

**DEVELOPMENT OF NEW CHEMICAL
BIOLOGICAL TOOLS TO PROBE SPLICE
SITE SELECTION**

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

By

Helen Lewis MSc (Leicester)

Department of Chemistry

University of Leicester

September 2012

ABSTRACT

Development of new chemical biological tools to probe splice site selection

Helen Lewis

RNA splicing is a key process in gene expression and regulation in Eukaryotes and involves the processing of pre-mRNA sequences into mature mRNA. Pre-mRNA consists of exons (protein coding regions) and introns (non-protein coding regions). The introns of the pre-mRNA are excised and the exons ligate to form mature mRNA ready for export from the nucleus. Within the pre-mRNA there are numerous splice sites, some of which are conserved whilst others are alternative splice sites. RNA splicing can follow two different pathways: constitutive or alternative and in humans around 90% of pre mRNA is alternatively spliced which accounts for the formation of multiple isoforms of a single gene.¹ The regulation of splicing involves cis-acting factors which are enhancer/silencer sequences within the pre-mRNA and trans-acting factors comprising of cellular factors including RNA and proteins, combined together they enhance or silence splicing. A major challenge in the field is to determine the interplay between the various factors associated with the promotion or silencing of specific splice sites. Two putative models for the utilization of splice sites have been proposed; firstly, a looping mechanism whereby the enhancers randomly collide with each other by 3D diffusion forming a loop.² Secondly, enhancer mediated splicing occurs by a cooperative protein binding process.³ However, due to the limitations of current biochemical tools, it is not possible to determine the exact mode of action. In this thesis, a new chemical biological approach has been developed which addresses this question, involving the construction of tripartite RNA constructs separated by a non-RNA tether. Using these model systems, compelling evidence is provided which demonstrates splice site selection does not proceed via a looping mechanism which is the widely accepted model in the field.

ACKNOWLEDGEMENTS

Firstly I would like to thank my Chemistry supervisor Dr Glenn Burley and my Biochemistry supervisor Professor Ian Eperon for providing me with this opportunity, and for their continued help and encouragement throughout the course of my PhD. I would also like to thank Gerry Griffith, Graham Eaton, Mick Lee, Shairbanu Ibrahim, Andrew Bottrill for their technical support and help for the various NMR, MS and RP-HPLC machines used over the course of the last few years. I would like to give special thanks to the Organic synthesis group, in particular the Burley group members especially Andrew Perrett and the Eperon group for their help and advice and for making the time in and out of the laboratory enjoyable. I am also very appreciative to the support given to me from my family. Lastly I would like to thank Leverhulme Trust for providing the funding for this PhD.

ABBREVIATIONS

ACN	Acetonitrile
Alt	Alternative
ATP	Adenosine triphosphate
Bpt	Branchpoint
BTT	5'-benzylthio-1-H-tetrazole
CPG	Controlled pore glass
CTP	Cytidine triphosphate
Cy5	Cyanine 5 dye
DIEA	N,N-Diisopropylethylamine
DMF	Dimethylformamide
DNA	Deoxyribonucleic acid
DMSO	Dimethyl sulfoxide
DMTr	Dimethoxytrityl
d/s	Downstream

DTT	Dithiothreitol
EDC	N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride
EDTA	Ethylenediaminetetraacetic acid
ESE	Exonic splicing enhancer
ESI	Electrospray ionisation
ESS	Exonic splicing silencer
ETT	Ethylthiotetrazole
FITC	Fluorescein isothiocyanate
Fmoc	9-Fluorenylmethoxycarbonyl
FTU	Fluorescein thiourea
GNA	Glycol nucleic acid
GTP	Guanosine triphosphate
HEG	Hexaethylene glycol
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
hnRNP	Heterogeneous nuclear ribonucleoproteins

HOBt	Hydroxybenzotriazole
IR	Infra-red
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
MALDI	Matrix-assisted laser desorption/ionisation
mRNA	Messenger RNA
MS	Mass spectrometry
NHS	N-hydroxysuccinimide
NMR	Nuclear magnetic resonance
nt	Nucleotide
NTP	Nucleoside triphosphate
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PK	Proteinase K
Ppt	Polypyrimidine tract

Pre-mRNA	Pre-cursor messenger RNA
RNA	Ribonucleic acid
RP-HPLC	Reverse-phase high performance liquid chromatography
r.t.	Room temperature
SDS	Sodium dodecyl sulphate
SF1	Splicing factor 1
snRNP	Small nuclear ribonucleoprotein
SR	Serine arginine
SRSF1	Serine arginine splicing factor (also known as SF2/ASF)
TAE	Tris-acetate-EDTA
TBDMS	Tert-butyldimethylsilyl
TBE	Tris-borate-EDTA
TBTA	1-(1-benzyltriazol-4-yl)-N,N-bis[(1-benzyltriazol-4-yl)methyl]methanamine
TC	Thiomorpholine-carbothioate
TG	Tris-glycine

THF	Tetrahydrofuran
THPTA	Tris(3-hydroxypropyltriazolylmethyl)amine
TOF	Time of Flight
u/s	Upstream
UTP	Uridine triphosphate

CONTENTS

ABSTRACT	I
ACKNOWLEDGEMENTS	II
ABBREVIATIONS.....	III
CONTENTS	VIII
1 INTRODUCTION	1
1.1 THE MECHANISM OF ALTERNATIVE RNA SPLICING.....	3
1.2 THE SPLICEOSOME	6
1.3 ALTERNATIVE SPLICING	10
1.4 REGULATORY PROCESSES USED BY HIGHER EUKARYOTES IN	
ALTERNATIVE RNA SPLICING.....	13
1.4.1 Splicing Regulators	13
1.4.1.1 <i>Splicing Silencers.....</i>	<i>14</i>
1.4.1.2 <i>Splicing Enhancers</i>	<i>19</i>
1.4.1.2.1 <i>SR Proteins</i>	<i>19</i>
1.4.1.2.2 <i>Intronic Splicing Enhancers</i>	<i>20</i>
1.4.1.2.3 <i>Exonic Splicing Enhancers</i>	<i>21</i>
1.4.1.2.3.1 <i>RNA looping – 3-dimensional diffusion model.....</i>	<i>25</i>
1.4.1.2.3.2 <i>Protein propagation – 1-dimensional model</i>	<i>26</i>
1.5 AIMS.....	27
1.5.1 Method Development.....	27
1.5.2 Hypothesis to be Tested	29

1.6	DEVELOPING METHODOLOGY FOR THE CONSTRUCTION OF	
	TRIPARTITE TRANSCRIPTS.....	30
1.6.1	Enzymatic Synthesis of Modified RNA.....	32
1.6.1.1	<i>Incorporation of functional groups on the 5' end of transcripts</i>	<i>34</i>
1.6.1.2	<i>Incorporation of functional groups onto 3' end</i>	<i>36</i>
1.6.2	Solid Phase Synthesis.....	37
1.6.3	Bioconjugation Chemistry for the Ligation of the RNA Strands.....	41
1.6.3.1	<i>Amide coupling and thiol-maleimide coupling.....</i>	<i>43</i>
1.6.3.2	<i>Staudinger Ligation</i>	<i>44</i>
1.6.3.3	<i>Click Chemistry.....</i>	<i>46</i>
1.6.3.3.1	<i>The copper-catalysed Huisgen [3+2] cycloaddition: click chemistry</i>	
	<i>.....</i>	<i>47</i>
1.6.3.3.2	<i>Copper-free click chemistry.....</i>	<i>50</i>
1.7	STRATEGY	52
2	DEVELOPMENT OF SPLICING ASSAY WITH AN EXONIC SPLICING	
	ENHANCER AT THE 5' END	54
2.1	INTRODUCTION	54
2.1.1	Splicing Assays to be Tested.....	54
2.2	RESULTS AND DISCUSSION	56
2.2.1	β -globin Construct.....	56
2.2.1.1	<i>Construction of the model transcripts</i>	<i>57</i>
2.2.2	SRSF1 Gene	61
2.2.3	Adeno WW Virus.....	64
2.3	SUMMARY.....	70

2.4	EXPERIMENTAL.....	71
2.4.1	Preparation of PCR Fragments for Adenovirus and Beta-Globin Constructs.....	71
2.4.2	PCR Amplification Optimisation of Genomic DNA-SF2/ASF cryptic Intron	72
2.4.3	Radioactive Transcription of Amplified PCR products	73
2.4.4	Splicing Assays ¹⁴⁵	74
2.4.5	Synthesis of Radioactive Ladder ¹⁴⁴	74
2.4.6	Phenol Chloroform Extraction ¹⁴⁴	75
2.4.7	Ethanol Precipitation	75
2.4.8	Buffers	75
3	DEVELOPMENT OF INITIATORS TO INCORPORATE INTO RNA TRANSCRIPTS FOR THE CONSTRUCTION OF TRIPARTITE RNA SYSTEMS.....	76
3.1	INTRODUCTION	76
3.1.1	Aim of Chapter 3.....	78
3.1.2	G-Initiators	79
3.2	RESULTS AND DISCUSSION	80
3.2.1	Synthesis of G-Initiators (G1-G3).....	80
3.2.2	Synthesis of G-Initiator (G1)	84
3.2.3	Click Reaction with G-Initiator (G1) and Coumarin Azide (33)	85
3.2.4	Click Reaction between G-Initiator (G1) and Biotin azide (35)	87
3.2.5	THPTA ligand verses TBTA ligand	88
3.2.5.1	<i>Synthesis of tris(hydroxypropyl)triazolylmethyl-amine (THPTA)</i>	<i>89</i>

3.2.5.2	<i>Click Reaction between G-Initiator (G1) and fluorophore azides (33) & (43) with THPTA ligand</i>	92
3.2.6	Incorporation of G-Initiator (G1) into A4.....	95
3.2.7	Inhibition Effect of Transcription with the Addition of G-Initiator (G1)	100
3.2.8	Fluorescence as a Means of Detection	101
3.2.9	Development of G-Initiator (G4)	101
3.2.10	Synthesis of 2'OMe ESE-azides (ESE1 & ESE11)	104
	3.2.10.1 <i>Solid phase synthesis of ESE-amines</i>	104
	3.2.10.2 <i>Synthesis of ESE-azides via NHS coupling</i>	107
3.2.11	Determining the Presence of the Azide on the ESE using Click Chemistry	107
3.2.12	Incorporation of G-Initiator into 44 Nucleotide Transcript (44-mer) ...	111
3.2.13	Click Reaction of Incorporated 44-mer Transcript and Various Azides	113
3.2.14	Optimisation of the Click Reaction between 44 Nucleotide Transcript and 2'OMe ESE-Azides	115
3.2.15	Investigation into the Action of the Phosphorothioate in the Click Reactions	118
3.2.16	Incorporation of G-Initiator During Cold Transcription and Subsequent Ligations to Fluorophores	122
	3.2.16.1 <i>FTU-Azide (43)</i>	122
	3.2.16.2 <i>Cy5 Azide (49)</i>	123
3.3	SUMMARY	125
3.4	EXPERIMENTAL	127

3.4.1	General Procedures	127
3.4.2	Synthesis of Hexanyl Phosphoramidite (50) ¹⁵⁷	128
3.4.3	Synthesis of Guanosine Initiator (G1).....	129
3.4.4	Synthesis of Guanosine Initiator (G4).....	130
3.4.5	Synthesis of 2'Methoxy-ESE-Amines	131
3.4.6	Synthesis of NHS-Azide (47) ¹⁵⁸	133
3.4.7	NHS-Azide (47) Coupling to ESE-Amines	133
3.4.8	Synthesis of 3-Azido-1-Propanol (38) ¹²⁴	136
3.4.9	Synthesis of 3-Azido-propyl Acetate (39) ¹²⁴	136
3.4.10	Synthesis of Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA (42)) ¹²⁴	137
3.4.11	Synthesis of Azide modified Fluorescein Thiourea (FTU-N ₃ (43)).....	138
3.4.12	Radioactive Transcription for Incorporation of G Initiator (G1) into Ad1WW (A4).....	139
3.4.13	Biotin Pull Down Assay.....	140
3.4.14	Radioactive Transcription Optimisation for the Incorporation of G- Initiators (G1 & G4) into 44 Nucleotide RNA Transcript	140
3.4.15	Cold Transcription of into 44 Nucleotide RNA Incorporating G-Initiator (G1)	140
3.4.16	Click Reactions between RNA-Alkynes and Cargo-Azides using Cu(I)TBTA Catalyst, (Table 3.10).....	141
3.4.17	Click Reactions between RNA-Alkynes and Cargo-Azides using Cu(I)THPTA Catalyst, (Table 3.11)	141
3.4.18	Click Reactions between (G1) and Fluorophore-Azides (33)/(43) using Finn Method ¹²⁴	141

3.4.19	Click Reactions Testing the Copper Catalysts (Table 3.12)	142
3.4.20	Buffers	142
4	SPLICING OF THE TRIPARTITE TRANSCRIPTS	149
4.1	INTRODUCTION	149
4.1.1	Aim of Chapter 4	149
4.2	RESULTS AND DISCUSSION	151
4.2.1	Incorporation of G-Initiator (G4) into Ad1WW Transcript (A4)	151
4.2.2	Splicing Reactions of the Tripartite Constructs with ESE1-4	154
4.2.2.1	<i>Preparation of the tripartite constructs with ESE1-4</i>	<i>154</i>
4.2.2.2	<i>Splicing of tripartite constructs using ESE1-4</i>	<i>157</i>
4.2.3	Synthesis of Naturally Occurring 2'Hydroxy ESEs	161
4.2.3.1	<i>Design and synthesis of 2'-hydroxy ESE-azides by solid phase</i>	<i>164</i>
4.2.3.2	<i>Confirming the presence of an azide group on the ESEs by click chemistry</i>	<i>167</i>
4.2.3.3	<i>Solution phase coupling of NHS-azide (47)</i>	<i>168</i>
4.2.4	Splicing Reactions of the Tripartite Constructs using ESE5-10	171
4.2.4.1	<i>Preparation of the tripartite constructs using ESE5-10</i>	<i>171</i>
4.2.4.2	<i>Splicing of the tripartite constructs ESE5-8</i>	<i>171</i>
4.2.5	The Influence of the 7-Methylguanosine Cap on the Alternative Splicing of A3	174
4.2.6	Comparative Analysis of Spliceosome Complex Formation for A3, A4 and ESE8-A4	177
4.3	SUMMARY	180
4.4	EXPERIMENTAL	182

4.4.1	Radioactive Transcription of G-Initiator (G4) Incorporated into Ad1WW (A4)	182
4.4.2	Click Reactions between G4-A4 and ESEs	182
4.4.3	Splicing Reaction for Transcripts Bioconjugated to Enhancer Sequences and Controls ¹⁴⁵	182
4.4.4	Splicing Reaction for Native gel Analysis ¹⁶⁷	183
4.4.5	Splicing of Transcripts for Native gel Analysis with the Addition of U6 Blocked Oligonucleotide ¹⁶⁷	184
4.4.6	Radioactive Transcription of Amplified PCR products using [α - ³² P] UTP	184
4.4.7	Synthesis of ESE-Amines	184
4.4.8	Solid phase NHS-Azide to ESE-Amine on CPG Resin	185
4.4.9	NHS-Azide Coupling to ESE-Amines	187
4.4.10	Buffers	188
5	CONCLUSION AND FUTURE WORK.....	189
5.1	FURTHER WORK.....	192
5.1.1	Synthesis of the 3' Alkyne Modified Di-Cytosine (61)	193
5.1.2	Ligation of (61) to 44 Nucleotide Transcript	195
5.1.3	Experimental	199
5.1.3.1	<i>Synthesis of dinucleotide-alkyne (61)</i>	<i>199</i>
5.1.3.2	<i>RNA ligation between dinucleotide-alkyne (61) and transcript</i>	<i>199</i>
5.1.3.3	<i>Buffers.....</i>	<i>199</i>
6	APPENDIX.....	200
7	REFERENCES	223

1 INTRODUCTION

Prior to 1978 it was considered that one gene produced one protein. However, in 1978, one year after splicing was discovered, Walter Gilbert hypothesized that one gene can generate distinct protein products by variations in the splicing pattern. This biological process is now known as alternative splicing (Figure 1.1).^{4,5} Over the years, analyses of genomes of higher Eukaryotes have taken place through various biological and computational techniques. In 1990 the Human Genome Project; an international collaboration was officially launched; its aim was to discover all the estimated human genes and sequence the entire genome. By its completion in 2003, it was reported that the number of genes was significantly lower than previously expected; which ranged from 40,000-150,000 genes to between 20,000-25,000 genes.^{6,7} Today it is widely accepted that most protein coding genes in mammals produce many proteins due to alternative splicing. It is thought that there are at least 500,000 human proteins, which are all encoded by the mere 20,000-25,000 protein coding genes. The diversification of proteins found in higher eukaryotes is attributed to regulation at the splicing level.^{4,8}

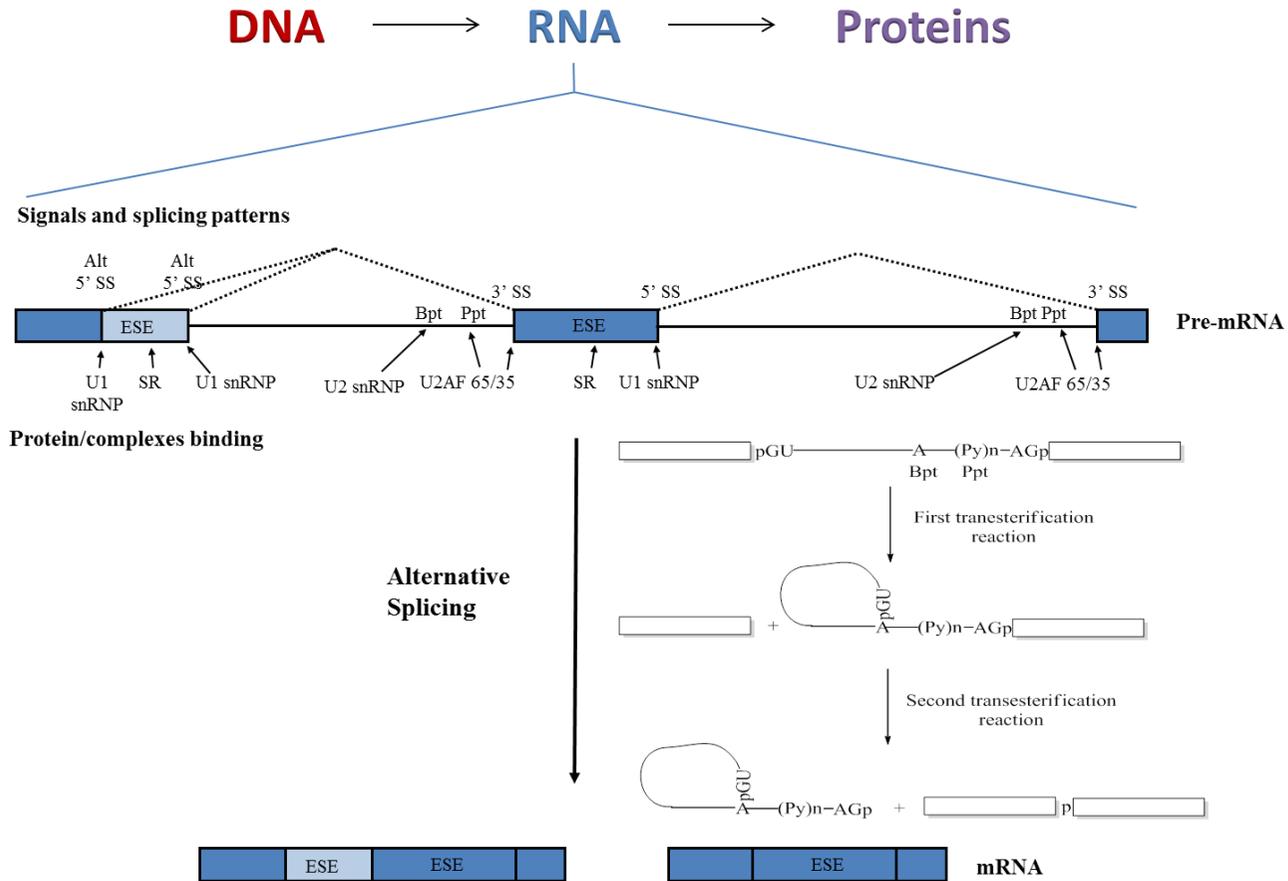


Figure 1.1 Overview of pre-mRNA processing, showing essential signals and protein components; exons are represented by boxes and introns by black solid line, splicing patterns by dashed line. Signals: 5' SS, 5' splice site; Alt, Alternative; 3' SS, 3' splice site; Bpt, branchpoint; Ppt/(Py)_n, polypyrimidine tract. Proteins: snRNP, small nuclear ribonucleoproteins (U1 & U2); auxiliary factors U2AF 65/35; SR proteins.

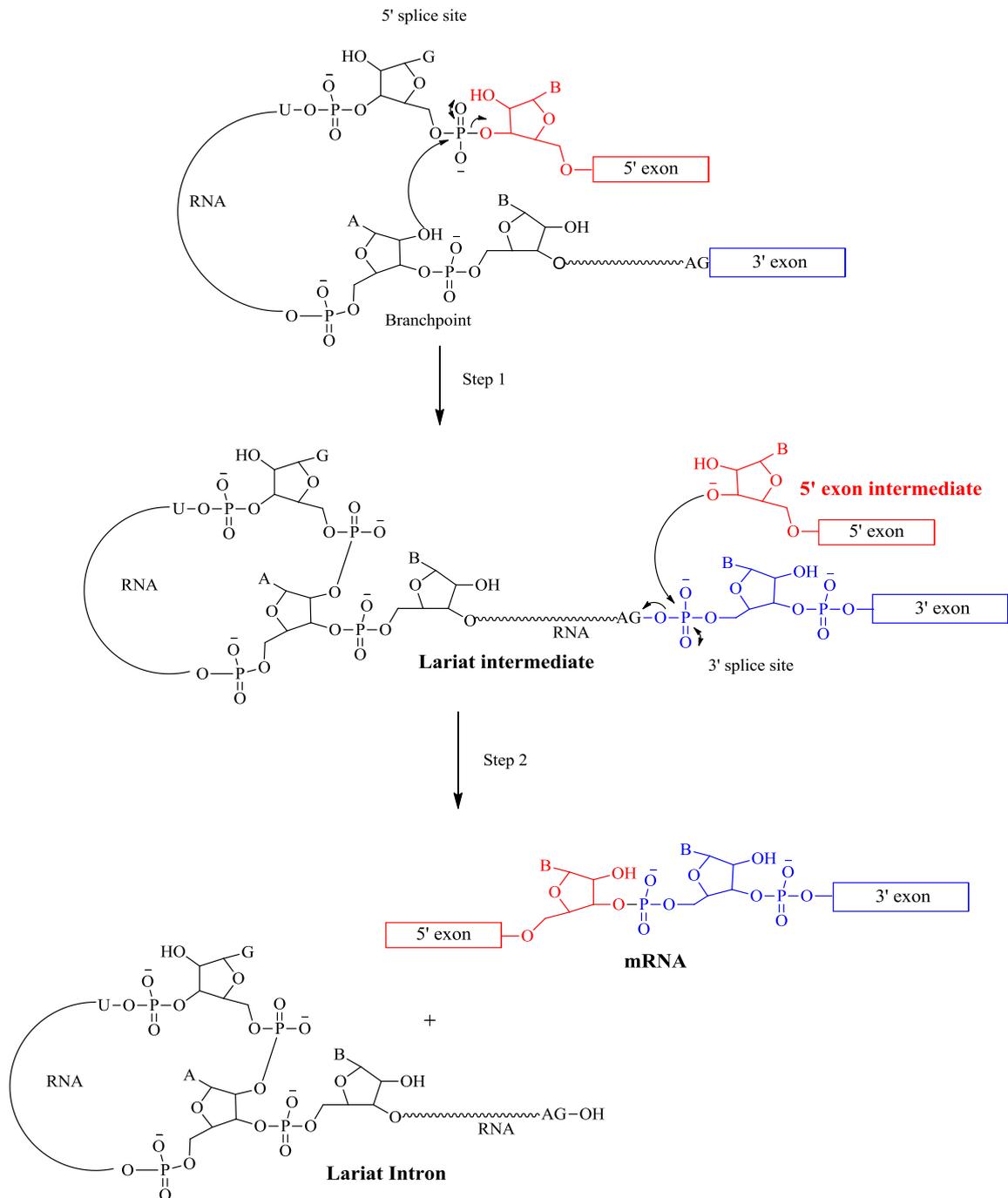
1.1 THE MECHANISM OF ALTERNATIVE RNA SPLICING

During transcription both protein coding (exons) and non-protein coding (introns) regions of the gene embedded in DNA are converted into RNA in the nucleus. This transcribed RNA is known as precursor messenger RNA (Pre-mRNA) (Figure 1.1). Before translation can occur in the cytoplasm, the pre-mRNA undergoes a process of intronic excision and the stitching together of exon sequences to form mature mRNA sequences in the nucleus. The process caters for the formation of several mature mRNA isoforms and is known as RNA splicing (Figure 1.1); it is catalysed by the macromolecular machine known as the spliceosome. Most introns are spliced during transcription, while the downstream RNA is still being synthesised.

Within pre-mRNA there are a number of conserved sequences that promote and repress splicing. To initiate splicing a number of intron-defining splicing factors are essential including a guanosine-uridine (GU) dinucleotide at the 5' splice site, an adenosine-guanosine (AG) dinucleotide at the 3' splice site and an adenosine (A) situated at the branch point (Figure 1.1). The branch point is located between 17-40 nucleotides upstream of the 3' splice site and in higher eukaryotes is followed by a polypyrimidine tract (ppt, Figure 1.1).^{9,10} The excision of introns from pre-mRNA transcripts occurs by two consecutive transesterification reactions where by the dinucleotides at the splice sites and the A at the branch point are integral components in these transesterification reactions (Scheme 1.1). The first transesterification occurs at the 5' splice site with the nucleophilic attack of the 5' phosphodiester bond of the guanosine (G) by the 2' hydroxyl group of the branch point A, generating a 5' exon intermediate and a lariat intermediate (Scheme 1.1, step 1). The second transesterification reaction is between the 3' hydroxyl group of the 5' exon intermediate which attacks the 3' phosphodiester bond

of the G at the 3' splice site affording the mature mRNA product and the lariat intron (Scheme 1.1, step 2).¹¹⁻¹⁵

Scheme 1.1 Overview of the transesterification reactions that occur in alternative RNA splicing. Nucleophilic attack of the 5' phosphodiester bond of G at the 5' splice site by the 2' hydroxyl group of the branchpoint A forms two intermediates; a 5' exon intermediate with a free 3'OH (step 1, shown in red) and a lariat intermediate (step 1 shown in black & blue). The 3' hydroxyl group of the 5' exon attacks the 3' phosphodiester bond of G at the 3' splice site to afford the mature mRNA product and the lariat intron.



1.2 THE SPLICEOSOME

The two transesterification reactions involved in the excision of the intron and reattachment of the exons to form mature mRNA are catalysed by a macromolecular complex known as the spliceosome. The spliceosome consists of 5 small nuclear ribonuclear proteins (snRNPs), including U1, U2, U4/U6 and U5 along with numerous other non-snRNP factors.^{14,16,17} Recent advances in proteomics, purification methods and mass spectrometry have enabled the identification of new proteins which had no prior connection to splicing *in vitro*, and over 300 spliceosomal proteins have been identified to date.^{18,19} Hence, the process of splicing is incredibly complex, with many protein binding events occurring at discrete time points during spliceosome complex formation.

Throughout the splicing process the spliceosome accumulates at several stages, known as the H, E, A, B, B* and C complexes (Scheme 1.2).^{3,10,14} Each of these complexes plays an important role in the formation of the active B* complex by binding snRNPs and SR proteins at different stages. The initial complex formed is the H-complex which is an ATP independent complex comprising heterogeneous nuclear ribonucleoproteins (hnRNPs) which act as activators and suppressors of splicing. The H-complex forms on RNA whether or not it contains functional splice sites (Scheme 1.2, i).^{20,21} Although the H complex is not a necessary precursor to splicing, it is important as it may compete and regulate excision of areas of the pre-mRNA especially in relation to alternative splicing.^{3,22}

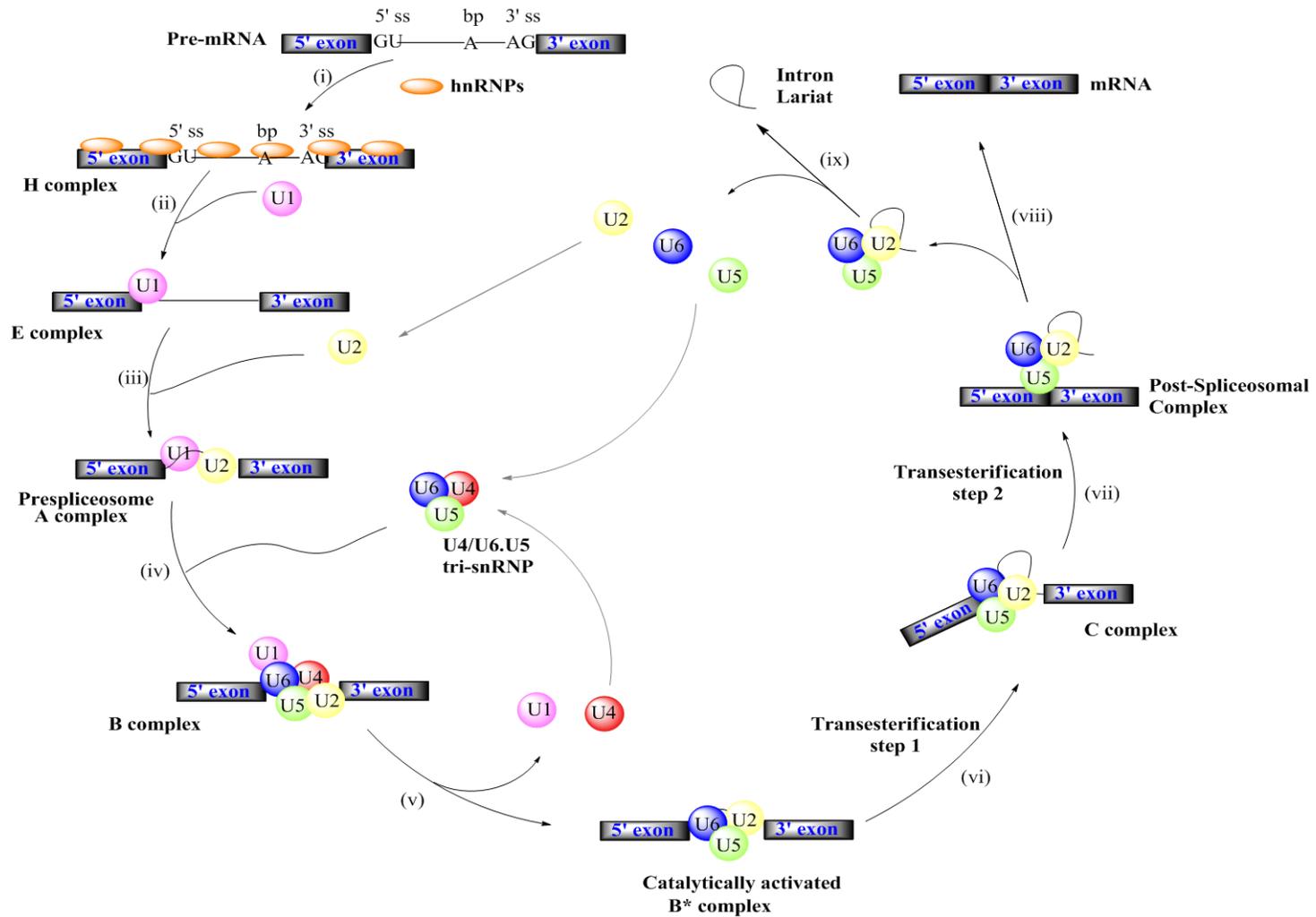
The assembly of the E complex follows (Scheme 1.2, ii), with the ATP-independent binding of U1 snRNP to the 5' splice site of the intron which binds through base pair

interactions and is stabilised by serine-arginine rich (SR) proteins.¹⁴ The splicing factor SF1 binds to the branchpoint and is stabilised by the U2 auxiliary factor (U2AF) heterodimer which is recruited to the polypyrimidine tract and the terminal intronic AG at the 3' splice site.²³⁻²⁵ Additionally many of the non-spliceosomal proteins found in the H complex appear in the E complex and bind the RNA, not to specific snRNPs.²¹

The E complex is rapidly succeeded by the formation of the prespliceosomal A complex (Scheme 1.2, iii) in which the ATP-dependent association of U2 snRNP occurs at the 3' splice site by base pairing to the branchpoint sequences.²⁶⁻²⁸ The binding of the U2 snRNP is thought to bulge the branchpoint A, positioning this nucleotide in a suitable geometry to act as the nucleophile in the first transesterification step.²⁶ The remaining ATP-dependent U4/U6 and U5 tri-snRNPs enter the complex and base pair with each other; U6 base pairs with U2 and U5 forms bonds with the 5' exon. The addition of the tri-snRNP forms the B complex (Scheme 1.2, iv).^{22,28} Consequently there is a reduction in binding affinity of U1 snRNP to the 5' splice site at this point.²⁶ However, the B complex is not catalytically active even though it has all the components required for splicing; a rearrangement of the snRNPs in the complex is required to form the catalytic B* complex (Scheme 1.2, v). U1 snRNP dissociates from the 5' splice site and leaves the spliceosome. U4 acts as a chaperone to U6 which is believed to be intrinsically reactive, and in order to form an active complex U4 unwinds from U6 and leaves the spliceosome. The U6 is then free to base pair to the 5' splice site along with more extensive base pairing with snRNA within the U2 forming the active site. The loss of U1 and U4 from the spliceosome and the binding of the U6 to the 5' splice site forms the activated B* complex which initiates the first transesterification where the 2'-hydroxyl group of the A at the branch point attacks the phosphodiester bond at the 5' splice site (Scheme 1.2, vi). The C complex is formed at this stage in the splicing cycle and is

ready to catalyse the second transesterification reaction (Scheme 1.2, vii). The U5 snRNP which is bound to the 5' exon holds the 5' exon in the complex enabling the second transesterification reaction to take place.^{22,28,29} Once the mRNA and lariat intron (Scheme 1.2, viii) are formed the remaining U2, U5 and U6 are released, completing the catalytic cycle (Scheme 1.2, ix).¹⁴

Scheme 1.2 Schematic representation of the spliceosome catalytic cycle; U1 snRNP (pink); U2 snRNP (yellow); U4 snRNP (red); U5 snRNP (green); U6 snRNP (blue); ss, splice site;, bp, branch point.^{10,14}



1.3 ALTERNATIVE SPLICING

Alternative splicing is the process where genes are spliced in different arrangements producing mRNA and proteins which are structurally and functionally diverse.⁴ In higher Eukaryotes approximately 70% of genes encode for two or more mRNA isoforms contributing to proteomic diversity signifying the importance of alternative splicing.³⁰ There are numerous ways in which alternative splicing can occur as shown in Figure 1.2.^{3,5,31} The most abundant type in mammals is through exon skipping (Figure 1.2 A) and accounts for nearly 40% of alternative splicing events.^{1,32,33} Intron retention on the other hand is more common in lower metazoans and protozoa (Figure 1.2 B).³⁴ An archetypal example of exon skipping is the survival of motor neuron (SMN) genes whereby exon 7 is frequently skipped in SMN2 but is not skipped for SMN1. There are 11 nucleotide differences that exist between the genomic sequences of SMN1 and SMN2. In particular a single base mutation from C to T at position 6 in exon 7 and A to G and position 100 in intron 10 are crucial in alternative splicing of SMN2.³⁵⁻³⁸ Additionally, the strength of the splice site determines whether exon skipping or inclusion can take place. A strong splice site is defined as one which has high binding affinity to U1 snRNP arising to constitutive splicing (exon inclusion) whereas a weak splice site has suboptimal binding with the U1 snRNP giving rise to alternative splicing (exon skipping).³⁴

Alternative (cryptic) 5' and 3' splice sites (Figure 1.2 C & D) are the next most abundant form of alternative splicing. Approximately 19% of alternative splicing occurs via the use of cryptic 3' splice sites and almost 9% due to cryptic 5' splice sites.^{33,34} Eperon *et al.* (1986) proposed that the intrinsic strength of the competing 5' splice sites along with the distance between them could determine splice site selection.³⁹ More

recently it was proposed that alternative 5' splice site selection only occurs if the alternative splice sites belong to the same subclass i.e. both splice sites need to be strong, and for genes with one weak and one strong splice site only the strong splice site would be used.⁴⁰ Alternative splice sites that observe weak splicing signals are stimulated by additional exonic and intronic splicing regulatory sequences that are situated in the exon and the flanking introns.^{34,41}

Other alternative splicing patterns are observed but much less frequently in higher eukaryotes than those discussed in the preceding paragraphs. These include intron retention which generally occurs in genes that consist of short introns (< 275 nucleotides) and is a result of intron definition. Intron definition occurs when the splicing machinery fails to recognise the weak splice sites flanking the short introns and therefore the introns are retained during splicing (Figure 1.2 B).⁴² Cryptic introns are also found in exons where weak cryptic 5' and 3' splice sites are located in an exon and are alternatively spliced creating a shorter mRNA product (Figure 1.2 E).^{43,44} Additional alternative splicing patterns include mutually exclusive splicing whereby only 1 exon is included in a group of exons (Figure 1.2 G), alternative poly-adenylation sites are located within the gene arising from the termination of transcription at different points (Figure 1.2 G) and alternative promoter regions during transcription giving rise to varying pre-mRNA lengths (Figure 1.2 H).^{3,5}

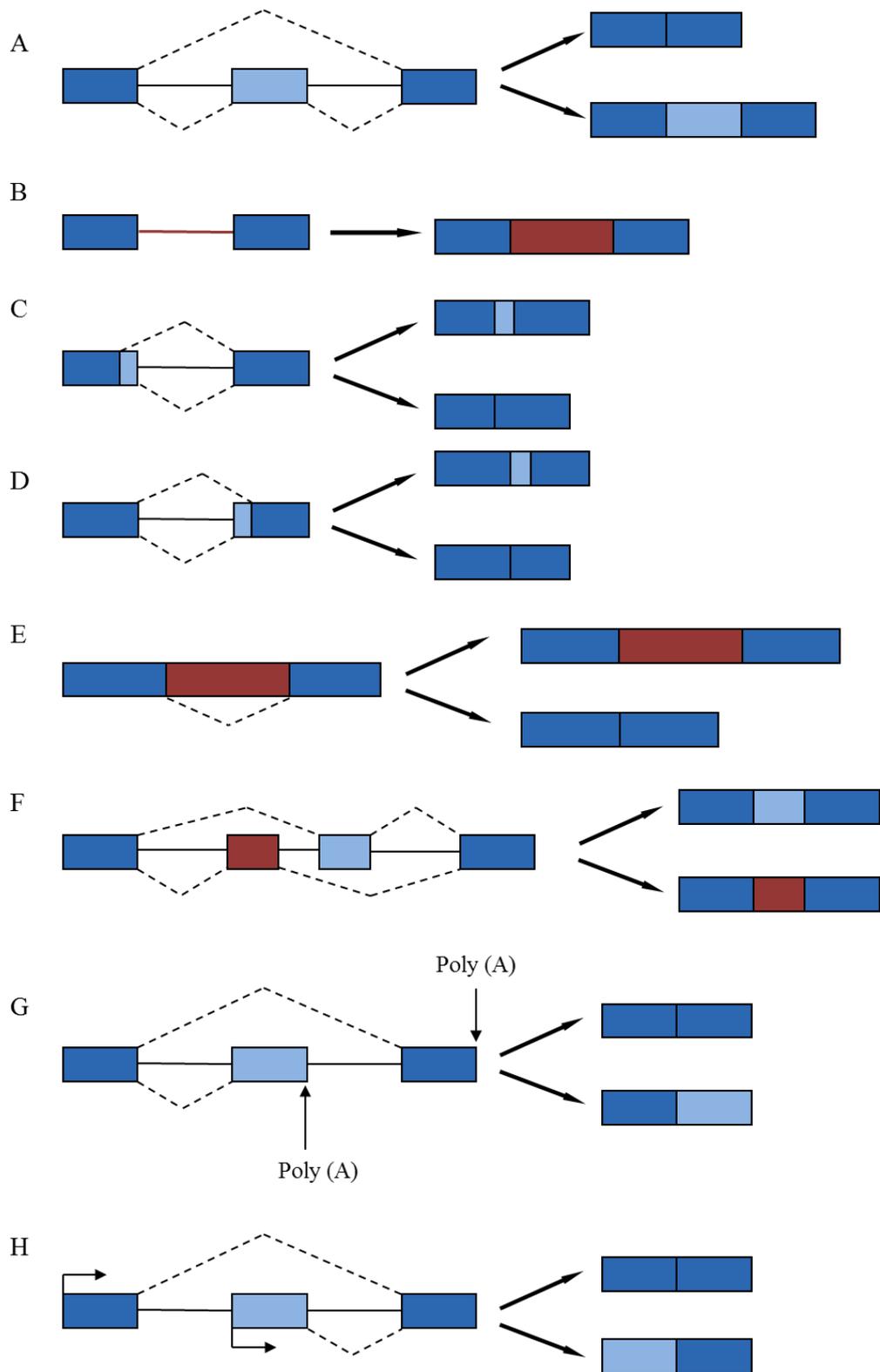


Figure 1.2 Schematic representation of alternative splicing patterns occurring in higher eukaryotes; exons are defined by boxes, introns by solid lines and splicing patterns by dashed lines. (A) Cassette exons (light blue) can be included or excluded from the mRNA. (B) Inclusion of an Intron (red). (C & D) Alternative 5' and 3' splice sites creating mRNA of differing lengths. (E) Cryptic intron (red) located in exon can included or excluded. (F) Mutually exclusive splicing involves the inclusion of only 1 exon in a group of exons. (G) Alternative poly (A) sites terminating transcription at different points. (H) Alternative promoters leading to different start points of the 5' exon.

1.4 REGULATORY PROCESSES USED BY HIGHER EUKARYOTES IN ALTERNATIVE RNA SPLICING

The complexity of alternative splicing in human tissues has been surveyed in recent years by high-throughput screening, and it was estimated that approximately 95% of multi-exon genes undergo alternative splicing.⁴⁵ On average seven alternative splicing events per multi-exon human gene were observed, although the exact number of mRNA and protein isoforms generated have not been calculated.^{45,46} As a consequence of the complexity of alternative splicing, regulation of splice site selection is essential within the RNA splicing machinery.

1.4.1 Splicing Regulators

The regulation of splicing involves both cis- and trans-regulatory elements. Cis-regulatory elements include the conserved sequences at the 5' splice site, 3' splice site and the branch site as previously discussed (Section 1.1). Additional cis-factors can be divided into four sub-categories: exonic splicing enhancers (ESEs), intronic splicing enhancers (ISEs), exonic splicing silencers (ESSs) and intronic splicing silencers (ISSs). Both constitutive and alternative splicing are moderated by these short and diverse sequences that either promote or suppress the assembly of the spliceosome complex, usually during the pre-spliceosomal complexes of E and A.^{3,10,47}

Trans-acting factors include the serine-arginine family (SR) of proteins which generally bind to ESEs facilitating the binding of the U1 and U2 snRNP to the pre-mRNA, and heterogeneous nuclear ribonucleoproteins (hnRNPs), some of which bind to ISSs and ESSs suppressing the formation of the early stages of the spliceosome.

1.4.1.1 Splicing Silencers

Splice site recognition is achieved in several ways and involve ESSs, ISSs and a myriad of proteins predominantly associated with the hnRNP family. The hnRNP proteins bind to the conserved silencer sequences in the pre-mRNA and suppress splicing. The mechanism by which they exert the inhibition effects is not clearly understood, but there have been two possible methods proposed:^{3,23,31,48}

1. Sterically blocking the snRNPs from binding the splice sites and branch point, or by blocking the SR proteins binding to ESEs through cooperative binding (Figure 1.3a).

2. Silencer proteins bind ISSs in the two introns flanking an exon and forming associations with each other looping out the exon causing exon skipping (Figure 1.3b).

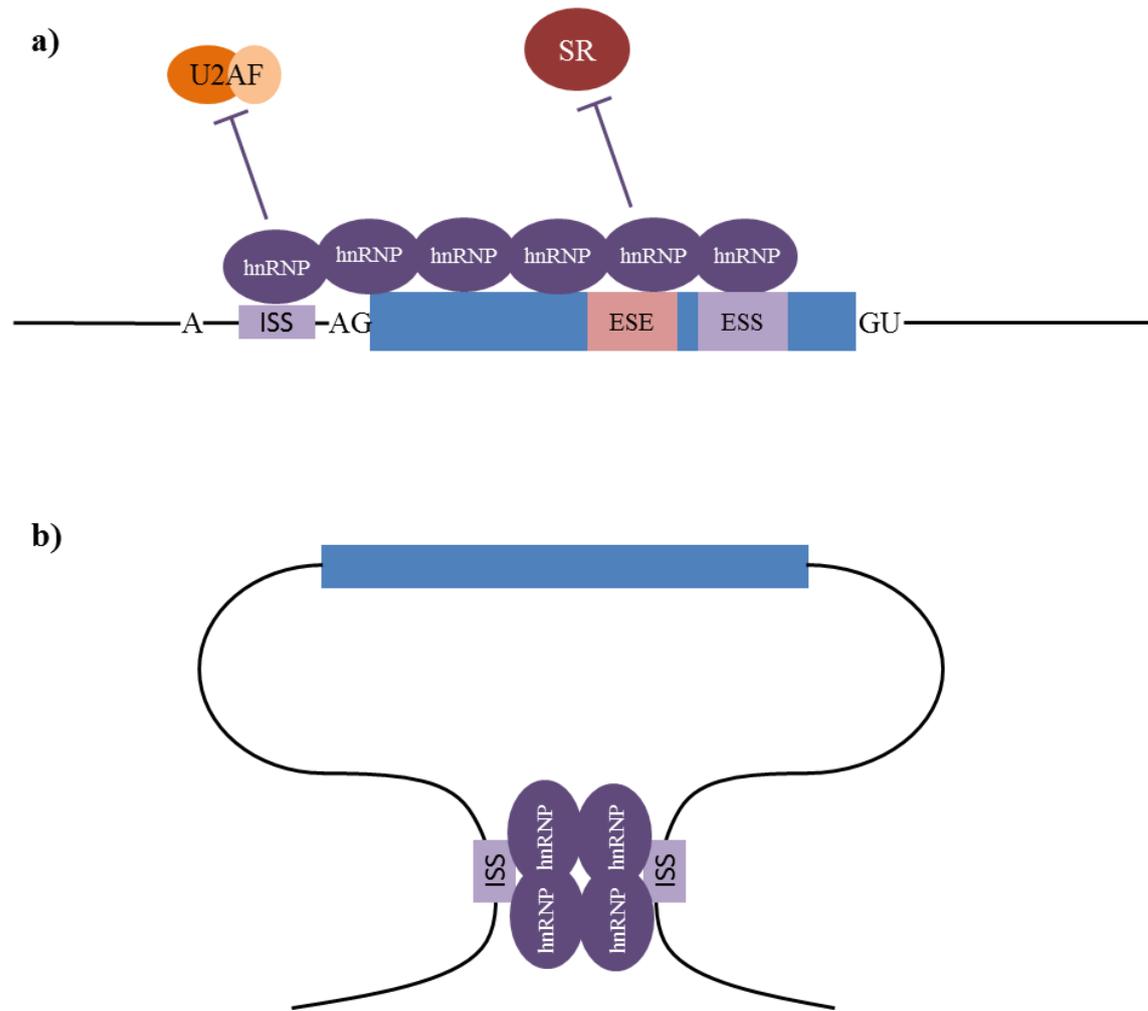


Figure 1.3 Diagrammatic representation of the two proposed mechanisms of silencers. Exon – blue rectangle, Introns black lines, Purple circles hnRNPs (A) The binding of hnRNPs smother the pre-mRNA preventing activator proteins (SR/U2AF) from binding silencing splicing (B) The hnRNPs are located across the transcript and bind together forming a loop preventing activation of splicing.

The process of splicing inhibition by sterically blocking snRNPs and ESEs occurs in the HIV-1 gene where 30 different HIV-1 mRNA isoforms are achieved by alternative splicing. The cis-acting factors, ISSs and ESSs, are integral in regulating alternative splicing of the HIV-1 gene. The removal of the second tat intron was shown to be inhibited by the binding of hnRNP A1 protein which interacted specifically with an ISS sequence located between -22 to -44 nucleotides upstream of the 3' splice site, overlapping one of the three alternative branchpoints (Figure 1.4). The binding of hnRNP A1 to this site suggests that the U2 snRNP is physically blocked from binding

to the pre-mRNA and therefore suppressing splicing (Figure 1.4). Furthermore there are two additional ESSs located in the exon downstream of the second tat intron (third tat exon) both of which bind hnRNP A1 and overlapping the ESE located in the exon. This finding suggests possible cooperative binding of the hnRNP A1 proteins bound to the ESSs which block the ESE from binding to the SR proteins required to initiate splicing.⁴⁹

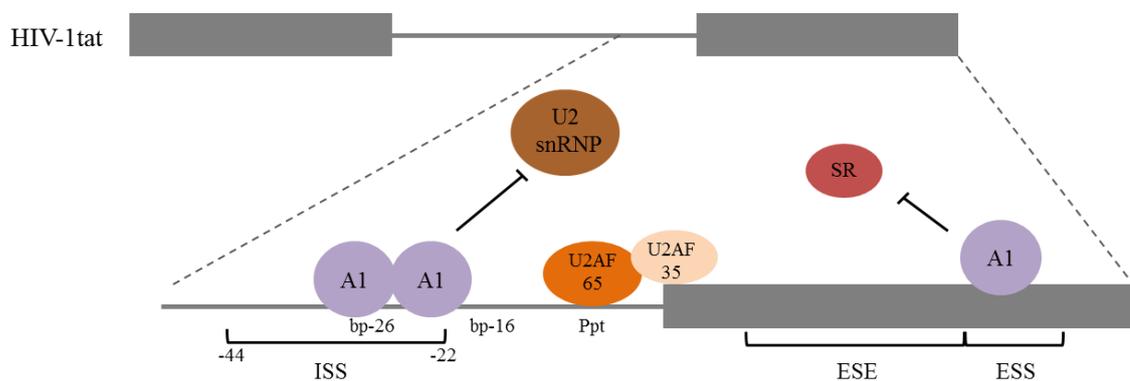


Figure 1.4 Diagrammatic representation of the regulation of HIV-1 tat gene by the binding of hnRNP A1 proteins to an ISS preventing U2 snRNP binding to the branch point (bp) and to an ESS inhibiting SR protein binding. Ppt: polypyrimidine tract.⁴⁹

More recently an *in vivo* study into the splicing of the SMN2 gene showed that the nuclear RNA binding protein Sam68 when bound to an ESS in exon 7 acted as a crucial regulator of SMN2 alternative splicing. The study found that Sam68 promoted skipping of exon 7 by binding to the ESS which was identified as UAGAGA, and associating with hnRNP A1 through cooperative binding suppressing splicing at the 3' splice site flanking exon 7, promoting exon 7 skipping. Exon 7 inclusion was rescued however when the binding of Sam68 to the ESS was inhibited.⁵⁰ This study demonstrated the importance of the association of RNA-binding proteins to ESSs in suppressing splicing.

Despite this evidence suggesting that ESS and ISS bind proteins that block the binding of snRNP and SR proteins which are essential in initiating splicing, this does not

explain how ISSs exert their function when they are over 100 to 200 nucleotides away from their respective splice sites. The RNA binding protein hnRNP A1 is a splicing regulator which binds to ESSs and ISSs silencing splicing of a variety of genes.^{49,51,52} hnRNP A1 is produced by the skipping of alternative exon 7B from the A1 pre-mRNA alternative splicing is observed with the skipping of exon 7B. Blanchette & Chabot discovered that the introns flanking exon 7B contained conserved elements which were high affinity A1 binding sites the promoted exon 7B exclusion (Figure 1.5). It was proposed that once the A1 proteins were bound to the two ISSs they came together looping out exon 7B and in turn bringing together the 5' splice site flanking exon 7 and the 3' splice site flanking exon 8 mediating skipping of exon 7B (Figure 1.5).^{51,53}

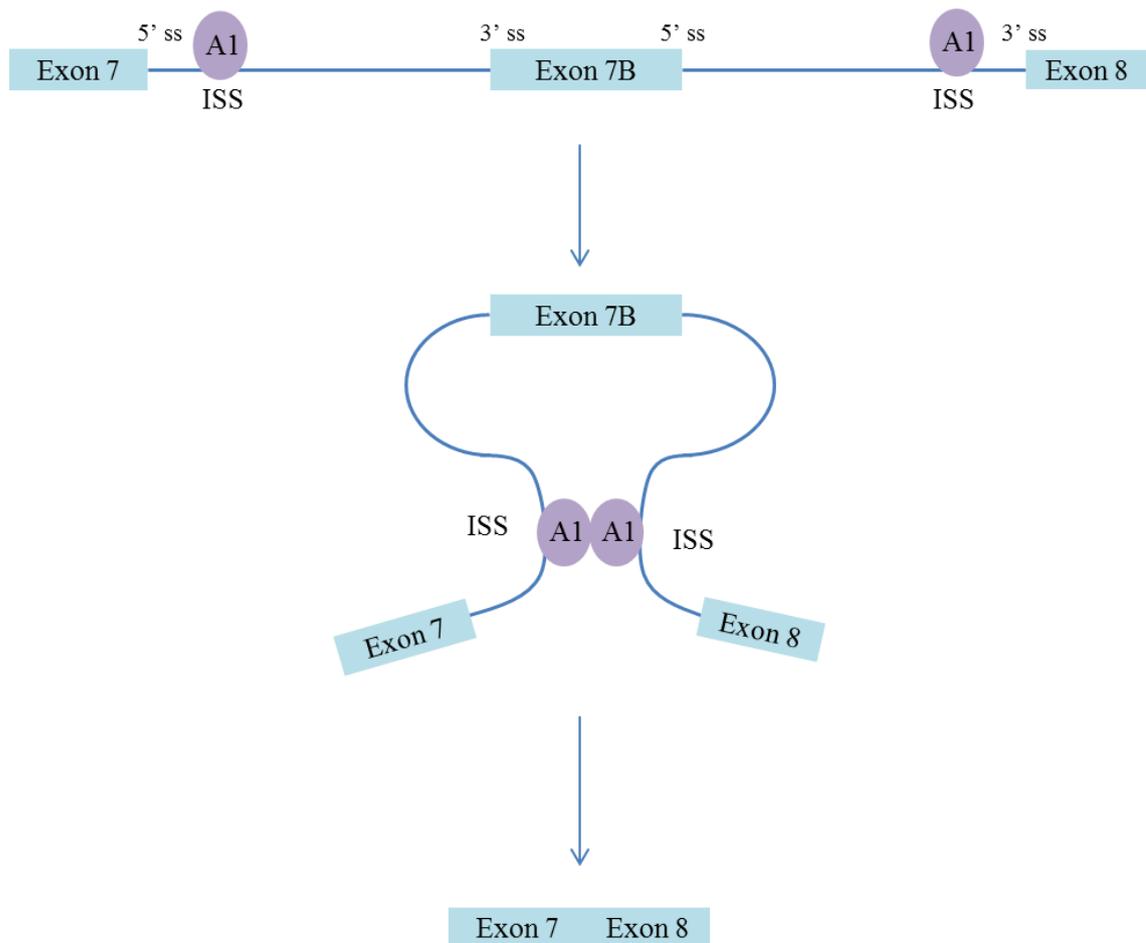


Figure 1.05 A model for the role of alternative splicing in the A1 protein gene. hnRNP A1 proteins bind to ISS in the flanking introns of exon 7B looping it out and bringing together exon 7 and exon 8.⁵¹

The research undertaken into silencers has demonstrated the importance of silencers in the regulation of splice site selection. However, the mode in which they exert their effects is still unclear. The research to date has so far provided evidence that both mechanisms proposed (Figure 1.3) as discussed in the preceding paragraphs occur. The field therefore requires model systems and a testing regime which can ascertain the molecular processes involved in splice site selection.

1.4.1.2 Splicing Enhancers

Pre-mRNA contains short diverse sequences within both exons and introns that exhibit functions that promote splicing at both 5' and 3' splice sites. SR proteins generally bind to these sequences promoting pre-spliceosome assembly.⁵⁴ Numerous investigations have taken place in identifying enhancer sequences but how these sequences exert their effect is still unclear and will be discussed in greater detail later.²³

1.4.1.2.1 *SR Proteins*

SR proteins are splicing factors that function in both constitutive and alternative splicing. They contain an N-terminal consisting of one or two RNA recognition motif (RRM) domains, that interact with the pre-mRNA and a C terminal RS domain; which is a region rich in serine and arginine dipeptides that interacts with proteins and RNA.^{9,55,56}

SR proteins were first discovered in the early 1990's by several groups. One of the first identified was the SRSF1 protein (also known as SF2/ASF) which was found to be integral in 5' splice site selection.^{54,57} Other SR proteins that have been identified include SRSF2 (also known as SC35), and SRSF3 (also known as SRp35). Both proteins have been shown to directly interact with the branch point promoting U2 snRNP association to the pre-mRNA and with the U1 snRNP at the 5' splice site stimulating pre-spliceosome assembly.⁵⁴ The function of SR proteins will be discussed in more detail in the exonic splicing enhancer section.

1.4.1.2.2 *Intronic Splicing Enhancers*

The studies into ISEs in general have been limited, although they are often located in introns of regulated exons and are necessary in the regulation of splicing. Interestingly SR proteins which generally bind to enhancer sequences are not involved in their regulation, however several other proteins such as hnRNPF, hnRNPH, KSRP and NOVA have been shown to stimulate splicing when bound to ISEs.^{23,58,59}

Several different ISE sequences have been determined such as the AU rich regions found in drosophila, which influences 5' splice site selection.⁶⁰ In 1997 McCullough and Schuler analysed the AU rich intronic sequences in intron 4 of β -con glycinin that contained two identical 5' splice sites in order to determine how the AU rich region exerted their effect. They discovered that when the AU region was located between the two 5' splice sites, the splice site located upstream of the AU region (distal site) was stimulated during splicing. The AU region promoted splice site recognition at the upstream site rather than masking the downstream 5' splice site, however, how this region promoted its effects was not clear.⁶⁰ A triplet of guanines (GGG) is commonly found in close proximity to a 5' splice site which is of predictive value for identification of exons. In particular the human α -globin gene was studied as it contained a number of GGG sequences within its intron. McCullough *et al.* discovered that these sequences stimulated splicing to the distal 5' splice site and instead of binding to SR-proteins which is common in ESEs, the sequences interacted directly with the U1 snRNP aiding the formation of the early prespliceosome complexes.⁶¹

The mechanism of action of ISEs is still poorly understood due to the complexity of splicing, and the limitations in biochemical technology there is no effective way to

determine the exact mode of action. What is known however is that they generally promote splicing to an adjacent splice site which is usually upstream of the ISE sequence. The ISEs either interact directly with the U1 snRNP in the early spliceosome complex of either E or A, or they interact with proteins such as KSRP or NOVA, which then exert their effects on splice site selection.⁵⁸⁻⁶²

1.4.1.2.3 Exonic Splicing Enhancers

In contrast to ISEs, a great deal of work has been done on ESEs and determining what cellular factors contribute to their ability to stimulate splice site recognition facilitating splicing. The first enhancer sequence identified was in exon ED111A of the fibronectin gene in 1987 where a centrally located 81 base pair region within a 270 nucleotide exon was deleted from the fibronectin gene, the wild type fibronectin generated two mRNA isoforms, however, with the 81 base pair deletion splicing only generated one mRNA isoform. More crucially however, was the reinsertion of the 81 base pair fragment into the fibronectin gene in the anti-sense orientation, did not reconstitute alternative splicing activity and, as with the 81 base pair deletion from the fibronectin gene only one mRNA isoform was generated. These findings demonstrated that this 81 base pair sequence was essential in the alternative splicing pathway of the fibronectin gene, and critically the orientation of the sequence was paramount.⁶³ Subsequently the fibronectin gene was investigated further in order to ascertain the structural and functional features of the ESE. The sequence was found to contain a nine nucleotide purine-rich sequence GGAGGAGAC, which exhibited properties that stimulated splicing at the upstream 3' splice site. The same sequence stimulated splicing in the human β -globin gene. This ESE sequence was found to interact with SR proteins which are required for A-complex formation and possibly the stabilisation of U2 snRNP binding.⁶⁴

Subsequent findings in alternative splicing of higher eukaryotes have found that SR proteins frequently bind to ESEs. SR proteins have one or two RNA recognition (RRM) domains which bind to pre-RNA, and C terminal domains that are enriched with arginine and serine (RS domains). These RS domains are highly phosphorylated enabling nuclear import of the SR proteins into the nucleus as well mediating spliceosome assembly.⁶⁵ Xiao and Manley showed the SR protein SRSF1 (also known as ASF/SF2) binds selectively to the U1 snRNP in an RS-domain dependent, phosphorylation enhanced manner, aiding in the assembly of the prespliceosome complex E and A.⁶⁶

Research into the function of ESEs has led to different hypotheses on how ESEs exert their effects on stimulating splicing with the aid of SR proteins. Tian *et al.* (1994) and Yue *et al.* (1997) proposed the U2AF recruitment model, which involves an SR protein bound to an ESE facilitates the binding of U2AF₆₅ to the polypyrimidine tract through an interaction with U2AF₃₅. which then forms interactions with the U1 snRNP at the 5' splice site, resulting in bringing together the spliceosome complex E which facilitates the rapid formation of the A complex (Figure 1.6a).⁶⁷⁻⁶⁹ Another model, proposed by Buzik *et al.* (1995), shows that a 3' splice site is stimulated by a downstream splicing factor where a SR protein binds to an ESE and interacts with the U1 snRNP stabilising binding to the 5' splice site at the opposite end of the intron; the U1snRNP then interacts with the U2AF at the polypyrimidine tract and the 3' splice site, stabilising U2 snRNP binding (Figure 1.6b).^{68,70,71}

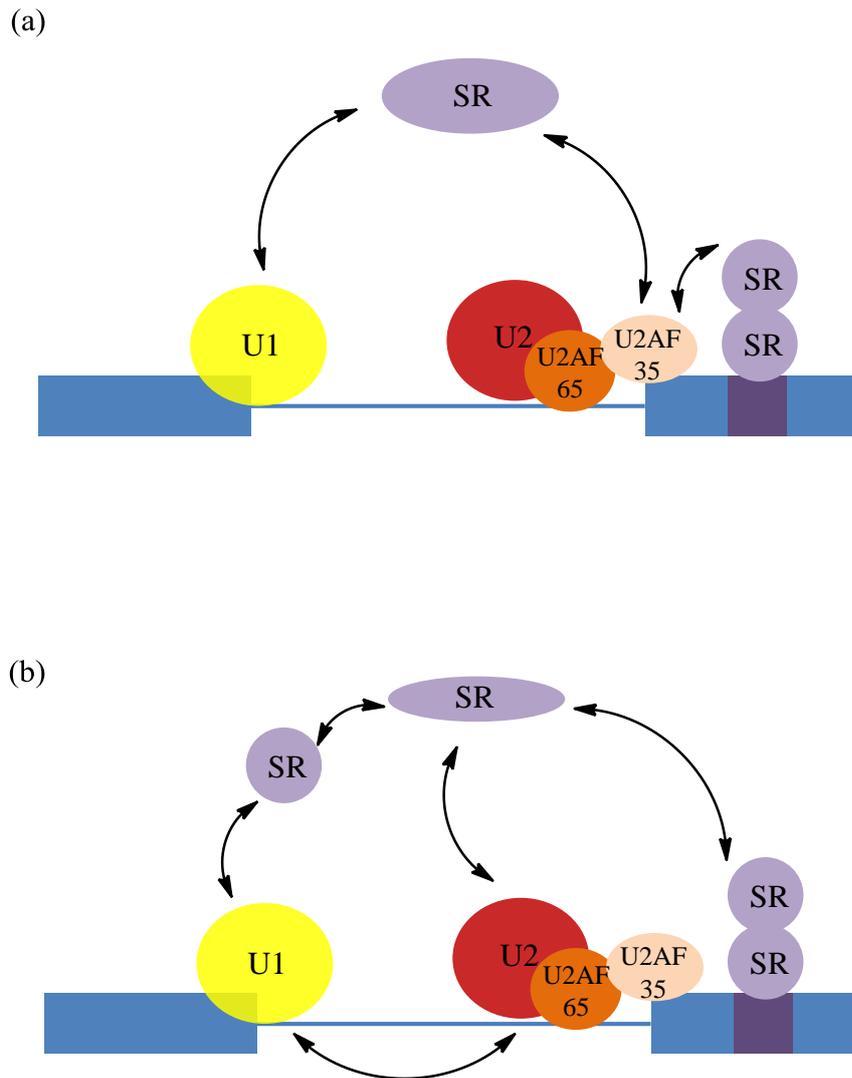


Figure 1.6 Putative models of exonic enhancer functions. (a) SR protein binds to ESE and interacts directly with U2AF_{65/35} splicing factor stabilising binding of U2 snRNP. Interactions then occur between the U2AF and another SR protein which interacts and stabilises the binding of U1 snRNP. (b) SR protein binds to ESE and interacting with the U1 snRNP at the 5' splice site. U1 snRNP then interacts and stabilises the binding of U2 snRNP to the pre-mRNA.⁶⁸

ESEs predominantly exert their effects on the 3' splice sites adjacent to ESE sequence, however there have been examples in which the ESE exerts its effects on the 5' splice site flanking the exon. It was found that activation of a weak 5' splice sites were activated via an MS2 enhancer sequence located upstream of the 5' splice site. Additionally splicing was enhanced 4 fold when SR proteins which specifically bound the MS2-ESE were added to the splicing reaction.^{72,73} These results verify the importance of ESEs but also the importance of SR proteins which bind to sequences

which stimulate splice site selection. In recent years the advancing bio-technology and computational technology has enabled the prediction of ESEs by computational methods, Fairbrother *et al.* for example has developed RESCUE-ESE which predicts sequences with ESE statistical analyses. These predicted ESEs are rich in purines and are repeat sequences enabling further studies into enhancers.^{41,74,75}

One major question confounding the field is how ESEs exert their effects on both the 3' and 5' splice sites? As previously mentioned there are models in relation to the SR binding to either U2AF directly or to the U1 snRNP which then interacts with the U2AF stabilising the spliceosome formation. However these models do not propose how these interactions take place. There are 2 major hypotheses:

1. RNA looping by 3-dimensional diffusion of the SR proteins to the splicing factors. (Figure 1.7a).^{2,47}
2. Protein propagation model; SR proteins cooperatively bind across the RNA forming interactions with splicing factors (Figure 1.7b).^{17,31}

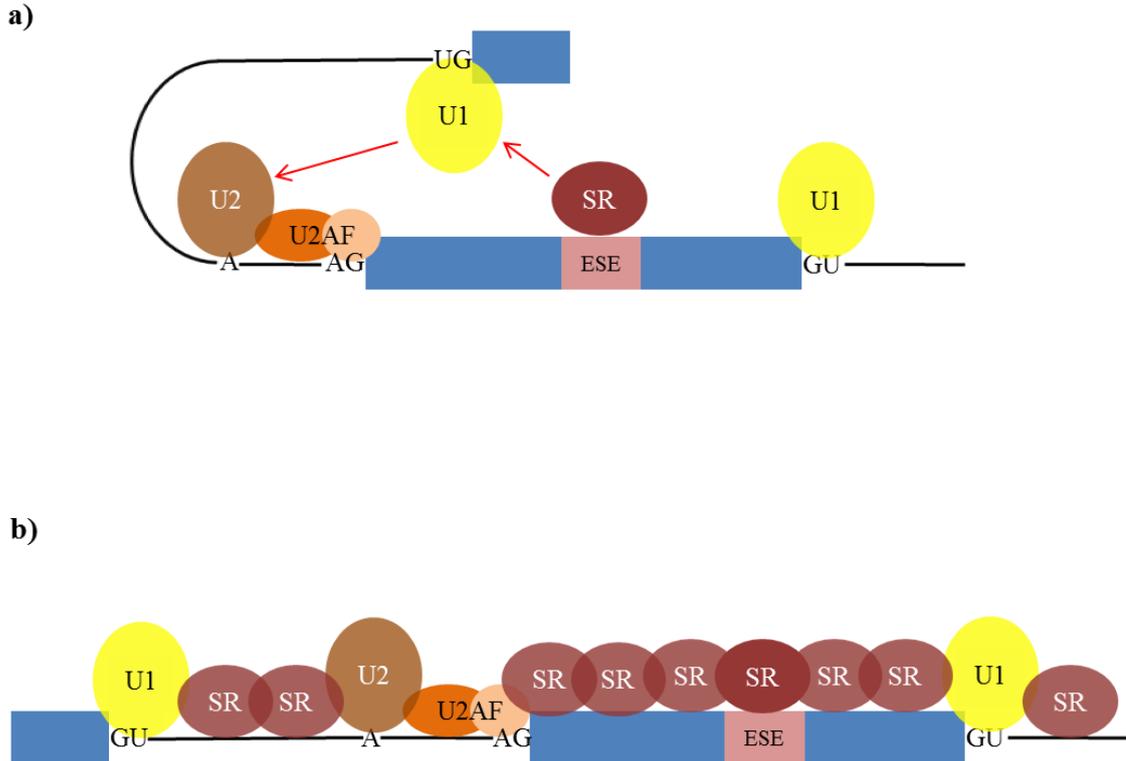


Figure 1.7 Diagrammatic representation of the two proposed mechanism of enhancer activity. (a) the looping model - SR proteins collide with activating proteins by 3D diffusion (red arrows) creating a loop. (b) the propagation model - A family of SR proteins cooperatively bind across the pre-mRNA and eventually binding the snRNPs.

1.4.1.2.3.1 RNA looping – 3-dimensional diffusion model

RNA looping is the most widely accepted model for the mechanism by which ESEs exert their effects on splice sites situated several hundred nucleotides away. An SR protein binds to the ESE and due to the flexibility of RNA is able to loop around coming into direct contact with either U1 or U2 snRNP or U2AF; interacting with these splicing factors via 3 dimensional diffusion (Figure 1.7a). The critical evidence for the RNA looping model came about when Graveley *et al.* (1998), investigated the factors that determined the strength of splicing enhancers of pre-mRNA.⁷⁶ They analysed model substrates where an RS domain was tethered to an ESE and found that the rate of splicing (r) related to the number of nucleotides (n) between the 3' splice site and the ESE-RS domain complex [$r = kn^{-3/2}$]. The relationship between the rate of splicing and

the ESE-SR complex holds true regardless of which splice site (5' or 3') is being targeted and the data is most consistent with RNA looping model.⁷⁶ Furthermore, Shen *et al.* showed by RNA-protein UV cross-linking experiments that when an ESE was tethered to an RS domain, the RS domain was in contact with the branchpoint. The interaction of the RS domain with the branchpoint could be promoting pre-spliceosome assembly.⁹

1.4.1.2.3.2 Protein propagation – 1-dimensional model

Despite the evidence for the 3-dimensional diffusion model there is also competing evidence which suggests ESE-SR complexes are exerting their effects by 1-dimensional protein propagation model. Initial binding of an SR protein to the ESE is followed by the binding of other co-activators such as the large SRm160 RS domain protein across the RNA which was observed by Caceres *et al.* (1993) (Figure 1.7b).^{17,57} Chiara *et al.* (2001) identified numerous exonic sites within the pre-mRNA to which SR proteins could be cross-linked implying that numerous SR proteins across the RNA are involved in splicing.⁷⁷ In particular two members of the SR protein family were found to bind to different sites upstream of the 5' splice site, providing one of the first examples of 5' splice sites stimulated by ESE-SR protein complexes upstream of the 5' splice site.⁷⁷ Sciabica *et al.* also found that the binding of SR protein to ESE reduced the length of the pre-mRNA rendering the rate data⁷⁶ inconclusive.⁷⁸ The data was reanalysed the results suggest that $r = \sim kn^{-5/2}$ which does not support the 3-dimensional diffusion model.⁷⁹

An additional argument for the propagation model is the presence of numerous proteins found in the nuclear extract *in vitro* that bind to the RNA with low specificity including

SR proteins. These proteins are likely to bind between the ESE and the splice sites and could play an integral part in promoting splicing.

