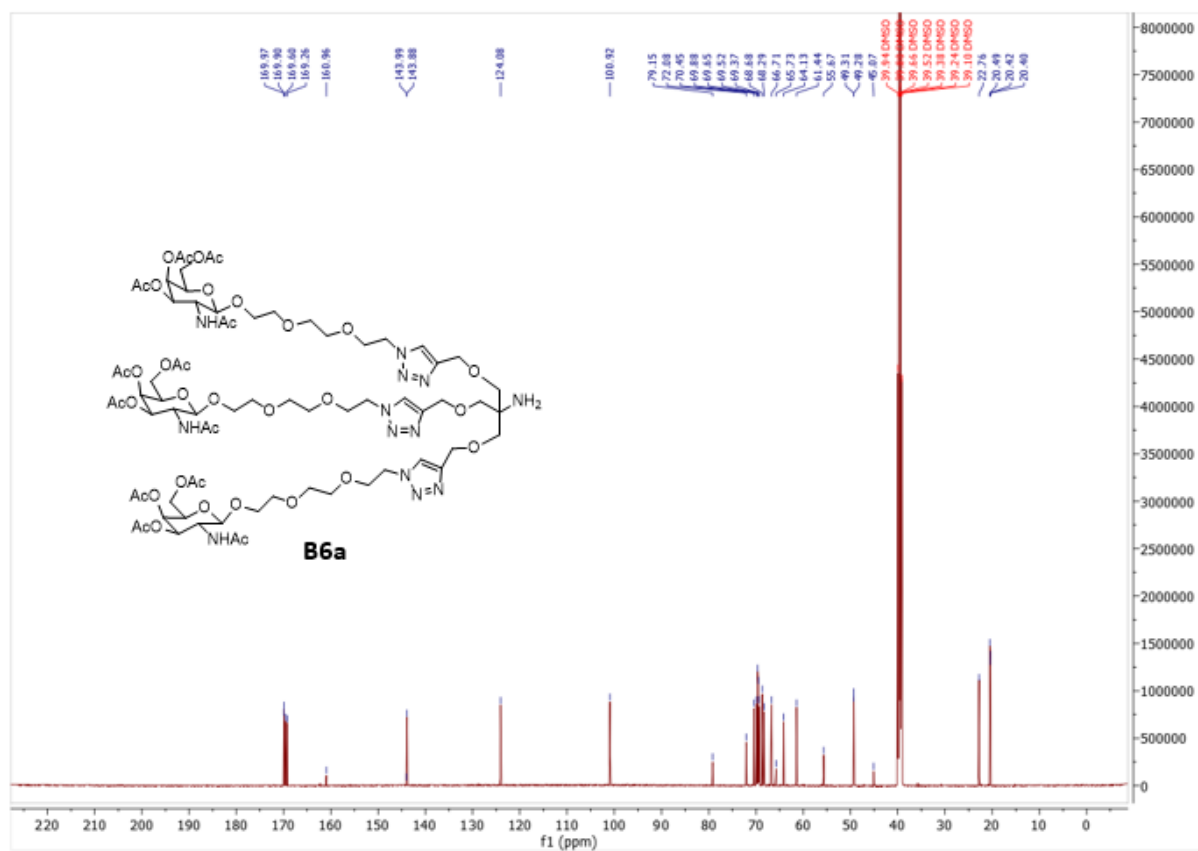
¹H and ¹³C NMR spectra of compound A2

¹H and ¹³C NMR spectra of compound A2a

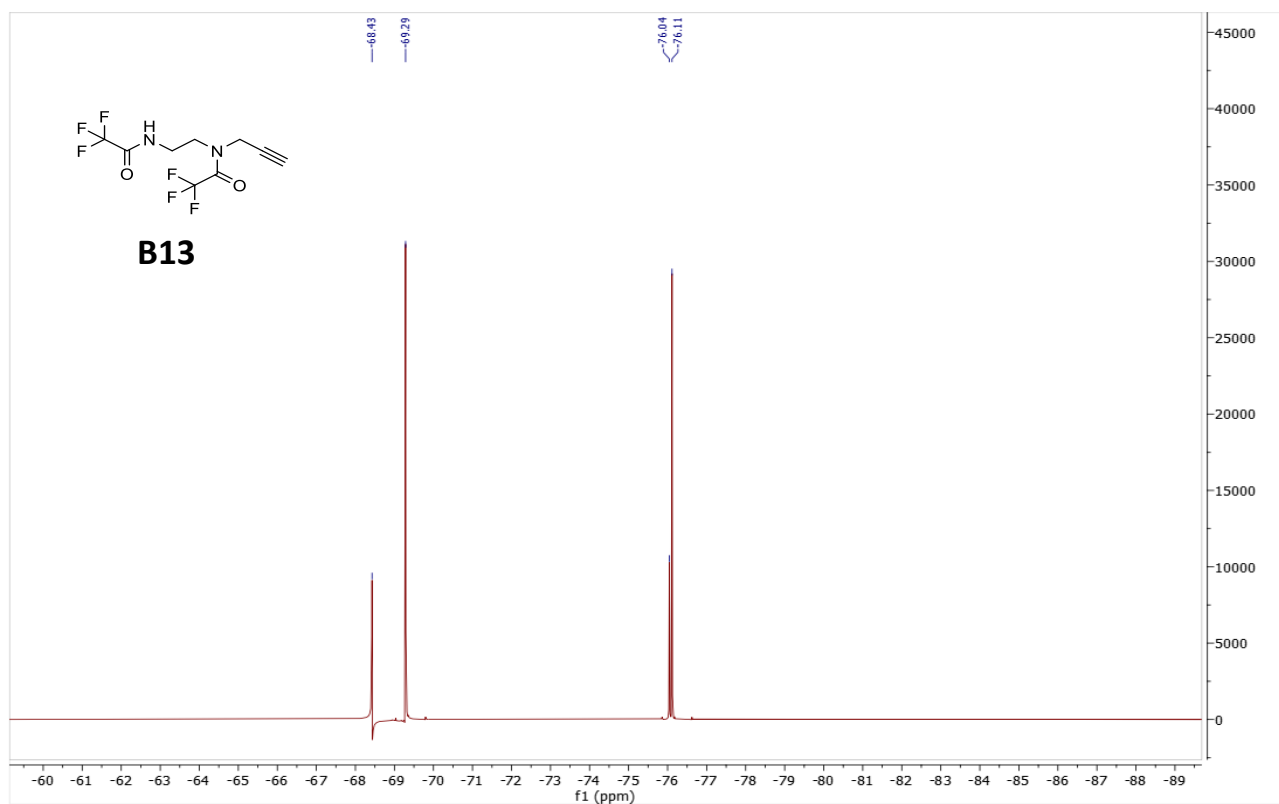
¹H and ¹³C NMR spectra of compound A3

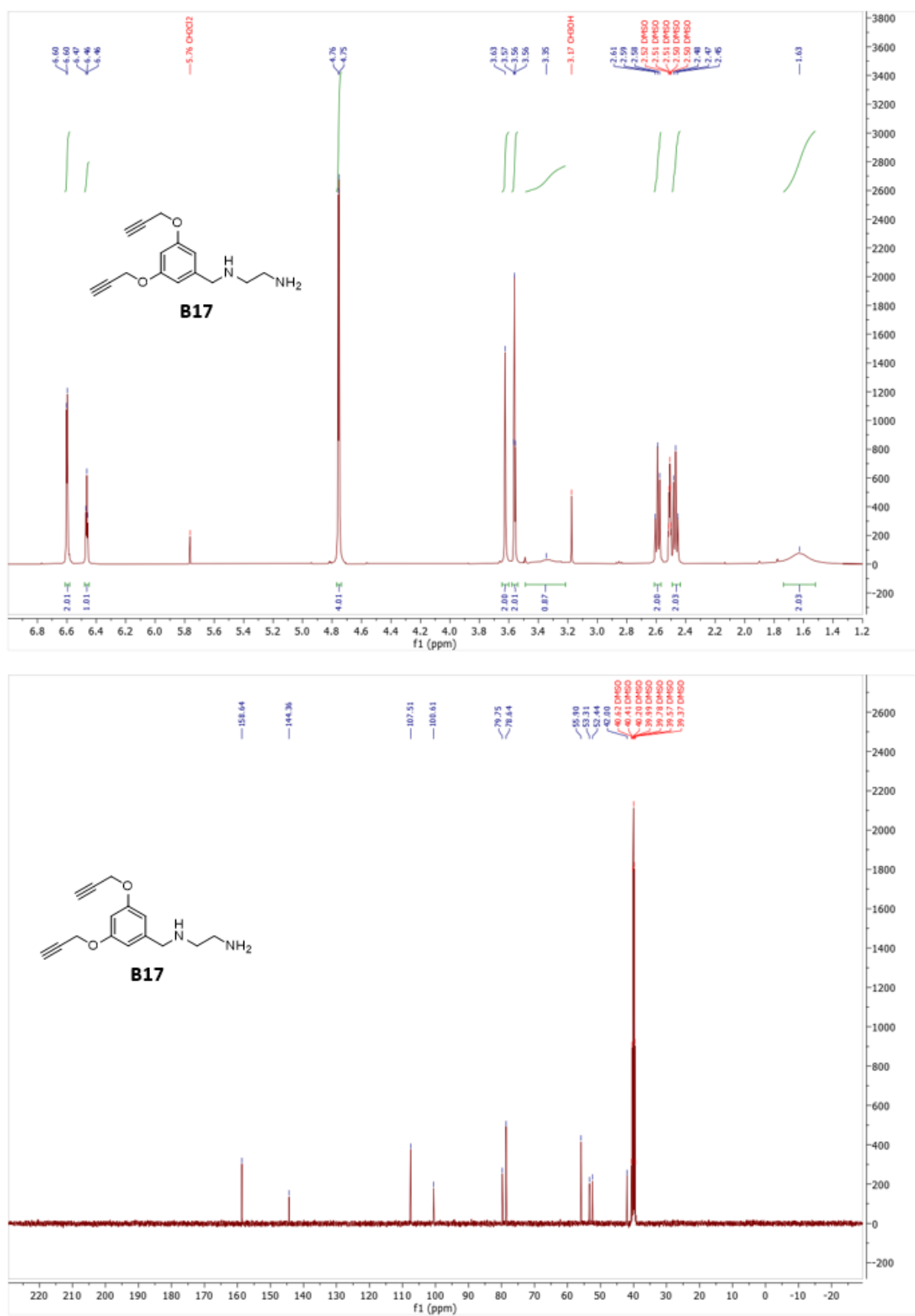


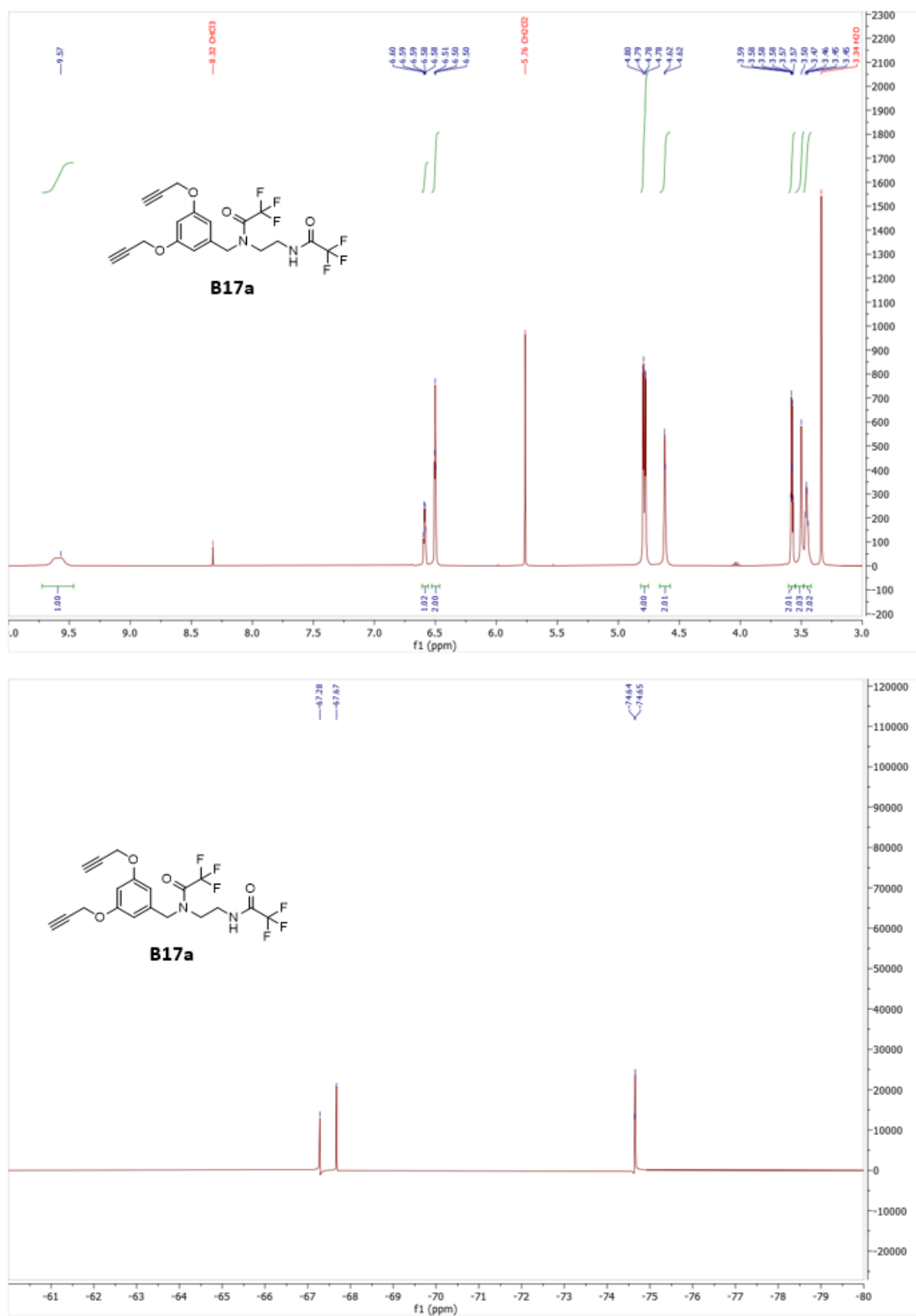




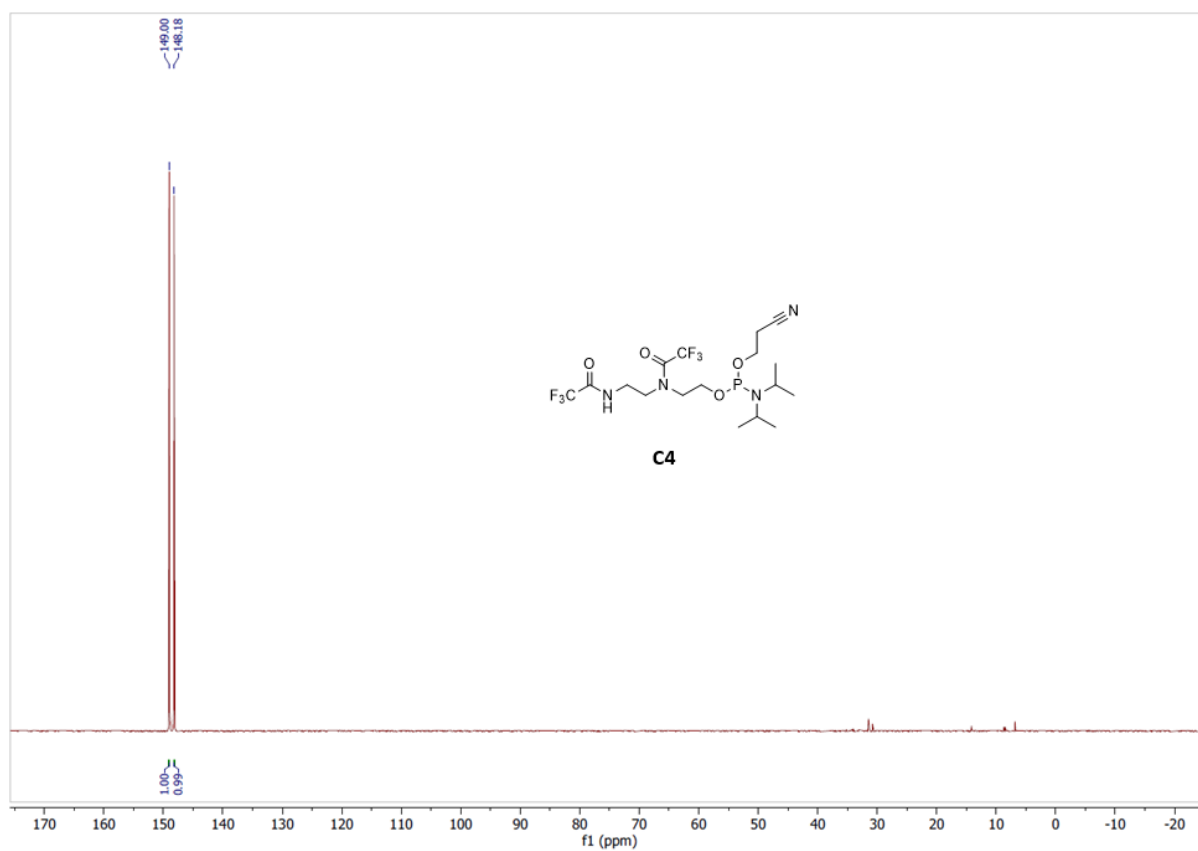
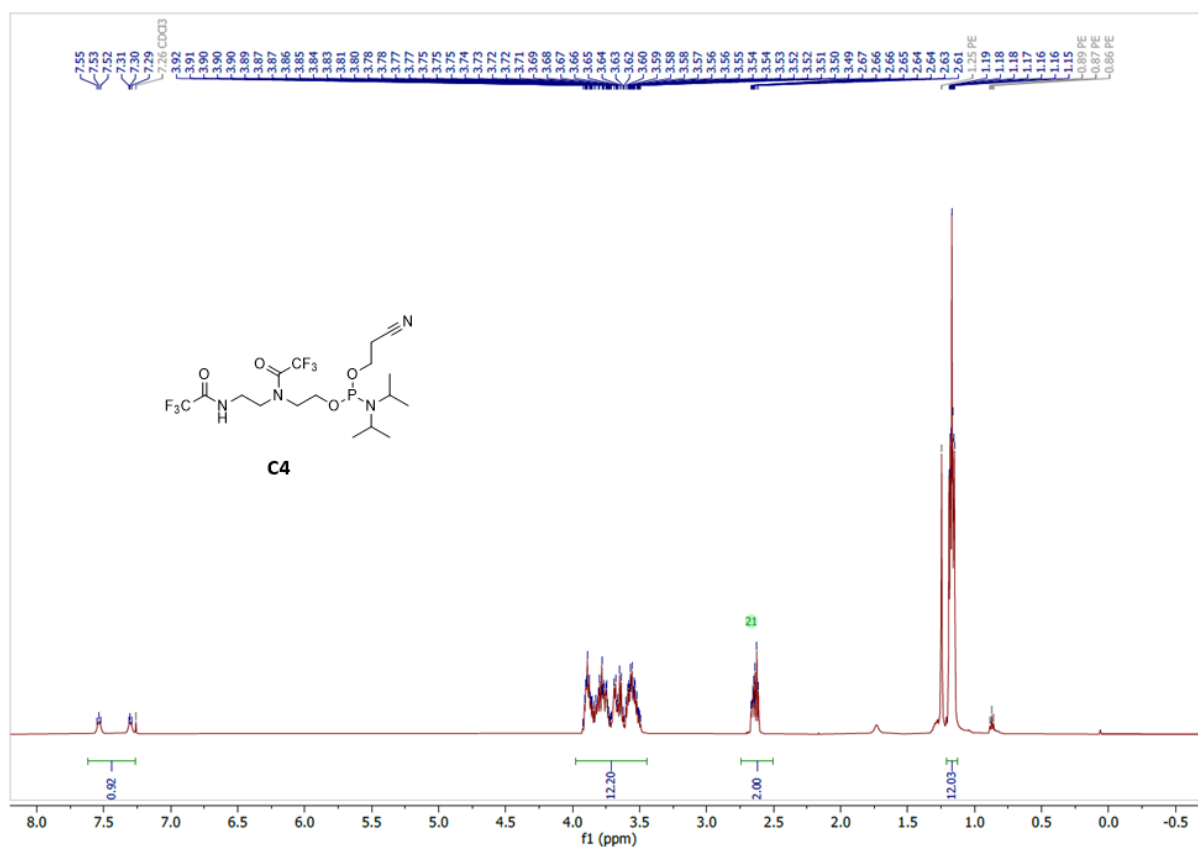




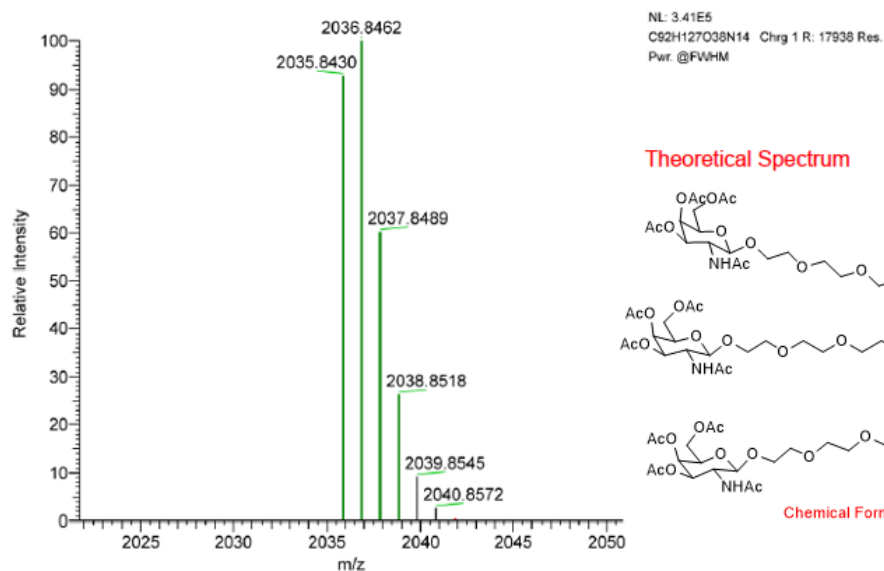
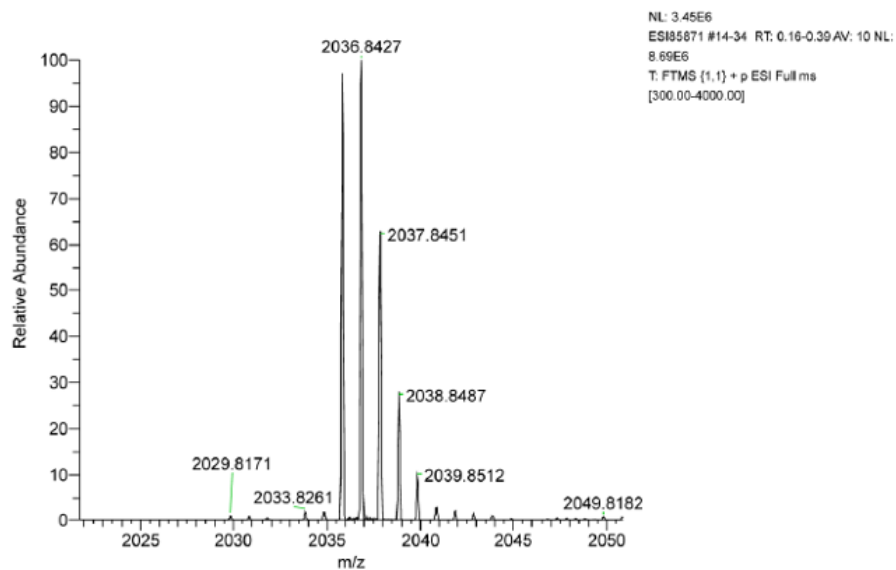
