University of Nottingham
UK | CHINA | MALAYSIA

Synthesis, detection, and quantification of modified nucleotides in RNA

Francis Baron

School of Chemistry

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

September 2019

Acknowledgements

Firstly, I would like to thank my supervisors Professors Rupert Fray and Professor Christopher Hayes for their great supervision over the last four years.

Thanks to Matt, Mustafa, Charlotte, Katie, Mandeep, Michael, Edward, Ben, Paulina, and Rebecca from the Hayes group, and to Zsuzsa, Eleanor, Sam, Mi, Alex, Ed, Doha, Nathan, Cameron, Martina, and Michele, from the Fray group, and all of the masters and project students I have worked with over the last few years. Particular thanks to Zsusza, Eleanor and Cameron who did some of the experiments in this thesis in our collaborative research projects.

I would like to thank the analytical staff and service staff from the departments of Chemistry and Plant Science, and thanks to Professor Neil Oldham for your help and expertise with the analysis of my oligonucleotides.

Thanks to Professor Zheng Yuan for supervising my PIPS research project in Shanghai, Miriam Columbi for helping me get to China, and to all of my colleagues at Shanghai Jiao Tong University, especially Li Chun, for all your help during my placement.

Last but not least, thanks to all of my friends and family, to everyone in my DTP cohort, and especially to Sarah for all you help over the last four years.

Contents

Abbreviations

1D TLC	one-dimensional thin layer chromatography
2D TLC	two-dimensional thin layer chromatography
^{32}P	Phosphorus-32
A	Adenosine
A^P	Adenosine 3’ phosphate
^PA	Adenosine 5’ phosphate
app	Apparent
ATP	Adenosine triphosphate
br	Broad
bz	Benzoyl
d	Doublet
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
DEAE	Diethylaminoethyl
dGTP	Deoxyguanosine triphosphate
dT	Deoxythymidine
dTTP	Deoxythymidine triphosphate
DMTr	Dimethoxytrityl
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate

C	Cytosine
C^P	Cytosine 3’ phosphate
CPS	counts per second
^PC	Cytosine 5’ phosphate
CE	Cyanoethyl
CTP	Cytosine triphosphate
G	Guanosine
G^P	Guanosine 3’ phosphate
^PG	Guanosine 5’ phosphate
GTP	Guanosine triphosphate
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
hr	Hour
kb	Kilo-base pair
LCMS	Liquid chromatography–mass spectrometry
lncRNA	long non-coding ribonucleic acid
m	Multiplet
Me	Methyl
MiCLIP	m^6A individual-nucleotide-resolution cross-linking and immunoprecipitation
MeCN	Acetonitrile
MeRIP-Seq	Methylated RNA immunoprecipitation and sequencing

M	Molar
min	Minute
miRNA	Micro ribonucleic acid
mol	Moles
mRNA	Messenger ribonucleic acid
m^6A	N^6-methyladenosine
ncRNA	non-coding ribonucleic acid
NMR	Nuclear magnetic resonance
nt	Nucleotide
NTP	Nucleotide triphosphate
PCR	Polymerase chain reaction
PhSeT	5-phenylselenylmethyl-thymidine
PhSe-dTTP	5-phenylselenylmethyl-thymidine triphosphate
PNK	Polynucleotide kinase
q	Quartet
RBP	RNA binding protein
RF	Retention factor
rRNA	Ribosomal ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
RNA	Ribonucleic acid
RNase	Ribonuclease

SCARLET	Site-specific cleavage and radioactive-labelling followed by ligation-assisted extraction and thin-layer chromatography
siRNA	Short interfering ribonucleic acid
t	*tert*
t	triplet
T	Thymine
TBDMS	*tert*-Butyldimethylsilyl
TBDPS	*tert*-Butyldiphenylsilyl
TLC	Thin layer chromatography
tRNA	transfer ribonucleic acid
UV	Ultraviolet
U	Uracil
U^{P}	Uracil 3’ phosphate
^{P}U	Uracil 5’ phosphate
UTP	Uridine triphosphate
UTR	Untranslated region
YTHDF	YTH domain protein family
ZBP	Zipcode binding protein 1

Abstract

Background: *N*^{6}-methyladenosine (m^{6}A) is the most abundant internal modification of mRNA. m^{6}A regulates almost every stage in the mRNA life cycle with essential roles in splicing,[1] polyadenylation,[2] nuclear export,[3,4] stability,[5] translation,[6–8] and degradation.[5,9] m^{6}A is essential for normal development of eukaryotic organisms, and abnormal levels of m^{6}A have been associated with the development of disease including various types of cancer.[10] Many research groups are working on artificially manipulating m^{6}A levels for the therapeutic treatment of disease.[11]

In order to fully understand the m^{6}A modification and develop associated therapeutics, methods to accurately detect and quantify m^{6}A are required. MeRIP-seq[12,13] and miCLIP[2,14] are the most commonly used methods of m^{6}A sequencing, however, they are severely limited by (a) their inability to reliably quantify m^{6}A levels, and (b) their use of an anti m^{6}A antibody which is known to have problems with off-target binding.[15] There is a significant demand for new methods of m^{6}A sequencing and quantification.[11]

m^{6}A modified oligonucleotides are useful tools for the study of m^{6}A. They can be used for various applications including the identification and study of m^{6}A writer, reader, and eraser proteins, and can be used as substrates to test methods of m^{6}A sequencing. Synthesis of m^{6}A modified oligonucleotides requires an m^{6}A phosphoramidite which is currently produced via a complicated seven-step chemical synthesis.[16] Alternate methods for the synthesis of the m^{6}A phosphoramidite will reduce the cost and increase availability of m^{6}A modified oligonucleotides.

Research aims: The aims of this thesis are to (a) synthesise m^6A modified oligonucleotide probes to study the role of the modification in mRNA, and (b) develop new methods for the detection and quantification of m^6A.

Results: We have developed a method for the single step synthesis of m^6A phosphoramidite and synthesised a number of m^6A modified oligonucleotides which will prove to be valuable tools for research into m^6A.

We have developed various new methods for the detection and quantification of m^6A. We have developed a method involving site specific radiolabelling and thin layer chromatography (TLC) which we have used to probe m^6A levels in a number of RNA transcripts. We have identified a phenyl selenide modified thymidine (PhSeT) nucleoside that selectively crosslinks with adenosine[17] but not m^6A in both mononucleotide and interstrand contexts. This molecule was unsuitable for m^6A sequencing due to the unreliable nature of the crosslinking reaction. However, we have demonstrated that a PhSeT nucleoside triphosphate (PhSe-dTTP) is able to selectively inhibit reverse transcription reactions at sites of m^6A in the RNA template.

Conclusions: Due to the significant challenge of detecting and quantifying m^6A, the methods we have developed have had varying levels of success. The PhSe-dTTP reverse transcription-based method is by far the most promising. Efforts to develop this technology as a transcriptome wide method of m^6A sequencing are currently in progress.

Chapter 1: An overview of m^6A research

Overview: m^6A is the most prevalent internal modification of messenger RNA that plays an essential role in the regulation of gene expression. In this chapter we introduce the concept of the epitranscriptome and give a brief history of research into m^6A. We discuss existing methods of m^6A detection and quantification, introduce the m^6A writer, reader, and eraser proteins, and discuss the biological significance of the m^6A modification. Finally, we summarise the major challenges facing future research into m^6A and the epitranscriptome.

1.1 Introduction

1.1.1 The central dogma of molecular biology, epigenetics, and the epitranscriptome

In 1957 the concept of the central dogma of molecular biology was first introduced, describing the relationship between DNA, RNA and protein (Figure 1).[18] Eukaryotic cells store their genetic information within the nucleus in the form of DNA. DNA is transcribed into a messenger RNA (mRNA) template, which is exported to the cytoplasm and translated into the corresponding amino acid sequence of a specific protein. It is now known that gene expression is under tight regulatory control by a vast number of parallel and overlapping mechanisms to ensure that protein and RNA are at an optimum level to respond to their constantly changing environment.[19]

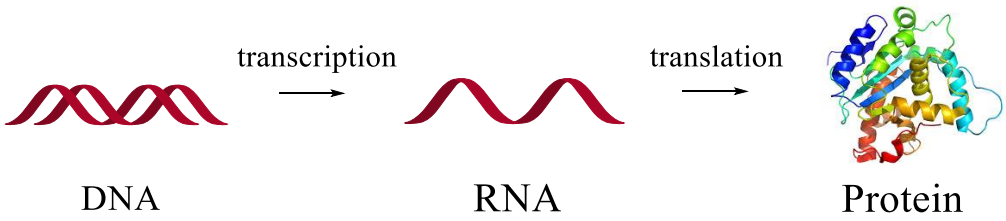


Figure 1: The central dogma of molecular biology. DNA is transcribed into an mRNA template which is translated into a specific protein.[18]

1.1.1.1 Nucleic acid

Nucleotides are the building blocks of DNA and RNA and are joined to form long chains called polynucleotides. Nucleotides are composed of three distinct components, a pentose sugar, a nitrogenous base, and a phosphate group. The main difference between DNA and RNA is that RNA contains a ribose sugar and DNA contains a deoxyribose sugar (Figure 2). The extra 2' hydroxyl group in the ribose sugar makes RNA more reactive than DNA and provides RNA with unique structural properties.[20] The nitrogenous bases in DNA are adenine, guanine, cytosine, and thymine, whereas in RNA thymine is replaced with uracil (Figure 3).[21]