1.5 AIMS

The objective of this thesis was to determine the mechanisms by which enhancers activate splicing. The literature to date has reported two plausible modes of action; proteins exerting their effect by 3-dimensional diffusion (Figure 1.7a) or 1-dimensional propagation of proteins along the pre-mRNA (Figure 1.7b). Both models are considerably different: the former relies on a through-space interaction between distant sites, whereas the latter is reliant on multiple protein-protein and protein-RNA interactions to define the location of splice sites. Currently, there is no general test to determine the mechanism of splice site selection, which is a considerable limitation in understanding this critical biological process. The main objective therefore is to develop new techniques which could test the effects of ESEs and RNA binding proteins on splice site selection at both the 5' and 3' splice sites.

1.5.1 Method Development

This project tackled this challenge by tethering an ESE sequence via linkers to pre-mRNA transcripts at both the 5' and 3' end. Using *in vitro* splicing assays the effects of these non-RNA linkers will be measured (Figure 1.8). It is envisaged that the linkers will maintain the length and connectivity of RNA; however, they will lack the capacity to bind to the RNA binding proteins. If they reproduce splicing effects of the natural sequence then we can deduce that splicing is regulated by a looping mechanism.

A linker strategy is therefore required which enables tethering of an ESE to a transcript. Polyethylene glycol (PEG) (Figure 1.8) would be a suitable linker as it is hydrophilic, neutral in charge, linear and is non-toxic.⁸⁰ The polyether backbone is inert in biological systems as well as most chemical reaction conditions. Furthermore a prime property of PEG is its ability to repel macromolecules, therefore the proteins cannot bind across and therefore propagation of proteins along the RNA will be impeded.⁸¹ In addition, a previous study by Pasman *et al.* (1996), showed that when a PEG linker consisting of 10 ethylene glycol repeats was inserted into an intron 24 nucleotides upstream of the branch point, splicing of the transcript still took place which showed that PEG is not detrimental to the splicing process.²

There are several other linkers that could be synthesised and inserted into between the ESE and the RNA transcript. These include polyamides (Figure 1.8) which could modulate the rigidity. Polyphosphates (Figure 1.8) provide the means to mimic the phosphodiester backbone of RNA and glycol nucleic acids (GNA) (Figure 1.8) which are structurally the most simplified phosphodiester containing nucleic acid analogues.⁸²⁻

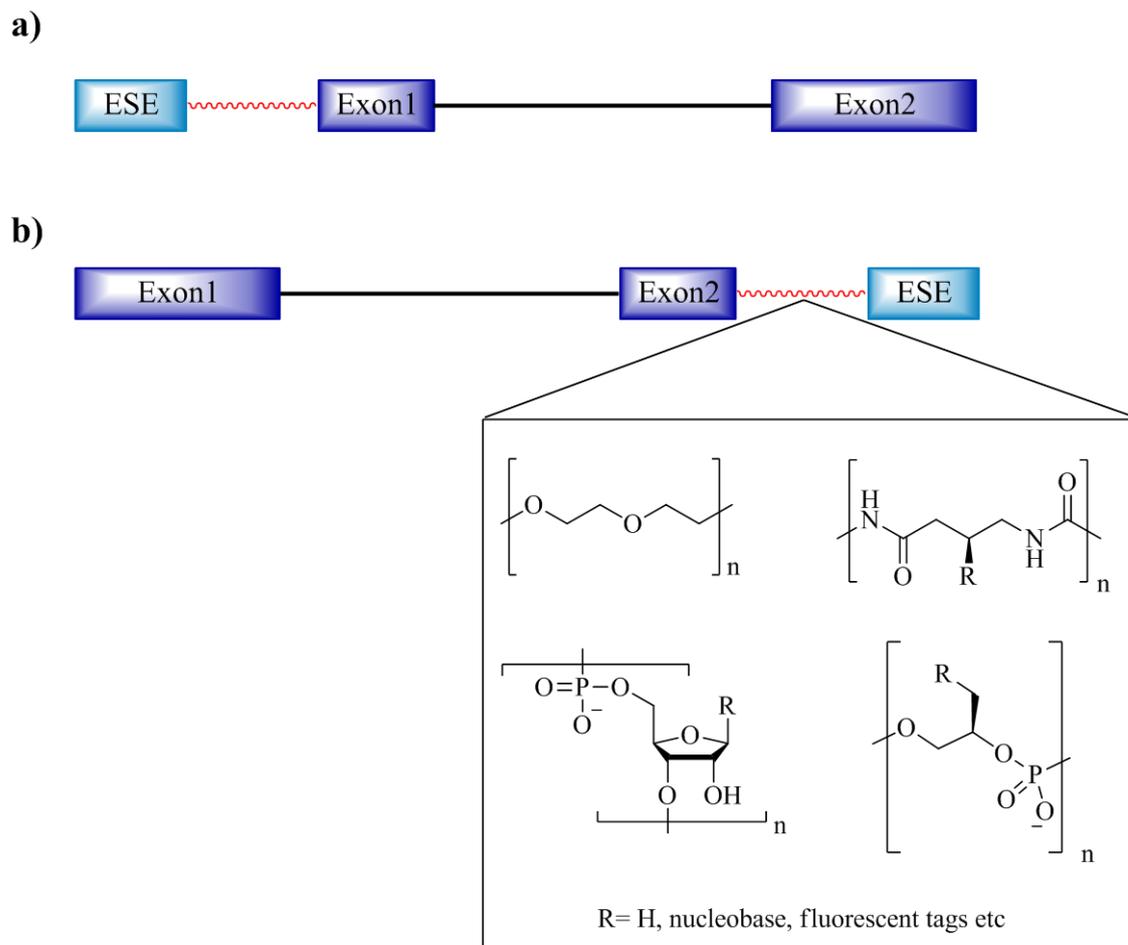


Figure 1.8 Schematic representation of the synthetic RNA tripartite systems to be investigated. a) The ESE is located at 5' end separated from transcript by a flexible tether (red wavy line). b) ESE is tethered to the 3' end. Potential tethers are highlighted in box; PEG, polyamide, polyphosphate, GNA.

1.5.2 Hypothesis to be Tested

The mechanism by which ESEs exert their effects of splice site selection and activation is still unclear. The studies undertaken so far have brought about two very different models; the 3-dimensional diffusion model and the 1-dimensional propagation model as discussed in Section 1.4.1.2.3. The aim of this was to test the 3-dimensional diffusion (looping) model for the mechanism by an ESE stimulating a splice site. A fundamental property of the looping model is that the outcome is determined by the probability of a direct encounter between components at the two sites, and therefore the nature of the

connection between them is immaterial as long as it does not restrict conformational flexibility.

1.6 DEVELOPING METHODOLOGY FOR THE CONSTRUCTION OF TRIPARTITE TRANSCRIPTS

In order to test how ESEs exert their effect on alternative splicing an RNA tripartite system, where a short ESE will be tethered to a longer portion of RNA using a combination of synthetic and enzymatic techniques, will be constructed. ESE sequences will be synthesised by automated solid phase synthesis whereas the longer portion of RNA will be enzymatically synthesised during transcription by T7 RNA polymerase. Bioconjugation of these strands will form a tripartite structure ready for splicing (Figure 1.9).

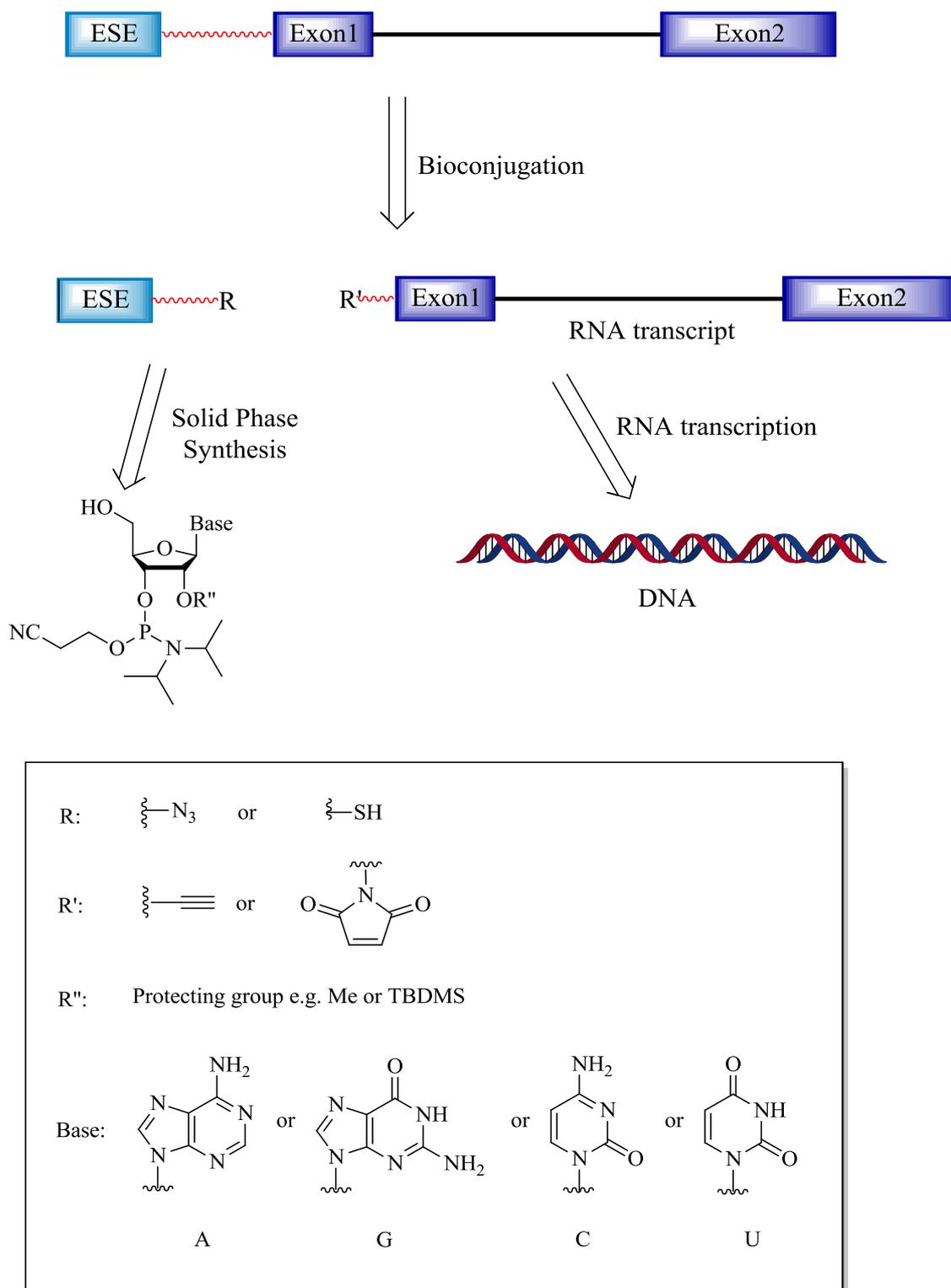


Figure 1.9 Retrosynthetic route for the construction the of the RNA tripartite constructs. A: adenosine, G: guanosine, C: cytidine, U: uridine.

1.6.1 Enzymatic Synthesis of Modified RNA

Transcription is the process biological systems use to transcribe DNA sequences into RNA. The standard method for producing pre-mRNA for *in vitro* splicing studies is to use T7 RNA polymerase which catalyses the synthesis of RNA in the 5' to 3' direction from a DNA template. At the beginning of transcription a 7-methyl-guanosine (m^7G) cap is incorporated at the 5' end of the pre-mRNA by the addition of the diguanosine triphosphate (**1**) which is recognised by the T7 polymerase. The m^7G cap protects the RNA from degradation from RNases. Transcription is the most efficient method for synthesising long strands of RNA and transcripts can be easily radiolabelled with ^{32}P by the addition of [α - ^{32}P] NTP into the transcription reaction.

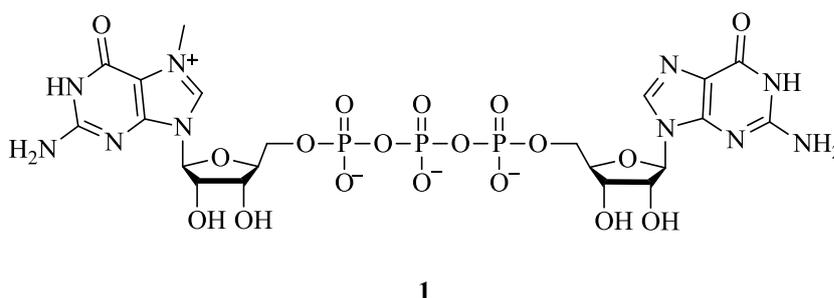


Figure 1.10 7-methyl-diguanosine triphosphate (**1**) used to cap the 5' end of RNA during transcription.

Modification of the transcript at either the 3' or the 5' end is required to conjugate it to a non-RNA tether. In recent years several groups have developed methodology to label transcripts at the 5' end using modified monophosphate guanosines (**2-7**). Compounds (**2-7**) were incorporated into RNA during the transcription process at the 5' end (Scheme 1.3a).^{80,85-93} Additionally RNA could be modified at the 3' end by ligating modified dinucleotides using T4 RNA ligase (Scheme 1.3b).^{94,95}

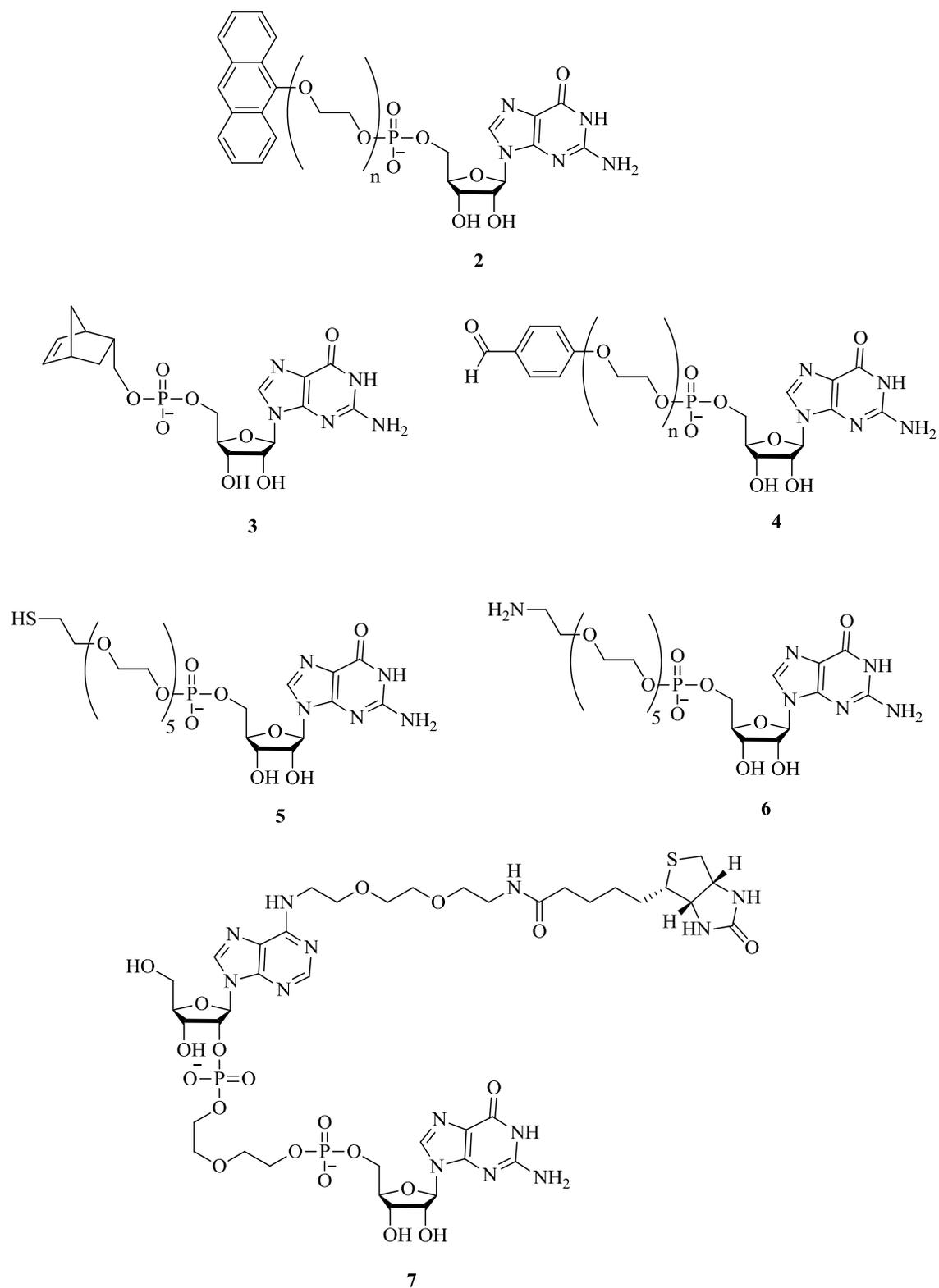
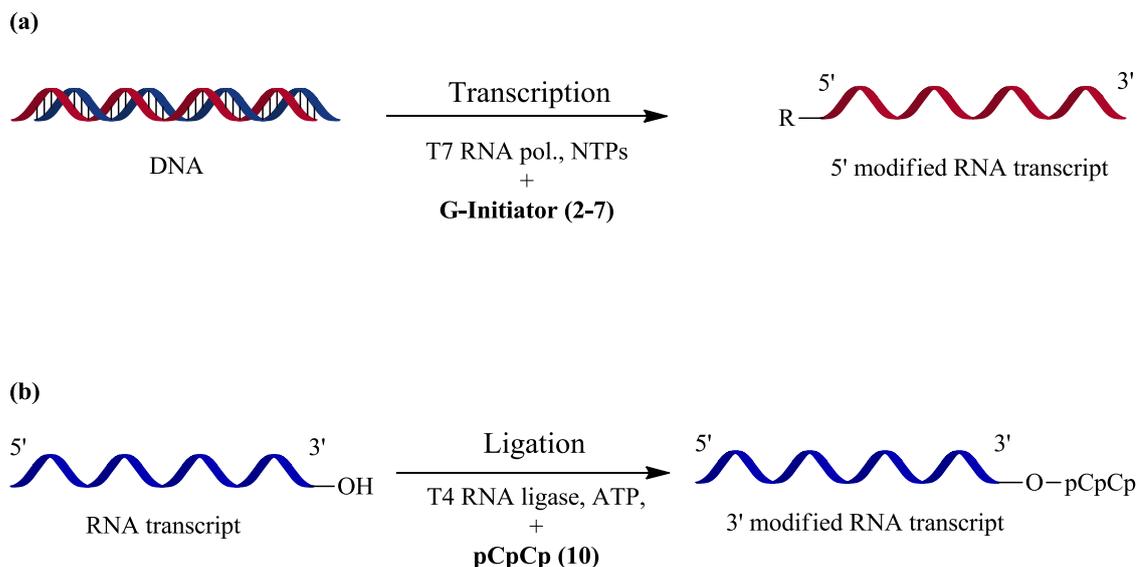


Figure 1.11 Representative examples of G-initiators synthesised and incorporated into RNA transcripts.^{86,88,92,96}

Scheme 1.3 Schematic representation of enzymatic incorporation using G-initiators (2-7). Reaction (a): Transcription; 2-7 is used as a Cap in the transcription of RNA; R = G-initiators (2-7). Reaction (b): Ligation; (10) is ligated on a RNA transcript by T4 RNA ligase; R' = pCpCp (10).



1.6.1.1 Incorporation of functional groups on the 5' end of transcripts

During *in vitro* transcription the pre-mRNA is capped with m^7G that is selectively incorporated at the 5' end after initiation of transcription. The modified guanosine cannot be incorporated into the rest of the pre-mRNA during the elongation process as only nucleotide triphosphates (NTPs) are utilized.⁸⁵ The process of capping provides a plausible route for incorporating modified guanosines (G-initiators) into mRNA by replacing the usual m^7G cap with a guanosine monophosphate derivative.

Jäschke *et al.* (1999), developed one of the first G-initiators, whereby an anthracene was tethered to the 5' end of the guanosine via a PEG linker (2). Replacement of the m^7G cap with G-initiator (2) during transcription of a 159 nucleotide transcript enabled 57 % incorporation.⁸⁸ A variety of research groups have expanded the diversity of 5' functionality in transcripts ranging from fluorophores (8) to biotin (9) tags. An array of

functional groups have been used, including thiols for thiol-maleimide coupling, amines for NHS-ester coupling, aldehydes that can be reacted with amines and hydrazines and a norbornene which underwent a Diels Alder reaction with a tetrazine fluorophore (**8**). The efficiency of incorporation of these G-initiators (**5**) and (**6**) enabled 65% and 60% incorporation into a 10 nucleotide transcript respectively, whereas (**3**) enabled 70% incorporation into a 19 nucleotide transcript and (**4**) enabled 52% into a 44 nucleotide transcript.^{85–89,96,97} Wolf *et al.* developed a much larger G-initiator (**7**) comprising a PEG-tethered adenosine that was modified at the N⁶ position with biotin, but incorporation rates were low (2%).⁹²

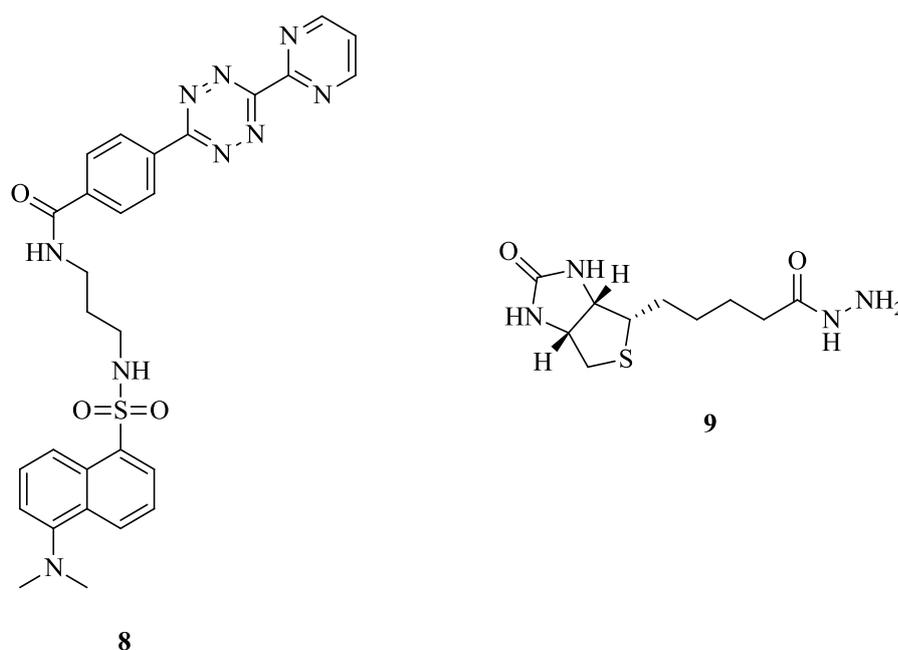


Figure 1.12 Tetrazine fluorophore (**8**) and biotin hydrazine (**9**) used to conjugate to G-initiators incorporated into RNA.⁸⁷

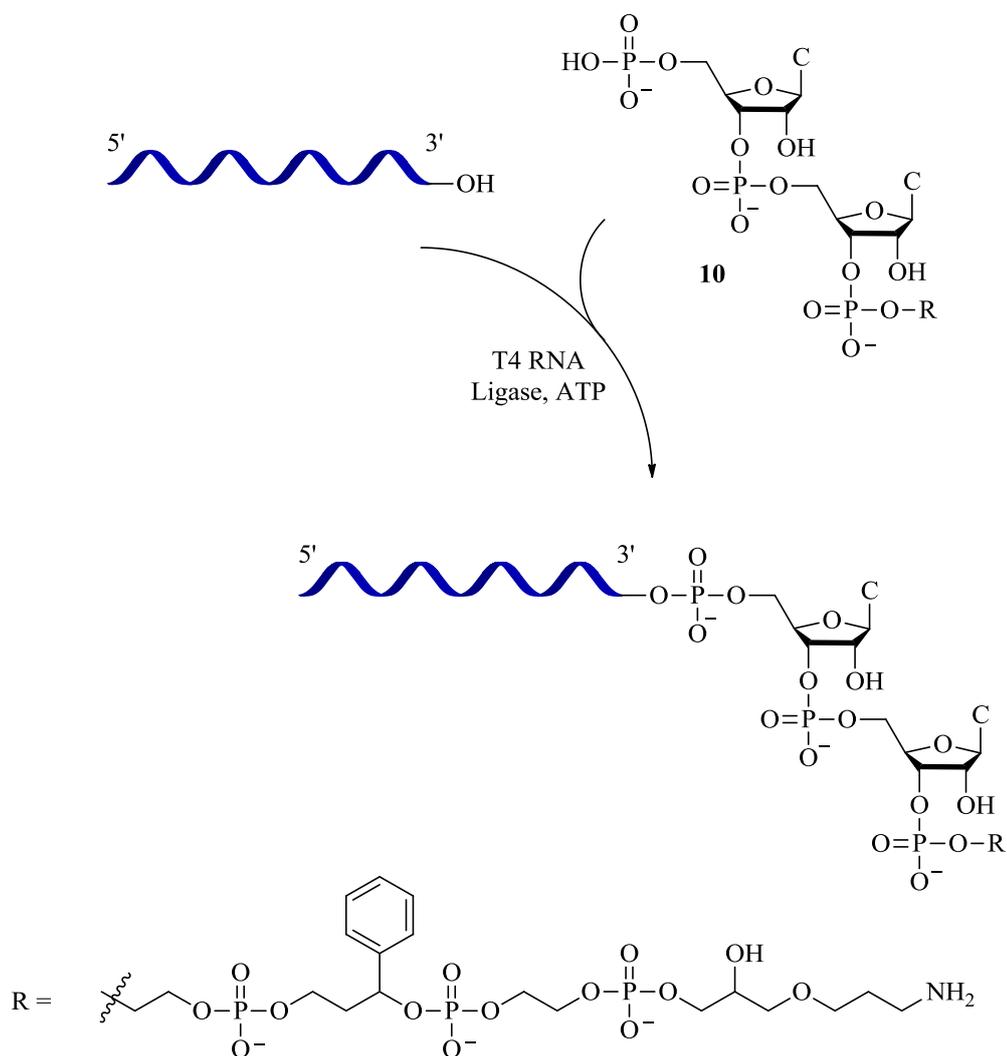
The research to date has shown the potential of G-initiators for modifying the RNA transcript at the 5' end as the modification can be incorporated during the transcription process with ease. However, the design of the initiator will be important as the initiators synthesised have all been incorporated into RNA with varying degrees of efficiency.

1.6.1.2 Incorporation of functional groups onto 3' end

The introduction of modifications onto the 3' end of a transcript has not been studied in as much detail as incorporation of modified guanosines into the 5' end. Early studies showed that T4 RNA ligase catalyses the formation of an internucleotide phosphodiester bond at the 3' end in an ATP-dependent manner. The reaction occurred between a donor 5' terminal phosphate group and a 3' terminal hydroxyl group (Scheme 1.4). The minimum requirement for the acceptor was that it had to contain at least a tri-nucleoside diphosphate and an effective donor could be any mononucleoside 3'.5'-bisphosphate (pNp).^{98,99}

Phosphorylated dinucleotides (**10**) comprising two cytidines which were phosphorylated at the 5' end (pCpCp) and modified with a photocleavable linker at the 3' end have been prepared by Jäschke *et al.* This pCpCp (**10**) was ligated to RNA transcripts in 80% yield obtained (Scheme 1.4).^{94,95}

Scheme 1.4 Schematic representation of the ligation of a dinucleotide (**10**) to RNA using T4 RNA ligase. 3'-OH of RNA transcript is ligated to the 5'-phosphate group of **10**.^{94,95}



To conclude two enzymatic methods (transcription & ligation) provide plausible routes for incorporating modified bases into long strands of RNA at either the 5' or 3' ends. The modified enzymatic RNA would then be ready to bioconjugate to modified ESEs creating the RNA tripartite constructs.

1.6.2 Solid Phase Synthesis

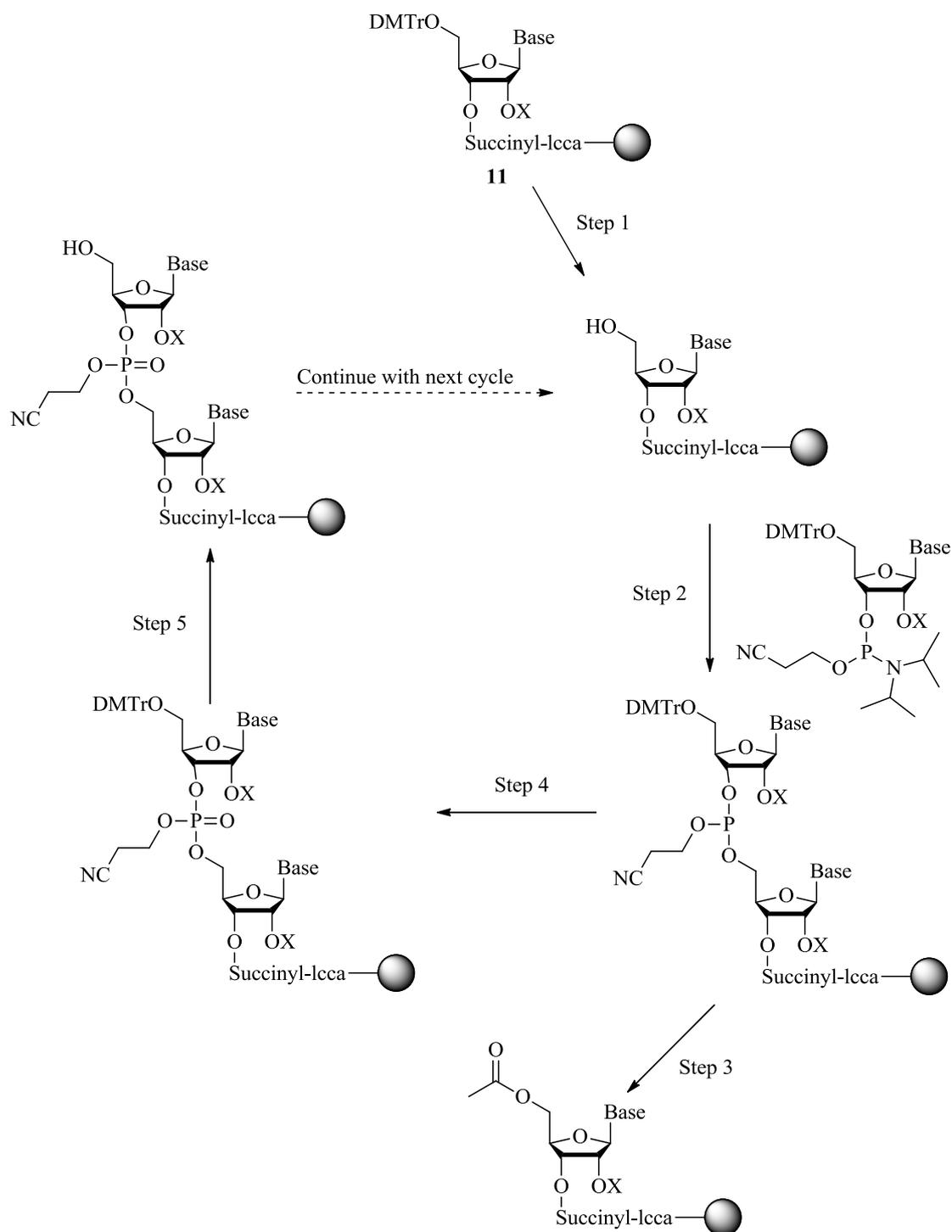
The first chemical synthesis of a dinucleotide phosphate and a dinucleotide with natural internucleotide linkages was reported over 50 years ago.¹⁰⁰ Since then synthesis has

improved dramatically especially after Letsinger and Lunsford reported that P(III) was considerably more reactive than P(V) phosphorylating agents.¹⁰¹ Further to this discovery the first deoxynucleoside phosphoramidite was synthesised revolutionising the synthesis of oligonucleotides.¹⁰² The synthesis of RNA is more complicated than DNA due to the additional hydroxyl group located at the 2' position of the ribose ring, which requires selective protection during synthesis. The 2' OH protecting group used extensively has been TBDMS. The large steric bulk of the TBDMS group contributed to the decrease in coupling efficiency.¹⁰³ As a result of the reduction in coupling efficiency, only strands of up to 50 nucleotides can be prepared before the yield becomes unbearably low.

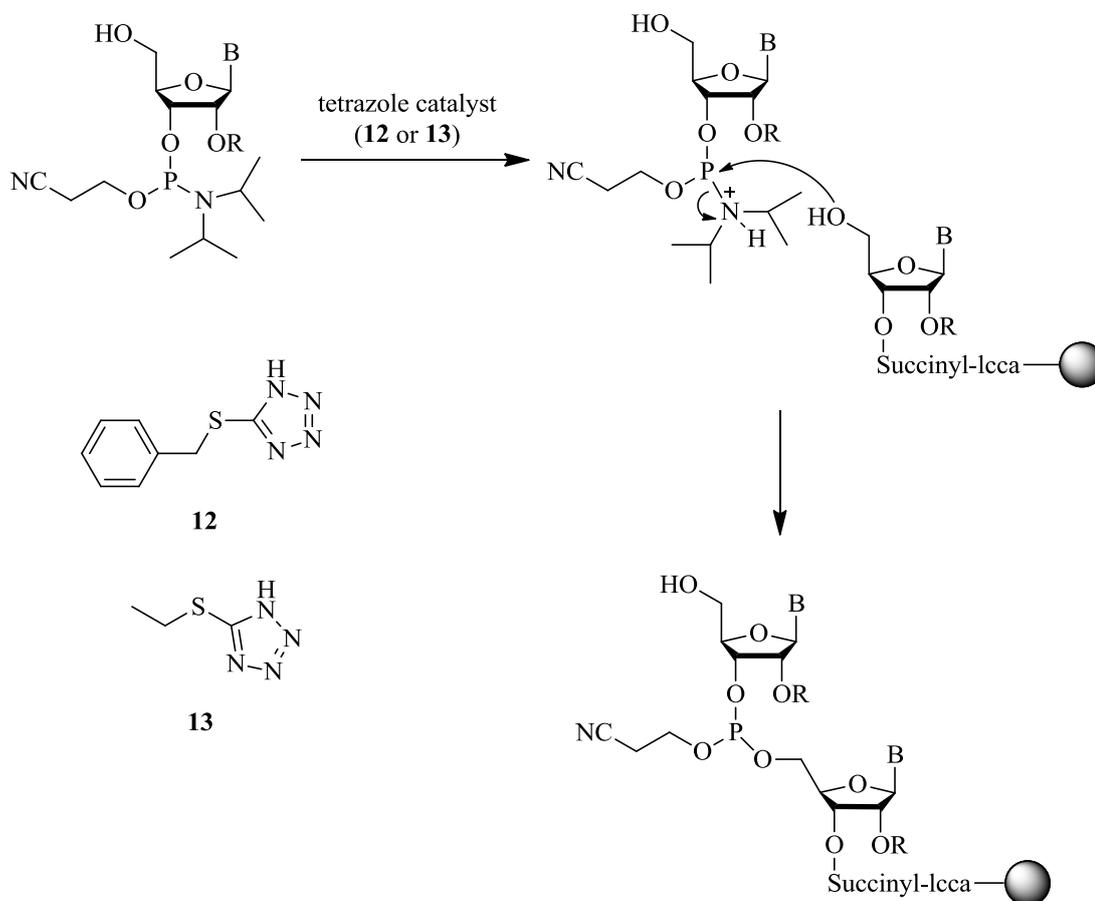
Solid phase synthesis of RNA proceeds from 3' → 5' (Scheme 1.5). The first base or modifier at the 3' end is attached via a succinyl group to a controlled pore glass (CPG) support (Scheme 1.5, **(11)**). The 5' hydroxyl group is protected with the acid-labile 4'-dimethoxytrityl (DMTr) group and is removed in step 1 of the synthetic cycle (Scheme 1.5, step 1). The DMTr group is removed by the addition of 3% trichloroacetic acid in DCM. Coupling of the next base/modifier phosphoramidite occurs (Scheme 1.5, step 2). The phosphoramidite is activated by either 5-benzylthio-1-H-tetrazole (**(12)**) or 5-ethylthio-1-H-tetrazole (**(13)**). The tetrazole protonates the diisopropylamine group, which is then displaced by nucleophilic substitution by the 5'OH of the support bound nucleoside (Scheme 1.6).^{104,105} A capping step is then included in the cycle in order to reduce complicated mixtures of RNA products (Scheme 1.5, step 3). Oxidation of the phosphotriester with aqueous I₂, forms protected version of the natural P(V) phosphodiester of nucleotides (Scheme 1.5, step 4). Deprotection of the DMTr group yields a 5'OH (Scheme 1.5, step 5). The process is then repeated until the addition of each base/modifier is complete. Once the sequence is complete the bases are

deprotected and cleaved from the solid support with aqueous ammonia and purified by RP-HPLC.

Scheme 1.5 Solid phase synthesis of oligoribonucleotides with first base (B) attached to a CPG resin. B indicates bases A, C, G & U, X = TBDMS, Me, TC. Step 1: deprotection of the DMTr group; Step 2: coupling of the base (B) phosphoramidite; Step 3: capping of unreacted bases; Step 4: oxidation of phosphotriester; Step 5: deprotection of DMTr



Scheme 1.6 Activation of the phosphoramidite by a tetrazole (12 or 13) catalyst. B indicates bases: A, G, C or U and R = protecting group e.g. Me, or TBDMS.

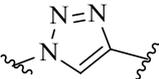
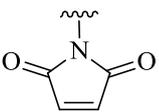
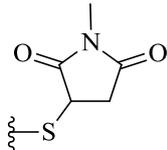
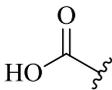
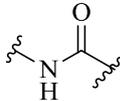
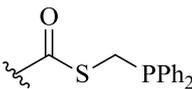
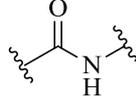
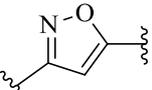
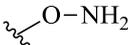
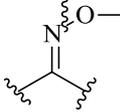
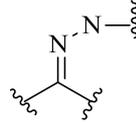
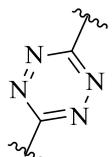
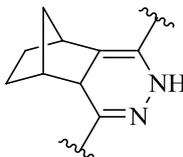


1.6.3 Bioconjugation Chemistry for the Ligation of the RNA Strands

In order to form the tripartite RNA constructs whereby a non-flexible RNA linker such as PEG is inserted between the ESE and the splice site bioconjugation of the two strands is required (Figure 1.8). The ESE will be modified by solid phase synthesis as discussed in Section 1.6.2 with a functional group such as an azide and the RNA transcript will be modified enzymatically (Section 1.6.1) with the corresponding functional group, in this case an alkyne. These modifications will enable the two strands to be conjugated together forming the RNA tripartite structure.

There are a several viable bioconjugation techniques suitable for conjugating biomolecules to biotin and fluorescent tags for detection both *in vitro* and *in vivo*. These methods include thiol-maleimide coupling, amide coupling, Staudinger ligation, oxime and hydrazone formation, nitrile oxide-alkyne cycloaddition, Diels-Alder and ‘click’ chemistry (Table 1.1); although in recent years click chemistry has become the method of choice due to its bio-orthogonal nature in particular.^{87,97,106-109}

Table 1.1 Bioconjugation techniques suitable for the conjugation of biomolecules.

Reagents		Products	Conjugation Reaction
R	R'		
			Click Chemistry
			Thiol Maleimide Coupling
			Amide Coupling
			Staudinger Ligation
			Nitrile oxide-alkyne cycloaddition
			Oxime formation
			Hydrazone formation
			Diels Alder

1.6.3.1 Amide coupling and thiol-maleimide coupling

Reactive amino (-NH₂) and sulfhydryl (-SH) groups can be incorporated into RNA during solid phase synthesis at both the 3' and 5' end of the strands using phosphoramidites such as (14) and (15). The reactive amino group is reacted with an N-hydroxysuccinimide ester (NHS-ester, (16), Figure 1.13b) whereas the incorporated sulfhydryl group is reacted with maleimide derivatives ((17), Figure 1.13c). Amide coupling is typically easier than thiol-maleimide coupling as the intermolecular disulfide bridges form between the strands of RNA during deprotection of the bases, which require a mild oxidant to free the sulfhydryl group. However, amide coupling presents its own set of problems as the reaction mixture cannot contain any buffers with free amines such as Tris, as the NHS-ester will degrade and potentially cross react attributing to low yields of the reaction, additionally the pH needs to be optimal 8.0-9.0, too low the amino group will be protonated and reactivity will be low, too high the NHS-ester will hydrolyse faster than it reacts with the oligonucleotide.¹¹⁰⁻¹¹⁴

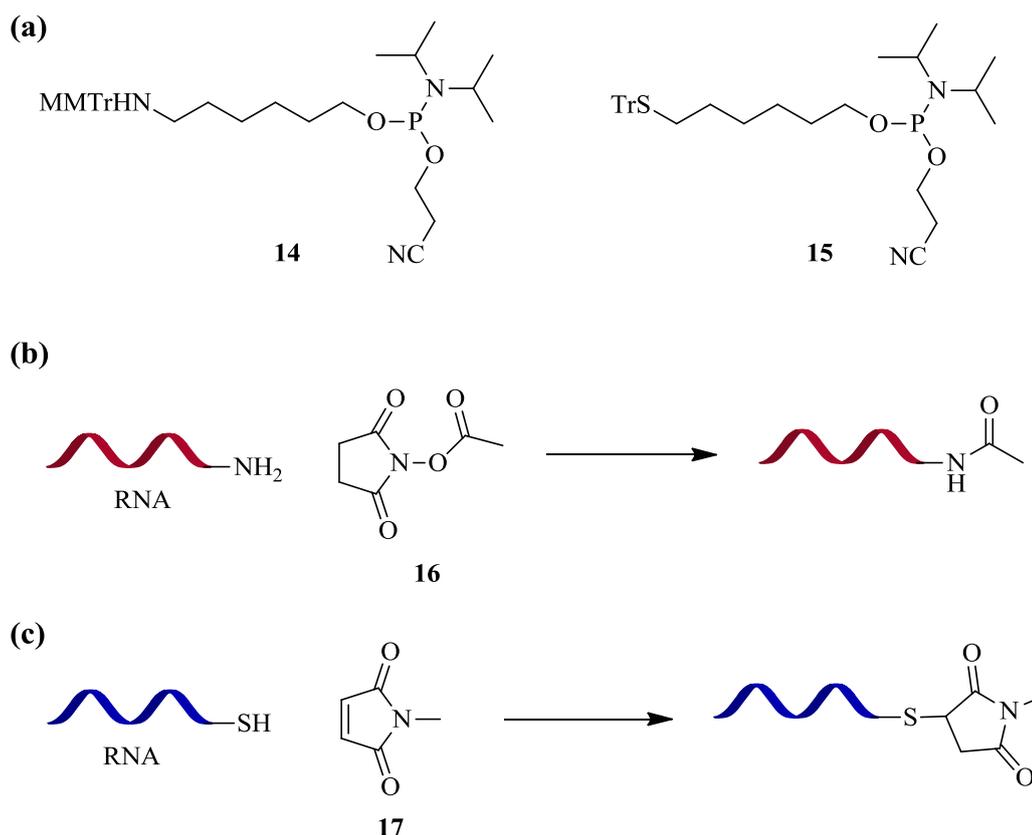


Figure 1.13 (a) Amino and sulfhydryl modified phosphoramidites (**14** and **15** respectively). (b) Schematic representation of NHS coupling to the amino modified RNA transcript. (c) Schematic representation of thiol-maleimide coupling.

1.6.3.2 Staudinger Ligation

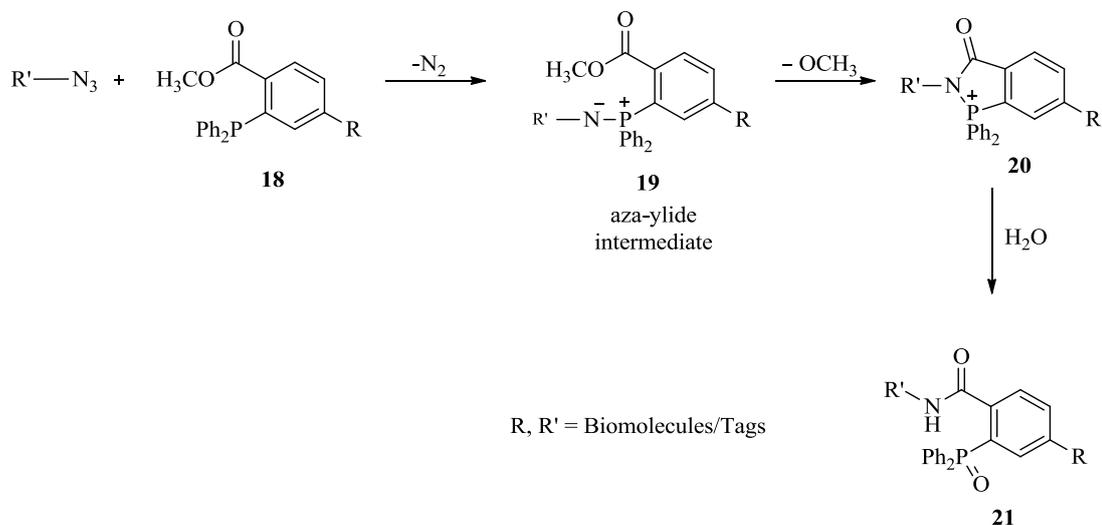
The potential of the Staudinger ligation as a bioconjugation method was first recognised by Bertozzi in 2000, where the ligation between phosphines and azides formed a stable amide bond (**21**) (Scheme 1.7a).¹¹⁵ To avoid hydrolysis of the aza-ylide intermediate (**19**), phosphine reagents were designed that had an ester group placed in the ortho position (**18**). This ester group captures the aza-ylide intermediate (**19**) by cyclisation (**20**). Subsequent hydrolysis of this cyclic intermediate (**20**) produced a stable amide bond; however, the structure also included a phosphine oxide within the structure (**21**).^{116,117}

To eliminate the phosphine oxide from the structure both Bertozzi and Raines independently developed new phosphine reagents that included thioester phosphines (**22**). The intermediate aza-ylide (**23**) undergoes an intermolecular reaction producing an amidophosphonium salt (**24**) and upon hydrolysis gives an amide (**25**), Scheme 1.7b).^{106,118} This reaction is known as the traceless Staudinger ligation.

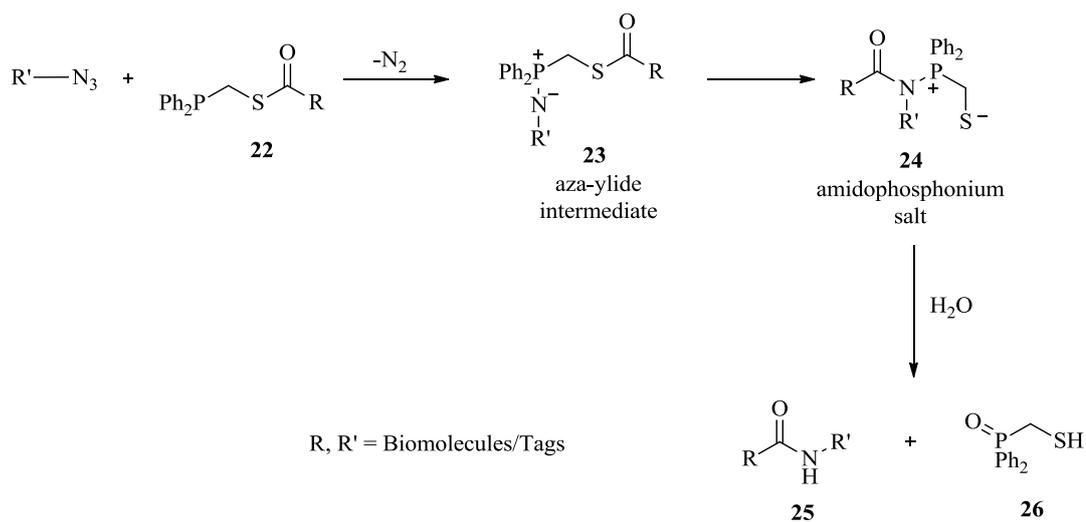
Despite its potential as a method to bioconjugated biomolecules, there are structural drawbacks. The Staudinger ligation suffers from slow reaction kinetics, and a requirement for high concentrations of triarylphosphines (> 250 μM).¹¹⁶ The reaction with alkyl azides displays second order kinetics where the rate determining step is the attack of the phosphine on the azide. Improving the reaction kinetics by increasing the nucleophilicity of the phosphine has resulted in the increased susceptibility of phosphine oxidation in air disrupting the desired ligation.¹¹⁶ Furthermore, the less stable thioesters are generally difficult to synthesise, rendering these reactions less favourable compared to other methods developed.^{116,117}

Scheme 1.7 a) Non-traceless Staudinger ligation. b) Traceless Staudinger ligation.

(a)



(b)



1.6.3.3 Click Chemistry

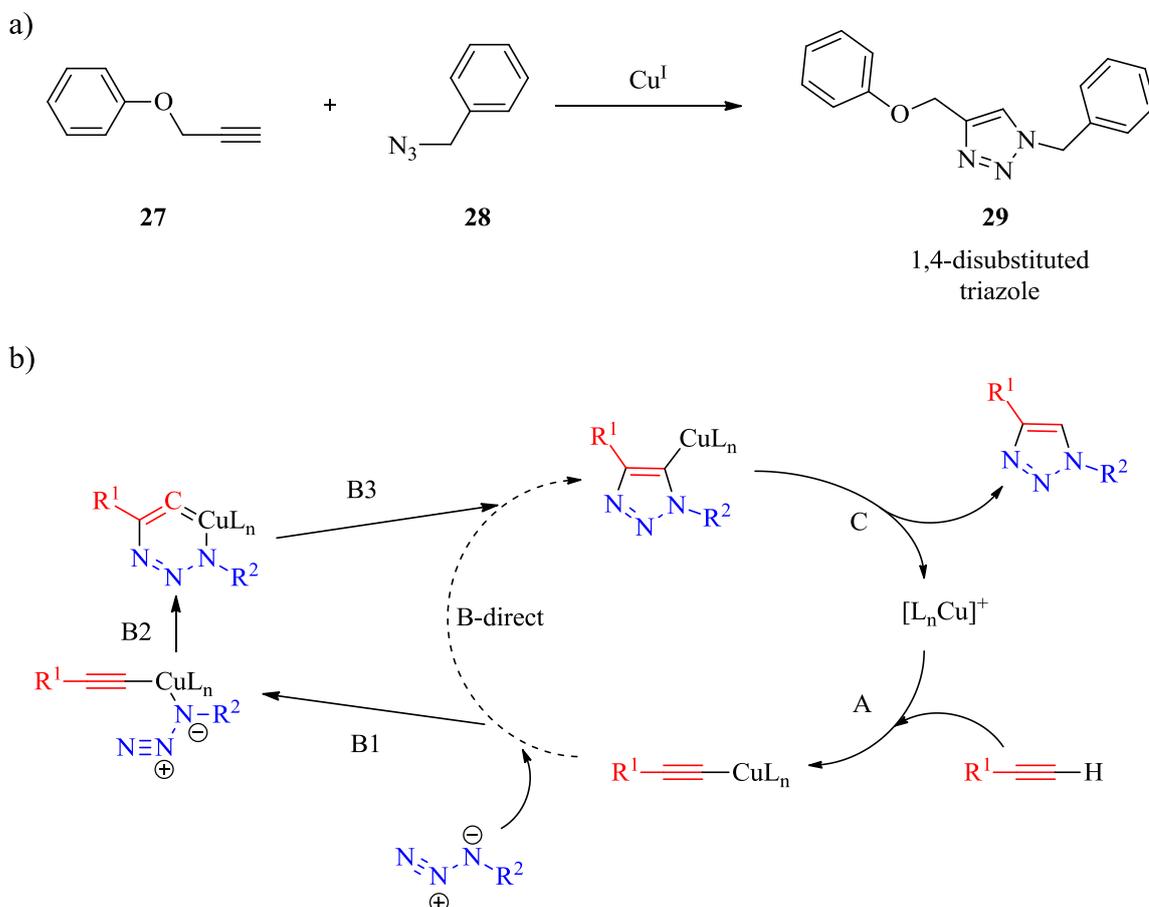
The term “click” chemistry refers to any reaction that is extremely rapid, highly selective, high yielding with little or no by-products, performed in mild reaction conditions and solvents at ambient temperatures.^{113,119} In recent times the term click chemistry generally refers to the copper catalysed alkyne-azide cycloaddition (CuAAC). These reactions are advantageous in biological systems due their bio-orthogonal nature.

Of all the bio-orthogonal techniques developed the azide functionality is the most popular as it is metabolically inert, and participates in selective reactions with other abiotic groups. The “bio-orthogonality” has been exploited extensively *in vitro* and more recently *in vivo*.¹¹⁵

1.6.3.3.1 The copper-catalysed Huisgen [3+2] cycloaddition: click chemistry

Copper (I) catalysed alkyne-azide cycloaddition (Scheme 1.8a) was first reported by Sharpless in 2002.¹²⁰ The reaction is based on the Huisgen [3+2] cycloaddition where an alkyne is reacted with an organic azide at elevated temperatures forming a triazole group resulting in a mixture of 1,4 and 1,5 regioisomers. Sharpless found that by adding a catalytic amount CuSO₄ in the presence of sodium ascorbate to a reaction containing phenyl propargyl ether (**27**) and benzylazide (**28**) resulted in the exclusive formation of 1,4-disubstituted triazole (**29**). Furthermore the reaction took place at room temperature in aqueous conditions resulting in high yields, with an increased reaction rate.¹²⁰

Scheme 1.8 a) Copper Catalysed alkyne-azide cycloaddition b) Mechanism for the copper catalysed click reaction proposed by Sharpless. R^1 and R^2 are biomolecules and probes. The catalytic cycle has two proposed mechanisms: B-direct which is the less favoured concerted [2+3] cycloaddition or the step wise, annealing sequence (B1 \rightarrow B2 \rightarrow B3) which proceeds via a 6 membered copper-containing intermediate.¹²⁰



The discovery of copper-catalysed click chemistry has revolutionised the bioconjugation of biomolecules and in particular DNA and RNA, as the azide group possesses high intrinsic activity, is highly selective, is stable in water and does not interfere with biochemical processes occurring in the cell.^{121,122} Additionally, incorporation of alkyne groups in DNA and RNA via solid phase synthesis as phosphoramidites is commercially available, and azide groups can be incorporated efficiently through NHS-azide coupling to a primary amine.

There are many examples in the literature that have utilised the copper click reaction in DNA. Carell *et al.* incorporated alkyne 5-position-modified-pyrimidines and modified purines into double stranded DNA enzymatically using DNA polymerase during PCR (polymerase chain reaction); the incorporated alkynes reacted efficiently with various azides including coumarin azide.^{123–126}

The compatibility of the copper click reaction with DNA would suggest that it could be applied to the bioconjugation of RNA, although RNA is more susceptible to degradation than DNA due to the 2' hydroxyl group on the ribose ring which could be problematic. However, it is only in the last few years that extensive research has begun to take place. Finn *et al.* (2009), reported a copper click reaction between RNA modified at the 5' end with an alkyne group and a coumarin azide in the presence of the water soluble ligand tris(3-hydroxypropyltriazolylmethyl)amine (THPTA).¹²⁷ The ligand was added to the reaction mixture in order to stabilise the Cu(I) species, the ligand along with other copper binding ligands were also found to accelerate the copper click reaction.^{127,128} Furthermore El-Sagheer and Brown (2010) clicked two short strands of RNA modified at the 5' and 3' ends with an azide and alkyne group respectively together in the presence of the THPTA ligand.¹⁰³ Likewise Parades *et al.* (2011) reported the click reaction between two strand of RNA modified at the 3' and 5' end, but this time, the modified bases were incorporated enzymatically into RNA transcripts.¹¹⁴

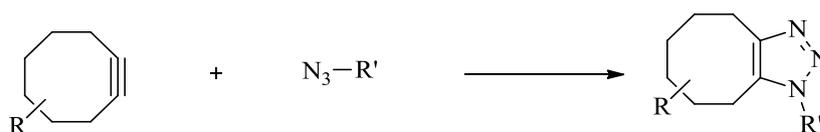
Copper-catalysed click chemistry is a viable route for bioconjugating two strands of RNA together. The azide and alkyne moieties can be incorporated readily into RNA at both the 5' and the 3' end and the reaction itself is highly efficient and high yielding. There is one main drawback to the copper click reaction: copper (I) is highly toxic to

cells and, although there are examples in the literature where copper click bioconjugations have been used *in vivo*, it is not the most appropriate method.^{129,130}

1.6.3.3.2 Copper-free click chemistry

The toxicity of copper on both bacterial and mammalian cells meant that the copper catalysed alkyne-azide cycloaddition was not desirable bioconjugation method for *in vivo* labelling. Bertozzi and colleagues began to look into alternative activated alkynes to overcome the need for copper and proceeded to develop the ring-strained cyclooctynes. They initially found that a simple cyclooctyne reacted with organic azides in physiological conditions and in the absence of auxiliary reagents generated the desired triazole product albeit slowly (Scheme 1.9). Importantly, the cyclooctyne was stable in mild acidic conditions and to prolonged exposure to biological nucleophiles, so it could circumvent the need for copper catalysis.¹³¹ However the cyclooctyne-azide ligation results in the formation of two regioisomers.

Scheme 1.9 Copper-free cyclooctyne-azide click reaction where R and R' are biomolecules and probes



Further research was undertaken into cyclooctynes, a difluoromethylene moiety was added which is an electron withdrawing group (Figure 1.14, DIFO, **(30)**), dibenzocyclooctynes (Figure 1.14, DIBO, **(31)**), where the aromatic rings contribute to the ring strain of the alkyne increasing the reactivity rate, and an amide incorporated into the ring of a dibenzocyclooctyne (Figure 1.14, BARAC, **(32)**) were synthesised and

reacted with cell surface azides. All were efficient in forming the triazole without the aid of copper with (32) giving a 10 fold higher signal after just 1 minute.¹³²⁻¹³⁴

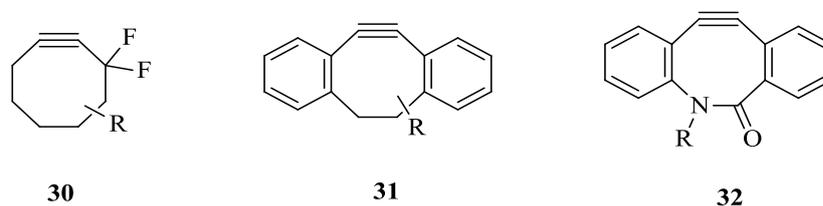


Figure 1.14 Cyclooctynes designed for efficient copper-free click reactions. R represents biomolecules and probes.

The development of the cyclooctynes presented another potential method in the bioconjugation of two strands of RNA together and in 2010 van Delf *et al.* synthesised a dibenzocyclooctyne phosphoramidite and incorporated it to the 5' end of a 16-mer oligoribonucleotide by solid phase synthesis and successfully conjugated it to an azido-peptide.¹³⁵ This reaction demonstrates the potential of copper-free click reactions for the bioconjugation of RNA. However, as the cyclooctynes are bulky they could significantly perturb splicing.

Of all the bioconjugation techniques discussed the copper click reaction was chosen as the most desirable method, as both the alkyne and azides are bio-orthogonal to each other dispelling any cross-reactivity that are associated with amide-NHS or thiol-maleimide coupling and unlike the Staudinger ligation the reactants are readily available and do not require inert atmospheres during conjugation. Furthermore the triazole group that is formed imparts minimal steric bulk relative to copper-free versions such as the bulky cyclooctynes. Plus you get a regiopure isomer which simplifies the analysis of the splicing assay. Despite the labile nature of the RNA, copper-catalysed click chemistry has been shown as a viable bioconjugation technique as long as a copper-stabilising ligand is present.^{103,127}

1.7 STRATEGY

In order to determine the mechanism by which ESEs exert their effects on splice sites methodology was required which enabled the insertion of a flexible non-RNA linker into the pre-mRNA transcript between the ESE and the splice site. It was decided to primarily focus on the ESEs which stimulated the 5' splice site of a mini-gene due to the techniques already developed by Jäschke *et al.* involving the incorporation of G-initiators which incorporate into the 5' end of a transcript (Section 1.6.1.1). The ESE would therefore be modified at the 3' end with an appropriate functional group. PEG linkers were chosen as the tether as it is hydrophilic, neutral in charge and cannot bind RNA binding proteins (Section 1.5.1).

Copper-catalysed alkyne-azide cycloaddition (Scheme 1.8) was chosen as the most desirable method for bioconjugation as the reaction is extremely rapid, bio-orthogonal and occurs at ambient temperatures in aqueous solutions. In addition the copper click reaction does not require bulky hydrophobic reactants that could significantly perturb splicing. The G-initiator was therefore designed with an alkyne group at the 5' end (**G1-G4**) and the ESE was designed with an azide group at the 5' end (**ESE1**).

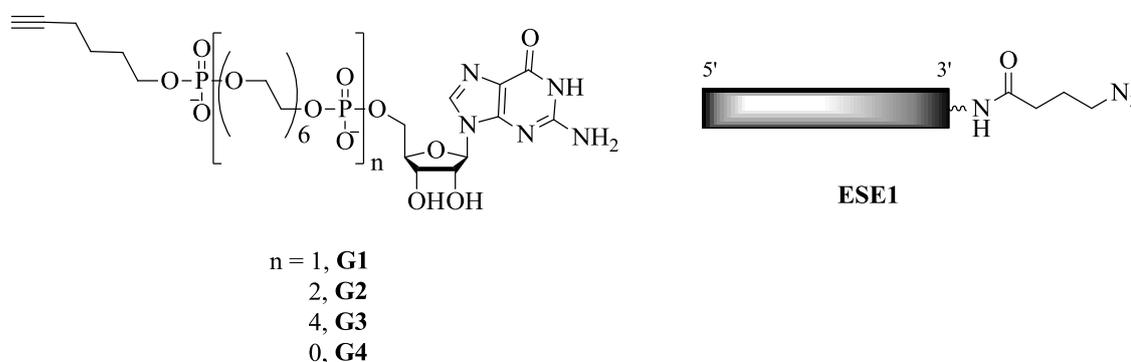


Figure 1.15 Diagrammatic representation of G-Initiators (G1-G4) and ESE1 synthesised in this thesis.

The thesis will be split into 3 major sections:

1. Development of a splicing assay whereby an ESE at the 5' end stimulates splicing of the 5' splice site. (Chapter 2).
2. Synthesis and incorporation of the G-initiators (**G1-G4**) into RNA transcripts and subsequent construction of the tripartite RNA systems. (Chapter 3).
3. Splicing assays of the tripartite structures; determining whether the splicing pattern is maintained with the addition of the linker between the 5' splice site and the ESE. (Chapter 4).

2 DEVELOPMENT OF SPLICING ASSAY WITH AN EXONIC SPLICING ENHANCER AT THE 5' END

2.1 INTRODUCTION

The simplest method for producing a pre-mRNA containing a flexible linker by click chemistry would be to attach an ESE and the linker to the 5' or 3' end of the substrate pre-mRNA. The first requirements, therefore, are to identify splicing reactions that are affected by a terminal ESE and to establish a method of attaching the linker and the ESE. It was decided not to look at the stimulation of splicing, as for a 3' ESE, but at the effect of the ESE on splice site selection, therefore targeting ESEs at the 5' end.

2.1.1 Splicing Assays to be Tested

This decision entailed finding a suitable assay for an ESE at the 5' end of transcript. Three different constructs were tested concurrently which all had different properties, but potentially could exhibit differences in splicing depending on the sequence adjacent to the 5' splice site:

- β -globin – 2 exon mini gene with a potential ESE in the first 20 nucleotides adjacent to the 5' splice site (Figure 2.1a).
- SRSF1 gene – exon 4 has an optional auto-regulated intron with an enhancer located between 24 to 40 nucleotide upstream of the 5' splice site (Figure 2.1b).
- Adeno 1 WW – a 2 exon mini gene that has 2 identical alternative 5' splice site that when spliced produces 2 mRNA isoforms: the shorter upstream mRNA and the longer downstream mRNA (Figure 2.1c). The insertion of an ESE upstream

of both upstream splice sites could potentially stimulate upstream splice site selection discussed in more detail later.

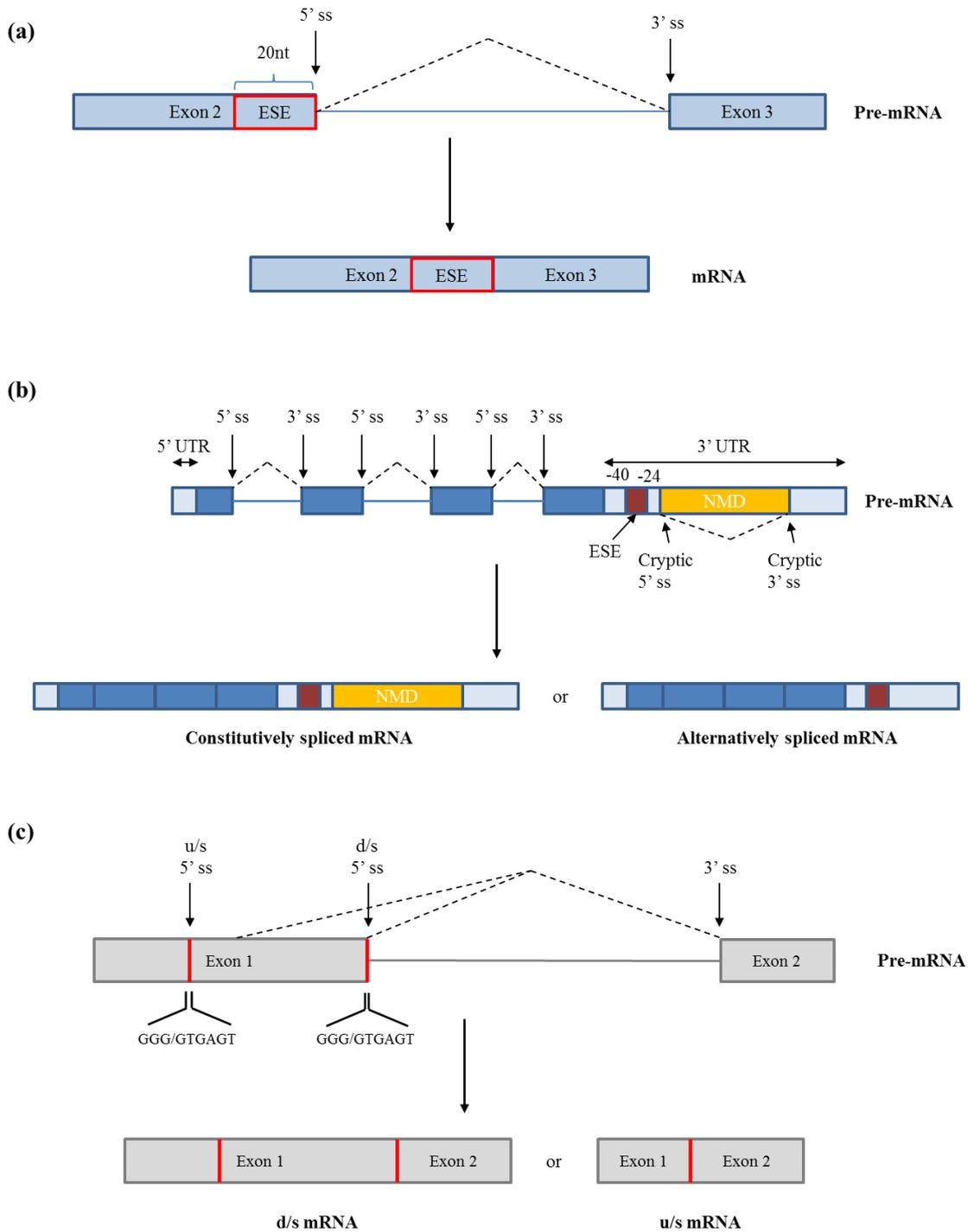


Figure 2.1 Diagrammatic representation of mini-genes to be tested. (a) β-globin gene (b) SFSR1 gene (c) Ad1WW gene.

2.2 RESULTS AND DISCUSSION

2.2.1 β -globin Construct

Within exon 2 of the β -globin mini-gene there is a potential ESE sequence in the first 20 nucleotide adjacent to the 5' splice site. In order to determine if the sequence was actually an ESE and activated splicing at the 5' splice site, the sequence needed to be replaced with a random sequence that had no known ESE activity itself. The hypothesis is that by replacing the natural sequence, splicing should at least be partially if not fully inhibited and therefore during splicing the amount of mRNA produced should have significantly reduced (Figure 2.2).

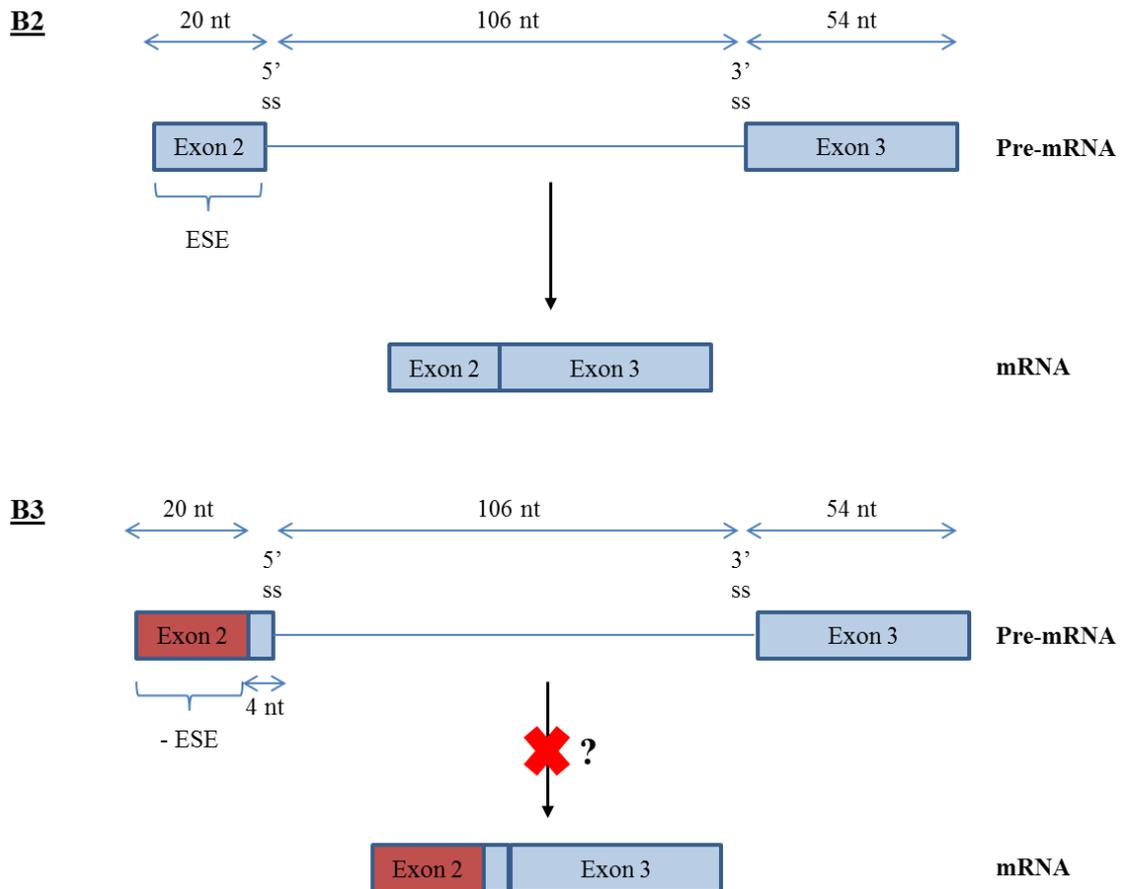


Figure 2.2 A diagrammatic representation of splicing of the β -globin mini-gene constructs designed, exons blue boxes and introns blue lines; B2 Natural pre-mRNA sequence of β -globin, exon 2 containing the potential ESE splicing to give mRNA as shown, B3 β -globin exon 2 replaced with a random sequence highlighted by red box potentially inhibiting splicing .

2.2.1.1 Construction of the model transcripts

In order to undertake this study, forward primers were designed so that the DNA constructs could be amplified with the relevant sequences at the 5' end. The first primer designed contained the last 20 nucleotide of exon 2 (Table 2.1, T7-ESE-BGE2-last 20), the second primer contained a random 20 nucleotide sequence, the last 4 nucleotides of exon 2 and the first 21 nucleotides of the intron (Table 2.1, T7-ESE-BGE2-20N-Intron). Both forward primers contained the T7 promoter region ready for transcription and the reverse primer (Table 2.1, P17) was the same for both constructs. The two different DNA constructs (**B2** and **B3**, Figure 2.2) were obtained by PCR.