¹H and ¹³C NMR spectra of compound B17



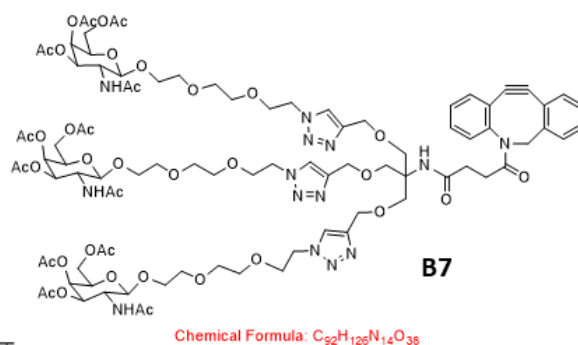
¹H and ¹⁹F NMR spectra of compound B17a

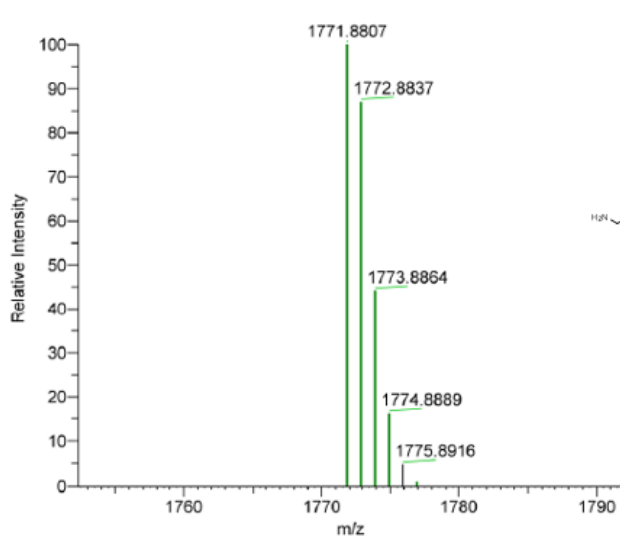
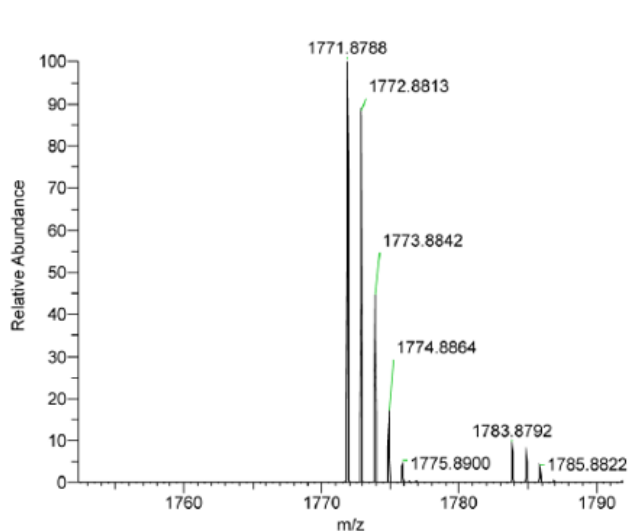
¹H and ³¹P NMR spectra of compound C4

6.3 High-Resolution Mass Spectra

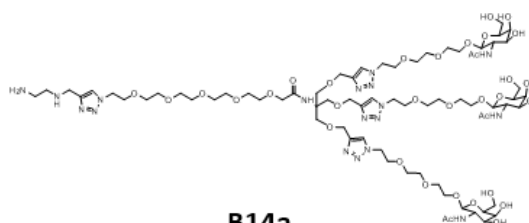


Theoretical Spectrum

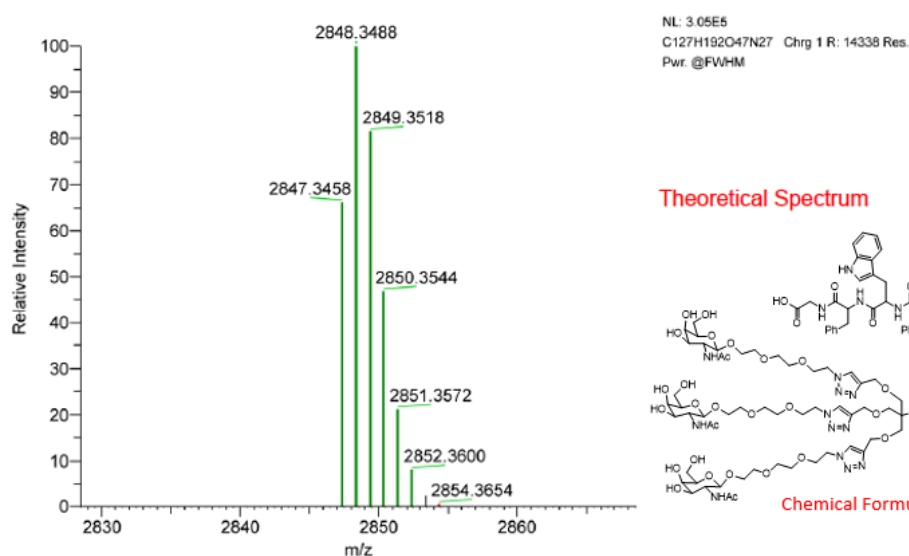
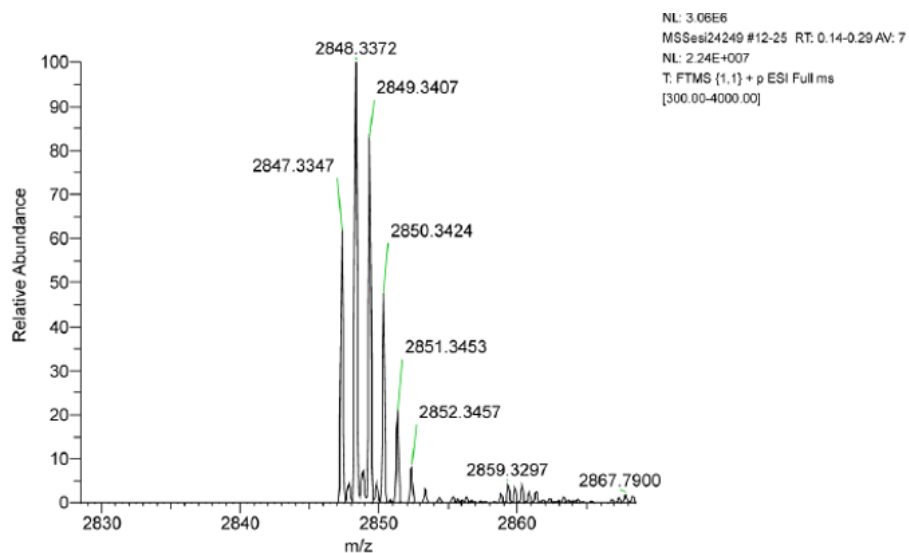