Figure 2: Structures of DNA and RNA. DNA contains a deoxyribose sugar and RNA contains a ribose sugar.[21]

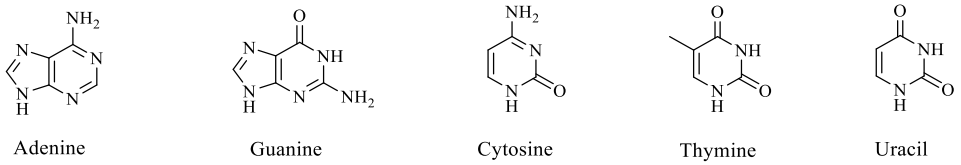


Figure 3: Structures of the 5 nitrogenous bases found in DNA and RNA.[21]

The characteristics of a particular DNA or RNA molecule are dependent on its unique sequence of nucleotides and the structures that they form. DNA is often composed of two complementary strands that are coiled together to form a double helix which is tightly packaged around histone proteins to form a

condensed structure known as chromatin. RNA on the other hand forms much more complex and variable three-dimensional structures.[19]

1.1.1.2 Types of RNA molecules

There are a number of different types of RNA molecules. RNAs can be categorised into two distinct groups, coding and non-coding RNA. mRNA is referred to as coding RNA as it contains the information to make a specific protein. Instead of providing a template for protein synthesis, non-coding RNAs fulfil a range of important structural and enzymatic functions. The most abundant non-coding RNAs are transfer RNA (tRNA) and ribosomal RNA (rRNA) that both function in protein synthesis. Other non-coding RNAs include short interfering RNA (siRNA), microRNA (miRNA) long non-coding RNA (lncRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), and a number of ribozymes that possess various enzymatic activities.[20] tRNAs are the most abundant of all RNA molecules with 10s of millions of transcripts in each human cell.[22]

1.1.1.3 mRNA structure

mRNA is first synthesised in the form of primary mRNA (pre-mRNA). In order to fulfil its role as a template for protein synthesis, the pre-mRNA must first undergo a number of processing steps to form mature mRNA (Figure 4).[20] A cap structure composed of N^7-methylguanosine (m^7G) is added to the 5' end of the mRNA.[23] m^7G is joined by an unusual triphosphate linkage that it joins the 5' hydroxyl group of the first nucleotide to the 5' hydroxyl group of m^7G (Figure 5).[21] Non-coding intronic segments within the pre-mRNA are

subsequently removed in a process called splicing,[24,25] and a series of adenosine nucleotides referred to as the poly(A) tail are added to the 3’ end of the mRNA[26]

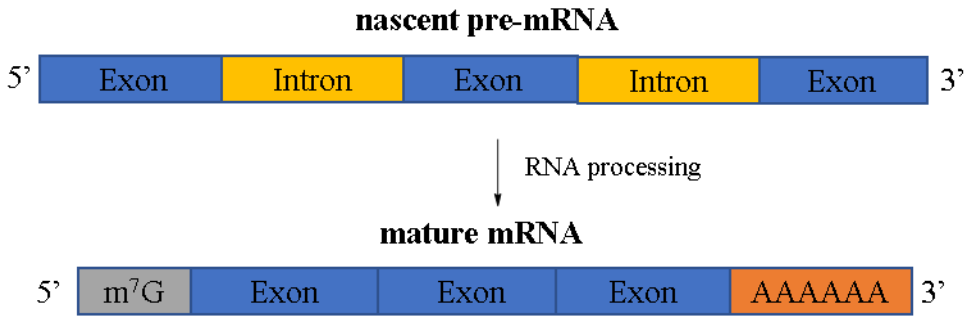


Figure 4: mRNA processing. The nascent pre-mRNA is converted into mature mRNA by the removal of introns (splicing) and the addition of the 5' m7G cap and the 3' poly(A) tail.[20]

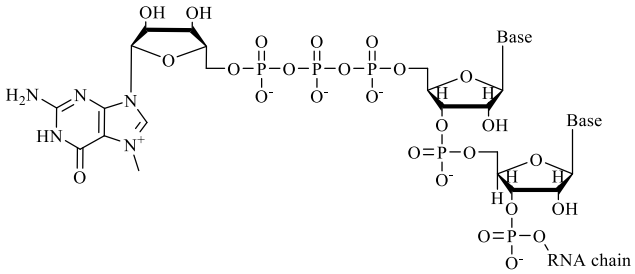


Figure 5: Structure of the mRNA 5' cap. m^7G is joined to the 5’ most nucleotide by a triphosphate linkage.[21]

The 5’- m^7G cap and 3’- poly(A) tail fulfil a number of important functions. They protect the mRNA from degradation, and are required for the initiation of protein synthesis, nuclear export and optimal splicing of the mRNA.[23,27–29]

1.1.1.4 Transcriptome

Human cells contain approximately 19,000 protein coding genes,[30] however, only a small fraction of these genes will be actively expressed in the form of mRNA at any one time. The transcriptome is a term used to describe the total pool of mRNA in a cell or an organism at a given point in time. The transcriptome is highly dynamic and is constantly changing as new mRNA is being transcribed and destroyed. This constant turnover of mRNA allows the cell to rapidly respond its protein output in response to its environment.[20]

1.1.1.5 Epigenetics

Epigenetics was initially defined as heritable changes to phenotype that are not caused by alterations to the DNA sequence. More recently the term epigenetics has expanded to include all protein and nucleic acid modifications irrespective of their heritability.[31] Epigenetic modifications can be found throughout the central dogma of molecular biology.[19] The most prominently studied epigenetic modifications include histone modifications, chromatin remodelling[32] and DNA methylation.[33]

1.1.1.6 Epitranscriptome

The epitranscriptome is a term used to collectively describe modifications to RNA.[13,34] The majority of RNA modifications reside in non-coding RNA and are required for the highly specific structural conformations of these RNA species. Relatively few modifications have been identified in mRNA.

1.1.1.7 Nucleotide modifications

All types of nucleic acids have been shown to contain modified nucleotides in addition to the unmodified nucleotides A, G, C, T, and U. These nucleotide modifications play important regulatory roles by impacting the structure of the nucleic acid and affecting the way in which it interacts with other biological molecules.[21] The majority of nucleotide modifications are methylations. The small size of this modification means it is non-mutagenic, and does not unintentionally inhibit the translational machinery.[35]

1.1.1.7.1 Modified nucleotides in DNA

The most common and well-studied DNA modification is 5-methylcytosine (5mC) (Figure 6).[21] This modification, discovered in 1951,[36] is primarily located within the promoter of genes at CpG sites (CpG islands). 5mC is recognised by 'reader proteins' that specifically bind to the nucleotide and regulate the expression of the gene.[37,38] Promoter localised 5mC represses gene expression, whereas, 5mC that is located in the gene body has been shown to promote gene expression,[39] and regulate the positioning of the nucleosome.[40] 5mC methylation is a dynamic and reversible modification that is converted back to unmodified cytosine by dioxygenase enzymes.[41] 5mC functions in a range of biological processes including X-chromosome inactivation[42] and genomic imprinting.[43]

Another DNA modification, 6mA (Figure 6) has been the subject of relatively little research. 6mA functions as a mechanism of cellular defence in bacteria,[44] and also regulates gene expression in prokaryotes and eukaryotes.[45,46]

Additional DNA modifications have also been identified including 4mC,[47] 5hmC,[41] 5fC,[41] and 5caC.[48]

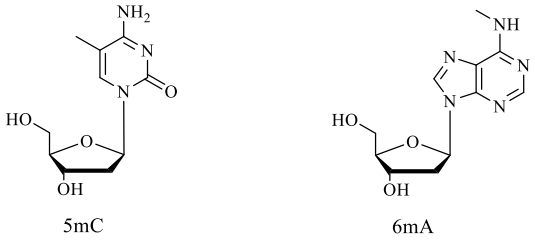


Figure 6: Structures of the modified nucleotides in DNA: 5mC and 6mA.[21]

1.1.1.7.2 Modified nucleotides in non-coding RNA

Over 150 unique chemical modifications have been identified in RNA of which more than 60 have been identified in mammals.[49] RNA modifications are considerably more prevalent that DNA modifications, however, despite this they have historically been the subject of relatively little research. The majority of RNA modifications reside in tRNA and rRNA molecules. tRNAs are particularly highly modified with an average of 13 modified nucleotides per tRNA molecule (Figure 7).[22]