The PCR fragments of **B2** and **B3** were transcribed into RNA using [α - 32 P] GTP to radiolabel the guanosines throughout the sequence in order to visualise splicing products by phosphor-imaging. The transcription of **B2** and **B3** produced single bands when visualised by polyacrylamide gel electrophoresis, the bands were excised from the gel and the RNA was eluted using RNA elution buffer overnight at 4°C, the RNA was then purified by ethanol precipitation and resuspended in water. The transcripts were then spliced; the transcripts were incubated in HeLa cell nuclear extract which contains a mix of SR proteins in the presence of 3.2 mM MgCl₂, 50 mM K-glutamate, 1.5 mM rATP and 20 mM creatine phosphate at 30°C. The splicing reaction was monitored over a 2 hour time course in order to determine splicing efficiency at various time points. Samples were taken at 0, 15, 30, 60, 90 and 120 minutes which were treated with proteinase K (PK) enzyme at 37°C for 30 minutes to remove the proteins. The splicing reactions were then purified by ethanol precipitation and analysed by gel electrophoresis (Figure 2.3).

In the splicing reaction of **B2** which has the potential ESE present in exon 2 adjacent to the 5' splice site pre-mRNA (P) is spliced into mRNA (m) over the 2 hour time course (Figure 2.3a), the amount of mRNA (m) produced increased over time. The intron lariats (L) were visible after just 15 minutes, indicating the splicing reaction had begun. The mRNA (m) was visible at the 30 minute time point (Figure 2.3a). The amount of RNA was quantitated for both the pre-mRNA and mRNA, the results showed that the most significant increase in mRNA production occurs between 30-60 minutes from with a 40% increase. After 120 minutes a total of 76% of the pre-mRNA had been converted into mRNA (Figure 2.3c). **B3** where the potential ESE was replaced with a random sequence was expected to demonstrate a sizable reduction in mRNA production during splicing compared to **B2**, however this was not observed over the 2 hour time course

(Figure 2.3b & c). Analysis of the gel showed the same splicing pattern as **B2**, at 0 minutes only pre-mRNA (P) was observed and as with **B2**, the intron lariats (L) were visible at the 15 minute time point indicating the splicing reaction had begun. At 30 minutes an additional mRNA (m) band is present (Figure 2.3b). The amount of mRNA produced increased over time with a significant increase in of 40% occurring between 30-60 minutes (Figure 2.3c).

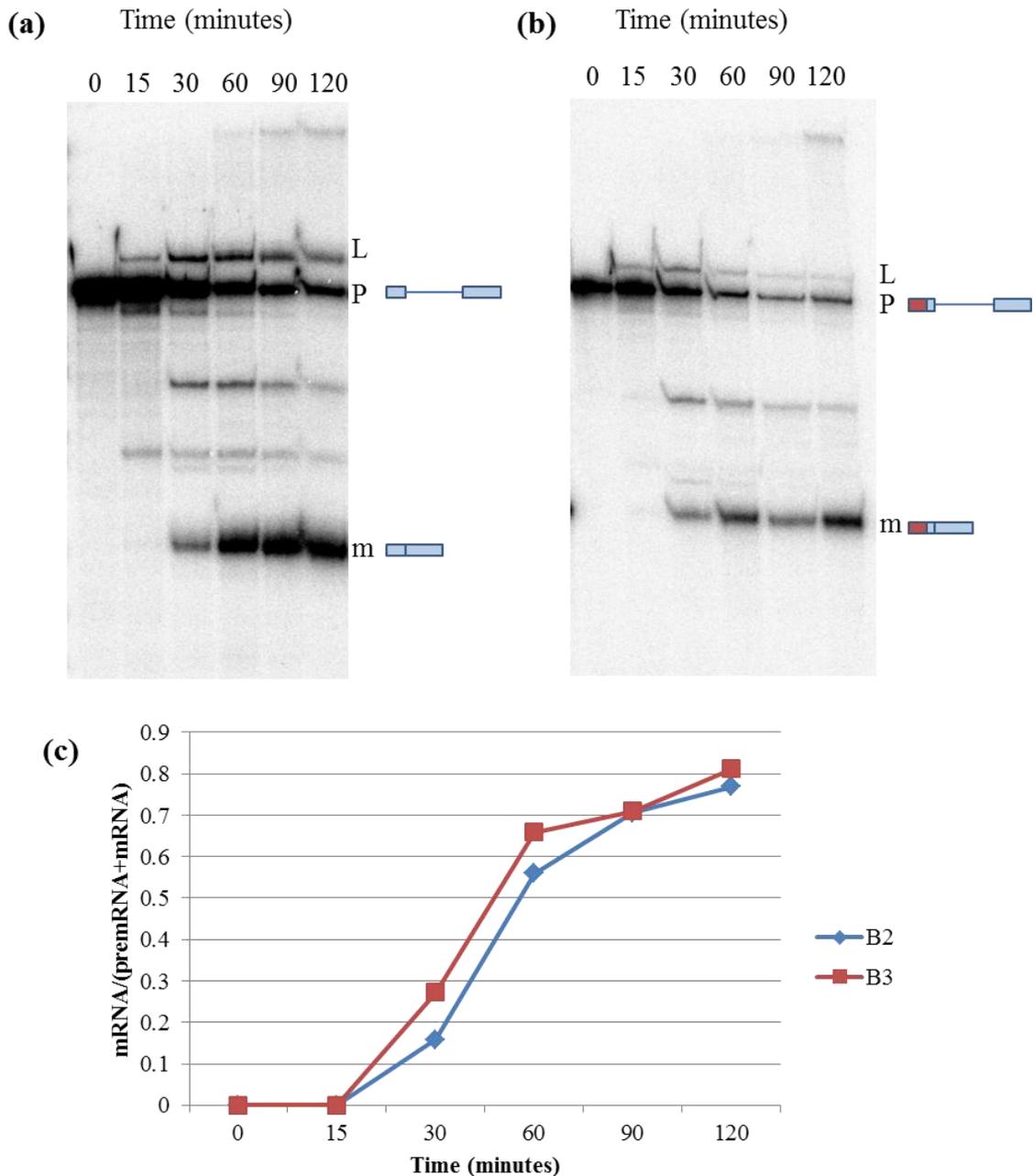


Figure 2.3 Splicing of B2 and B3 over a two hour time course; (a) An 8% denaturing polyacrylamide gel for the splicing of B2, amount of mRNA produced increasing over time; L shows lariat by products of splicing, P is unspliced pre-mRNA, m is spliced mRNA (b) An 8% denaturing polyacrylamide gel for the splicing of B3 labelled as of (a), mRNA increasing over time.; (c) Proportion of mRNA produced over the time course for B2 & B3.

These results do not support the presence of the ESE. Designing sequences which have no enhancing properties is extremely difficult. ESEs were initially characterised by purine rich sequences but after further analysis additional AC-rich motifs and pyrimidine rich motifs emerged as ESEs; therefore there was no guarantee that the

sequence that was designed to replace the natural sequence did not exhibit enhancing elements in its own right.^{74,136,137} To conclude either way a range of constructs with varying sequences adjacent to the 5' splice site in exon 2 would need to be tested, to compare splicing efficiency.

The similarity in splicing efficiency of both **B2** and **B3** constructs, however, meant that this assay was not going to be suitable to determine how ESEs exert their effect in promoting splicing.

2.2.2 SRSF1 Gene

The SRSF1 gene is found on chromosome 17 and codes for SRSF1 which is an essential sequence specific splicing factor involved in the splicing of pre-mRNA. The 3' UTR (untranslated region) of the gene located in exon 4 contains an alternative intron that is normally retained in mature RNA. It is spliced in response to high levels of SRSF1 and triggers nonsense mediated mRNA decay (NMD) (Figure 2.4a).¹³⁸ NMD detects and degrades transcripts that contain a termination codon more than fifty nucleotides upstream of a spliced intron.¹³⁹ A cryptic intron is a region within a conserved exon that contains weak cryptic splice sites that is recognised by the spliceosome enabling it to be alternatively spliced (Figure 1,2 E). The alternative splicing of the cryptic intron of the SRSF1 gene produces the RNA transcript and NMD.⁴⁴

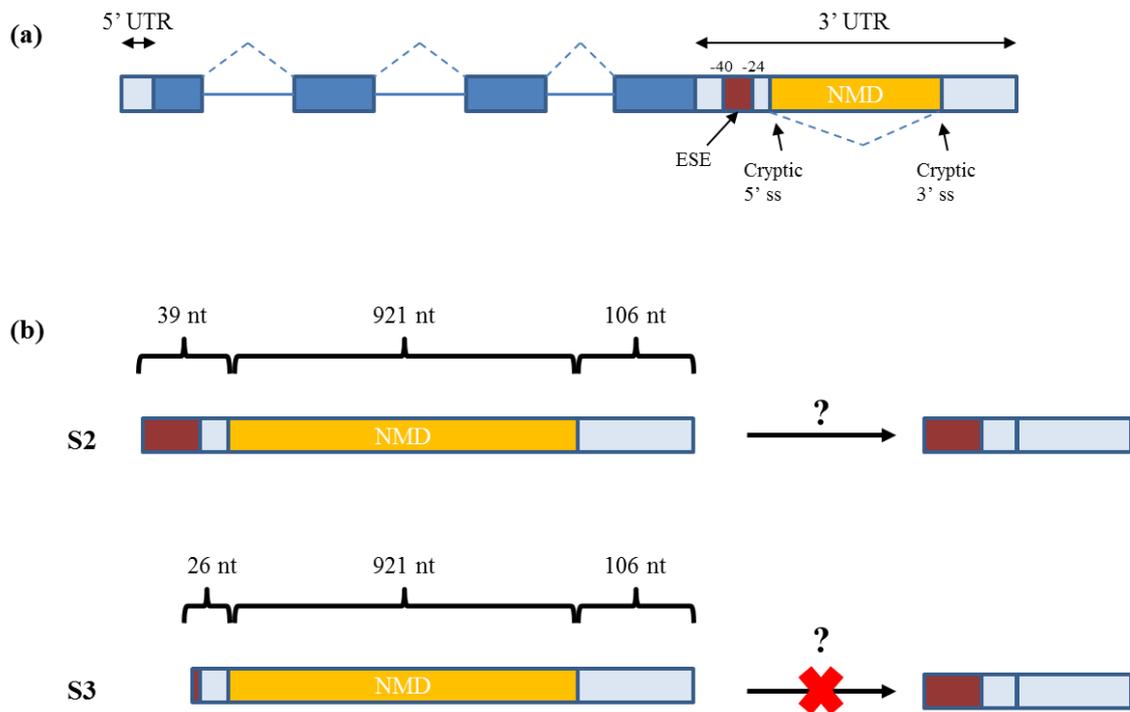


Figure 2.4 A diagrammatic representation of the SRSF1 gene and its cryptic intron in exon 4; (a) SRSF1 gene with its splicing patterns represented by dashed lines, exons: blue boxes, intron: blue lines, ESE: red box, NMD region/cryptic intron: yellow box; (b) S2: Exon 4 with the cryptic intron with the potential ESE sequence splicing to produce mRNA, S3: Exon 4 with the cryptic intron without the potential ESE sequence which should not splice.

The cryptic intron provides a potential assay for splicing at the weak cryptic 5' splice site as there is a possible enhancer located between -24 to -40 nucleotides upstream of the cryptic 5' splice site (Figure 2.4a).⁴⁴ The primers were designed to include (Table 2.1, T7-SF2-40) and exclude (Table 2.1, T7-SF2-24) this region in exon 4, with a common primer at the 3' end of the cryptic intron (Table 2.1, SF2-R). The hypothesis as with the β -globin assay was that the cryptic intron will be spliced out of the construct containing the potential ESE in the -24 to -40 nt region (Figure 2.4b, **S2**) and will not splice for the one without the -24 to -40 region (Figure 2.4b, **S3**).

Splicing of **S2** was carried out initially without the addition of SRSF1 over a 2 hour time course (Figure 2.5a), resulting in no splicing products being formed. In order to promote splicing SRSF1 was subsequently added into the splicing reaction and

monitored over a 2 hour time course, producing exactly the same results, the cryptic intron had not been spliced out and only the pre-mRNA was present (Figure 2.5b).

The lack of splicing of **S2** could be a result of a number of factors; the cryptic splice site is intrinsically a weak splice site and therefore will not splice readily, the concentration of the components of the splice site may not have been optimum for splicing to occur (for example the concentration of Mg^{2+} can have a significant impact as either too low or too high concentrations can decrease splicing efficiency, moreover the optimal concentration differs depending on the system).¹⁴⁰ Another factor may have been due to the presence of the Sam68 splicing factor in the nuclear extract. Sam68 has been shown to inhibit the alternative splicing of the 3'UTR in which the NMD is retained, therefore splicing of this region is inhibited.⁴⁴ Furthermore, even though a cryptic intron splices *in vivo*, it does not necessarily mean it can be reciprocated *in vitro*. As a result the SRSF1 gene was not a plausible assay to insert a linker between the ESE and the 5' splice site.

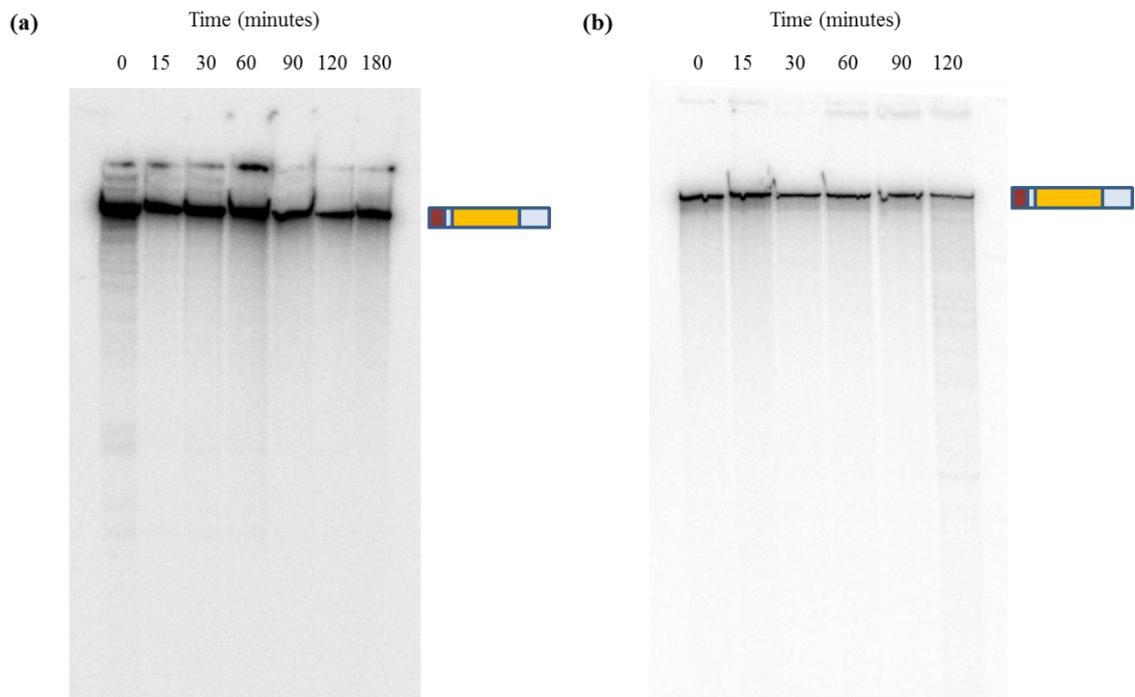


Figure 2.5 Splicing reaction of S2 over a 2 hour time course; (a) without addition of SRSF1; (b) with the addition of SRSF1.

2.2.3 Adeno WW Virus

An adenovirus-based construct containing two alternative 5' splice sites in exon 1 of a two exon mini-gene was also investigated to determine whether splice site selection could be manipulated with the addition of a known ESE upstream of the alternative 5' splice sites. The hypothesis was that an ESE inserted upstream of the upstream 5' splice site would shift splicing from the downstream 5' splice site to the upstream (nearest) 5' splice site (Figure 2.6).

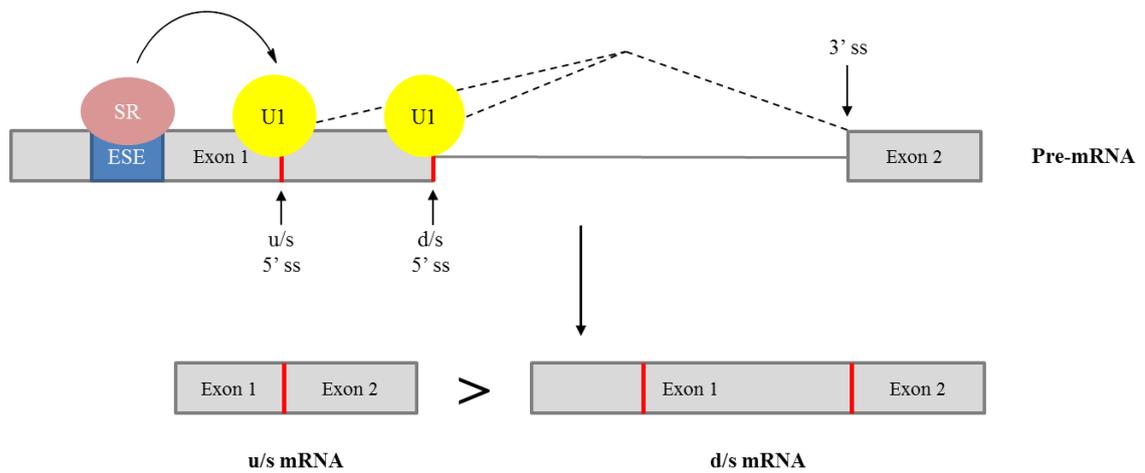


Figure 2.6 A diagrammatic representation of the hypothetical splicing patterns of the Ad1WW gene when an ESE is inserted upstream of the 5' splice sites. The SR protein bound to the ESE stimulates splicing predominantly at the upstream splice site.

The sequence of the two alternative 5' splice sites of the Ad1WW construct are identical (GGG/GTGAAGT) separated by 93 nt and splices to produce two mRNA isoforms (Figure 2.7a).¹⁴¹ The sequence used as the enhancer comprised a series of repeats based on 5'GGA/ This sequence has been shown in previous work from this laboratory to act as an enhancer.^{36,142}

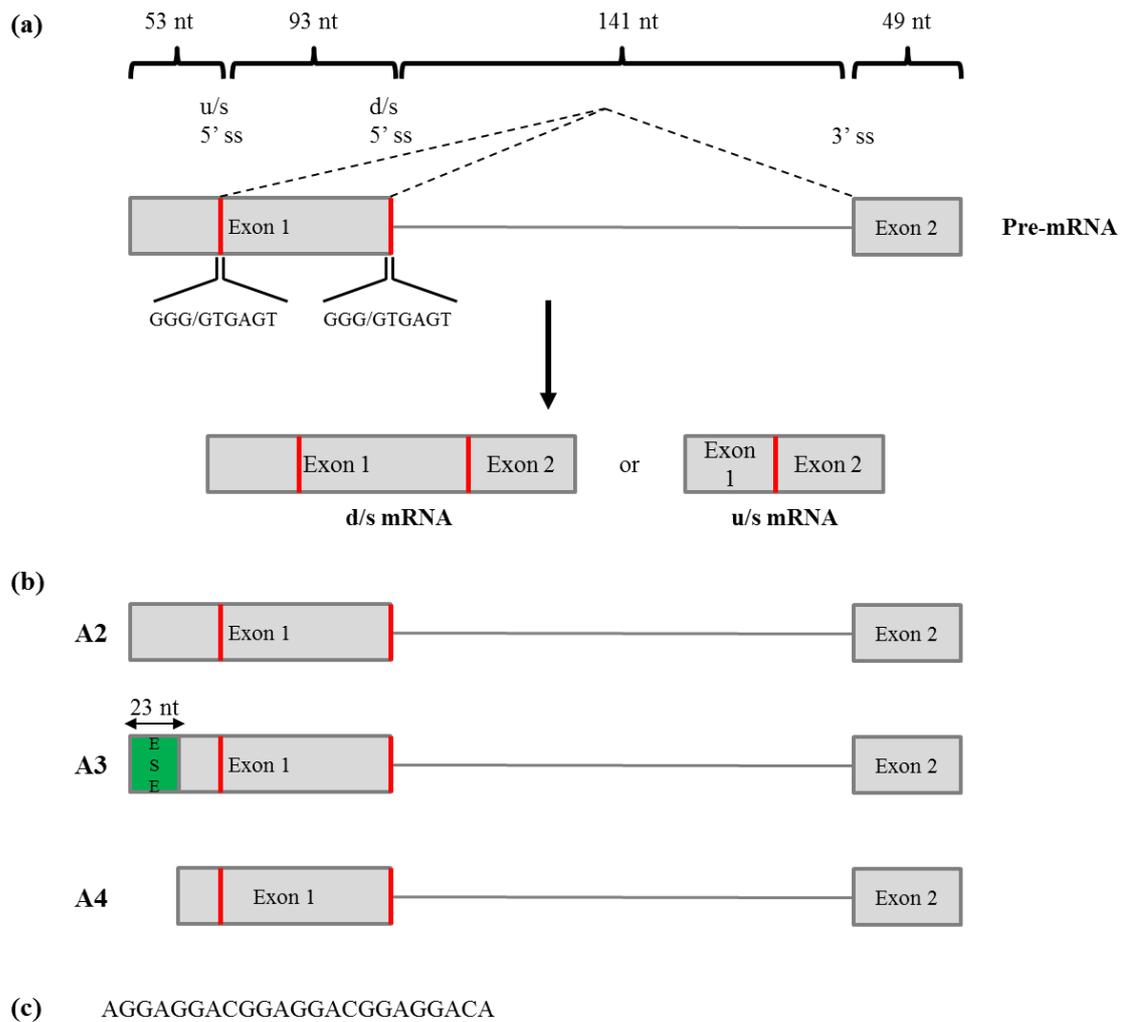


Figure 2.7 A diagrammatic representation of Ad1WW¹⁴¹ with exons (grey boxes), intron (grey lines); (a) Splicing of pre-mRNA at the alternative 5' splice sites (red lines) producing 2 mRNA isoforms; (b) A2: Natural sequence of Ad1WW 336 nt in length, A3: Ad1WW with ESE (green box) replacing natural sequence 337 nt in length, A4: Ad1WW 314 nt in length; (c) Sequence of ESE.

The natural pre-mRNA sequence (Figure 2.7b, A2) 30 nt upstream of the upstream 5' splice site was modified with the 5'GGA ESE (Figure 2.7b, A3). Transcription of the PCR products was carried out radiolabelling the guanosines in the transcript with [α -³²P] GTP. The transcripts were purified by gel extraction ready for splicing. The splicing assays were monitored over a 2 hour time course with time points being taken at 0, 15, 30, 60, 90 and 120 minutes and once treated with proteinase K (PK) enzyme which is used to digest the proteins from the nuclear extract, the splicing products were analysed by gel electrophoresis and phosphor-imaged. The splicing of A2 produced the

two mRNA isoforms (Figure 2.8a) after 30 minutes with the percentage of total mRNA increasing over time. At 0 and 15 minutes only pre-mRNA was visible, at 30 minutes intron lariats were observed (Figure 2.8a). Analysis of the gel showed that at 30 minutes 98% of the RNA present was still pre-mRNA with 2% accounting for upstream mRNA (Figure 2.8b). At 60 minutes the splicing of pre-mRNA produced both the upstream and downstream mRNA and after 120 minutes 41% of the pre-mRNA had been converted into upstream mRNA and 14% into downstream mRNA (Figure 2.8b). The analysis of splicing of **A3** showed the pre-mRNA spliced at the upstream splice site only with a band only being observed for upstream mRNA isoform (Figure 2.8b & 2.8c). Further analysis of the splicing products demonstrated that splicing was overall more efficient with 66% of the total pre-mRNA being spliced compared with 54% for **A2** (Figure 2.8a & 2.8b).

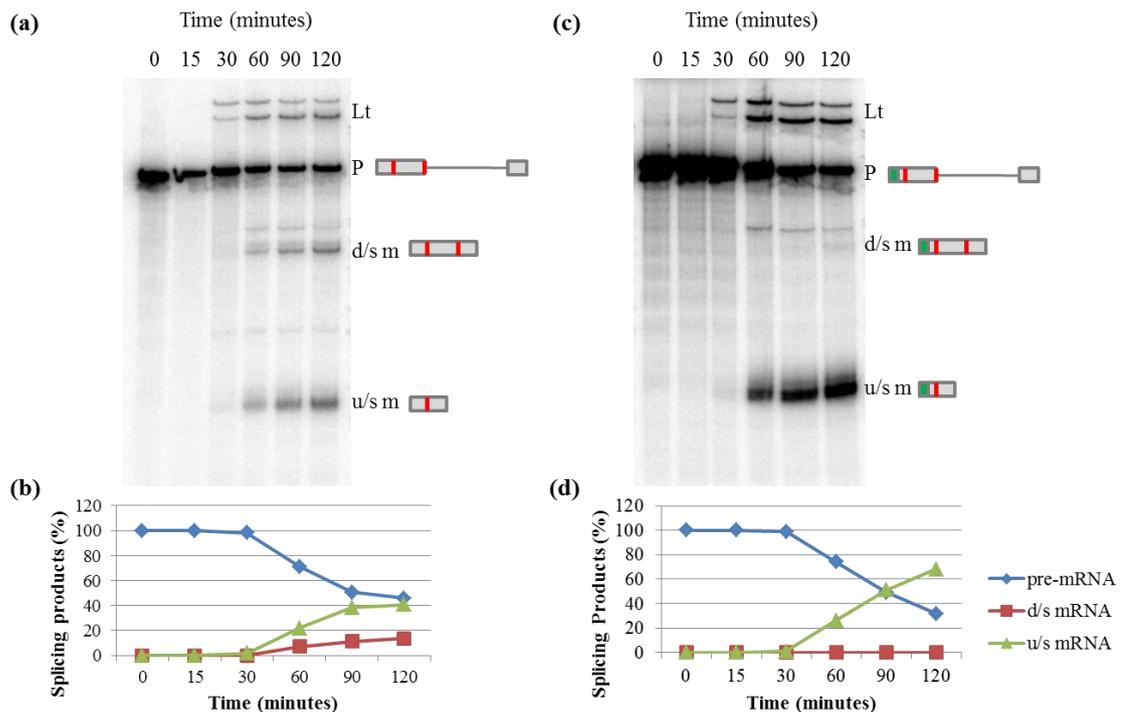


Figure 2.8 Splicing of A2 and A3 over a two hour time course; (a) A2, 8% polyacrylamide gel with pre-mRNA, d/s and u/s mRNA, Lt shows lariats by products of splicing, P is unspliced pre-mRNA, d/s m shows downstream mRNA spliced at 5' downstream splice site, u/s m shows upstream mRNA spliced at 5' upstream splice site (b) A2 Corresponding histogram showing the percentage of the spliced products over the 2 hour time course. (C) A3, as of (a), (d) A3 as of (b).

The results obtained in this splicing reaction indicated that the ESE had an effect on splice site selection, with the upstream splice site being selected over the downstream site. However, an additional control was needed to determine if it was the ESE having the desired effect or if it was due to the loss of the natural sequence. A further construct was designed so that the transcript began where the ESE was added (Figure 2.7b, **A4**) using primers T7-Ad-CAT & Ad1 Trx R (Table 2.1). **A2**, **A3** and **A4** were all transcribed radiolabelling with [α - 32 P] GTP ready for splicing. The splicing reactions were monitored over a 2 hour time course with time points taken at 0, 15, 30, 60, 90 and 120 minutes. When analysed by gel **A2** and **A3** spliced as expected with downstream & upstream mRNA being observed for **A2** (Figure 2.9a) with a total mRNA of 49% of which 46% arose from upstream mRNA with 3% from downstream mRNA and only upstream mRNA (66%) produced for **A3** (Figure 2.9b). Splicing of **A4** followed the

same pattern as **A2** with both the downstream and upstream mRNA isoforms being produced (Figure 2.9c). Analysis of the bands showed that after 2 hours a total of 48% of pre-mRNA was spliced with 45% arising from the upstream splice site and 3% from the downstream splice site (Figure 2.9c).

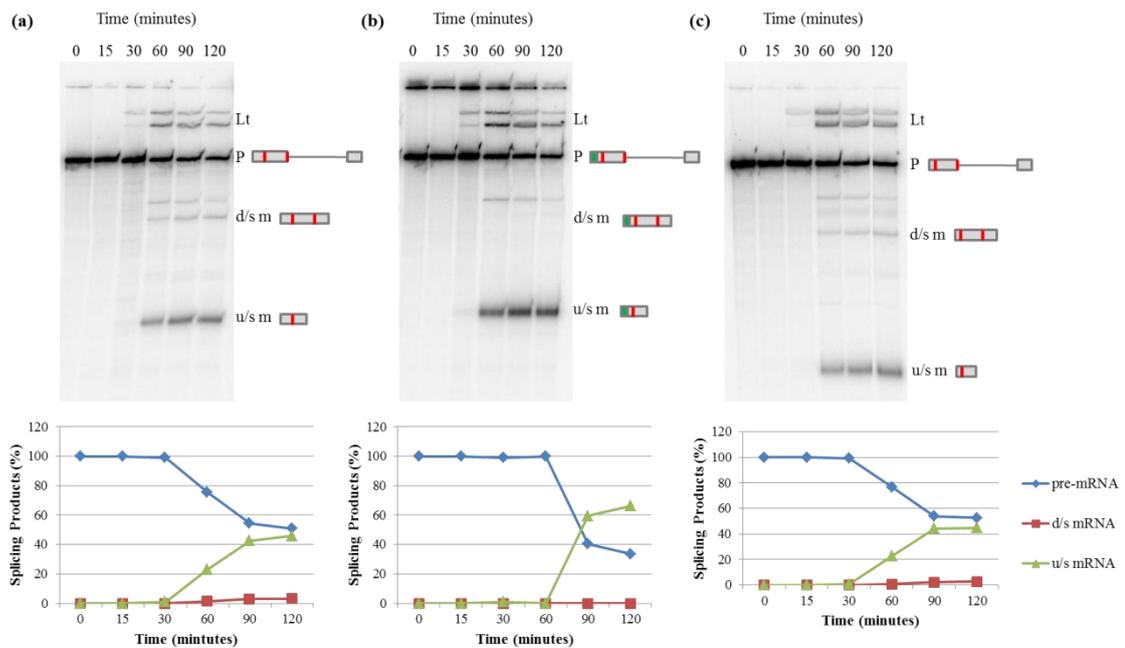


Figure 2.9 Splicing of A2, A3 and A4 over a two hour time course; (a) A2, 8% polyacrylamide gel with pre-mRNA, d/s and u/s mRNA highlighted percentages of splicing shown in line graph below, Lt shows lariat by product of splicing, P shows pre-mRNA, d/s m shows downstream mRNA spliced at downstream 5' splice site, u/s m shows upstream mRNA spliced at upstream 5' splice site. (b) A3, as for A2 ; (c) A4 as for A2

The results from these three splicing assays have demonstrated that the addition of an ESE upstream of the upstream splice site has an impact on splice site selection; in particular the ESE enhances splicing at the upstream 5' splice site. As a result it would be an ideal assay to use to determine whether enhancers exert their effect in a 3-dimensional manner in relation to SR proteins; a flexible linker could feasibly be inserted between the ESE and the main transcript.

2.3 SUMMARY

The assays with the various mini-genes developed and tested within this chapter have exhibited differing results. The β -globin 2 exon mini-gene with a potential ESE adjacent to the 5' splice site was investigated by replacing the potential ESE sequence with a random sequence, and comparing the splicing efficiency. However, when splicing was analysed the two constructs (**B2** & **B3**) spliced in virtually the same manner and therefore was not going to be a suitable assay in the investigation of ESE's and splice site selection (Section 2.2.1).

The SRSF1 gene has a cryptic intron in the 3' UTR on exon 4 and a potential enhancer between -24 to -40 nt upstream of the cryptic 5' splice site. Two constructs were developed which included and excluded this region (**S2** & **S3**, Section 2.2.2). A splicing reaction for **S2** showed that no spliced products were produced. Unfortunately the assay was ill-fated from the start as PCR amplification of **S3** was fruitless as no amplification of the DNA occurred which could be due to the lack of optimisation of different factors such as the concentration of $MgCl_2$ which chelates with dNTPs and therefore a low concentration of $MgCl_2$ can result into a decrease in free Mg^{2+} which facilitate the DNA polymerase.^{143,144}

The Ad1WW construct (Section 2.2.3) on the other hand proved to be an appropriate assay to progress with. The two alternative 5' splice sites in exon 1 gave rise to two mRNA isoforms being produced during splicing, the addition of the ESE upstream of the upstream 5' splice site (**A3**) enhanced splicing at the upstream splice site with splicing being shifted from the downstream splice site to the upstream splice site compared with the natural sequence (**A2**). The enhanced splicing at the upstream 5'

splice site with the addition of the ESE provides a suitable *in vitro* assay to test the influence of inserting a flexible non-RNA linker such as PEG between the main transcript and the ESE to determine the effect on splicing at the two alternative 5' splice sites. Inserting the linker at this point will establish whether the enhancer exerts the same effect on splicing and if upstream splicing is enhanced it would infer that the ESE exerts its effects by 3-dimensional diffusion so that the SR protein that binds to the ESE loops around interacting with U1 snRNP at the upstream 5' splice site. As a result the Ad1WW assays can be and will be used to investigate the mechanism of the ESE.

2.4 EXPERIMENTAL

2.4.1 Preparation of PCR Fragments for Adenovirus and Beta-Globin Constructs

PCR were carried out using G-Storm Thermal Cycler as in accordance to the manufacturer's guidelines. The Adenovirus (Ad1 WW) constructs **A1**, **A2**, **A3** and **A4** were PCR-amplified with reverse primer Ad1 Trx R and the forward primers As1 Trx F T7-Ad-50, T7-GGA-Ad-30 and the T7-Ad-CAT respectively (Table 2.1.). The β -globin constructs **B1**, **B2** and **B3** were amplified using the reverse primer P17 with the forward primers P16, T7-ESE-BGE2-last 20 and T7-ESE-BGE2-20N-intron respectively (Table 2.1). 10-50 μ l Reactions were prepared using either Go-Taq (Promega) or Phusion™ Hot Start (New England Biolabs) polymerases as per DNA polymerase's manufacturer's instructions.^{145,146} To ensure specific amplification, the PCR amplicons were analysed on 2% agarose gels and subsequently purified by phenol-chloroform extraction and Ethanol precipitation ready for transcription.¹⁴⁷

2.4.2 PCR Amplification Optimisation of Genomic DNA-SF2/ASF cryptic Intron

PCR (10 µl/temperature) were carried out using a temperature gradient from 54-62°C as per the manufacturer's instructions. The genomic-DNA (SF2/ASF cryptic intron region) was amplified using the oligonucleotides T7-SF2-40 with SF2-R and SF2-24 with SF2-R respectively (Table 2.1). PCR amplification was performed using GoTaq (Promega) according to the manufacturer's instructions.¹⁴⁵ The PCR amplicons were analysed on a 1% Agarose gel and further reactions were carried out using the optimum temperature.

Table 2.1 Oligonucleotide Sequences for PCR amplification of Ad1WW, β-globin and SRSF1 constructs. Sequences underlined are the T7 promoter region.

Name	Sequence
Ad1 Trx F	TAATACGACTCACTATAGGGAACAAAAGCTTGTATGCCT
T7-Ad-50	<u>AAATTAATACGACTCACTATAGGGTTCGTCCTCACTCTC</u> TTCCGCAT
T7-GGA-Ad-30	<u>AAATTAATACGACTCACTATAGGGAGGAGGACGGAGGA</u> CGGAGGACATCGCTGTCTGCGAGGGCCA
T7-Ad-CAT	<u>AAATTAATACGACTCACTATAGGCATCGCTGTCTGCGAG</u> GGCCAGCTGTT
Ad1 Trx R	GATCCAAGAGTACTGGAAAGACCGCGA
P16	<u>AAATTAATACGACTCACTATAGGGCTGCTGGTTGTCTAC</u> CCA

T7-ESE-BGE2- last 20	<u>AAATTAATACGACTCACTATAGGGATCCTGAGAACTTC</u> AGGGTGAT
T7-ESE-BGE2- 20N-intron	<u>AAATTAATACGACTCACTATAGGGATCCTGAGAACTTC</u> AGGGTGAT
P17	AACTTACCTGCCAAAATGATGAGACA
T7-SF2-40	<u>AAATTAATACGACTCACTATAGGGAGGATTGAGGAGGA</u> TCAGATCAATAA
T7-SF2-24	<u>AAATTAATACGACTCACTATAGGGAATCAGATCAATAA</u> TGGAGGCAATGGTA
SF2-R	CCATGAATCCTGGTAATTCATCCT

2.4.3 Radioactive Transcription of Amplified PCR products

All radioactive transcripts were generated using T7 RNA polymerase and labelled with ^{32}P in 10 μl reactions containing 1x T7 hot transcription buffer, NTP's (0.05 mM GTP, 0.5 mM ATP, 0.5 mM CTP and 0.5 mM UTP; Promega), diguanosine triphosphate (0.22 U, Roche), DTT (10 mM), [α - ^{32}P] GTP (25-50 μCi), PCR product (25-150 ng/ μl), T7 polymerase (5% v/v) and RNasin® (Promega)/RNaseout™ (Invitrogen) (5% v/v). The reaction mixture was incubated at 37°C for 1-2 hour and then analysed on a 6% denaturing polyacrylamide gel. The gel was exposed to x-ray film and the band of RNA was cut out from the gel RNA elution buffer (300 μl) was added to the excised band of RNA and left eluting overnight at 4°C. The eluent was removed from the gel and ethanol precipitated and resuspended in water ready for use.

2.4.4 Splicing Assays¹⁴⁸

Splicing reactions were initially performed using MgCl₂ (2 mM), KCl (3.5 mM), rATP (2 mM), phosphocreatine (10 mM), radiolabelled transcript and HeLa cell Nuclear extract (24.5 µl) made up to a final volume of 70 µl. The reactions were incubated at 30°C between 2-3 hours depending on time course. Time points were taken at 0, 15, 30, 60, 90, 120 and 180 minutes and stored in 1x PK buffer (40 µl) and stored on ice. After the last time point was taken PK enzyme (2 µl) was added to each time point and incubated at 37°C for 45 minutes. The time points were then ethanol precipitated, resuspended in formamide dyes and run on 6-10% denaturing polyacrylamide gels. (Volumes were adjusted according to final volume required).

2.4.5 Synthesis of Radioactive Ladder¹⁴⁷

The radioactive ladder was prepared by digesting the plasmid PBR322 (0.5 µg/mg) with the enzyme HPAII incubating at 37°C for 30 minutes. The digested plasmid was then treated with antarctic phosphatase for 15 minutes at 37°C. The phosphatase was then deactivated by incubating for 15 minutes at 65°C. A T4 kinase reaction comprised of 1x T4 Kinase buffer, 2% v/v T4 Kinase and [γ -³²P] ATP was added to 10% of total digest mixture, mixed and incubated at 37°C for 30 minutes. The free nucleotides were removed by passing through a GE healthcare S-300 column. The mixture was diluted with formamide dyes and denatured for 2-5 minutes at 80°C and then loaded onto a denaturing polyacrylamide gel. The gel was dried and exposed to a Phosphor screen for analysis.

2.4.6 Phenol Chloroform Extraction¹⁴⁷

The PCR amplicons were made up to 100 μ l with RNase Free H₂O, 200-300 μ l phenol/chloroform/isoamyl alcohol 25:24:1 solution (Sigma) was added, vortexed and centrifuged for 5 minutes at 13000 rpm. The organic layer was removed and the process was repeated.

2.4.7 Ethanol Precipitation

To the 100 μ l of DNA/RNA 3x volume of absolute ethanol was added along with one-tenth the volume of sodium acetate (for DNA). The mixture was mixed gently and the centrifuged for 15-30 minutes at 130000 rpm. The supernatant was removed and the DNA/RNA was resuspended in 200 μ l 70% ethanol, mixed and centrifuged for 10-30 minutes at 130000 rpm. The supernatant was removed, dried and resuspended in H₂O.

2.4.8 Buffers

- 1 x T7 Hot Transcription Buffer: 20 mM Tris-HCl pH7.5, 3 mM MgCl₂, 1 mM spermidine-HCl, 5 mM NaCl.
- RNA elution buffer: 1 mM EDTA, 0.2% w/v SDS, 0.5M sodium acetate pH4.0.
- PK buffer: 100 mM Tris-HCl pH7.5, 12.5 mM EDTA, 150 mM NaCl, 1% w/v SDS.
- 1 x TBE: 89 mM Tris base, 89 mM boric acid, 2 mM EDTA.
- T4 Kinase buffer : 70 mM Tris-HCl pH 7.6, 10 mM MgCl₂, 5 mM dithiothreitol.
- Formamide dyes : 90% v/v formamide, 50 mM EDTA, bromophenol blue, xylene cyanol.

3 DEVELOPMENT OF INITIATORS TO INCORPORATE INTO RNA TRANSCRIPTS FOR THE CONSTRUCTION OF TRIPARTITE RNA SYSTEMS

3.1 INTRODUCTION

A splicing assay was developed in Chapter 2 which demonstrated the importance of the presence of an ESE in 5' splice site selection. The Ad1WW contained two alternative 5' splice sites separated by 93 nucleotides in exon 1 of the 2 exon mini-gene (Figure 2.7a). Splicing of the natural **A2** and **A4** (Figure 2.7b) resulted in the production of the downstream and upstream mRNA isoforms (Figure 2.9). It was proposed that by inserting an ESE adjacent to the upstream 5' splice site, the ESE would stimulate splicing activity to the proximal upstream 5' splice site. Previous studies have demonstrated the importance of ESEs on 5' splice site selection. For example Humphery *et al.* (1995) showed a 32 nucleotide splicing enhancer regulated the usage of competing 5' splice sites, the ESE was situated between the upstream and downstream splice site and stimulated the upstream splice site but only in the presence of the downstream splice site.¹⁴⁹ Hence, an artificial enhancer sequence replaced the natural sequence upstream of the 5' splice site of the Ad1WW construct (Figure 2.7b, **A3**). The splicing of **A3** resulted in the upstream 5' splice site being utilised solely compared with both splice sites for the natural **A2** and **A4** constructs (Figure 2.9). The artificial enhancer was shown to stimulate the upstream 5' splice site shifting splicing from the downstream splice site to the upstream splice site. As a result this assay was ideal to develop a tool to investigate the mechanism of ESEs.

Incorporation of a flexible non-RNA linker between two sites of RNA has previously been demonstrated by Pasman *et al.* in 1996 when they inserted a PEG linker into an intron of the pre-mRNA, using oligodeoxynucleotide-directed ligations. Splicing still occurred, with the presence of the PEG linker supporting a 3-dimensional diffusion model whereby the RNA loops around enabling interaction of proteins located at various points across the RNA (Figure 1.7a).² However, the insertion of DNA into the pre-mRNA would compromise the interpretation of the results if the actions of the ESE were inhibited, since the cells contain a number of DNA-binding proteins. These DNA binding proteins could bind to the DNA and in turn smother the enhancer sequence preventing the association of SR proteins resulting in the inhibition of splicing. To overcome this limitation it was decided to develop a novel method in which a G-initiator is incorporated into an RNA transcript at the 5' and in turn bioconjugated to an ESE modified at the 3' end (Figure 3.1).

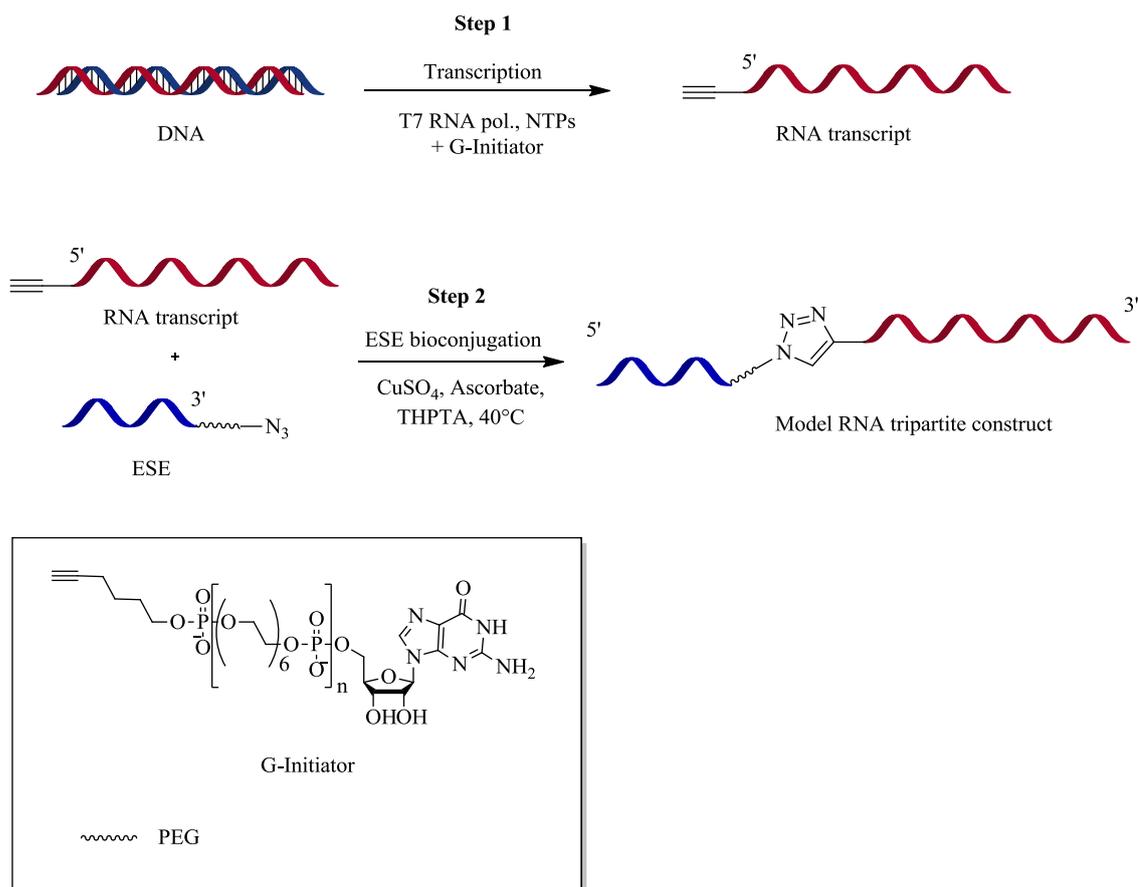


Figure 3.1 Diagrammatic representation of the incorporation of G-initiators into RNA (step 1) and the subsequent bioconjugation reaction (step 2).

3.1.1 Aim of Chapter 3

The objective of this chapter was to develop methodology to construct RNA tripartite structures. The assay developed in Chapter 2 (Section 2.2.3) was based on the 5' splice site with the ESE located upstream of the 5' splice site (Figure 2.7b, **A3**). In order to insert a linker between the ESE and the RNA transcript, the ESE will need to be modified at the 3' end and the transcript will need to be modified at the 5' end with the relevant modifications for bioconjugation.

The objective will be broken up into the following subaims:

- Design and synthesis of G-initiators alkyne-modified at the 5' end and enzymatically incorporate into RNA during transcription.
- Synthesis of azide-modified ESEs modified at the 3' end with linkers of varying lengths using solid phase synthesis methods.
- Bioconjugation of the 5' modified transcript with the modified ESE to produce RNA tripartite structures.

3.1.2 G-Initiators

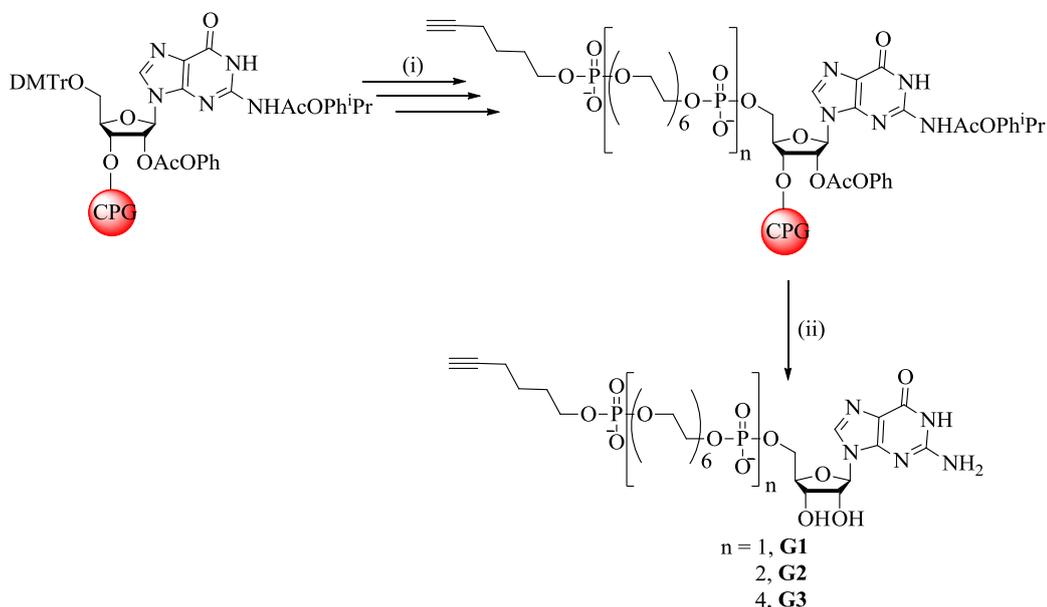
It has been previously shown that various G-initiators can be incorporated into RNA enzymatically using T7 polymerase-mediated transcription.^{80,86,88,90–93,103,113,150} These initiators have been modified with various functional groups ranging from an anthracene ring coupled via a PEG linker to a 5' monophosphate guanosine (**2**) to 5' amino (**6**) and 5' sulfhydryl (**5**) modified guanosines (Figure 1.11).^{88,91,93,150} The initiators can function as both a fluorescent tag and as a reactive moiety for thiol-maleimide and amine-NHS couplings. It was decided however, to develop an initiator which incorporated an alkyne at the 5' terminus which could then be reacted to a 3' azido-modified ESE using the copper catalysed Huisgen [3+2] cycloaddition (click chemistry) to form the triazole linkage between the ESE and the transcript (Figure 3.1, step 2). Click chemistry was chosen as the most desirable bioconjugation technique as both the alkyne and azide functional groups are bio-orthogonal to each other.¹²¹ Other bio-orthogonal bioconjugation reactions are available such as copper free click chemistry, but, these require bulky and hydrophobic reactants that could significantly perturb splicing.

3.2 RESULTS AND DISCUSSION

3.2.1 Synthesis of G-Initiators (G1-G3)

G-Initiators **G1-G3** were synthesised by standard solid phase synthesis using phosphoramidite coupling (Section 1.6.2). In order to incorporate a non-RNA linker into the tripartite system it was decided to insert a PEG_n linker into the G-initiator (**G1-G3**) between the monophosphate guanosine and the alkyne phosphoramidite (Scheme 3.1). Previous studies by Jäschke *et al.* have shown that the length of the PEG had little impact on incorporation into a transcript if it was less than PEG₁₄. Therefore due to commercial availability a hexaethylene glycol (HEG) was subsequently used.⁸⁸

Scheme 3.1 Synthesis of G-Initiators (G1-G3) using the iPr-Pac-G RNA SynBase™ CPG solid support: (i) detritylation, coupling of unincorporated phosphoramidite and oxidation of newly coupled phosphoramidite. (ii) Deprotection and cleavage with 35% NH₄OH, 50°C, overnight.



It was decided to initially synthesise three different G-initiators (**G1-G3**) which incorporated 1, 2 and 4 HEG linkers. A 0.2 μmol iPr-Pac-G RNA SynBase™ CPG 1000/110 column was used in the synthesis of all three initiators. After solid phase

synthesis the initiators were cleaved from the resin and purified by RP-HPLC and analysed by MALDI. The absorbance at 260 nm was recorded to determine the number of moles obtained for each (Table 3.1). G-initiators (**G1-G3**) when analysed by MALDI showed a mass which was 2 hydrogens less than the expected mass in the negative spectrum. This loss in mass could be due to the lack of calibration of each sample, and as they are all consistently the same it was assumed that the initiators had been made correctly.

Table 3.1 Data obtained for the synthesis of G-initiators (G1-G3) after purification and desalting

Initiator	Abs ₂₆₀	Amount (μmol)	Concentration in 100 μl (mM)	Expected Mass m/z	Actual Mass m/z
G1	2.24	0.02	0.18	787.2624	785.0951
G2	6.02	0.05	0.50	1131.3950	1129.1052
G3	3.24	0.03	0.27	1819.6604	1817.1500

NMR analysis was not possible due to the small scale of the synthesis. In order to confirm the presence of a terminal alkyne group for (**G1-G3**) a click reaction was carried out using coumarin azide (**33**). (**33**) is an excellent fluorogenic probe for the click reaction as (**33**) is not fluorescent in its azide form due to the electron rich α -nitrogen group. However, once the azide reacts with the alkyne group forming a triazole ring (**34**) it fluoresces at 478nm (Scheme 3.2).¹⁵¹

removed from the reaction the additional peaks in the UV/vis region were confirmed to be from the (33) and the TBTA ligand.

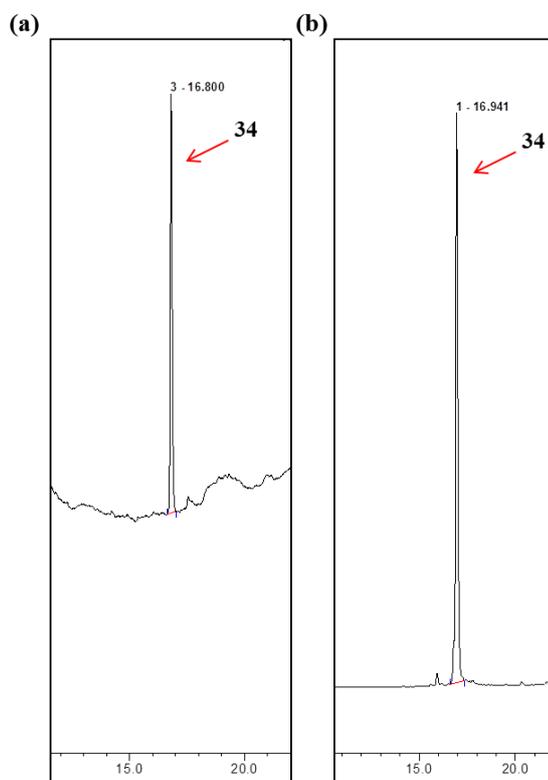


Figure 3.2 RP-HPL chromatogram for the click reaction between G-initiator (G1) and coumarin-N₃, conditions were as described in Section 3.4.1; (a) UV/vis chromatogram at wavelength 260 nm, peak observed at 16.8 minutes (b) emission chromatogram at wavelength 492 nm, peak observed at 16.94 minutes.

The initial synthesis of all three initiators (G1-G3) gave relatively low yields 10%, 25% and 15% respectively, giving concentrations of between 0.18-0.5 mM in 100 ul of water (Table 3.1). Given that earlier studies have shown that a minimum of at least 4 mM is required to incorporate into the transcript, these quantities are subsequently insufficient.^{86,92} In order to increase the yield it was decided to synthesise the G-initiators using multiple (10 in total) 1.0 μ mol iPr-Pac-G RNA SynBase™ CPG 1000/110 columns and combining them after purification. Due to the need of ten columns per initiator it was decided to synthesise the G-initiator which had only one HEG phosphoramidite linker between the base and the alkyne (G1) to begin with, due

to the previous study by Jäschke *et al.* which showed anything under PEG₁₄ incorporates efficiently. Therefore G-initiator (**G3**) and probably G-initiator (**G2**) are not likely to incorporate as efficiently as G-initiator (**G1**).⁸⁸

3.2.2 Synthesis of G-Initiator (G1)

G-Initiator (**G1**) was synthesised on a much larger scale (10 x 1 μ mol syntheses) with the 3'-terminal guanosine used as previously described (Section 3.2.1). After purification by semi preparative RP-HPLC all the respective fractions were combined. After lyophilisation the absorbance of the (**G1**) was measured at 260 nm to be 0.528 in a 1:100 dilution and the amount was calculated to be 43.93 μ mol with an overall concentration of 43.93 mM in 100 μ l, attaining 44% yield. The purity was reanalysed by analytical RP-HPLC (Figure 3.3a) and the mass was determined to be 788.2520 by LC/MS (Figure 3.3b).

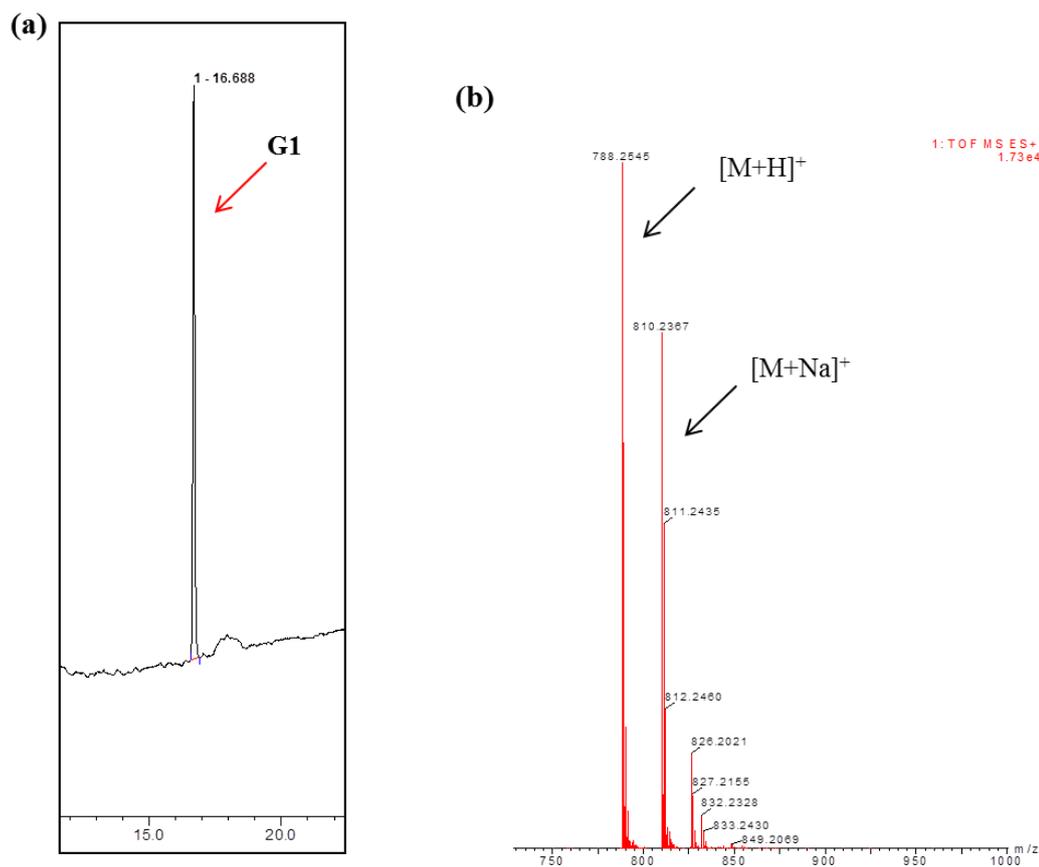


Figure 3.3 RP-HPLC and Mass spectra of the purified G-initiator (G1), RP-HPLC conditions were as described in Section 3.4.1; (a) RP-HPL chromatogram in UV/vis region of 260 nm, peak observed at 16.6 minutes; (b) LC/MS of (G1) which shows m/z values that correspond to expected molecular weight.

3.2.3 Click Reaction with G-Initiator (G1) and Coumarin Azide (33)

Coumarin azide (**33**) was reacted with G-Initiator (**G1**) by copper catalysis using the TBTA ligand (Scheme 3.2) to confirm the presence of the alkyne functional group. The reaction was analysed by analytical RP-HPLC at the wavelengths 260 nm and 492 nm, the absorption wavelength of RNA and emission wavelength of (**34**) respectively. At 260 nm the RP-HPL chromatogram displayed 3 major peaks at 18.58, 20.28 and 23.4 minutes (Figure 3.4a) and at 492 nm the chromatogram showed only 1 peak at 18.60 minutes (Figure 3.4b). This indicated that the click product (**34**) was present and the marginal delay in retention time is due lag time between to two detectors. The peak at

23.4 minutes had previously been investigated (Section 3.2.1) and can account for the TBTA ligand. The reaction was re-run on the RP-HPLC this time spiking the reaction with starting G-initiator 1 to determine if there had been a shift between from the starting material (**G1**) to the product (**34**) (Figure 3.4c & 3.4d), an additional peak in the 260 nm wavelength was also observed at 16.57 minutes which would further clarify the formation of the product (**34**) (Figure 3.4c). However, this does not explain the additional peak at 20.28 minutes and as there are no additional by-products no additional peaks should be present. As the reaction was run in 30% DMSO, a sample of DMSO was also analysed by RP-HPLC and at 260nm a peak with the retention time of 20.3 minutes was also observed which would then account for the peak in the previous spectra.

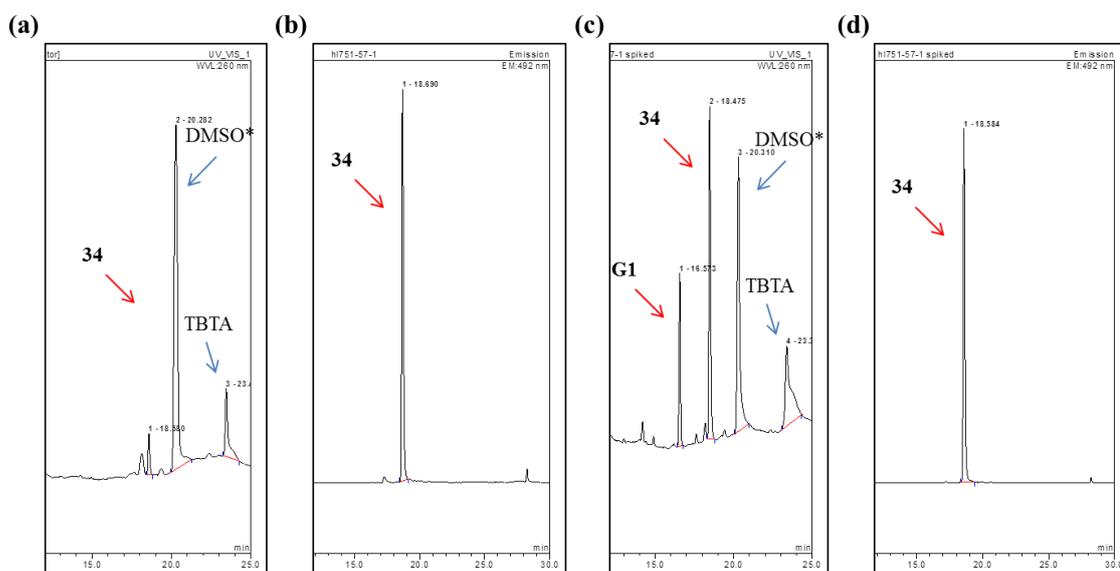
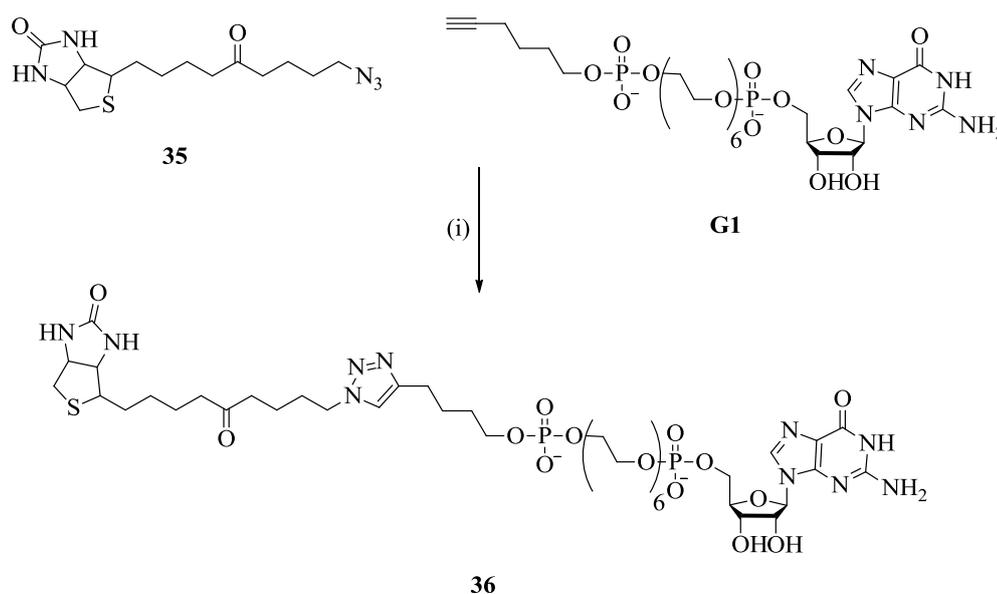


Figure 3.4 RP-HPLC spectra of the for the click reaction between G-initiator (**G1**) and coumarin-N₃ (**33**), conditions were as described in Section 3.4.1; (a) chromatogram for the UV/vis wavelength 260 nm (b) corresponding chromatogram for the emission wavelength at 492nm, peak at 18.5 minutes corresponds to click products, (c) chromatogram at 260 nm for the reaction spiked with (**G1**), (d) corresponding chromatogram at 492 nm. (DMSO*: contamination in DMSO)

3.2.4 Click Reaction between G-Initiator (G1) and Biotin azide (35)

Biotin azide (**35**) was then reacted with the G-initiator (**G1**) using the Cu(I)TBTA catalyst (Scheme 3.3). The reaction was analysed by RP-HPLC at 260 nm and a peak was observed at 16.96 minutes (Figure 3.5a) which was collected, and concentrated *in vacuo*, desalted and analysed by MALDI. A mass for the clicked product (**36**) was observed at 1110.5323 (Figure 3.5b) with the expected mass being 1112.4015. The loss of protons can be accounted for the potential negative charges on the two phosphate groups.

Scheme 3.3 Click reaction between Biotin-N₃ (**35**) and G-initiator **G1**; (i) Cu(I), TBTA



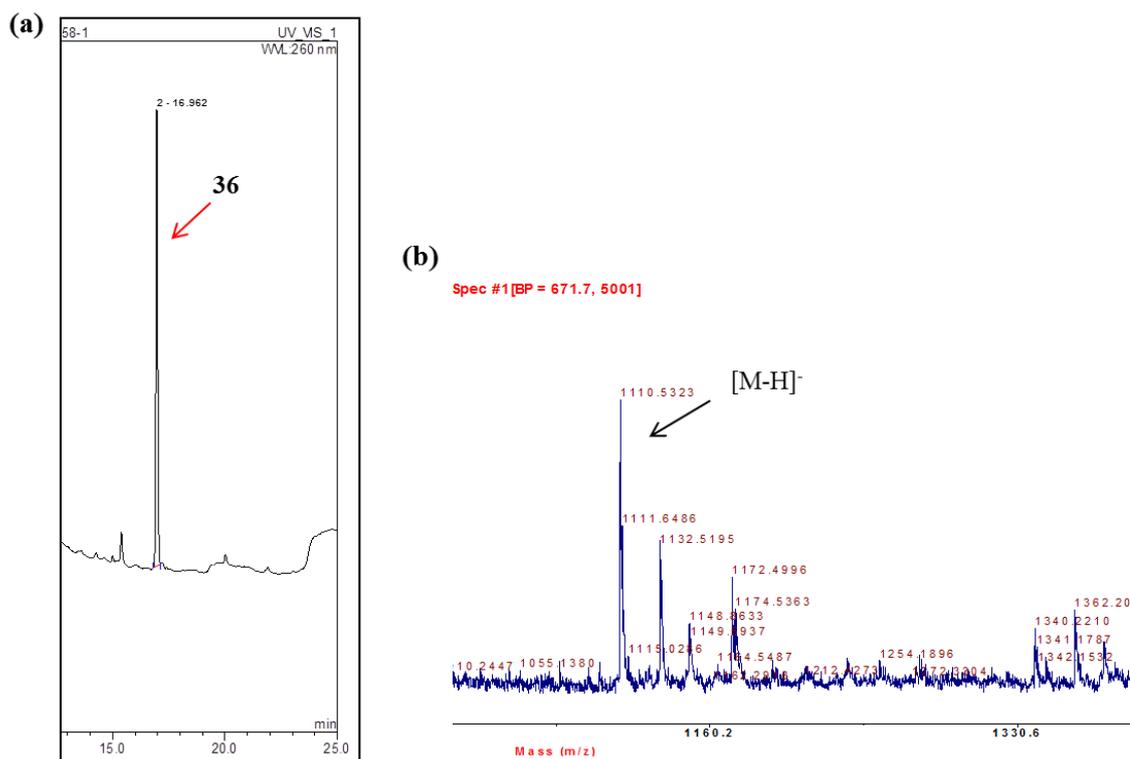


Figure 3.5 RP-HPLC and MALDI spectra for the reaction between biotin- N_3 (**35**) and G-initiator (**G1**), RP-HPLC conditions were as described in Section 3.4.1; (a) RP-HPL chromatogram for the wavelength 260 nm, (b) MALDI from the RP-HPLC fraction collected at 16.9 minutes, m/z of 1110.5323 is observed.