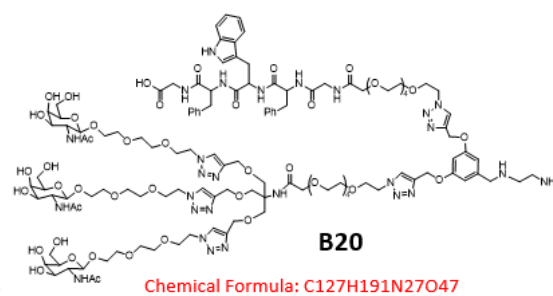
Theoretical spectrum



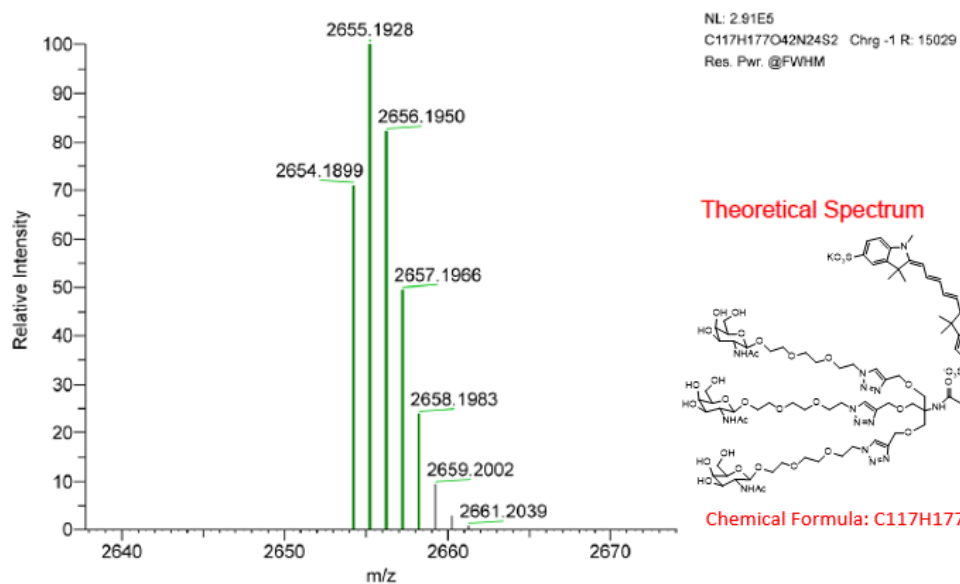
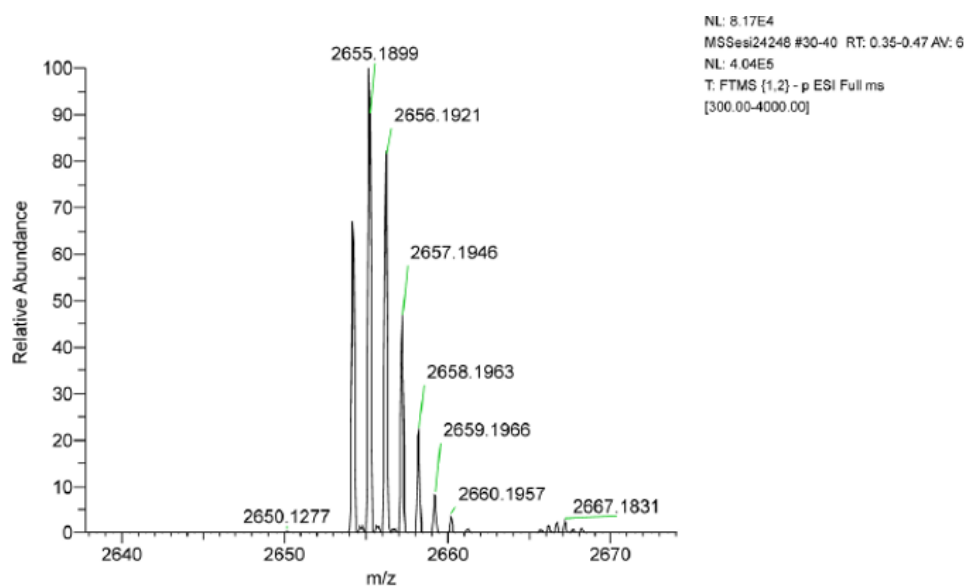
Chemical Formula: C72H126N18O33



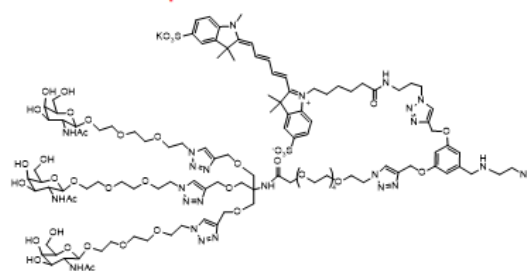
Theoretical Spectrum



Chapter – 6 Appendix



Theoretical Spectrum



Chemical Formula: C117H177N24O42S2-

6.4 Supplementary Experimental Data

Oligonucleotide	Sequence (5'-to-3')
ON4.1	CCACCACCACCACCACCACUAAUAGUGAAGGCGGCCGCAGCUCGCUUUGUUUAAAC
ON4.2	XTGATTCATTCTAATAGCAGC
ON4.3	GCTGCTATTAGAATGAATCAGTTTAAACAAAGCGAGCTGC
ON4.4	GCTGCTATTAGAATGAATCAYGTTTAAACAAAGCGAGCTGC

Table S1 | List of oligonucleotide sequences in RNA-EDA-oligo conjugation studies

X = N-ethyl ethylenediamine; Y = C3 spacer

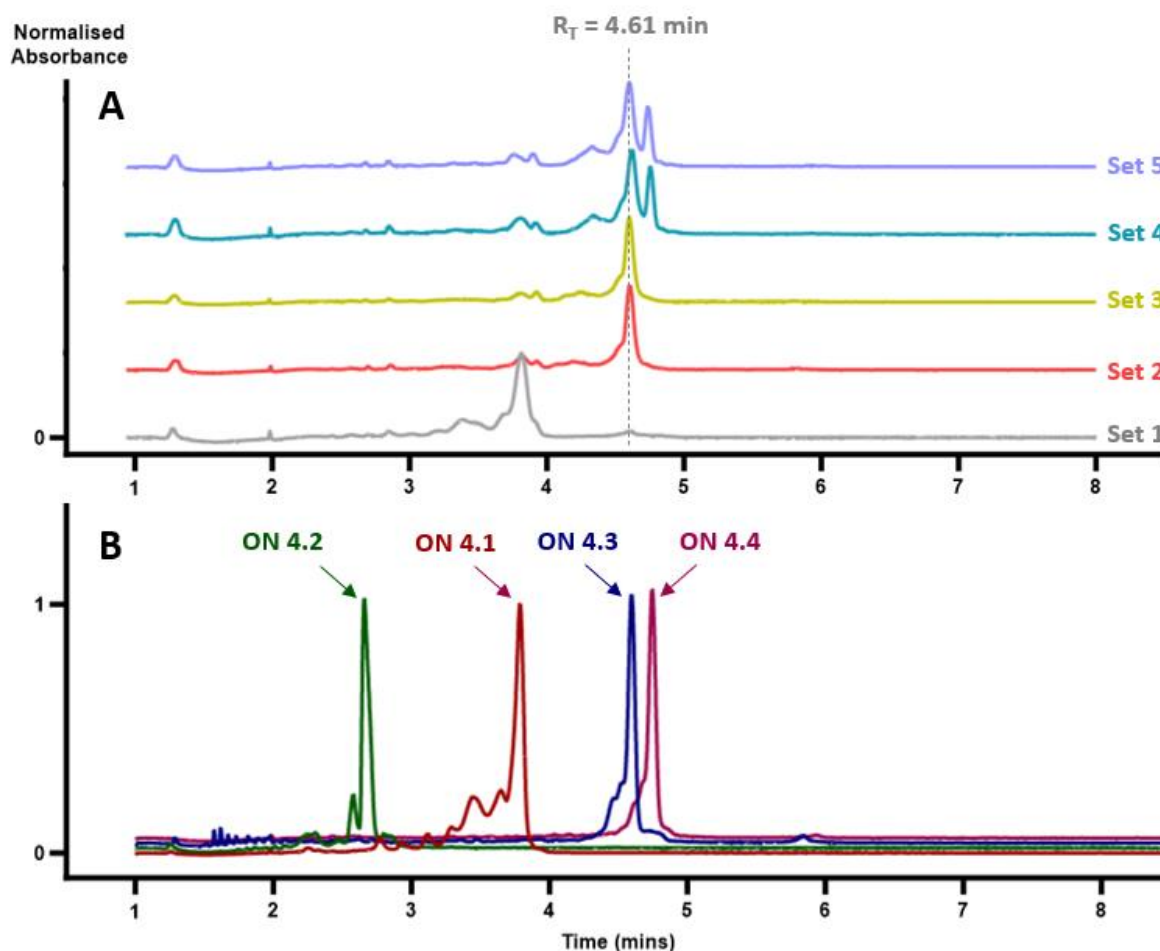


Figure S1 | UPLC results from screening of RNA-EDA-oligo conjugation conditions

(A) Overlaid UPL chromatograms show the retention profiles of the isolated RNA species after precipitation under various conditions with reductive amination. General reaction condition: RNA (final conc. = 10 μ M) was first incubated with NaIO_4 (1.5 mM) with or without a splint (1 eq.) in darkness for 30 mins, then to the mixture was added various components to a final conc.: KH_2PO_4 buffer (100 mM, pH 6.0), 5'-EDA oligo (100 μ M, 10 eq.) and $\text{NaBH}_3(\text{CN})$ (20 mM) for 2 hours at different temperatures. Set 1: no splint was used, reacted at 25°C; set 2: ODN4.3 as splint, reacted at 25°C; set 3: ODN4.3 as splint, reacted at 37°C; set 4: ODN4.4 as splint, reacted at 25°C; set 5: ODN4.4 as splint, reacted at 37°C. The RNA species was isolated by Monarch[®] RNA Cleanup Kit. Ligated product was identified at the major peak at $R_t = 4.61$ min. (B) Retention profiles of the crude oligonucleotides employed in this study: ON4.1 – 56-mer RNA (calc. mass = 17797; found mass = 17799); ON4.2 – 20-mer 5'-EDA oligo (calc. mass = 6256; found mass = 6257); ON4.3 – 40-mer splint (calc. mass = 12352; found mass = 12351); ON4.4 – 40-mer splint with internal C3 spacer (calc. mass = 12488; found mass = 12490).

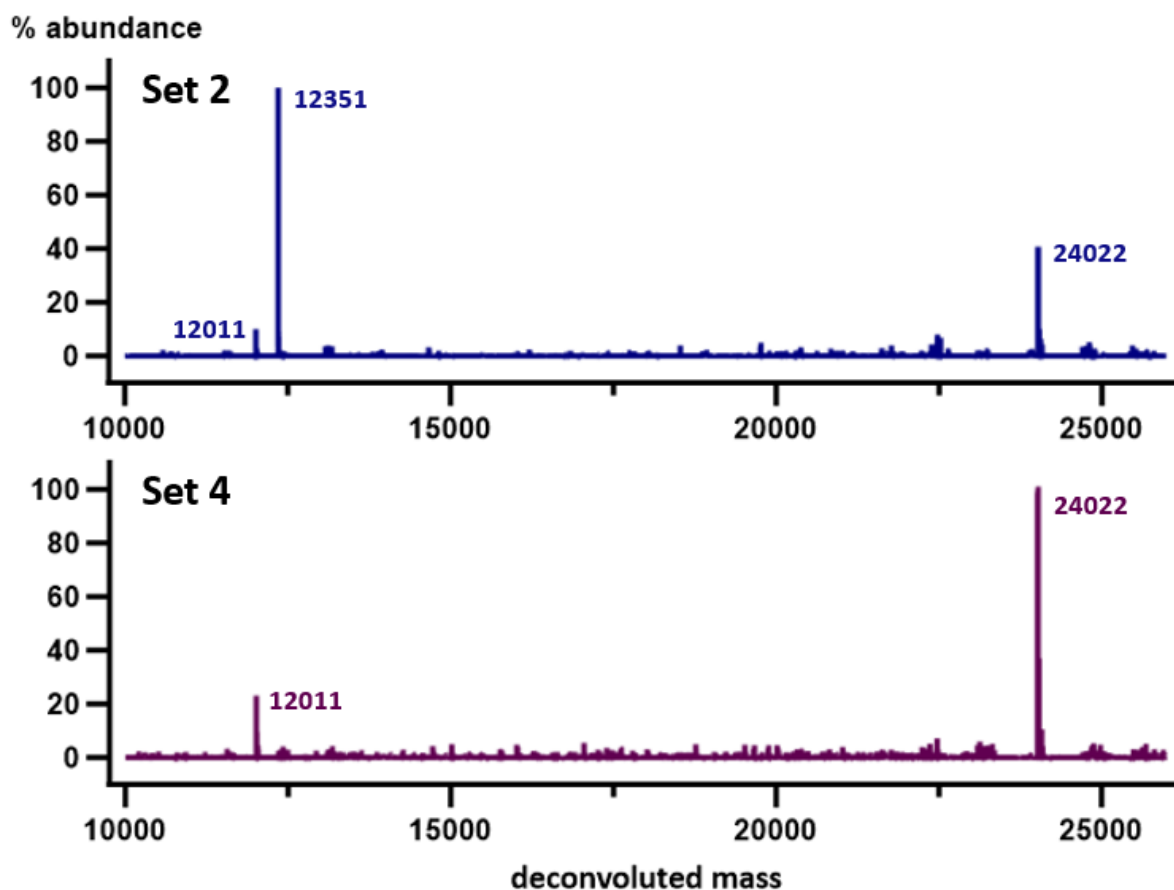


Figure S2 | MS results from screening of RNA-EDA-oligo conjugation reaction

Integration of the peak at $R_t = 4.61$ mins of corresponding chromatograms (Figure S1) over the RNA conjugation reaction indicates the presence of the ligation product (calc. mass = 24019). In set 2 (and set 3), the splint oligo ON4.2 (calc. mass = 12352) was eluted at the same R_t as the product whereas in set 4 (and set 5), the peak indicates the presence exclusively of the product.

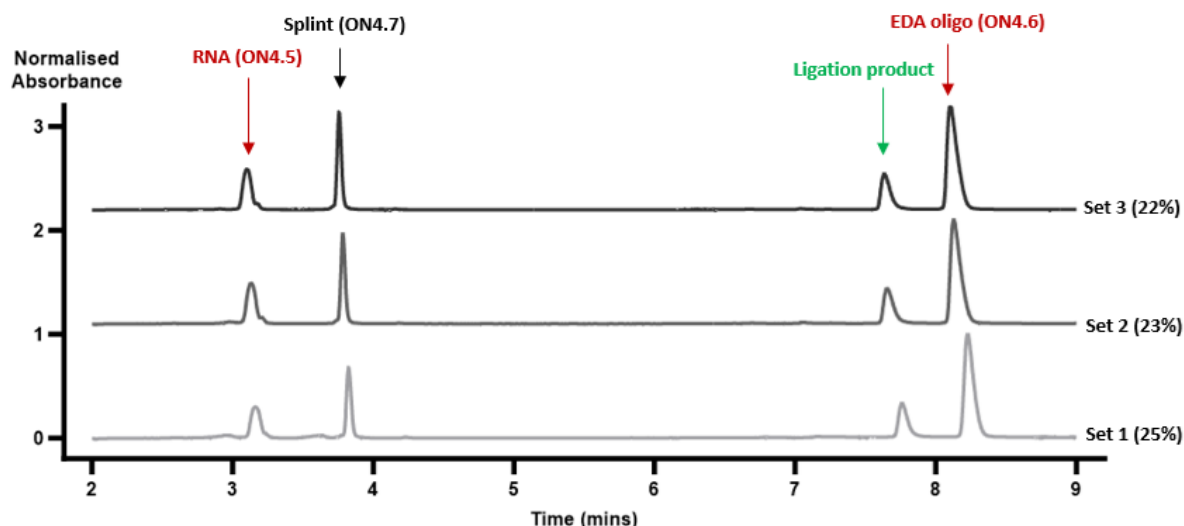


Figure S3 | UPLC results from RNA-EDA-oligo conjugation at annealing conditions

Overlaid UPL chromatograms show the retention profiles of the isolated RNA species after precipitation under various conditions with reductive amination. Before injecting to UPLC-MS, to the mixture was added Et3N.EDTA (final conc. = 50 mM) to remove sodium adducts at early elution. Conversion % in parentheses was calculated as the ratio between area under the product peak and that of the RNA starting material and product. General oxidation-reductive amination condition: RNA (final conc. = 5 μ M) was annealed with reactant oligos from 95°C to 25°C, incubated with NaIO₄ (1.5 mM) in KH₂PO₄ buffer (100 mM, pH 6.0) in darkness for 30 mins, then to the mixture was added NaBH₃(CN) (20 mM) for 2 hours at r.t. Set 1: splint (1 eq.) was added during oxidation, annealing rate = -0.4°C/s and EDA-oligo was added after oxidation; set 2: splint (1 eq.) and EDA-oligos were added during oxidation, annealing rate = -0.4°C/s; set 3: splint (1 eq.) and EDA-oligos were added during oxidation, annealing rate = -0.1°C/s.