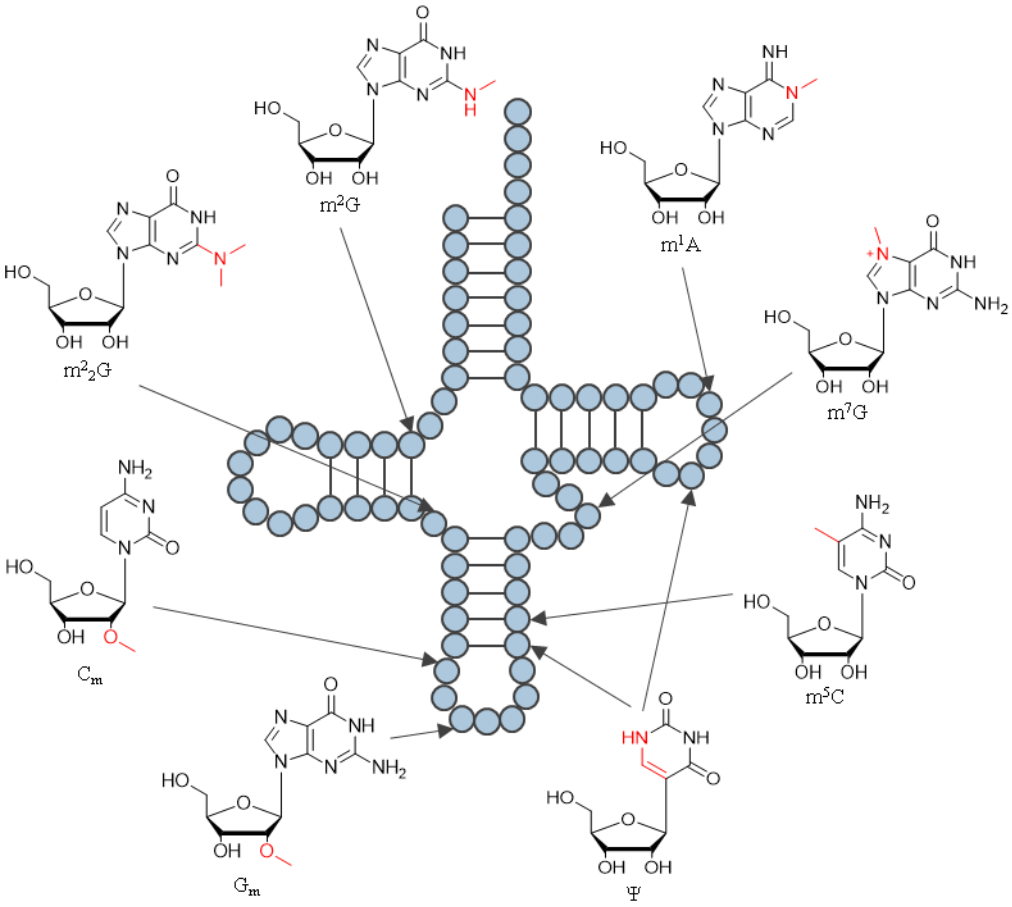


Figure 7: Modified nucleotides in yeast phenylalanine tRNA.13 out of the 94 nucleotides are modified.[50]

1.1.1.8 Modified nucleotides in mRNA

Despite the prevalence of modifications in non-coding RNA, very few nucleotide modifications have been identified in mRNA. mRNA modifications were first identified in the 1970s,[51] however, they were thought to have limited biological significance due to technological limitations preventing their study. With the advancement of RNA sequencing technologies, the presence of a small number of modified nucleotides in mRNA have now been confirmed and their fundamental biological roles are beginning to be uncovered.[35]

1.1.1.9 m^6A_m: methylation of the 5' cap

In addition to the 5'- m^7G cap, certain transcripts in higher eukaryotes are further modified at their 5'- end. The ribose sugar of the first 2 nucleotides adjacent to m^7G can be methylated (N_m). Cap-0 transcripts contain no N_m ribose methylation, Cap-1 transcripts are N_m methylated at the first nucleotide, and Cap-2 transcripts are N_m methylated at the first 2 nucleotides.[52] CMTR1 and CMTR2 catalyse the methylation of Cap-1[53] and Cap-2[54] nucleotides respectively (Figure 8).

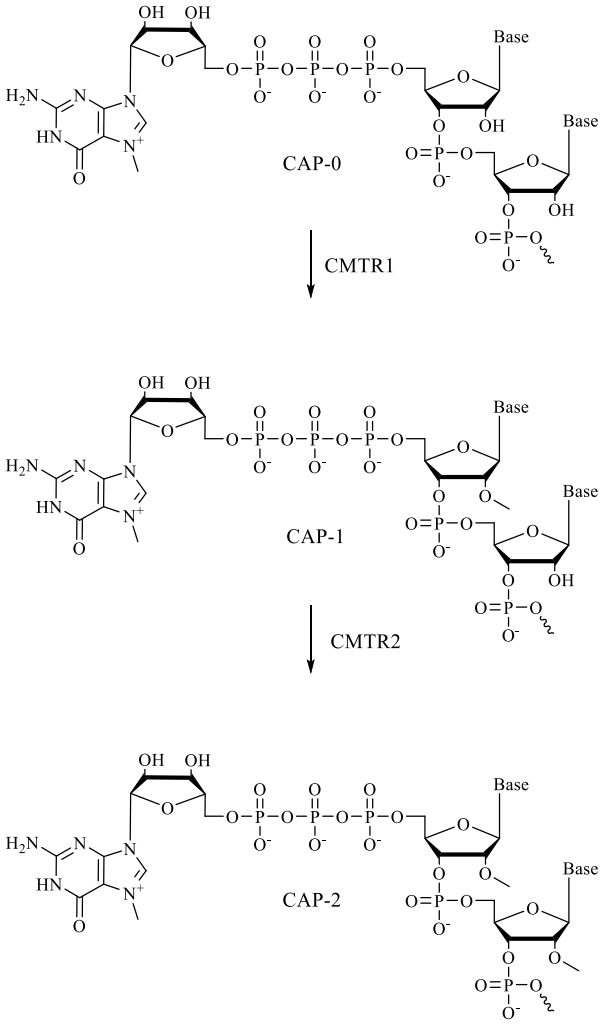


Figure 8: Ribose methylation of cap adjacent nucleotides. Cap-1 transcripts are N_m methylated at the first nucleotide by CMTR1[53].Cap-2 transcripts are also N_m methylated at the second nucleotide by CMTR2.[54]

If the first nucleotide of a Cap-1 transcript is an adenosine (A_m) then this nucleotide can be further methylated to m^6A_m, these transcripts are referred to as being Cap-1 m^6A_m modified (Figure 9).[29,55,56] This methylation is catalysed by the methyltransferase PCF1.[57–59] The cap adjacent modifications have been shown to play a number of roles in the regulation of gene expression, including stabilisation of the RNA and promoting translation. N_m methylation is also used as a mechanism to distinguish self from non-self RNA.[60,61]

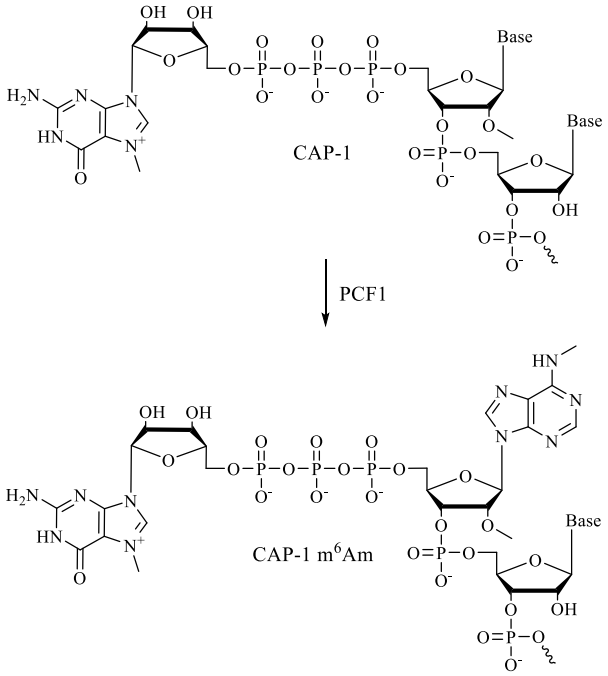


Figure 9: m^6A_m methylation. If the first nucleotide of a Cap-1 transcript is adenosine, then this nucleotide can be further methylated to m^6A_m by PCF1.[57–59]

In addition to m^7G methylation, it has been shown that the 5' guanosine cap can be hypermethylated to 2,7-dimethylguanosine ($m^{2,7}G$), and 2,2,7-trimethylguanosine ($m^{2,2,7}G$) in certain viral mRNAs and eukaryotic snoRNAs.[62]

1.1.1.10 Internal mRNA modifications

In addition to modifications adjacent to the 5' cap, a number of internal modified nucleotides have also been identified in mRNA. Our knowledge of internal modifications to mRNA is limited. Pseudouridine (Ψ),[63,64] 5-methylcytidine (m^5C),[51,65] 5-hydroxymethylcytidine (hm^5C),[66] N^4-acetylcytidine (Ac^4C),[67] Inosine (I),[68] N^6-methyladenosine (m^6A),[51] N^1-methyladenosine (m^1A),[69] and 2′-O-methylation (Nm)[70] are currently the only

internal modified nucleotides to have been identified in mRNA (Figure 10). However, due to limitations in the sequencing technologies of these nucleotides there is an ongoing debate about how abundant these modifications really are, and whether these modifications have any biological relevance or are they just non-functional artefacts present in the mRNA.[71]