3.2.5 THPTA ligand versus TBTA ligand

RNA is insoluble in organic solvents and needs to be dissolved in aqueous solutions. A significant drawback in using the TBTA ligand (Figure 3.6) is that it is not water soluble and requires dissolution in 25% tert-butanol in DMSO. Developed protocols use a mix of DMSO and water for the click reaction between oligonucleotides and azides in the presence of copper(I) and TBTA. DMSO however, is not compatible in RNA splicing reactions and has been shown to affect alternative splicing in some genes, therefore it would be preferable to use a ligand that is completely water soluble.¹⁵⁴ Hence it was decided to synthesise and subsequently use THPTA (Figure 3.6) which is a water soluble ligand and has been shown to work with RNA alkynes and (**33**).¹²⁷

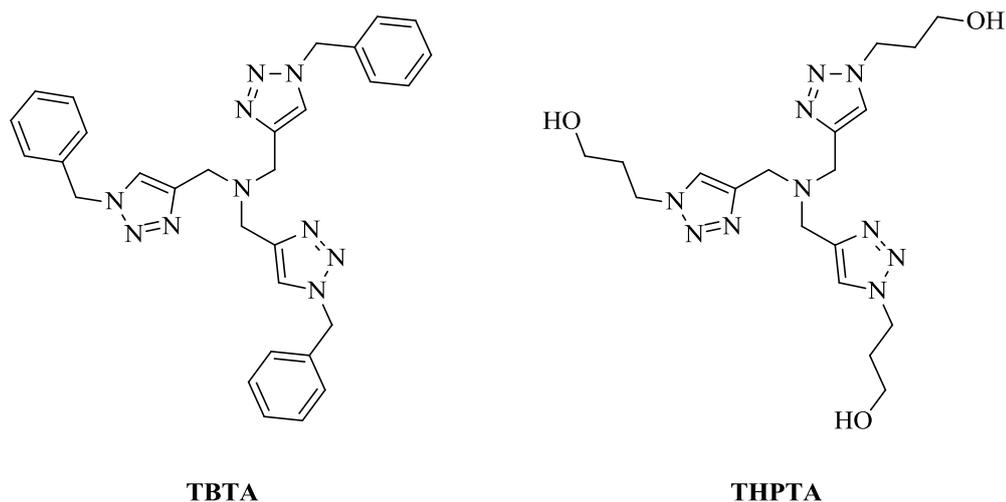


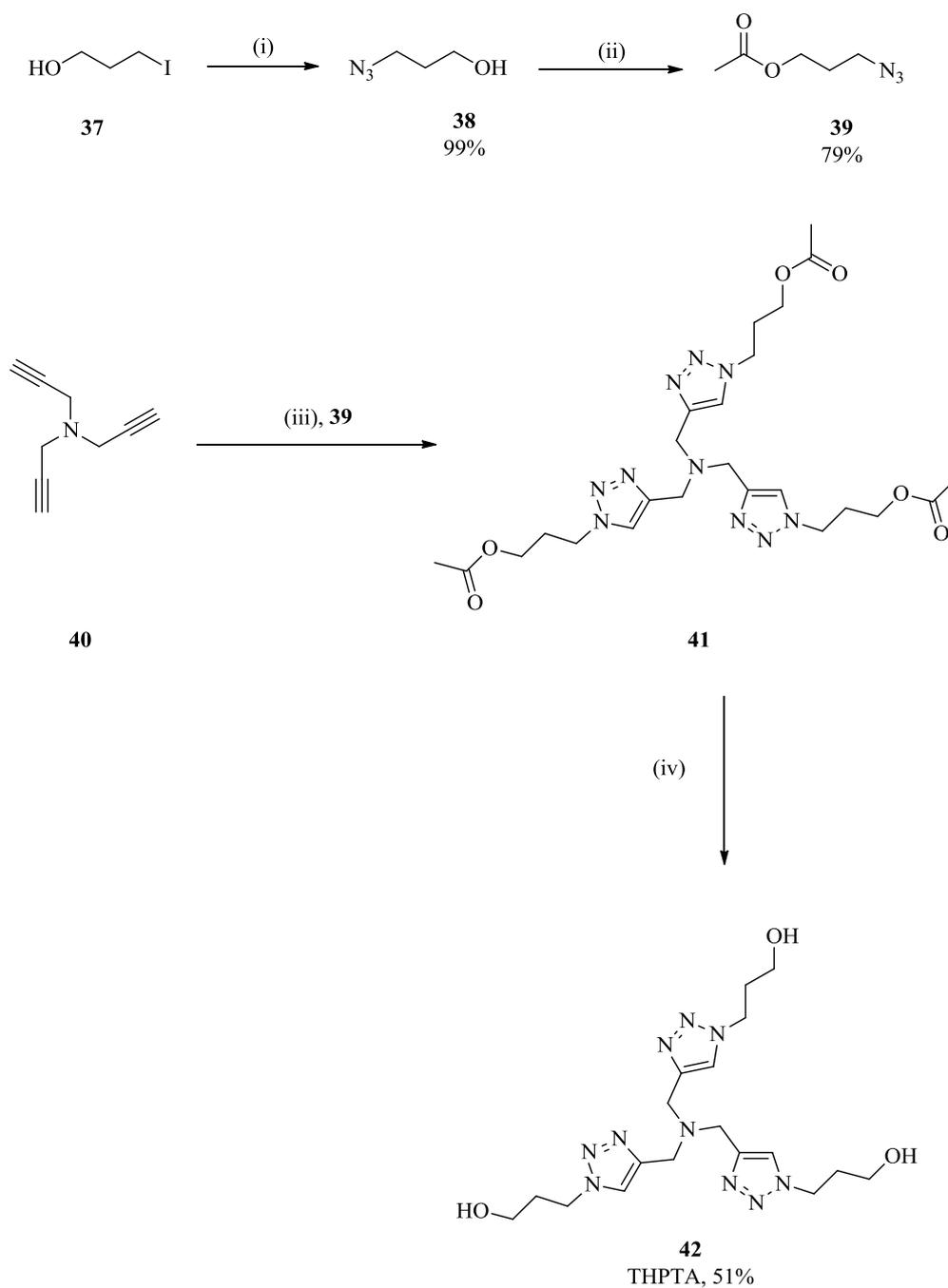
Figure 3.6 TBTA and THPTA ligands used in copper click reactions.

3.2.5.1 Synthesis of tris(hydroxypropyl)triazolylmethyl-amine (THPTA)

THPTA was initially synthesised as described by Finn in 2009 replacing the 3-bromopropan-1-ol with the iodo derivative (**37**).¹²⁷ The synthesis of the 3-azido-propyl acetate (**39**) was performed over 2 steps (Scheme 3.4). In the first step, (**37**) was reacted with NaN_3 overnight obtaining intermediate 3-azido-propan-1-ol (**38**) attaining 99% yield. (**38**) was subsequently reacted with acetic anhydride (Scheme 3.4, step ii) overnight at room temperature. However analysis by ^1H NMR revealed the reaction had not gone to completion, the reaction mixture was reduced to half the volume and the temperature was increased to 40°C to drive the reaction to completion. NMR analysis confirmed the production of pure (**39**) attaining an overall yield of 79% which is considerably more than the 63% achieved by Finn *et al.* However, the formation of (**41**) between tripropargylamine to 3-azido-propyl acetate in the presence of copper(I) acetate (CuOAc) as described by Finn *et al.*¹²⁷ was unsuccessful. This could be due to the quality of the CuOAc which may have partially oxidised to Cu^{II} which would therefore not act as an efficient catalyst in the click reaction. The CuOAc was replaced with

copper(I) trifluoromethanesulfonate which produced pure THPTA (**42**) with an overall yield of 51%. Copper trifluoromethanesulfonate was not the most efficient catalyst as the reaction was not complete after 2 days using 5 mol% when analysed by NMR which showed that the mono, di and tris-(hydroxypropyl)triazolymethyl-amine were all present, and therefore an additional 20 mol% was added to drive the reaction. Despite the inefficiency of the reaction the yield produced enough for subsequent click reactions.

Scheme 3.4 Synthesis of tris(hydroxypropyl)triazolylmethyl-amine (THPTA); (i) sodium azide (10 equiv.), water, (ii) acetic anhydride (1 equiv.), triethylamine (1 equiv.), DCM, (iii) (39) (5 equiv.), Cu(I)trifluoromethanesulfonate (5 mol%), 2,3,5 collidine (1 equiv.), (iv) 35% ammonium hydroxide solution in methanol.



3.2.5.2 Click Reaction between G-Initiator (G1) and fluorophore azides (33) & (43) with THPTA ligand

Click reactions between G-initiator (**G1**) and a fluorophore azide (**33**) or (**43**) were carried out using the THPTA ligand. As previously discussed, the Finn group had developed the THPTA ligand along with a method to conjugate a biomolecule such as RNA to a corresponding azide such as coumarin azide (**33**) (Scheme 3.2).¹²⁷ Initially the Finn method was carried out at both room temperature (25°C) and at 40°C for the reaction between G-initiator (**G1**) and (**33**). When analysed by RP-HPLC the chromatograms at both wavelengths (260nm & 492nm) showed no peaks at the expected retention time of around 19 minutes for the click product (**34**), nor was there a peak present for (**G1**) in the UV/vis region of 260 nm (Table 3.2; 1 & 2). This however, could be due to the fact there were peaks present at 3.4 and 3.8 minutes with high absorbances that could potentially smother the signal of the initiator, but this does not explain why a peak in the emission region of 492 nm was not observed. Since the Finn method came up with negative results using (**33**), the reaction was repeated using the same conditions as the previous initiator/coumarin (**G1/33**) click reactions (Section 3.2.3) with both TBTA and THPTA (Table 3.2; 3&4). RP-HPLC analysis showed that for both reactions there were two defined peaks albeit low absorbances at 16.6 and 19.2 minutes in the UV/vis region at 260 nm which is the expected retention time for the gradient used (5-50% buffer B over 20 minutes) for the initiator (**G1**) and the click product (**34**) (Appendix, Figures 6.3 and 6.4). A peak was likewise present at 19.3 minutes for the emission wavelength of 492 nm, however, the absorbance measured for the THPTA reaction (Table 3.2; 3) and for the TBTA reaction (Table 3.2; 4) were extremely low. Although there are two definitive peaks in the UV/vis region the low

absorbance for the emission wavelength renders these results inconclusive and needed investigating further.

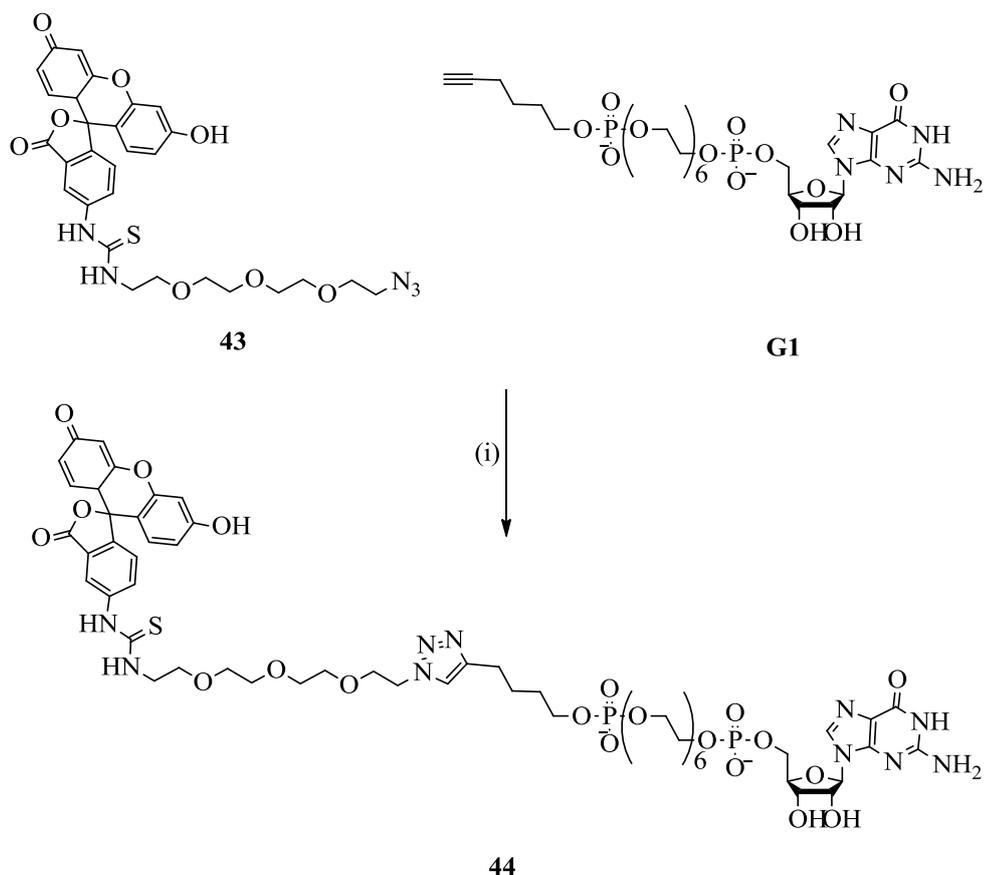
For further analysis fluorescein thiourea (FTU) azide (**43**) was used as the fluorophore as it is highly fluorescent before and after bio-conjugation (Scheme 3.5). The three reactions were repeated and analysed by RP-HPLC at 260 nm and 521nm. The Finn method gave similar results for (**43**) as it had for (**33**) (Table 3.2; 7), a peak was not present at the expected retention time of approximately 19 minutes, nor was a peak seen for the initiator but there were peaks at 3.4, 4.1 and 26.2 minutes (Appendix, Figure 6.7). The peak at 26 minutes was determined to be that of (**43**) and a corresponding peak was visible for the emission wavelength at 521 nm. The chromatograms for the UV/vis region at 260 nm there were two peaks observed for the reactions using the TBTA method for both ligands (Table 3.2; 6 & 7), the peak for (**43**) at approximately 26 minutes and an additional peak at 20.1 minutes. The chromatogram for the emission wavelength also shows two peaks which correspond to the peaks in the UV/vis region, confirming that the reaction had taken place and the conversion of both reactions from initiator (**G1**) to click product (**44**) appear to be 100%. The intensity of the peak at 20.1 minutes at 521 nm for THPTA (Table 3.2; 6) was greater than a 1000 mV and had reached the limit of the detection.

Table 3.2 Various click reactions between fluorophore azides and G-initiator comparing THPTA and TBTA ligands. All reactions were at 40°C unless otherwise stated (Appendix, Figures 6.1 to 6.7).

Reaction	Azide	Ligand	Method	Retention time at 260nm (minutes)	Retention time at emission wavelength* (minutes)	Percentage conversion from 1 to click product (%)
1	(33)	THPTA	Finn (r.t) ¹²⁷ Section 3.4.18	3.43 3.85	none	0
2	(33)	THPTA	Finn ¹²⁷ Section 3.4.18	3.46 3.85	none	0
3	(33)	THPTA	Section 3.4.17	16,64 19.18	19.29	44
4	(33)	TBTA	Section 3.4.16	16.65 19.18	19.28	37
5	(43)	TBTA	Section 3.4.16	20.11 26.26	20.13 26.25	100
6	(43)	THPTA	Section 3.4.17	20.14 26.26	20.19 26.31	100
7	(43)	THPTA	Finn (r.t) ¹²⁷ Section 3.4.18	3.49 4.10 26.27	26.32	0

* Wavelength maxima for Coumarin (33) and FTU (43) are 492 nm and 521 nm respectively

Scheme 3.5 Click reaction between FTU-N₃ (**43**) and G-initiator (**G1**), (i) Cu(I), TBTA.



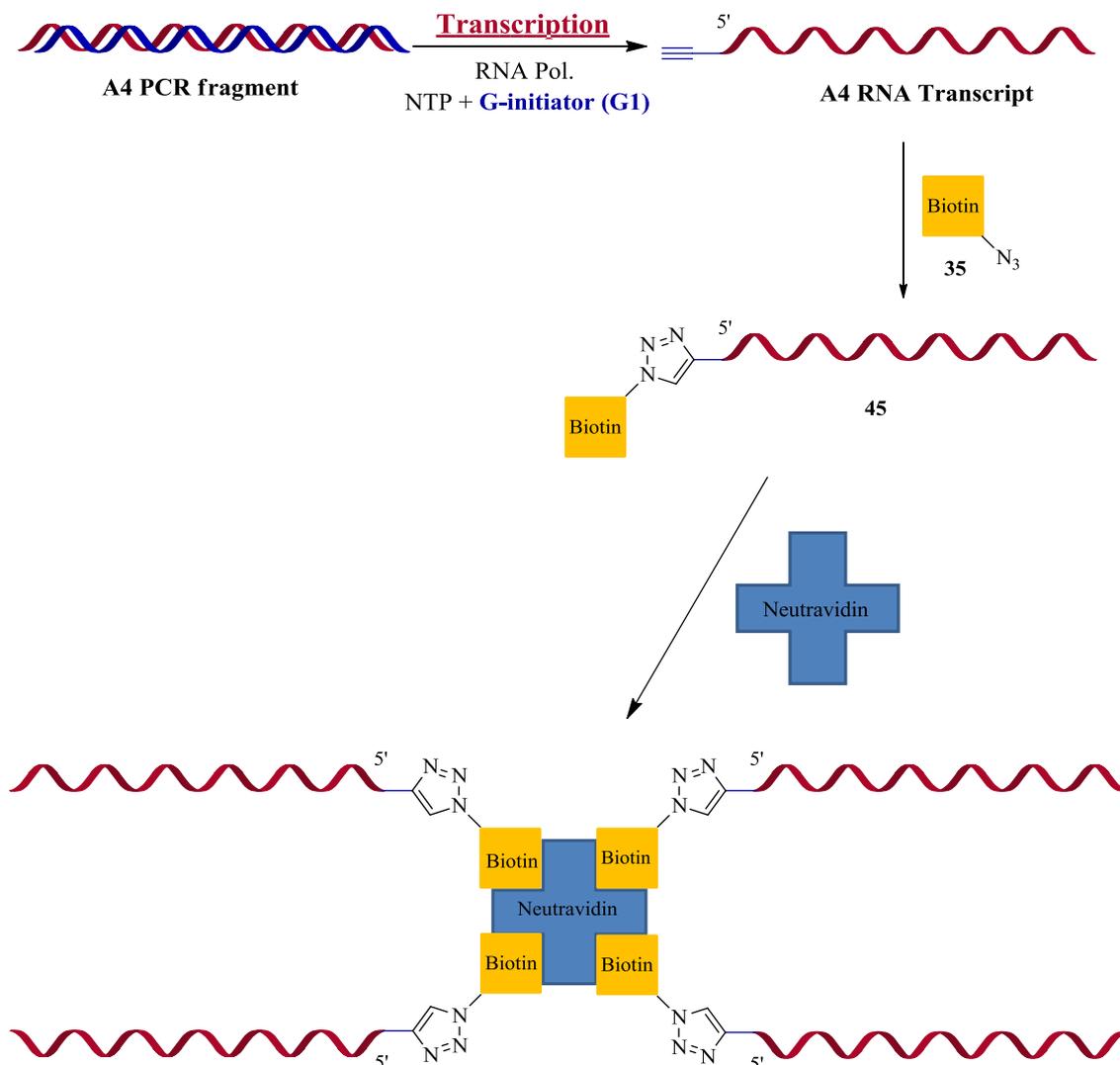
These results have verified that the Finn method was not a viable route as it either did not work or could not be monitored by RP-HPLC for this particular reaction. Furthermore, there is little difference in reactivity between TBTA and THPTA. The conversions from initiator (**G1**) to click product (**34**) for the reaction with (**33**) are 44% and 37% for THPTA and TBTA respectively and quantitative conversion was observed when reacted with FTU-azide. Consequently THPTA is an ideal ligand for subsequent reactions due to its water solubility and efficiency.

3.2.6 Incorporation of G-Initiator (**G1**) into **A4**

In order to construct the RNA tripartite system based on **A3** (Figure 2.7b), the linker needed to be inserted between the ESE and the remaining RNA transcript. **A4** (Figure

2.7b) 5' end begins where the ESE sequence ends, therefore incorporating (**G1**) into **A4** enabled the modified ESE to be conjugated to the transcript at the correct site. Hence, initial incorporation tests were performed using the **A4** PCR fragment and (**G1**). During transcription a 7-methylguanosine (m⁷G cap) is used to modify the 5' end of the transcript in order to protect the RNA from 5' exonucleases, it is also known to promote splicing *in vitro*.¹⁵⁵ However, in this case the m⁷G is not required, replacing it with the initiator during transcription, using the same conditions as normal transcription with low GTP (0.05 mM) and 1 mM of (**G1**). The incorporation was tested by a click ligation of the alkyne-modified transcript with (**35**) and a subsequent pull down assay using NeutrAvidin™ beads (Scheme 3.6).

Scheme 3.6 Schematic representation of the incorporation of G-initiator (G4) into transcript (A4) by T7 polymerase transcription. The subsequent click reaction with (35) and the biotin pull down assay using NeutrAvidin™.



Transcription of **A4** with the G-initiator (**G1**) was carried out alongside **A4 + m⁷G** and **A4 - m⁷G** as controls for the biotin pull down assay to determine the incorporation of (**G1**) into **A4**. The radioactivity of each transcript was recorded by scintillation counting and the number of moles of each transcript was calculated by the number of guanosines present within the transcript ranging from 8-70 fmoles.

Each transcript (0.03 nM) was reacted with (35) (1 μ M) in the presence of Cu(I) TBTA and Cu(I) THPTA forming conjugated transcript (45) (Scheme 3.6). Due to the low molarity of each transcript a large excess of (35) was used, this was to ensure that enough azide (35) was present after several dilutions from the stock solution. After 40 minutes the reactions were added to the NeutrAvidin™ beads and after an hour of incubation the samples were washed with wash buffer until the control samples were no longer deemed radioactive using a Geiger counter as control samples should not be bound to the biotin azide. Gel electrophoretic analysis of reactions (Table 3.3) is presented in (Figure 3.7). Bands were observed in every lane, and surprisingly the control with **A4 + m⁷G** (Figure 3.7a, lane 1) had a more intense band than the other reactions. The bands were quantified to determine the number of counts of each band (Figure 3.7c). The reactions which used THPTA were all similar in their radioactivity, comparable results were seen for the TBTA reactions (Figure 3.7c).

Table 3.3 Reactions between biotin azide (35) and various A4 transcripts in the presence of Cu(I) and ligand TBTA/THPTA.

Reaction	Transcript	Azide	Ligand
1	A4 + m⁷G	35	TBTA
2	A4 - m⁷G	35	TBTA
3	A4 + G1	35	TBTA
4	A4 + m⁷G	35	THPTA
5	A4 - m⁷G	35	THPTA
6	A4 + G1	35	THPTA

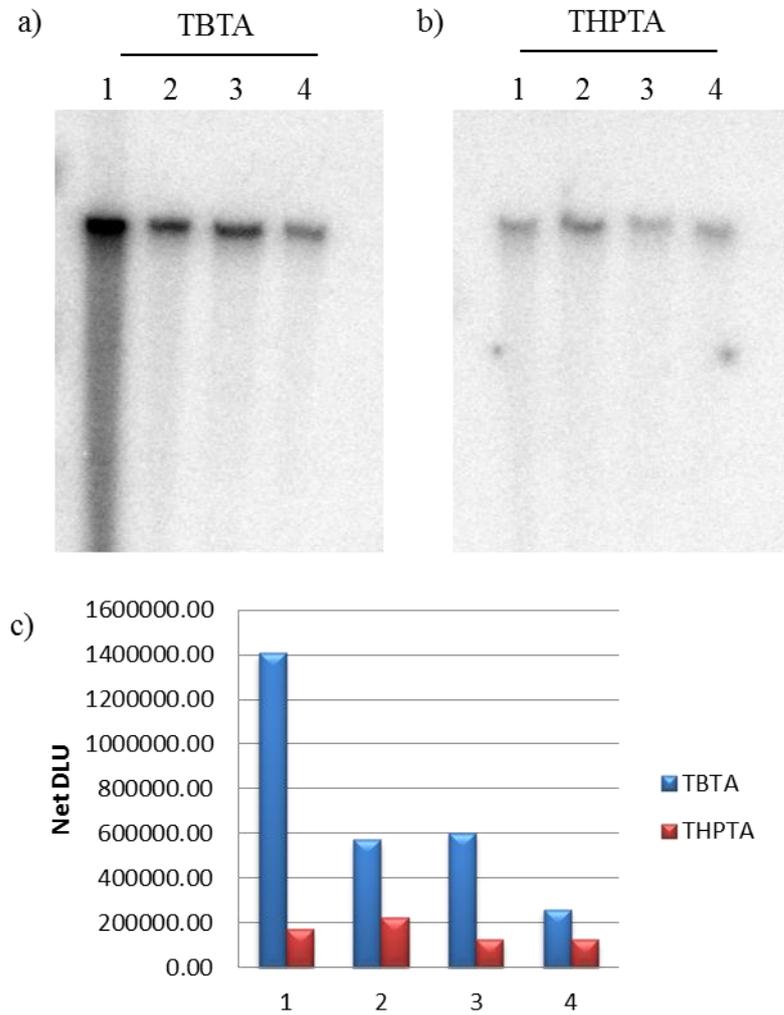


Figure 3.7 Biotin pull down assay of the click reaction between G-initiator (G1) incorporated A4 transcript; (a) polyacrylamide gel for the click reaction using TBTA ligand, lane 1 A4+m⁷G + (35), lane 2 A4-m⁷G + (35) lane 3&4 A4+1 + (35), (b) polyacrylamide gel for the click reaction using THPTA ligand, lane 1 A4+m⁷G + (35), lane 2 A4-m⁷G + (35) lane 3&4 A4+1 + (35), (c) a histogram presenting quantified radioactive counts of each band.

These results indicate that the initiator had either not incorporated into the transcript (or if it had it was in minimal quantities) or the length of the RNA transcript inhibited the click reaction but either way further studies needed to be implemented.

3.2.7 Inhibition Effect of Transcription with the Addition of G-Initiator (G1)

The first test was to see if the initiator at varying concentrations had an inhibitory effect on transcription. Various transcriptions were set up with concentrations ranging from 0-0.8 mM and the transcripts were run on a polyacrylamide gel and quantified the number of counts each transcript. The results obtained were comparable indicating that the initiator (**G1**) did not inhibit transcription (Figure 3.8). Therefore, if the initiator was incorporating into the transcript the biotin conjugation and subsequent biotin pull down assay is not sufficient enough to determine incorporation efficiency. As a result it was decided to look for another detection method. One plausible route is through the use of fluorophores.

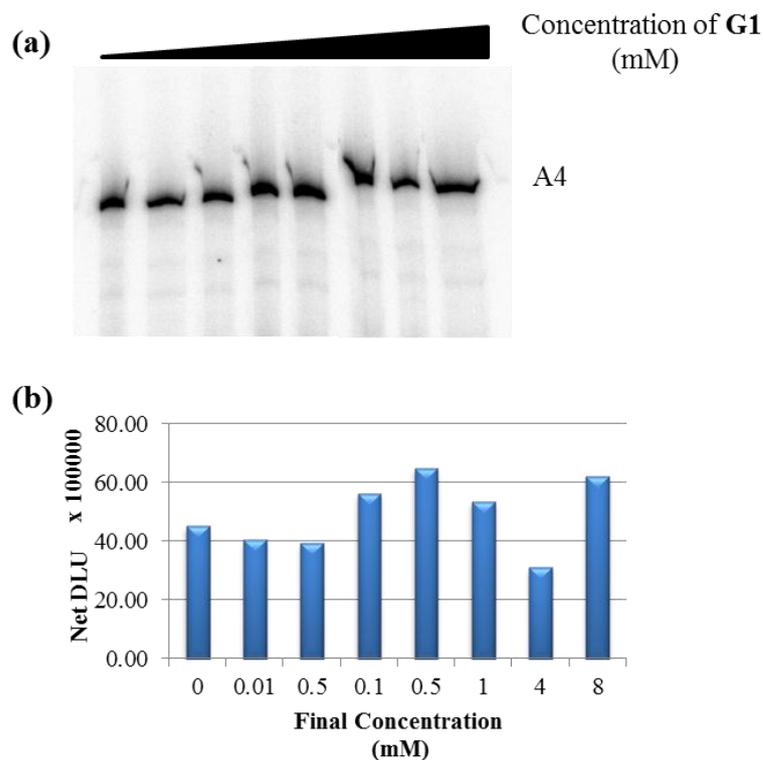


Figure 3.8 Inhibition effect of transcription with the addition of G-initiator (G1); (a) polyacrylamide gel of transcripts increasing concentration of initiator from 0-0.8mM, (b) histogram representing the radioactivity of each transcript for the varying concentrations of G-initiator.

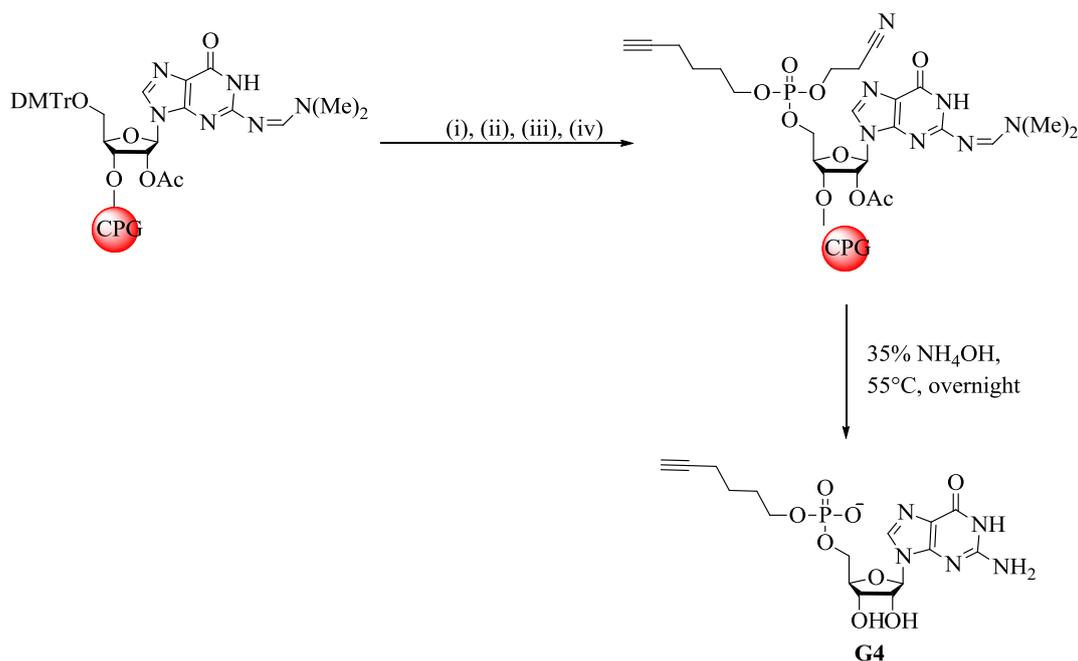
3.2.8 Fluorescence as a Means of Detection

The click reactions between **A4+G1** and biotin azides followed by the biotin pull down assay indicated that the incorporation of (**G1**) into **A4** had not occurred although the results were ambiguous as bands of RNA were present for the control reactions (Sections 3.2.6). It was therefore decided to investigate whether it was possible to use a fluorophore-azide to click to the transcript. A β -globin oligonucleotide modified with Alexa 488 fluorophore was used to determine the minimum concentration required for the detector. The oligonucleotide was run on a 15% polyacrylamide gel with concentrations ranging from 0.01 – 1000 nM. The oligonucleotides that were less than 1 nM could not be detected and as a result this method could not be used as the RNA transcripts were always less than 0.5 nM.

3.2.9 Development of G-Initiator (**G4**)

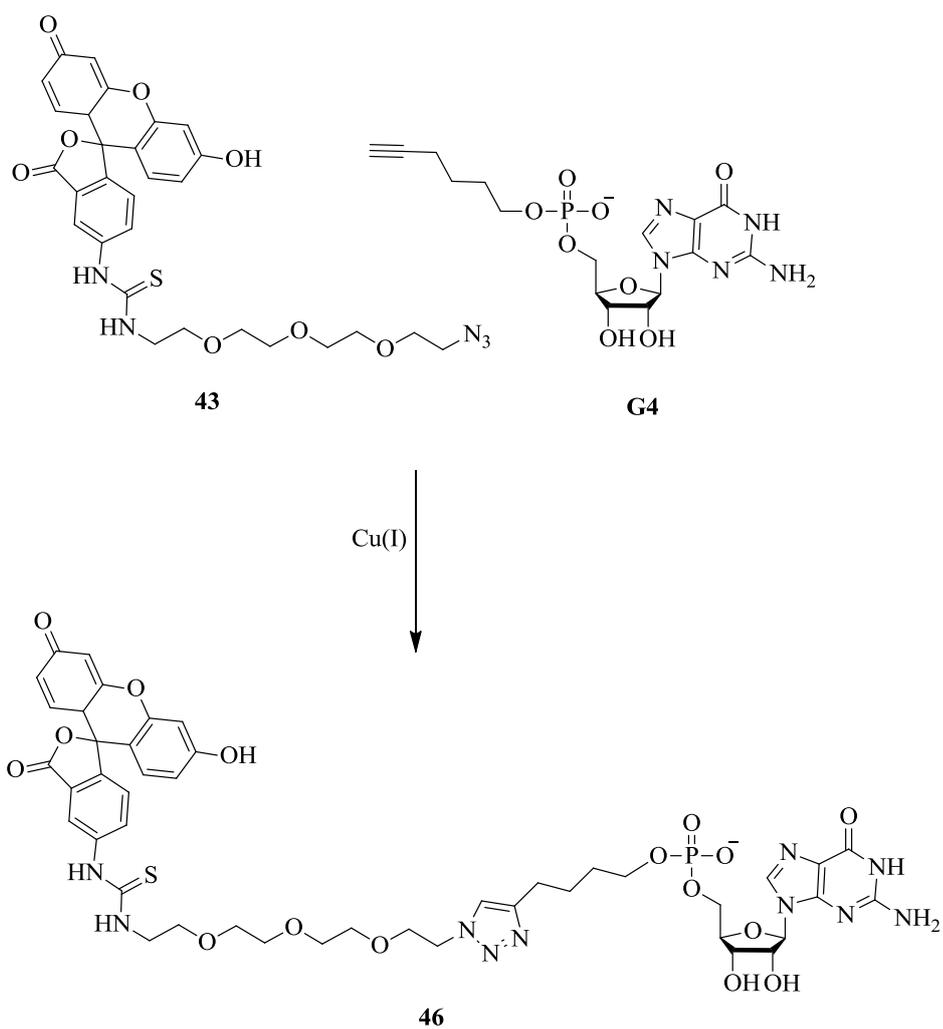
An additional initiator (**G4**) was designed due to the inconclusive results of the incorporation of (**G1**), this time removing the PEG linker which may be the cause of the lack of initiation. The 5'-hexanyl phosphoramidite was coupled to the G base by solid phase synthesis like the previous initiators (**G1-G3**) synthesised (Scheme 3.7). The absorbance at 260 nm after purification was 3.13 AU for a 1 in 100 dilution and the number of moles calculated to be 2.61 μ moles affording a 26% yield. The yield was considerably lower than (**G1**) at 44% but was most likely due to the coupling efficiency during the solid phase synthesis. The structure and purity was confirmed by RP-HPLC (Figure 3.9a), LC/MS and NMR analysis.

Scheme 3.7 Synthesis of G-Initiator (G4) on CPG solid support: (i) detritylation, (ii) coupling of hexanyl phosphoramidite, (iii) capping of unincorporated base, (iv) oxidation of newly coupled phosphoramidite.



Further confirmation that the coupling of the hexanyl phosphoramidite had taken place was achieved by reacting (G4) with FTU-azide using copper(I) TBTA click chemistry (Scheme 3.8). The reaction was monitored by analytical RP-HPLC in the UV/vis region of 260 nm and at the emission wavelength for fluorescein; 521 nm. A peak at 19.32 and 19.42 minutes was observed on the chromatograms for 260 nm and 521 nm respectively (Figure 3.9b & c) indicative of the click reaction was successful.

Scheme 3.8 Click reaction between G-initiator (G4) and FTU-N₃ (43).



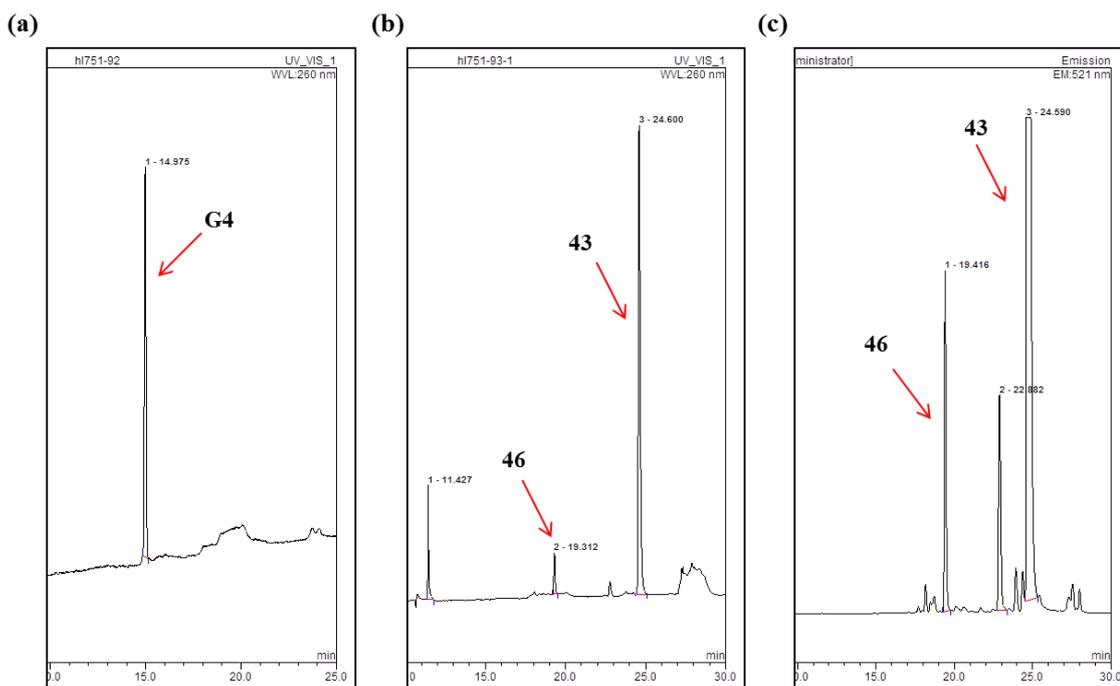


Figure 3.9 RP-HPL chromatograms of G-initiator (G4) and subsequent click reaction with (43), conditions were as described in Section 3.4.1; (a) chromatogram of purified G-initiator (G4) at 260 nm, (b) chromatogram at 260 nm for click reaction between (G4) and (43), peak at 19.3 minutes corresponds to click product (46), (c) corresponding chromatogram at 521 nm for click reaction with peak at 194 minutes.

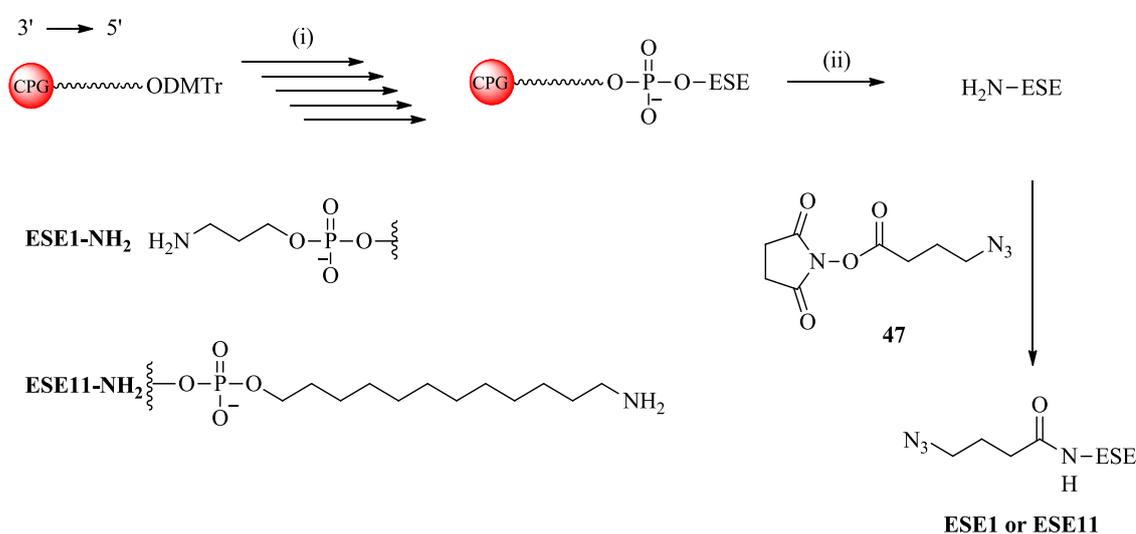
3.2.10 Synthesis of 2'OMe ESE-azides (ESE1 & ESE11)

3.2.10.1 Solid phase synthesis of ESE-amines

ESE1-NH₂ and **ESE11-NH₂** were synthesised by automated solid phase synthesis (Section 1.6.2) amino-modified at the 3' and 5' ends respectively using different modifiers (Scheme 3.9) and then NHS-azide (**47**) coupled to give the subsequent azide (**ESE1** or **ESE11**). Azides cannot be used directly during solid phase synthesis as they are unstable to P^{III} and will not survive the phosphoramidite conditions.¹⁰³ **ESE1** has been designed so that an azide was incorporated into the 3' end in order to bioconjugate the ESE to the G-initiators incorporated into the 5' end of the transcript during transcription. **ESE11** on the other hand was designed to bioconjugate to the transcript at the 3' end where a pCpC alkyne (Scheme 5.2) had been ligated. 2'methoxy RNA was

originally chosen over natural RNA due to its enhanced enzymatic stability to hydrolysis and nucleases. Moreover, previous splicing studies have used 2'methoxy oligonucleotides to bind to pre-mRNA with high affinity without having a negative effect on splicing patterns.^{36,142,156}

Scheme 3.9 Synthesis of ESE1 & ESE11 by solid phase using amino modified CPG solid supports and NHS-coupling of NHS-azide (47). Reagents and conditions. (i) detritylation(3% trichloroacetic acid in DCM), coupling of phosphoramidite (Spacer-CE Phosphoramidite 18 0.12M,5'-MMT-Amino-Modifier C12-CE Phosphoramidite 0.12M,OMe-iPr-Pac-G-CE Phosphoramidite 0.1M, 2'-OMe-Ac-C-CE Phosphoramidite 0.1M, 2'-OMe-Pac-A-CE Phosphoramidite), capping of unincorporated base (Cap A THF/Pyridine/acetic acid (8:1:1) & Cap B 10% Methylimidazole in THF), oxidation of phosphoramidite (iodine in THF/pyridine/water (7:2:1) 0.02M), repeated for addition of each base, (ii) deprotection and cleavage following LinkTechnologies protocols.¹⁵⁷



The ESE-amines were cleaved and deprotected from the resin and desalted and analysed by gel electrophoresis (Figure 3.10a). From the absorbance value at 260 nm the amount of **ESE1-NH₂** was calculated to be 0.13 μmole, 13% overall yield and **ESE11-NH₂** was 0.19 μmole, 19% overall yield (Table 3.4). A single band was observed on the gel when stained with SYBR gold for both the ESE-amines (Figure 3.10a).

Table 3.4 Amounts and concentrations obtained for ESE-amines and ESE-azides

	Abs ₂₆₀	Amount (μ mole)	Concentration (mM)
ESE1-NH₂	3.84	0.13	1.35
ESE11-NH₂	5.39	0.19	1.88
ESE1	2.83	0.10	0.99
ESE11	2.63	0.09	0.92

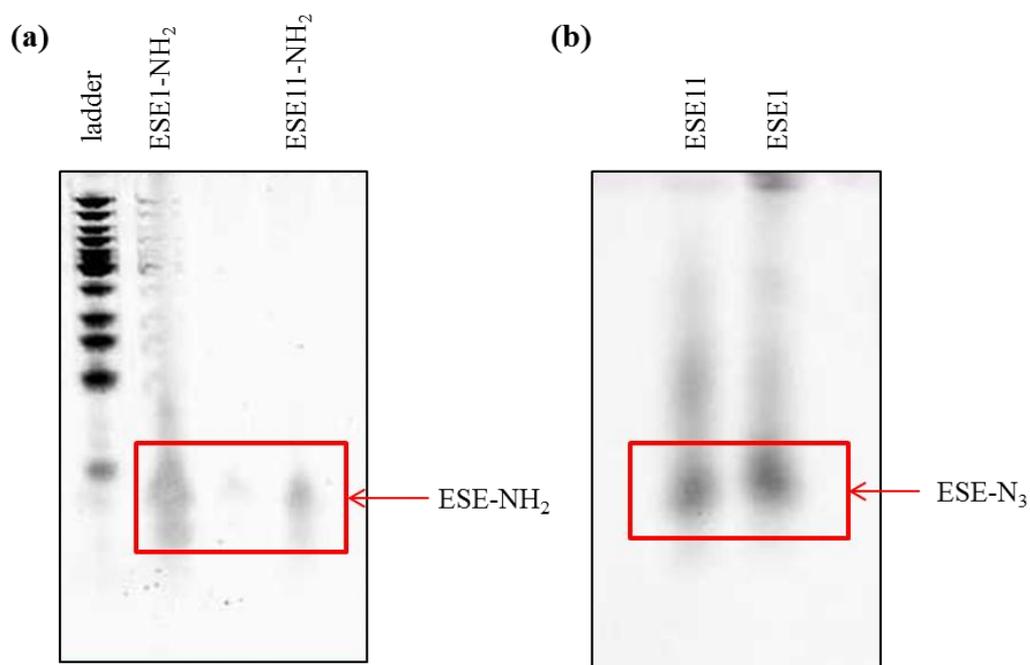


Figure 3.10 Polyacrylamide gels representing the (a) ESE-amines and (b) ESE-azides

3.2.10.2 Synthesis of ESE-azides via NHS coupling

The ESE-Amines were then both coupled to the NHS-azide (**47**), to give **ESE1** and **ESE11** resulting in overall yield of 77% and 47% respectively (Table 3.4) and analysed by gel electrophoresis. Bands corresponding to ligated products of **ESE1** and **ESE11** were observed (Figure 3.10b). A significant amount of streaking for both sequences were observed which could be attributed to ESE being too concentrated when loaded onto the gel. The presence of the azide group was not confirmed by gel electrophoresis since the molecular weight is not substantially different to see a shift on the gel. Mass spectrometric analysis of **ESE1** showed a broad peak with molecular weight $[M-H]^-$ of 8598.4555 the expected mass was 8544.2984 (Appendix, Figure 6.21), either lack of calibration or due to the formation of adducts with Na^+ or K^+ leading to multiple peaks and a broad signal.¹⁵⁸ However, peaks were not observed for **ESE11** which could be due to the high content of guanosines in the sequence (12 out of the 23 bases). Fragmentation of oligonucleotides occurs readily with the most prominent type occurring due to purine loss; therefore, this may have prevented an accurate mass being obtained.¹⁵⁹

3.2.11 Determining the Presence of the Azide on the ESE using Click Chemistry

The ESE-azides (**ESE1** & **ESE11**) were subsequently reacted with a 21 nt DNA alkyne (21-mer) and a 19 nucleotide RNA alkyne (19-mer) (Appendix, Table 6.1) to confirm the NHS coupling had taken place (Figure 3.11). A number of reactions were undertaken changing the molarity of each reagent as well as the ligand used (Table 3.4) to confirm the azide was attached to the ESE.

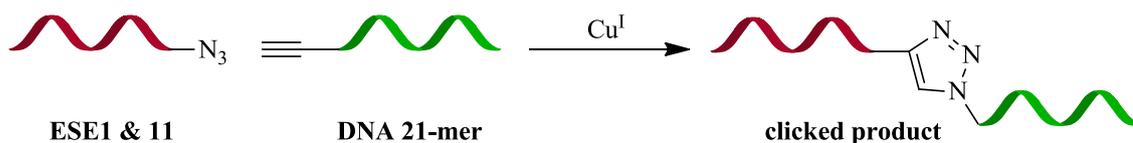


Figure 3.11 Click reaction between the ESEs (**ESE1** and **ESE11**) with DNA 21-mer producing clicked product.

The click reactions between the 21-mer and **ESE1** & **ESE11** (Table 3.5, reactions 5 & 6) in the presence of the THPTA ligand did not afford the desired ligation product, which was confirmed by gel electrophoretic analysis (Figure 3.10e & 3.9f). Reaction 1 where 5 μM of both alkyne and azide oligonucleotides were used, a click product was visible with **ESE1** (Figure 3.12a) and although **ESE11** a band is visible which has a slower electrophoretic mobility than either the 21-mer and **ESE11**, it has migrated faster than the **ESE1** clicked product which was unexpected as the mass of **ESE11** is comparable to **ESE1** and therefore the clicked products should migrate through the gel at the same rate, therefore the results are inconclusive (Figure 3.12a). The bands in reaction 1 are all barely visible which would indicate that the concentration of the starting reagents were too dilute hence the concentrations of both azides were consequently increased to 15 μM in reaction 2 which this time afforded click products for both ESEs (Figure 3.12b) confirming the azide was present on **ESE1** and **ESE11**.

Reactions 3 and 4 were undertaken in order to determine whether an excess of either the alkyne or azide was preferential, with these particular concentrations, having the azide in excess was preferential over the alkyne (Figure 3.12c & 3.12d) when using the ligand TBTA.

Table 3.5 Click Chemistry conditions for reactions between the DNA alkyne (DNA 21-mer) and ESE-azides

Reaction	DNA-Alkyne (21-mer) Concentration (μM)	ESE-Azide Concentration (μM)	Cu(I) Concentration (mM)
1	5	5.	5
2	15	15	0.15
3	5	15	0.15
4	15	5	0.15
5	5	15	0.15 (THPTA)
6	15	5	0.15 (THPTA)

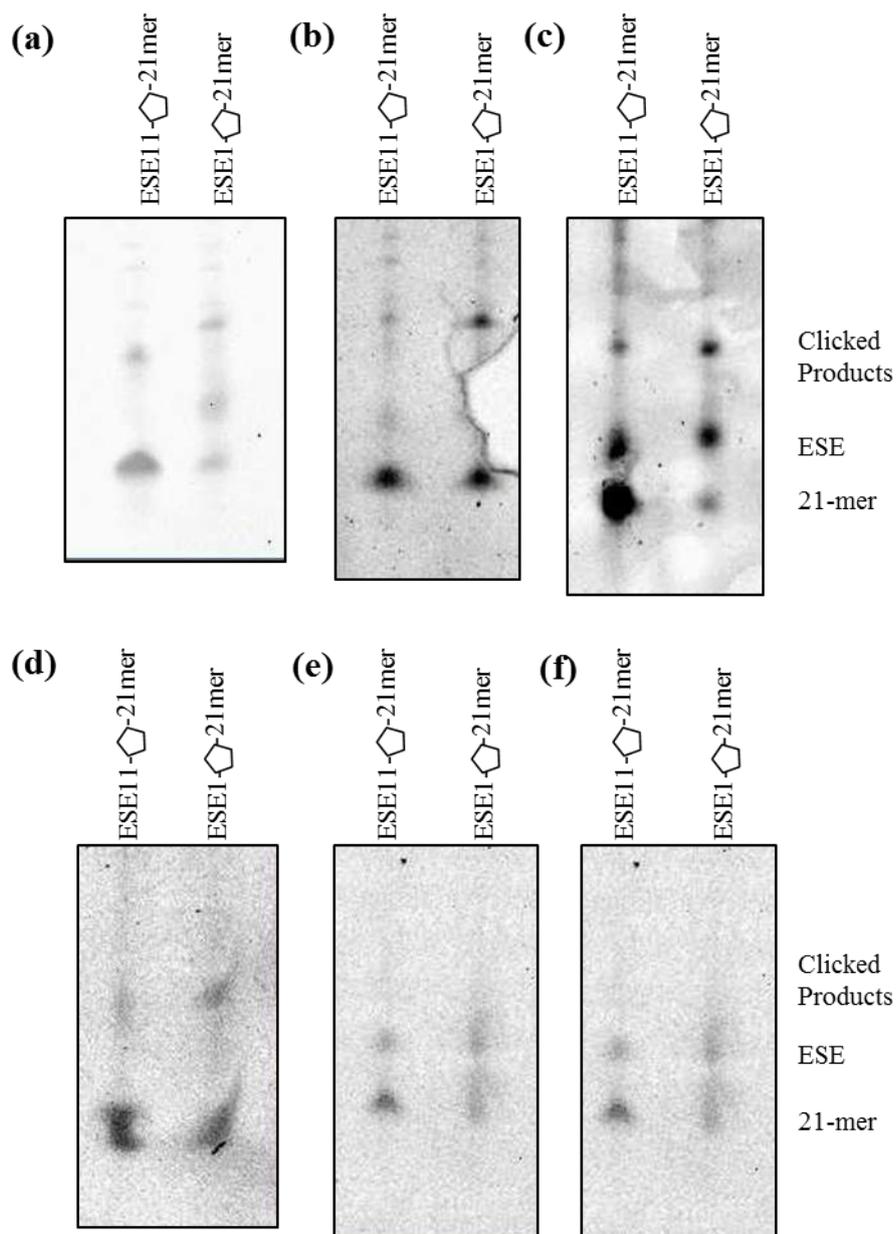


Figure 3.12 Polyacrylamide gels for the click reactions in Table 3.5; (a) reaction 1, (b) reaction 2, (c) reaction 3, (d) reaction 4, (e) reaction 5, (f) reaction 6.

Due to the lack of activity for the click reactions using the THPTA ligand, an additional reaction was undertaken, this time reacting (**G4**) together with **ESE1**; the reagents used to prepare the splicing constructs. (**G4**) was in excess (50 μM) compared to **ESE1** (15 μM), the reaction was run on large 10% polyacrylamide gel to enable the reaction to run for longer on the gel enabling a 1 nucleotide shift. The reaction was run alongside controls to verify each band, it was assumed that due to the size of (**G4**), it eluted off the

bottom of the gel (Figure 3.13, lane 2). However, the reaction with all the reagents (Figure 3.13, lane 1) a band is visible which has a slower electrophoretic mobility than the ESE (Figure 3.13, lanes 3 & 4) due to the increase in molecular weight of the click **ESE1-G4** product. The reaction appeared to have gone to completion as a band is not present for **ESE1** (Figure 3.13, lane 1). This reaction corroborates that THPTA is a viable ligand and that **ESE1** was efficiently coupled to the NHS-azide (**47**). Optimisation of the click reaction was still required for the reactions between the transcript incorporated with the alkyne and **ESE1** as the alkyne transcript was the limiting reagent.

G Initiator (G4) (50uM)	+	+	+	-
ESE1 (15uM)	+	-	+	+
Cu(THPTA) (50mM)	+	+	-	+

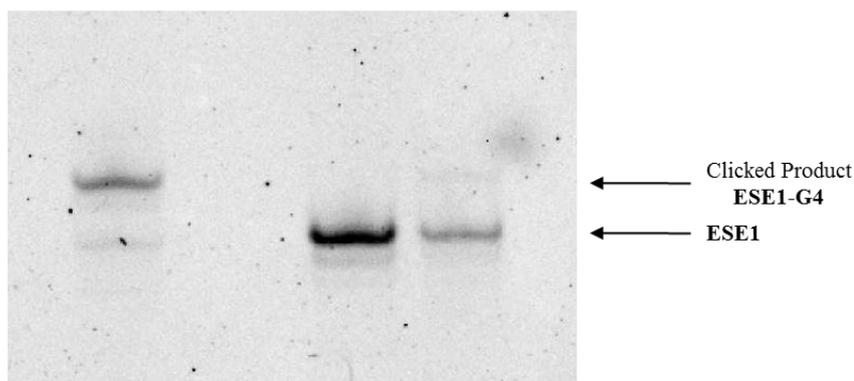


Figure 3.13 Polyacrylamide gel showing the click reaction between **ESE1** and **G-initiator (G4)**.

3.2.12 Incorporation of G-Initiator into 44 Nucleotide Transcript (44-mer)

The previous experiments incorporating **G-initiator (G1)** into **A4** were insufficient in clarifying the initiator had been incorporated in the construct as the one base difference between **A4** and **G1-A4** was not sufficient for separation by gel electrophoresis. It was decided to investigate the efficiency incorporation of both initiators (**G1** & **G4**) by using a much shorter strand of RNA, in this case a 44-mer (Appendix, Table 6.1).

As previously stated the concentration of initiator used during transcription depended on each individual structure but they ranged from 3-5 mM.^{80,86–88,90–92,103,150} As a result it was decided to probe the concentrations required for maximum incorporation into the transcript. A concentration test was set up with a replacing the m⁷G with the initiator over a range of concentrations (0, 0.1, 0.2, 0.3, 0.4 & 0.5 mM) increasing the incubation time to 4 hours (Figure 3.14).⁸⁶

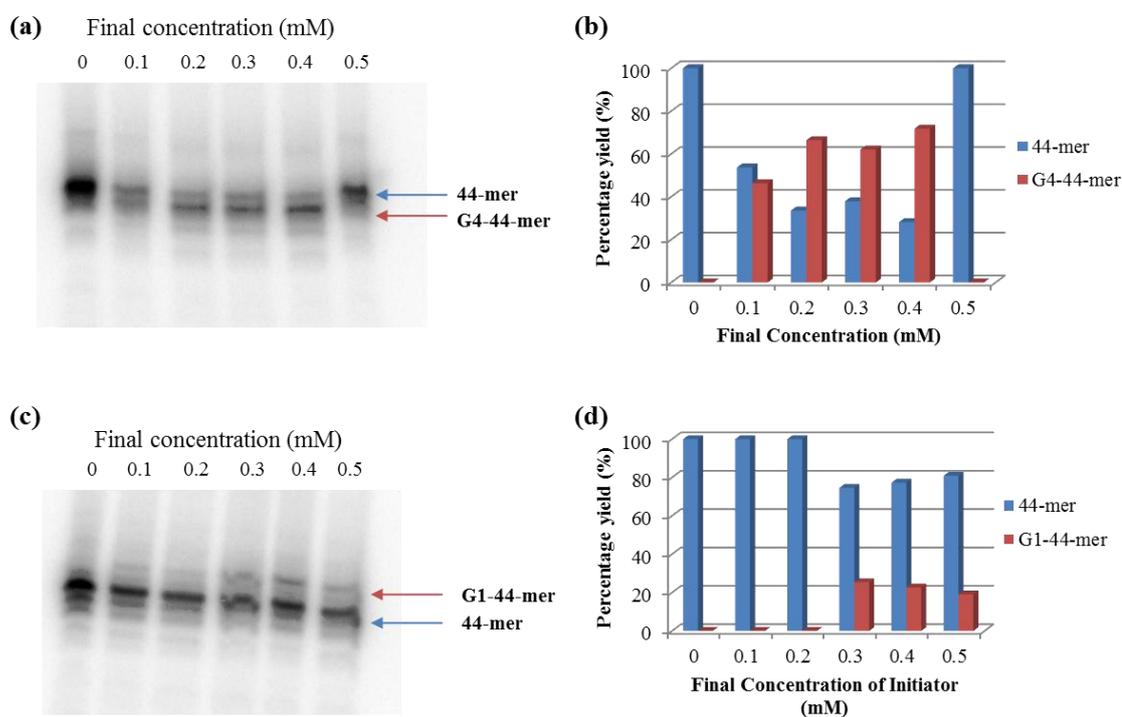


Figure 3.14 Concentration test for the incorporation of G-initiators into a 44-mer transcript; (a) polyacrylamide gel for G-initiator (G4), (b) A corresponding histogram representing the percentage incorporated into the transcript, (c) A polyacrylamide gel for G-initiator (G1) and (d) the corresponding histogram showing the amount of G1 incorporated into transcript.

Upon analysis of the 10% polyacrylamide gel (Figure 3.14), both initiators could be seen to be incorporated into the 44-mer transcript but at different rates and concentrations. (G1) was slow to incorporate: with the addition of 0.1 and 0.2 mM of (G1) no incorporation was observed, with the addition of 0.3 mM on the other hand there was a marked increase to 25% of transcript incorporating (G1). Interestingly there

was a slight decrease in incorporation when 0.4 and 0.5 mM of **1** was added (Figure 3.14, c & d). G-initiator (**G4**) when analysed (Figure 3.14, a & b) incorporated into the transcript with the addition of only 0.1 mM (**G4**) (46% incorporation) and increasing the concentration to 0.4 mM maximum incorporation of 72% was achieved. Surprisingly there was a considerable reduction in the incorporation of (**G4**) when the concentration was increased to 0.5 mM. This was not unsurprising as a number of studies have shown that increasing the concentration of an initiator can have a positive effect initially, but higher concentrations resulted in a decrease in the yield of transcript.^{91,92} The results confirmed that the initiator without the PEG linker (**G4**) incorporated into the 44-mer transcript more efficiently, obtaining 3-fold more incorporated transcript. To ensure incorporation into **A4**, (**G4**) was chosen as the exemplar initiator to use in splicing assays. Interestingly, when run on the gel, the transcript with (**G4**) incorporated migrated further down the gel than the transcript without (Figure 3.14, a). This was unforeseen for two reasons: (i) the molecular weight of the transcript plus initiator is greater than the transcript alone and (ii) the accumulated negative charge on the transcript without initiator is greater by 2 due to the triphosphate on the 5' end than the equivalent transcript with initiator. Why the transcript with initiator (**G4**) migrates further than without is an anomaly and cannot be explained.

3.2.13 Click Reaction of Incorporated 44-mer Transcript and Various Azides

The 44 nucleotide transcript incorporating G-initiator (**G4**) was gel-purified and then reacted with (**43**) and **ESE1** in the presence of Cu(I) and TBTA. (**43**) is insoluble in aqueous solutions, and 20% DMSO was required within the reaction, therefore TBTA was used over THPTA. The reactions were run alongside controls and analysed by gel

electrophoresis (Figure 3.15). For the reaction with **ESE1** and the **44-mer**, two bands were observed, the lower band migrates to the same point on the gel as the controls and is therefore the **44-mer**, the other band is located considerably higher up the gel indicating that the length of the RNA is significantly longer, concluding the reactions had worked (Figure 3.15a). The reaction with the **44-mer** and **(43)** also exhibited two bands on the gel, although, this time the shift between the two are minimal (Figure 3.15b). These preliminary results demonstrated that the G-initiator (**G4**) incorporated into transcript and the subsequent click reactions were successful. However, the reactions did not go to completion and was not very efficient and therefore required optimization.

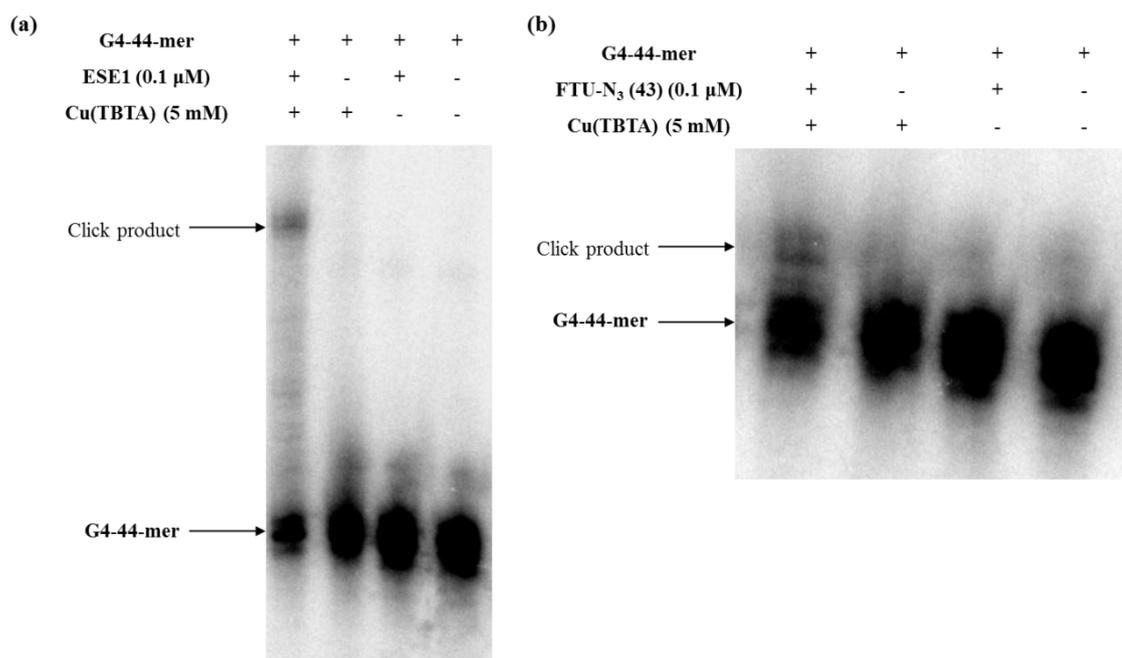


Figure 3.15 Click reaction between the 44mer incorporated with G-Initiator (G4) and an azide; (a) A polyacrylamide gel for the reaction with ESE1 and its controls, (b) a polyacrylamide gel for the reaction with FTU-N₃ and the corresponding controls.

3.2.14 Optimisation of the Click Reaction between 44 Nucleotide Transcript and 2'OMe ESE-Azides

The preliminary results with the hot transcript and **ESE1** confirmed the incorporation of the G-initiator (**G4**) and furthermore the ESE was ligated to the transcript inefficiently. As the reaction was inefficient optimisation of the conditions was necessary. Firstly the optimal time for the reaction was determined taking into account the highest yield with the least amount of degradation. A 12 hour time course was initially set up taking a sample every hour to analyse. The resultant gel showed that even after 1 hour the transcript had degraded immensely and after 4 hours the RNA had all but completely degraded. A one hour time course was then undertaken where samples of the reaction between **ESE1** and the **G4-44-mer** transcript were taken at 10, 20, 40 and 60 minutes. The results showed that 20 minutes was the optimum time before degradation became too much of a factor although only 12% was converted into product (Figure 3.16a). An equivalent 2'methoxy phosphorothioate ESE (**ESE3**) was also ligated to the **G4-44-mer** transcript and showed that 20 minutes was also the optimum time for the reaction before degradation played its part. Surprisingly, however, the product obtained with the phosphorothioate was substantially more than **ESE1** at 44% conversion (Figure 3.16b), 4 times more efficient than the 2'methoxy ESE (**ESE1**).

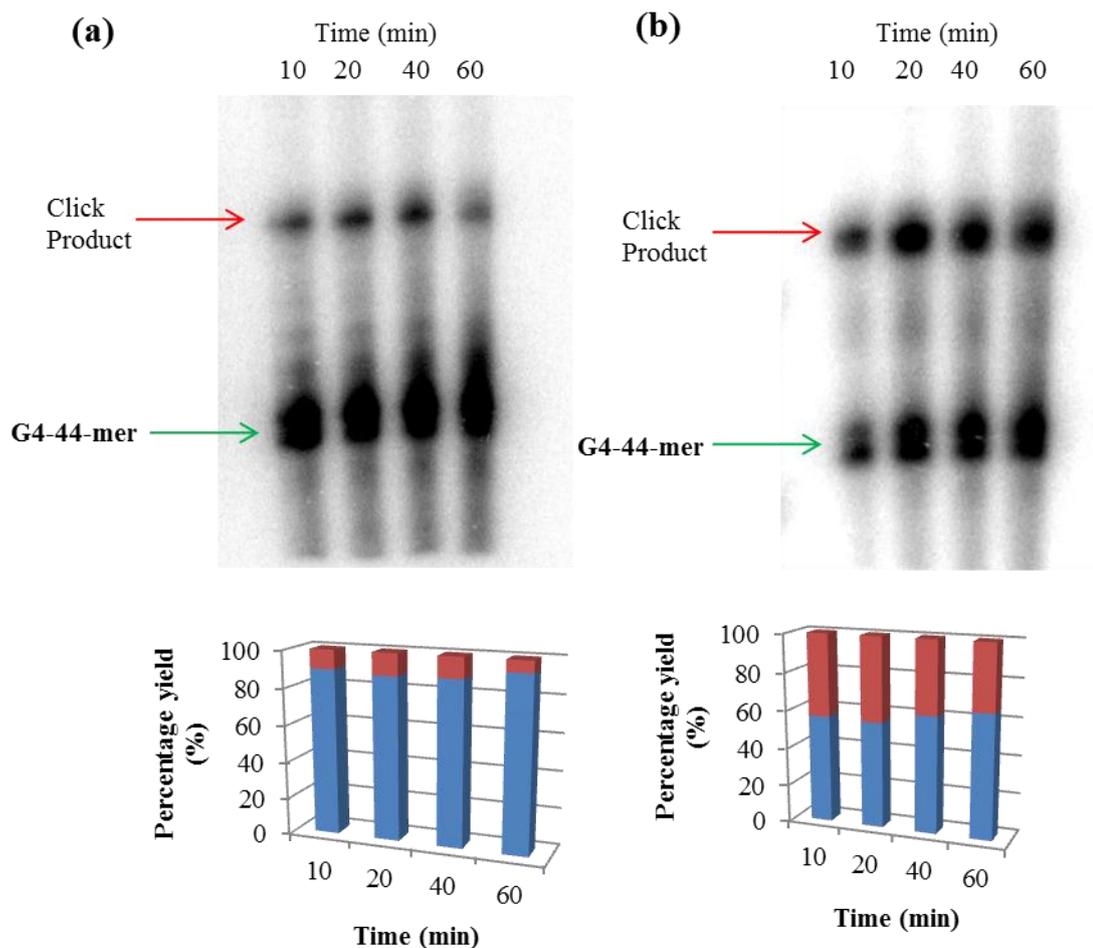


Figure 3.16 A 60 minute time course for the click reaction between ESE1 & ESE3 and 44-mer transcript incorporated with G-initiator (G4); (a) reaction between ESE1, maximum yield of 12% was obtained after 20 minutes, (b) reaction between ESE3, maximum yield of 44% was obtained after 20 minutes.

The optimum concentration of the **ESE1** was then established by testing the different concentrations ranging from 0.1 to 100 μM (Figure 3.17) This test showed that 10 μM was the optimum concentration, attaining approximately 40% yield (Figure 3.17b).

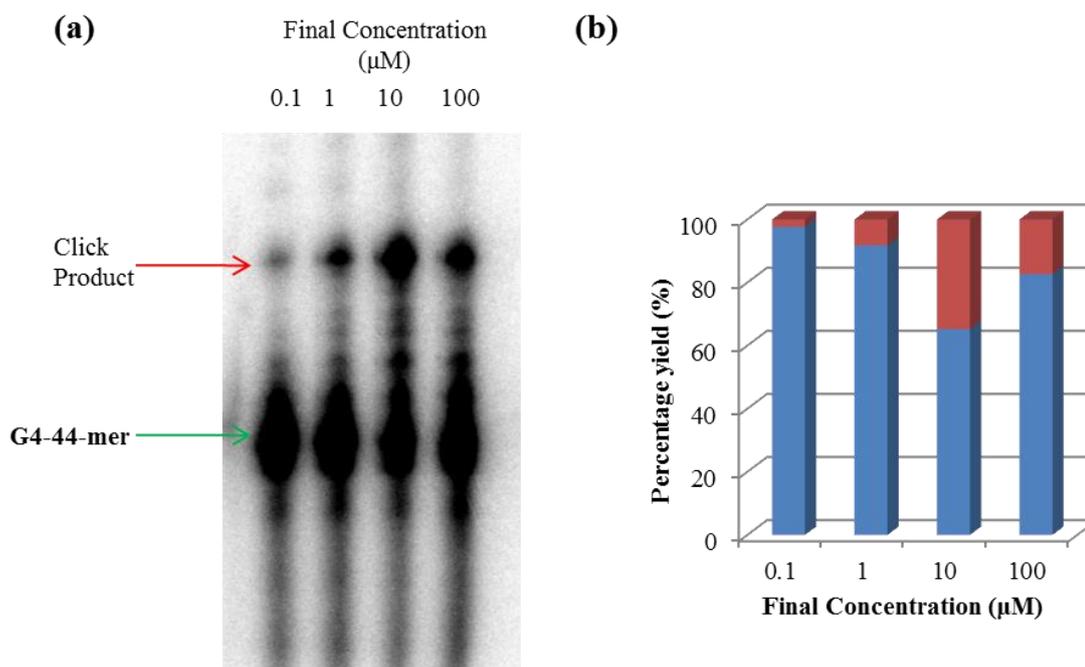


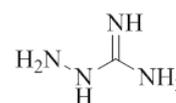
Figure 3.17 Concentration test for the click reaction between ESE1 and 44 nt transcript incorporated with G-initiator (G4); (a) polyacrylamide gel for the click reactions (b) a histogram showing the various yields obtained (%), maximum yield is obtained with the addition of 10 μM.

Degradation of the RNA at 20 minutes was relatively high and possibly contributes to the loss of product. RNaseOUT™ is a Recombinant Ribonuclease Inhibitor which is a potent inhibitor of RNases A, B and C that degrades the RNA. Therefore it was decided to add it into the reaction to help combat degradation. The results showed that it had a positive effect and degradation decreased considerably when you compare to the reaction without (Figure 3.18, lane 4).

Aminoguanidine was used by the Finn group to scavenge the reactive oxygen species of the ascorbate oxidation which may cause degradation of the RNA.¹²⁷ It was therefore tested in this reaction to determine if it had any effect. The results showed that aminoguanidine alone did inhibit degradation although not to the same extent as RNaseOUT™ (Figure 3.18, lane 2). A further reaction was undertaken adding both RNaseOUT™ and aminoguanidine (48). There was little difference between the

reaction with just RNaseOUT™ and the reaction with both. However for gel purification purposes it was decided for future reactions both (48) and RNaseOUT™ were used.

G4-44-mer	+	+	+	+
ESE1	+	+	+	+
Cu(I)THPTA	+	+	+	+
48	-	+	+	-
RNaseOUT™	-	-	+	+



48

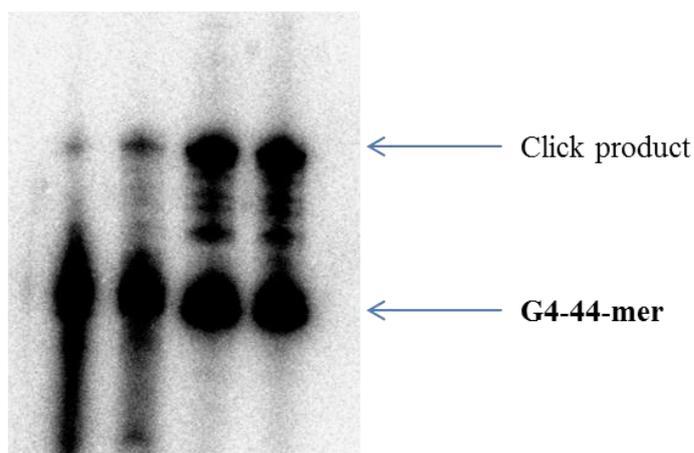


Figure 3.18 A polyacrylamide gel for the click reactions between ESE1 and 44-mer transcript incorporated with G-initiator (G4), aminoguanidine (48) and RNaseOUT™ have been added to prevent degradation of the RNA.