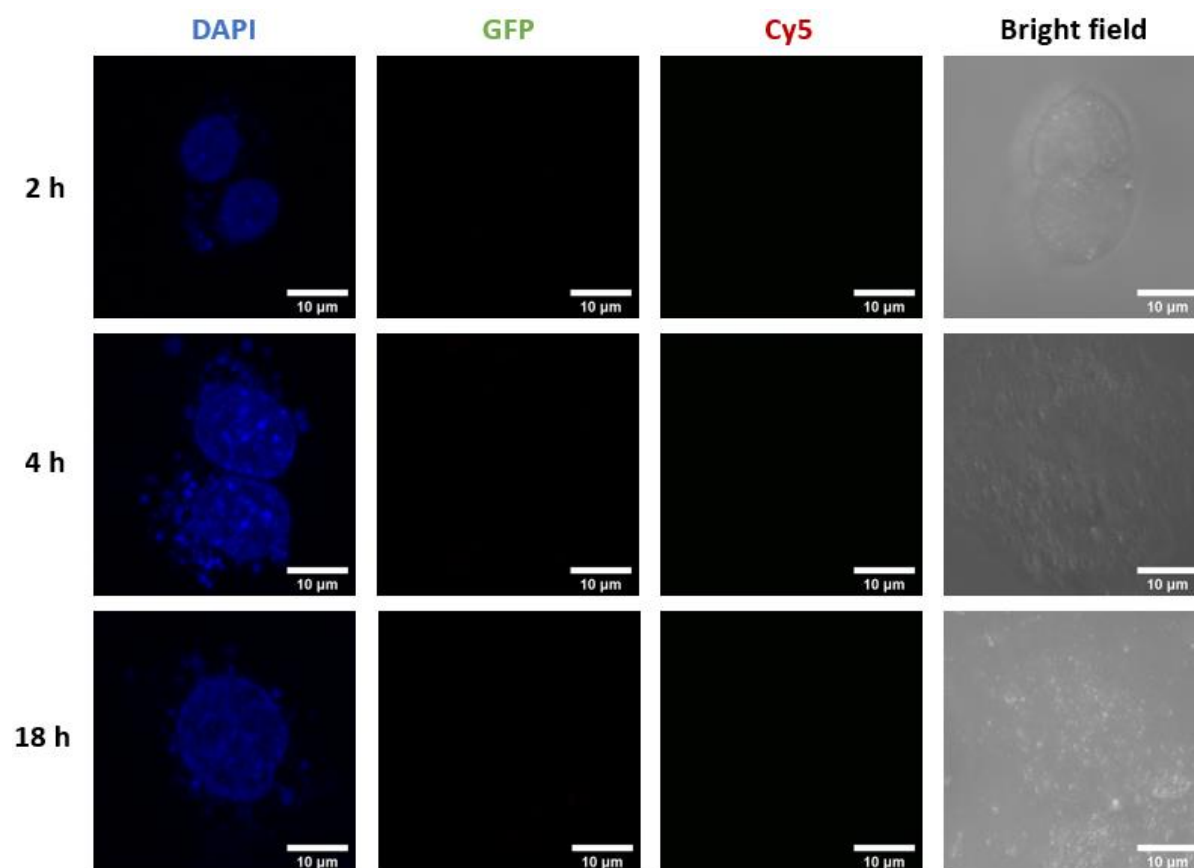


Figure S4 | Images of RNP localisation in HEPG2 cells in split channels (untreated control)

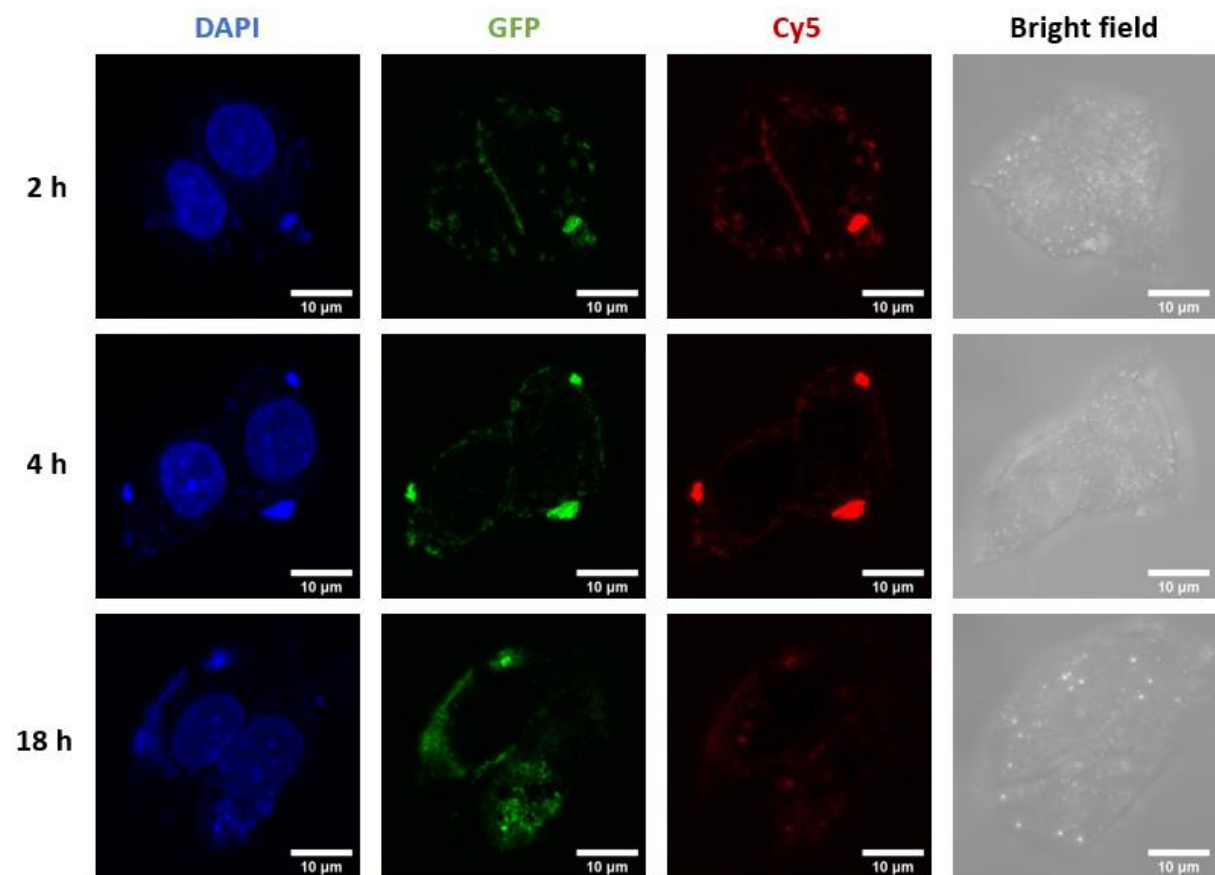


Figure S5 | Images of RNP localisation in HEPG2 cells in split channels (eGFP-Cas9:ON3.5-EDA-GalNAc-Cy5)

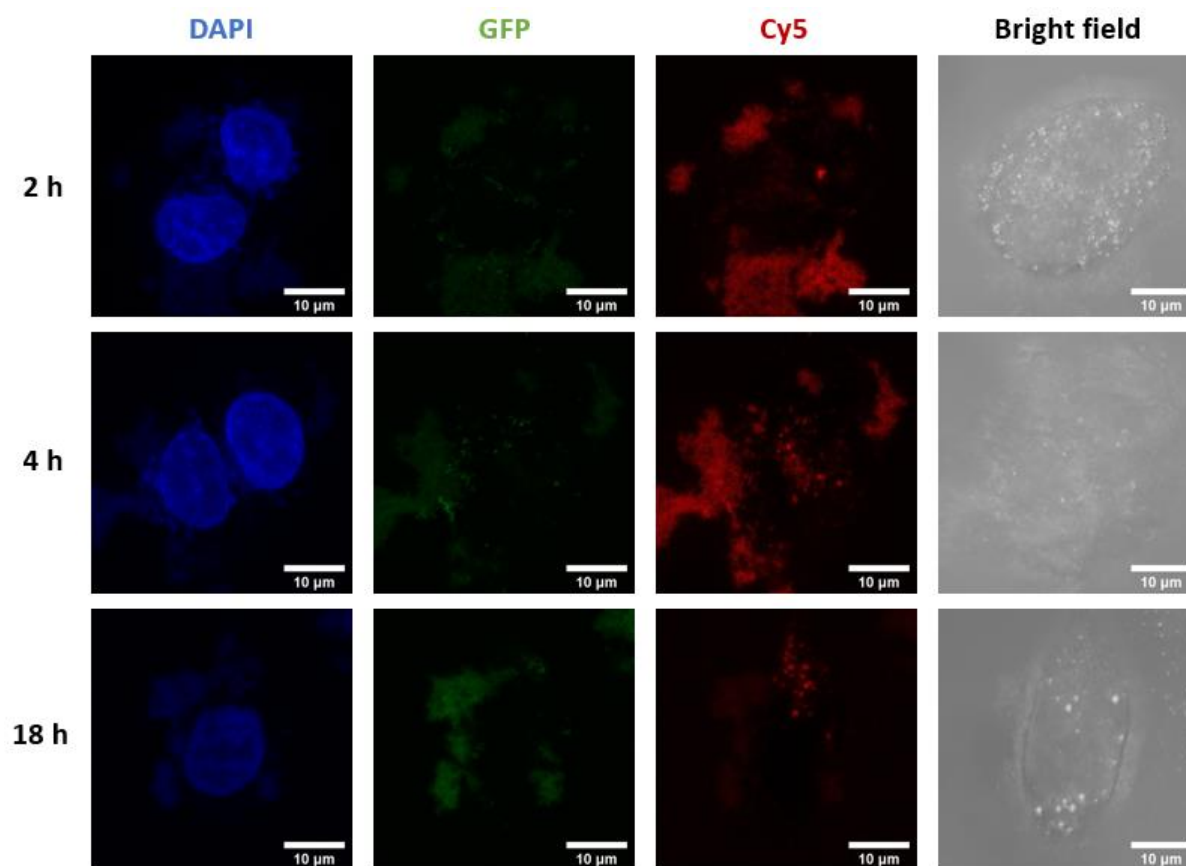


Figure S6 | Images of RNP localisation in HEPG2 cells in split channels (eGFP-Cas9:ON3.5-EDA-GalNAc-Cy5 with ppTG21)

References

- (1) Blackburn, G. M.; Gait, M. J.; Loakes, D.; Williams, D. Chapter 2. DNA and RNA Structure. In *Nucleic Acids in Chemistry and Biology*; Royal Society of Chemistry: Cambridge, 2007; pp 13–76.
- (2) Egli, M. Sugar Pucker and Nucleic Acid Structure. **2018**, 309–315.
- (3) Neidle, S.; Sanderson, M. The Building Blocks of DNA and RNA. *Princ. Nucleic Acid Struct.* **2022**, 29–51.
- (4) Richmond, T. J.; Davey, C. A. The Structure of DNA in the Nucleosome Core. *Nature* **2003**, 423 (6936), 145–150.
- (5) CHARGAFF, E.; LIPSHITZ, R.; GREEN, C. Composition of the Desoxypentose Nucleic Acids of Four Genera of Sea-Urchin. *J. Biol. Chem.* **1952**, 195 (1), 155–160.
- (6) Franklin, R. E.; Gosling, R. G. The Structure of Sodium Thymonucleate Fibres. I. The Influence of Water Content. *Acta Crystallogr.* **1953**, 6 (8), 673–677.
- (7) FRANKLIN, R. E.; GOSLING, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature* **1953**, 171 (4356), 740–741.
- (8) WATSON, J. D.; CRICK, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, 171 (4356), 737–738.
- (9) Lindman, B.; Medronho, B.; Alves, L.; Norgren, M.; Nordenskiöld, L. Hydrophobic Interactions Control the Self-Assembly of DNA and Cellulose. *Q. Rev. Biophys.* **2021**, 54 (2016).
- (10) Karabiyik, H.; Sevinçek, R.; Karabiyik, H. π -Cooperativity Effect on the Base Stacking Interactions in DNA: Is There a Novel Stabilization Factor Coupled with Base Pairing H-Bonds? *Phys. Chem. Chem. Phys.* **2014**, 16 (29), 15527–15538.
- (11) Kertesz, M. Pancake Bonding: An Unusual Pi-Stacking Interaction. *Chem. - A Eur. J.* **2019**, 25 (2), 400–416.
- (12) Barciszewski, J.; Clark, B. F. C. Why RNA? In *RNA Biochemistry and Biotechnology*; Springer Netherlands: Dordrecht, 1999; pp 1–9.
- (13) Neidle, S.; Sanderson, M. RNA Structures and Their Diversity. In *Principles of Nucleic Acid Structure*; Elsevier, 2022; pp 287–346.
- (14) Lesnik, E. A.; Freier, S. M. Relative Thermodynamic Stability of DNA, RNA, and DNA:RNA Hybrid Duplexes: Relationship with Base Composition and Structure. *Biochemistry* **1995**, 34 (34), 10807–10815.
- (15) Crick, F. H. On Protein Synthesis. *Symp. Soc. Exp. Biol.* **1958**, 12, 138–163.
- (16) Crick, F. H. Central Dogma of Molecular Biology. *Nature* **1970**, 227 (5258), 561–563.
- (17) Revyakin, A.; Liu, C.; Ebright, R. H.; Strick, T. R. Abortive Initiation and Productive Initiation by RNA Polymerase Involve DNA Scrunching. *Science (80-.).* **2006**, 314 (5802), 1139–1143.
- (18) Goldman, S. R.; Ebright, R. H.; Nickels, B. E. Direct Detection of Abortive RNA Transcripts in Vivo. *Science (80-.).* **2009**, 324 (5929), 927–928.