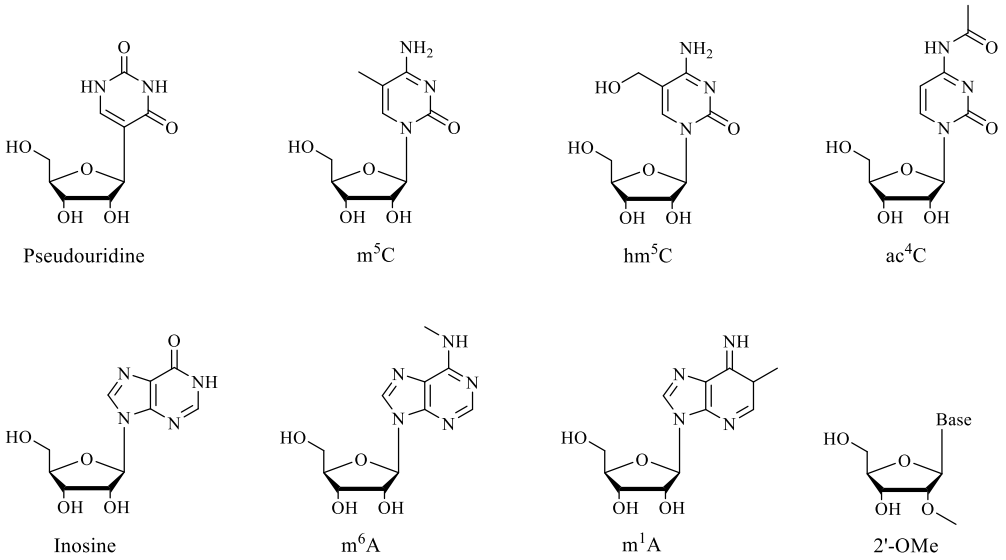


Figure 10: Internal modifications of mRNA. Pseudouridine (Ψ),[63,64] 5-methylcytidine (m^5C),[51,65] 5-hydroxymethylcytidine (hm^5C),[66] N^4-acetylcytidine (ac^4C),[67] inosine (I),[68] N^6-methyladenosine (m^6A),[51] N^1-methyladenosine (m^1A),[69] and 2′-O-methylation (Nm).[70]

1.1.1.11 m^6A

m^6A is the most abundant internal modification of mRNA. It is formed by the addition of a methyl group to the *N*-6 position of an adenosine base (Figure 11). m^6A is highly conserved throughout eukaryotes. Many components of the m^6A methyltransferase complex have been identified (Figure 11),[72–74] along with RNA binding proteins that specifically target m^6A.[5,6,75,76] There is also

evidence to suggest that the modification is reversible as two potential m^{6}A demethylase enzymes, ALKBH5 and FTO have been identified (Figure 11).[77]

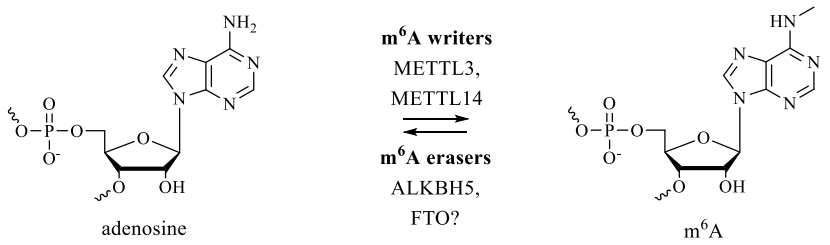


Figure 11: m^{6}A methylation is catalysed by a methyltransferase complex consisting of METTL3,[72] and METTL14,[73] together with WTAP and additional complex members. m^{6}A can be converted back to adenosine by the demethylase enzymes ALKBH5 and FTO.[77]

m^{6}A plays important regulatory roles at almost every stage in the life of an mRNA including splicing,[1] polyadenylation,[2] nuclear export,[3,4] stability,[5] translation,[6–8] and degradation.[5,9] The importance of m^{6}A in gene regulation is demonstrated by the devastating effects caused by overexpression and mutagenesis of the writers, erasers, and readers of m^{6}A. m^{6}A is essential for normal development of eukaryotic organisms, and is involved in the maintenance and differentiation of stem cells,[73,78] the circadian clock,[79,80] and the immune response.[81–83] Abnormal levels of m^{6}A have been associated with the development of a number of cancers including acute myeloid leukaemia (AML),[84,85] glioblastoma,[86,87] and breast cancer.[88]

1.1.1.11.1 m^{6}A stoichiometry

There are many copies of each mRNA transcript. For a specific mRNA only a fraction of the transcripts will contain the m^{6}A modification. For example, a

specific nucleotide position in bovine prolactin pre-mRNA contains the m^{6}A modification in 24 % of transcripts and adenosine in the remaining 76 % (Figure 12).[89] Early estimates suggested that m^{6}A sites are methylated in between 20-90 % of transcripts,[90] however, m^{6}A fractions that are significantly lower have since been identified. [91]

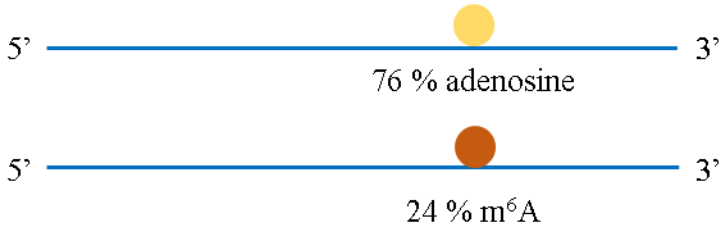


Figure 12: m^{6}A stoichiometry. A specific nucleotide position in bovine prolactin pre-mRNA contains the m^{6}A modification in 24 % of transcripts and adenosine in the remaining 76 % of transcripts.[89]

The stoichiometry of an m^{6}A site is very important. The amount of a transcript that needs to be methylated will depend on the particular role of the m^{6}A residue. If the m^{6}A mark is signalling for transcript to be degraded, then a high m^{6}A stoichiometry would be required to influence the levels of the particular mRNA within the total mRNA pool. If, however, the m^{6}A mark is signalling for the transcript to be preferentially translated, or alternatively spliced into a different isoform, then even a very low m^{6}A stoichiometry could have a very important biological significance. Furthermore, methylation of RNA comes as a reasonably high energetic cost, so it is likely that m^{6}A methylation, even when at low levels, is playing an important biological function.[92]

1.1.1.12 A brief history of m^6A research

1.1.1.12.1 Discovery of m^6A

m^6A methylation of mRNA was first discovered in mouse cells in 1974.[93] Despite being known about for over forty years, m^6A has only recently come to the forefront of molecular biological research. At the time of its discovery, many modified nucleotides had already been identified in tRNA and rRNA, however difficulties in purifying mRNA from non-coding RNA made analysis of mRNA methylation impossible.

Thankfully, stringent mRNA purification techniques using poly(T) cellulose were soon developed which allowed analysis of mRNA methylation. RNA methylation was measured by growing cells on [3H]-methionine which was incorporated into S-adenosylmethionine (SAM) which is the methyl donor for nucleotide methylation. Using this method, methylation of the 5' cap at the 2' ribose position was identified along with m^6A at internal positions within the mRNA.[93] Soon after its discovery in mice, m^6A was identified in human cells,[51] viruses,[94,95] and plants.[96,97]

Not long after its discovery m^6A was shown to be restricted to the consensus sequence of [A/G]-m^6A-C.[98] This indicated that m^6A was a targeted modification and not an artefact of unrelated chemistry. Subsequent studies have expanded the consensus sequence to the DRACH consensus motif ([G/A/U][G/A][m^6AC[U>A>C]).[99–101]

1.1.1.12.2 Difficulties in m^6A research

Despite being discovered in the 1970s, studies regarding m^6A were limited. This is because technological limitations made detection and analysis of m^6A extremely difficult. Additionally, many academics in the field remained sceptical about whether the identified m^6A actually originated from contaminating non-coding RNA due to doubts over the quality of the mRNA purification techniques. The m^6A modification was also believed to be biologically irrelevant as no functional role had been attributed to the modification.[102]

1.1.1.12.3 Revival of m^6A research

Over the next few decades a small number of research groups continued to pursue studies into the m^6A modification with the m^6A methyl transferase complex being partially identified in 1997.[72] The discovery of the methyltransferase protein allowed research into the biological function of m^6A and was shown to be an essential process in eukaryotes.[103]

Interest in m^6A was suddenly revived in 2012 after a number of important discoveries. The protein FTO (Fat mass and obesity associated) was shown to act as an m^6A demethylase[77] which led to the hypothesis that m^6A is a dynamically regulated modification that can be continuously tweaked in order to regulate gene expression. Soon after this discovery, two research groups developed a novel method for the transcriptome wide mapping of m^6A using a combination of m^6A immunoprecipitation and RNA sequencing (m^6A-Seq/MeRIP-Seq).[12,13] These experiments showed for the first time the abundance of the modification and made certain that the m^6A originated from

mRNA. These advances brought m^6A back into the limelight and led to significant advances in our understanding of m^6A and the epitranscriptome.

1.1.2 Detection and quantification of m^6A

Since its discovery, it has proved extremely difficult to accurately detect and quantify m^6A. Accurate transcriptome wide mapping and quantification of m^6A will be hugely beneficial to our understanding of the modification, allowing mutagenesis of individual m^6A residues which will help determine the functional activity of the modification in different cellular and sequence contexts.