3.2.15 Investigation into the Action of the Phosphorothioate in the Click Reactions

It was shown in Section 3.2.14 that the reactions with the 2'-OMe phosphorothioates ESE (**ESE3**) and G4-44 gave considerably higher yields (44%) than the equivalent 2' OMe phosphate ESE (**ESE1**, 12%). The reason why the reaction was more efficient was unclear. Cu(I) ion is a soft metal and therefore prefers to bond with a soft ligand. The phosphorothioates of **ESE3** could possibly be acting as a ligand stabilising Cu(I),

accelerating the click reaction. Therefore it was decided to undertake a number of experiments to try and ascertain whether the phosphorothioate was acting as a ligand.

In order to do this a series of click reactions were undertaken whereby **G4-44-mer** was reacted with both **ESE1** and **ESE3** in the presence of a variety of copper species: Cu(I)THPTA, Cu₂(I)SO₄, Cu(II)SO₄ and no copper (Figure 3.19). The reactions between **G4-44-mer** and **ESE1** (Figure 3.19, lane1) and **ESE3** (Figure 3.19, lane 5) with the addition of Cu(I)THPTA attributed to two bands visible on the gel, the unreacted **G4-44-mer** and the clicked product which was expected. The addition of Cu₂(I)SO₄ to the reactions with **G4-44-mer** and **ESE1** (Figure 3.19, lane2) and **ESE3** (Figure 3.19, lane 6) shows complete degradation of the RNA, hence the Cu(I) without a ligand is highly toxic to the RNA. This also indicates that the phosphorothioate is not acting as a ligand and stabilising the Cu(I) complex as all the RNA had degraded and no product was visible (Figure 3.19, lane 6). As expected, only the **G4-44-mer** band was present in the reactions without any copper (Figure 3.19, lane 4 and lane 8). Surprisingly, the reactions with Cu(II)SO₄ for both ESEs (**ESE1** and **ESE3**) formed two bands when analysed by gel electrophoresis. The band with faster electrophoretic mobility was assigned as unreacted **G4-44-mer**, and the band with the slower electrophoretic mobility migrated to the same location as the clicked product observed in the reactions with Cu(I)THPTA (Figure 3.19, lane1 and lane 3), indicating that the click reaction had taken place in the presence of Cu(II)SO₄. This was unexpected considering that the click reaction should only take place in the presence of Cu(I) or in the presence of a highly strained alkyne or extremely high temperatures.

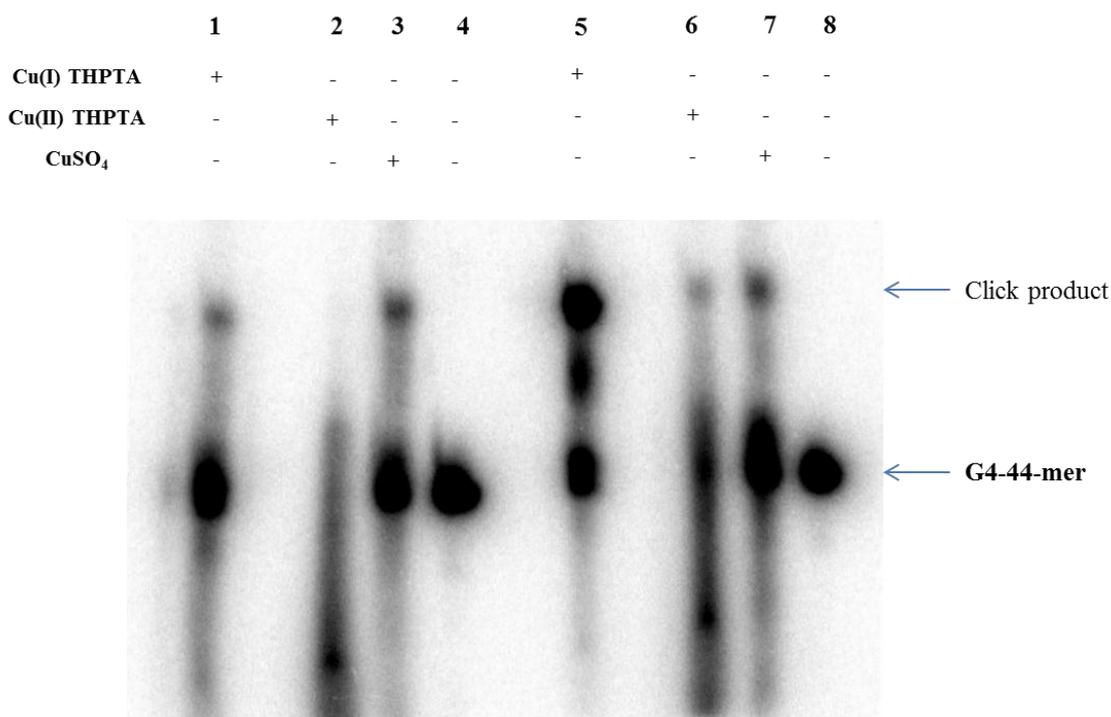


Figure 3.19 Polyacrylamide gel showing the reactions between G4-44 nucleotide transcript and ESE1 (lanes 1-4) and ESE3 (lanes 5-6) with a variety of copper species.

An additional reaction was undertaken with FTU-N₃ (**43**) and a 2' OMe phosphorothioate RNA modified with an alkyne group at the 5' end (RNA1, Appendix, Table 6.1) and a 2'OMe phosphate RNA modified with an alkyne group (RNA2, Appendix, Table 6.1) varying the copper species added: Cu(I)THPTA, Cu₂(I)SO₄, Cu(II)SO₄ and no Cu. The reactions were analysed by polyacrylamide gel and firstly imaged at 521 nm the emission wavelength of (**43**) (Figure 3.20). A band was seen for all the reactions which were determined to be (**43**). Additional bands were seen in the reaction between RNA1 and RNA2 and (**43**) with the addition of Cu(I)THPTA which was expected (Figure 3.20, lane 1 and lane 5). There was no reaction with Cu₂(I)SO₄ with either RNA1 or RNA2 (Figure 3.20, lane 2 and lane 6) nor was there a reaction with no Cu added (Figure 3.20, lane 4 and lane 8). The reaction with Cu(II)SO₄ saw a reaction between RNA2 (2'OMe phosphate RNA) and (**43**), (Figure 3.20, lane7) but not between RNA1 (2'OMe phosphorothioate RNA) and (**43**), (Figure 3.20, lane 3). The gel

was stained with SYBR gold and reimaged at 260 nm and bands were visible in all the lanes, the reactions with Cu(I)THPTA had gone to completion (Figure 3.20, lane 1 and lane5), but the reaction with Cu₂(I)SO₄ and RNA2 (Figure 3.20, lane 7) showed two bands implying the reaction had not gone to completion.

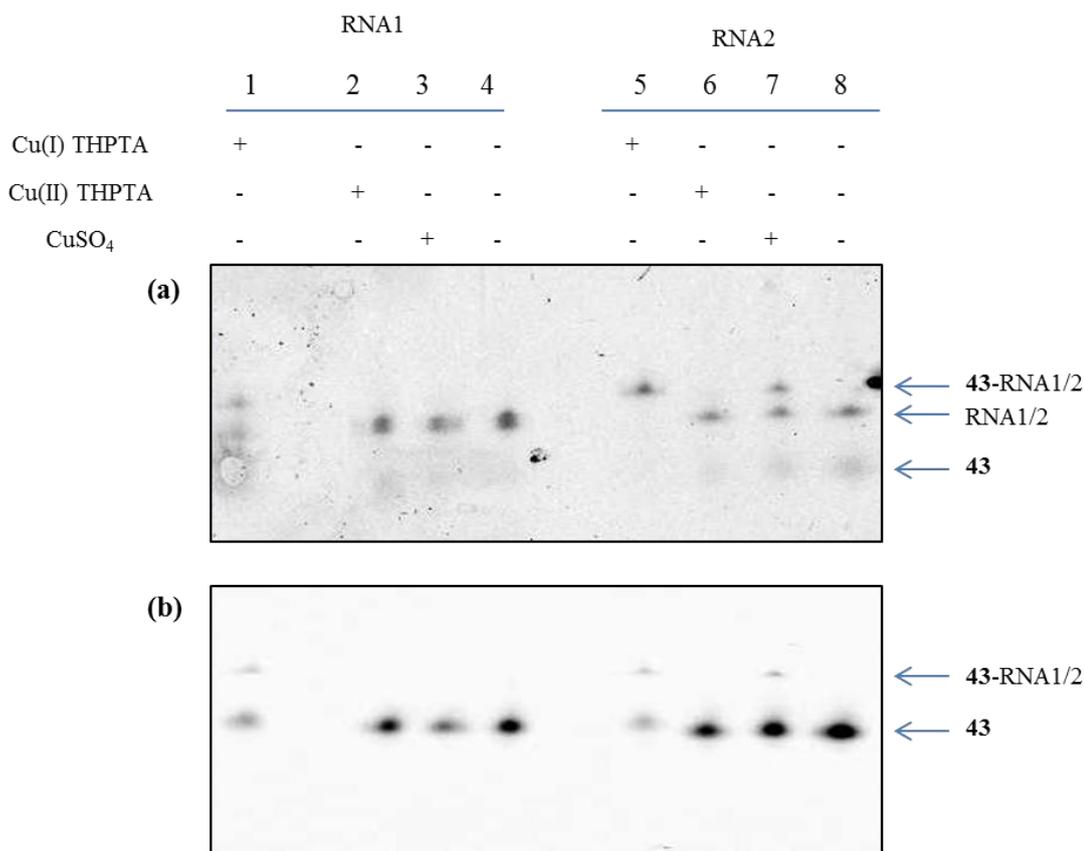


Figure 3.20 Polyacrylamide gels showing the reactions between FTU-N3 and RNA1 and RNA2 with a variety of Copper species. (a) gel stained with SYBR gold and imaged at 260 nm. (b) gel imaged at 521 nm for fluorescein (FTU).

As a result of this experiment, it did not appear that the phosphorothioates were acting as ligands; only RNA2 (2'OMe phosphate RNA) reacted in the presence of Cu(II)SO₄. These findings are extremely interesting but are unexplained. There are a number of possibilities why the Cu(II)SO₄ reaction is taking place. Firstly, (48) is added into the reaction, which may be acting as a reducing agent. However it does not explain why the RNA was not degraded in the same way as with Cu₂(I)SO₄, which was reduced by

sodium ascorbate in the experiment with the **G4-44-mer** transcript (Figure 3.19). Secondly, it could be due a small amount of Cu(I) being present in the batch of Cu(II)SO₄ and as it is added in excess, there may be enough Cu(I) present to catalyse the reaction. To determine why the reaction with Cu(II)SO₄ worked further analysis is required.

3.2.16 Incorporation of G-Initiator During Cold Transcription and Subsequent Ligations to Fluorophores

Along with incorporating the G-initiator (**G4**) into a hot transcript, it was also incorporated into a cold transcript for additional confirmation that incorporation had occurred. Cold transcription produces significantly more RNA than hot transcription by a factor of 1×10^4 at least. This could subsequently be used to ligate to fluorophore-azides (FTU (**43**) and Cy5 (**49**) which can be detected by their emission wavelengths).

3.2.16.1 FTU-Azide (43)

44-mer was initiated with G-initiator (**G4**) and reacted with an excess of (**43**) (Scheme 3.10) and analysed by gel electrophoresis. The gel was first imaged at 521nm (the emission wavelength for fluorescein (FTU)) and one band was observed on the gel in the lane with all the reactants in the mixture. No other bands were observed in any of the control lanes concluding the click ligation was successful (Figure 3.21a). The gel was then stained with SYBR gold and re-imaged at 260 nm, where bands were observed in all 4 lanes with the **44-mer** present (Figure 3.21b). Both the unincorporated (**44-mer**) and incorporated transcript (**G4-44-mer**) were observed in the controls (Figure 3.21b, lane 2-4) whereas a band for the FTU labelled transcript (**(43)-44-mer**) could be clearly seen along with the **44-mer** (Figure 3.21b, lane 1).

Scheme 3.10 Click reaction between G4-44 nucleotide transcript and FTU-azide (43).

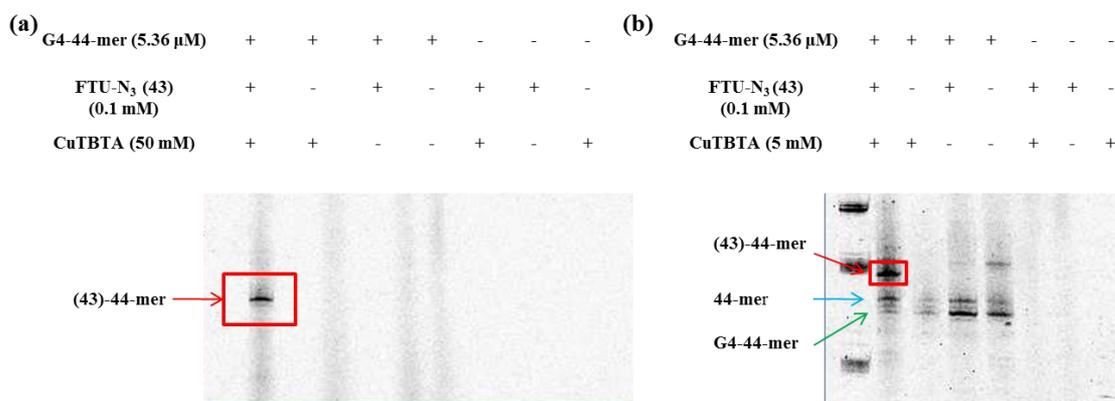
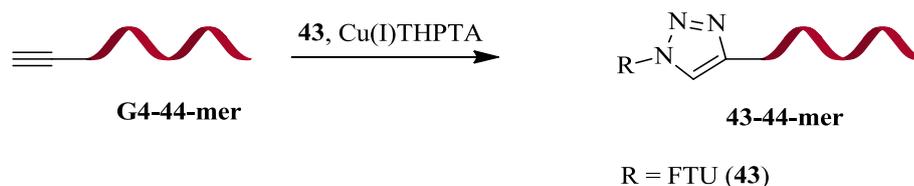


Figure 3.21 Click reaction between the cold transcription transcript incorporated with G-initiator (G4-44-mer) and (43); (a) polyacrylamide gel imaged at 521 nm, band highlighted is the click product, (b) same gel stained with SYBR gold and imaged at 260 nm, bands can be seen in all lanes with RNA present in the reaction, band highlighted is the click product.

3.2.16.2 Cy5 Azide (49)

Cy5-N₃ (49) was also click ligated to the 44mer in the presence of Cu(I) and TBTA (Scheme 3.11). (49) is extremely sensitive to photo-bleaching and therefore needed to be kept in the dark as much as possible. When imaged at 670 nm a band was visible for the reaction containing all the reagents but not in the controls (Figure 3.22b); to visualise the RNA the gel was stained with SYBR gold, and bands were visible for both the unincorporated and incorporated transcripts in the lanes for the controls (Figure 3.22a). Only one band is visible in lane 1 which migrated to the same point at the unincorporated 44-mer. Due to the band being visible at 670 nm it is possible that the

(49)-44-mer could migrate at the same rate as the unincorporated 44mer due to the increase in mass.

Scheme 3.11 Click reaction between the G4-44-mer nucleotide transcript and Cy-5 azide (49).

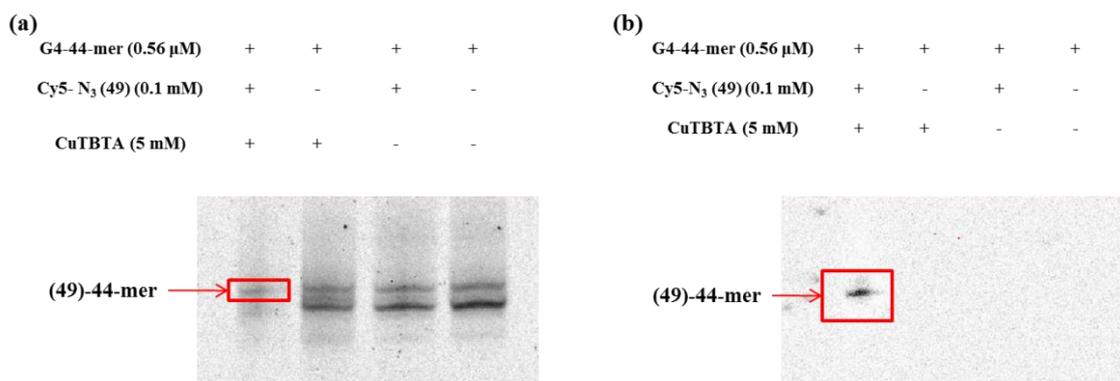
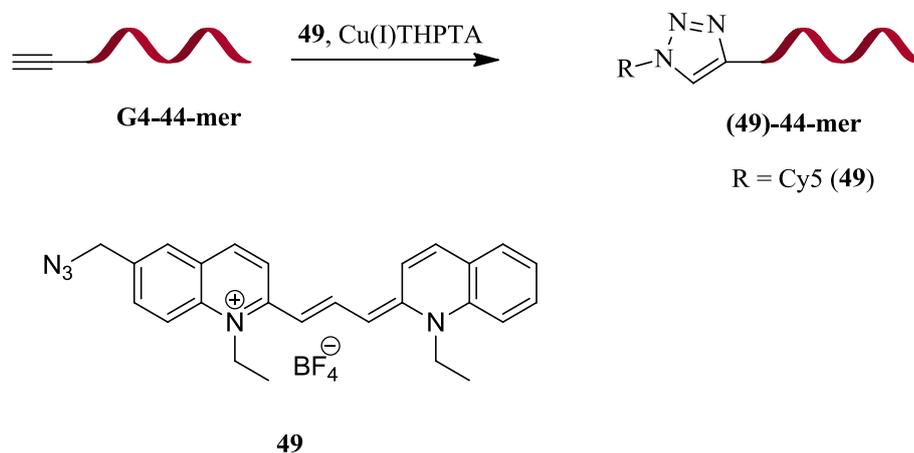


Figure 3.22 Click reaction with Cy5 azide (49) and cold transcription transcript incorporated with G-initiator (G4); (a) polyacrylamide gel stained with SYBR gold and imaged at 260 nm, band highlighted is the Cy5 labelled transcript ((49)-44-mer), (b) same gel imaged at 670 nm to detect, one band present for the reaction highlighted.

The reactions with the cold transcripts and the fluorophores (43) & (49) confirmed that the initiator had successfully incorporated into the 44mer and that fluorophores could be used as a means of detection in the future.

3.3 SUMMARY

The aim of this chapter was to design and incorporate G-initiators that could be incorporated into RNA enzymatically during transcription and in turn used to bioconjugate via click chemistry to a synthetic ESE which had been modified at the 3' end of the sequence. The G-initiators designed had an alkyne group at the 5' terminus attached to the guanosine via a linker of varying lengths of PEG. G-initiators (**G1**) and (**G4**) were successfully synthesised obtaining sufficient amounts to be used for transcription. (**G1**) and (**G4**) were both incorporated into a 44-mer transcript at various concentrations. However (**G4**) incorporated more efficiently than (**G1**) with 72% compared to 25% of total incorporation (Figure 3.14). This is most likely due to the additional PEG linker in (**G1**), either due to its flexibility or its increase in size. **ESE1** and **ESE11** were synthesised with a HEG linker and an amino modifier modification at the 3' end; the amine group was then NHS-coupled to (**47**), forming the ESE azide required for the click reactions.

The incorporation of (**G4**) into the **44-mer** was significantly higher yielding than (**G1**); it was therefore decided to proceed with (**G4**) for the bioconjugation reactions in order to obtain the tripartite RNA constructs with the utmost yield. The incorporated **44-mer** transcript was bioconjugated to **ESE1**, **ESE3**, FTU-N₃ (**43**) and Cy5-N₃ (**49**). Interestingly the reaction between the transcript and phosphorothioate ESE (**ESE3**) obtained a much higher yield than its equivalent **ESE1**.

Optimisation of the click reaction between transcript and **ESE1** was carried out in order to obtain the maximum yield possible with the least amount of degradation. Firstly the length of the reaction was investigated and it found that 20 minutes was the optimal

time for maximum yield before degradation became too significant. The concentration of the ESE was probed, the findings showed that a final concentration of 10 μM resulted in the highest overall yield over a range of concentrations, with approximately 40% click product obtained (Figure 3.17). In order to reduce degradation, (**48**) and RNaseoutTM were added to the reaction. Degradation was significantly reduced attributing to higher yields of click products (Figure 3.18). As a result both reagents were added to all subsequent click reactions, using the optimised conditions stated in Table 3.6.

Table 3.6 Optimised conditions for the click reaction between G-initiator (4**) incorporated transcript and ESE-N₃. Reactions were carried out at 40°C for 20 minutes.**

Reagent	Final Concentration
	-
ESE-N ₃	10 μM
Cu(I)THPTA	5mM
Aminoguanidine (48)	40mM
RNaseout TM	2% v/v

Optimisation of both the incorporation of the G-initiators (**G1/G4**) into the RNA transcript and the subsequent click reactions were essential. The results obtained in this chapter in relation to the incorporation and click ligation of G-initiator (**G4**) to **ESE1** and **ESE3** can be applied to the Ad1WW construct to determine whether or not the linker has an effect on the splicing pattern.

3.4 EXPERIMENTAL

3.4.1 General Procedures

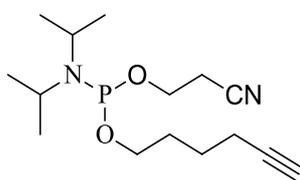
Reagents used in the solid phase synthesis of RNA were supplied by Link technologies and Glen Research. All other reagents were supplied by Sigma Aldrich, Acros Organics, Alfa Aesar and Fisher Scientific unless stated otherwise.

^1H , ^{13}C and ^{31}P NMR spectra were chronicled on a *Bruker DPX 300*, *Bruker DPX 400* or *Bruker DPX 500* spectrometer as stated, chemical shifts were recorded in parts per million (ppm, δ), J coupling constants are stated in Hertz (Hz), with multiplicities abbreviated as singlet (s), doublet (d), triplet (t), quartet (q), quintet (qu) and multiplet (m). Mass spectra were recorded using a micromass Quattra LC spectrometer for ESI, Voyager-DE STR using hydroxypicolinic acid matrix for MALDI-TOF, and Xevo QToF mass spectrometer for LC/MS and accurate mass. IR spectra were recorded using a Perkin Elmer FT-IR with ATR attachment.

Analytical and semi-preparative RP-HPLC were run using the Dionex Ultimate 3000, with the Dionex RF 2000 fluorescence detector. For analytical RP-HPLC a Phenomenex, Clarity 5 micron Oligo RP, 250 x 4.6 mm column was used. A standard 30 minute program using Buffer A (0.1M triethylammonium acetate pH7.2 in water) and Buffer B (0.1M triethylammonium acetate pH7.2 in 80:20 acetonitrile/water) was used for all RNA compounds analysed by RP-HPLC. The flow rate was 1 ml/minute with a gradient as follows: 0-4 minutes held at 5% buffer B, then a linear gradient from 5-50% buffer B for 20 minutes, 50-90% for a further 2 minutes, after 1 minute at 90%

buffer B, the gradient was returned to 5% buffer B over 1 minute and was then held at 5% buffer B for the remaining 2 minutes. For semi-preparative RP-HPLC a Phenomenex, Clarity 5 micron Oligo RP, 250 x 10 mm column was used with a flow rate of 4 ml/minute. A standard 32 minute program was used using the same buffers and gradient as the analytical program, holding at 5% buffer B for an additional 2 minutes at the end of the run.

3.4.2 Synthesis of Hexanyl Phosphoramidite (**50**)¹⁶⁰



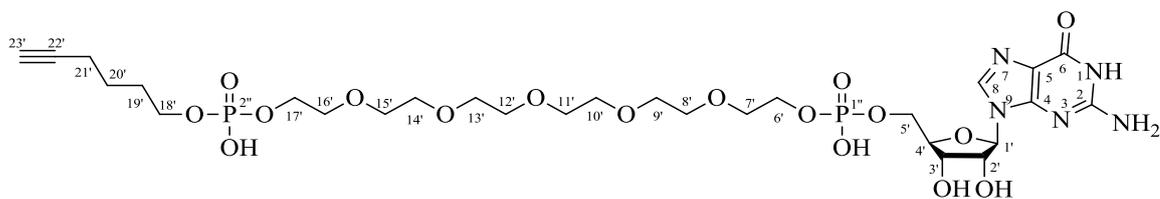
50

To a solution of 5-hexyn-1-ol (200 mg, 2.04 mmol, 225 μ l), and DIEA (1.3 eq, 2.65 mmol, 462 μ l) in dry DCM (2 ml) 2-cyanoethyl-N,N-diisopropyl-chlorophosphoramidite (1.5eq, 3.06 mmol, 683 μ l) was added under argon and left stirring for 2 hours. The mixture was then purified by flash column chromatography in 30% ethyl acetate in hexane, silica was deactivated with pyridine, 60 ml column volume fractions were collected and the solvent was removed under high vacuum resulting in (**50**), a colourless oil (0.422g, 77% yield).

¹HNMR (300MHz, C₆D₆), δ 1.22 (t/2d, J=6.48, 12H), δ 1.52-1.86 (m, 6H), δ 1.86 (t, J=2.65, 1H), δ 2.08 (dt, J=6.86 & 2.65, 2H), δ 3.30-3.50 (m, 2H), δ 3.52-3.78 (m, 2H).

³¹P NMR (300MHz, C₆D₆) δ 147.99.

3.4.3 Synthesis of Guanosine Initiator (G1)



Guanosine was modified at the 5' end by automated solid phase synthesis on an ABI DNA synthesiser. iPr-Pac-G RNA SynBase™ CPG 1000/110 (10 x 1 μmol), Spacer-CE Phosphoramidite 18 (HEG) (0.12 M) and 5'-hexanyl phosphoramidite (0.12 M) were prepared and attached to the machine ready for synthesis. When synthesis was complete the resin was removed from the column and put into a glass vial, 35% ammonium solution (2 ml) was added to the resin and heated overnight at 55°C to deprotect and cleave the G-initiator. The solution was removed from the resin and concentrated down to dryness and redissolved in H₂O (100 μl). The product was analysed and purified by reverse phase RP-HPLC and mass confirmed by MALDI. The purified product was then freeze-dried x3 and dissolved in H₂O (100 μl). The concentration was calculated to be 4.40 μmol (44% yield).

¹H NMR (500MHz, D₂O) δ1.48-1.55 (m, 2H, *H20'*). δ1.62-1.69 (m, 2H, *H19'*), δ2.16 (td, J=7.0, 2.7, 2H, *H21'*), δ2.27 (t, J=2.6, 1H, *H23'*), δ3.51-3.67 (m, 20H, *H7'-17'*), δ3.79-3.86 (m, 4H, *H6'* & *H18'*), δ3.89-3.93 (m, 2H, *H17'*), δ4.02 (t, J=3.9, 2H, *H5'*) 4.22-4.26 (m, 1H, *H4'*), δ4.43, (dd, J=3.7, 1H, *H3'*) δ4.74 (t, J=5.6, 1H, *H2'*), δ5.85 (d, J=5.9, 1H, *H1'*) δ8.02 (s, 1H, *H8*) (Appendix, Figure 6.8).

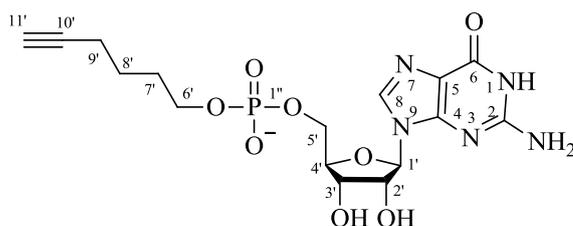
¹³C NMR (500MHz, D₂O) δ17.23 (CH₂, *C21'*), δ24.14 (CH₂, *C20'*) δ28.94 (CH₂, *C19'*) δ64.64 (CH₂-O-P, *C6'* & *C17'*), δ64.91 (CH₂-O-P, *C5'*), δ65.65 (CH₂-O-P, *C18'*),

δ 69.28-70.20 (CH₂-O, C7'-C17'), δ 70.42 (CH-OH, C3'), δ 73.38 (CH-OH, C2'), δ 83.67 (CH, C4'), δ 85.89 (C, C22'), δ 86.86 (O-CH-N, C1'), δ 116.29 (C, C5), δ 137.53 (N-C=N, C8), δ 151.89 (C, C4), δ 153.93 (C, C2), δ 158.88 (CO, C6) (Appendix, Figure 6.9).

³¹P{¹H} NMR (500MHz, D₂O) δ 0.28 (PO₄H, P1'') δ 0.73 (PO₄H, P2'') (Appendix, Figure 6.10).

LC/MS TOF MS ES+ calculated C₂₈H₄₇N₅O₁₇P₂ [M+H] = 788.2442, found 788.2545 (Appendix, Figure 6.11).

3.4.4 Synthesis of Guanosine Initiator (G4)



G4

The synthesis of G-Initiator was carried

out by solid phase synthesis as described in section 3.4.3, using a commercially available dmf-G RNA SynBase™ CPG 1000/110 (10 x1 μ mol) solid support. The final concentration of the G-Initiator product was calculated to be 2.61 μ mol (26% yield).

¹H NMR (500MHz, D₂O) δ 1.25-1.31 (m, 2H, H8'), 1.43-1.50 (qu, J=7.1, 2H, H7'), 1.95-2.00 (m, 2H, H9'), 2.17 (t, J=2.56, 1H, H11'), 3.58-3.70 (m, 2H, H6'), 3.97-4.00 (m, 2H, H5'), 4.22-4.26 (m 1H, H4') 4.43-4.46 (m, 1H, H3'), 4.75-4.79 (m, 1H, H2'), 5.85 (d, J=5.73 1H, H1'), 8.01 (s, 1H, H8) (Appendix, Figure 6.12).

^{13}C NMR (500MHz, D_2O) δ 17.11 (CH_2 , $C9'$), 24.04 (CH_2 , $C8'$), 28.89 (CH_2 , $C7'$), 64.85 (CH_2 , $C5'$), 65.61 (CH_2 , $C6'$), 69.00 (CH , $C11'$), 70.38 (CH , $C3'$), 73.20 (CH , $C2'$), 83.8 (CH , $C4'$), 86.78 (CH , $C1'$), 137.44 (CH , $C8$) (Appendix, Figure 6.13).

$^{31}\text{P}\{^1\text{H}\}$ NMR (500MHz, D_2O) δ 0.35 (s, PO_4H , PI') (Appendix, Figure 6.14).

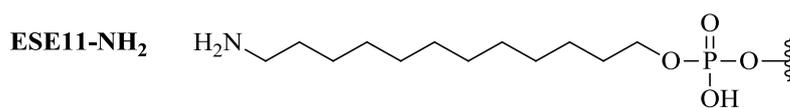
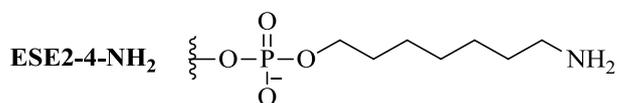
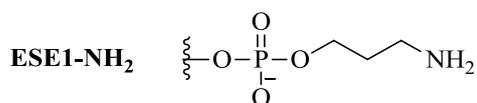
LC/MS TOF MS ES+ calculated $\text{C}_{16}\text{H}_{22}\text{N}_5\text{O}_8\text{P}$ $[\text{M}+\text{H}] = 444.1133$ found 444.1269 (Appendix, Figure 6.15), LC/MS TOF MS ES- calculated $\text{C}_{16}\text{H}_{22}\text{N}_5\text{O}_8\text{P}$ $[\text{M}-\text{H}] = 442.1133$ found 442.1125 (Appendix, Figure 6.16).

3.4.5 Synthesis of 2'Methoxy-ESE-Amines

All RNA was synthesised by automated solid phase using the ABI DNA synthesiser. Cleavage and deprotection was undertaken according to the manufacturer's instructions.¹⁵⁷

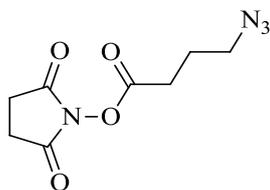
Table 3.7 Sequences for 2'-methoxy ESE-amines synthesised

Name [†]	3' and 5' Modified RNA Enhancer sequences synthesised (5'-3')
ESE1-NH₂	A ₀ G ₀ G ₀ A ₀ G ₀ G ₀ A ₀ C ₀ G ₀ G ₀ A ₀ G ₀ G ₀ A ₀ C ₀ G ₀ G ₀ A ₀ G ₀ G ₀ A ₀ C ₀ A ₀ -HEG-NH ₂
ESE2-NH₂	A ₀ G ₀ G ₀ A ₀ G ₀ G ₀ A ₀ C ₀ G ₀ G ₀ A ₀ G ₀ G ₀ A ₀ C ₀ G ₀ G ₀ A ₀ G ₀ G ₀ A ₀ C ₀ A ₀ -NH ₂
ESE3-NH₂	A _{os} G _{os} G _{os} A _{os} G _{os} G _{os} A _{os} C _{os} G _{os} G _{os} A _{os} G _{os} G _{os} A _{os} C _{os} G _{os} G _{os} A _{os} G _{os} G _{os} A _{os} C _{os} A _{os} -HEG-NH ₂
ESE4-NH₂	A _{os} G _{os} G _{os} A _{os} G _{os} G _{os} A _{os} C _{os} G _{os} G _{os} A _{os} G _{os} G _{os} A _{os} C _{os} G _{os} G _{os} A _{os} G _{os} G _{os} A _{os} C _{os} A _{os} -NH ₂
ESE11-NH₂	NH ₂ -HEG-A ₀ G ₀ G ₀ A ₀ G ₀ G ₀ A ₀ C ₀ G ₀ G ₀ A ₀ G ₀ G ₀ A ₀ C ₀ G ₀ G ₀ A ₀ G ₀ G ₀ A ₀ C ₀ A ₀



[†] ESE2-4 were synthesised by Andrew Perrett.

3.4.6 Synthesis of NHS-Azide (**47**)¹⁶¹



47

4-Azidobutanoic acid (0.296 g, 2.30 mmol) was dissolved in anhydrous dichloromethane under argon and cannulated into a flask containing N-hydroxysuccinimide (0.369 g, 3.95 mmol), and HOBt (0.426 g, 3.95mmol). The reaction was cooled to 0°C and left stirring under argon for 5 minutes. EDC (0.605 g, 3.95 mmol) was added to the mixture and left to warm to r.t. and left stirring overnight. The solution was concentrated down and the NHS-azide (**47**) was purified by flash column chromatography 1:1 petroleum ether 40:60 / ethyl acetate $r_f = 0.31$. The fractions were combined and the solvent was evaporated off to obtain (**47**), a colourless oil (0.217 g, 42% yield).

¹HNMR (300MHz, CDCl₃) δ2.02 (qu, J=6.89, 2H), δ2.72 (t, J=7.18, 2H), δ2.84 (s, 2H), δ3.44 (t, J=6.55, 2H), NMR also shows a trace of starting material still present.

¹³CNMR (300MHz, CDCl₃) δ24.23, δ25.57, δ28.10, δ49.97, δ167.90, δ168.97.

3.4.7 NHS-Azide (**47**) Coupling to ESE-Amines

RNA-NH₂ solution was concentrated down to dryness and resuspended in a solution of NaHCO₃ (25 mM, 90 μl). A solution of NHS-N₃ (**47**) (100 eq, 210 μl) in acetonitrile was added to the RNA-NH₂, mixed thoroughly and left shaking overnight. An additional amount of NHS-N₃ (100 eq, 20-40 μl) was added to the solution and left shaking for a further 5 hours to ensure the reaction had gone to completion. The

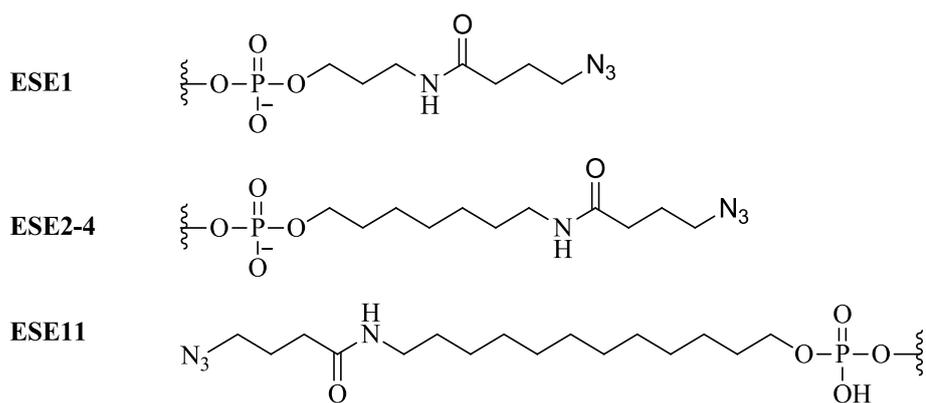
solution was removed in a vacuum concentrator, and resuspended in 100 μ l of water. The product was purified and desalted by a GE healthcare Nap-25 column, freeze-dried and dissolved in 100 μ l of water.

Table 3.8 Concentration of ESE obtained for the reaction between ESE-NH₂ and NHS-N₃, (determined by UV at 260 nm)

ESE	Number of moles of ESE-NH₂ used (μmol)	Concentration of ESE obtained (mM)
ESE11	0.019	0.92
ESE1	0.013	0.99

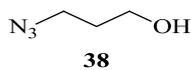
Table 3.9 Sequences for 2'-methoxy ESE-azides synthesised

Name [‡]	3' and 5' Modified RNA Enhancer sequences synthesised (5'-3')
ESE1	A ₀ G ₀ G ₀ A ₀ G ₀ G ₀ A ₀ C ₀ G ₀ G ₀ A ₀ G ₀ G ₀ A ₀ C ₀ G ₀ G ₀ A ₀ G ₀ G ₀ A ₀ C ₀ A ₀ -HEG-N ₃
ESE2	A ₀ G ₀ G ₀ A ₀ G ₀ G ₀ A ₀ C ₀ G ₀ G ₀ A ₀ G ₀ G ₀ A ₀ C ₀ G ₀ G ₀ A ₀ G ₀ G ₀ A ₀ C ₀ A ₀ -NH ₂
ESE3	A _{os} G _{os} G _{os} A _{os} G _{os} G _{os} A _{os} C _{os} G _{os} G _{os} A _{os} G _{os} G _{os} A _{os} C _{os} G _{os} G _{os} A _{os} G _{os} G _{os} A _{os} C _{os} A _{os} -HEG-N ₃
ESE4	A _{os} G _{os} G _{os} A _{os} G _{os} G _{os} A _{os} C _{os} G _{os} G _{os} A _{os} G _{os} G _{os} A _{os} C _{os} G _{os} G _{os} A _{os} G _{os} G _{os} A _{os} C _{os} A _{os} -N ₃
ESE11	N ₃ -HEG-A ₀ G ₀ G ₀ A ₀ G ₀ G ₀ A ₀ C ₀ G ₀ G ₀ A ₀ G ₀ G ₀ A ₀ C ₀ G ₀ G ₀ A ₀ G ₀ G ₀ A ₀ C ₀ A ₀



[‡] ESE2-4 were synthesised by Andrew Perrett.

3.4.8 Synthesis of 3-Azido-1-Propanol (38)¹²⁷

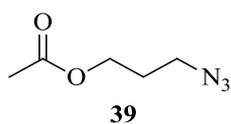


3-Iodopropanol (4.86 g, 26.1 mmol, 2.5 ml), and sodium azide (16.9 g, 261 mmol) were dissolved in water (150 ml) and refluxed at 90°C overnight. The mixture was extracted with dichloromethane (3 x 100 ml). The combined organic layers were dried over MgSO₄, filtered and concentrated down obtaining 3-azido-1-propanol as a pale yellow oil (2.60 g, 99% yield).

¹HNMR (400MHz, CDCl₃) δ 3.75 (q, J=5.6, 2H, H1), δ3.45 (t, J=6.6 2H, H3), δ1.85 (qu, J=6.21, 2H, H2), δ1.5 (t, J=5.16, 1H, OH), additional peaks are due to traces of water and dichloromethane.

IR broad stretch at 3343.5 cm⁻¹ (OH); sharp peak at 2089.2 cm⁻¹ (N₃).

3.4.9 Synthesis of 3-Azido-propyl Acetate (39)¹²⁷



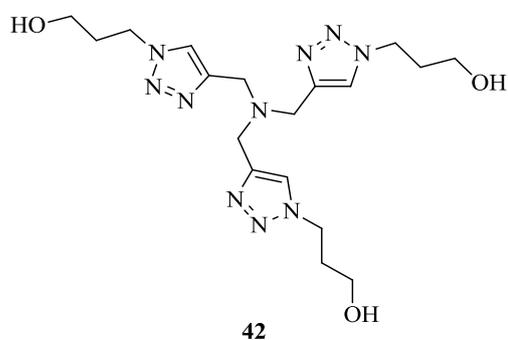
3-Azido-1-propanol (2.6 g, 25.7 mmol), acetic anhydride (24.7 mmol, 2.43 ml) and triethylamine (25.7 mmol, 3.6 ml) were dissolved in dichloromethane (150 ml) and left stirring at r.t overnight. After 20 hours the reaction mixture was then concentrated down to half the reaction volume and left refluxing at 40°C overnight. Additional acetic anhydride (0.2 eq, 5.1 mmol, 0.48 ml) and triethylamine (0.2 eq, 5.1 mmol, 0.72 ml) were added to the mixture and refluxed 40°C for a further 48 hours. The reaction mixture was then washed with a 0.1M sodium hydroxide (x2), followed by 0.1M HCl (x2) and finally with brine (x1). The organic

layer was dried with MgSO_4 , filtered and concentrated down resulting in a yellow oil (2.95 g, 80% yield).

$^1\text{H NMR}$ (400MHz, CDCl_3) δ 4.17 (t, $J=6.30$, wa 2H), δ 3.41 (t, $J=6.64$, 2H), δ 2.1 (s, 3H), δ 1.93 (qu, $J=6.49$, 2H).

IR sharp peak 2094.03cm^{-1} (N_3), sharp peak 1736.17 (C=O).

3.4.10 Synthesis of Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA (42))¹²⁷



To a solution of 3-azido-propyl acetate (2.54 g, 17.8 mmol) and tripropargylamine (3.55 mmol, 0.480 ml) in acetonitrile (100 ml), 2,3,5-collidine (3.55 mmol, 0.461 ml) was followed by copper(I) trifluoromethanesulfonate (5 mol%) turning the solution green. The reaction mixture was left stirring at r.t. for 2 days, the reaction had not gone to completion, additional copper(I) trifluoromethanesulfonate (20 mol%) was added and stirred at 40°C overnight. The solution was concentrated down to give a blue/green solid (3.23 g). This was treated with ammonia in methanol (2M, 100 ml) and stirred overnight at 40°C . The solution was concentrated down and then dissolved in water (100 ml) and washed with dichloromethane. The aqueous layer was collected and the water was evaporated off to give a blue/green solid (2.19 g). The copper ions were removed by passing the complex dissolved in 40ml of water through a DOWEX 50 treated with HCl ion exchange

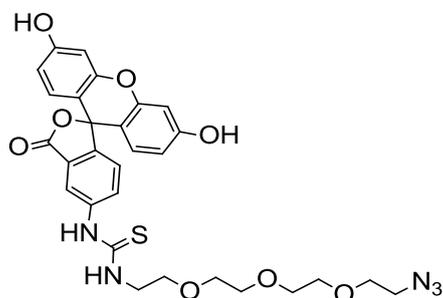
column. The extracted THPTA was concentrated down and the residue dried under high vacuum. Acetonitrile was added to the residue and sonicated to remove any impurities, the off white solid was filtered, dried and collected (0.79 g, 51% yield).

$^1\text{H NMR}$ (400MHz, DMSO-d_6) δ 8.02 (s, CH), δ 4.65 (t, $J=4.97$, 3H, OH) δ 4.41 (t, $J=7.10$, 6H, CH_2), δ 3.63 (s, 6H, CH_2), δ 3.40 (q, $J=5.86$, 6H, CH_2), δ 1.97 (qu, $J=6.56$, 6H, CH_2).

$^{13}\text{C NMR}$ (400MHz, DMSO-d_6) δ 143.35 (CO), δ 123.91 (CH), δ 57.44 (CH_2), δ 47.07 (CH_2), δ 46.47 (CH_2), δ 32.95 (CH_2).

MS ESI/TOF ES+ calculated $\text{C}_{18}\text{H}_{30}\text{N}_{10}\text{O}_3$ $[\text{M}+\text{H}] = 435.50$, found 435.

3.4.11 Synthesis of Azide modified Fluorescein Thiourea (FTU- N_3 (43))



43

To a solution of fluorescein isothiocyanate (100 mg, 0.26 mmol) in DMF (1500 μl), O-(2-aminoethyl)-O'-(2-azidoethyl)diethylene glycol (0.33 mmol, 66.25 μl) and DIEA (0.26 mmol, 44.7 μl) were added, the mixture was then sonicated and left shaking at r.t. overnight. The DMF was removed in a vacuum concentrator at 45°C until dry. The product was purified by flash column chromatography with a gradient of 10%-25% methanol in chloroform. The relevant

fractions were combined and concentrated down obtaining FTU-N₃ a sticky fluorescent yellow gum (0.155 g, 70% yield).

¹HNMR (500MHz, MeOD) δ3.32-3.34 (m, 4H), δ3.62-3.75 (m, 13H), δ3.84 (bs, 2H), δ6.55-6.57 (dd, J=8.72, 2.43, 2H), δ6.69-6.71 (dd, J=8.65, 2.40, 4H), δ7.17-7.19 (dd, J=8.2, 0.45, 1H), δ7.80-7.81 (d, J=7.85, 2H) δ8.17 (d, J=1.8, 1H). Additional peaks are from DMF and impurities.

¹³CNMR (500MHz, MeOD) δ45.54 (CH₂), δ49.87 (CH), δ51.76 (CH₂), δ71.1-71.7 (CH₂), δ103.55 (CH), δ111.53 (C), δ113.66 (CH), δ119.75 (CH), δ125.74 (CH), δ130.35 (CH), δ132.42 (CH), δ142.48 (C), δ154.21 (C), δ161.48 (C), δ164.89 (C), δ171.18 (C), δ182.94 (C). Additional peaks are seen between δ11-40 which are from DMF and impurities.

MS ESI/TOF ES+ calculated C₂₉H₂₉N₅O₈S [M+H] = 607.63, found 608.

MS ESI/TOF ES- calculated C₂₉H₂₉N₅O₈S [M+H] = 607.63, found 606.

3.4.12 Radioactive Transcription for Incorporation of G Initiator (G1) into Ad1WW (A4)

Transcription of Ad1WW (**A4**) whereby a G-Initiator (**G1**) was incorporated at the 5' end of the transcript was undertaken. Transcription was performed as described in section 2.4.3, replacing 0.2 U diguanosine triphosphate with 10mM G-Initiator (**G1**). The molarities of the transcripts were calculated using a scintillation counter.

3.4.13 Biotin Pull Down Assay

NeutrAvidin Agarose Beads (2-10 eq, 20 µg/ml Biotin capacity) were added to an eppendorf and blocked with tRNA (10 mg/ml) in wash buffer and left tumbling for 2 hours, 4°C. The beads were washed with wash buffer x3. The click reaction product (**45**) was then added to the blocked beads along with wash buffer and left to tumble for 1 hour 4°C. The beads were washed until no radioactivity was present in the negative controls. The beads (10-20 µl) were then added to formamide dyes and heated at 95°C for 25 minutes, then loaded onto a 6% denaturing polyacrylamide gel.

3.4.14 Radioactive Transcription Optimisation for the Incorporation of G-Initiators (G1 & G4) into 44 Nucleotide RNA Transcript

The transcriptions were carried out as stated in section 2.x with a reaction volume of 2.5 µl, replacing the NTP mix concentration (1 mM GTP, 4 mM ATP, 4 mM CTP, 4 mM UTP). The 0.2 U diguanosine triphosphate was replaced with G-Initiators (**G1** & **G4**) at a variety of concentrations (0-5 mM). The reactions were then incubated for 4 hours and then run on a denaturing polyacrylamide gel, dried and exposed to a phosphor screen for analysis. Subsequent transcriptions with G-initiator (**G4**) were then carried out using the optimum concentration (4 mM) and purified by gel electrophoresis.

3.4.15 Cold Transcription of into 44 Nucleotide RNA Incorporating G-Initiator (G1)

A reaction mixture containing cold transcription buffer (10 µl), DTT (10 µl, 0.1mM) NTP's (20 µl, 5 mM GTP, 20 mM ATP, 20 mM CTP, 20 mM UTP), PCR fragment (1 ng/µl, 2.5 µl), T7 polymerase (5 µl), RNaseout (5 µl, Invitrogen), G-Initiator (**G1**) (20

μl , 20 mM) made up to 100 μl with H_2O . The reaction mixture was incubated at 37°C for 4 hours. The transcript was then purified using GE healthcare S-300 column as per manufacturer's instructions, phenol/chloroform extracted and ethanol precipitated, resuspending the RNA in H_2O .

3.4.16 Click Reactions between RNA-Alkynes and Cargo-Azides using Cu(I)TBTA Catalyst, (Table 3.10)

RNA-Alkyne and Cargo- N_3 was added to an eppendorf tube and mixed. CuBr (1-2 mg, 7-14 μmol) was dissolved in TBTA (0.05 M). Cu(I) TBTA (5 mM) was added to the reaction mixture, vortexed briefly. The reaction was left shaking at 400 rpm at 40°C for 20-60 minutes.

3.4.17 Click Reactions between RNA-Alkynes and Cargo-Azides using Cu(I)THPTA Catalyst, (Table 3.11)

RNA-Alkyne and Cargo- N_3 was added to an eppendorf tube and mixed. CuSO_4 (1-2 mg, 4-8 μmol) was dissolved THPTA solution (0.05 M). Sodium ascorbate (3 eq, 12-24 μmol) was added to reduce the Cu(II) to Cu(I). The Cu(I)THPTA solution (5 mM) was added to the reaction mixture, vortexed briefly. The reaction was left shaking at 400 rpm at 40°C for 10-60 minutes.

3.4.18 Click Reactions between (G1) and Fluorophore-Azides (33)/(43) using Finn Method¹²⁷

To a solution of (G1) (0.01 mM) in HEPES buffer (100 mM), (33)/(43) (0.02 mM) was added followed by a pre-mixed CuSO_4 (0.1 mM) and THPTA (0.5 mM) and (48) (5

mM). Finally sodium ascorbate (5 mM) was added to the reaction mixture, which was then mixed thoroughly and left reacting for 1 hour at r.t./40°C. The reactions were analysed by RP-HPLC.

3.4.19 Click Reactions Testing the Copper Catalysts (Table 3.12)

To a solution of RNA-alkyne, **ESE1** (10 μ M) / **(43)** (0.1 mM) was added, followed by Cu catalyst (5 mM), RNaseout™ (2% v/v) and **(48)** (40 mM), the reactions were left shaking 400 rpm, at 40°C for 20 minutes. The reactions were analysed by polyacrylamide gel electrophoresis.

3.4.20 Buffers

- RNA Wash Buffer: 100 mM NaCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA.

Table 3.10 Click chemistry reactions using Cu(I)TBTA catalyst (5 mM) at 40°C, shaking at 400 rpm for 20-60 minutes.

Reaction	Alkyne	Concentration	Azide	Concentration	Solvent
1	G1	12 μ M	33	0.6 mM	30% v/v DMSO in H ₂ O
2	G1	26 μ M	33	0.6 mM	30% v/v DMSO in H ₂ O
3	G1	26 μ M	35	0.6 mM	30% v/v DMSO in H ₂ O
4	G1	26 μ M	33	0.6 mM	30% v/v DMSO in H ₂ O
5	G1	26 μ M	43	0.6 mM	30% v/v DMSO in H ₂ O
6	G1-A4	0.3 nM	35	5 nM	20% v/v DMSO in H ₂ O
7	G4	16 μ M	43	0.6 mM	30% v/v DMSO in H ₂ O
8	21-mer	5 μ M	ESE1	5 μ M	H ₂ O

9	21-mer	5 μ M	ESE9	5 μ M	H ₂ O
10	21-mer	15 μ M	ESE1	15 μ M	H ₂ O
11	21-mer	15 μ M	ESE9	15 μ M	H ₂ O
12	21-mer	5 μ M	ESE1	15 μ M	H ₂ O
13	21-mer	5 μ M	ESE9	15 μ M	H ₂ O
14	21-mer	15 μ M	ESE1	5 μ M	H ₂ O
15	21-mer	15 μ M	ESE9	5 μ M	H ₂ O
16	G4-44-mer	0.3-1 nM	43	0.1 μ M	H ₂ O
17	G4-44-mer	0.3-1 nM	ESE1	0.1 μ M	H ₂ O
18	G4-44-mer	5.36 μ M	43	0.1 mM	20% v/v DMSO in H ₂ O
19	G4-44-mer	0.56 μ M	49	0.1 mM	20% v/v DMSO in H ₂ O

Table 3.11 Click chemistry reactions using Cu(I)THPTA catalyst in H₂O, at 40°C, 400 rpm for 10-60 minutes.

Reaction	Alkyne	Concentration	Azide	Concentration	Additional reagents/conditions
1	G1	26 μM	33	0.6 mM	-
2	G1	26 μM	43	0.6 mM	-
3	G1-A4	0.3 nM	35	5 nM	20% v/v DMSO in H ₂ O
4	DNA 21-mer	5 μM	ESE1	15 μM	-
5	DNA 21-mer	5 μM	ESE9	15 μM	-
6	DNA 21-mer	15 μM	ESE1	5 μM	-
7	DNA 21-mer	15 μM	ESE9	5 μM	-
8	G4	50 μM	ESE1	15 μM	-
9	G4-44-mer	0.3-1 nM	ESE1	1 μM	-

10	G4-44-mer	0.3-1 nM	ESE3	1 μ M	-
11	G4-44-mer	0.3-1 nM	ESE1	0.1 μ M	-
12	G4-44-mer	0.3-1 nM	ESE1	1 μ M	-
13	G4-44-mer	0.3-1 nM	ESE1	10 μ M	-
14	G4-44-mer	0.3-1 nM	ESE1	100 μ M	-
15	G4-44-mer	0.3-1 nM	ESE1	10 μ M	2% v/v RNaseout™
16	G4-44-mer	0.3-1 nM	ESE1	10 μ M	2% v/v RNaseout™, (48) 40 mM
17	G4-44-mer	0.3-1 nM	ESE1	10 μ M	(48) 40 mM

Table 3.12 Click Reactions for the varying copper catalysts in the presence of 2% v/v RNaseout™ and (48) 40 mM at 40°C, 400 rpm for 20 minutes.

Reaction	Alkyne	Concentration	Azide	Concentration	Copper catalyst	Concentration
1	G4-44-mer	0.3-1 nM	ESE1	10 μM	Cu(I)THPTA	5 mM
2	G4-44-mer	0.3-1 nM	ESE1	10 μM	Cu ₂ SO ₄	5 mM
3	G4-44-mer	0.3-1 nM	ESE1	10 μM	CuSO ₄	5 mM
4	G4-44-mer	0.3-1 nM	ESE3	10 μM	Cu(I)THPTA	5 mM
5	G4-44-mer	0.3-1 nM	ESE3	10 μM	Cu ₂ SO ₄	5 mM
6	G4-44-mer	0.3-1 nM	ESE3	10 μM	CuSO ₄	5 mM
7	RNA1	5.36 μM	43	0.1 mM	Cu(I)THPTA	5 mM
8	RNA1	5.36 μM	43	0.1 mM	Cu(II)THPTA	5 mM
9	RNA1	5.36 μM	43	0.1 mM	CuSO ₄	5 mM

10	RNA2	5.36 μ M	43	0.1 mM	Cu(I)THPTA	5 mM
11	RNA2	5.36 μ M	43	0.1 mM	Cu(II)THPTA	5 mM
12	RNA2	5.36 μ M	43	0.1 mM	CuSO ₄	5 mM

4 SPLICING OF THE TRIPARTITE TRANSCRIPTS

4.1 INTRODUCTION

The Adenovirus WW construct described in Chapter 2 (Section 2.2.3) was established as a suitable transcript to ascertain how ESEs exert their effects on the 5' splice sites (Figure 2.7). Splicing of this Ad1WW construct produced 2 mature RNA spliced products due to the presence of the two alternative 5' splice sites located in exon 1. The upstream splice site produces the shorter upstream mRNA isoform whereas splicing at the downstream splice site produces the much longer downstream mRNA isoform (Figure 2.7a). The natural Ad1WW transcripts (**A2** and **A4**, Figure 2.7b) splice at both splice sites, producing the two mRNA isoforms with more upstream mRNA generated than downstream mRNA (Figure 2.9a & 2.9c), but the addition of an ESE sequence upstream of the 5' splice site (**A3** transcript, Figure 2.7b) shifts the splicing pattern from the downstream splice site to the upstream splice site producing predominantly upstream mRNA (Figure 2.9b). Thus an ESE exerts its effect by preferentially selecting the upstream splice site (i.e. the proximal 5' splice site).

4.1.1 Aim of Chapter 4

The aim of this chapter is to determine how ESEs exert their effects in stimulating splicing. As discussed in Section 1.4.1.2.3, the predominant model is that there is 3-dimensional diffusion, whereby an SR protein binds to the ESE and loops around interacting with splicing factors U1 and U2 snRNP (Figure 1.7a). There are a number of alternatives, of which the most often described is a 1-dimensional propagation model,

where an SR binds to an ESE and cooperatively binds to RNA binding proteins across the RNA stabilising the U1 snRNP and U2 snRNP (Figure 1.7b).

In order to test the first mechanism by inserting a flexible non-RNA linker between the ESE and its target splice site, the G-initiator (**G4**) could be incorporated into the 5' end of the **A4** transcript and ligated via copper click chemistry to ESEs modified with a HEG linker and an azide group at the 3' end as developed in Chapter 3 (Figure 4.1). The tripartite ESE-conjugated constructs will then be spliced to ascertain their splicing preference. If splicing occurs preferentially at the upstream splice site it would imply that the ESE is exerting its effects following the looping mechanism (Section 1.4.1.2.3) because the ESE is still able to interact with the U1 snRNP at the upstream 5' splice site stimulating splicing. However, if splicing at the upstream 5' splice site is inhibited it would suggest that the mechanism does not involve looping and that the signal from the SR proteins and RNA binding proteins cannot cross the PEG linker.

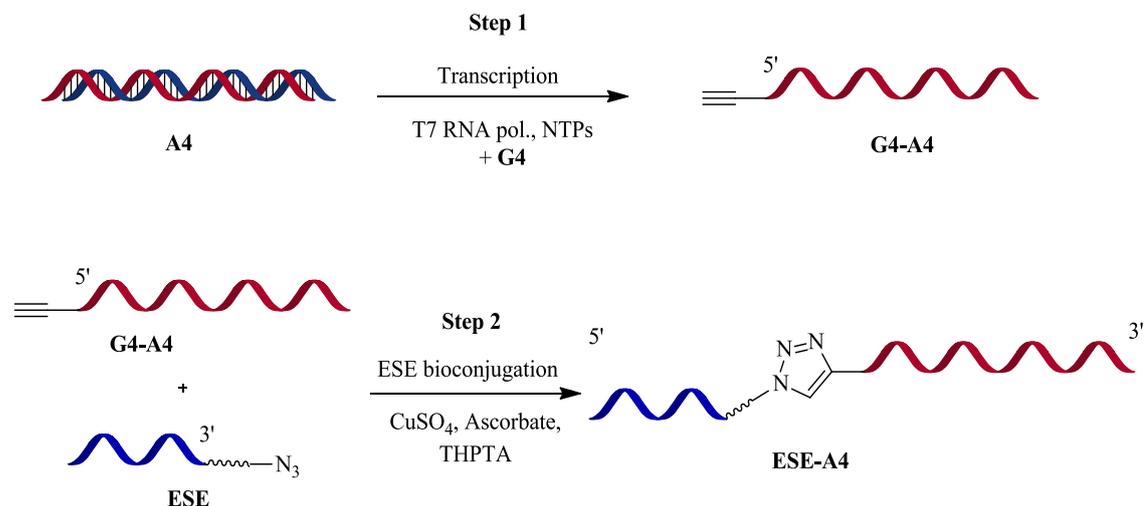


Figure 4.1 Diagrammatic representation of the incorporation of G4 into A4 during transcription (step 1) and the subsequent bioconjugation reaction (step 2).

4.2 RESULTS AND DISCUSSION

4.2.1 Incorporation of G-Initiator (G4) into Ad1WW Transcript (A4)

G-initiator (**G4**) was incorporated into the short 44 nucleotide transcript and bioconjugated by click chemistry to **ESE1** and **ESE3** (Section 3.2.12). The next step was to examine the incorporation into the Ad1WW transcript (**A4**) and to subsequently bioconjugate the ESEs (**ESE1** and **ESE3**). G-initiator (**G4**) was incorporated into the transcript using the optimized conditions determined in Section 3.2.14; the m⁷G cap was replaced with 0.4 mM of (**G4**), alongside 0.1 mM GTP and 0.4 mM ATP, CTP & UTP and was radiolabelled with [α -³²P] GTP. **A4** is 315 nucleotide in length, approximately 7 times longer than the 44-mer, meaning that it is difficult to visualise the 1 base difference between the G-initiator (**G4**) incorporated **A4** (**G4-A4**) and **A4** by gel electrophoresis. Therefore determining the amount of incorporation of (**G4**) was not viable nor was purifying the incorporated **A4** (**G4-A4**) from pre-mRNA lacking the alkyne modification of (**G4**).

In order to determine incorporation of G-initiator (**G4**) had taken place the transcript was reacted with **ESE1** and **ESE3** using the optimized conditions described in Section 3.3 (Table 3.6) and analysed by gel electrophoresis (Figure 4.2a). The transcripts degraded extensively despite the addition of aminoguanidine and RNaseoutTM. Both of these compounds have been shown to suppress degradation when the ESEs were reacted with the **44-mer** (Section 3.2.14). Despite the observed degradation a band corresponding to the putative click product was observed. Two bands were visible for the reaction between **ESE1** and **G4-A4**, a major band which was the unreacted **G4-A4** and a faint minor band located above the unreacted **G4-A4** which is the clicked product

(Figure 4.2a, lane 1). The reaction between **ESE3** and **G4-A4** also presented 2 bands with the major band being the unreacted **G4-A4** and the minor band located above was the clicked product (Figure 4.2a, lane 2). The bands were quantified (Figure 4.1b). **ESE3** had phosphorothioate backbone reacted marginally better than **ESE1** which has the natural phosphate backbone with yields of 1.6% and 0.6% respectively, these findings were consistent with the findings in Section 3.2.14.

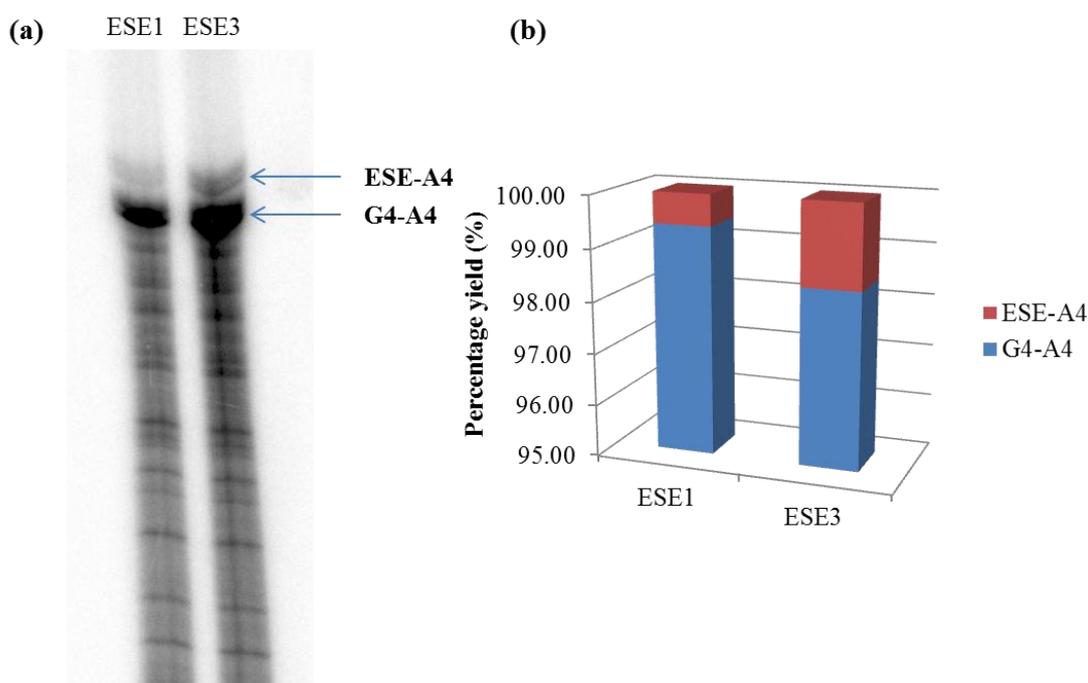


Figure 4.2 A Click reaction between transcript **G4-A4** and **ESE1/ESE3** with the addition of aminoguanidine and RNaseOUT™. (a) 6 % polyacrylamide gel for the reactions, bands were present for the **ESE-A4** product and for the unreacted **G4-A4** labelled. (b) A Histogram of the quantitated gel, demonstrating percentage of **ESE-A4** to unreacted **G4-A4**.

Click chemistry was chosen as a method for ligation due to its bio-orthogonal nature, the reagents are highly selective in highly complex biological system and side reactions do not occur.¹²¹ Additionally the reactions are generally high yielding and readily react efficiently at room temperature. Moreover they are water compatible which is a requirement when ligating two RNA strands together.^{114,119,121} However, the click reaction between **G4-A4** and **ESE1** and **ESE3** was not high yielding after a 20 minute

incubation at 40°C as less than 2% product was produced (Figure 4.2b), which could be due to the following:

1. The length of the RNA may have an overall impact on the click reaction, Humenik *et al* in 2007 reported on the click reaction between alkyne modified oligodinucleotides of varying lengths and a proteins of various sizes modified with an azide, and found the conversion rate decreased as the size of the macromolecules increased.^{162,163} A similar pattern was seen in the click reactions undertaken in this study, where G-initiator (**G4**) reacted with the **ESE1** with a quantitative conversion (Figure 3.13). The **44-mer** incorporated G-initiator (**G4**) when reacted with the same ESE (**ESE1**) generated yields of up 50% (Figure 3.17) and finally **G4-A4** reacted with **ESE1** generated a yield of less than 2% (Figure 4.2). This could inherently be due to the bulkiness of the starting RNA. The amount of G-initiator (**G4**) incorporated into the transcript **A4** could not be determined. However, from previous results (Section 3.2.12) only 72% of G-initiator (**G4**) was incorporated into the 44-mer using the same conditions (Figure 3.14 a & b), and in all probability the amount of incorporation will not exceed this amount, in fact it is possible that a decrease in incorporation could have occurred, but due to the increased length of **A4** (315-mer), separation by gel electrophoresis was not possible and as a result the actual conversion of **G4-A4** to the clicked product **ESE1-A4** and **ESE3-A4** could not be calculated.
2. The amount of RNA degradation was so significant it impacted on the amount of product observed as an increase in degradation reduced the amount of intact RNA produced. Furthermore, the concentration of transcripts (0.3-0.5 nM) was extremely low with reaction volumes typically around 5 µl meaning that the probability of the azide and the alkyne coming into contact with each other was

greatly reduced compared to the similar reaction between **ESE1** and G-initiator (**G4**) (Section 3.2.11, Figure 3.13) in which the concentration of (**G4**) was increased by 1×10^4 to $50 \mu\text{M}$.

3. Single stranded RNA molecules have the ability to fold into highly elaborate structures such as hairpin loops and pseudo knots and this secondary structure may have an impact on the overall reaction. The **G4-A4** transcript may fold such that the alkyne is not readily available to react with the azide.¹⁶⁴

A combination of all these factors discussed could potentially have a bearing on the overall reaction between the ESEs (**ESE1** & **ESE3**) and the alkyne incorporated transcript (**G4-A4**) and could lead to the decrease in the overall yield produced. Despite this a sufficient amount of the click product was obtained by gel electrophoresis and isolation by excision from the gel.

4.2.2 Splicing Reactions of the Tripartite Constructs with ESE1-4

4.2.2.1 Preparation of the tripartite constructs with ESE1-4

The slower electrophoretic mobility of tripartite structure enabled separation of the products ESE-A4 from the starting alkyne transcript **G4-A4** by gel electrophoresis. However, the amount of transcript per click reaction was doubled so that more [α -³²P] GTP was present in order to detect the products when exposed to an x-ray film, enabling the product to be excised and then extracted from the gel using RNA elution buffer. The extracted RNA was ethanol precipitated and resuspended in a minimal amount of water (2-10 μl) depending on the overall radioactivity when analysed by the Geiger counter ready for splicing.

ESE1-4 were bioconjugated to **G4-A4** to produce four different tripartite constructs (Figure 4.3). The ESEs were all based on 2'-OMe RNA, with **ESE3** & **ESE4** containing a phosphorothioate backbone, **ESE1** & **ESE3** were modified at the 3' end with a HEG linker and the azide, whereas **ESE2** & **ESE4** were modified with only the azide (Table 3.9). When the triazole group was formed the linker is approximately equivalent to 5-6 nucleotides for **ESE1** & **ESE3**, and 3 nucleotides for **ESE2** & **ESE4** (Figure 4.3).