References

- (19) Des Georges, A.; Hashem, Y.; Unbehauen, A.; Grassucci, R. A.; Taylor, D.; Hellen, C. U. T.; Pestova, T. V.; Frank, J. Structure of the Mammalian Ribosomal Pre-Termination Complex Associated with ERF1•eRF3•GDPNP. *Nucleic Acids Res.* **2014**, 42 (5), 3409–3418.
- (20) Lagerkvist, U. “Two out of Three”: An Alternative Method for Codon Reading. *Proc. Natl. Acad. Sci. U. S. A.* **1978**, 75 (4), 1759–1762.
- (21) Dill, K. A. Dominant Forces in Protein Folding. *Biochemistry* **1990**, 29 (31), 7133–7155.
- (22) Newberry, R. W.; Raines, R. T. Secondary Forces in Protein Folding. *ACS Chem. Biol.* **2019**, 14 (8), 1677–1686.
- (23) Nick Pace, C.; Martin Scholtz, J.; Grimsley, G. R. Forces Stabilizing Proteins. *FEBS Lett.* **2014**, 588 (14), 2177–2184.
- (24) Srinivas, U. S.; Tan, B. W. Q.; Vellayappan, B. A.; Jeyasekharan, A. D. ROS and the DNA Damage Response in Cancer. *Redox Biol.* **2019**, 25 (October 2018), 101084.
- (25) Rastogi, R. P.; Richa; Kumar, A.; Tyagi, M. B.; Sinha, R. P. Molecular Mechanisms of Ultraviolet Radiation-Induced DNA Damage and Repair. *J. Nucleic Acids* **2010**, 2010.
- (26) Wyatt, M. D.; Pittman, D. L. Methylating Agents and DNA Repair Responses: Methylated Bases and Sources of Strand Breaks. *Chem. Res. Toxicol.* **2006**, 19 (12), 1580–1594.
- (27) Huang, R.; Zhou, P. K. *DNA Damage Repair: Historical Perspectives, Mechanistic Pathways and Clinical Translation for Targeted Cancer Therapy*; Springer US, 2021; Vol. 6.
- (28) Hossain, M. A.; Lin, Y.; Yan, S. Single-Strand Break End Resection in Genome Integrity: Mechanism and Regulation by APE2. *Int. J. Mol. Sci.* **2018**, 19 (8), 1–13.
- (29) Roberts, T. C.; Langer, R.; Wood, M. J. A. Advances in Oligonucleotide Drug Delivery. *Nat. Rev. Drug Discov.* **2020**, 19 (10), 673–694.
- (30) Tran, P.; Weldemichael, T.; Liu, Z.; Li, H. Y. Delivery of Oligonucleotides: Efficiency with Lipid Conjugation and Clinical Outcome. *Pharmaceutics* **2022**, 14 (2).
- (31) Hyjek, M.; Figiel, M.; Nowotny, M. RNases H: Structure and Mechanism. *DNA Repair (Amst).* **2019**, 84 (July), 102672.
- (32) Crooke, S. T. Molecular Mechanisms of Antisense Oligonucleotides. *Nucleic Acid Ther.* **2017**, 27 (2), 70–77.
- (33) Monia, B. P.; Lesnik, E. A.; Gonzalez, C.; Lima, W. F.; McGee, D.; Guinosso, C. J.; Kawasaki, A. M.; Cook, P. D.; Freier, S. M. Evaluation of 2'-Modified Oligonucleotides Containing 2'-Deoxy Gaps as Antisense Inhibitors of Gene Expression. *J. Biol. Chem.* **1993**, 268 (19), 14514–14522.
- (34) Aartsma-Rus, A.; Straub, V.; Hemmings, R.; Haas, M.; Schlosser-Weber, G.; Stoyanova-Beninska, V.; Mercuri, E.; Muntoni, F.; Sepodes, B.; Vroom, E.; Balabanov, P. Development of Exon Skipping Therapies for Duchenne Muscular Dystrophy: A Critical Review and a Perspective on the Outstanding Issues. *Nucleic Acid Ther.* **2017**, 27 (5), 251–259.

References

- (35) Dominski, Z.; Kole, R. Restoration of Correct Splicing in Thalassemic Pre-mRNA by Antisense Oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (18), 8673–8677.
- (36) Voinnet, O. Induction and Suppression of RNA Silencing: Insights from Viral Infections. *Nat. Rev. Genet.* **2005**, *6* (3), 206–220.
- (37) Kong, X.; Yang, M.; Le, B. H.; He, W.; Hou, Y. The Master Role of siRNAs in Plant Immunity. *Mol. Plant Pathol.* **2022**, *23* (10), 1565–1574.
- (38) Tuschl, T.; Elbashir, S. M.; Harborth, J.; Lendeckel, W.; Yalcin, A.; Weber, K. Duplexes of 21-Nucleotide RNAs Mediate RNA Interference in Cultured Mammalian Cells. *Nature* **2001**, *411* (May), 494–498.
- (39) Song, E.; Lee, S. K.; Wang, J.; Ince, N.; Ouyang, N.; Min, J.; Chen, J.; Shankar, P.; Lieberman, J. RNA Interference Targeting Fas Protects Mice from Fulminant Hepatitis. *Nat. Med.* **2003**, *9* (3), 347–351.
- (40) Ishino, Y.; Shinagawa, H.; Makino, K.; Amemura, M.; Nakamura, A. Nucleotide Sequence of the *iap* Gene, Responsible for Alkaline Phosphatase Isoenzyme Conversion in *Escherichia coli*, and Identification of the Gene Product. *J. Bacteriol.* **1987**, *169* (12), 5429–5433.
- (41) Mojica, F. J. M.; Díez-Villaseñor, C.; García-Martínez, J.; Soria, E. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J. Mol. Evol.* **2005**, *60* (2), 174–182.
- (42) Bolotin, A.; Quinquis, B.; Sorokin, A.; Dusko Ehrlich, S. Clustered Regularly Interspaced Short Palindrome Repeats (CRISPRs) Have Spacers of Extrachromosomal Origin. *Microbiology* **2005**, *151* (8), 2551–2561.
- (43) 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* (3), 653–663.
- (44) Brouns, S. J. J.; Jore, M. M.; Lundgren, M.; Westra, E. R.; Slijkhuis, R. J. H.; Snijders, A. P. L.; Dickman, M. J.; Makarova, K. S.; Koonin, E. V.; van der Oost, J. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* (80-.). **2008**, *321* (5891), 960–964.
- (45) 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* (80-.). **2007**, *315* (5819), 1709–1712.
- (46) 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* (80-.). **2013**, *339* (6121), 823–826.
- (47) 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* (80-.). **2013**, *339* (6121), 819–823.
- (48) Jinek, M.; East, A.; Cheng, A.; Lin, S.; Ma, E.; Doudna, J. RNA-Programmed Genome Editing in Human Cells. *Elife* **2013**, *2013* (2), 1–9.
- (49) Adli, M. The CRISPR Tool Kit for Genome Editing and Beyond. *Nat. Commun.* **2018**, *9* (1).

References

- (50) 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), 1173–1183.
- (51) 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* (80-.). **2019**, *365* (6459), 1301–1305.
- (52) Chaudhary, N.; Weissman, D.; Whitehead, K. A. mRNA Vaccines for Infectious Diseases: Principles, Delivery and Clinical Translation. *Nat. Rev. Drug Discov.* **2021**, *20* (11), 817–838.
- (53) Choe, J.; Lin, S.; Zhang, W.; Liu, Q.; Wang, L.; Ramirez-Moya, J.; Du, P.; Kim, W.; Tang, S.; Sliz, P.; Santisteban, P.; George, R. E.; Richards, W. G.; Wong, K. K.; Locker, N.; Slack, F. J.; Gregory, R. I. mRNA Circularization by METTL3–EIF3h Enhances Translation and Promotes Oncogenesis. *Nature* **2018**, *561* (7724), 556–560.
- (54) Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149–2154.
- (55) Andrews, B. I.; Antia, F. D.; Brueggemeier, S. B.; Diorazio, L. J.; Koenig, S. G.; Kopach, M. E.; Lee, H.; Olbrich, M.; Watson, A. L. Sustainability Challenges and Opportunities in Oligonucleotide Manufacturing. *J. Org. Chem.* **2021**, *86* (1), 49–61.
- (56) Michelson, A. M.; Todd, A. R. Nucleotides Part XXXII. Synthesis of a Dithymidine Dinucleotide Containing a 3': 5'-Internucleotidic Linkage. *J. Chem. Soc.* **1955**, *22* (2), 2632–2638.
- (57) Hall, R. H.; Todd, A.; Webb, R. F. 644. Nucleotides. Part XLI. Mixed Anhydrides as Intermediates in the Synthesis of Dinucleoside Phosphates. *J. Chem. Soc.* **1957**, 3291.
- (58) Gilham, P. T.; Khorana, H. G. Studies on Polynucleotides. I. A New and General Method for the Chemical Synthesis of the C 5 "-C 3 " Internucleotidic Linkage. Syntheses of Deoxyribo-Dinucleotides 1. *J. Am. Chem. Soc.* **1958**, *80* (23), 6212–6222.
- (59) Reese, C. B. The Chemical Synthesis of Oligo- and Poly-Nucleotides by the Phosphotriester Approach. *Tetrahedron* **1978**, *34* (21), 3143–3179.
- (60) Letsinger, R. L.; Ogilvie, K. K. Nucleotide Chemistry. XIII. Synthesis of Oligothymidylates via Phosphotriester Intermediates. *J. Am. Chem. Soc.* **1969**, *91* (12), 3350–3355.
- (61) Froehler, B. C.; Ng, P. G.; Matteucci, M. D. Synthesis of DNA via Deoxynucleoside H-Phosphonate Intermediates. *Nucleic Acids Res.* **1986**, *14* (13), 5399–5407.
- (62) Garegg, P. J.; Lindh, I.; Regberg, T.; Stawinski, J.; Strömberg, R.; Henrichson, C. Nucleoside H-Phosphonates. III. Chemical Synthesis of Oligodeoxyribonucleotides by the Hydrogenphosphonate Approach. *Tetrahedron Lett.* **1986**, *27* (34), 4051–4054.
- (63) Caruthers, M. H. Gene Synthesis Machines: DNA Chemistry and Its Uses. *Science* (80-.). **1985**, *230* (4723), 281–285.
- (64) Brown, T.; Brown (Jnr), T. *Solid phase oligonucleotide synthesis*. Nucleic Acids Book. <https://atdbio.com/nucleic-acids-book/Solid-phase-oligonucleotide-synthesis>.

References

- (65) *Deprotection Guide*. Glen Report.
<https://www.glenresearch.com/media/productattach/import/>.
- (66) Clavé, G.; Reverte, M.; Vasseur, J. J.; Smietana, M. Modified Internucleoside Linkages for Nuclease-Resistant Oligonucleotides. *RSC Chem. Biol.* **2021**, 2 (1), 94–150.
- (67) Bano, N.; Ehlinger, C.; Yang, T. yuan; Swanson, M.; Allen, S. Considerations in the Immunogenicity Assessment Strategy for Oligonucleotide Therapeutics (ONTs). *AAPS J.* **2022**, 24 (5).
- (68) Bost, J. P.; Barriga, H.; Holme, M. N.; Gallud, A.; Maugeri, M.; Gupta, D.; Lehto, T.; Valadi, H.; Esbjörner, E. K.; Stevens, M. M.; El-Andaloussi, S. Delivery of Oligonucleotide Therapeutics: Chemical Modifications, Lipid Nanoparticles, and Extracellular Vesicles. *ACS Nano* **2021**, 15 (9), 13993–14021.
- (69) Khvorova, A.; Watts, J. K. The Chemical Evolution of Oligonucleotide Therapies of Clinical Utility. *Nat. Biotechnol.* **2017**, 35 (3), 238–248.
- (70) Deleavey, G. F.; Damha, M. J. Designing Chemically Modified Oligonucleotides for Targeted Gene Silencing. *Chem. Biol.* **2012**, 19 (8), 937–954.
- (71) McKenzie, L. K.; El-Khoury, R.; Thorpe, J. D.; Damha, M. J.; Hollenstein, M. Recent Progress in Non-Native Nucleic Acid Modifications. *Chem. Soc. Rev.* **2021**, 50 (8), 5126–5164.
- (72) Kole, R.; Krainer, A. R.; Altman, S. RNA Therapeutics: Beyond RNA Interference and Antisense Oligonucleotides. *Nat. Rev. Drug Discov.* **2012**, 11 (2), 125–140.
- (73) Choung, S.; Kim, Y. J.; Kim, S.; Park, H. O.; Choi, Y. C. Chemical Modification of SiRNAs to Improve Serum Stability without Loss of Efficacy. *Biochem. Biophys. Res. Commun.* **2006**, 342 (3), 919–927.
- (74) Monia, B. P.; Johnston, J. F.; Sasmor, H.; Cummins, L. L. Nuclease Resistance and Antisense Activity of Modified Oligonucleotides Targeted to Ha-Ras. *J. Biol. Chem.* **1996**, 271 (24), 14533–14540.
- (75) Cummins, L. L.; Owens, S. R.; Risen, L. M.; Lesnik, E. A.; Freier, S. M.; Mc Gee, D.; Cook, C. J.; Cook, P. D. Characterization of Fully 2'-Modified Oligoribonucleotide Hetero-and Homoduplex Hybridization Andnuclease Sensitivity. *Nucleic Acids Res.* **1995**, 23 (11), 2019–2024.
- (76) Sheng, L.; Rigo, F.; Frank Bennett, C.; Krainer, A. R.; Hua, Y. Comparison of the Efficacy of MOE and PMO Modifications of Systemic Antisense Oligonucleotides in a Severe SMA Mouse Model. *Nucleic Acids Res.* **2020**, 48 (6), 2853–2865.
- (77) Manoharan, M.; Akinc, A.; Pandey, R. K.; Qin, J.; Hadwiger, P.; John, M.; Mills, K.; Charisse, K.; Maier, M. A.; Nechev, L.; Greene, E. M.; Pallan, P. S.; Rozners, E.; Rajeev, K. G.; Egli, M. Unique Gene-Silencing and Structural Properties of 2'-Fluoro-Modified SiRNAs. *Angew. Chemie - Int. Ed.* **2011**, 50 (10), 2284–2288.
- (78) Singh, S. K.; Wengel, J. Universality of LNA-Mediated High-Affinity Nucleic Acid Recognition LNA (Locked Nucleic Acid) Is a Novel Class of Nucleic Acid Mimic Structurally Closely Resembling RNA ; Incorporation of Three LNA Monomers Together with Six Ribonucleotide Monomers Affo. *Chem. Commun.* **1998**, 1247–