1.1.2.1 Quantifying total m^6A levels

A number of methods can be used to quantify the total levels of m^6A in an mRNA sample. These include the use of thin layer chromatography (TLC),[74,104] mass spectrometry,[105] and HPLC.[105] These are reliable methods of m^6A quantification; however, they are time consuming and labour intensive, and it is difficult to ensure that all contaminating non-coding RNAs have been removed from the sample. Most importantly these methods are unsuitable for the site-specific detection and quantification of m^6A.[92,104]

1.1.2.2 Site-specific m^6A sequencing

RNA sequencing is a commonly used molecular biology technique which determines the nucleotide sequence of RNA molecules.[106] Certain modified nucleotides can be directly identified by RNA sequencing for example inosine,

which is reverse transcribed as if it were guanosine, allowing site specific detection of the modification.[107] However, standard RNA sequencing technologies are unable to differentiate between adenosine and m^6A.

Another possible method of m^6A sequencing would be to chemically treat the nucleotide to form another known base. This method is used to sequence the 5mC/m^5C modification in DNA and RNA, whereby the nucleic acid is treated with bisulphite which converts cytosine residues to uracil, whilst 5mC residues are unaffected. The C to U mutations in the cDNA are identified by sequencing.[108] However, currently no chemical treatment has been shown to convert m^6A to another common base.

1.1.2.3 Antibody based transcriptome mapping of m^6A

1.1.2.3.1 Me-Rip Seq

In 2012 a new ground-breaking method of m^6A sequencing called MeRIP-Seq was developed. MeRIP-Seq involves fragmentation of the mRNA followed by immunopreciptation using an anti m^6A antibody. RNA sequencing and is then used is used to identify fragments that contain m^6A (Figure 13).[12,13]

The main benefit of this method is that it provides the sequence context of the modified nucleotide, which ensures that the m^6A originated from mRNA and not contaminating non-coding RNA. Similar methods have since been developed for the detection of other mRNA modifications including m^1A and inosine.[69,109]

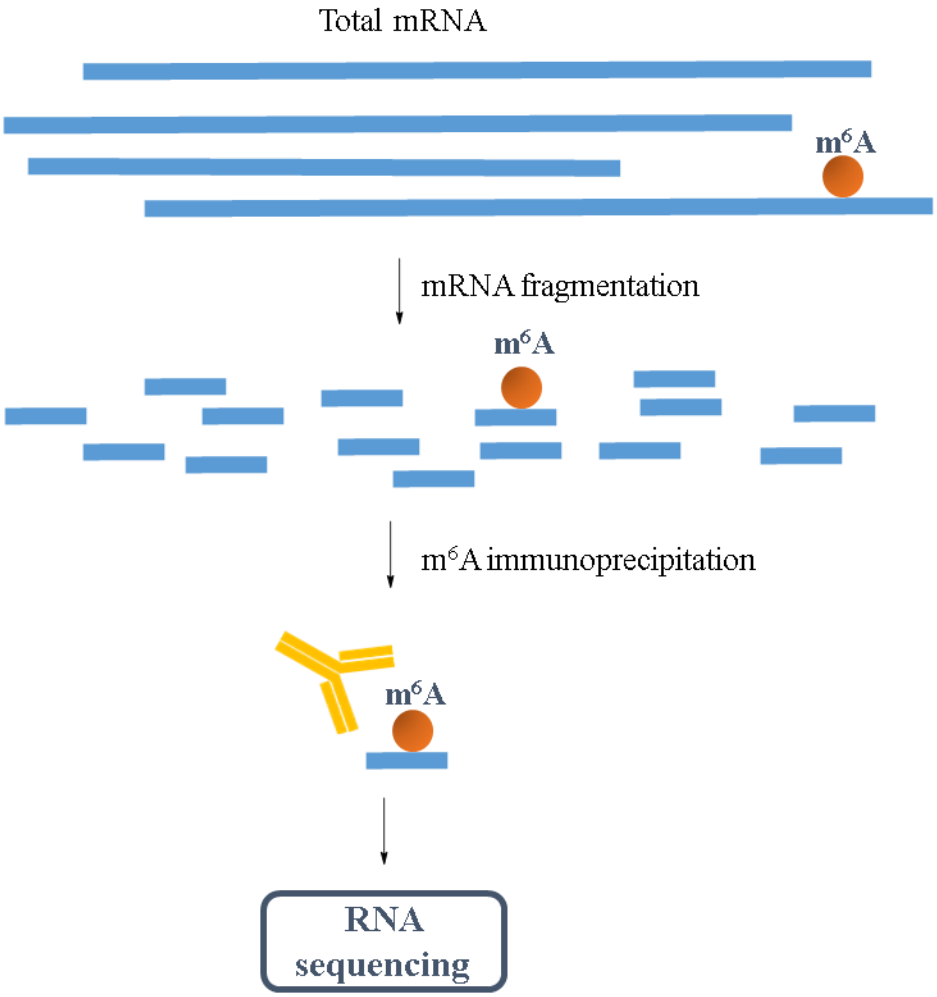


Figure 13: MeRIP-Seq sequencing of m^6A. The RNA is fragmented and immunoprecipitated using an anti m^6A antibody. m^6A containing fragments are then identified by RNA sequencing.[12,13]

Me-RIP-Seq was the first method to give a global view of m^6A throughout the transcriptome, identifying approximately 12,000 m^6A sites in the mRNA and lncRNA of human cells[12]. This showed for the first time the extent of the m^6A modification and led to a plethora of studies in the field. An interesting observation from these results is that only ~25% of transcripts appeared to contain the m^6A mark and ~10 % of the m^6A consensus sites are methylated showing that the modification is highly selective. m^6A sites were found to be

primarily enriched in the 3' UTR however m^6A peaks are also found at exon junctions, and the 5' UTR (many of these 5' UTR m^6A peaks have since been identified as m^6A_m).[12–14]

Unfortunately, there are a number of limitations to the MeRip-Seq method. Firstly, it is unable to detect the precise location of the modification as is limited to a resolution of approximately 200 nucleotides, the exact location can only be predicted by identifying the m^6A consensus motif (DRACH) within the region known to contain m^6A.[13] Secondly, MeRIP-Seq is unable to quantify m^6A levels. Additionally, Me-Rip Seq is also unable to identify multiple closely located m^6A sites (m^6A clusters).[13]

1.1.2.3.2 MiCLIP Seq

In an attempt to identify m^6A at a single nucleotide resolution, an improved version of MeRIP-Seq, named miCLIP (m^6A individual-nucleotide-resolution cross-linking and immunoprecipitation) was developed.[2,14] The method involves UV induced crosslinking between m^6A and the anti m^6A antibody, followed by reverse transcription. The crosslinking results in mutational signatures at the site at which the antibody was bound which are used to identify m^6A at a single nucleotide resolution. However, as with MeRIP-Seq, MiCLIP is also unable to quantify levels of m^6A.

A similar method has also been used to detect m^5C in which the m^5C writer, NSun2 is crosslinked to its binding sites on the mRNA. This results in mutational signatures that are identified by sequencing.[110]

1.1.2.3.3 Problems with the m^6A antibody

It has recently come to light that the anti m^6A antibodies used in MeRIP-Seq and MiCLIP are not as specific as it was first believed. The antibody has been shown to display off-target binding to purine rich regions, generating false positive results,[111] and additionally is unable to differentiate between m^6A and m^6A_m.[14] The abundance and location of m^6A peaks has varied considerably between studies and experimental repeats using the MeRIP-Seq method, however, this may be partly due to the unpredictable nature of the m^6A modification and not entirely representative of the quality of the method.[92]

1.1.2.4 Antibody independent m^6A detection

Due to the lack of specificity of the anti-m^6A antibody, antibody independent methods of m^6A detection are in high demand. A number of such methods have been developed in recent years.

1.1.2.4.1 SCARLET

One such antibody independent method of m^6A detection is SCARLET (site-specific cleavage and radioactive-labelling followed by ligation-assisted extraction and thin-layer chromatography).[90] The SCARLET method (described in detail in sections 3.1.2) involves radiolabelling a single A/m^6A nucleotide and quantifying the relative levels of adenosine and m^6A using thin layer chromatography. The main benefit of SCARLET is its ability to accurately detect and quantify m^6A at a single nucleotide resolution. However, this method is extremely time consuming, labour intensive and is only capable of analysing a single nucleotide site at once.[112] Despite its obvious drawbacks,

SCARLET will likely play an important role in research into m^6A. SCARLET can be used to test and validate new m^6A detection methods, and can be used for small scale research into m^6A sites, as it is currently the most accurate method of m^6A analysis.

1.1.2.4.2 4Se-dTTP based m^6A detection

In 2018, Hong and co-workers developed an antibody independent method of m^6A detection. They utilised a chemically modified nucleoside triphosphate (4Se-dTTP) (Figure 14), which causes truncation of reverse transcription specifically at sites of m^6A in the RNA template. The reverse transcription products are sequenced and compared with a non-methylated control (generated using the demethylase FTO) resulting in accurate mapping of the m^6A modification.[113]

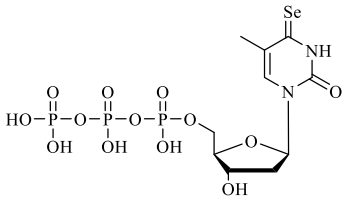


Figure 14: Structure of 4Se-dTTP. Used for the detection of m^6A in RNA.[113]

1.1.2.4.3 m^6A sensitive nuclease enzyme

MazF is an *E.coli* toxin that possesses sequence specific endoribonuclease activity. MazF makes a targeted cleavage at the 5' of the first A in the sequence motif 'ACA',[114] however, is sensitive to m^6A and is unable to cleave RNA in the 'm^6A-CA' sequence context.[115] This specific enzyme has been combined with RNA sequencing to map m^6A throughout the transcriptome in two independent methods 'm^6A-REF-seq'[116] and 'MASTER-seq'.[91]

These methods are both limited to the identification of m^6A residues within the m^6A-CA sequence context which accounts for only a fraction of the total m^6A found in mRNA. Also, MazF has not yet been tested on RNA substrates containing other modifications within the 'ACA' sequence motif such as m^1A, m^5C, A_m, m^6A_m, so may also exhibit sensitivity towards these modifications.