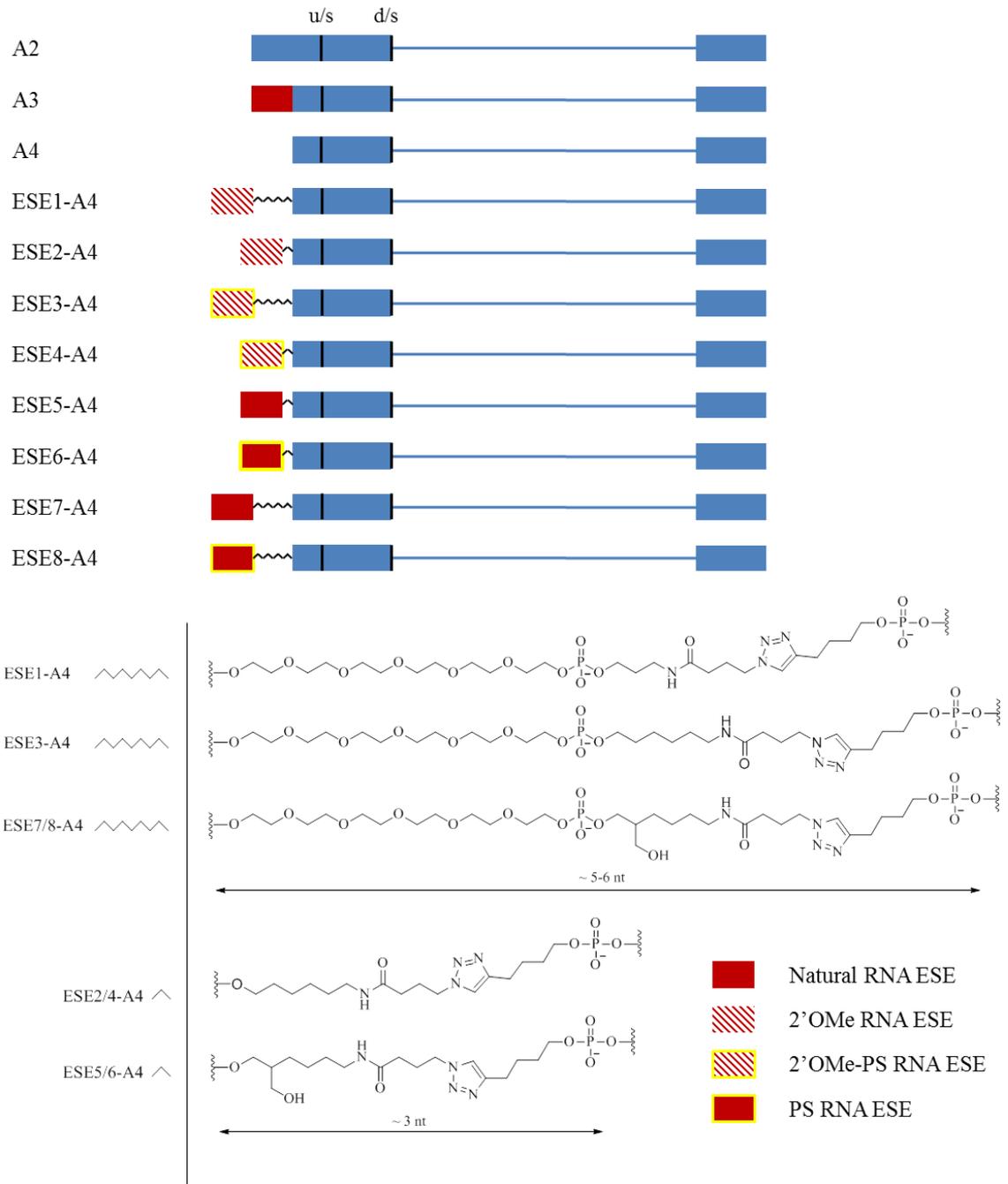


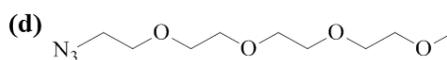
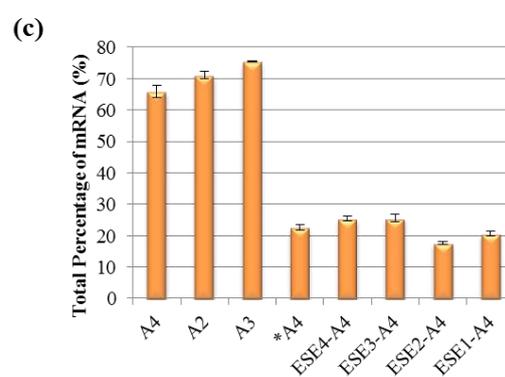
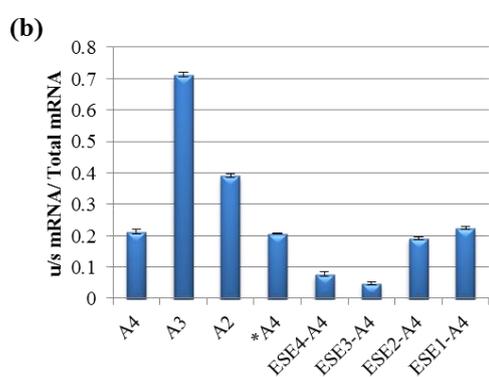
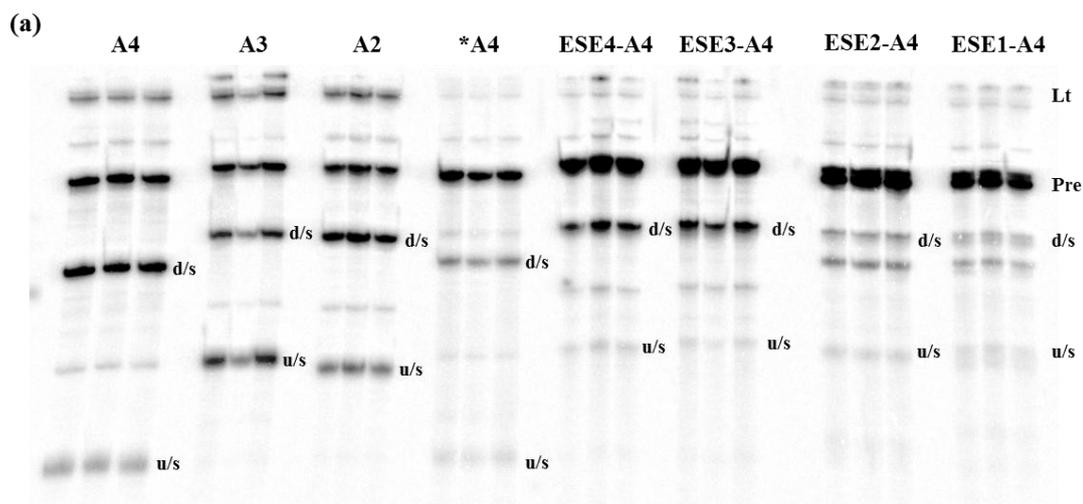
Figure 4.3 Schematic representation of the varying adeno virus constructs used in the splicing assays. The various linkers are approximately equivalent to 5-6 nt (ESE1, ESE3, ESE7 & ESE8) or to 3 nt (ESE2, ESE4, ESE5 & ESE6).

4.2.2.2 Splicing of tripartite constructs using ESE1-4

RNA splicing patterns of the tripartite pre-mRNAs (Figure 4.3) were tested by incubating these constructs in HeLa cell nuclear extract in the presence of 3.2 mM MgCl₂, 50 mM K-glutamate, 1.5 mM rATP and 20 mM creatine phosphate for 90 minutes. For comparative analysis of splicing **A2**, **A3** and **A4** pre-mRNA (Figure 4.3) were spliced at the same time. An additional control where **G4-A4** was bioconjugated to O-(2-azidoethyl)-O'-methyl-triethylene glycol (**51**) forming ***A4** was also spliced to determine if the triazole group had an overall effect on splicing. The splicing reactions were all performed in triplicate and the products were separated by gel electrophoresis and exposed to a phosphor-screen and analysed.

After splicing for 90 minutes the majority of spliced mRNA from transcripts **A2** and **A4** was derived from the downstream splice site producing downstream mRNA (Figure 4.4a). The ratio of upstream mRNA to total mRNA was 0.39 and 0.21 respectively (Figure 4.4b), whereas the presence of the ESE (Figure 4.3) in transcript **A3** shifts splicing substantially to the upstream splice site (Figure 4.4a); with a ratio of upstream mRNA to total mRNA of 0.71 (Figure 4.4b). These results are comparable to the initial splicing assays discussed in Section 2.2.3, in which the ESE located in **A3** shifts splicing from the downstream to the upstream splice site although splicing occurs predominantly at the upstream splice site (Figure 2.9). The difference in overall splicing preferences could be due to the change in batches of the HeLa cell nuclear extract: variations in splicing factors such as SR proteins and hnRNPs amongst others can occur and will impact splicing. Strikingly, none of the ESE tripartite structures enhanced the use of the upstream splice site. The upstream mRNA produced was barely visible on the gel for all 4 tripartite constructs (Figure 4.4a, **ESE1-A4**, **ESE2-A4**, **ESE3-A4** & **ESE4-**

A4), the ratio of upstream mRNA to total mRNA for **ESE1-A4**, **ESE2-A4**, **ESE3-A4** and **ESE4-A4** was 0.22, 0.19, 0.04 and 0.08 respectively (Figure 4.4b). The ratio of upstream mRNA to total mRNA for both **ESE1-A4** and **ESE2-A4** is comparable to **A4** (Figure 4.4b). The amount of upstream mRNA produced for the ESEs containing the phosphorothioate backbone (**ESE3-A4** & **ESE4-A4**) was considerably less (4% and 8% respectively) than the amount produced for **A4** (21%) and the ESEs with the phosphate backbone (**ESE1-A4** and **ESE2-A4**, 22% and 19% respectively) (Figure 4.4c). **ESE1-A4** and **ESE2-A4** were susceptible to degradation, giving rise to the double bands of product mRNA (Figure 4.4a), **ESE3-A4** and **ESE4-A4** on the other hand did not degrade as only one band of product mRNA was visible (Figure 4.a), this is most likely caused by the extra stability of the PS bond to hydrolysis by nucleases.^{36,165} Interestingly, the overall splicing efficiency for the ESE tripartite constructs decreased considerably. The mRNA produced after 90 minutes for **A2**, **A3** and **A4** constructs ranged from 65-75% compared to only 18-25% of mRNA for **ESE1-**, **ESE2-**, **ESE3-** & **ESE4-A4** (Figure 4.4c). The triazole group is the most likely cause of the deficiency. This was confirmed by ***A4** which was bioconjugated to the triethylene glycol linker (**51**), with only 23% of the pre-mRNA spliced producing both upstream and downstream mRNA products (Figure 4.4c). However, the triazole group did not have any effect on splice site selection as the ratio of upstream mRNA to the total amount of RNA is 0.21 which is exactly the same as the precursor **A4** (Figure 4.4b).



51

Figure 4.4 a) 10 % polyacrylamide gel electrophoresis of splicing reaction products. Reactions with the radiolabelled RNA sequences were done in triplicate for 90 mins. A2, A3 & A4 are unconjugated transcripts; ESE1-ESE4 are tripartite molecules with RNA A conjugated to the 2'OMe ESE sequences in Table 3.9. *A4 is A4 conjugated to O-(2-Azidoethyl)-O'-methyl-triethylene glycol but not an ESE. u/s and d/s show product mRNAs formed by splicing to u/s ss and d/s ss; Pre is unspliced pre-mRNA; Lt shows lariat by-products of splicing. (b) Means and standard deviations for the proportion of mRNA spliced to u/s splice site. (c) Means and standard deviations of the percentage of the mRNA produced at both the u/s and d/s ss. (d) O-(2-azidoethyl)-O'-methyl-triethylene glycol (51) used to form *A4.

Table 4.1 Data obtained for the percentage of mRNA produced and the ratio of upstream mRNA to total mRNA from splicing assay. (Figure 4.4).

Construct	upstream mRNA : Total mRNA Ratio	Percentage of mRNA produced (%)
A2	0.39	65
A3	0.71	71
A4	0.21	76
*A4	0.21	23
ESE1-A4	0.22	21
ESE2-A4	0.19	18
ESE3-A4	0.04	26
ESE4-A4	0.08	25

It was clear from these findings that the artificially tethered ESE sequences (Figure 4.4, **ESE1-A4** to **ESE4-A4**) were unable to reproduce the action of a naturally occurring ESE (Figure 4.4, **A3**) which enhances the utilization of the upstream splice site. Interestingly the phenomenon was seen irrespective of the linker whether it contained a short alkyl chain (**ESE2** and **ESE4**) which is equivalent to approximately 3 nucleotides (Figure 4.3) or a single HEG linker plus the short alkyl chain (**ESE1** and **ESE3**), approximately equivalent to 5-6 nucleotides (Figure 4.3). Moreover, the composition of the ESE backbone did not affect splice site selection: both the ESEs containing phosphates (**ESE1** and **ESE2**, Figure 4.3) and the ESEs containing phosphorothioates

(**ESE3** and **ESE4**, Figure 4.3) spliced primarily to the downstream splice site following the same pattern as the natural constructs **A2** and **A4** (Figure 4.4).

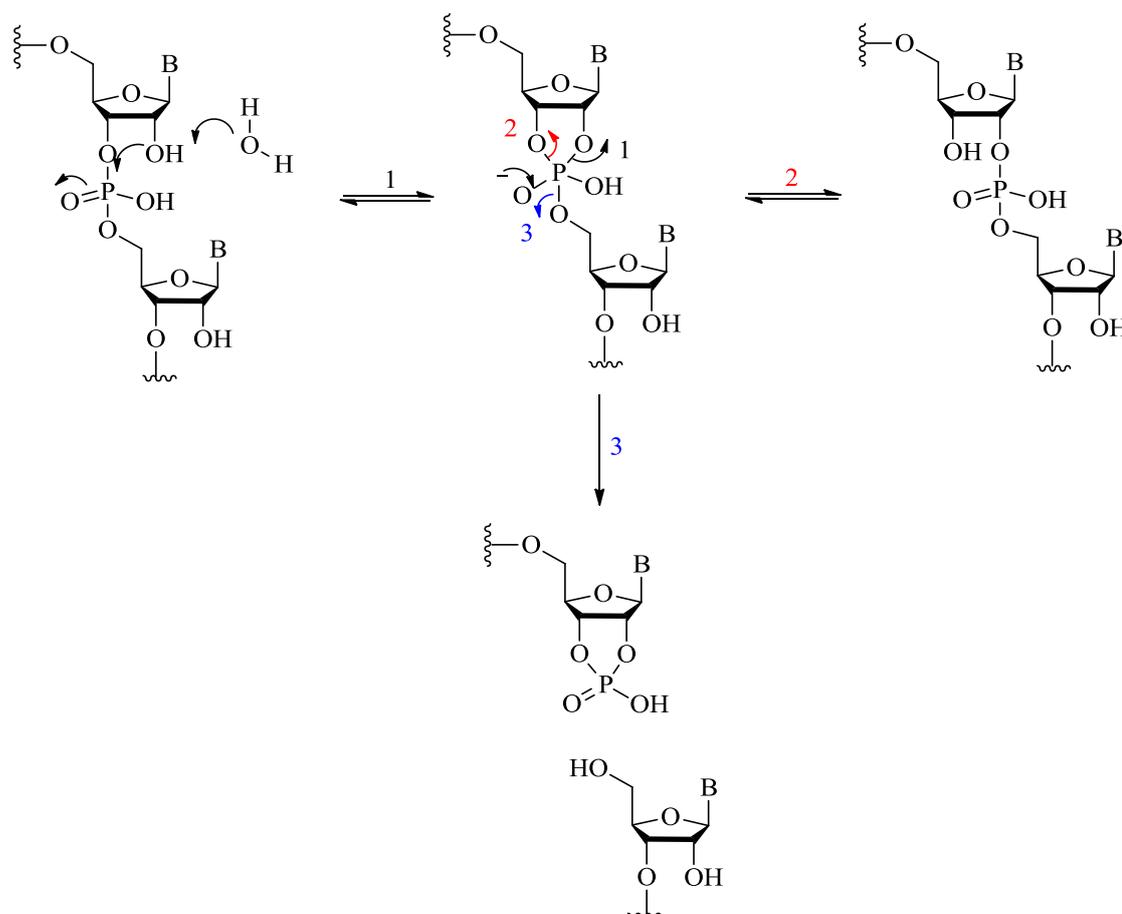
Since the splicing of the RNA tripartite constructs (**ESE1-A4** to **ESE4-A4**, Figure 4.4) did not splice in the same manner as the construct containing the artificial ESE (**A3**, Figure 4.3), it infers that the ESE does not promote splicing by the looping method discussed. If the ESE had exerted its effects by the “looping” model then the RNA would have still been able to loop around bypassing the PEG linker interacting with the U1snRNP. However, the findings were not conclusive due to the presence of the 2'OMe modification on every base throughout the synthetic ESE. This modification itself could be the route cause for the ESE not exerting its effect as they are not naturally found in genes *in vivo*, therefore further experiments were required using ESE based on the naturally occurring 2'OH RNA.

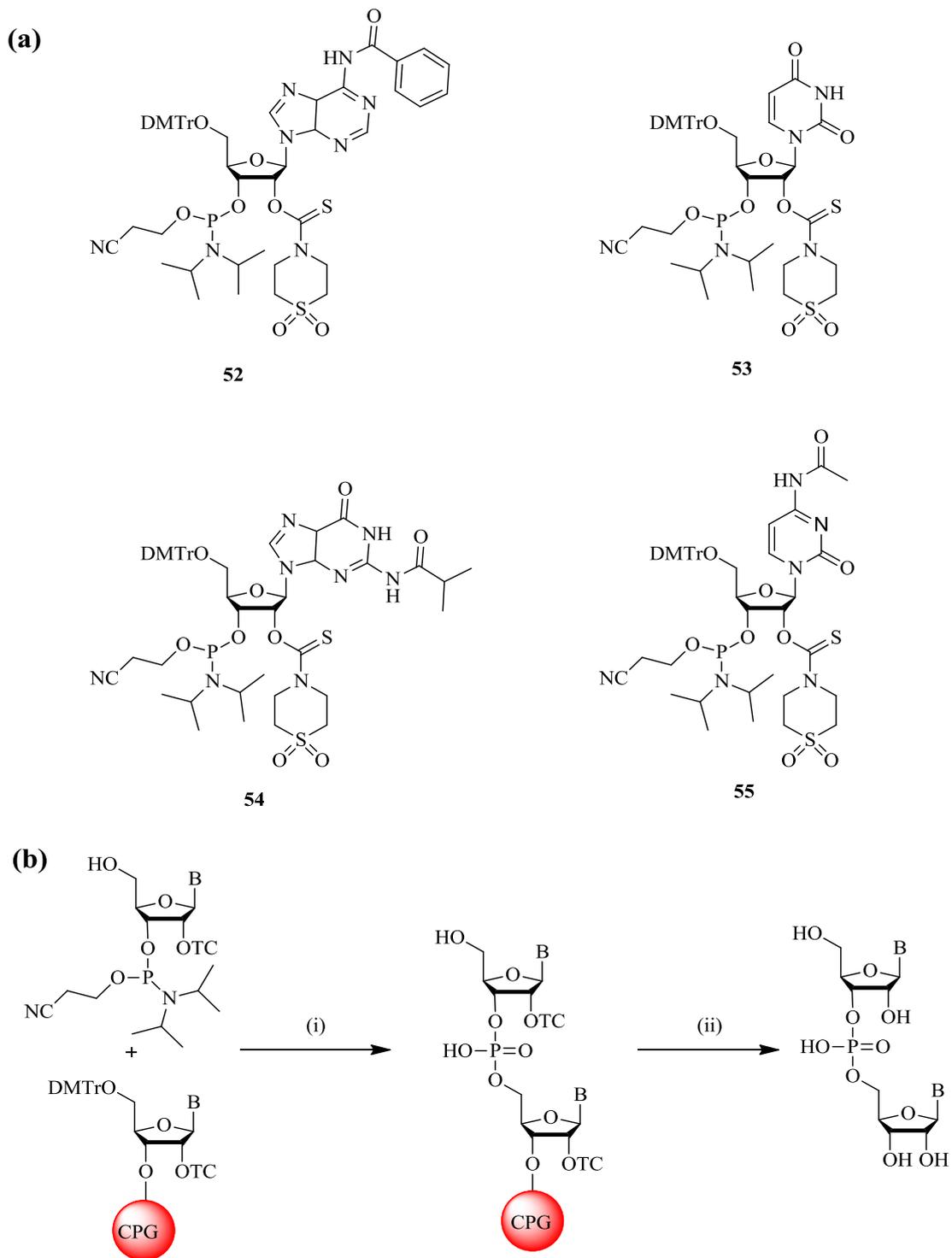
4.2.3 Synthesis of Naturally Occurring 2'Hydroxy ESEs

RNA synthesis can be extremely challenging as the free 2'OH group can react with the 3'P resulting in either cleavage of the oligonucleotide or rearrangement (Scheme 4.1). The most widely accepted method for synthesising RNA is the use of TBDMS protecting group on the 2'O site. However, it is not an efficient synthesis with coupling times between 12-15 minutes for each base using BTT activator. Cleavage and deprotection of the oligonucleotide is time consuming 2 step process requiring relatively harsh conditions; the initial cleavage and deprotection of the nucleobases occurs using concentrated aqueous ammonia at 65°C 10 minutes, which can partially deprotect the TBDMS group resulting in the cleavage of the oligo or a rearrangement (Scheme 4.1). The second step is the removal of the TBDMS group which is most

commonly achieved by using N-methylpyrrolidone/triethylamine/triethylamine trihydrofluoride for 3 hours at 65°C. TC-RNA chemistry (Figure 4.5a) provides an alternative route to RNA synthesis, the coupling times are greatly reduced to 3 minutes using ETT activator, and only requires a mild two step deprotection and cleavage process which is complete after 2 hours at room temperature (Figure 4.5b).¹⁵⁷ The thiomorpholine-carbothioate (TC) protecting group was designed to deprotect slower than the protecting groups on the nucleobases reducing the risk of rearrangement and cleavage.¹⁵⁷ For these reasons it was decided to synthesise the ESEs using the TC-RNA phosphoramidites.

Scheme 4.1 A schematic representation of the internal base hydrolysis resulting in either (1 & 2) a 2'/3' phosphate rearrangement or (3) cleavage of the backbone.



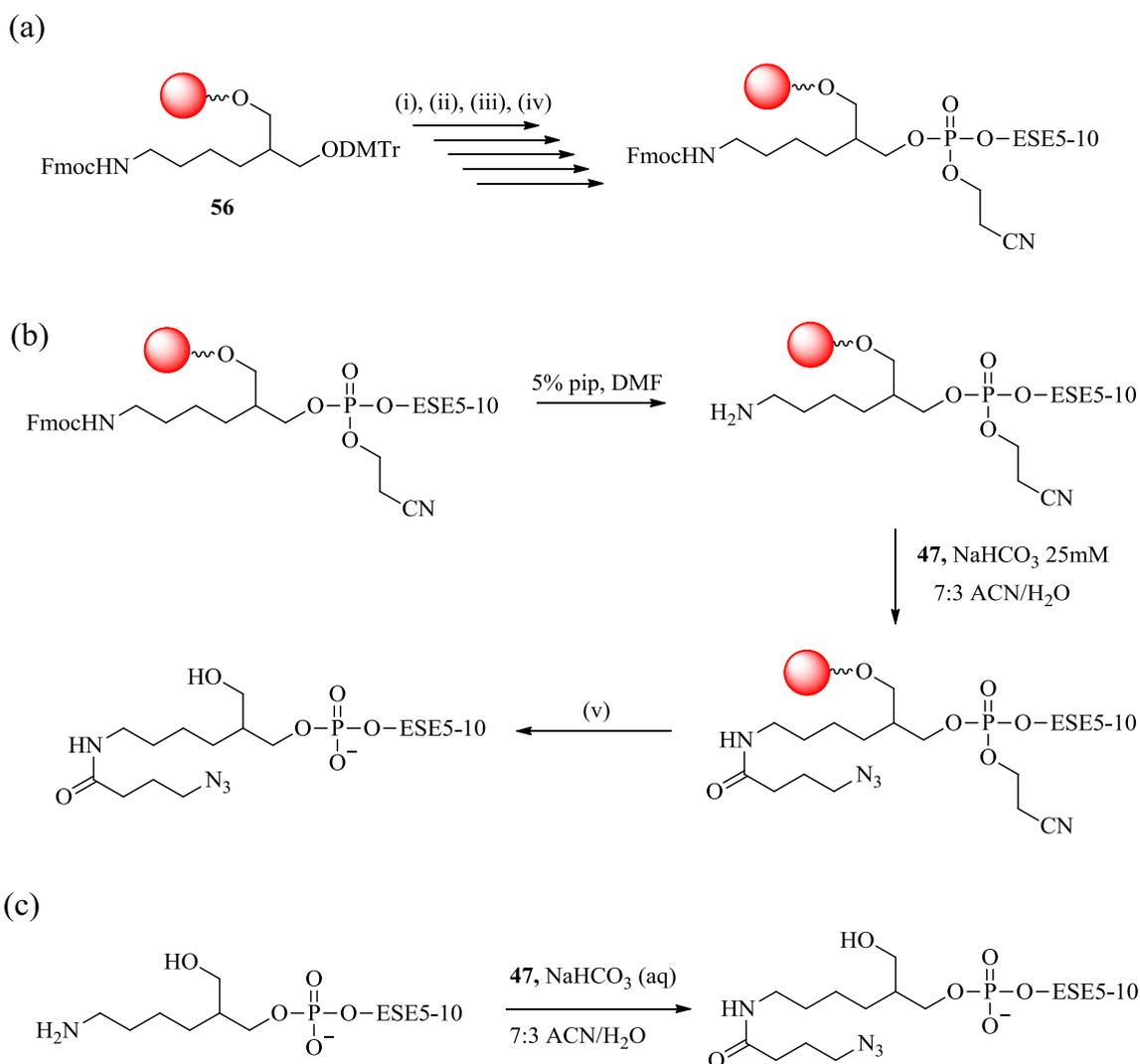


4.2.3.1 Design and synthesis of 2'-hydroxy ESE-azides by solid phase

Six ESEs (Table 4.4) were designed and synthesised to resemble **ESE1-4** (Table 3.9), to limit degradation of the ESEs, **ESE5** and **ESE7** were designed with phosphorothioates replacing the phosphates for the first five bases at the 5' end as phosphorothioates are more stable to hydrolysis by nucleases,¹⁶⁵ the addition of the 5 phosphorothioates to this ESE at the 5' end has been used by Eperon *et al.* (2003) and has not shown inhibitory effects on the enhancer sequence.^{36,142} The splicing of the **ESE1-A4** and **ESE2-A4** showed 2 bands for all RNA products implying the lower bands were of the same length as **A4** (Figure 4.4a) which could be due to the linkers acting as a cap. It was decided to attach a HEG linker to the 5' end of **ESE9** and **ESE10**. **ESE5** and **ESE9** were designed to replicate **ESE2** whereas **ESE7** and **ESE10** were designed to replicate **ESE1**. **ESE6** and **ESE8** were designed with phosphorothioates throughout the sequence replicating **ESE4** and **ESE3** respectively.

ESE5-10 (Table 4.4) were synthesised by automated solid phase synthesis. Due to the mild deprotection conditions required for TC-RNA a 3'-amino-modifier C7 CPG (**56**) was used as the solid support (Scheme 4.2a). The amino group on the modifier was protected by an Fmoc group which enabled the coupling of the NHS-azide (**47**) whilst the RNA was still attached to the CPG. Fmoc was removed by 20% piperidine in DMF, leaving a free amine ready for coupling. (**47**) in a 0.025M solution of NaHCO₃ was added to the resin and left tumbling overnight, the ESE was then deprotected with ethylenediamine/toluene, and cleaved with 0.1M triethylammonium acetate buffer (Scheme 4.2b).

Scheme 4.2 Schematic representation of (a) solid phase synthesis of ESE5-10; Reagents and conditions. (i) detritylation (3% trichloroacetic acid in DCM), (ii) coupling of phosphoramidite (Spacer-CE Phosphoramidite 18 0.12M, Ac-C TC-RNA CE Phosphoramidite 0.1M, iBu-G TC-RNA CE Phosphoramidite 0.1M & Bz-A TC-RNA CE Phosphoramidite 0.1M), (iii) capping of unincorporated base (Cap A THF/Pyridine/acetic acid (8:1:1) & Cap B 10% Methylimidazole in THF), (iv) oxidation of phosphoramidite (iodine in THF/pyridine/water (7:2:1) 0.02M or phenylacetyl disulfide in acetonitrile/pyridine (1:1) 0.02M), repeated for addition of each base. (b) Solid phase coupling of NHS-N₃ (47), Fmoc group removed with 5% piperidine in DMF followed by coupling of 47 in 25 mM NaHCO₃, (v) Deprotection and cleavage with 1:1 ethylenediamine/toluene, and 0.1M triethylammonium acetate buffer.¹⁵⁷ (c) Solution phase NHS-N₃ (47) coupling.



The six ESEs were examined by gel electrophoresis (Figure 4.7a). Major bands were seen for all the ESEs (**ESE5-9**) (Figure 4.7a, band highlighted by red box) although the bands of RNA were extremely streaky above and below the band especially for **ESE8**.

The streakiness below the band may be due to degradation of the RNA from nucleases, but this does not explain why there was streaking above the band. This was possibly due to the secondary structure of the RNA. The ESE is rich in guanosines which can result in the formation of G-quadruplexes. G-quadruplexes are non-canonical interaction by Hoogsteen base pairing of a quartet of guanosines (Figure 4.6a) which can self-associate intra- or inter-molecularly (Figure 4.6b).^{166,167} The slower electrophoretic mobility of the G-quadruplex structure than the linear RNA may contribute to the streaking visible above the band.¹⁶⁸ However, another factor to the streaking may have been due to overloading of the RNA.

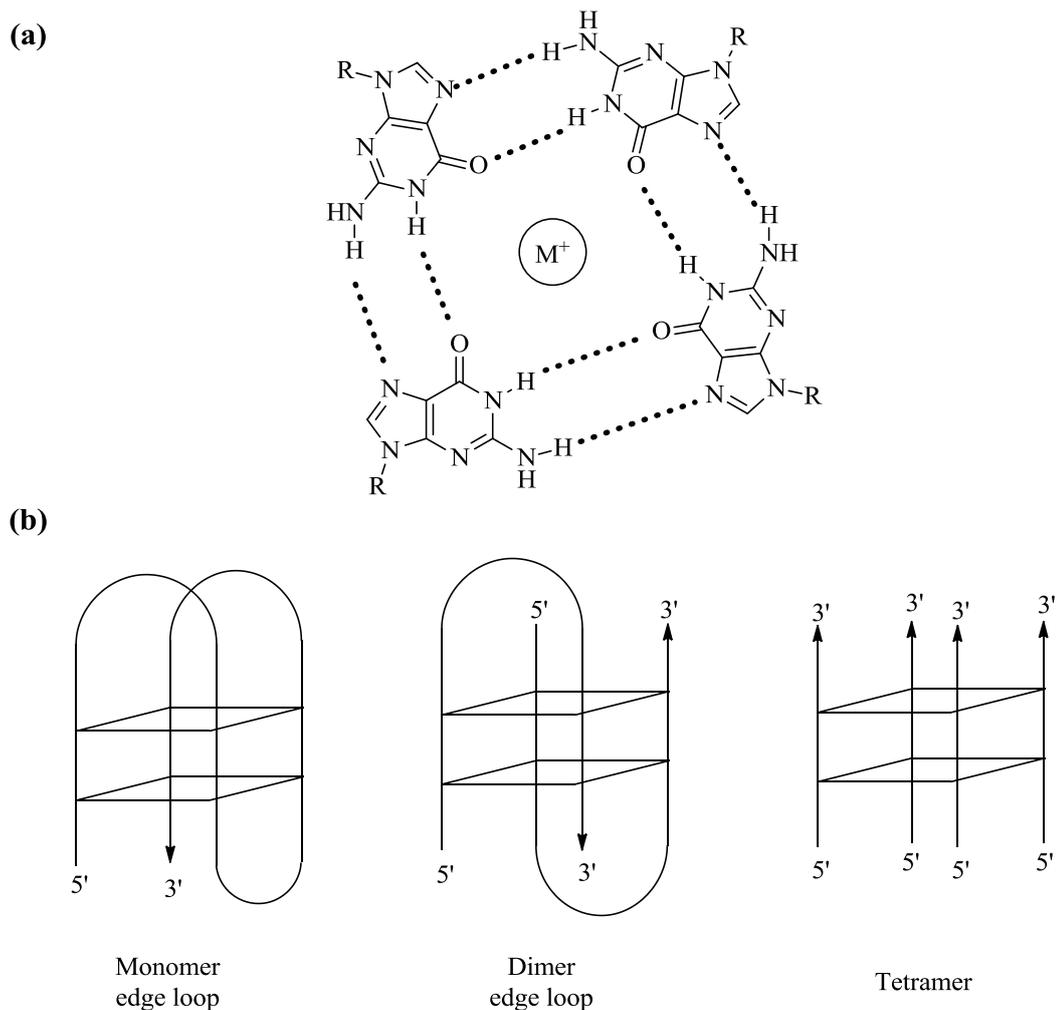


Figure 4.6 (a) Structure of the G-quartet base pairing via Hoogsteen hydrogen bonds forming the cyclic structure, metal ion stabilises the structure. R corresponds to the RNA bases. (b) Schematic representation of G-quadruplex structures which assemble using one (monomer), two (dimer) or 4 (tetramer) strands.¹⁶⁶

4.2.3.2 Confirming the presence of an azide group on the ESEs by click chemistry

The presence of the azide on the ESEs (**ESE5-10**) were determined by reacting the ESEs with the **44-mer** incorporating the G-Initiator (**G4**) (**G4-44-mer**). The reactions were undertaken in triplicate, concurrently to limit errors, in the presence of Cu(I) and THPTA at 40°C for 20 minutes and analysed by gel electrophoresis which was imaged by a phosphor-imager screen. The analysis of the gels appeared to show only starting **G4-44-mer** transcript was present in the reactions for **ESE5**, **ESE6**, **ESE9** and **ESE10** (Figure 4.7b, block 1, 2, 4 & 5) as there was only 1 band observed for these reactions

which had the same electrophoresis mobility as **G4-44-mer** transcript (Figure 4.7b, lane 7). Streaks of RNA were observed above the bands for the reaction between **ESE6** and **ESE8** and **G4-44-mer** transcript (Figure 4.7b, block 3 & 6), containing phosphorothioate linkages throughout the sequence, however it was not evident whether the reaction had actually worked. The results did suggest however, that (**47**) had not coupled to the ESE whilst on resin, and recoupling of (**47**) was required. The failure in the coupling of (**47**) to the RNA whilst on resin might be due to steric sequestration of the 3' amino group.

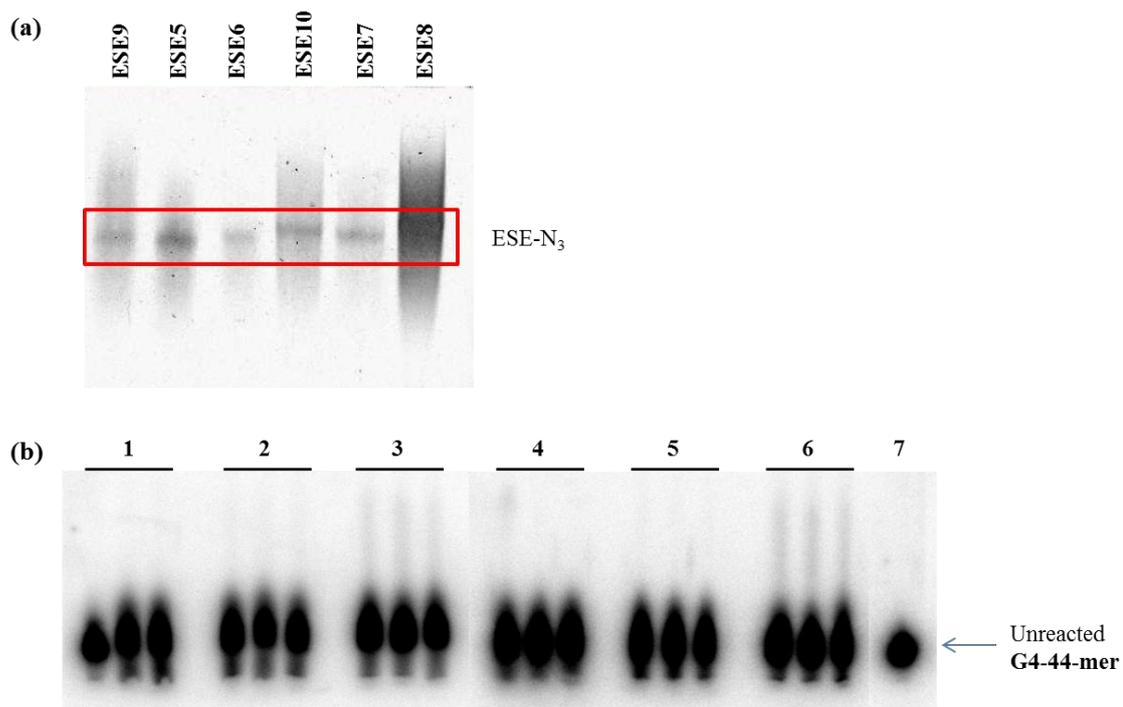


Figure 4.7 (a) ESE5-10 after NHS-N₃ was coupled to the RNA on solid phase,, ESE-N₃ highlighted in the red box. (b) Triplicate click reactions between ESEs and 4-44mer with Cu(I) THPTA; block 1: ESE9 and 4-44, block 2: ESE5 and 4-44, block 3: ESE6 and 4-44, block 4: ESE10 and 4-44, block 5: ESE7 and 4-44 block 6: ESE8 and 4-44, lane 7: 4-44.

4.2.3.3 Solution phase coupling of NHS-azide (**47**)

The ESEs (**ESE5-10**) were coupled with 100 equivalents of (**47**) with 0.025M NaHCO₃ in a 7:3 mix of acetonitrile/water. The reactions were left overnight and the salts and

unreacted (**47**) was removed by a GE healthcare Nap-25 column. The ESEs were re-suspended in 500 µl arising in overall concentrations ranging from 0.6-1.7 mM (Table 4.6).

The resulting ESEs were then reacted by the copper click reaction using Cu(I) and THPTA with **G4-44-mer** transcript to determine the existence of the azide. The reaction was then analysed by gel electrophoresis and imaged using the phosphor-imager. The reaction appeared to have worked, albeit not to completion for all six ESEs (**ESE5-10**), but an additional band appeared that migrated between the starting 44-mer and the clicked product that could be due to a drop off during solid phase synthesis creating 2 varying sequences or to RNA degradation which could have occurred during the click reaction itself (Figure 4.8a). The reactions between the **44-mer** and **ESE9** and **ESE10** were inefficient as the clicked product was barely visible, although upon analysis the amount of product were 2% and 1%, respectively (Figure 4.6b). The reactions between the ESEs with the 5 bases at the 5' end containing phosphorothioates (**ESE5** and **ESE7**) afforded higher yields of the click product with 7% and 14% respectively of click product being present (Figure 4.8b). The reaction between **ESE6** and **ESE8** with the **44-mer** afforded the highest yields of clicked product with 72% and 49% respectively of **G4-44-mer** transcript being converted (Figure 4.8b). The results also corroborate with the previous results (Section 3.2.13) inferring the phosphorothioates react more efficiently than the phosphate ESEs.

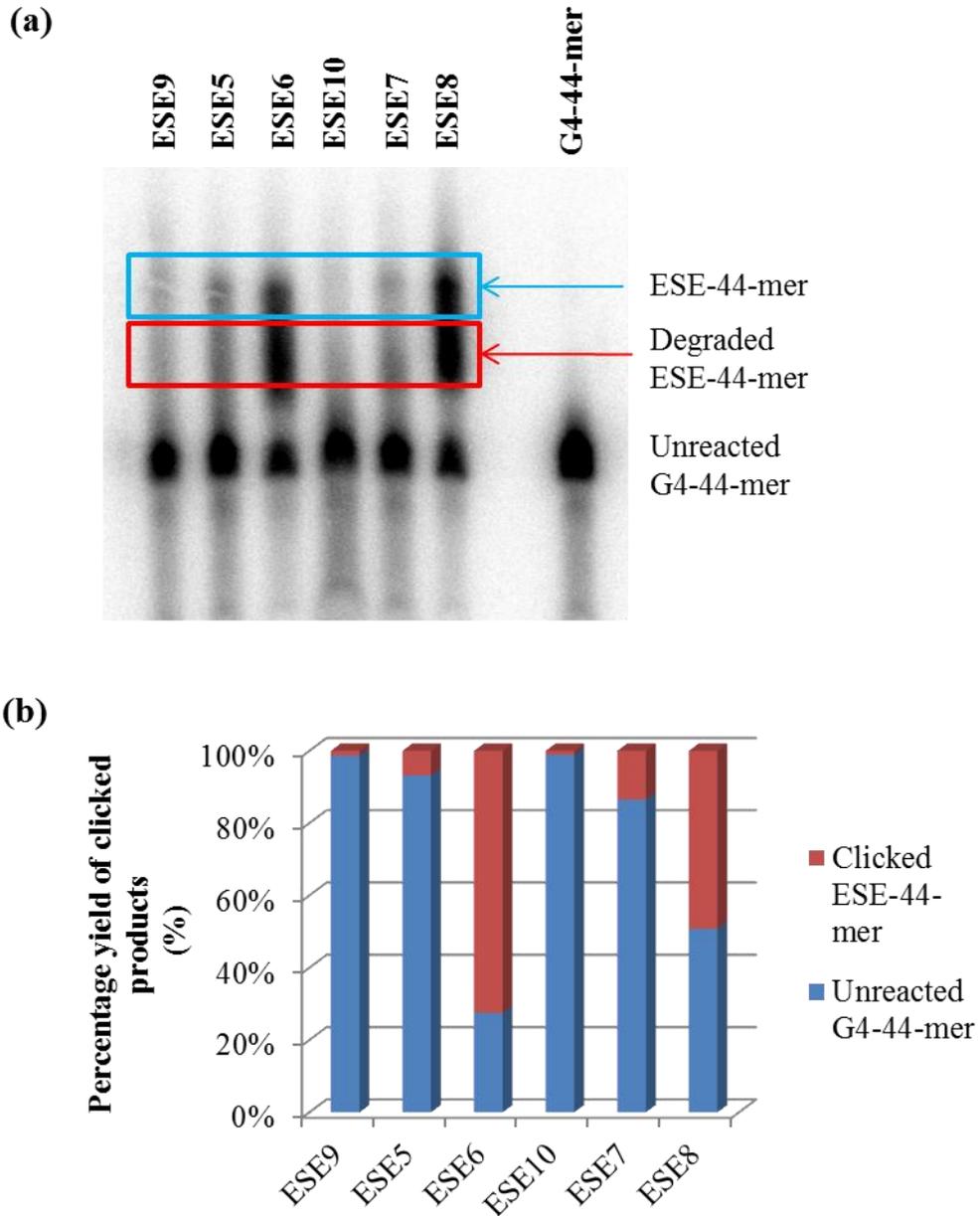


Figure 4.8 Click reaction between ESE5-10 and G4-44-mer after coupling of NHS-N₃ in solution. (a) Polyacrylamide gel of the reactions; ESE-44mer highlighted in blue box, possible degraded ESE clicked to G4-44-mer highlighted in red box, G4-44-mer run in separate lane to determine the unreacted G4-44-mer. (b) Histogram shows the percentage yield of all click products compared (red) to unreacted G4-44-mer (blue).

The results suggest that the azide is present in all six ESEs and therefore could be used to ligate to **A4** ready for splicing. Despite the presence of the additional band it was decided not to purify the ESEs directly, due to time constraints, but to purify the bands after the click reaction with **G4-A4**, which would be required anyway.

4.2.4 Splicing Reactions of the Tripartite Constructs using ESE5-10

4.2.4.1 Preparation of the tripartite constructs using ESE5-10

The ligations of the transcript **G4-A4** to the ESEs were undertaken using the optimized conditions described in Section 3.3 (Table 3.6). The ligated products were run on a 6% polyacrylamide gel for 4 hours at 10W to prevent the gel from overheating and in turn potentially reducing degradation. The gel was imaged using X-ray films to excise the required bands. Definitive bands were not visible for the tripartite constructs **ESE9-A4** and **ESE10-A4**, so purification from the starting **G4-A4** was not possible and these constructs were not used in the subsequent splicing reactions. However the ligation reactions between **G4-A4** and **ESE5-8**, bands were visible, for the tripartite constructs. As with the previous reactions in Section 4.2.3.3 there were 2 bands visible in addition to the starting **G4-A4** transcript; both the bands were cut out from the gel and the RNA was eluted using RNA elution buffer. After ethanol precipitation the resulting RNA was then dissolved in minimum volumes ready for splicing.

4.2.4.2 Splicing of the tripartite constructs ESE5-8

The RNA splicing preferences of the tripartite ESE-conjugated pre-mRNAs were investigated by incubation in a HeLa cell extract for 90 minutes using the same conditions as described in Section 4.2.2.2. The two bands excised from each of the click reactions between **G4-A4** and **ESE5-8** were spliced, although upon analysis it was clear that the lower bands of the two were not the full length ESE as it had a faster electrophoretic mobility than the controls **A2** and **A3**, whereas the upper band had the same mobility through the gel as **A2** and **A3** and these were the bands analysed in the splicing gels.

As expected from the previous results in Section 4.2.2.2 (Figure 4.4) the splicing of the controls showed that the transcript with the ESE incorporated into the 5' end (**A3**) shifted splicing from predominantly the downstream splice site to the upstream splice site producing more upstream mRNA than downstream mRNA (Figure 4.9a), with 83% of the mRNA arising from use of the upstream site compared with 48% and 31% for **A2** and **A4** respectively (Figure 4.9b). The splicing of the tripartite ESE-conjugated constructs on the other hand did not enhance splicing at the upstream splice site and the majority of the mRNA produced arose from the downstream splice site, irrespective of the linker length (Figure 4.9a & b). The upstream splice site accounted for only 16% and 17% of the spliced products in **ESE5** and **ESE7** respectively and 9% and 8% for **ESE6** and **ESE8** respectively (Figure 4.9b). These results are consistent with the results obtained in Section 4.2.2.2 where the ESEs sugars all contained 2'OMe groups and substantiates the conclusion that the ESE is not exerting its effect by 3-dimensional diffusional encounters between bound components.

The overall splicing efficiency of the tripartite structures was reduced significantly from 83-88% of pre-mRNA being spliced to mRNA for **A2**, **A3** and **A4** to a maximum of 52% of pre-mRNA spliced for **ESE6-A4** and only 37% of mRNA was produced for **ESE7-A4** (Figure 4.9c) and were consistent to the findings with the 2'OMe ESE used in Section 4.2.2.2. The decrease in splicing efficiency may have been due to the presence of the triazole group and is probable, as only 23% of mRNA was produced when a triethylene glycol azide (**51**) was conjugated to **G4-A4** (Section 4.2.2.2, Figure 4.4c). Another possible cause for the decrease in splicing efficiency may be due to the lack of m⁷G cap on the 5' end of the ESEs. The m⁷G cap is used to protect the RNA from 5' exonucleases and has been found to promote splicing in HeLa cell extract in adenovirus 2 constructs.¹⁵⁵ To determine the effect the lack of m⁷G cap has on splicing efficiency a

comparison between the same construct with or without m⁷G cap needed to be carried out.

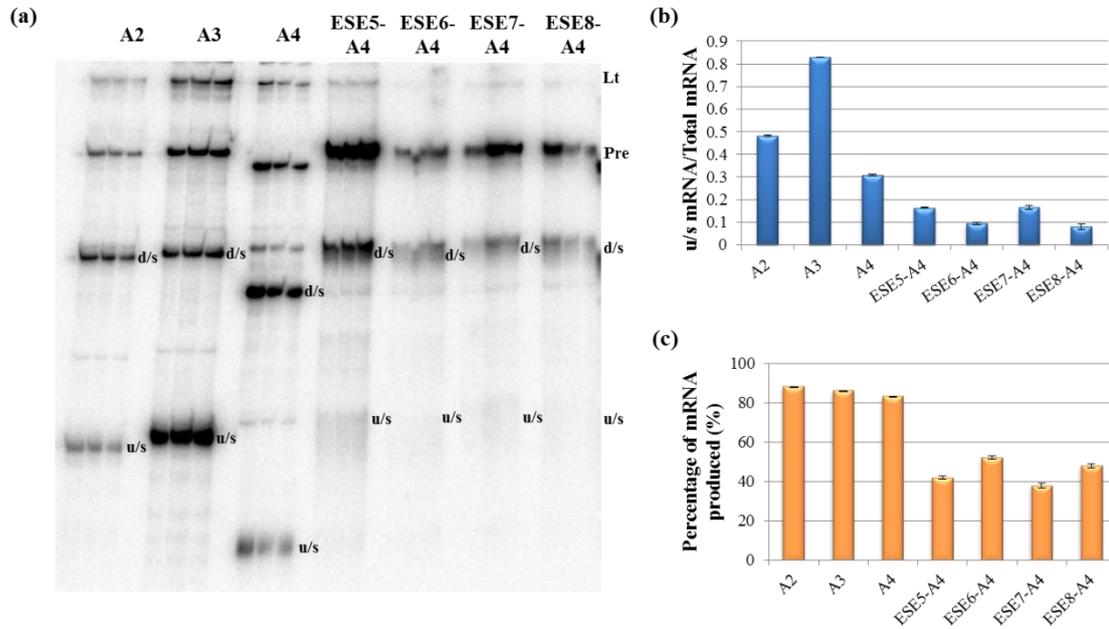


Figure 4.9 a) 10 % polyacrylamide gel electrophoresis of splicing reaction products. Reactions with the radiolabelled RNA sequences were done in triplicate for 90 mins. A2, A3 & A4 are unconjugated transcripts; ESE5-ESE8 are tripartite molecules with RNA A conjugated to the 2'OH ESE sequences in Table 4.4. u/s and d/s show product mRNAs formed by splicing to u/s ss and d/s ss; Pre is unspliced pre-mRNA; Lt shows lariat by-products of splicing. (b) Means and standard deviations for the proportion of mRNA spliced to u/s splice site. (c) Means and standard deviations of the overall yield of mRNA.

Table 4.2 Data obtained for the percentage of mRNA produced and the ratio of upstream mRNA to total mRNA from splicing assay. (Figure 4.9).

Construct	upstream mRNA : Total mRNA Ratio	Percentage of mRNA produced
A2	0.48	88
A3	0.83	86
A4	0.31	83
ESE5-A4	0.16	42
ESE6-A4	0.09	52
ESE7-A4	0.17	37
ESE8-A4	0.08	47

4.2.5 The Influence of the 7-Methylguanosine Cap on the Alternative Splicing of A3.

Splicing efficiency of a construct without the m⁷G cap needed to be determined to ascertain whether the lack of m⁷G cap in the tripartite ESE-conjugated constructs contributed to the decrease in splicing efficiency and to clarify that it has no impact on the way in which the ESE exerts its effects in splicing. An initial triplicate splicing test was carried out whereby **A3** which had been transcribed with and without m⁷G cap and labelled with [α -³²P] GTP was incubated for 90 minutes in HeLa cell extracts. The splicing reaction was then analysed by gel electrophoresis (Figure 4.10a) and the bands were quantified accounting for the number of guanosines in each transcript. Capped **A3** spliced predominantly at the upstream splice site with a ratio of 0.72 for upstream

mRNA to total mRNA (Figure 4.10b). The uncapped **A3** also spliced predominantly to upstream splice site with the upstream mRNA to total mRNA ratio calculated at 0.78 (Figure 4.10b), implying that the lack of cap does not impact on splice site selection.

Further tests were carried out to establish whether the cap has an effect on overall splicing efficiency. **A3** was transcribed with and without m⁷G; however the transcripts were labelled with [α -³²P] UTP instead of the [α -³²P] GTP. The splicing reactions were monitored over a 2 hour time course with time points taken at 0, 30, 60, 90 and 120 minutes. The splicing reactions were then analysed by gel electrophoresis (Figure 4.10c), quantitation of the bands was carried out accounting for the number of uridines within the transcript. Splicing occurred predominantly at the upstream splice site giving rise to more upstream mRNA being produced over downstream mRNA in relation to pre-mRNA (Figure 4.10d). The amount of pre-mRNA present after 30 minutes was equivalent for the two transcripts, but as the time course proceeded, the amount of product mRNA for both the upstream and downstream splice sites was much greater with capped **A3** than for uncapped **A3**, (Figure 4.10d).

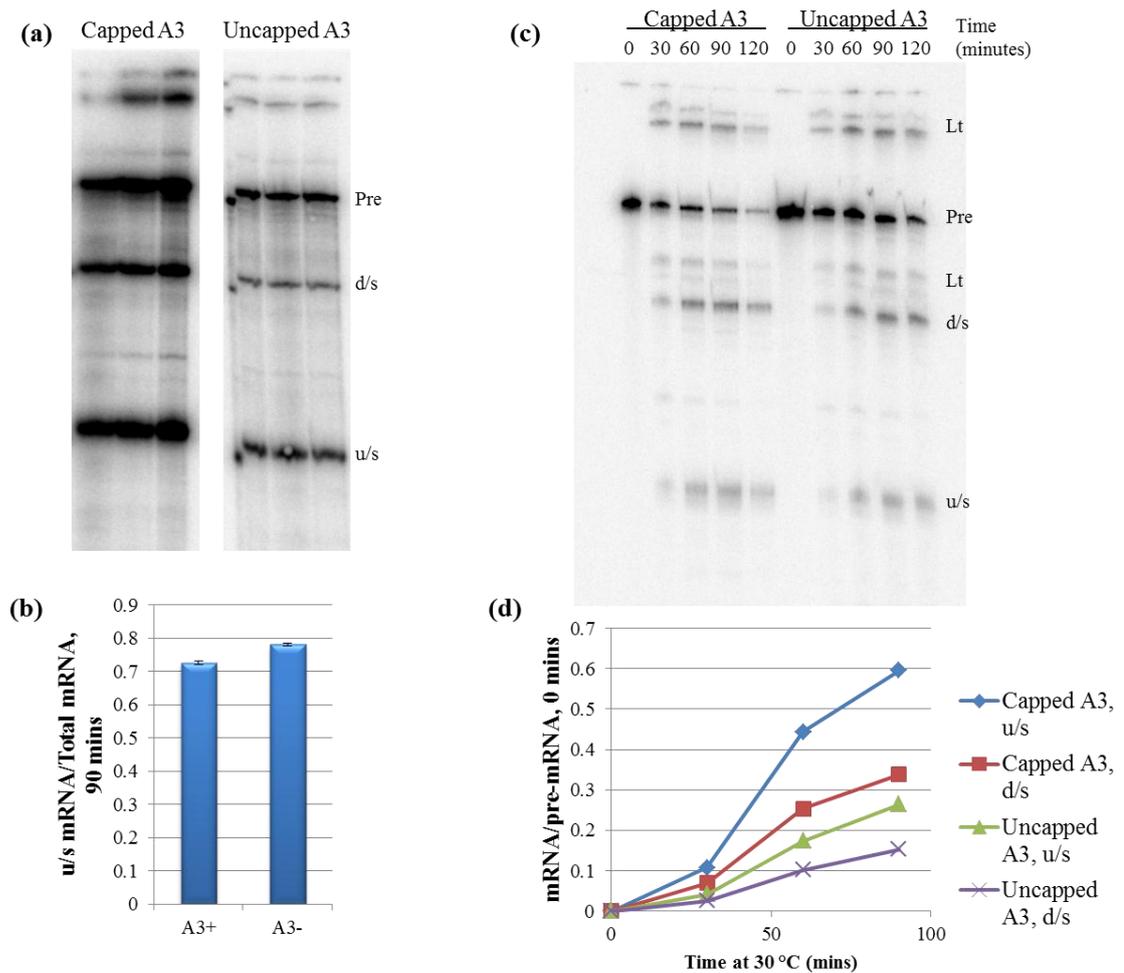


Figure 4.10 (a) 10% polyacrylamide gel for the splicing reaction of capped and uncapped A3. Reactions radiolabelled with [α - 32 P] GTP were done in triplicate, incubated for 90 mins; u/s and d/s show product mRNAs formed by splicing to u/s ss and d/s ss; Pre is unspliced pre-mRNA; Lt shows lariat by-products of splicing. (b) Means and standard deviations for the proportion of mRNA spliced to u/s splice site. (c) As with (a), reactions radiolabelled with [α - 32 P] UTP, incubated over 120 minute time course. (d) Proportion of initial pre-mRNA splicing to u/s and d/s ss for the capped and uncapped A3 after various times of incubation.

These results corroborate with the initial 90 minute splicing assay of capped/uncapped A3; splice site selection was not affected irrespective of the presence of m⁷G cap, but splicing efficiency is; the lack of m⁷G cap has a negative impact on splicing.

4.2.6 Comparative Analysis of Spliceosome Complex Formation for A3, A4 and ESE8-A4

It has been possible to assess the formation of the spliceosomal complex *in vitro* for a number of years.^{27,169} Non-denaturing gel electrophoresis is used to resolve the different complexes and at least four of them are distinct enough to be visible; the composition and order of appearance differ for each complex going from E→A→B→C (Scheme 1.2)^{27,170} In 2009 Kotaljich *et al.* ascertained that splice site selection between two alternative 5' splice sites was determined during the A-complex.²⁷ Spliceosome complex formation was compared between **A4** (the precursor to the tripartite ESE constructs), **A3** (the construct with the ESE incorporated into the 5' end) and **ESE8-A4** (Figure 4.3). As there was no difference in splice site selection for any of the tripartite ESE-conjugated constructs, it was decided to examine the spliceosome complex formation of only one of the ESEs. **ESE8** was chosen because the ligation of phosphorothioate ESEs was significantly more efficient.

To analyse the spliceosomal complexes, splicing reactions were undertaken in HeLa cell extract at 0, 2.5, 5, 7.5 10, 15, 30 and 60 minutes. Instead of treating with PK enzyme which is used to remove the proteins, the reactions were treated with heparin (4 ng/μl) to prevent nonspecific protein/RNA associations before running on an agarose gel.⁷³ The gel was imaged by a phosphor-imager screen. In order to identify the A complex, a separate splicing reaction was undertaken with **A3** in which the HeLa cell nuclear extract was incubated with α -U6-oligonucleotide (Appendix, Table 6.1) for 15 minutes at 30°C prior to the addition of **A3** RNA. The α -U6-oligonucleotide stalled complex formation at the A complex, and only A complex was present throughout the time course after heparin addition (Figure 4.11b).^{46,171} Analysis of the native gels

showed the A complex was present at all the time points, the B complex began to form after 2.5 minutes and the C complex was present after 30 minutes (Figure 4.11a).

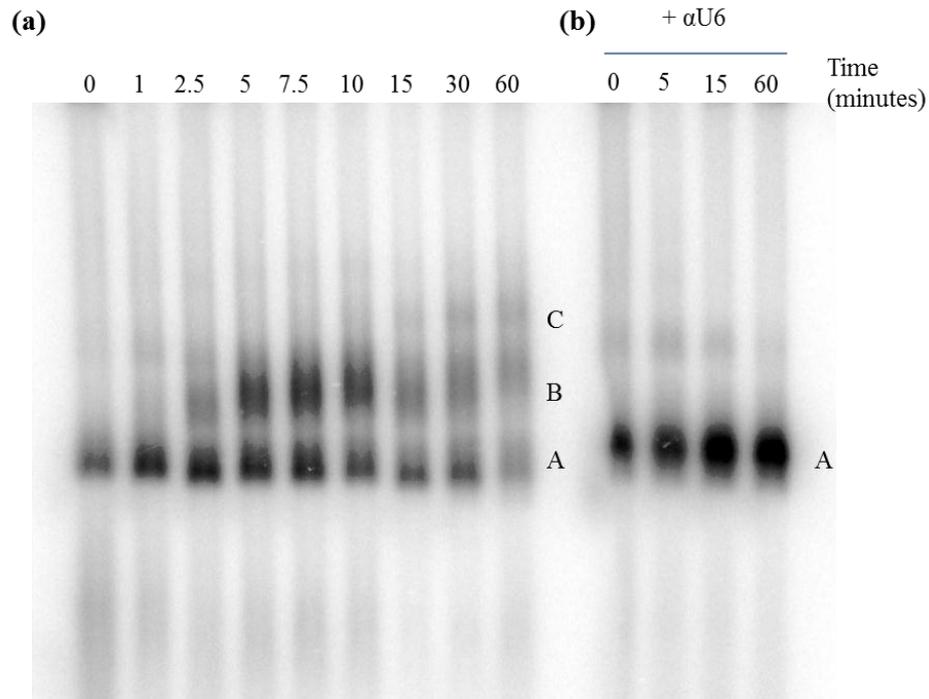


Figure 4.11 Non denaturing gel showing the spliceosome complex formation of the splicing reaction of A3 at 30°C. (a) HeLa nuclear cell extract used as in normal splicing reactions; (b) Nuclear extract treated with α -U6 stalling spliceosome at the A complex

The splicing reactions for the constructs **A3**, **A4** and **ESE8-A4** were analysed. After 2.5 minutes the B complex began to form for **A4** and **A3** (Figure 4.12a), comprising 60% and 23% of the total complexes respectively (Figure 4.12b & 4.12c). The amount of B complex increased over time till after 10 minutes 87% was present for **A4** (Figure 4.12b) and 74% for **A3** (Figure 4.12b). At 30 minutes the C complex was visible for both **A3** and **A4** contributing approximately 20% of all the complexes (Figure 4.12b & c). Interestingly, the formation of B complex stalls for the **ESE8-A4** tripartite constructs; it was only present after 5 minutes attributing to 52% of the complexes present (Figure 4.12d), although the formation of the C-complex from the B-complex

occurred at the same point in time as for **A3** and **A4**, reaching 12% of all the complexes present after 30 minutes (Figure 4.12d).

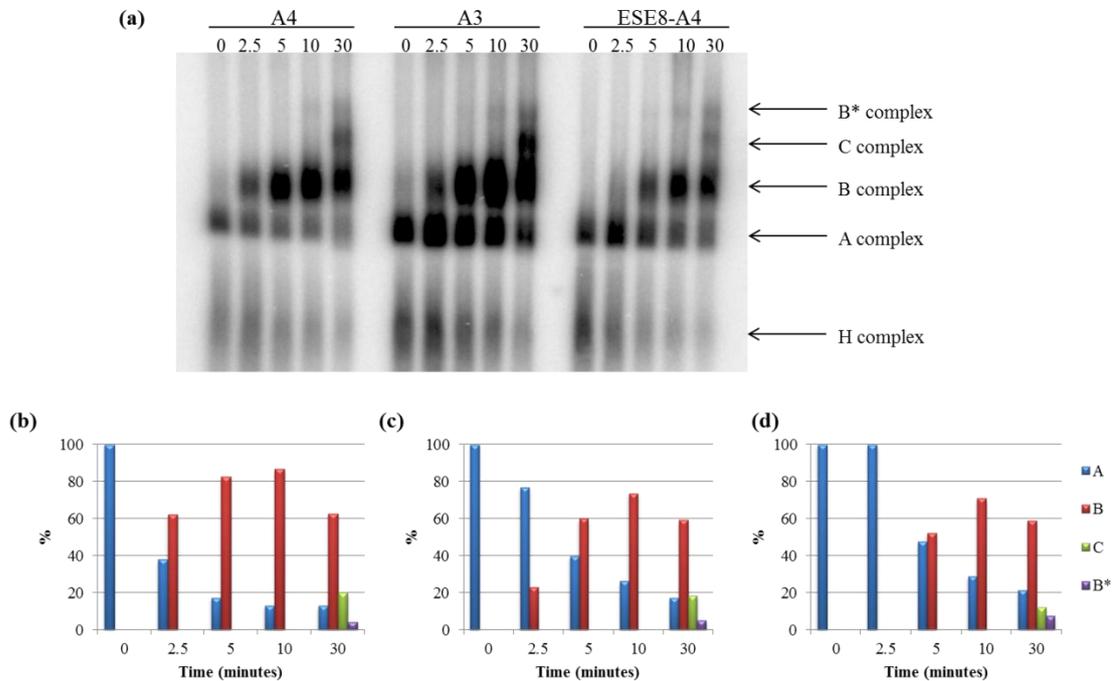


Figure 4.12 (a) Non-denaturing gel showing the spliceosome complex formation during the splicing reaction at 30°C of **A4**, **A3** and **ESE8-A4** over a 30 minute time course, treated with heparin at r.t for 30 minutes. (b) Histogram of the percentage of each complex present at the time points for transcript **A4**, A complex in blue, B-complex in red, C-complex in green and B*-complex in purple; (c) as (b) for transcript **A3**; (d) as (b) for **ESE8-A4**.

Despite there being no difference in when the A complex begins to transfer into the B-complex for **A3** and **A4**, there is overall more B complex present after 2.5 minutes for **A4** than **A3**. Kotaljich *et al.* demonstrated that it was in the A complex that the commitment to alternative 5' splice site pairing took place²⁷ and therefore the rate in which the A complex is converted into B complex may contribute to which splice site is selected predominately over the other. In this case the faster the conversion rate is the downstream splice site is chosen predominately and vice versa. This hypothesis however does not explain why the upstream splice site is not chosen predominately over the downstream splice site for the tripartite structures. The stalling of the formation

of B complex however, may contribute to the poor splicing efficiency for the **ESE-A4** tripartite constructs.

4.3 SUMMARY

The aim of this chapter was to determine whether a non-RNA linker inserted between ESE and the proximal 5' splice site, influences splice site selection. ESEs were designed using different linker lengths, one linker contained a short alkyl chain modified with an azide at the 3' end (**ESE2**, **ESE4**, **ESE5** and **ESE6**) that was equivalent to 3 nucleotides in length once ligated to alkyne on the 5' end of the transcript, the second linker contained a HEG linker which was conjugated to the alkyl-azide via a phosphodiester bond (**ESE1**, **ESE3**, **ESE7** and **ESE8**) and was equivalent to 5-6 nucleotides once the triazole was formed. **ESE1** to **ESE4** were 2'OMe based constructs (Table 3.9) and **ESE5-8** were 2'OH based constructs (Table 4.4).

The splicing assays showed that the majority of the mRNA derived from the downstream splice site for the naturally occurring **A2** and **A4** constructs. The addition of the ESE in **A3** on the other hand reversed preferences to the upstream splice site (Figures 4.4 & 4.9). Remarkably, none of ligated ESE tripartite constructs whether they were based on 2'OMe or 2'OH RNA chemistry were able to enhance the upstream splice site. These findings were also observed regardless of the length of the linker (Figure 4.4 and 4.9). It was clear that artificially tethered sequences were unable to reproduce the action of the integral ESE. These findings infer that the enhancer is not exerting its effects by the widely accepted 3-dimensional diffusion model. If the 3-dimensional diffusion method was in force the SR protein bound to the ESE would still be able to loop around and interact with U1 snRNP at the upstream 5' splice site

stimulating splicing to produce predominantly upstream mRNA. However, this is not the case, the upstream splice site was not enhanced and therefore 3-dimensional diffusion was not taking place or was not sufficient to enhance splice site usage.

The splicing efficiency of the tripartite ESE-conjugated constructs was significantly less than the natural RNA transcripts. The decrease in the overall splicing efficiency of the artificial tripartite constructs may be due to two factors. Firstly the presence of the triazole group may be a major factor, as splicing efficiency decreased when the triethylene-azide (**51**) was ligated to the **G4-A4** and spliced where only 23% of the pre-mRNA was spliced compared with 65% for **A4** (Figure 4.4c). The second factor could be due to the lack of m⁷G cap at the 5' end of the transcript. The m⁷G cap had been shown to be an integral factor in the splicing of adenovirus 2 when investigated by Konarska et al.¹⁵⁵ An additional investigation using the transcript **A3** was carried out +/- m⁷G which resulted in splicing efficiency decreasing for the transcript without the m⁷G cap, although the ratio of upstream mRNA to total mRNA did not vary (Figure 4.10). Interestingly, when the spliceosome complex was analysed, the formation of the B complex from the A complex was stalled with the artificial tripartite constructs, this may also attribute to the overall decrease in splicing efficiency but is not conclusive and further information is required.

The overall decrease in splicing efficiency for the tripartite **ESE-A4** (Figure 4.3) constructs compared to the natural Ad1WW (**A2**, **A3** and **A4**, Figure 4.3) does not alter the conclusion that tethered ESEs were not able to reproduce the action of the integral ESE in **A3**, which enhanced splicing at the upstream 5' splice site. Instead the ESEs reverted back to the natural **A2** and **A4** splicing patterns where the downstream splice

site was used predominantly, suggesting that the 3-dimensional diffusion model (Figure 1.7) was not being utilised and contiguous RNA is required.

4.4 EXPERIMENTAL

4.4.1 Radioactive Transcription of G-Initiator (G4) Incorporated into Ad1WW (A4)

The transcriptions were carried out as stated in Section 2.4.3 with a reaction volume of 2.5 μ l, replacing the NTP mix concentration (1 mM GTP, 4 mM ATP, 4 mM CTP, 4 mM UTP). The 0.2 U diguanosine triphosphate was replaced with G-Initiators (**G4**) (4 mM). The reactions were then incubated for 4 hours and purified by gel electrophoresis (Section 2.4.3).

4.4.2 Click Reactions between G4-A4 and ESEs

Transcripts with a 5' alkyne, ESE-azide (10 μ M), Cu(I)THPTA (5 mM), aminoguanidine (40 mM) and RNaseOUT™ (2 %, v/v) were mixed, vortexed briefly and left to react at 40 °C for 20 minute. The reactions were loaded on a 6% denaturing gel for electrophoresis. The gel was exposed to X-ray film. The portion of gel containing the click products was excised and the products were eluted by soaking overnight at 4 °C. The RNA was precipitated with ethanol and dissolved in water.

4.4.3 Splicing Reaction for Transcripts Bioconjugated to Enhancer Sequences and Controls¹⁴⁸

A 10 μ l splicing reaction was performed for the transcripts bioconjugated to the enhancer sequence. MgCl₂ (3.2 mM), K-glutamate (50 mM), rATP (1.5 mM),

Phosphocreatine (20 mM), RNA template (1-2 μ l), RNaseout™ (5% v/v) and lastly HeLa cell Nuclear extract (5 μ l, Cilbiotech) was added to an eppendorf and mixed thoroughly. The reaction was then incubated for 90 minutes at 30°C. PK enzyme (48 μ l) was added to PK buffer (1 ml) and 42 μ l of this mixture was then added to the splicing reaction upon completion of incubation. The mixture was then incubated for 30 minutes at 37°C. The splicing reaction was then ethanol precipitated. Formamide dyes (10 μ l) were added to the dried splicing reaction, denatured at 80°C for 1 minute and run on a denaturing polyacrylamide gel. The gel was dried and exposed to a phosphor screen for analysis.

4.4.4 Splicing Reaction for Native gel Analysis¹⁷⁰

A 10ul splicing reaction was prepared for the transcript whereby MgCl₂ (3.2 mM), K-glutamate (50 mM), rATP (0.15 mM), Phosphocreatine (20 mM), RNA template (1 μ l), RNaseOUT™ (5% v/v) and lastly HeLa cell Nuclear extract (5 μ l, Cilbiotech) was mixed together in an eppendorf and left at 30°C for up to 60 minutes. Time points were taken at intervals 0, 1, 2.5, 5, 7.5, 10, 15, 30 and 60 minutes in which 2 x 2 μ l were taken from the reaction and frozen immediately in liquid Nitrogen. Once the time course was complete, the time points were separated, one half was PK treated as described in the previous splicing reaction, the other half was treated with Heparin (0.5ul, 4 ng/ μ l) for 30 minutes at r.t. TG loading buffer was added to the reaction and loaded onto a 2% LMP agarose gel (Invitrogen) which was run in TG running buffer for 4°C at 100 V for 4 hours. The gel was then dried and exposed to a phosphor screen for analysis.

4.4.5 Splicing of Transcripts for Native gel Analysis with the Addition of U6 Blocked Oligonucleotide¹⁷⁰

α -U6 oligonucleotide (1 μ l, 10 μ M) was added to HeLa cell Nuclear Extract (Cilbiotech) and incubated for 15 minutes at 30°C in order to block the A complex during the splicing reaction. This was then added to the splicing reaction and following procedure in section 4.4.4.

4.4.6 Radioactive Transcription of Amplified PCR products using [α -³²P] UTP

The [α -³²P] UTP labelled transcripts were generated using T7 RNA polymerase in 10 μ l reactions containing 1x T7 hot transcription buffer, NTP's (0.05 mM GTP, 0.5 mM ATP, 0.5 mM CTP and 0.05 mM UTP; Promega), diguanosine triphosphate (0.22 U, Roche), DTT (10 mM), [α -³²P] UTP (25-50 μ Ci), PCR product (25-150 ng/ μ l), T7 polymerase (5% v/v) and RNaseout (Invitrogen) (5% v/v). The reaction mixture was incubated at 37°C for 1-2 hours and then analysed on a 6% denaturing polyacrylamide gel. The gel was exposed to x-ray film and the band of RNA was cut out from the gel RNA elution buffer (300 μ l) was added to the excised band of RNA and left eluting overnight at 4°C. The eluent was removed from the gel and ethanol precipitated and resuspended in water ready for use.

4.4.7 Synthesis of ESE-Amines

ESE5-10-NH₂ (Table 4.3) were synthesised by automated solid phase using an ABI394 DNA synthesiser using TC-RNA chemistry. RNA reagents were supplied by Link Technology, cleavage and deprotection was undertaken according to the manufacturer's instructions.¹⁵⁷

Table 4.3 Sequences for 2'-hydroxy ESE-amines synthesised.

Name	3' and 5' Modified RNA Enhancer sequences synthesised (5'-3')
ESE5-NH₂	A _s G _s G _s A _s G _s GACGGAGGACGGAGGACA-NH ₂
ESE6-NH₂	A _s G _s G _s A _s G _s G _s A _s C _s G _s G _s A _s G _s G _s A _s C _s G _s G _s A _s G _s G _s A _s C _s A _s -NH ₂
ESE7-NH₂	A _s G _s G _s A _s G _s GACGGAGGACGGAGGACA-HEG-NH ₂
ESE8-NH₂	A _s G _s G _s A _s G _s G _s A _s C _s G _s G _s A _s G _s G _s A _s C _s G _s G _s A _s G _s G _s A _s C _s A _s -HEG-NH ₂
ESE9-NH₂	HEG-AGGAGGACGGAGGACGGAGGACA-NH ₂
ESE10-NH₂	HEG-AGGAGGACGGAGGACGGAGGACA-HEG-NH ₂

4.4.8 Solid phase NHS-Azide to ESE-Amine on CPG Resin

RNA coupled to 3'-Amino modifier C7 CPG 100 (Link Technologies) was Fmoc deprotected with a solution of 20% piperidine in DMF for 3x5 minutes washing the resin after each deprotection with DMF x2 and acetonitrile x3. **(47)** was dissolved in 30% solution of NaHCO₃ (0.025 M) in acetonitrile added to the resin and left rocking overnight. The solution was removed and the resin was washed with acetonitrile x3 and dried. The RNA was then deprotected and cleaved following Link Technologies protocol attaining **ESE5-10** (Table 4.4) at concentrations stated (Table 4.5).¹⁵⁷

Table 4.4 Sequences for 2'-hydroxy ESE-azides synthesised.