- 1248.
- (79) Singh, S. K.; Nielsen, P.; Koshkin, A. a.; Wengel, J. LNA (Locked Nucleic Acids): Synthesis and High-Affinity Nucleic Acid Recognition Nucleic Acids), Is Introduced . Following the Watson – Crick Base Pairing Rules , LNA Forms Duplexes with Complements- Stabilities and Generally Improved Selectivities . *Chem. Commun.* **1998**, 455–456.
- (80) Hagedorn, P. H.; Persson, R.; Funder, E. D.; Albæk, N.; Diemer, S. L.; Hansen, D. J.; Møller, M. R.; Papargyri, N.; Christiansen, H.; Hansen, B. R.; Hansen, H. F.; Jensen, M. A.; Koch, T. Locked Nucleic Acid: Modality, Diversity, and Drug Discovery. *Drug Discov. Today* **2018**, 23 (1), 101–114.
- (81) Touznik, A.; Maruyama, R.; Hosoki, K.; Echigoya, Y.; Yokota, T. LNA/DNA Mixmer-Based Antisense Oligonucleotides Correct Alternative Splicing of the SMN2 Gene and Restore SMN Protein Expression in Type 1 SMA Fibroblasts. *Sci. Rep.* **2017**, 7 (1), 1–9.
- (82) Baker, Y. R.; Thorpe, C.; Chen, J.; Poller, L. M.; Cox, L.; Kumar, P.; Lim, W. F.; Lie, L.; McClorey, G.; Epple, S.; Singleton, D.; McDonough, M. A.; Hardwick, J. S.; Christensen, K. E.; Wood, M. J. A.; Hall, J. P.; El-Sagheer, A. H.; Brown, T. An LNA-Amide Modification That Enhances the Cell Uptake and Activity of Phosphorothioate Exon-Skipping Oligonucleotides. *Nat. Commun.* **2022**, 13 (1), 1–11.
- (83) Gilar, M.; Belenky, A.; Smisek, D. L.; Bourque, A.; Cohen, A. S. Kinetics of Phosphorothioate Oligonucleotide Metabolism in Biological Fluids. *Nucleic Acids Res.* **1997**, 25 (18), 3615–3620.
- (84) Crooke, S. T.; Vickers, T. A.; Liang, X. H. Phosphorothioate Modified Oligonucleotide-Protein Interactions. *Nucleic Acids Res.* **2021**, 48 (10), 5235–5253.
- (85) Laurent, Q.; Martinet, R.; Moreau, D.; Winssinger, N.; Sakai, N.; Matile, S. Oligonucleotide Phosphorothioates Enter Cells by Thiol-Mediated Uptake. *Angew. Chemie - Int. Ed.* **2021**, 60 (35), 19102–19106.
- (86) Mutisya, D.; Selvam, C.; Lunstad, B. D.; Pallan, P. S.; Haas, A.; Leake, D.; Egli, M.; Rozners, E. Amides Are Excellent Mimics of Phosphate Internucleoside Linkages and Are Well Tolerated in Short Interfering RNAs. *Nucleic Acids Res.* **2014**, 42 (10), 6542–6551.
- (87) Baraniak, D.; Boryski, J. Triazole-Modified Nucleic Acids for the Application in Bioorganic and Medicinal Chemistry. *Biomedicines* **2021**, 9 (6).
- (88) Lucas, R.; Zerrouki, R.; Granet, R.; Krausz, P.; Champavier, Y. A Rapid Efficient Microwave-Assisted Synthesis of a 3',5'-Pentathymidine by Copper(I)-Catalyzed [3+2] Cycloaddition. *Tetrahedron* **2008**, 64 (23), 5467–5471.
- (89) De Mesmaeker, A.; Waldner, A.; Lebreton, J.; Hoffmann, P.; Fritsch, V.; Wolf, R. M.; Freier, S. M. Amides as a New Type of Backbone Modification in Oligonucleotides. *Angew. Chemie Int. Ed. English* **1994**, 33 (2), 226–229.
- (90) Idziak, I.; Just, G.; Damha, M. J.; Giannaris, P. A. Synthesis and Hybridization Properties of Amide-Linked Thymidine Dimers Incorporated into Oligodeoxynucleotides. *Tetrahedron Lett.* **1993**, 34 (34), 5417–5420.

References

- (91) Epple, S.; Thorpe, C.; Baker, Y. R.; El-Sagheer, A. H.; Brown, T. Consecutive 5'- And 3'-Amide Linkages Stabilise Antisense Oligonucleotides and Elicit an Efficient RNase H Response. *Chem. Commun.* **2020**, 56 (41), 5496–5499.
- (92) Sugo, T.; Terada, M.; Oikawa, T.; Miyata, K.; Nishimura, S.; Kenjo, E.; Ogasawara-Shimizu, M.; Makita, Y.; Imaichi, S.; Murata, S.; Otake, K.; Kikuchi, K.; Teratani, M.; Masuda, Y.; Kamei, T.; Takagahara, S.; Ikeda, S.; Ohtaki, T.; Matsumoto, H. Development of Antibody-SiRNA Conjugate Targeted to Cardiac and Skeletal Muscles. *J. Control. Release* **2016**, 237, 1–13.
- (93) Zhang, H.; Yan, Z.; Wang, X.; Gaňová, M.; Chang, H.; Laššáková, S.; Korabecna, M.; Neuzil, P. Determination of Advantages and Limitations of QPCR Duplexing in a Single Fluorescent Channel. *ACS Omega* **2021**, 6 (34), 22292–22300.
- (94) Nazarenko, I.; Lowe, B.; Darfler, M.; Ikonomi, P.; Schuster, D.; Rashtchian, A. Multiplex Quantitative PCR Using Self-Quenched Primers Labeled with a Single Fluorophore. *Nucleic Acids Res.* **2002**, 30 (9), e37.
- (95) *Labels and Modifiers by Glen Research*. <https://www.glenresearch.com/browse/labels-and-modifiers>.
- (96) *Modified Oligo Synthesis Reagents | LGC, Biosearch Technologies*. <https://shop.biosearchtech.com/nucleic-acid-chemistry-reagents-and-instruments/modified-oligo-synthesis-reagents/c/modified-oligo-synthesis-reagents>.
- (97) Epple, S.; Modi, A.; Baker, Y. R.; Węgrzyn, E.; Traoré, D.; Wanat, P.; Tyburn, A. E. S.; Shivalingam, A.; Taemaitree, L.; El-Sagheer, A. H.; Brown, T. A New 1,5-Disubstituted Triazole DNA Backbone Mimic with Enhanced Polymerase Compatibility. *J. Am. Chem. Soc.* **2021**, 143 (39), 16293–16301.
- (98) Knouse, K. W.; Flood, D. T.; Vantourout, J. C.; Schmidt, M. A.; McDonald, I. M.; Eastgate, M. D.; Baran, P. S. Nature Chose Phosphates and Chemists Should Too: How Emerging P(V) Methods Can Augment Existing Strategies. *ACS Cent. Sci.* **2021**, 7 (9), 1473–1485.
- (99) Knouse, K. W.; deGruyter, J. N.; Schmidt, M. A.; Zheng, B.; Vantourout, J. C.; Kingston, C.; Mercer, S. E.; McDonald, I. M.; Olson, R. E.; Zhu, Y.; Hang, C.; Zhu, J.; Yuan, C.; Wang, Q.; Park, P.; Eastgate, M. D.; Baran, P. S. Unlocking P(V): Reagents for Chiral Phosphorothioate Synthesis. *Science (80-.)*. **2018**, 361 (6408), 1234–1238.
- (100) Von Matt, P.; De Mesmaeker, A.; Pielles, U.; Zürcher, W.; Altmann, K. H. 2'-Deoxyribo-PNAs: A Structurally Novel Class of Polyamide Nucleic Acids with Good RNA and DNA Binding Affinity. *Tetrahedron Lett.* **1999**, 40 (15), 2899–2902.
- (101) Isobe, H.; Fujino, T.; Yamazaki, N.; Guillot-Nieckowski, M.; Nakamura, E. Triazole-Linked Analogue of Deoxyribonucleic Acid (TLDNA): Design, Synthesis, and Double-Strand Formation with Natural DNA. *Org. Lett.* **2008**, 10 (17), 3729–3732.
- (102) Moses, J. E.; Moorhouse, A. D. The Growing Applications of Click Chemistry. *Chem. Soc. Rev.* **2007**, 36 (8), 1249–1262.
- (103) Castro, V.; Rodríguez, H.; Albericio, F. CuAAC: An Efficient Click Chemistry Reaction on Solid Phase. *ACS Comb. Sci.* **2016**, 18 (1), 1–14.
- (104) Fantoni, N. Z.; El-Sagheer, A. H.; Brown, T. A Hitchhiker's Guide to Click-Chemistry

References

- with Nucleic Acids. *Chem. Rev.* **2021**, *121* (12), 7122–7154.
- (105) Morvan, F.; Meyer, A.; Jochum, A.; Sabin, C.; Chevolot, Y.; Imbert, A.; Praly, J. P.; Vasseur, J. J.; Souteyrand, E.; Vidal, S. Fucosylated Pentaerythrityl Phosphodiester Oligomers (PePOs): Automated Synthesis of DNA-Based Glycoclusters and Binding to *Pseudomonas Aeruginosa* Lectin (PA-IIL). *Bioconjug. Chem.* **2007**, *18* (5), 1637–1643.
- (106) Bouillon, C.; Meyer, A.; Vidal, S.; Jochum, A.; Chevolot, Y.; Cloarec, J.; Praly, J.; Vasseur, J.; Morvan, F. Microwave Assisted “ Click ” Chemistry for the Synthesis of Multiple Labeled-Carbohydrate Oligonucleotides on Solid Support † A Versatile Approach Has Been Developed for the Multiple Labeling of Oligonucleotides . First , Three Linkers as a H - Phosphonate. *J. Org. Chem.* **2006**, *71* (3), 4700–4702.
- (107) Pourceau, G.; Meyer, A.; Vasseur, J. J.; Morvan, F. Synthesis of Mannose and Galactose Oligonucleotide Conjugates by Bi-Click Chemistry. *J. Org. Chem.* **2009**, *74* (3), 1218–1222.
- (108) *Oligo synthesis: Why IDT leads the oligo industry.*
<https://eu.idtdna.com/pages/education/decoded/article/oligo-synthesis-why-idt-leads-the-oligo-industry>.
- (109) Zhang, Y.; Kleiner, R. E. A Metabolic Engineering Approach to Incorporate Modified Pyrimidine Nucleosides into Cellular RNA. *J. Am. Chem. Soc.* **2019**, *141* (8), 3347–3351.
- (110) Xu, Y.; Karalkar, N. B.; Kool, E. T. Nonenzymatic Autoligation in Direct Three-Color Detection of RNA and DNA Point Mutations. *Nat. Biotechnol.* **2001**, *19* (2), 148–152.
- (111) Chen, J.; Baker, Y. R.; Brown, A.; El-Sagheer, A. H.; Brown, T. Enzyme-Free Synthesis of Cyclic Single-Stranded DNA Constructs Containing a Single Triazole, Amide or Phosphoramidate Backbone Linkage and Their Use as Templates for Rolling Circle Amplification and Nanoflower Formation. *Chem. Sci.* **2018**, *9* (42), 8110–8120.
- (112) El-Sagheer, A. H.; Sanzone, A. P.; Gao, R.; Tavassoli, A.; Brown, T. Biocompatible Artificial DNA Linker That Is Read through by DNA Polymerases and Is Functional in *Escherichia Coli*. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (28), 11338–11343.
- (113) Shivalingam, A.; Taemaitree, L.; El-Sagheer, A. H.; Brown, T. Squaramides and Ureas: A Flexible Approach to Polymerase-Compatible Nucleic Acid Assembly. *Angew. Chemie - Int. Ed.* **2020**, *59* (28), 11416–11422.
- (114) Landegren, U.; Kaiser, R.; Sanders, J.; Hood, L. A Ligase-Mediated Gene Detection Technique. *Science* (80-.). **1988**, *241* (4869), 1077–1080.
- (115) Lohman, G. J. S.; Chen, L.; Evans, T. C. Kinetic Characterization of Single Strand Break Ligation in Duplex DNA by T4 DNA Ligase. *J. Biol. Chem.* **2011**, *286* (51), 44187–44196.
- (116) Rossi, R.; Montecucco, A.; Ciarrocchi, G.; Biamonti, G. Functional Characterization of the T4 DNA Ligase: A New Insight into the Mechanism of Action. *Nucleic Acids Res.* **1997**, *25* (11), 2106–2113.
- (117) Winz, M. L.; Samanta, A.; Benzing, D.; Jäschke, A. Site-Specific Terminal and Internal Labeling of RNA by Poly(A) Polymerase Tailing and Copper-Catalyzed or

References

- Copper-Free Strain-Promoted Click Chemistry. *Nucleic Acids Res.* **2012**, *40* (10), 1–13.
- (118) Xu, W.; Chan, K. M.; Kool, E. T. Fluorescent Nucleobases as Tools for Studying DNA and RNA. *Nat. Chem.* **2017**, *9* (11), 1043–1055.
- (119) Damase, T. R.; Sukhovshin, R.; Boada, C.; Taraballi, F.; Pettigrew, R. I.; Cooke, J. P. The Limitless Future of RNA Therapeutics. *Front. Bioeng. Biotechnol.* **2021**, *9* (March), 1–24.
- (120) *Trilink mRNA synthesis*. https://www.trilinkbiotech.com/discovery-mrna-synthesis?utm_campaign=Custom-mRNA-Synthesis&utm_medium=cpc&utm_source=google&gad=1&gclid=Cj0KCQjw9fqkBhDSARIsAHlcQYQwpnAfKUOjEeyxikGdNEMKXFO4s-AMN9S9fUellVodPop0-YgJMncaAmHeEALw_wcB.
- (121) *Invitrogen™ TrueGuide™ Synthetic gRNA*. Invitrogen. <https://www.thermofisher.com/order/custom-oligo/crispr>.
- (122) Rosa, S. S.; Prazeres, D. M. F.; Azevedo, A. M.; Marques, M. P. C. mRNA Vaccines Manufacturing: Challenges and Bottlenecks. *Vaccine* **2021**, *39* (16), 2190–2200.
- (123) He, M.; Wu, X.; Mao, S.; Haruehanroengra, P.; Khan, I.; Sheng, J.; Royzen, M. Bio-Orthogonal Chemistry Enables Solid Phase Synthesis and HPLC and Gel-Free Purification of Long RNA Oligonucleotides. *Chem. Commun.* **2021**, *57* (35), 4263–4266.
- (124) Klöcker, N.; Weissenboeck, F. P.; van Dülmen, M.; Špaček, P.; Hüwel, S.; Rentmeister, A. Photocaged 5' Cap Analogues for Optical Control of mRNA Translation in Cells. *Nat. Chem.* **2022**, *14* (8), 905–913.
- (125) Passmore, L. A.; Collier, J. Roles of mRNA Poly(A) Tails in Regulation of Eukaryotic Gene Expression. *Nat. Rev. Mol. Cell Biol.* **2022**, *23* (2), 93–106.
- (126) Milisavljevič, N.; Perlíková, P.; Pohl, R.; Hock, M. Enzymatic Synthesis of Base-Modified RNA by T7 RNA Polymerase. A Systematic Study and Comparison of 5-Substituted Pyrimidine and 7-Substituted 7-Deazapurine Nucleoside Triphosphates as Substrates. *Org. Biomol. Chem.* **2018**, *16* (32), 5800–5807.
- (127) Espinasse, A.; Lembke, H. K.; Cao, A. A.; Carlson, E. E. Modified Nucleoside Triphosphates in Bacterial Research for in Vitro and Live-Cell Applications. *RSC Chem. Biol.* **2020**, *1* (5), 333–351.
- (128) 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), 1–11.
- (129) 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.
- (130) 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