1.1.2.4.4 Metabolic labelling of m^6A sites

Metabolic labelling involves the introduction of unnatural metabolites into the metabolic pathways of cells. These metabolites can have unique chemical features to aid the detection and study of particular biomolecules.[117] SAM is the methyl donor of m^6A methylation of mRNA. A number of research groups have developed SAM analogues for the *in vivo* detection of m^6A.

The first example of this was a SAM analogue with an aryl group in place of the methyl group.[118] mRNA was extracted from cells and incubated with the aryl-SAM and the methyltransferase proteins METTL3 and METTL14. The methyltransferase proteins arylated the mRNA at specific adenosine residues. The mRNA was then treated with iodine which promotes a cyclisation reaction (Figure 15). The product of this reaction is incapable of nucleotide base pairing so results in base mutations or truncations in reverse transcription reactions. Sequencing of cDNA products can be used to detect and quantify the exact location of the m^6A modification.[118]

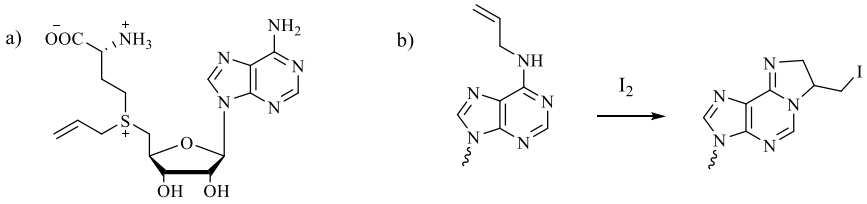


Figure 15: a) Aryl modified SAM. b) Aryl modified adenosine is reacted with iodine, the cyclised product results in reverse transcription mutations.[118]

More recently a similar study replaced the methyl group of SAM with a propargyl group, and also the sulphur atom with selenium (the sulphur-selenium substitution was used to improve the stability and activity of the modified SAM) (Figure 16).[119] The metabolite was introduced into HeLa cells and the propargyl group was incorporated into mRNA by the methyltransferase complex. The propargyl group is then reacted with azide-biotin and purified with streptavidin magnetic beads. The modification is sufficient to inhibit reverse transcription so sites of m^6A are identified by sequencing the cDNA products of the reverse transcription reaction.[120]

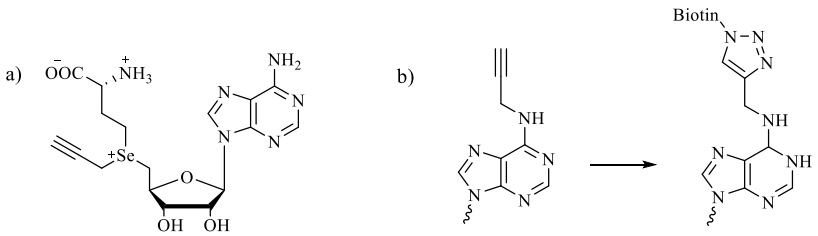


Figure 16: a) propargyl modified SAM. b) propyl modified adenosine is reacted with azide-biotin. This RNA modification results in stalling of reverse transcription reactions.[120]

1.1.2.5 Third generation m^{6}A sequencing

Third generation nucleic acid sequencing is a new class of sequencing methods in development.[121] Standard methods of nucleic acid sequencing involve fragmentation of the nucleic acid followed by amplification.[122] In contrast, third generation sequencing analyses the nucleotide sequence on a single molecule level, without the need for amplification. The most common third generation sequencing technologies are Nanopore and SMRT sequencing.[121] Both of these methods are being developed for the sequencing of nucleotide modifications including m^{6}A in mRNA.[123,124]

1.1.2.5.1 Nanopore

Nanopore sequencing involves transferring nucleic acid through a nanopore protein. Nanopore proteins are naturally found in the cell membrane where they facilitate the transport of molecules in and out of the cell. In nanopore sequencing the nanopore protein is placed into an electrically resistant membrane and an ionic current is passed through the nanopore. The current is disrupted as the nucleic acid chain passes through the nanopore. The characteristic changes in the current are used to determine the sequence of the nucleic acid.[121]

There is great potential for nanopore technology, however, nanopore sequencing currently has a high error rate when sequencing unmodified nucleotides and is not yet fully optimised for the sequencing of RNA.[121]

1.1.2.5.2 SMRT sequencing

Single molecule real time (SMRT) sequencing uses nucleoside triphosphates that are each linked to coloured fluorescent labels at their terminal phosphate. Nucleic acid is amplified using a DNA polymerase enzyme and as a nucleoside triphosphate is added to the growing chain the fluorescent label in cleaved. The released fluorescent label is passed through a zero-mode wave guide to a detector which determines the identity of the nucleotide based on the colour of the fluorescent signal.[121]

1.1.2.6 The future of m^6A detection

Despite many m^6A detection methods being developed in recent years there is still a demand for new methods of accurate transcriptome wide mapping and quantification of m^6A. Such methods are essential for the progression of our understanding of the m^6A modification.

1.1.3 The m^6A methyl transferase complex (m^6A writers)

m^6A is deposited onto mRNA transcripts by a multi component methyltransferase complex that uses SAM as the methyl donor.[51,125] In mammals the methyltransferase complex consists of the proteins METTL3, METTL14, WTAP, KIAA1429, RBM15, HAKAI, and Zc3h13.

1.1.3.1 METTL3

The first component of the methyltransferase complex to be identified was METTL3.[72] METTL3 is highly conserved in eukaryotes and since its initial discovery, functional homologs have been identified in Drosophila (Ime4),[126] Saccharomyces cerevisiae (IME4),[127] and Arabidopsis thaliana (MTA)[74]. Knockdown of METTL3 results in embryo lethality in both mice[128] and plants,[74] and results in increased levels of apoptosis in HepG2 (human liver carcinoma) cells,[12] and zebrafish embryos.[129] Drosophila flies lacking Ime4 survive to adulthood, however, have a range of developmental defects including the inhibition of oogenesis, abnormal walking speed, locomotion, the inability to fly, and altered male: female ratios due to disrupted sex determination in females.[126,130–132] Yeast can also survive without Ime4 however this results in the termination of sporulation.[133]

1.1.3.2 METTL14

20 years after the discovery of METTL3, a second methyltransferase protein, METTL14 was shown to function in m^{6}A methylation.[73] METTL3 and METTL14 join to form a stable heterodimer that has a considerably greater methyltransferase activity than either of the proteins individually.[73] Initial studies showed that METTL14 has a slightly higher methyltransferase activity than METTL3 *in vitro*. However, it has now come to light that METTL3 is the only catalytic component of the m^{6}A methyltransferase complex. It has been shown that METTL14 contains side chains within its putative SAM binding pocket that sterically inhibit an interaction with SAM. METTL14 also lacks

the essential aromatic residues required to facilitate an interaction with the adenosine substrate.[134,135]

Instead of catalysing the methylation reaction, METTL14 functions as a structural protein that stabilises the protein complex.[134,136] A recent study has shown that trimethylation of lysine 36 of histone H3 is a marker for m^6A methylation of newly transcribed mRNA. METTL14 directly interacts with the methylated lysine residue and recruits the methyltransferase complex to RNA polymerase II.[137]

1.1.3.3 WTAP

WILMS TUMOUR1-ASSOCIATED PROTEIN (WTAP), a protein originally known for its roles in alternative splicing[138] and embryo development[139] was the third protein to be implicated in the mammalian methyltransferase complex. WTAP was first linked to the methyltransferase complex in a study of its Arabidopsis homolog, FIP37 which was shown to interact with the MTA methylase.[74] Knock down of WTAP decreased m^6A levels to a greater extent than either METTL3 or METTL14. WTAP does not possess any methyltransferase activity, instead it appears to function in recruiting additional accessory proteins to the methyltransferase complex and targeting the complex to specific sites in the mRNA.[73,129]

1.1.3.4 Additional accessory proteins in the methyltransferase complex

A number of regulatory proteins also play important roles in the methyltransferase complex:

- **Slz1** is a yeast protein that functions in shuttling the methyltransferase complex from the cytoplasm to the nucleolus which is essential for methylation and subsequent meiotic induction.[15]
- **KIAA1429** interacts with the methyltransferase METTL3.[140] Silencing of KIAA1429 in mammals, plants, and flies results in significant reductions in m^6A levels.[131,140,141] However, the mechanism by which the protein is involved in m^6A regulation is yet to be fully understood.
- **Zc3h13** is a mammalian protein that recruits the methyltransferase complex to the nucleus. Depletion of Zc3h13 expression results in a significant reduction in m^6A levels.[142]
- **HAKAI** is an E3-ubiquitin ligase that interacts with the methyltransferase complex in mammals, Arabidopsis and Drosophila. Knockdown of HAKAI significantly decreases m^6A levels.[141,143,144]
- **RBM15 and RBM15B** are mammalian proteins that target the methyltransferase complex to specific sites in the mRNA.[145] RMB15 knockout in mice results in embryo lethality and a range of defects related to the heart, spleen, and vasculature.[146,147]

1.1.4 Demethylation of m^6A (m^6A erasers)

mRNAs are generally short lived with a high rate of turnover; they are constantly being transcribed and degraded allowing the cell to rapidly alter its protein output in response to changing conditions.[20] For this reason, m^6A was assumed to be a static modification that could only be removed by degradation

of the methylated transcript. However, recent evidence suggests that the modification may in fact be reversible.