Name	3' and 5' Modified RNA Enhancer sequences synthesised (5'-3')
ESE5	A ₅ G ₅ G ₅ A ₅ G ₅ GACGGAGGACGGAGGACA-N ₃
ESE6	A ₅ G ₅ G ₅ A ₅ G ₅ G ₅ A ₅ C ₅ G ₅ G ₅ A ₅ G ₅ G ₅ A ₅ C ₅ G ₅ G ₅ A ₅ G ₅ G ₅ A ₅ C ₅ A ₅ -N ₃
ESE7	A ₅ G ₅ G ₅ A ₅ G ₅ GACGGAGGACGGAGGACA-HEG-N ₃
ESE8	A ₅ G ₅ G ₅ A ₅ G ₅ G ₅ A ₅ C ₅ G ₅ G ₅ A ₅ G ₅ G ₅ A ₅ C ₅ G ₅ G ₅ A ₅ G ₅ G ₅ A ₅ C ₅ A ₅ -HEG-N ₃
ESE9	HEG-AGGAGGACGGAGGACGGAGGACA-N ₃
ESE10	HEG-AGGAGGACGGAGGACGGAGGACA-HEG-N ₃

Table 4.5 Mass of reagents and final concentration of ESE obtained for the reactions between ESE coupled to resin and NHS-N₃ (47)

ESE	Mass of RNA coupled to Resin (mg)	Mass of NHS-N₃ (47) (mg)	Final Concentration of ESE obtained in 500ul H₂O (mM)
ESE5	28.4	12.4	0.237
ESE6	28.3	13	0.257
ESE7	27.7	11.4	0.304
ESE8	-	11	0.214
ESE9	40.1	7.9	0.382
ESE10	29.2	14.2	0.308

4.4.9 NHS-Azide Coupling to ESE-Amines

RNA-NH₂ solution was concentrated down to dryness and resuspended in a solution of NaHCO₃ (25 mM, 90 µl). A solution of (47) (100 eq, 210 µl) in acetonitrile was added to the RNA-NH₂, mixed thoroughly and left shaking overnight. An additional amount of (47) (100 eq, 20-40 µl) was added to the solution and left shaking for a further 5 hours to ensure the reaction had gone to completion. The solution was removed in a vacuum concentrator, and resuspended in 100 µl of water. The product was purified and desalted by a GE healthcare Nap-25 column, freeze-dried and dissolved in 100-500 µl of water obtaining **ESE5-10** (Table 4.4) at concentrations stated (Table 4.6).

Table 4.6 Concentration of ESE obtained for the reaction between ESE-NH₂ and NHS-N₃ (47).

ESE	Number of moles used of ESE-NH₂ (μmol)	Concentration Obtained of ESE-N₃ (mM)
ESE5	0.118	0.823
ESE6	0.128	0.615
ESE7	0.152	0.824
ESE8	0.107	0.736
ESE9	-	1.74
ESE10	0.154	0.956

4.4.10 Buffers

- TG loading buffer: 50 mM Tris base, 50 mM glycine, 40% v/v glycerol, bromophenol blue, xylene cyanol.
- TG running buffer: 50 mM Tris base, 50 mM glycine.

5 CONCLUSION AND FUTURE WORK

RNA splicing is highly complex and involves hundreds of cis-acting and trans-acting splicing factors which contribute to its regulation. One such cis-acting splicing factor are ESEs which are usually purine-rich sequences located in exons at distances of up to several hundred nucleotides from the target splice sites.^{47,74} As discussed in Chapter 1 (Section 1.4.1.2.3) ESEs are important in stimulating splicing at both the 5' and 3' splice sites. SR proteins bind to ESEs stabilising the binding of the three canonical splicing signals: U1 snRNP, U2 snRNP and U2AF protein. The mechanism by which enhancers exert their effects has not been determined, although there have been two distinct models proposed:

1. The 3-dimensional diffusion model: where an SR protein associates with the ESE, directly encountering the target proteins (U1/U2 snRNP or U2AF) at the 5' and 3' splice sites, thereby creating a loop in the intervening RNA (Figure 1.7a).
2. The 1-dimensional propagation model: an SR protein binds to the ESE, interacting with the target proteins by cooperatively binding to other RNA binding proteins across the RNA (Figure 1.7b).

The objective of this study was to test the first model. In order to achieve this, a chemical biological tool was required to extrapolate whether or not the first model was being used during splicing.

The study primarily focused on how ESEs exerted their effects on 5' splice sites. A splicing assay was developed using the Ad1WW mini-gene, which contained two identical alternative 5' splice sites in exon 1, that when spliced produced 2 mRNA

isoforms (Figure 2.7a). A 23 nucleotide ESE based on GGA repeats^{36,142} replaced the natural sequence upstream of the 5' splice site (**A3**, Figure 2.7b). The initial findings showed that the natural sequences (**A2** and **A4**, Figure 2.7b) spliced producing the two mRNA isoforms (Figure 2.9a & c), whereas the addition of the ESE shifted splice site selection from the distal downstream splice site to the proximal upstream splice site producing only the upstream mRNA isoform (Figure 2.9b).

The enhanced splicing at the upstream 5' splice site due to the insertion of the ESE upstream of the 5' splice sites enabled a system to be developed that could determine whether the ESE was stimulating splicing using the 3-dimensional diffusion model. A non-flexible RNA linker (PEG) was inserted between the ESE and the RNA transcript (**A4**) using copper catalysed click chemistry to bioconjugate the two strands of RNA together (Figure 4.1). The hypothesis was that if splicing at the upstream 5' splice site was still enhanced with the addition of PEG which was unable to bind RNA binding proteins it would imply that the SR protein bound to the ESE was able to interact with the U1 snRNP at the upstream 5' splice site, and stimulate splicing there, thereby bypassing the linker by the formation of a loop. However, if splicing at the upstream splice site was inhibited then the linker was interfering with the ESE enhancing abilities, implying that continuous RNA was required for the ESE to exert its effects.

A 44 nucleotide model transcript was used to optimise the protocol for synthesising the RNA tripartite constructs. G-initiator (**G4**) was synthesised by solid phase synthesis using standard phosphoramidite chemistry (Scheme 3.7) where a hexanyl phosphoramidite was incorporated onto the 5' end of the guanosine affording at 26% yield. Incorporation of (**G4**) into the 44 nucleotide transcript (**44-mer**) occurred during transcription. Optimal concentration of (**G4**) was imperative to achieve maximum

incorporation; optimisation found that 0.4mM of (**G4**) resulted in optimal incorporation of 72% according to quantitation by gel electrophoresis (Figure 3.14).

The ESE-azides (**ESE1-8**) were prepared by solid phase synthesis using 3'-amino-modified supports which were coupled to NHS-N₃ (**47**) after aminolytic cleavage from the solid support. Three variables were investigated:

- i. ESE sugar structure (2'-OMe or 2'-OH)
- ii. Nature of the phosphate backbone (phosphodiester and phosphorothioate)
- iii. Non-RNA tether (short alkyl chain or HEG chain)

The click reaction between **G4-44** and **ESE1** was optimised in order to attain the highest yield with the least amount of degradation, in which 10 µM of ESE was required (Table 3.6). Interestingly the phosphorothioate **ESE3** appeared to enhance the yield of the click reaction compared to the corresponding phosphodiester **ESE1** attaining yields of 44% and 12% respectively (Figure 3.16). The click reaction between **G4-A4** and **ESE1-8** proceeded smoothly, albeit low yields and due to slower electrophoretic mobility were separated and purified by gel electrophoresis ready for splicing

The RNA splicing preferences of the tripartite of the ESE-conjugated pre-mRNAs (**ESE1-8-A4**, Figure 4.4 & 4.9) were tested alongside the controls (**A2**, **A3** and **A4**). The natural transcripts **A2** and **A4** spliced predominantly at the downstream splice site whereas **A3** (the transcript with the ESE incorporated upstream of the 5' splice sites) stimulated splicing to the upstream splice site, producing mostly upstream mRNA. Strikingly none of the ligated ESEs whether based on 2'-OMe or 2'-OH chemistry, were able to enhance splicing at the upstream splice site. This phenomenon was seen

irrespective of the length of the linker (Figures 4.4 and 4.9). From these findings it was clear that artificially tethered ESEs were unable to reproduce the action of the integral ESE.

As a result it was concluded that the stimulation of splicing at the proximal (upstream) 5' splice site by an ESE does not involve looping of the RNA, i.e. the 3-dimensional diffusion of the SR protein bound to the ESE and the U1 snRNP situated at the 5' splice site did not occur. It would appear that contiguous RNA is required for stimulation at the proximal 5' splice site by the ESE and therefore the propagation of proteins along the exon from the ESE to the U1 snRNP is most likely occurring.

However, further tests are still required to confirm that the 3-dimensional diffusion model is not being utilised in splicing. As only one ESE is being tested in this system it would be interesting and useful to see what would happen if different ESE sequences are used. Different ESEs may exert their effects in different ways. Nonetheless as a result of this study a tool has been developed whereby flexible non-RNA linkers can be inserted into the 5' end of pre-mRNA using click chemistry to conjugate two strands of RNA together,

5.1 FURTHER WORK

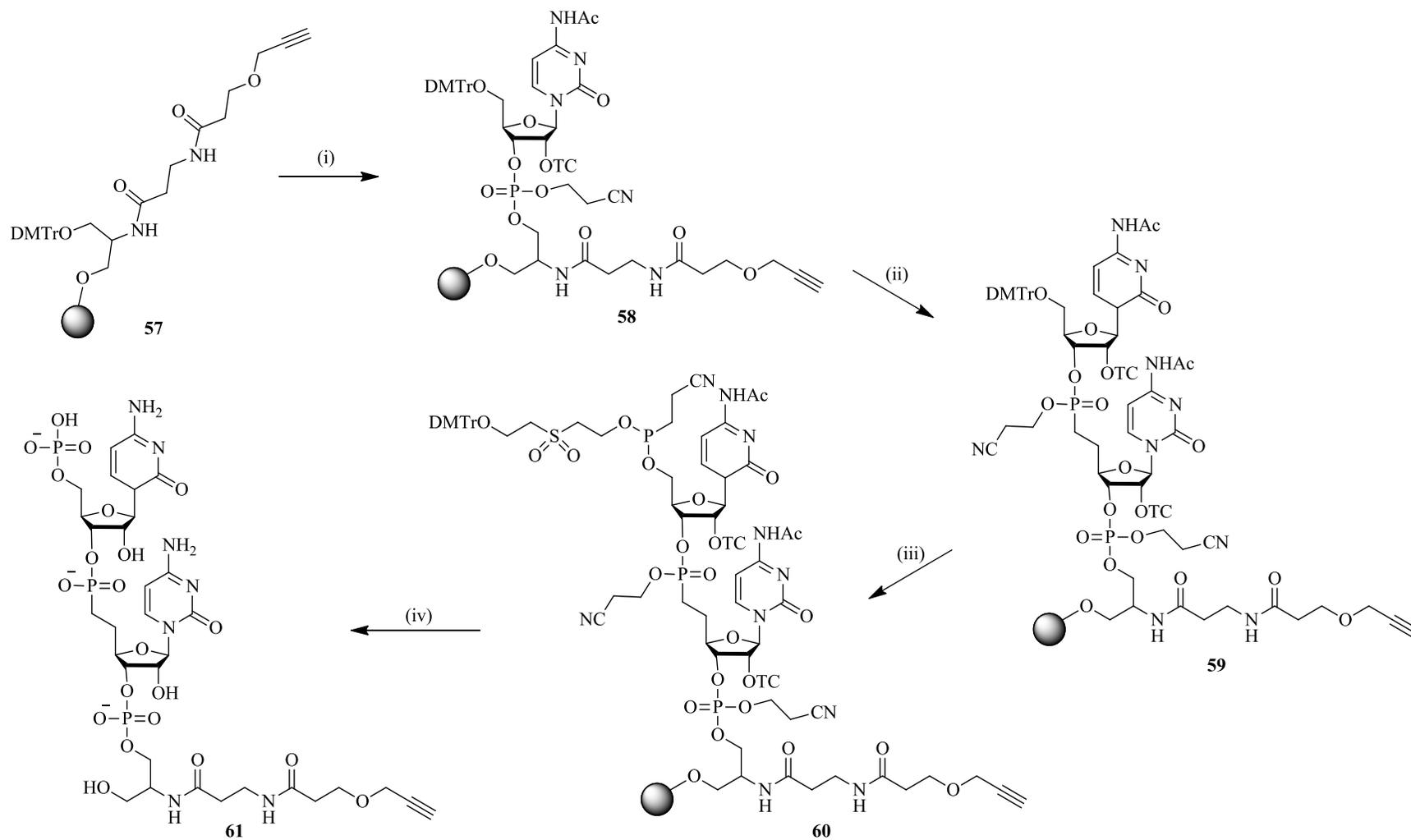
This study has focussed on the 5' end of the pre-mRNA, however ESEs may act differently in enhancing the 3' splice site, therefore a system is required to insert a non-RNA linker between the 3' splice site and the ESE (Figure 1.8b). As described in Chapter 1 (Section 1.6.1.2), modified dinucleotides (pNp) can be ligated to the 3' end of RNA by T4 RNA ligase. Jäschke *et al.* designed and synthesised modified dinucleotides (**10**) and ligated them to short strands of RNA.^{94,95} It was therefore decided to design

and synthesise by solid phase synthesis a dinucleotide (**61**) that incorporated an alkyne group to the 3' end.

5.1.1 Synthesis of the 3' Alkyne Modified Di-Cytosine (61)

The 3' alkyne modified di-cytosine (**61**) was synthesised by standard phosphoramidite chemistry using phosphoramidite coupling using 3'-alkyne modifier serinol CPG (**57**) (Glen research, Scheme 5.1). In order to obtain sufficient quantities 8 x 1 μ mol (**57**) solid supports were used. After purification by semi-preparative RP-HPLC, the respective fractions were combined, freeze-dried and redissolved in 100 μ l of water. The absorbance of (**61**) was measured at 260 nm to be 1.71, an overall concentration 12.13 mM, attaining 15% yield. The sample was reanalysed by RP-HPLC (Figure 5.1a) and the mass was determined by LC/MS to be 963.1879 (Figure 5.1b).

Scheme 5.1 Schematic representation of the solid phase synthesis of pCpC dinucleotide (**61**); (i) and (ii) coupling of Ac-C TC-RNA phosphoramidite (**55**), (iii) coupling of chemical phosphorylation agent, (iv) cleavage and deprotection attaining dinucleotide (**61**).



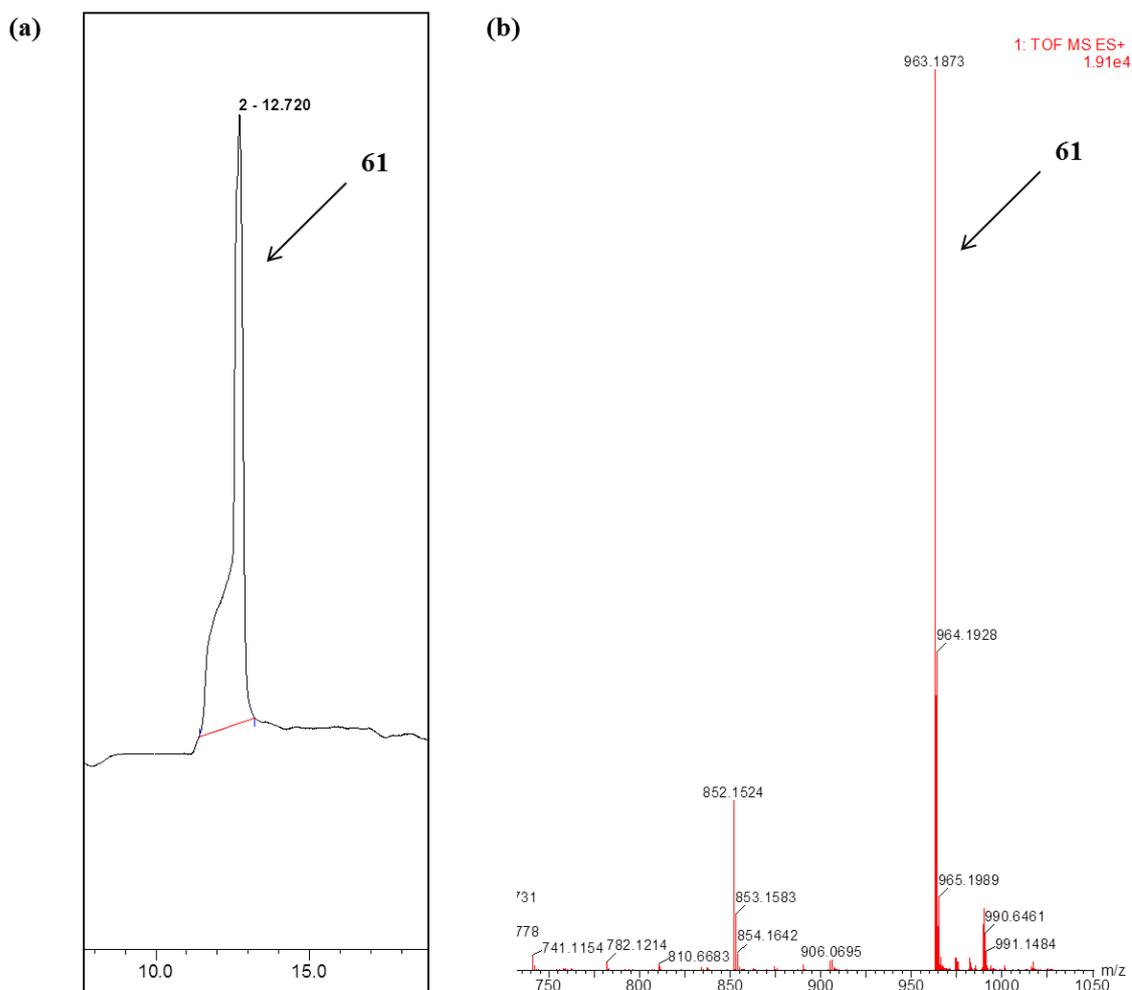


Figure 5.1 (a) RP-HPL chromatogram of purified dinucleotide (61**) and (b) corresponding LC/MS spectrum.**

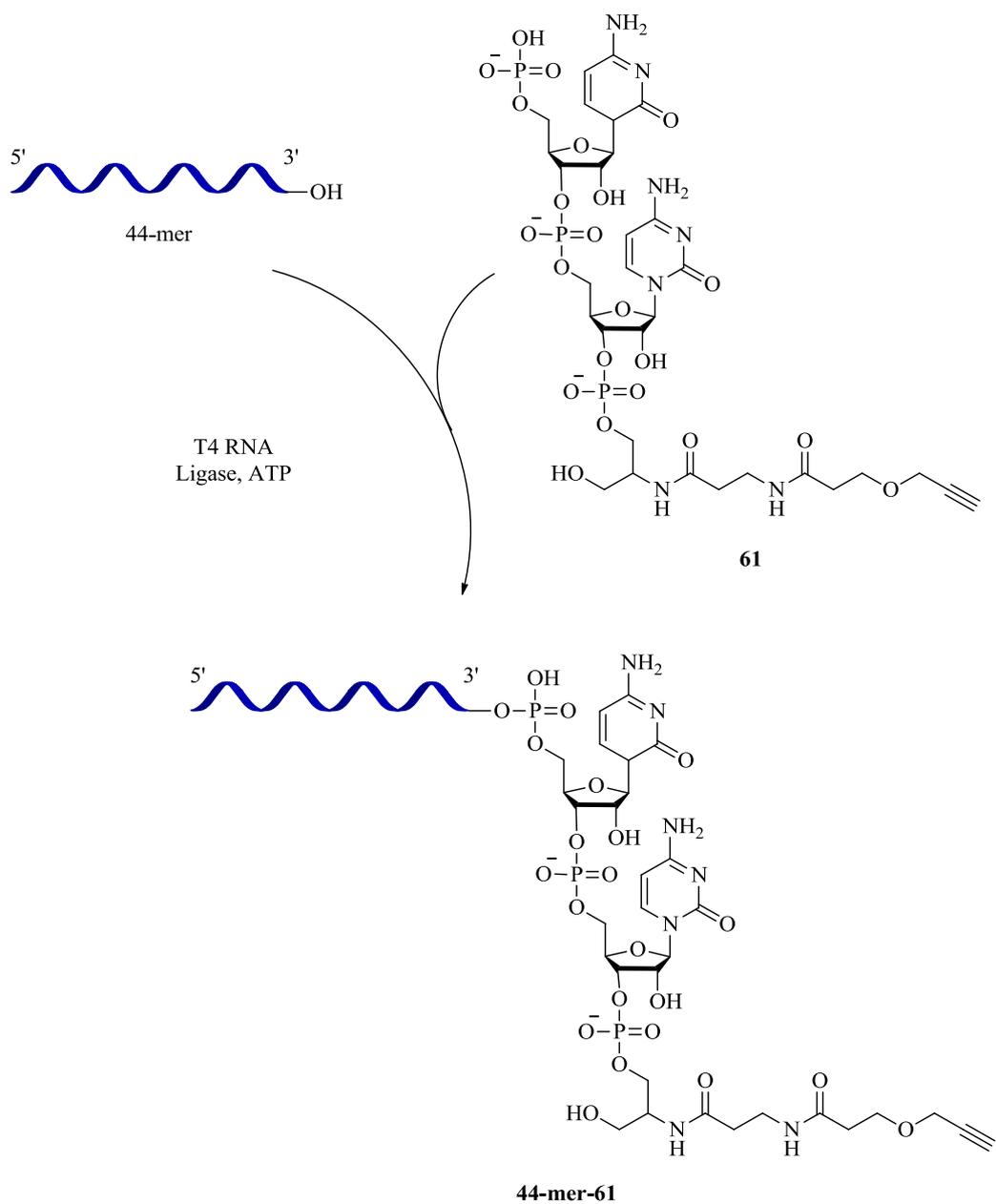
Due to time constraints and problems with the RP-HPLC confirmation of the presence of alkyne on (**61**) by the click reaction with FTU-N₃ (**43**) was not possible, it was therefore decided to ligate (**61**) to the 44 nucleotide transcript by T4 RNA ligase directly.

5.1.2 Ligation of (**61**) to 44 Nucleotide Transcript

(**61**) was ligated to the 44 nucleotide transcript which had been resuspended in RNA ligation buffer in the presence of T4 RNA ligase (4U) (Scheme 5.2). In order to determine the optimal concentration of (**61**) required, a concentration test was

undertaken whereby 0, 0.5, 1, 1.5, 2, 2.5 mM of (**61**) was added to the 44-mer and left incubating at 4°C overnight. The ligation reactions were run on a 10% polyacrylamide gel and analysed. Only one band was observed in the ligation reaction in the absence of (**61**) whereas two bands were observed for the ligation reaction where (**61**) was added for all the concentrations (Figure 5.2). The band with the slower electrophoretic mobility was thought to be the ligated transcript due to the increase in molecular weight. In order to confirm the presence of (**61**) at the 3' end of the transcript further experiments are required including the reaction between the transcript and an ESE-N₃ or fluorophore-N₃ via click chemistry.

Scheme 5.2 Schematic representation of the ligation between the 44 nucleotide transcript and the dinucleotide (61).



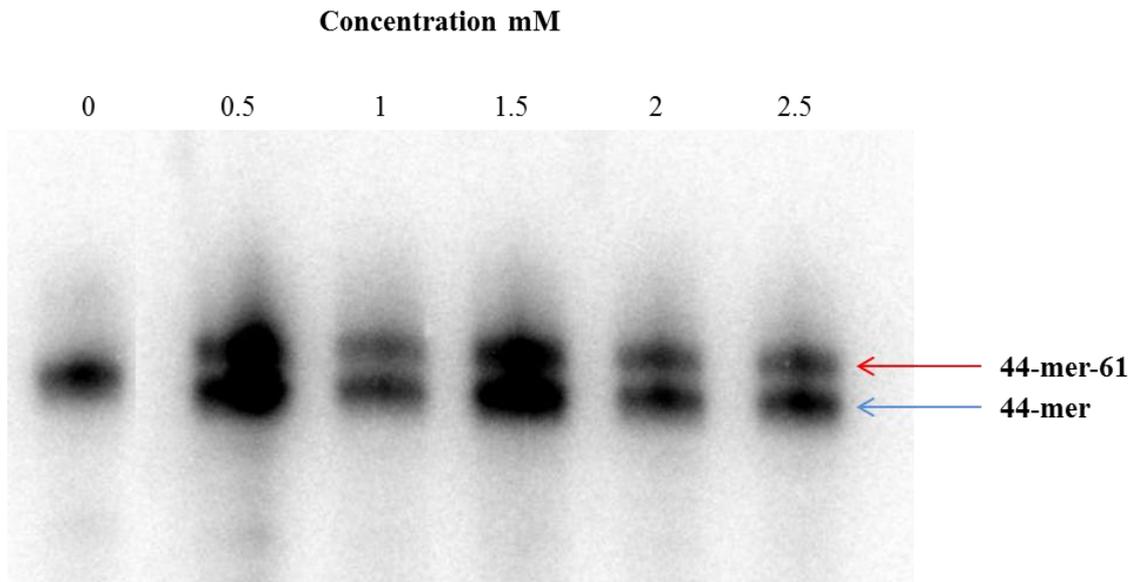


Figure 5.2 Concentration test for the ligation of dinucleotide (61) to the 44 nucleotide RNA transcript using T4 RNA ligase.

These preliminary results show that incorporation of the dinucleotide (**61**) occurred, although further optimisation is required to ensure maximum incorporation is achieved. The concentrations of (**61**) used so far did not produce fully incorporated **44-mer-61** and therefore higher concentrations of (**61**) may be required, additionally the incubation time may need to be increased. Upon optimisation of ligation of (**61**) into the 44-mer, conjugation to ESEs can be tested and optimised.

The preliminary data achieved so far represents a feasible method by which linkers can be inserted into the 3' end of pre-mRNA, enabling investigations into the mechanism of ESEs on the 3' splice site to take place.

5.1.3 Experimental

5.1.3.1 Synthesis of dinucleotide-alkyne (61)

The dinucleotide was synthesised using standard solid phase synthesis using 3'-alkyne modifier serinol CPG (**57**) (Glen research), Ac-C TC-RNA phosphoramidite (**55**) (0.1 M, Link technologies), Chemical phosphorylation agent (0.1 M, CPR, Link technologies).¹⁵⁷

5.1.3.2 RNA ligation between dinucleotide-alkyne (61) and transcript

RNA transcript in RNA ligation buffer (5 µl) was added to pCpC-Alkyne (5 µl, 0-2.5 mM). T4 RNA ligase (0.4 µl, 10 U/µl) was then added. The reaction mixture was mixed together and left incubating overnight at 4°C. The RNA ligation mixture was then made up to 100 µl and phenol/chloroform extracted and ethanol precipitated. Formamide dyes were added and the product was run on a denaturing polyacrylamide gel and analysed.

5.1.3.3 Buffers

- RNA ligation buffer: 0.2 mM ATP, 75 mM HEPES pH7.5, 30 mM MgCl₂, 4.5 mM DTT, 20% v/v DMSO.

6 APPENDIX

Table 6.1 Miscellaneous DNA and RNA sequences used.

Name	Sequence
DNA 21-mer	≡-TCGTTAGCATATATGGACATA
RNA1	≡-HEG-G _{os} A _{os} U _{os} U _{os} U _{os} U _{os} G _{os} U _{os} C _{os} U _{os} A _{os} A _{os} A _{os} A _{os} C _{os}
RNA2	≡- HEG-G _o A _o U _o U _o U _o U _o G _o U _o C _o U _o A _o A _o A _o A _o C _o
44-mer	GTAGGGGCCAGCCTACTGGCTGGTCCTCATGACCCTCTCTGCAG
α-U6	CUGUGUAUCGUUCCAAUUUU

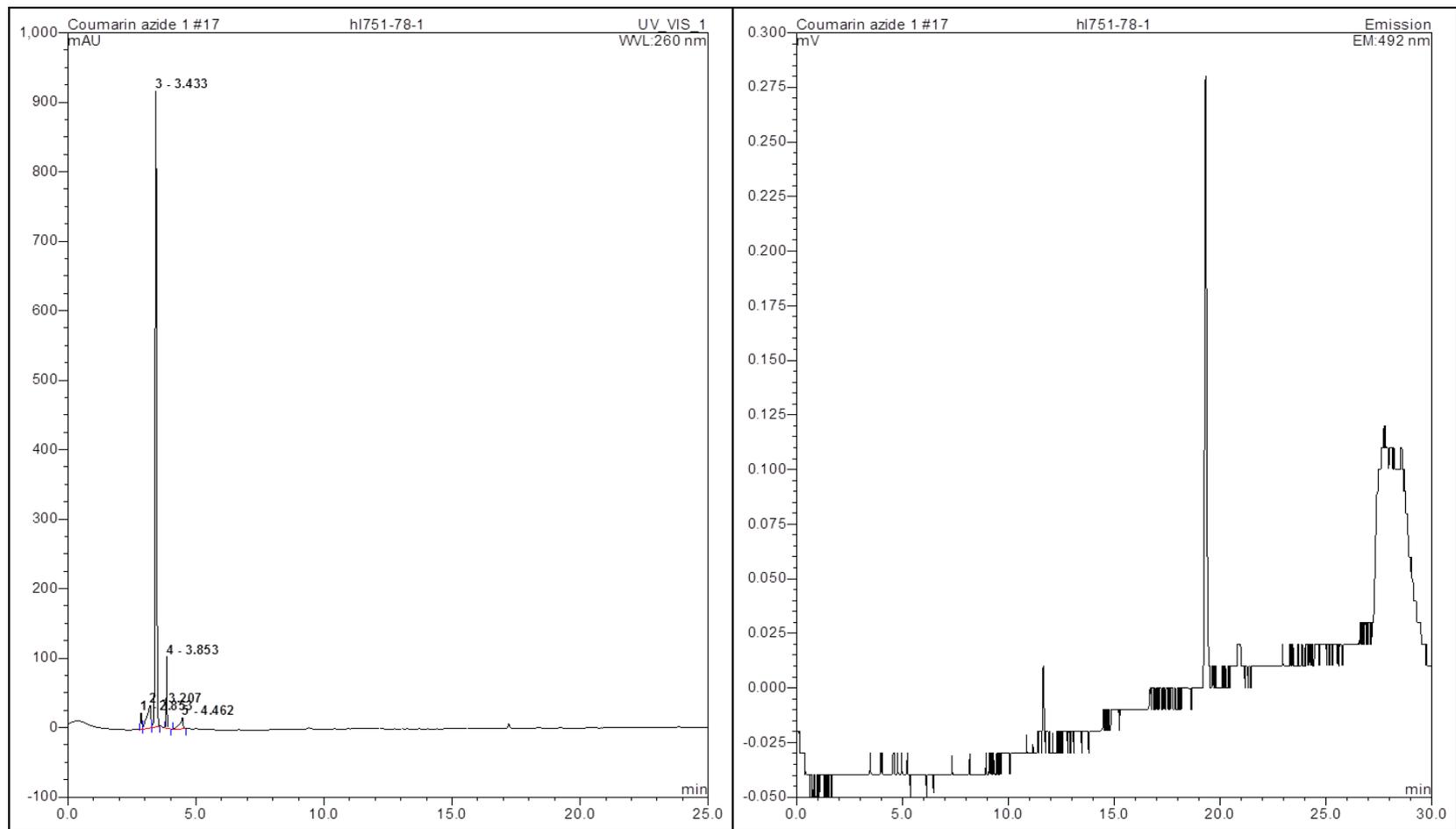


Figure 6.1 RP-HPL chromatograms for reaction between G1 and (43) in the presence of Cu(I)THPTA (Table 3.2, reaction 1). Left chromatogram at 260 nm and the corresponding chromatogram at 492 nm on the right.

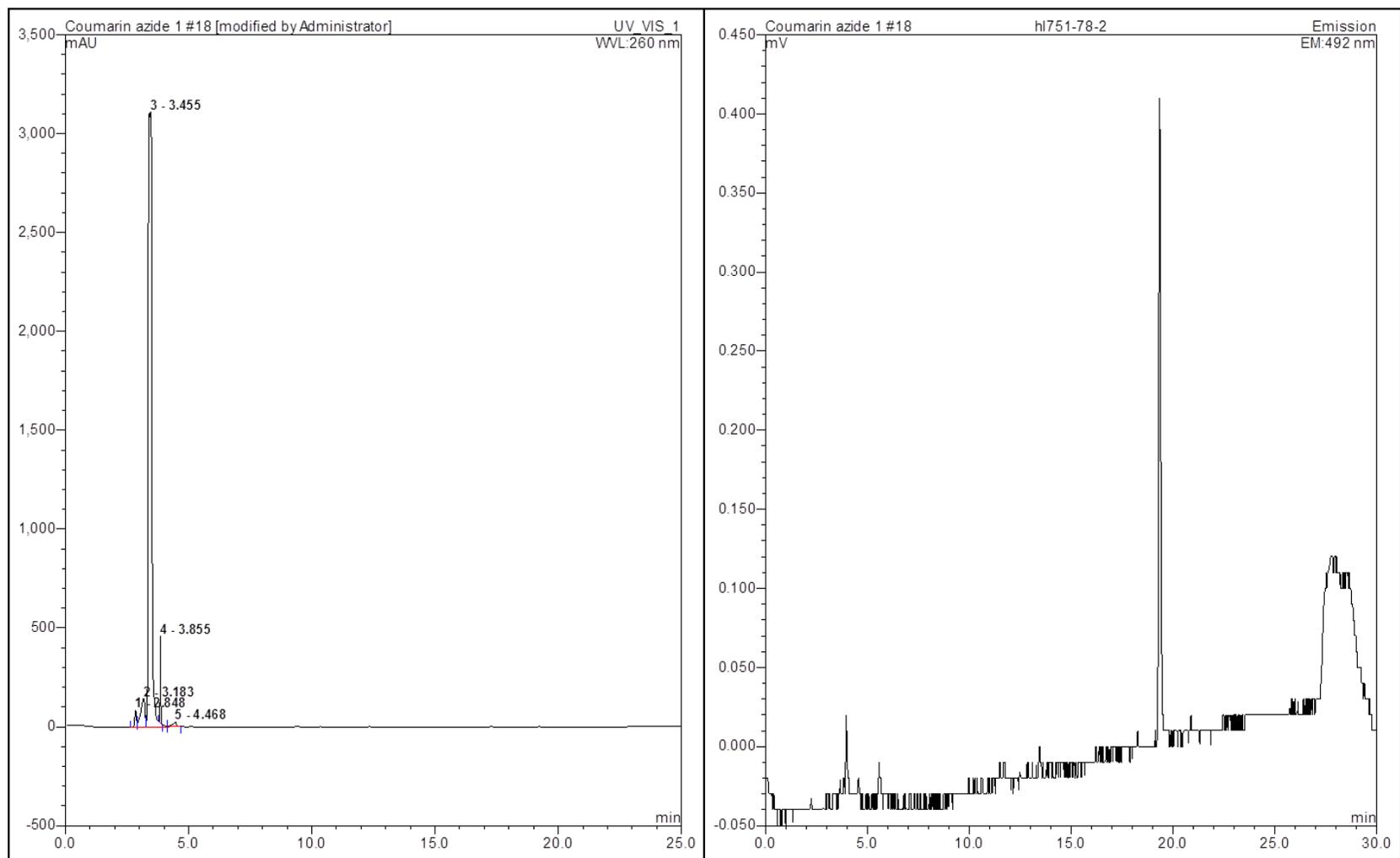


Figure 6.2 RP-HPL chromatograms for reaction between G1 and (43) in the presence of Cu(I)THPTA (Table 3.2, reaction 2). Left chromatogram at 260 nm and the corresponding chromatogram at 492 nm on the right.

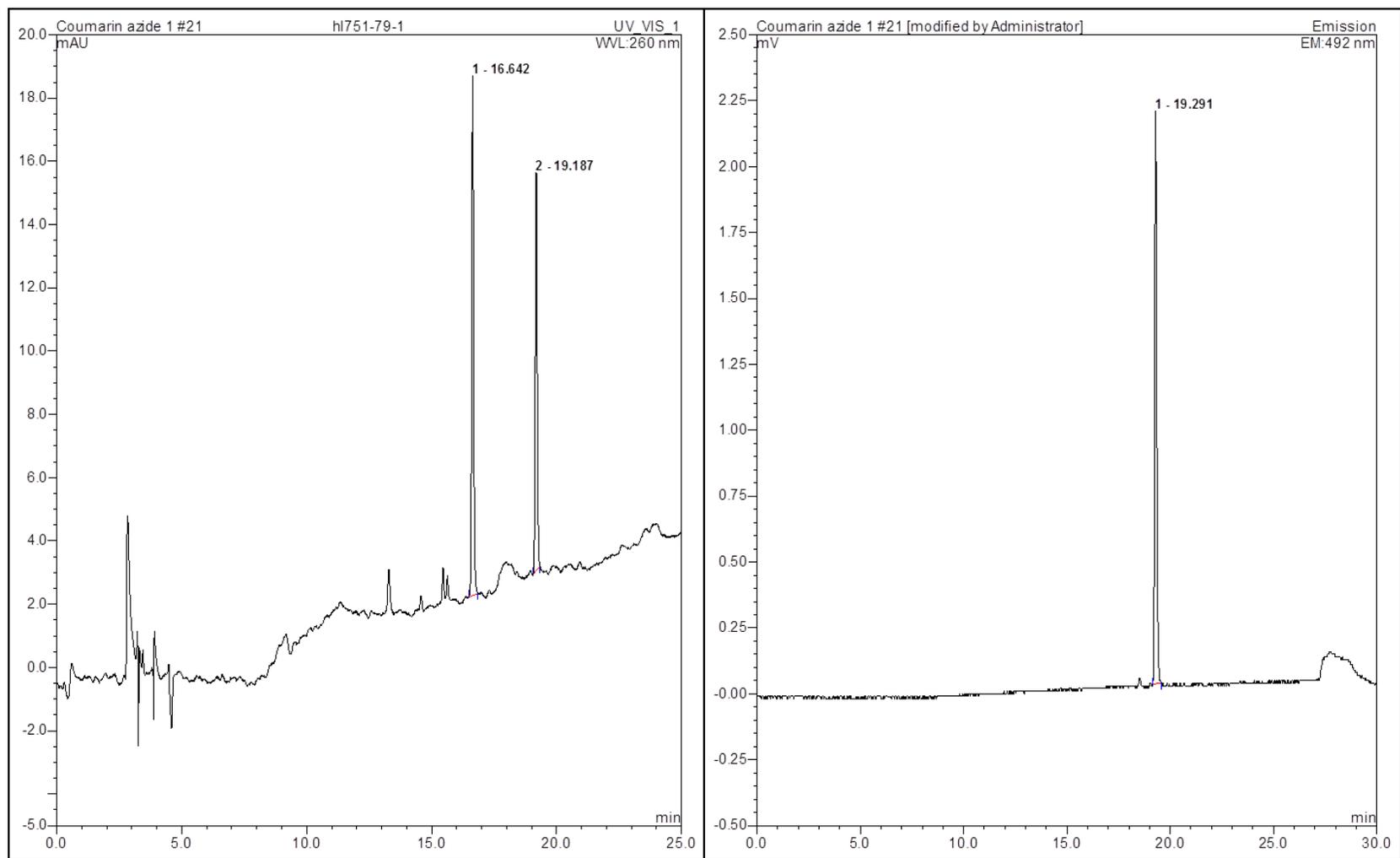


Figure 6.3 RP-HPL chromatograms for reaction between G1 and (43) in the presence of Cu(I)THPTA (Table 3.2, reaction 3). Left chromatogram at 260 nm and the corresponding chromatogram at 492 nm on the right.

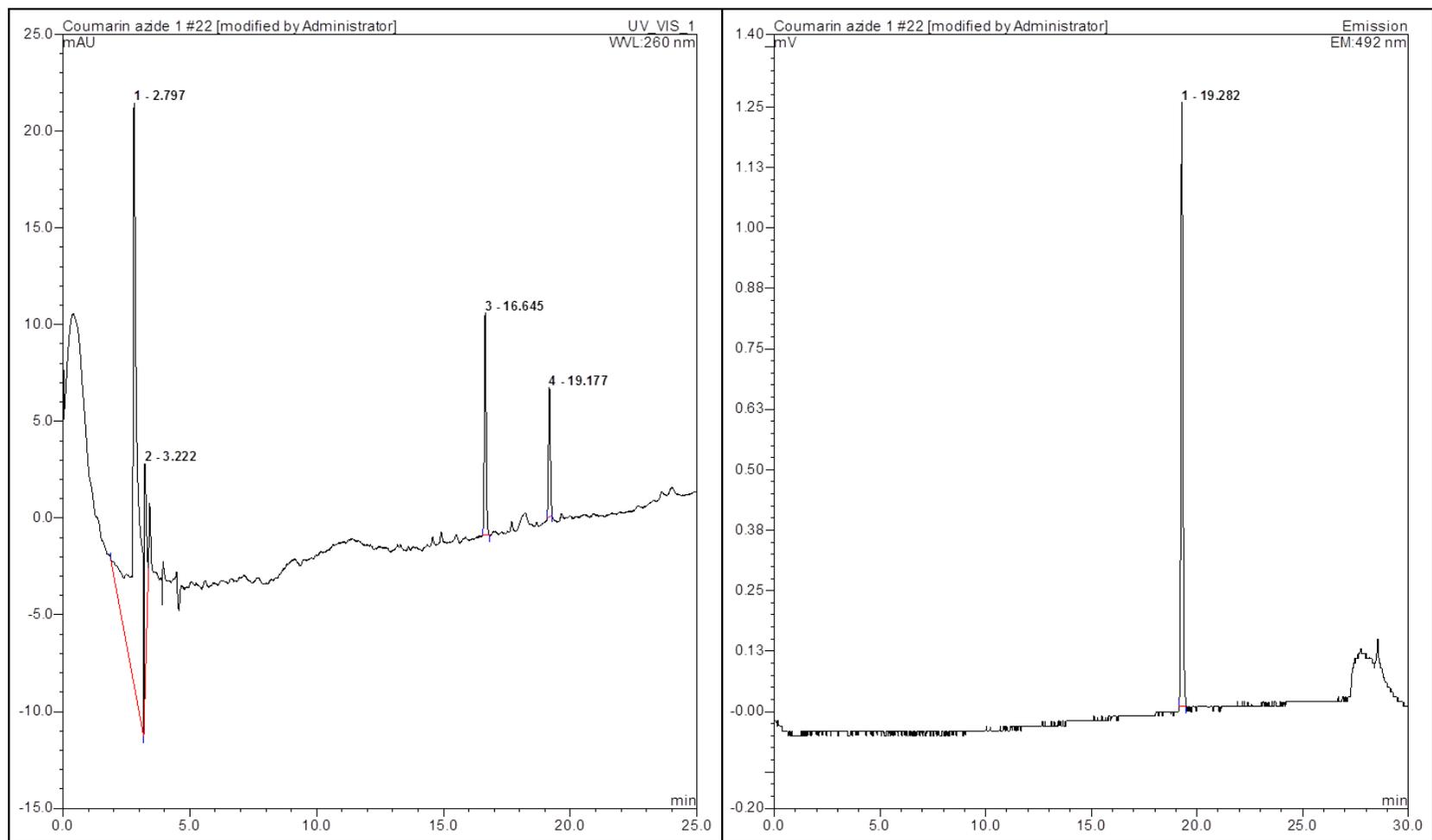


Figure 6.4 RP-HPL chromatograms for reaction between G1 and (43) in the presence of Cu(I)THPTA (Table 3.2, reaction 4). Left chromatogram at 260 nm and the corresponding chromatogram at 492 nm on the right.

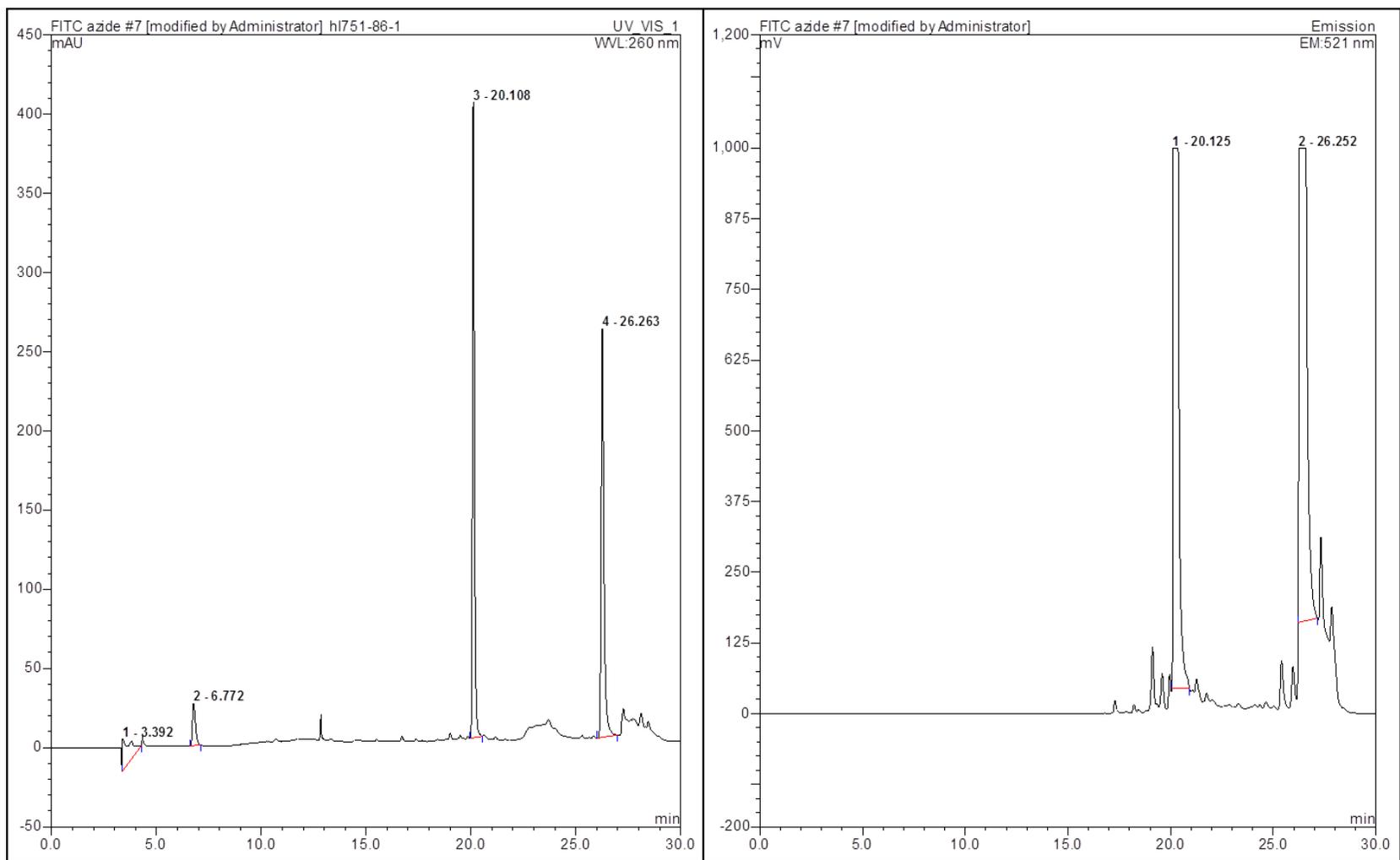


Figure 6.5 RP-HPL chromatograms for reaction between G1 and (43) in the presence of Cu(I)TBTA (Table 3.2, reaction 5). Left chromatogram at 260 nm and the corresponding chromatogram at 492 nm on the right.

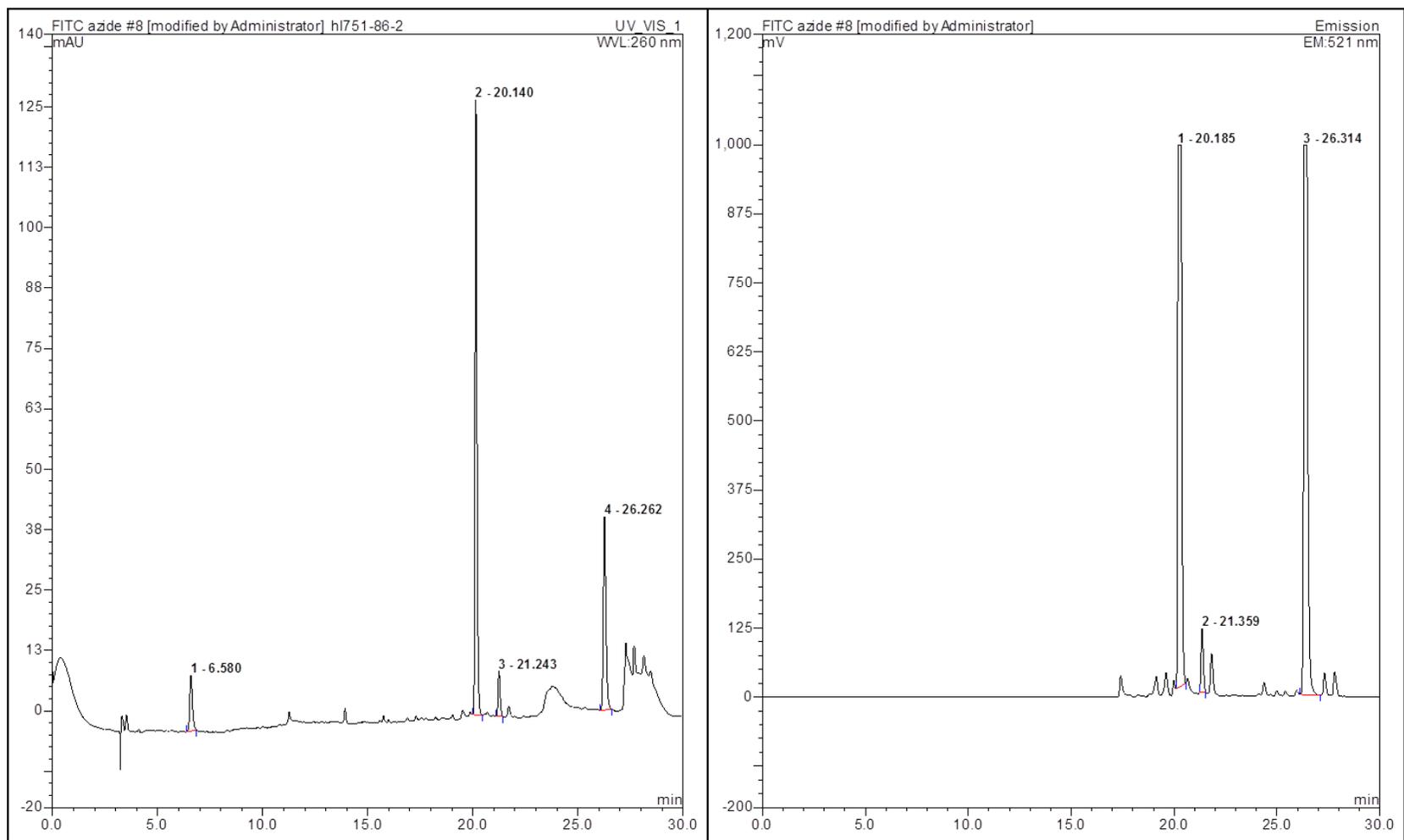


Figure 6.6 RP-HPL chromatograms for reaction between G1 and (43) in the presence of Cu(I)TBTA (Table 3.2, reaction 6). Left chromatogram at 260 nm and the corresponding chromatogram at 492 nm on the right.

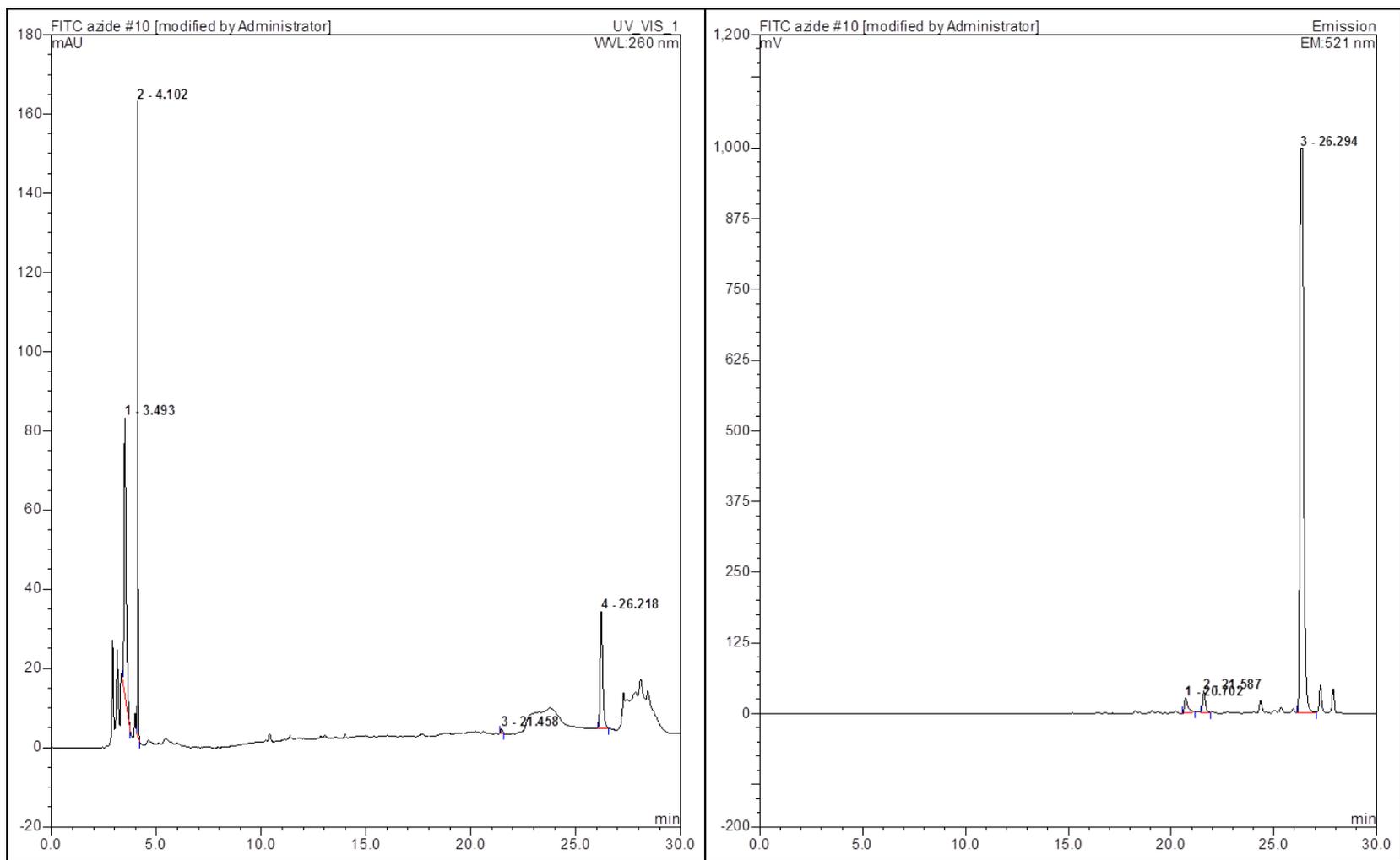


Figure 6.7 RP-HPL chromatograms for reaction between G1 and (43) in the presence of Cu(I)THPTA (Table 3.2, reaction 7). Left chromatogram at 260 nm and the corresponding chromatogram at 492 nm on the right.

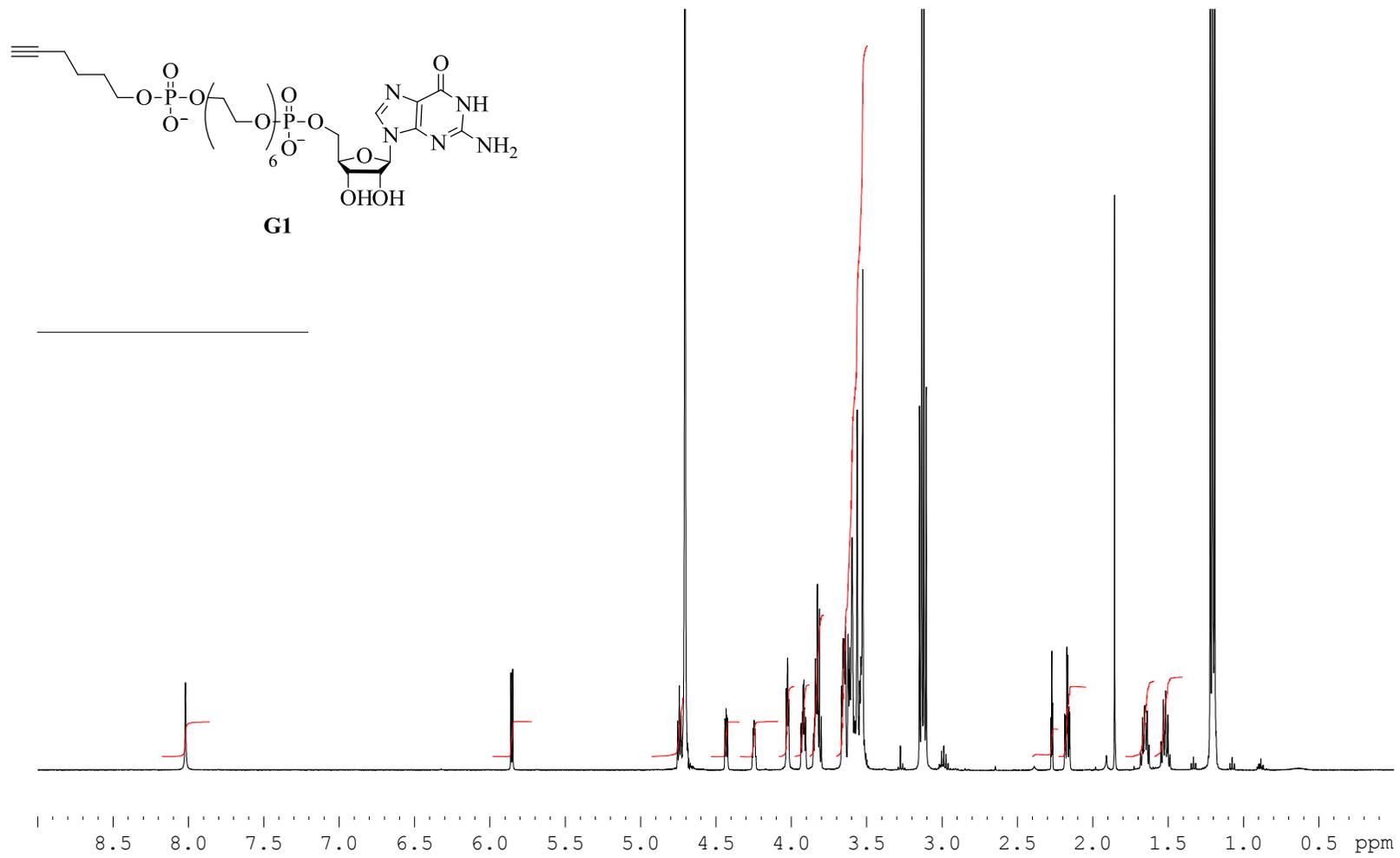


Figure 6.8 ¹H NMR (500MHz, D₂O) of G-Initiator (G1)

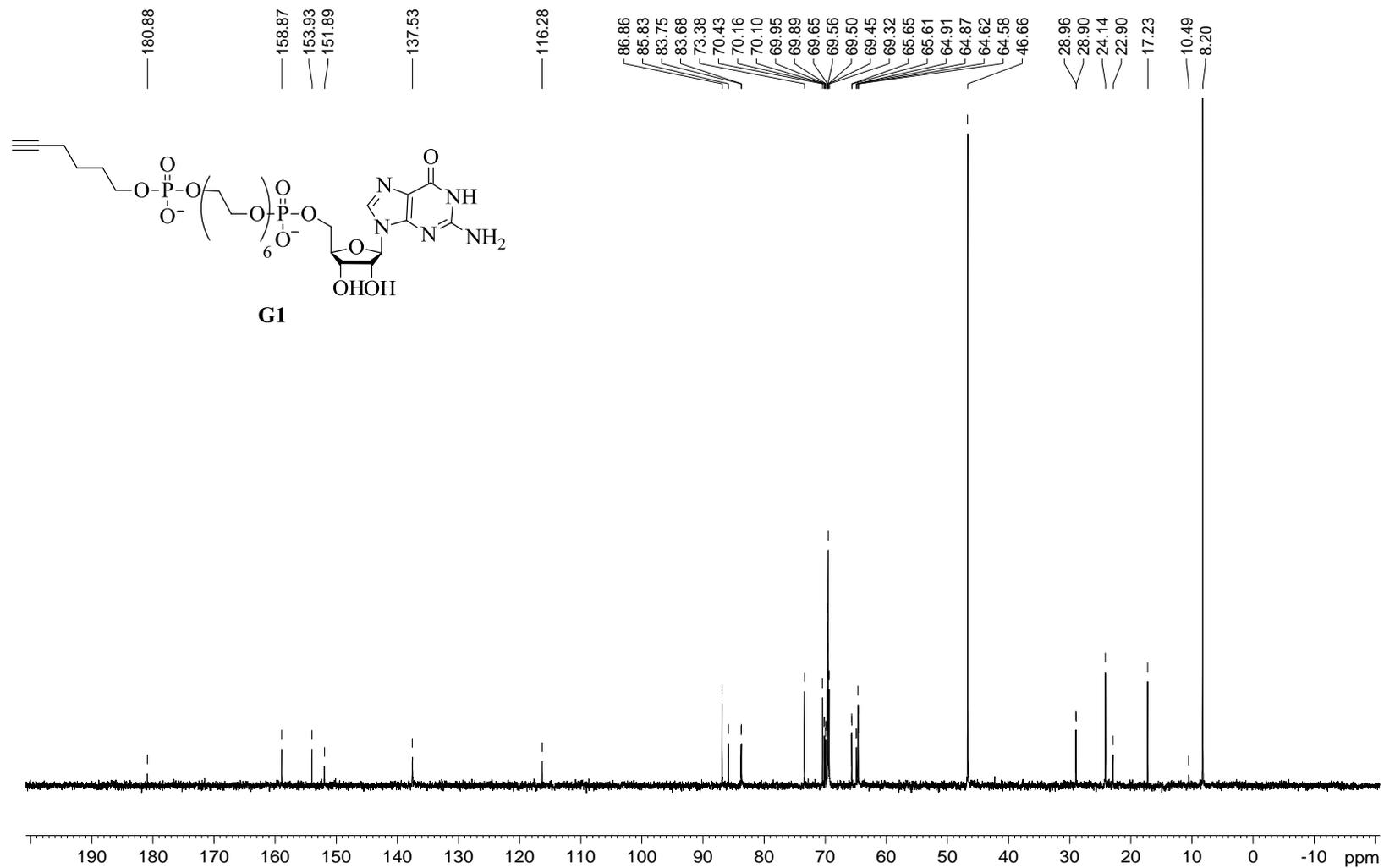


Figure 6.9 ¹³C (500MHz, D₂O) NMR of G-Initiator (G1)

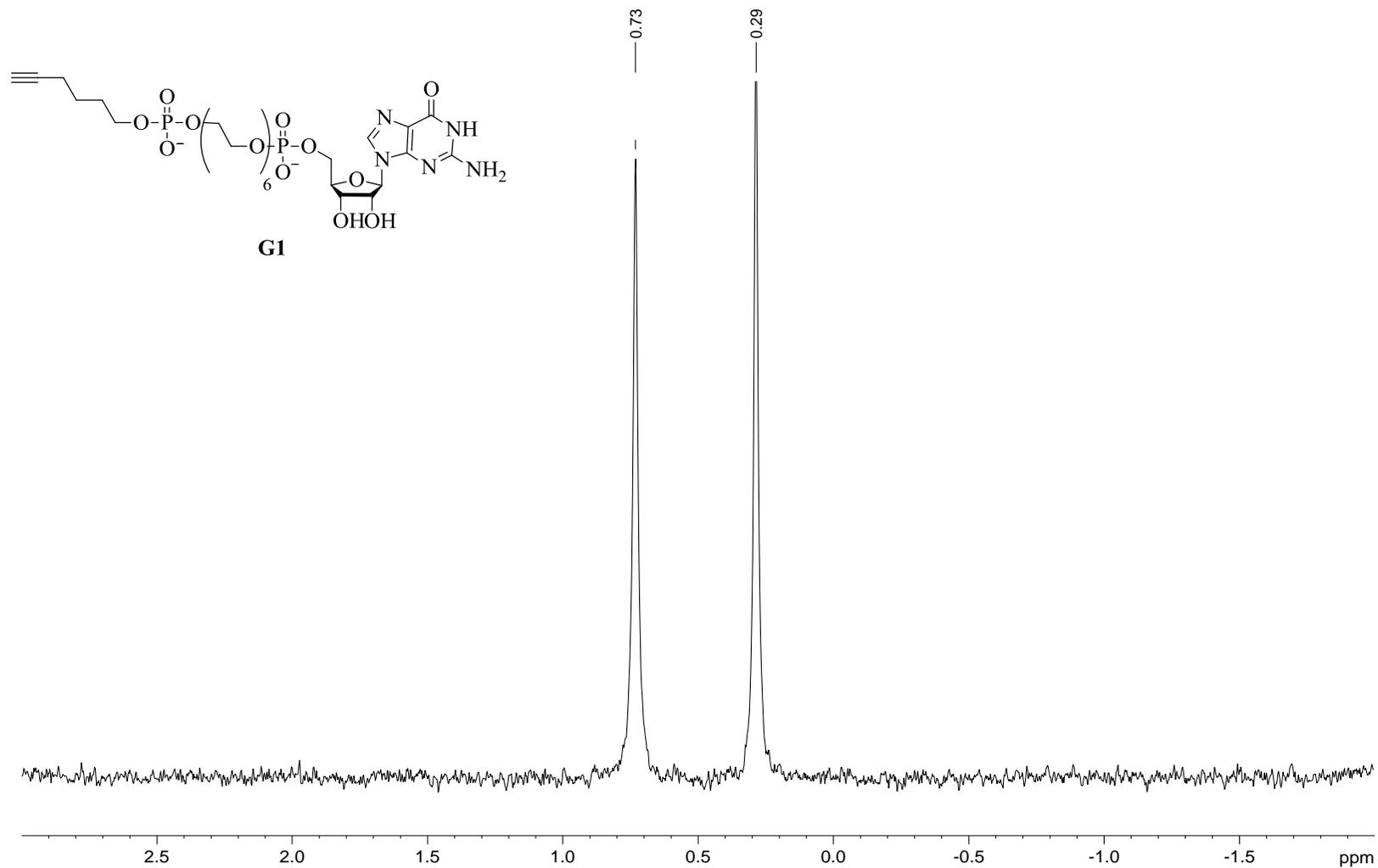


Figure 6.10 $^{31}\text{P}\{^1\text{H}\}$ NMR (500MHz, D_2O) of G-Initiator (G1)

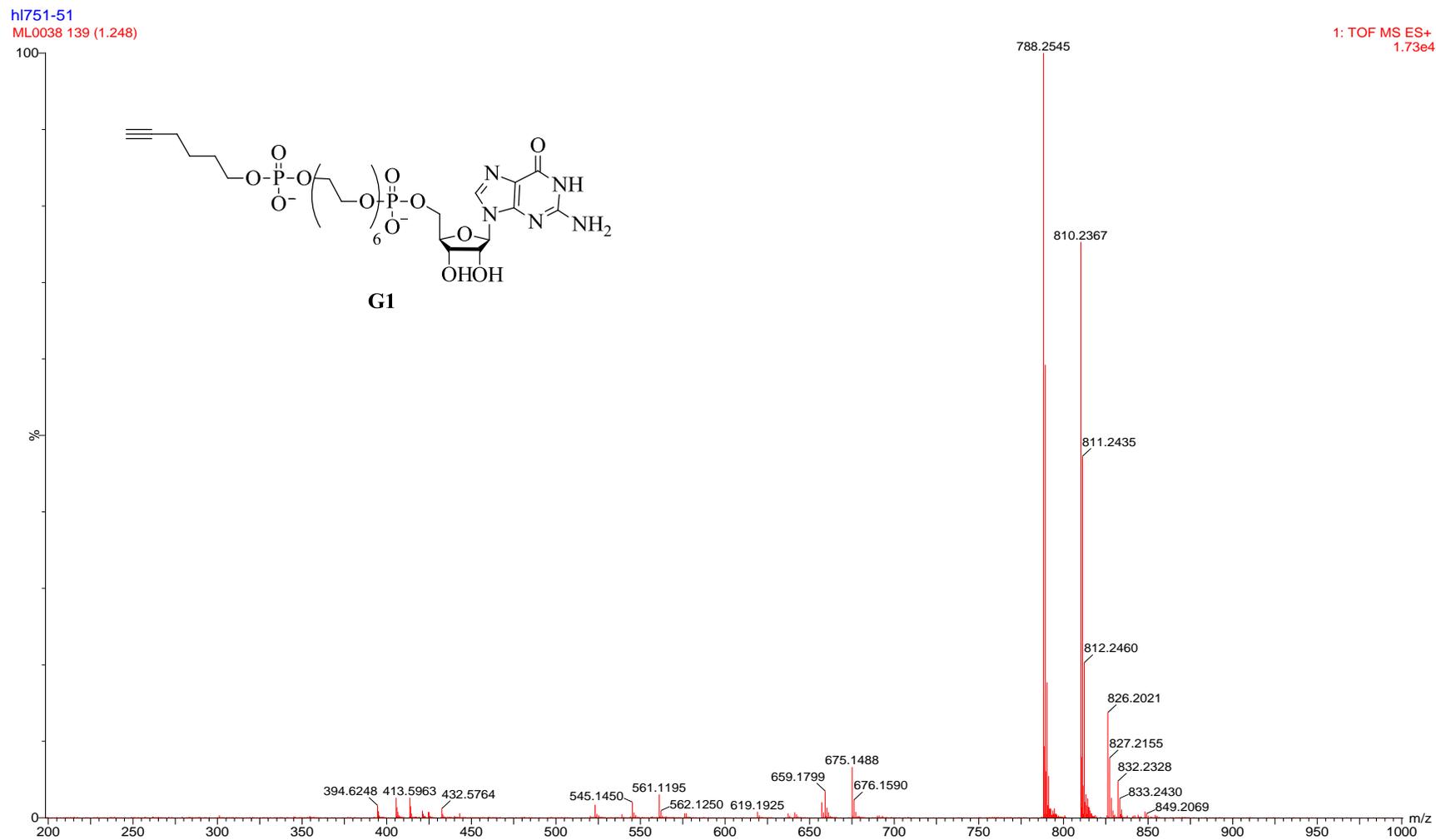


Figure 6.11 LC/MS of G-Initiator (G1)

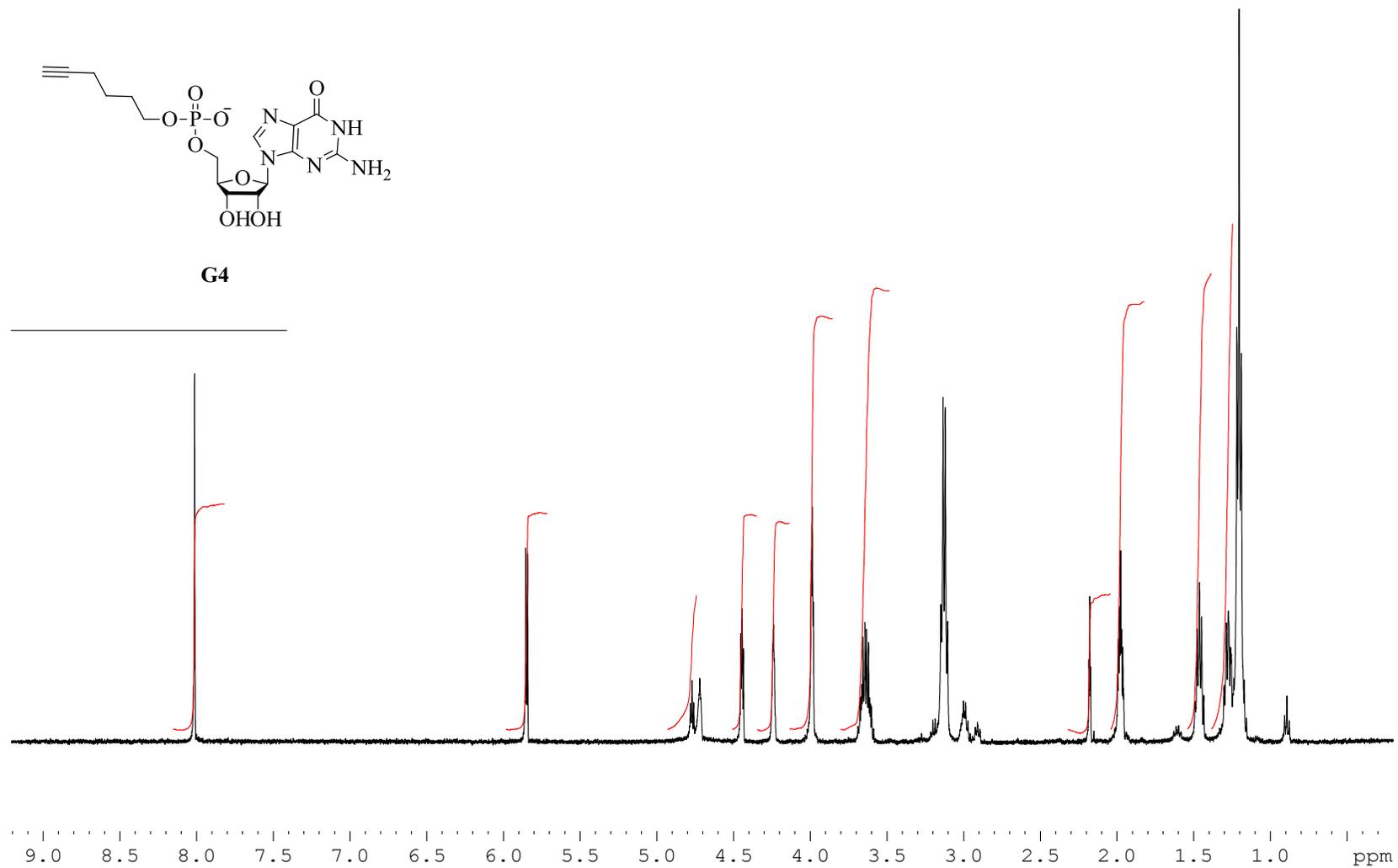


Figure 6.12 ¹H NMR (500MHz, D₂O) of G-Initiator (G4)