References

- Cells. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (51), E71110–E71117.
- (131) Taemaitree, L. Chemically Modified Guide RNAs to Expand CRISPR Functionality. **2018**, 251.
- (132) Fujino, T.; Suzuki, T.; Okada, K.; Kogashi, K.; Yasumoto, K. I.; Sogawa, K.; Isobe, H. Chimeric RNA Oligonucleotides Incorporating Triazole-Linked Trinucleotides: Synthesis and Function as mRNA in Cell-Free Translation Reactions. *J. Org. Chem.* **2016**, *81* (19), 8967–8976.
- (133) Yamamoto, J.; Ebisuda, S.; Kong, L.; Yamago, H.; Iwai, S. Post-Synthetic Modification of 3' Terminus of RNA with Propargylamine: A Versatile Scaffold for RNA Labeling through Copper-Catalyzed Azide-Alkyne Cycloaddition. *Chem. Lett.* **2017**, *46* (5), 767–770.
- (134) Zhao, M.; Steffen, F. D.; Börner, R.; Schaffer, M. F.; Sigel, R. K. O.; Freisinger, E. Site-Specific Dual-Color Labeling of Long RNAs for Single-Molecule Spectroscopy. *Nucleic Acids Res.* **2018**, *46* (3), 1–8.
- (135) Mamot, A.; Sikorski, P. J.; Siekierska, A.; De Witte, P.; Kowalska, J.; Jemielity, J. Ethylenediamine Derivatives Efficiently React with Oxidized RNA 3' Ends Providing Access to Mono and Dually Labelled RNA Probes for Enzymatic Assays and in Vivo Translation. *Nucleic Acids Res.* **2022**, *50* (1), 1–15.
- (136) Astruc, D.; Boisselier, E.; Ornelas, C. Dendrimers Designed for Functions: From Physical, Photophysical, and Supramolecular Properties to Applications in Sensing, Catalysis, Molecular Electronics, Photonics, and Nanomedicine. *Chem. Rev.* **2010**, *110* (4), 1857–1959.
- (137) Vutukuri, D. R.; Basu, S.; Thayumanavan, S. Dendrimers with Both Polar and Apolar Nanocontainer Characteristics. *J. Am. Chem. Soc.* **2004**, *126* (48), 15636–15637.
- (138) Tomalia, D. A.; Naylor, A. M.; Goddard, W. A. Starburst Dendrimers: Molecular-Level Control of Size, Shape, Surface Chemistry, Topology, and Flexibility from Atoms to Macroscopic Matter. *Angew. Chemie Int. Ed. English* **1990**, *29* (2), 138–175.
- (139) Boas, U.; Heegaard, P. M. H. Dendrimers in Drug Research. *Chem. Soc. Rev.* **2004**, *33* (1), 43–63.
- (140) Hudson, R. H. E.; Damha, M. J. Nucleic Acid Dendrimers: Novel Biopolymer Structures. *J. Am. Chem. Soc.* **1993**, *115* (6), 2119–2124.
- (141) Shchepinov, M. S.; Udalova, I. A.; Bridgman, A. J.; Southern, E. M. Oligonucleotide Dendrimers: Synthesis and Use as Polylabelled DNA Probes. *Nucleic Acids Res.* **1997**, *25* (22), 4447–4454.
- (142) Shchepinov, M. S.; Mir, K. U.; Elder, J. K.; Frank-Kamenetskii, M. D.; Southern, E. M. Oligonucleotide Dendrimers: Stable Nano-Structures. *Nucleic Acids Res.* **1999**, *27* (15), 3035–3041.
- (143) Chandela, A.; Ueno, Y. Design, Synthesis and Evaluation of Novel, Branched Trident Small Interfering RNA Nanostructures for Sequence-Specific RNAi Activity. *RSC Adv.* **2019**, *9* (59), 34166–34171.
- (144) Kim, Y.; Yin, P. Enhancing Biocompatible Stability of DNA Nanostructures Using Dendritic Oligonucleotides and Brick Motifs. *Angew. Chemie - Int. Ed.* **2020**, *59* (2),

References

- 700–703.
- (145) Distler, M. E.; Teplensky, M. H.; Bujold, K. E.; Kusmierz, C. D.; Evangelopoulos, M.; Mirkin, C. A. DNA Dendrons as Agents for Intracellular Delivery. *J. Am. Chem. Soc.* **2021**, *143* (34), 13513–13518.
- (146) Kawamoto, Y.; Liu, W.; Yum, J. H.; Park, S.; Sugiyama, H.; Takahashi, Y.; Takakura, Y. Enhanced Immunostimulatory Activity of Covalent DNA Dendrons. *ChemBioChem* **2022**, *23* (4), 1–7.
- (147) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chemie - Int. Ed.* **2001**, *40* (11), 2004–2021.
- (148) Kolb, H. C.; Sharpless, K. B. The Growing Impact of Click Chemistry on Drug Discovery. *Drug Discov. Today* **2003**, *8* (24), 1128–1137.
- (149) Haldón, E.; Nicasio, M. C.; Pérez, P. J. Copper-Catalysed Azide-Alkyne Cycloadditions (CuAAC): An Update. *Org. Biomol. Chem.* **2015**, *13* (37), 9528–9550.
- (150) Himo, F.; Lovell, T.; Hilgraf, R.; Rostovtsev, V. V.; Noodleman, L.; Sharpless, K. B.; Fokin, V. V. Copper(I)-Catalyzed Synthesis of Azoles. DFT Study Predicts Unprecedented Reactivity and Intermediates. *J. Am. Chem. Soc.* **2005**, *127* (1), 210–216.
- (151) Worrell, B. T.; Malik, J. A.; Fokin, V. V. Direct Evidence of a Dinuclear Copper Intermediate in Cu(I)-Catalyzed Azide-Alkyne Cycloadditions. *Science (80-.)*. **2013**, *340* (6131), 457–460.
- (152) Tornøe, C. W.; Christensen, C.; Meldal, M. Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *J. Org. Chem.* **2002**, *67* (9), 3057–3064.
- (153) Miller, G. P.; Kool, E. T. A Simple Method for Electrophilic Functionalization of DNA. *Org. Lett.* **2002**, *4* (21), 3599–3601.
- (154) Miller, G. P.; Kool, E. T. Versatile 5'-Functionalization of Oligonucleotides on Solid Support: Amines, Azides, Thiols, and Thioethers via Phosphorus Chemistry. *J. Org. Chem.* **2004**, *69* (7), 2404–2410.
- (155) Presolski, S. I.; Hong, V.; Cho, S. H.; Finn, M. G. Tailored Ligand Acceleration of the Cu-Catalyzed Azide-Alkyne Cycloaddition Reaction: Practical and Mechanistic Implications. *J. Am. Chem. Soc.* **2010**, *132* (41), 14570–14576.
- (156) Knauf, M. J.; Bell, D. P.; Hirtzer, P.; Luo, Z. P.; Young, J. D.; Katre, N. V. Relationship of Effective Molecular Size to Systemic Clearance in Rats of Recombinant Interleukin-2 Chemically Modified with Water-Soluble Polymers. *J. Biol. Chem.* **1988**, *263* (29), 15064–15070.
- (157) Juliano, R. L. The Delivery of Therapeutic Oligonucleotides. *Nucleic Acids Res.* **2016**, *44* (14), 6518–6548.
- (158) Hahn, J.; Wickham, S. F. J.; Shih, W. M.; Perrault, S. D. Addressing the Instability of DNA Nanostructures in Tissue Culture. *ACS Nano* **2014**, *8* (9), 8765–8775.
- (159) Seferos, D. S.; Prigodich, A. E.; Giljohann, D. A.; Patel, P. C.; Mirkin, C. A.

References

- Polyvalent DNA Nanoparticle Conjugates Stabilize Nucleic Acids. *Nano Lett.* **2009**, 9 (1), 308–311.
- (160) Kiełpiński, Ł. J.; Hagedorn, P. H.; Lindow, M.; Vinther, J. RNase H Sequence Preferences Influence Antisense Oligonucleotide Efficiency. *Nucleic Acids Res.* **2017**, 45 (22), 12932–12944.
- (161) Debacker, A. J.; Voutila, J.; Catley, M.; Blakey, D.; Habib, N. Delivery of Oligonucleotides to the Liver with GalNAc: From Research to Registered Therapeutic Drug. *Mol. Ther.* **2020**, 28 (8), 1759–1771.
- (162) Jin, C.; El-Sagheer, A. H.; Li, S.; Vallis, K. A.; Tan, W.; Brown, T. Engineering Enzyme-Cleavable Oligonucleotides by Automated Solid-Phase Incorporation of Cathepsin B Sensitive Dipeptide Linkers. *Angew. Chemie - Int. Ed.* **2022**, 61 (13).
- (163) Jansen, R.; Van Embden, J. D. A.; Gaastra, W.; Schouls, L. M. Identification of Genes That Are Associated with DNA Repeats in Prokaryotes. *Mol. Microbiol.* **2002**, 43 (6), 1565–1575.
- (164) Makarova, K. S.; Aravind, L.; Grishin, N. V.; Rogozin, I. B.; Koonin, E. V. A DNA Repair System Specific for Thermophilic Archaea and Bacteria Predicted by Genomic Context Analysis. *Nucleic Acids Res.* **2002**, 30 (2), 482–496.
- (165) Haft, D. H.; Selengut, J.; Mongodin, E. F.; Nelson, K. E. A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes. *PLoS Comput. Biol.* **2005**, 1 (6), 0474–0483.
- (166) 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* (80-.). **2007**, 315 (5819), 1709–1712.
- (167) Horvath, P.; Romero, D. A.; Coûté-Monvoisin, A. C.; Richards, M.; Deveau, H.; Moineau, S.; Boyaval, P.; Fremaux, C.; Barrangou, R. Diversity, Activity, and Evolution of CRISPR Loci in *Streptococcus Thermophilus*. *J. Bacteriol.* **2008**, 190 (4), 1401–1412.
- (168) Deveau, H.; Barrangou, R.; Garneau, J. E.; Labonté, 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), 1390–1400.
- (169) Chen, J. S.; Doudna, J. A. The Chemistry of Cas9 and Its CRISPR Colleagues. *Nat. Rev. Chem.* **2017**, 1 (10).
- (170) Le Rhun, A.; Escalera-Maurer, A.; Bratovič, M.; Charpentier, E. CRISPR-Cas in *Streptococcus Pyogenes*. *RNA Biol.* **2019**, 16 (4), 380–389.
- (171) Makarova, K. S.; Grishin, N. V.; Shabalina, S. A.; Wolf, Y. I.; Koonin, E. V. A Putative RNA-Interference-Based Immune System in Prokaryotes: Computational Analysis of the Predicted Enzymatic Machinery, Functional Analogies with Eukaryotic RNAi, and Hypothetical Mechanisms of Action. *Biol. Direct* **2006**, 1, 1–26.
- (172) Yosef, I.; Goren, M. G.; Qimron, U. Proteins and DNA Elements Essential for the CRISPR Adaptation Process in *Escherichia Coli*. *Nucleic Acids Res.* **2012**, 40 (12), 5569–5576.
- (173) Arslan, Z.; Hermanns, V.; Wurm, R.; Wagner, R.; Pul, Ü. Detection and

References

- Characterization of Spacer Integration Intermediates in Type I-E CRISPR-Cas System. *Nucleic Acids Res.* **2014**, 42 (12), 7884–7893.
- (174) Nuñez, J. K.; Lee, A. S. Y.; Engelman, A.; Doudna, J. A. Integrase-Mediated Spacer Acquisition during CRISPR-Cas Adaptive Immunity. *Nature* **2015**, 519 (7542), 193–198.
- (175) 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), 602–607.
- (176) Gesner, E. M.; Schellenberg, M. J.; Garside, E. L.; George, M. M.; MacMillan, A. M. Recognition and Maturation of Effector RNAs in a CRISPR Interference Pathway. *Nat. Struct. Mol. Biol.* **2011**, 18 (6), 688–692.
- (177) Chylinski, K.; Le Rhun, A.; Charpentier, E. The TracrRNA and Cas9 Families of Type II CRISPR-Cas Immunity Systems. *RNA Biol.* **2013**, 10 (5), 726–737.
- (178) 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* (80-.). **2012**, 337 (6096), 816–821.
- (179) Cho, S. W.; Kim, S.; Kim, J. M.; Kim, J. S. Targeted Genome Engineering in Human Cells with the Cas9 RNA-Guided Endonuclease. *Nat. Biotechnol.* **2013**, 31 (3), 230–232.
- (180) 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.
- (181) Chatterjee, N.; Walker, G. C. Mechanisms of DNA Damage, Repair, and Mutagenesis. *Environ. Mol. Mutagen.* **2017**, 58 (5), 235–263.
- (182) Rouet, R.; Thuma, B. A.; Roy, M. D.; Lintner, N. G.; Rubitski, D. M.; Finley, J. E.; Wisniewska, H. M.; Mendonsa, R.; Hirsh, A.; De Oñate, L.; Compte Barrón, J.; McLellan, T. J.; Bellenger, J.; Feng, X.; Varghese, A.; Chrnyk, B. A.; Borzilleri, K.; Hesp, K. D.; Zhou, K.; Ma, N.; Tu, M.; Dullea, R.; McClure, K. F.; Wilson, R. C.; Liras, S.; Mascitti, V.; Doudna, J. A. Receptor-Mediated Delivery of CRISPR-Cas9 Endonuclease for Cell-Type-Specific Gene Editing. *J. Am. Chem. Soc.* **2018**, 140 (21), 6596–6603.
- (183) Lino, C. A.; Harper, J. C.; Carney, J. P.; Timlin, J. A. Delivering Crispr: A Review of the Challenges and Approaches. *Drug Deliv.* **2018**, 25 (1), 1234–1257.
- (184) Yang, H.; Wang, H.; Shivalila, C. S.; Cheng, A. W.; Shi, L.; Jaenisch, R. XOne-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell* **2013**, 154 (6), 1370.
- (185) Horii, T.; Arai, Y.; Yamazaki, M.; Morita, S.; Kimura, M.; Itoh, M.; Abe, Y.; Hatada, I. Validation of Microinjection Methods for Generating Knockout Mice by CRISPR/Cas-Mediated Genome Engineering. *Sci. Rep.* **2014**, 4 (2), 1–6.
- (186) Chen, S.; Lee, B.; Lee, A. Y. F.; Modzelewski, A. J.; He, L. Highly Efficient Mouse Genome Editing by CRISPR Ribonucleoprotein Electroporation of Zygotes. *J. Biol.*

References

- Chem.* **2016**, 291 (28), 14457–14467.
- (187) Matano, M.; Date, S.; Shimokawa, M.; Takano, A.; Fujii, M.; Ohta, Y.; Watanabe, T.; Kanai, T.; Sato, T. Modeling Colorectal Cancer Using CRISPR-Cas9-Mediated Engineering of Human Intestinal Organoids. *Nat. Med.* **2015**, 21 (3), 256–262.
- (188) Samulski, R. J.; Muzyczka, N. AAV-Mediated Gene Therapy for Research and Therapeutic Purposes. *Annu. Rev. Virol.* **2014**, 1 (1), 427–451.
- (189) Wang, M.; Zuris, J. A.; Meng, F.; Rees, H.; Sun, S.; Deng, P.; Han, Y.; Gao, X.; Pouli, D.; Wu, Q.; Georgakoudi, I.; Liu, D. R.; Xu, Q. Efficient Delivery of Genome-Editing Proteins Using Bio-reducible Lipid Nanoparticles. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, 113 (11), 2868–2873.
- (190) Lv, H.; Zhang, S.; Wang, B.; Cui, S.; Yan, J. Toxicity of Cationic Lipids and Cationic Polymers in Gene Delivery. *J. Control. Release* **2006**, 114 (1), 100–109.
- (191) Wang, T.; Larcher, L. M.; Ma, L.; Veedu, R. N. Systematic Screening of Commonly Used Commercial Transfection Reagents towards Efficient Transfection of Single-Stranded Oligonucleotides. *Molecules* **2018**, 23 (10).
- (192) Roggenbuck, D.; Mytilinaiou, M. G.; Lapin, S. V.; Reinhold, D.; Conrad, K. Asialoglycoprotein Receptor (ASGPR): A Peculiar Target of Liver-Specific Autoimmunity. *Autoimmun. Highlights* **2012**, 3 (3), 119–125.
- (193) Huang, X.; Leroux, J. C.; Castagner, B. Well-Defined Multivalent Ligands for Hepatocytes Targeting via Asialoglycoprotein Receptor. *Bioconjug. Chem.* **2017**, 28 (2), 283–295.
- (194) Zhou, Y.; Teng, P.; Montgomery, N. T.; Li, X.; Tang, W. Development of Triantennary N-Acetylgalactosamine Conjugates as Degradable for Extracellular Proteins. *ACS Cent. Sci.* **2021**, 7 (3), 499–506.
- (195) Prakash, T. P.; Graham, M. J.; Yu, J.; Carty, R.; Low, A.; Chappell, A.; Schmidt, K.; Zhao, C.; Aghajan, M.; Murray, H. F.; Riney, S.; Booten, S. L.; Murray, S. F.; Gaus, H.; Crosby, J.; Lima, W. F.; Guo, S.; Monia, B. P.; Swayze, E. E.; Seth, P. P. Targeted Delivery of Antisense Oligonucleotides to Hepatocytes Using Triantennary N-Acetyl Galactosamine Improves Potency 10-Fold in Mice. *Nucleic Acids Res.* **2014**, 42 (13), 8796–8807.
- (196) Nair, J. K.; Willoughby, J. L. S.; Chan, A.; Charisse, K.; Alam, M. R.; Wang, Q.; Hoekstra, M.; Kandasamy, P.; Kelin, A. V.; Milstein, S.; Taneja, N.; O Shea, J.; Shaikh, S.; Zhang, L.; Van Der Sluis, R. J.; Jung, M. E.; Akinc, A.; Hutabarat, R.; Kuchimanchi, S.; Fitzgerald, K.; Zimmermann, T.; Van Berkel, T. J. C.; Maier, M. A.; Rajeev, K. G.; Manoharan, M. Multivalent N -Acetylgalactosamine-Conjugated siRNA Localizes in Hepatocytes and Elicits Robust RNAi-Mediated Gene Silencing. *J. Am. Chem. Soc.* **2014**, 136 (49), 16958–16961.
- (197) Rittner, K.; Benavente, A.; Bompard-Sorlet, A.; Heitz, F.; Divita, G.; Brasseur, R.; Jacobs, E. New Basic Membrane-Destabilizing Peptides for Plasmid-Based Gene Delivery in Vitro and in Vivo. *Mol. Ther.* **2002**, 5 (2), 104–114.
- (198) Sentmanat, M. F.; Peters, S. T.; Florian, C. P.; Connelly, J. P.; Pruett-Miller, S. M. A Survey of Validation Strategies for CRISPR-Cas9 Editing. *Sci. Rep.* **2018**, 8 (1), 1–8.