1.1.4.1 FTO

In 2012, Jia and colleagues demonstrated that the human protein, Fat mass and obesity associated protein (FTO) is able to demethylate m^6A *in vitro*. They also found that knockdown of FTO in human cell lines increased m^6A levels and overexpression of FTO decreased the levels of m^6A.[148] This led to the hypothesis that m^6A is not a static modification, but is instead a dynamic process, governed by the activities of the m^6A methyltransferase and demethylase proteins. *In vitro* studies of FTO dependent demethylation of m^6A showed that FTO oxidises m^6A to form N^6-hydroxymethyladenosine (hm^6A) and N^6-formyladenosine (f^6A) which converted back to adenosine with a half-life of 2.5 hrs (Figure 17). Both hm^6A and f^6A were identified in poly(A)+ RNA from human cells and mouse tissues.[149]

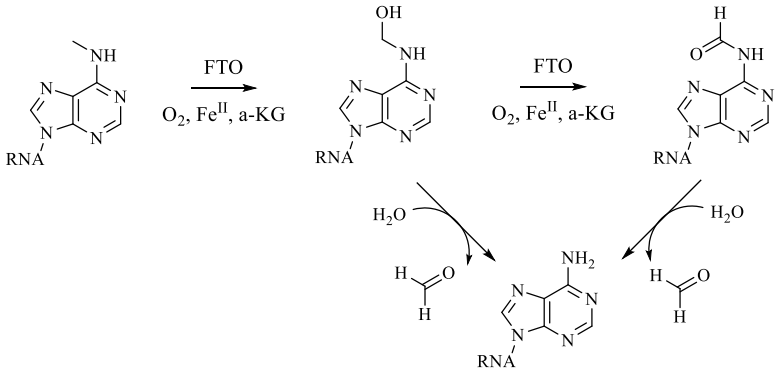


Figure 17: FTO oxidises m^6A to form hm^6A and f^6A *in vitro*. These molecules were converted into adenosine under physiologically relevant conditions.[149]

However, there is evidence to suggest that FTO is not actually an m^6A demethylase. The catalytic activity of FTO towards m^6A is considerably lower than that of similar dioxygenase enzymes[150] and FTO does not preferentially bind to, or demethylate m^6A within its consensus sequence, DRACH.[150,151] Furthermore, MeRIP-Seq analysis of FTO knockout mice exhibited only a small increase in the number of methylated transcripts. This indicates that the *in vitro* m^6A demethylation activity of FTO is not specific towards m^6A.[150]

It has since come to light that FTO actually functions as the demethylase of m^6A_m. FTO demethylates m^6A_m with a catalytic efficiency around 100 times higher than it does m^6A.[152] It was also shown that FTO will specifically demethylate m^6A_m when incubated with a mixture of m^6A and m^6A_m substrates.[152]

1.1.4.2 ALKBH5

Another candidate m^6A demethylase is ALKBH5 which demethylates m^6A but not m^6A_m *in vitro*.[77,152] ALKBH5 overexpression has been shown to decrease m^6A levels and conversely its knockdown increases m^6A levels. ALKBH5 knockdown mice have been observed to have spermatogenesis defects but are otherwise normal.[77] ALKBH5 is a nuclear localised protein, whilst mRNA transcripts are transported to the cytoplasm soon after transcription, this means the m^6A demethylation activity of ALKBH5 must take place either on the nascent pre-mRNA, or soon after the mature mRNA is synthesised.[77]

FTO and ALKBH5 are members of the non-heme Fe(II)- and α-KG-dependent dioxygenase AlkB protein family. FTO and ALKBH5 have been shown to have unique structural elements, a nucleotide recognition lid domain, and L1 loop domain, these features may provide the methylases with specificity towards m^6A_m and m^6A respectively.[153,154]

1.1.4.3 Functional role of mRNA demethylation

Only a small percentage of transcripts are demethylated, despite the ubiquitous expression of FTO and ALKBH5.[155] Future research will attempt to understand the mechanisms that determine which transcripts are demethylated.

1.1.4.1 Programable RNA methylation

Liu and colleagues (2019) developed a method for the manipulation of m^6A levels at specific sites within the mRNA, without effecting the primary nucleotide sequence.[156] They engineered an artificial m^6A writer protein by fusing the catalytic domains of METTL3 and METTL14 to CRISPR-Cas9. In a similar manner, they engineered an artificial m^6A eraser proteins using either ALKBH5 or FTO. These artificial writer and eraser proteins are targeted to specific sites within the mRNA using a sequence specific guide RNA (sgRNA).[156] This allows the study of single m^6A sites and will be a valuable tool for uncovering the role of the m^6A modification.

These artificial writer and eraser proteins are predominantly localised in the cytoplasm,[156] whereas METTL3 and METTL14 are naturally located in the nucleus.[72,73] m^6A regulates a number of nuclear localised processing events that occur early in the life of the mRNA (such as splicing,[1] polyadenylation,[2]

and nuclear export)[3,4] which will not be affected by these artificial m^{6}A writer and eraser proteins.

1.1.4.2 Future research into m^{6}A writers and erasers

Over the last few years our understanding of the writers and erasers of m^{6}A has improved significantly. Despite this there are still a number of unanswered questions. Many scientists in the field have predicted that there is much greater complexity to the methyltransferase and demethylase complexes and it is likely that in the coming years additional proteins and regulatory RNAs will be identified as writers and erasers of m^{6}A.[92]. Identifying these proteins and the precise mechanism by which they act will be of great importance. In order to achieve this, it will be important to have a method to accurately detect and quantify m^{6}A.

1.1.5 m^{6}A mechanism of action (m^{6}A readers)

Many biological functions have been attributed to m^{6}A with regulatory roles at almost every stage in the life cycle of the mRNA including splicing,[1] polyadenylation,[2] nuclear export,[3,4] stability,[5] translation,[6–8] and degradation.[5,9] However, the exact mechanisms by which m^{6}A acts remains elusive. This is largely due to technological limitations, for instance, methods to artificially manipulate m^{6}A levels at specific nucleotide positions are limited,[156] and methods for the detection and quantification of m^{6}A require significant improvement. Compounding the problem is the fact that the

function of an individual m^6A residue appears to depend on a range of factors including the location of the modification within the transcript, the specific tissue the mRNA resides in, the levels of stress that the cell may be facing, and the mechanism by which m^6A is recognised.[92]

1.1.5.1 m^6A functions via RNA binding proteins

RNA binding proteins are proteins that specifically recognise and interact with RNA molecules and play fundamental roles in post transcriptional processes.[20] RNA binding proteins interact with RNA molecules via their RNA binding domains which recognise RNA molecules based on either their nucleotide sequence or their three-dimensional tertiary structure.[20]

There are a number of mechanisms by which RNA binding proteins could potentially interact with m^6A: 1) proteins may directly bind to m^6A. 2) proteins binding may be inhibited by the modification. 3) Changes to the mRNA secondary structure caused by the presence or absence of m^6A could regulate binding between the protein and the RNA.