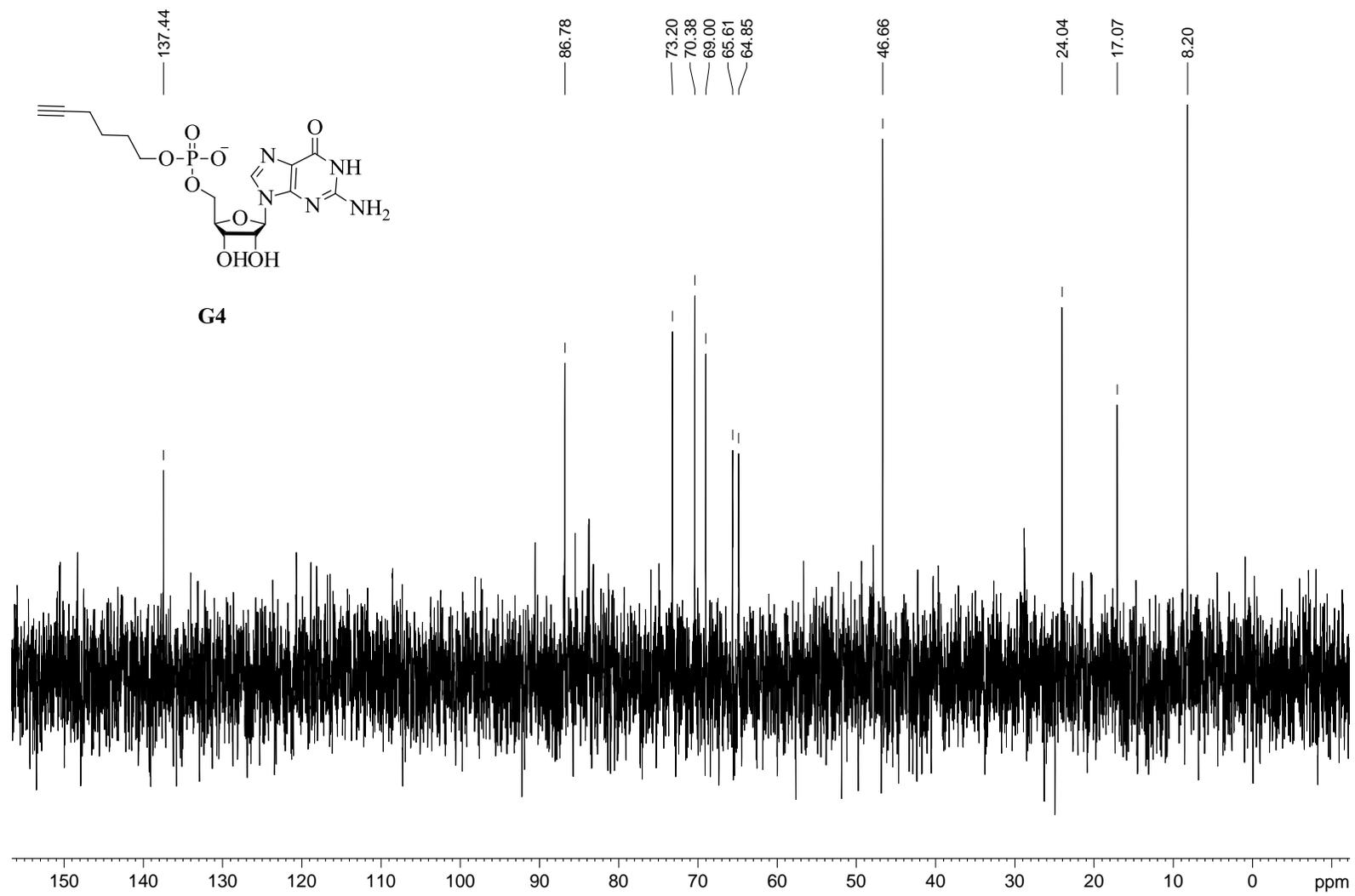


Figure 6.13 ¹³C NMR (500MHz, D₂O) of G-Initiator (G4)

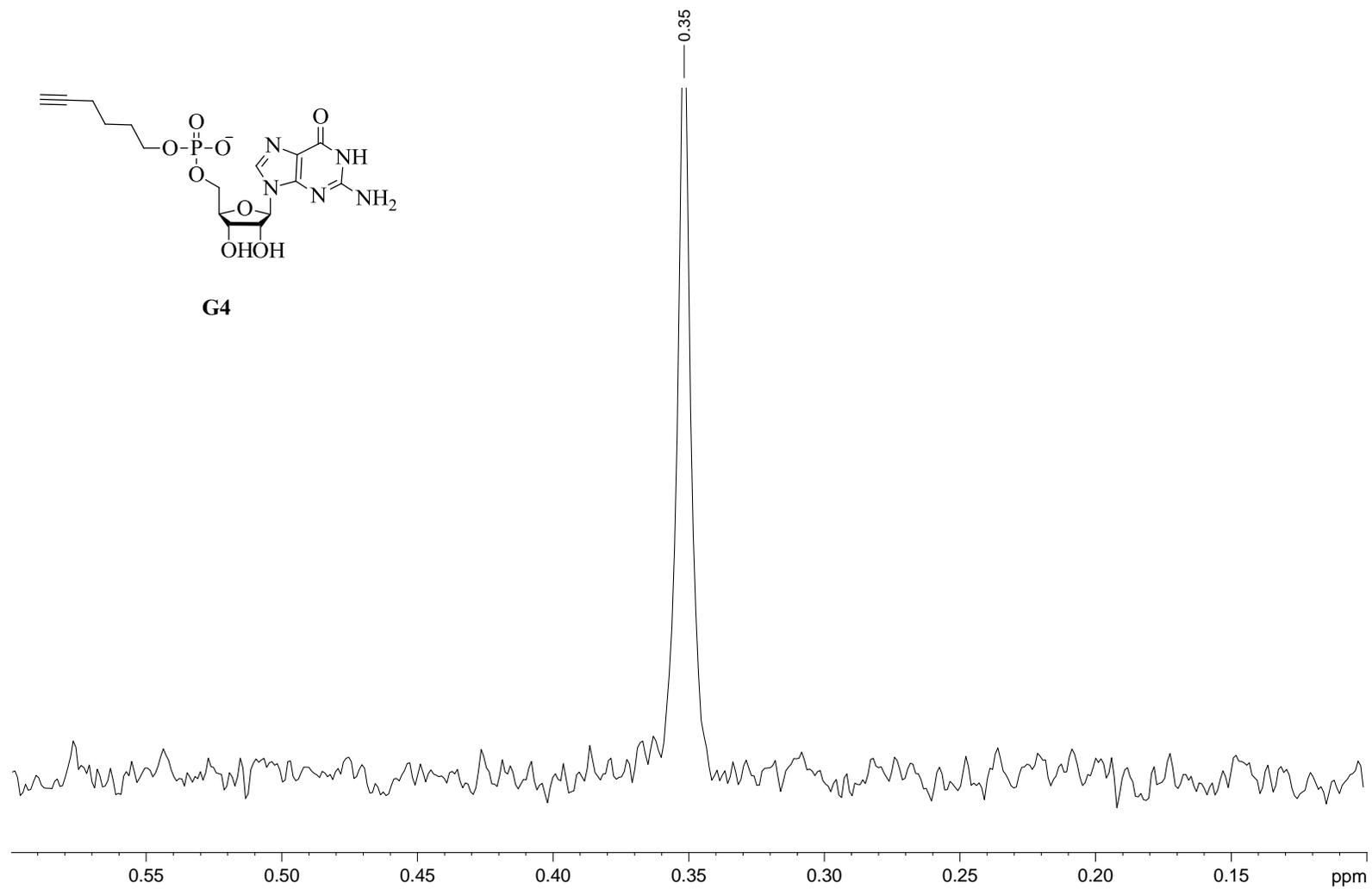


Figure 6.14 $^{31}\text{P}\{^1\text{H}\}$ NMR (500MHz, D_2O) of G-Initiator (G4)

HL 751-92

GE HL 11 126 (1.140) Cm (121:126-85:98)

1: TOF MS ES+
3.30e3

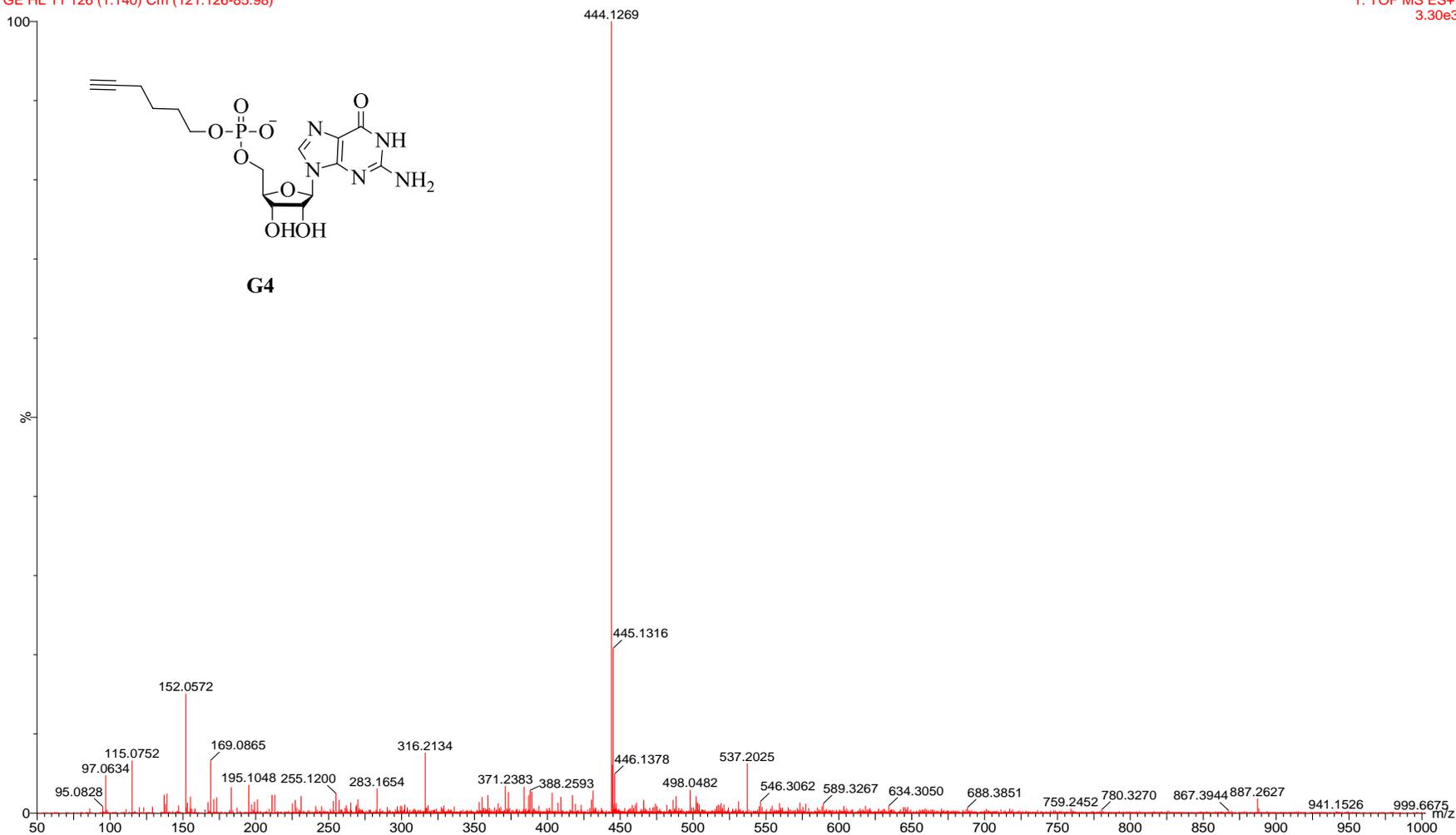


Figure 6.15 LC/MS of G-Initiator (G4)

HL 751-92

GE HL 11a 71 (1.126) Cm (71:74-57:62)

1: TOF MS ES-
3.43e3

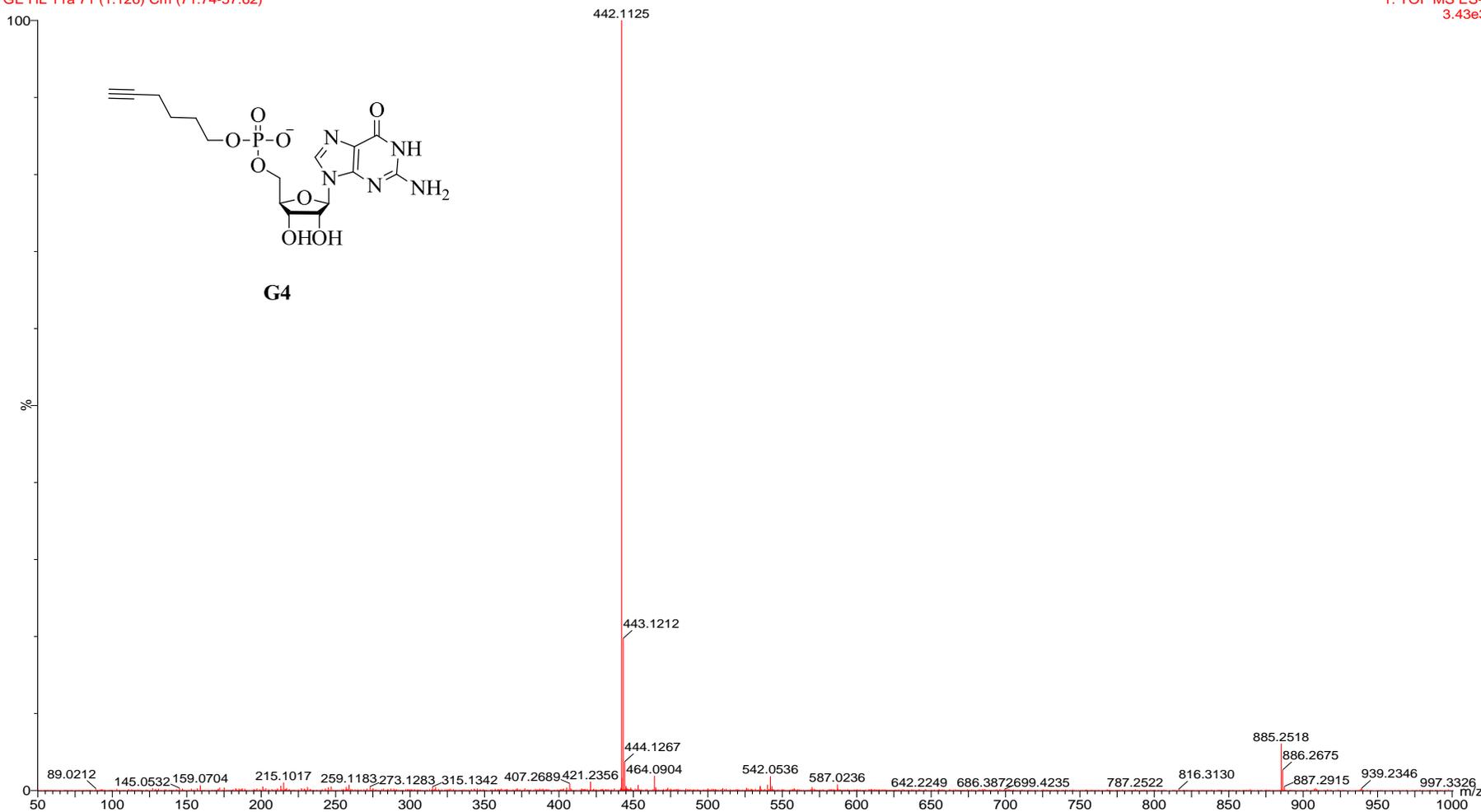


Figure 6.16 LC/MS of G-Initiator (G4)

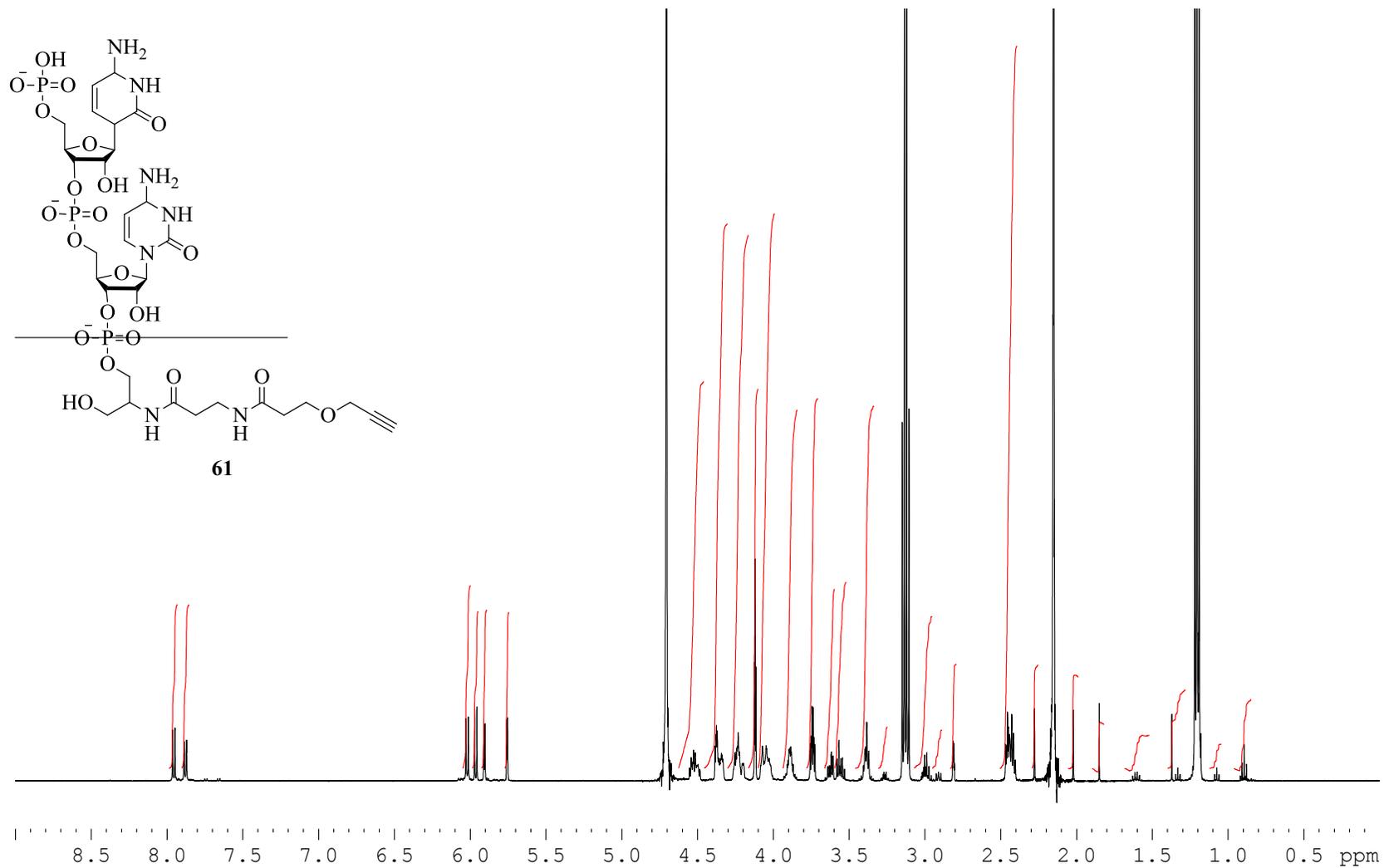


Figure 6.17 ^1H NMR (500MHz, D_2O) of pCpC-alkyne (61)

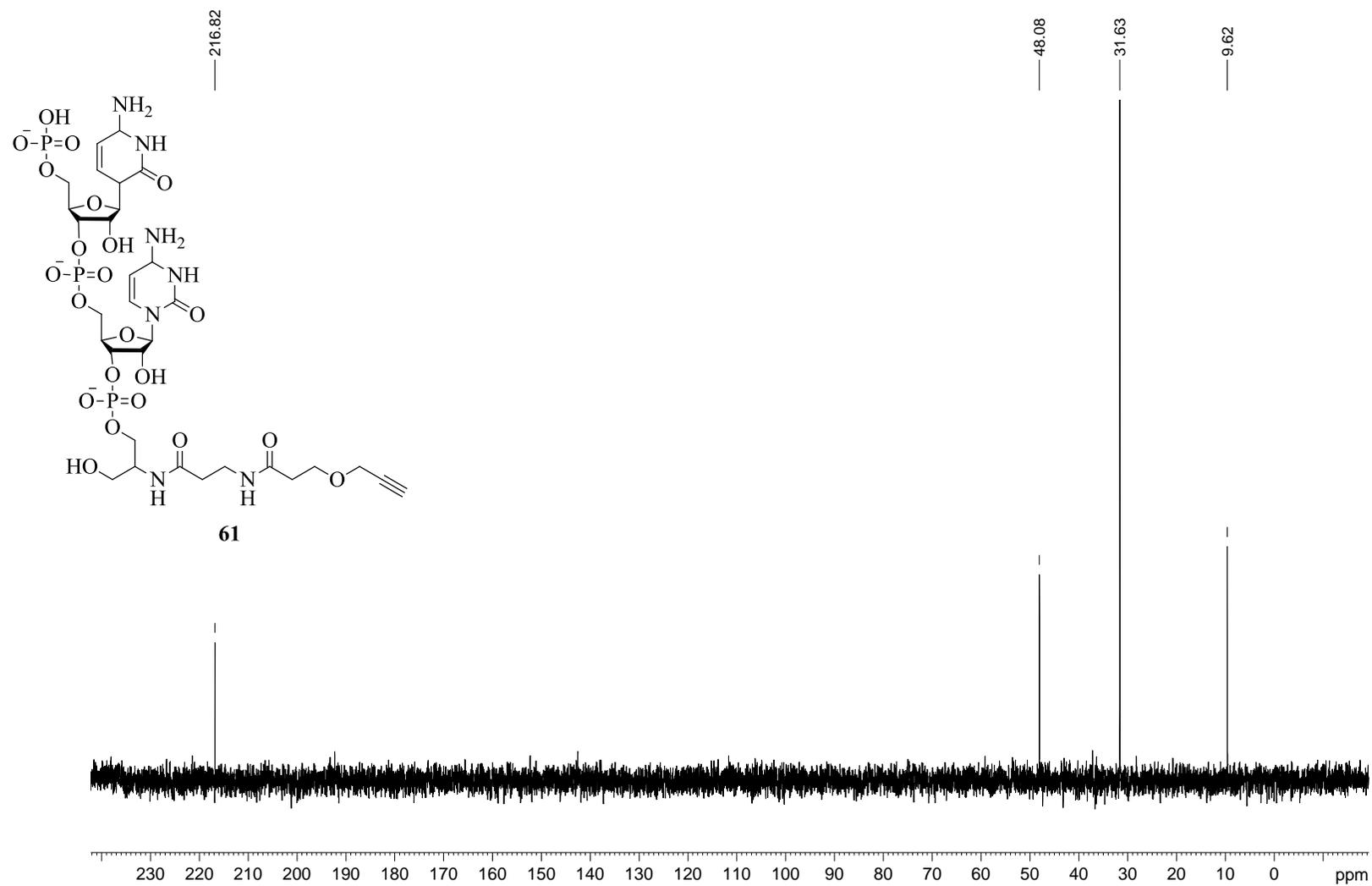


Figure 6.18 ^{13}C NMR (500MHz, D_2O) of pCpC-alkyne (61)

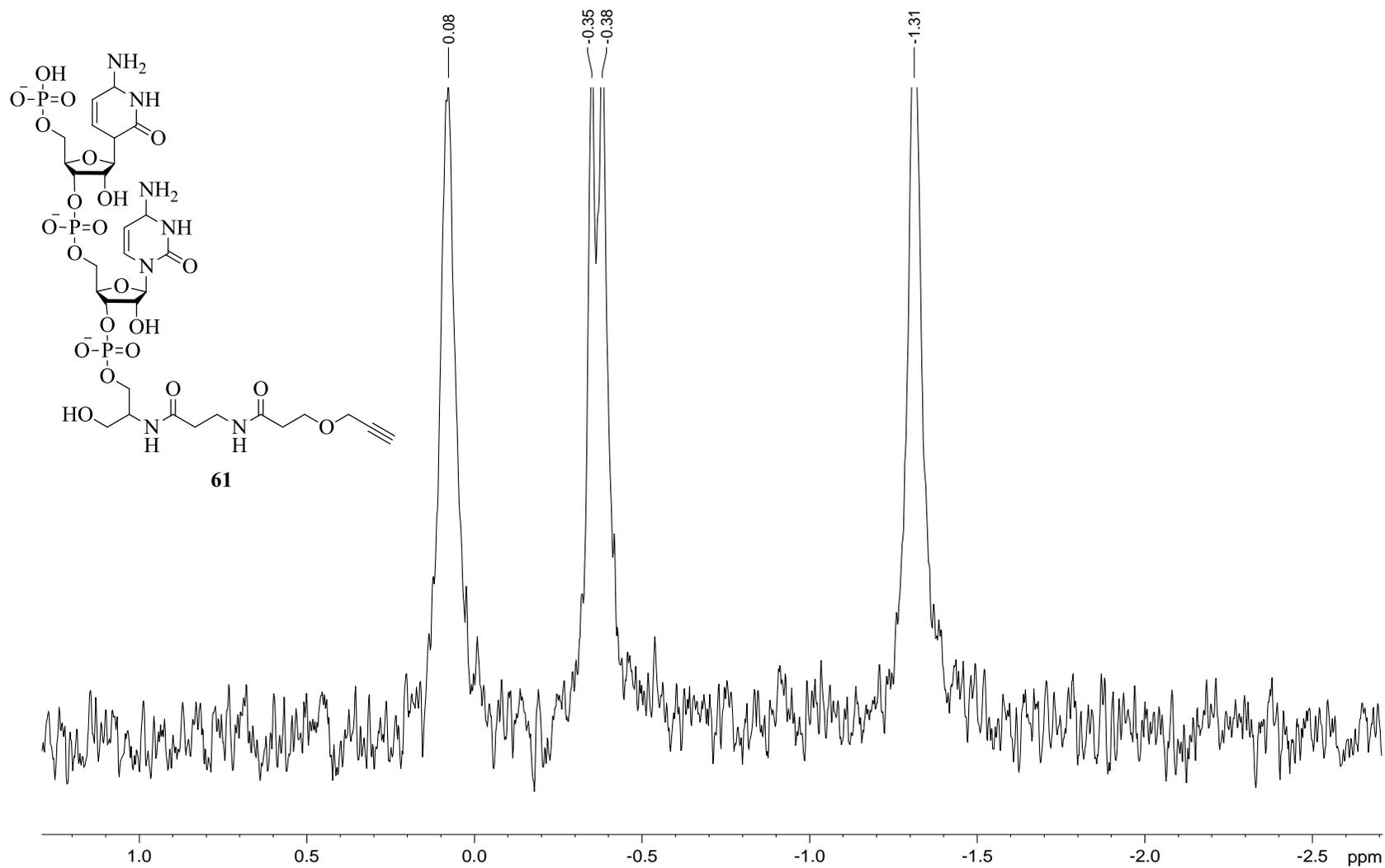


Figure 6.19 $^{31}\text{P}\{^1\text{H}\}$ NMR (500MHz, D_2O) of pCpC-alkyne (61)

hl752-83-1

KS-HL-1 42 (0.395) Cm (39:53)

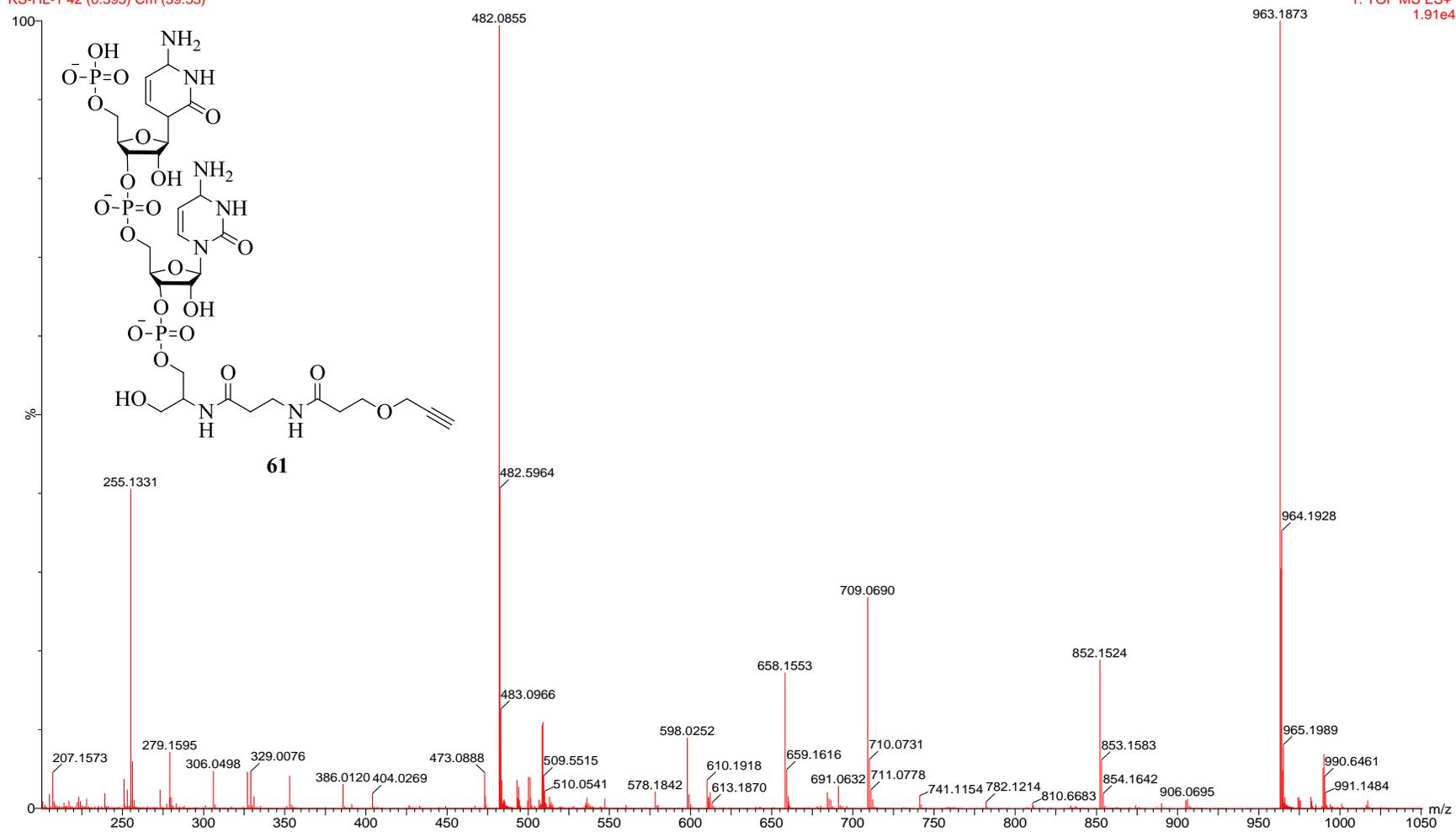


Figure 6.20 LC/MS of pCpc-alkyne (61)

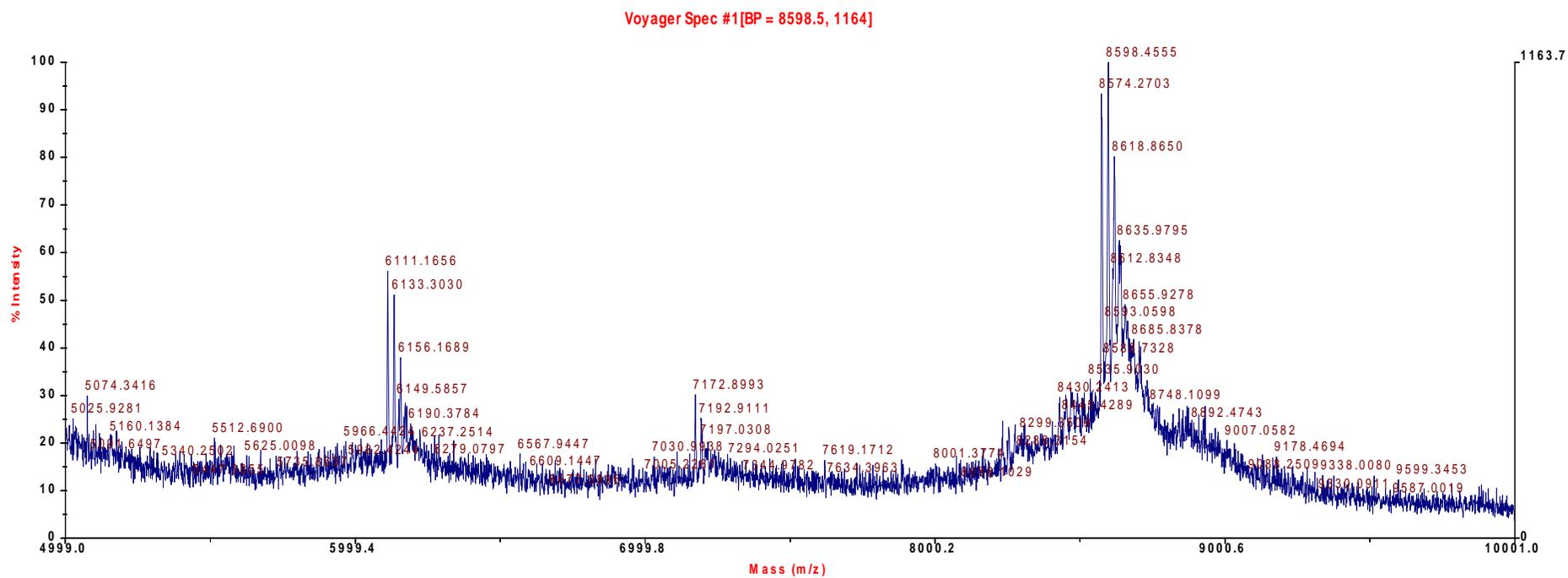


Figure 6.21 MALDI spectrum for ESE1

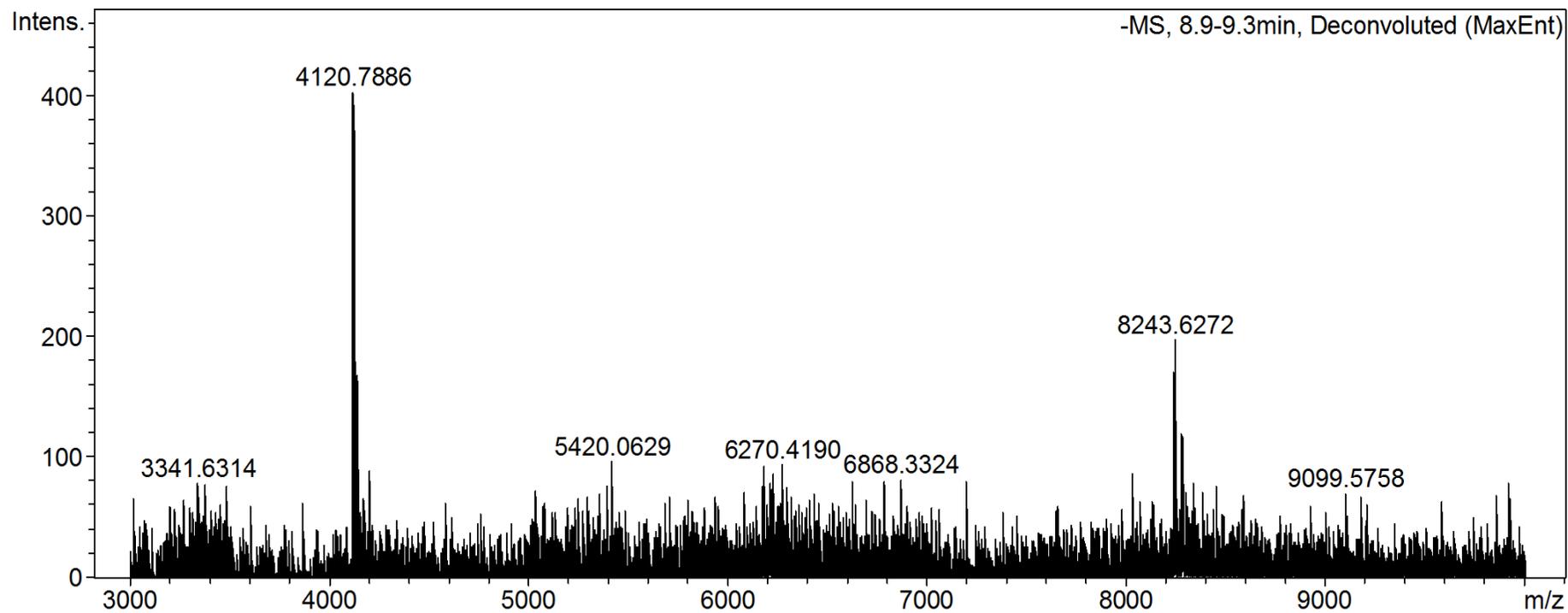


Figure 6.22 RP-HPLC-MS spectrum for ESE2.

7 REFERENCES

- (1) Keren, H.; Lev-Maor, G.; Ast, G. Alternative splicing and evolution: diversification, exon definition and function. *Nature reviews. Genetics* **2010**, *11*, 345–55.
- (2) Pasman, Z.; Garcia-Blanco, M. A. The 5' and 3' splice sites come together via a three dimensional diffusion mechanism. *Nucleic acids research* **1996**, *24*, 1638–45.
- (3) Matlin, A. J.; Clark, F.; Smith, C. W. J. Understanding alternative splicing: towards a cellular code. *Nature reviews. Molecular cell biology* **2005**, *6*, 386–98.
- (4) Blencowe, B. J. Alternative splicing: new insights from global analyses. *Cell* **2006**, *126*, 37–47.
- (5) Xing, Y.; Lee, C. Alternative splicing and RNA selection pressure--evolutionary consequences for eukaryotic genomes. *Nature reviews. Genetics* **2006**, *7*, 499–509.
- (6) International Human Genome Sequencing Consortium. Finishing the Euchromatic Sequence of the Human Genome. *Nature* **2004**, *431*, 931–45.
- (7) National Human Genome Research Institute <http://www.genome.gov/12011238>.
- (8) Pray, L. A. Eukaryotic Genome Complexity. *Nature Education* **2008**, *1*.
- (9) Shen, H.; Kan, J. L. C.; Green, M. R. Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Molecular cell* **2004**, *13*, 367–76.
- (10) Will, C. L.; Lührmann, R. Spliceosome Structure and Function. *Cold Spring Harbor perspectives in biology* **2010**.
- (11) Grabowski, P.; Padgett, R.; Sharp, P. Messenger RNA splicing in vitro: an excised intervening sequence and a potential intermediate. *Cell* **1984**, *37*, 415–427.
- (12) Konarska, M. M.; Grabowski, P. J.; Padgett, R. A.; Sharp, P. A. Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. *Nature* **1985**, *313*, 552–557.
- (13) Moore, M.; Sharp, P. Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* **1993**, *365*, 364–368.
- (14) Wahl, M. C.; Will, C. L.; Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **2009**, *136*, 701–18.

- (15) Ruskin, B.; Krainer, A. R.; Maniatis, T.; Green, M. R. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell* **1984**, *38*, 317–31.
- (16) Jurica, M. S. Searching for a wrench to throw into the splicing machine. *Nature chemical biology* **2008**, *4*, 3–6.
- (17) Longman, D.; McGarvey, T.; McCracken, S.; Johnstone, I. L.; Blencowe, B. J.; Cáceres, J. F. Multiple interactions between SRm160 and SR family proteins in enhancer-dependent splicing and development of *C. elegans*. *Current biology : CB* **2001**, *11*, 1923–33.
- (18) Jurica, M. S.; Moore, M. J. Pre-mRNA splicing: awash in a sea of proteins. *Molecular cell* **2003**, *12*, 5–14.
- (19) Sanford, J. R.; Cáceres, J. F. Pre-mRNA splicing: life at the centre of the central dogma. *Journal of cell science* **2004**, *117*, 6261–3.
- (20) Hartmuth, K.; Urlaub, H.; Vornlocher, H.-P.; Will, C. L.; Gentzel, M.; Wilm, M.; Lührmann, R. Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 16719–24.
- (21) Reed, R. Protein composition of mammalian spliceosomes assembled in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **1990**, *87*, 8031–5.
- (22) *Alternative Splicing in the Postgenomic Era*; Blencowe, B. J.; Graveley, B. R., Eds.; 623rd ed.; Springer Science+Business Media, 2007.
- (23) Chen, M.; Manley, J. L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature reviews. Molecular cell biology* **2009**, *10*, 741–54.
- (24) Abovich, N.; Rosbash, M. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **1997**, *89*, 403–12.
- (25) Liu, Z. Structural Basis for Recognition of the Intron Branch Site RNA by Splicing Factor 1. *Science* **2001**, *294*, 1098–1102.
- (26) Reed, R. Initial splice-site recognition and pairing during pre-mRNA splicing. *Current opinion in genetics & development* **1996**, *6*, 215–20.
- (27) Kotlajich, M. V; Crabb, T. L.; Hertel, K. J. Spliceosome assembly pathways for different types of alternative splicing converge during commitment to splice site pairing in the A complex. *Molecular and cellular biology* **2009**, *29*, 1072–82.
- (28) House, A. E.; Lynch, K. W. Regulation of alternative splicing: more than just the ABCs. *The Journal of biological chemistry* **2008**, *283*, 1217–21.

- (29) Bessonov, S.; Anokhina, M.; Will, C. L.; Urlaub, H.; Lührmann, R. Isolation of an active step I spliceosome and composition of its RNP core. *Nature* **2008**, *452*, 846–50.
- (30) Yu, Y.; Maroney, P. A.; Denker, J. A.; Zhang, X. H.; Olexander, D.; Lührmann, R.; Jankowsky, E.; Chasin, L. A.; Nilsen, T. W. Dynamic regulation of alternative splicing by silencers that modulate 5' splice site competition. *Cell* **2008**, *135*, 1224–1236.
- (31) Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annual review of biochemistry* **2003**, *72*, 291–336.
- (32) Alekseyenko, A. V.; Kim, N.; Lee, C. J. Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes. *RNA (New York, N.Y.)* **2007**, *13*, 661–70.
- (33) Sugnet, C. W.; Kent, W. J.; Ares, M.; Haussler, D. Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* **2004**, *77*, 66–77.
- (34) Kim, E.; Goren, A.; Ast, G. Alternative splicing: current perspectives. *BioEssays : news and reviews in molecular, cellular and developmental biology* **2008**, *30*, 38–47.
- (35) Martins de Araújo, M.; Bonnal, S.; Hastings, M. L.; Krainer, A. R.; Valcárcel, J. Differential 3' splice site recognition of SMN1 and SMN2 transcripts by U2AF and U2 snRNP. *RNA (New York, N.Y.)* **2009**, *15*, 515–23.
- (36) Skordis, L. A.; Dunckley, M. G.; Yue, B.; Eperon, I. C.; Muntoni, F. Bifunctional antisense oligonucleotides provide a trans-acting splicing enhancer that stimulates SMN2 gene expression in patient fibroblasts. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 4114–9.
- (37) MacKenzie, A. Genetic therapy for spinal muscular atrophy. *Nature biotechnology* **2010**, *28*, 235–7.
- (38) Faustino, N. A.; Cooper, T. A. Pre-mRNA splicing and human disease. *Genes & development* **2003**, *17*, 419–37.
- (39) Eperon, L. P.; Estibeiro, J. P.; Eperon, I. C. The role of nucleotide sequences in splice site selection in eukaryotic pre-messenger RNA. *Nature* **1986**, *324*, 280–282.
- (40) Roca, X.; Sachidanandam, R.; Krainer, A. R. Determinants of the inherent strength of human 5' splice sites. *RNA (New York, N.Y.)* **2005**, *11*, 683–98.

- (41) Goren, A.; Ram, O.; Amit, M.; Keren, H.; Lev-Maor, G.; Vig, I.; Pupko, T.; Ast, G. Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. *Molecular cell* **2006**, *22*, 769–81.
- (42) Sakabe, N. J.; De Souza, S. J. Sequence features responsible for intron retention in human. *BMC genomics* **2007**, *8*, 59.
- (43) Haseloff, J.; Siemering, K. R.; Prasher, D. C.; Hodge, S. Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 2122–7.
- (44) Valacca, C.; Bonomi, S.; Buratti, E.; Pedrotti, S.; Baralle, F. E.; Sette, C.; Ghigna, C.; Biamonti, G. Sam68 regulates EMT through alternative splicing–activated nonsense-mediated mRNA decay of the SF2/ASF proto-oncogene. *The Journal of cell Biology* **2010**, *191*, 87–99.
- (45) Pan, Q.; Shai, O.; Lee, L. J.; Frey, B. J.; Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics* **2008**, *40*, 1413–5.
- (46) Hodson, M. J.; Hudson, A. J.; Cherny, D.; Eperon, I. C. The transition in spliceosome assembly from complex E to complex A purges surplus U1 snRNPs from alternative splice sites. *Nucleic acids research* **2012**, 1–13.
- (47) Wang, Z.; Burge, C. B. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA (New York, N.Y.)* **2008**, *14*, 802–13.
- (48) Graveley, B. Alternative splicing: regulation without regulators. *Nature structural & molecular biology* **2009**, *16*, 13–15.
- (49) Tange, T. O.; Damgaard, C. K.; Guth, S.; Valcárcel, J.; Kjems, J. The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *The EMBO journal* **2001**, *20*, 5748–58.
- (50) Pedrotti, S.; Bielli, P.; Paronetto, M. P.; Ciccocanti, F.; Fimia, G. M.; Stamm, S.; Manley, J. L.; Sette, C. The splicing regulator Sam68 binds to a novel exonic splicing silencer and functions in SMN2 alternative splicing in spinal muscular atrophy. *The EMBO journal* **2010**, *29*, 1235–47.
- (51) Blanchette, M.; Chabot, B. Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *The EMBO journal* **1999**, *18*, 1939–52.
- (52) Abdul-Manan, N.; Williams, K. R. hnRNP A1 binds promiscuously to oligoribonucleotides: utilization of random and homo-oligonucleotides to discriminate sequence from base-specific binding. *Nucleic acids research* **1996**, *24*, 4063–70.

- (53) Nasim, M.; Eperon, I. A double-reporter splicing assay for determining splicing efficiency in mammalian cells. *Nature protocols* **2006**, *1*, 1022–1028.
- (54) Long, J. C.; Cáceres, J. F. The SR protein family of splicing factors: master regulators of gene expression. *The Biochemical journal* **2009**, *417*, 15–27.
- (55) Graveley, B. R. Sorting out the complexity of SR protein functions. *RNA (New York, N.Y.)* **2000**, *6*, 1197–211.
- (56) Shen, H.; Green, M. R. A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. *Molecular cell* **2004**, *16*, 363–73.
- (57) Cáceres, J. F.; Krainer, A. R. Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. *The EMBO journal* **1993**, *12*, 4715–26.
- (58) Dredge, B.; Darnell, R. Nova regulates GABAA receptor $\gamma 2$ alternative splicing via a distal downstream UCAU-rich intronic splicing enhancer. *Molecular and cellular biology* **2003**, *23*, 4687–4700.
- (59) Min, H.; Turck, C. W.; Nikolic, J. M.; Black, D. L. A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. *Genes & Development* **1997**, *11*, 1023–1036.
- (60) McCullough, A. J.; Schuler, M. a Intronic and exonic sequences modulate 5' splice site selection in plant nuclei. *Nucleic acids research* **1997**, *25*, 1071–7.
- (61) McCullough, A. J.; Berget, S. M. An intronic splicing enhancer binds U1 snRNPs to enhance splicing and select 5' splice sites. *Molecular and cellular biology* **2000**, *20*, 9225–9235.
- (62) Pastor, T.; Talotti, G.; Lewandowska, M. A.; Pagani, F. An Alu-derived intronic splicing enhancer facilitates intronic processing and modulates aberrant splicing in ATM. *Nucleic acids research* **2009**, *37*, 7258–67.
- (63) Mardon, H. J.; Sebastio, G.; Baralle, F. E. A role for exon sequences in alternative splicing of the human fibronectin gene. *Nucleic acids research* **1987**, *15*, 7725–7734.
- (64) Lavigne, A.; La Branche, H.; Kornblihtt, A. R.; Chabot, B. A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. *Genes & Development* **1993**, *7*, 2405–2417.
- (65) Zhong, X.-Y.; Ding, J.-H.; Adams, J. A.; Ghosh, G.; Fu, X.-D. Regulation of SR protein phosphorylation and alternative splicing by modulating kinetic interactions of SRPK1 with molecular chaperones. *Genes & development* **2009**, *23*, 482–95.

- (66) Xiao, S. H.; Manley, J. L. Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes & Development* **1997**, *11*, 334–344.
- (67) Tian, M.; Maniatis, T. A splicing enhancer exhibits both constitutive and regulated activities. *Genes & Development* **1994**, *8*, 1703–1712.
- (68) Blencowe, B. J. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends in biochemical sciences* **2000**, *25*, 106–10.
- (69) Yue, B. G.; Akusjärvi, G. A downstream splicing enhancer is essential for in vitro pre-mRNA splicing. *FEBS letters* **1999**, *451*, 10–4.
- (70) Li, Y.; Blencowe, B. J. Distinct factor requirements for exonic splicing enhancer function and binding of U2AF to the polypyrimidine tract. *The Journal of biological chemistry* **1999**, *274*, 35074–9.
- (71) Bruzik, J. P.; Maniatis, T. Enhancer-dependent interaction between 5' and 3' splice sites in trans. *Proceedings of the National Academy of Sciences of the United States of America* **1995**, *92*, 7056–7059.
- (72) Hicks, M. J.; Mueller, W. F.; Shepard, P. J.; Hertel, K. J. Competing upstream 5' splice sites enhance the rate of proximal splicing. *Molecular and cellular biology* **2010**, *30*, 1878–86.
- (73) Hicks, M. J.; Lam, B. J.; Hertel, K. J. Analyzing mechanisms of alternative pre-mRNA splicing using in vitro splicing assays. *Methods (San Diego, Calif.)* **2005**, *37*, 306–13.
- (74) Fairbrother, W. G.; Yeh, R.-F.; Sharp, P. a; Burge, C. B. Predictive identification of exonic splicing enhancers in human genes. *Science (New York, N.Y.)* **2002**, *297*, 1007–13.
- (75) Zhang, X. H.-F.; Arias, M. A.; Ke, S.; Chasin, L. A. Splicing of designer exons reveals unexpected complexity in pre-mRNA splicing. *RNA (New York, N.Y.)* **2009**, *15*, 367–76.
- (76) Graveley, B. R.; Hertel, K. J.; Maniatis, T. A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *The EMBO journal* **1998**, *17*, 6747–56.
- (77) Chiara, M. D.; Gozani, O.; Bennett, M.; Champion-Arnaud, P.; Palandjian, L.; Reed, R. Identification of proteins that interact with exon sequences, splice sites, and the branchpoint sequence during each stage of spliceosome assembly. *Molecular and cellular biology* **1996**, *16*, 3317–26.
- (78) Sciabica, K. S.; Hertel, K. J. The splicing regulators Tra and Tra2 are unusually potent activators of pre-mRNA splicing. *Nucleic acids research* **2006**, *34*, 6612–20.

- (79) Lewis, H.; Perrett, A. An RNA Splicing Enhancer that Does Not Act by Looping. *Angewandte Chemie* **2012**, Manuscript Accepted.
- (80) Jäschke, A.; Fürste, J. P.; Nordhoff, E.; Hillenkamp, F.; Cech, D.; Erdmann, V. a. Synthesis and properties of oligodeoxyribonucleotide—polyethylene glycol conjugates. *Nucleic Acids Research* **1994**, *22*, 4810–4817.
- (81) Zalipsky, S. Chemistry of polyethylene glycol conjugates with biologically active molecules. *Advanced Drug Delivery Reviews* **1995**, *16*, 157–182.
- (82) Zhang, L.; Peritz, A.; Meggers, E. A simple glycol nucleic acid. *Journal of the American Chemical Society* **2005**, *127*, 4174–5.
- (83) Schlegel, M. K.; Essen, L.-O.; Meggers, E. Duplex structure of a minimal nucleic acid. *Journal of the American Chemical Society* **2008**, *130*, 8158–9.
- (84) Schlegel, M. K.; Peritz, A. E.; Kittigowittana, K.; Zhang, L.; Meggers, E. Duplex formation of the simplified nucleic acid GNA. *Chembiochem : a European journal of chemical biology* **2007**, *8*, 927–32.
- (85) Fiammengo, R.; Musílek, K.; Jäschke, A. Efficient preparation of organic substrate-RNA conjugates via in vitro transcription. *Journal of the American Chemical Society* **2005**, *127*, 9271–6.
- (86) Schlatterer, J. C.; Jäschke, A. Universal initiator nucleotides for the enzymatic synthesis of 5'-amino- and 5'-thiol-modified RNA. *Biochemical and biophysical research communications* **2006**, *344*, 887–92.
- (87) Schoch, J.; Ameta, S.; Jäschke, A. Inverse electron-demand Diels-Alder reactions for the selective and efficient labeling of RNA. *Chemical communications (Cambridge, England)* **2011**, 12536–12537.
- (88) Seelig, B.; Jäschke, A. Ternary conjugates of guanosine monophosphate as initiator nucleotides for the enzymatic synthesis of 5'-modified RNAs. *Bioconjugate chemistry* **1999**, *10*, 371–8.
- (89) Sengle, G.; Jenne, A.; Arora, P. S.; Seelig, B.; Nowick, J. S.; Jäschke, A.; Famulok, M. Synthesis, incorporation efficiency, and stability of disulfide bridged functional groups at RNA 5'-ends. *Bioorganic & medicinal chemistry* **2000**, *8*, 1317–29.
- (90) Williamson, D.; Cann, M. J.; Hodgson, D. R. W. Synthesis of 5'-amino-5'-deoxyguanosine-5'-N-phosphoramidate and its enzymatic incorporation at the 5'-termini of RNA molecules. *Chemical communications (Cambridge, England)* **2007**, 5096–8.
- (91) Williamson, D.; Hodgson, D. R. W. Preparation and purification of 5'-amino-5'-deoxyguanosine-5'-N-phosphoramidate and its initiation properties with T7 RNA polymerase. *Organic & biomolecular chemistry* **2008**, *6*, 1056–62.

- (92) Wolf, J.; Dombos, V.; Appel, B.; Müller, S. Synthesis of guanosine 5'-conjugates and their use as initiator molecules for transcription priming. *Organic & biomolecular chemistry* **2008**, *6*, 899–907.
- (93) Zhang, L.; Sun, L.; Cui, Z.; Gottlieb, R. L.; Zhang, B. 5'-Sulfhydryl-Modified RNA: Initiator Synthesis, in Vitro Transcription, and Enzymatic Incorporation. *Bioconjugate Chemistry* **2001**, *12*, 939–948.
- (94) Hausch, F.; Jäschke, A. Libraries of multifunctional RNA conjugates for the selection of new RNA catalysts. *Bioconjugate chemistry* **1997**, *8*, 885–90.
- (95) Hausch, F.; Jäschke, A. Multifunctional dinucleotide analogs for the generation of complex RNA conjugates. *Tetrahedron* **2001**, *57*, 1261–1268.
- (96) Wahl, F.; Jäschke, A. PEG-tethered guanosine acetal conjugates for the enzymatic synthesis of modified RNA. *Biochemical and biophysical research communications* **2012**, *417*, 1224–6.
- (97) Pfander, S.; Fiammengo, R.; Kirin, S. I.; Metzler-Nolte, N.; Jäschke, A. Reversible site-specific tagging of enzymatically synthesized RNAs using aldehyde-hydrazine chemistry and protease-cleavable linkers. *Nucleic acids research* **2007**, *35*, e25.
- (98) England, T. E.; Uhlenbeck, O. C. 3'-Terminal labelling of RNA with T4 RNA ligase. *Nature* **1978**, *275*, 560–561.
- (99) Barrio, J. R.; Barrio, M. C.; Leonard, N. J.; England, T. E.; Uhlenbeck, O. C. Synthesis of modified nucleoside 3',5'-bisphosphates and their incorporation into oligoribonucleotides with T4 RNA ligase. *Biochemistry* **1978**, *17*, 2077–81.
- (100) Reese, C. B. Oligo- and poly-nucleotides: 50 years of chemical synthesis. *Organic & biomolecular chemistry* **2005**, *3*, 3851–68.
- (101) Letsinger, R. L.; Lunsford, W. B. Synthesis of thymidine oligonucleotides by phosphite triester intermediates. *Journal of the American Chemical Society* **1976**, *98*, 3655–61.
- (102) Beaucage, S. L.; Caruthers, M. H. Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Letters* **1981**, *22*, 1859–1862.
- (103) El-Sagheer, A. H.; Brown, T. New strategy for the synthesis of chemically modified RNA constructs exemplified by hairpin and hammerhead ribozymes. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 15329–34.
- (104) Berner, S.; Mühlegger, K.; Seliger, H. Studies on the role of tetrazole in the activation of phosphoramidites. *Nucleic acids research* **1989**, *17*, 853–864.

- (105) Vargeese, C.; Carter, J.; Yegge, J.; Krivjansky, S.; Settle, a; Kropp, E.; Peterson, K.; Pieken, W. Efficient activation of nucleoside phosphoramidites with 4,5-dicyanoimidazole during oligonucleotide synthesis. *Nucleic acids research* **1998**, *26*, 1046–50.
- (106) Saxon, E.; Armstrong, J. I.; Bertozzi, C. R. A “ Traceless ” Staudinger Ligation for the Chemoselective Synthesis of Amide Bonds. *Organic letters* **2000**, *2*, 2141–2143.
- (107) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angewandte Chemie (International ed. in English)* **2001**, *40*, 2004–2021.
- (108) Singh, I.; Heaney, F. Metal free, “click and click-click” conjugation of ribonucleosides and 2'-OMe oligoribonucleotides on the solid phase. *Organic & biomolecular chemistry* **2010**, *8*, 451–6.
- (109) Algay, V.; Singh, I.; Heaney, F. Nucleoside and nucleotide analogues by catalyst free Huisgen nitrile oxide-alkyne 1,3-dipolar cycloaddition. *Organic & biomolecular chemistry* **2010**, *8*, 391–7.
- (110) Qin, P. Z.; Pyle, A. M. Site-specific labeling of RNA with fluorophores and other structural probes. *Methods: A Companion to Methods in Enzymology* **1999**, *18*, 60–70.
- (111) Solomatin, S.; Herschlag, D. Methods of Site-Specific Labeling of RNA with Fluorescent Dyes. In *Methods in Enzymology*; Elsevier Inc., 2009; Vol. 469, pp. 47–68.
- (112) Kalia, J.; Raines, R. T. Advances in Bioconjugation. *Current Organic Chemistry* **2010**, *14*, 138–147.
- (113) Paredes, E.; Evans, M.; Das, S. R. RNA labeling, conjugation and ligation. *Methods (San Diego, Calif.)* **2011**, *54*, 251–9.
- (114) Paredes, E.; Das, S. R. Click chemistry for rapid labeling and ligation of RNA. *Chembiochem : a European journal of chemical biology* **2011**, *12*, 125–31.
- (115) Saxon, E.; Bertozzi, C. R. Cell Surface Engineering by a Modified Staudinger Reaction. *Science* **2000**, *287*, 2007–2010.
- (116) Sletten, E. M.; Bertozzi, C. R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angewandte Chemie (International ed. in English)* **2009**, *48*, 6974–98.
- (117) Van Berkel, S. S.; Van Eldijk, M. B.; Van Hest, J. C. M. Staudinger ligation as a method for bioconjugation. *Angewandte Chemie (International ed. in English)* **2011**, *50*, 8806–27.

- (118) Nilsson, B. L.; Kiessling, L. L.; Raines, R. T. Staudinger ligation: a peptide from a thioester and azide. *Organic letters* **2000**, *2*, 1939–41.
- (119) Hein, C.; Liu, X. Click chemistry, a powerful tool for pharmaceutical sciences. *Pharmaceutical research* **2008**, *25*, 2216–2230.
- (120) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective “ligation” of azides and terminal alkynes. *Angewandte Chemie (International ed. in English)* **2002**, *41*, 2596–9.
- (121) Best, M. D. Click chemistry and bioorthogonal reactions: unprecedented selectivity in the labeling of biological molecules. *Biochemistry* **2009**, *48*, 6571–84.
- (122) Van Dijk, M.; Rijkers, D. T. S.; Liskamp, R. M. J.; Van Nostrum, C. F.; Hennink, W. E. Synthesis and applications of biomedical and pharmaceutical polymers via click chemistry methodologies. *Bioconjugate chemistry* **2009**, *20*, 2001–16.
- (123) Burley, G. A.; Gierlich, J.; Mofid, M. R.; Nir, H.; Tal, S.; Eichen, Y.; Carell, T. Directed DNA metallization. *Journal of the American Chemical Society* **2006**, *128*, 1398–9.
- (124) Gramlich, P. M. E.; Wirges, C. T.; Gierlich, J.; Carell, T. Synthesis of modified DNA by PCR with alkyne-bearing purines followed by a click reaction. *Organic letters* **2008**, *10*, 249–51.
- (125) Gierlich, J.; Burley, G. A.; Gramlich, P. M. E.; Hammond, D. M.; Carell, T. Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA. *Organic letters* **2006**, *8*, 3639–42.
- (126) Gierlich, J.; Gutmiedl, K.; Gramlich, P. M. E.; Schmidt, A.; Burley, G. A.; Carell, T.; Triphosphates, A.; Chemistry, C. Synthesis of highly modified DNA by a combination of PCR with alkyne-bearing triphosphates and click chemistry. *Chemistry (Weinheim an der Bergstrasse, Germany)* **2007**, *13*, 9486–94.
- (127) Hong, V.; Presolski, S. I. S.; Ma, C.; Finn, M. G. Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angewandte Chemie (International ed. in English)* **2009**, *48*, 9879–83.
- (128) 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. *Journal of the American Chemical Society* **2010**, *132*, 14570–6.
- (129) Jao, C. Y.; Salic, A. Exploring RNA transcription and turnover in vivo by using click chemistry. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 15779–84.

- (130) Hong, V.; Steinmetz, N. F.; Manchester, M.; Finn, M. G. Labeling live cells by copper-catalyzed alkyne--azide click chemistry. *Bioconjugate chemistry* **2010**, *21*, 1912–6.
- (131) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *Journal of the American Chemical Society* **2004**, *126*, 15046–7.
- (132) Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V; Miller, I. a; Lo, A.; Codelli, J. a; Bertozzi, C. R. Copper-free click chemistry for dynamic in vivo imaging. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 16793–7.
- (133) Ning, X.; Guo, J.; Wolfert, M. A.; Boons, G.-J. Visualizing metabolically labeled glycoconjugates of living cells by copper-free and fast huigen cycloadditions. *Angewandte Chemie (International ed. in English)* **2008**, *47*, 2253–5.
- (134) Jewett, J. C.; Sletten, E. M.; Bertozzi, C. R. Rapid Cu-free click chemistry with readily synthesized biarylazacyclooctynones. *Journal of the American Chemical Society* **2010**, *132*, 3688–90.
- (135) Van Delft, P.; Meeuwenoord, N. J.; Hoogendoorn, S.; Dinkelaar, J.; Overkleeft, H. S.; Van der Marel, G. A.; Filippov, D. V Synthesis of oligoribonucleic acid conjugates using a cyclooctyne phosphoramidite. *Organic letters* **2010**, *12*, 5486–9.
- (136) Coulter, L.; Landree, M.; Cooper, T. Identification of a new class of exonic splicing enhancers by in vivo selection. *Molecular and cellular* **1997**, *17*, 2143–2150.
- (137) Schaal, T. D.; Maniatis, T. Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Molecular and cellular biology* **1999**, *19*, 1705–1719.
- (138) Lareau, L. F.; Inada, M.; Green, R. E.; Wengrod, J. C.; Brenner, S. E. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **2007**, *446*, 926–9.
- (139) Maquat, L. E. Nonsense-mediated mRNA decay in mammals. *Journal of cell science* **2005**, *118*, 1773–6.
- (140) Cooke, C.; Hans, H.; Alwine, J. C. Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal. *Molecular and cellular biology* **1999**, *19*, 4971–4979.
- (141) O'Mullane, L.; Eperon, I. C. The pre-mRNA 5' cap determines whether U6 small nuclear RNA succeeds U1 small nuclear ribonucleoprotein particle at 5' splice sites. *Molecular and cellular biology* **1998**, *18*, 7510–7520.

- (142) Owen, N.; Zhou, H.; Malygin, A. a; Sangha, J.; Smith, L. D.; Muntoni, F.; Eperon, I. C. Design principles for bifunctional targeted oligonucleotide enhancers of splicing. *Nucleic acids research* **2011**, *39*, 7194–208.
- (143) Kramer, M. F.; Coen, D. M. Enzymatic amplification of DNA by PCR: standard procedures and optimization. *Current protocols in cytometry / editorial board, J. Paul Robinson, managing editor ... [et al.]* **2006**, Appendix 3, Appendix 3K.
- (144) Roux, K. H. Optimization and troubleshooting in PCR. *Cold Spring Harbor protocols* **2009**, 2009, pdb.ip66.
- (145) Go Taq polymerase protocol
[http://www.promega.com/~media/files/resources/protocols/product information sheets/g/gotaq dna polymerase m300.pdf?la=en](http://www.promega.com/~media/files/resources/protocols/product%20information%20sheets/g/gotaq%20dna%20polymerase%20m300.pdf?la=en) (accessed Jun 18, 2012).
- (146) Phusion Hot start protocol
<http://www.neb.com/nebecomm/products/protocol537.asp> (accessed Jun 18, 2012).
- (147) Hodson, M. J. Use of single molecule methods to reveal the mechanisms of splice site selection, University of Leicester, 2011.
- (148) Eperon, I. C.; Krainer, A. R. Splicing of Messenger RNA Precursors. In *RNA Processing: a Practical Approach*; Haims, D.; Higgins, S., Eds.; Oxford University Press., 1994.
- (149) Humphrey, M. B.; Bryan, J.; Cooper, T. A.; Berget, S. M. A 32-nucleotide exon-splicing enhancer regulates usage of competing 5' splice sites in a differential internal exon. *Molecular and cellular biology* **1995**, *15*, 3979–88.
- (150) Seelig, B.; Jaschke, A.; Berlin, F. U. Site-Specific Modification of Enzymatically Synthesized RNA: Transcription Initiation and Diels-Alder Reaction. *Tetrahedron Letters* **1997**, *38*, 7729–7732.
- (151) Sivakumar, K.; Xie, F.; Cash, B. M.; Long, S.; Barnhill, H. N.; Wang, Q.; Brandon, M. A fluorogenic 1, 3-dipolar cycloaddition reaction of 3-azidocoumarins and acetylenes. *Organic* **2004**, *6*, 4603–4606.
- (152) Chan, T. R.; Hilgraf, R.; Sharpless, K. B.; Fokin, V. V Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Organic letters* **2004**, *6*, 2853–5.
- (153) Devaraj, N.; Miller, G.; Ebina, W.; Kakaradov, B.; Collman, J. P.; Kool, E. T.; Chidsey, C. E. D. Chemoselective covalent coupling of oligonucleotide probes to self-assembled monolayers. *Journal of the American Chemical Society* **2005**, *127*, 8600–8601.
- (154) Bolduc, L.; Labrecque, B.; Cordeau, M.; Blanchette, M.; Chabot, B. Dimethyl sulfoxide affects the selection of splice sites. *The Journal of biological chemistry* **2001**, *276*, 17597–602.

- (155) Konarska, M. M.; Padgett, R. A.; Sharp, P. A. Recognition of cap structure in splicing in vitro of mRNA precursors. *Cell* **1984**, *38*, 731–6.
- (156) Barabino, S. M.; Sproat, B. S.; Ryder, U.; Blencowe, B. J.; Lamond, a I. Mapping U2 snRNP--pre-mRNA interactions using biotinylated oligonucleotides made of 2'-OMe RNA. *The EMBO journal* **1989**, *8*, 4171–8.
- (157) http://www.linktech.co.uk/support/knowledge_base/446_use-protocol-information-for-products-in-2010-product-guide <http://www.linktech.co.uk/>.
- (158) Pielers, U.; Zürcher, W.; Schär, M.; Moser, H. E. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: a powerful tool for the mass and sequence analysis of natural and modified oligonucleotides. *Nucleic acids research* **1993**, *21*, 3191–6.
- (159) Kirpekar, F.; Nordhoff, E.; Kristiansen, K.; Roepstorff, P.; Lezius, a; Hahner, S.; Karas, M.; Hillenkamp, F. Matrix assisted laser desorption/ionization mass spectrometry of enzymatically synthesized RNA up to 150 kDa. *Nucleic acids research* **1994**, *22*, 3866–70.
- (160) No Title <http://www.glenresearch.com/ProductFiles/10-1908.html>.
- (161) Kumar, R.; El-Sagheer, A.; Tumpane, J.; Lincoln, P.; Wilhelmsson, L. M.; Brown, T. Template-directed oligonucleotide strand ligation, covalent intramolecular DNA circularization and catenation using click chemistry. *Journal of the American Chemical Society* **2007**, *129*, 6859–64.
- (162) Humenik, M.; Huang, Y.; Wang, Y.; Sprinzl, M. C-terminal incorporation of bio-orthogonal azide groups into a protein and preparation of protein-oligodeoxynucleotide conjugates by Cu⁺-catalyzed cycloaddition. *Chembiochem : a European journal of chemical biology* **2007**, *8*, 1103–6.
- (163) Meldal, M. Polymer “Clicking” by CuAAC Reactions. *Macromolecular Rapid Communications* **2008**, *29*, 1016–1051.
- (164) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell*; Fourth Edi.; Garland Science, 2002.
- (165) Behlke, M. a Chemical modification of siRNAs for in vivo use. *Oligonucleotides* **2008**, *18*, 305–19.
- (166) Saccà, B.; Lacroix, L.; Mergny, J.-L. The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic acids research* **2005**, *33*, 1182–92.
- (167) Kumari, S.; Bugaut, A.; Huppert, J.; Balasubramanian, S. An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nature chemical biology* **2007**, *3*, 218–221.

- (168) Sun, D.; Hurley, L. H. Biochemical Techniques for the Characterisation of G-Quadruplex Structures: EMSA, DMS Footprinting and DNA Polymerase Stop Assay. *Methods in Molecular Biology* **2010**, *608*, 65–79.
- (169) Will, C. L.; Lührmann, R. Protein functions in pre-mRNA splicing. *Current opinion in cell biology* **1997**, *9*, 320–8.
- (170) Das, R.; Reed, R. Resolution of the mammalian E complex and the ATP-dependent spliceosomal complexes on native agarose mini-gels. *RNA* **1999**, 1504–1508.
- (171) Kuhn, A. N.; Van Santen, M. A.; Schwienhorst, A.; Urlaub, H.; Lührmann, R. Stalling of spliceosome assembly at distinct stages by small-molecule inhibitors of protein acetylation and deacetylation. *RNA (New York, N.Y.)* **2009**, *15*, 153–75.