References

- (199) Monestier, M.; Charbonnier, P.; Gateau, C.; Cuillel, M.; Robert, F.; Lebrun, C.; Mintz, E.; Renaudet, O.; Delangle, P. ASGPR-Mediated Uptake of Multivalent Glycoconjugates for Drug Delivery in Hepatocytes. *ChemBioChem* **2016**, *17* (7), 590–594.
- (200) Dellinger, D. J.; Timár, Z.; Myerson, J.; Sierzchala, A. B.; Turner, J.; Ferreira, F.; Kupihár, 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), 11540–11556.
- (201) Yoshimura, Y.; Fujimoto, K. Ultrafast Reversible Photo-Cross-Linking Reaction: Toward in Situ DNA Manipulation. *Org. Lett.* **2008**, *10* (15), 3227–3230.
- (202) Filippova, J.; Matveeva, A.; Zhuravlev, E.; Stepanov, G. Guide RNA Modification as a Way to Improve CRISPR/Cas9-Based Genome-Editing Systems. *Biochimie* **2019**, *167*, 49–60.
- (203) Rozners, E. Chemical Modifications of CRISPR RNAs to Improve Gene-Editing Activity and Specificity. *J. Am. Chem. Soc.* **2022**, *144* (28), 12584–12594.
- (204) Mu, W.; Zhang, Y.; Xue, X.; Liu, L.; Wei, X.; Wang, H. 5' Capped and 3' PolyA-Tailed SgRNAs Enhance the Efficiency of CRISPR-Cas9 System. *Protein Cell* **2019**, *10* (3), 223–228.
- (205) Nahar, S.; Sehgal, P.; Azhar, M.; Rai, M.; Singh, A.; Sivasubbu, S.; Chakraborty, D.; Maiti, S. A G-Quadruplex Motif at the 3' End of SgRNAs Improves CRISPR-Cas9 Based Genome Editing Efficiency. *Chem. Commun.* **2018**, *54* (19), 2377–2380.
- (206) Lönn, P.; Kacsinta, A. D.; Cui, X. S.; Hamil, A. S.; Kaulich, M.; Gogoi, K.; Dowdy, S. F. Enhancing Endosomal Escape for Intracellular Delivery of Macromolecular Biologic Therapeutics. *Sci. Rep.* **2016**, *6* (March), 1–9.
- (207) Sanhueza, C. A.; Baksh, M. M.; Thuma, B.; Roy, M. D.; Dutta, S.; Préville, C.; Chrnyk, B. A.; Beaumont, K.; Dullea, R.; Ammirati, M.; Liu, S.; Gebhard, D.; Finley, J. E.; Salatto, C. T.; King-Ahmad, A.; Stock, I.; Atkinson, K.; Reidich, B.; Lin, W.; Kumar, R.; Tu, M.; Menhaji-Klotz, E.; Price, D. A.; Liras, S.; Finn, M. G.; Mascitti, V. Efficient Liver Targeting by Polyvalent Display of a Compact Ligand for the Asialoglycoprotein Receptor. *J. Am. Chem. Soc.* **2017**, *139* (9), 3528–3536.
- (208) Ramanathan, A.; Robb, G. B.; Chan, S. H. mRNA Capping: Biological Functions and Applications. *Nucleic Acids Res.* **2016**, *44* (16), 7511–7526.
- (209) Jackson, R. J.; Hellen, C. U. T.; Pestova, T. V. The Mechanism of Eukaryotic Translation Initiation and Principles of Its Regulation. *Nat. Rev. Mol. Cell Biol.* **2010**, *11* (2), 113–127.
- (210) Liu, D.; Brockman, J. M.; Dass, B.; Hutchins, L. N.; Singh, P.; McCarrey, J. R.; MacDonald, C. C.; Graber, J. H. Systematic Variation in mRNA 3'-Processing Signals during Mouse Spermatogenesis. *Nucleic Acids Res.* **2007**, *35* (1), 234–246.
- (211) Lutz, C. S. Alternative Polyadenylation: A Twist on mRNA 3' End Formation. *ACS Chem. Biol.* **2008**, *3* (10), 609–617.
- (212) Zhang, Y.; Liu, L.; Qiu, Q.; Zhou, Q.; Ding, J.; Lu, Y.; Liu, P. Alternative

References

- Polyadenylation: Methods, Mechanism, Function, and Role in Cancer. *J. Exp. Clin. Cancer Res.* **2021**, *40* (1), 1–19.
- (213) Gallie, D. R. The Cap and Poly(A) Tail Function Synergistically to Regulate mRNA Translational Efficiency. *Genes Dev.* **1991**, *5* (11), 2108–2116.
- (214) Trotta, C. R.; Miao, F.; Arn, E. A.; Stevens, S. W.; Ho, C. K.; Rauhut, R.; Abelson, J. N. The Yeast tRNA Splicing Endonuclease: A Tetrameric Enzyme with Two Active Site Subunits Homologous to the Archaeal tRNA Endonucleases. *Cell* **1997**, *89* (6), 849–858.
- (215) Jo, B.-S.; Choi, S. S. Introns: The Functional Benefits of Introns in Genomes. *Genomics Inform.* **2015**, *13* (4), 112.
- (216) NUDEL, U.; SOREQ, H.; LITTAUER, U. Z.; MARBAIX, G.; HUEZ, G.; LECLERCQ, M.; HUBERT, E.; CHANTRENNE, H. Globin mRNA Species Containing Poly(A) Segments of Different Lengths: Their Functional Stability in *Xenopus* Oocytes. *Eur. J. Biochem.* **1976**, *64* (1), 115–121.
- (217) Wilson, T.; Treisman, R. Removal of Poly(A) and Consequent Degradation of c-Fos mRNA Facilitated by 3' AU-Rich Sequences. *Nature* **1988**, *336* (6197), 396–399.
- (218) Tang, T. T. L.; Stowell, J. A. W.; Hill, C. H.; Passmore, L. A. The Intrinsic Structure of Poly(A) RNA Determines the Specificity of Pan2 and Caf1 Deadenylation. *Nat. Struct. Mol. Biol.* **2019**, *26* (6), 433–442.
- (219) Kühn, U.; Pieler, T. *Xenopus* Poly(A) Binding Protein: Functional Domains in RNA Binding and Protein-Protein Interaction. *J. Mol. Biol.* **1996**, *256* (1), 20–30.
- (220) Rissland, O. S.; Subtelny, A. O.; Wang, M.; Lugowski, A.; Nicholson, B.; Laver, J. D.; Sidhu, S. S.; Smibert, C. A.; Lipshitz, H. D.; Bartel, D. P. The Influence of MicroRNAs and Poly(A) Tail Length on Endogenous mRNA-Protein Complexes. *Genome Biol.* **2017**, *18* (1), 1–18.
- (221) Schäfer, I. B.; Yamashita, M.; Schuller, J. M.; Schüssler, S.; Reichelt, P.; Strauss, M.; Conti, E. Molecular Basis for Poly(A) RNP Architecture and Recognition by the Pan2-Pan3 Deadenylation. *Cell* **2019**, *177* (6), 1619–1631.e21.
- (222) Decker, C. J.; Parker, R. A Turnover Pathway for Both Stable and Unstable mRNAs in Yeast: Evidence for a Requirement for Deadenylation. *Genes Dev.* **1993**, *7* (8), 1632–1643.
- (223) Sonenberg, N.; Hinnebusch, A. G. Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell* **2009**, *136* (4), 731–745.
- (224) Minshall, N.; Reiter, M. H.; Weil, D.; Standart, N. CPEB Interacts with an Ovary-Specific EIF4E and 4E-T in Early *Xenopus* Oocytes. *J. Biol. Chem.* **2007**, *282* (52), 37389–37401.
- (225) Chekulaeva, M.; Hentze, M. W.; Ephrussi, A. Bruno Acts as a Dual Repressor of Oskar Translation, Promoting mRNA Oligomerization and Formation of Silencing Particles. *Cell* **2006**, *124* (3), 521–533.
- (226) Stowell, J. A. W.; Webster, M. W.; Kögel, A.; Wolf, J.; Shelley, K. L.; Passmore, L. A. Reconstitution of Targeted Deadenylation by the Ccr4-Not Complex and the YTH Domain Protein Mmi1. *Cell Rep.* **2016**, *17* (8), 1978–1989.

References

- (227) Webster, M. W.; Chen, Y. H.; Stowell, J. A. W.; Alhusaini, N.; Sweet, T.; Graveley, B. R.; Collier, J.; Passmore, L. A. mRNA Deadenylation Is Coupled to Translation Rates by the Differential Activities of Ccr4-Not Nucleases. *Mol. Cell* **2018**, *70* (6), 1089–1100.e8.
- (228) Lim, J.; Kim, D.; Lee, Y. S.; Ha, M.; Lee, M.; Yeo, J.; Chang, H.; Song, J.; Ahn, K.; Kim, V. N. Mixed Tailing by TENT4A and TENT4B Shields mRNA from Rapid Deadenylation. *Science* (80-.). **2018**, *361* (6403), 701–704.
- (229) Chang, H.; Lim, J.; Ha, M.; Kim, V. N. TAIL-Seq: Genome-Wide Determination of Poly(A) Tail Length and 3' End Modifications. *Mol. Cell* **2014**, *53* (6), 1044–1052.
- (230) Legnini, I.; Alles, J.; Karaikos, N.; Ayoub, S.; Rajewsky, N. FLAM-Seq: Full-Length mRNA Sequencing Reveals Principles of Poly(A) Tail Length Control. *Nat. Methods* **2019**, *16* (9), 879–886.
- (231) Kim, D.; Lee, Y. suk; Jung, S. J.; Yeo, J.; Seo, J. J.; Lee, Y. Y.; Lim, J.; Chang, H.; Song, J.; Yang, J.; Kim, J. S.; Jung, G.; Ahn, K.; Kim, V. N. Viral Hijacking of the TENT4–ZCCHC14 Complex Protects Viral RNAs via Mixed Tailing. *Nat. Struct. Mol. Biol.* **2020**, *27* (6), 581–588.
- (232) Strzelecka, D.; Smietanski, M.; Sikorski, P. J.; Warminski, M.; Kowalska, J.; Jemielity, J. Phosphodiester Modifications in mRNA Poly(A) Tail Prevent Deadenylation without Compromising Protein Expression. *Rna* **2020**, *26* (12), 1815–1837.
- (233) Hollenstein, M. DNA Catalysis: The Chemical Repertoire of DNazymes. *Molecules* **2015**, *20* (11), 20777–20804.
- (234) Fukuda, T.; Matsumoto, E.; Onogi, S.; Miura, Y. Aggregation of Alzheimer Amyloid β Peptide (1–42) on the Multivalent Sulfonated Sugar Interface. *Bioconjug. Chem.* **2010**, *21* (6), 1079–1086.
- (235) Piron, F.; Oprea, C.; Cisma, C.; Terec, A.; Roncali, J.; Grosu, I. Synthesis of Podands with Cyanurate or Isocyanurate Cores and Terminal Triple Bonds. *Synthesis (Stuttg).* **2010**, No. 10, 1639–1644.
- (236) Li, J.; Zhang, J.; Zhang, Q.; Bai, Z.; Zhao, Q.; He, D.; Wang, Z.; Chen, Y.; Liu, B. Syntheses and Anti-Cancer Activity of CO-Releasing Molecules with Targeting Galactose Receptors. *Org. Biomol. Chem.* **2018**, *16* (43), 8115–8129.
- (237) Medina, S. H.; Tiruchinapally, G.; Chevliakov, M. V.; Durmaz, Y. Y.; Stender, R. N.; Ensminger, W. D.; Shewach, D. S.; Elsayed, M. E. H. Targeting Hepatic Cancer Cells with Pegylated Dendrimers Displaying N-Acetylgalactosamine and SP94 Peptide Ligands. *Adv. Healthc. Mater.* **2013**, *2* (10), 1337–1350.
- (238) Chabre, Y. M.; Contino-Pépin, C.; Placide, V.; Tze, C. S.; Roy, R. Expeditive Synthesis of Glycodendrimer Scaffolds Based on Versatile TRIS and Mannoside Derivatives. *J. Org. Chem.* **2008**, *73* (14), 5602–5605.
- (239) Fu, C. W.; Tsai, H. E.; Chen, W. S.; Chang, T. T.; Chen, C. L.; Hsiao, P. W.; Li, W. S. Sialyltransferase Inhibitors Suppress Breast Cancer Metastasis. *J. Med. Chem.* **2021**, *64* (1), 527–542.
- (240) Abronina, P. I.; Zinin, A. I.; Orlova, A. V.; Sedinkin, S. L.; Kononov, L. O. An Easy

References

- Access to Asymmetrically Substituted Oligoethylene Glycols from 18-Crown-6. *Tetrahedron Lett.* **2013**, 54 (34), 4533–4535.
- (241) Cai, H.; Jiang, G.; Shen, Z.; Fan, X. Effects of Dendron Generation and Salt Concentration on Phase Structures of Dendritic-Linear Block Copolymers with a Semirigid Dendron Containing PEG Tails. *Macromolecules* **2012**, 45 (15), 6176–6184.
- (242) Xiao, S.; Chen, X.; Ye, Q.; Chen, K.; Xiao, W.; Guan, X.; Huang, B.; Liu, G.; Wei, H.; Peng, Y. Prop-2-Ynyloxybenzyloxy Substituted Phthalocyanine-Based Polymeric Nanoparticles: Synthesis, Photophysical Properties and in Vitro PDT Efficacy. *J. Coord. Chem.* **2020**, 73 (8), 1232–1244.