To date a number of m^6A binding proteins (m^6A readers) have been identified that either directly or indirectly bind to m^6A modified RNA. Reasonable progress has been made in understanding how these proteins recognise m^6A and the biological role that they play. m^6A readers include: YTHDF1,[6] YTHDF2,[5] YTHDC1,[75] IMP1,[157] and HNRNPA2B1.[158]

1.1.5.2 YTH protein family

The YTH domain family are a group of proteins that selectively bind to m^6A containing transcripts.[159] These proteins all consist of a low complexity N terminal domain and a characteristic C terminal YTH domain. The YTH domain contains a hydrophobic pocket that is the acceptor site of the methyl group of m^6A.[75] Mammals have five YTH proteins, compared with thirteen in Arabidopsis, and just two in Drosophila. The mammalian YTH proteins are YTHDC1 and YTHDC2 which are nuclear localised, and YTHDF1, YTHDF2, and YTHDF3 which are located in the cytoplasm.[159]

1.1.5.3 YTHDF1, YTHDF2 and YTHDF3

YTHDF2 was the first m^6A reader to be discovered. YTHDF2 specifically binds to modified transcripts and targets them for degradation.[5] YTHDF2 recruits the CCR4-NOT deadenylase complex which functions in the removal of the mRNA poly(A) tail, which is an important step in mRNA degradation.[9]

At a certain point during embryo development, the maternal genome is replaced by the zygotic genome. This process involves removal of all components related to the maternal genome.[160] YTHDF2 plays a key role in this process by targeting m^6A modified maternal transcripts for degradation.[161]

YTHDF1 has the opposite role to YTHDF2 as it promotes protein synthesis of m^6A containing transcripts.YTHDF1 facilitates ribosome binding to the mRNA and recruits a number of translation initiation factors.[6]

YTHDF3 plays a role in both degradation and translation of m^{6}A modified transcripts by interacting with either YTHDF2 or YTHDF1. YHDF1 and YTHDF2 share a large number of their target RNA targets and appear to work together in order to positively and negatively regulate the levels of m^{6}A methylated transcripts.[162,163]

1.1.5.4 YTHDC1: an m^{6}A dependent regulator of splicing

YTHDC1 is a nuclear localised m^{6}A reader that regulates the splicing of nascent mRNA. It was first suggested that m^{6}A was involved in splicing by Salditt-Georgieff in 1976 who observed that the average number of m^{6}A residues per transcript was higher in nuclear pre mRNA (~4 residues) than in mature cytoplasmic mRNA (~2 residues) and suggested that the m^{6}A methylated nucleotides were lost during splicing.[164] m^{6}A levels are significantly higher in transcripts that undergo alternative splicing and m^{6}A peaks accumulate at the 5' and 3' splice sites of nascent pre-mRNA.[12,13]

YTHDC1 regulates alternative splicing of m^{6}A modified transcripts by recruiting various splicing factors to the mRNA.[1] YTHDC1 recruits SRSF3 to the m^{6}A modified mRNA to promote exon inclusion and recruits SRSF10 to have the exon removed from the mature transcript.[1,165] YTHDC1 also promotes transport of mature m^{6}A modified transcripts from the nucleus to the cytoplasm. This nuclear exported is facilitated through interactions with SRSF3 and NXF1 (nuclear RNA export factor 1).[3]

1.1.5.5 IMP proteins

IMP1, IMP2, and IMP3 are a novel class of m^6A reader proteins in humans. The IMP proteins have been shown to stabilise m^6A containing transcripts, resulting in increased levels of transport and translation. Thousands of high confidence methylated targets of the IMP proteins have been identified.[157]

1.1.5.1 eIF3 – Cap independent translation

Under normal conditions, protein translation is initiated by the eukaryotic initiation factor 4E (eIF4E) that binds to the 5'- m^7G cap and recruits the 43S ribosomal complex.[166] However, during cellular stress, translation can occur independently of the 5'- cap in an m^6A dependent manner.[167] Eukaryotic initiation factor 3 (eIF3) binds to an m^6A residue at the 5'- of certain transcripts known as an m^6A induced ribosome engagement site (MIRES).[8] eiF3 recruits the 43S ribosomal complex to the mRNA to initiate translation. This process occurs independently of the m^7G cap and eIF4E and is known as cap independent translation.

During cellular stress increased levels of m^6A are observed in the 5' UTR.[8] It is hypothesised that cap independent translation allows a different set of proteins to be synthesised as a response to cellular stress, whereby m^6A marks a subset of transcripts for preferential translation.[168]

1.1.5.2 METTL3 as an m^6A reader

In a recent study, METTL3 has been shown to be both a reader and writer of m^6A. METTL3 appears to enhance translation of m^6A containing transcripts.

The methyltransferase activity of METTL3 comes from the C terminus of the protein, whereas the reader function is from the N terminus.[169]

1.1.5.3 HNRNPA21B

HNRNPA2B1 has been identified as a nuclear reader of m^{6}A. This protein has been linked to alternative splicing of m^{6}A containing mRNAs. HNRNPA2B1 has also been shown to regulate the processing of m^{6}A containing miRNAs, by interacting with the miRNA processing enzyme DGCR8.[158]

1.1.5.4 Indirect m^{6}A readers (the m^{6}A switch)

m^{6}A has been shown to have significant effects on RNA secondary structure and has been shown to destabilise both RNA/RNA and RNA/DNA base pairing.[154] This destabilisation is caused by the *N*^{6}-methylamino group of m^{6}A being forced to rotate into a high energy *anti* conformation in order to base pair with uracil (Figure 18).[154,170] A single m^{6}A residue is sufficient to destabilise an RNA duplex and promote the formation of single stranded RNA.[171]

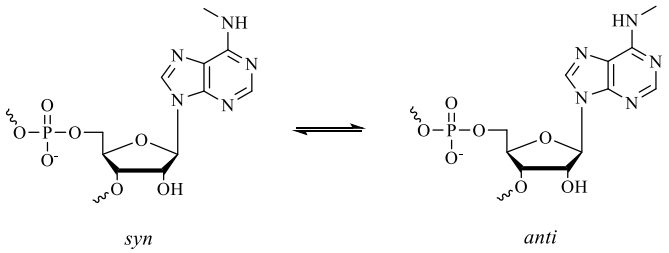


Figure 18: m^{6}A base pairing. In order to base pair with uracil, the *N*^{6}-methylamino group of m^{6}A must rotate from its natural *syn* conformation to the energetically less favourable *anti* conformation.[170]

The changes in RNA structure caused by m^6A can alter the interactions between the mRNA and RNA binding proteins. The binding motifs on RNA transcripts can be hidden within the complex folds of the RNA molecule. Therefore, mRNA structure can allow or inhibit access to the RNA binding protein. This recently discovered phenomenon of proteins indirectly binding to m^6A is known as the m^6A switch (Figure 19).[172,173]

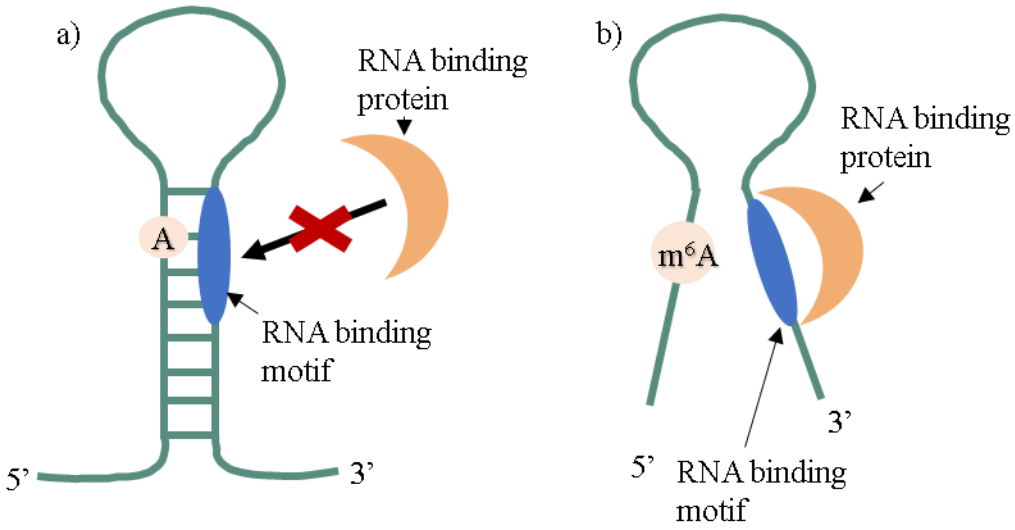


Figure 19: The m^6A switch. a) the RNA binding protein is unable to access the RNA binding motif. b) m^6A destabilises the RNA base pairing. The RNA binding motif is now single stranded allowing the RNA binding protein to bind (Adapted from Liu and Pan 2016).[174]

HNRNPC is an example of an indirect m^6A reader that functions in post transcriptional processing of mRNA transcripts. HNRNPC binds to U tract motifs (repeats of at least 5 uridines), however, these motifs are often located in the stem of stem loop structures which prevents HNRNPC binding.[175] m^6A methylation within the stem destabilises the base pairing, resulting in a single

stranded U tract motif, facilitating HNRNPC binding.[173,176,177]. Approximately 40,000 HNRPC based m^{6}A switches have been identified in mRNA.[173]

1.1.5.5 m^{6}A can inhibit RNA binding proteins

The m^{6}A modification has been shown to inhibit certain RNA binding proteins which preferentially interact with unmodified transcripts. G3BP1 is a stress granule protein that is inhibited by the presence of m^{6}A in transcripts. The m^{6}A dependent binding of G3BP1 appears to stabilise the unmodified RNA transcripts, preventing their degradation.[178]

1.1.6 Biological consequences of m^{6}A

The ability to profile m^{6}A levels throughout the transcriptome combined with our understanding of m^{6}A writers, erasers, and readers has provided the tools for elucidating the biological role of m^{6}A. Over the last few years a number of studies have linked m^{6}A to biological functions and disease.

1.1.6.1 Stem cell reprograming

Stem cells are undifferentiated cells which undergo self-renewal and are able to differentiate into other cell types.[179] There is a clear link between m^{6}A and the regulation of stem cells, however, the mechanism of action is not yet understood. The majority of embryonic stem cell related transcripts are m^{6}A methylated which has been shown to reduce the half-life and abundance of these transcripts.[128] The mechanism by which m^{6}A destabilises these transcripts is unknown; however, it is possible that an m^{6}A binding protein such as YTHDF2 selectively targets these transcripts for degradation.[180]

Undifferentiated stem cells are in either a 'naïve', or 'primed' state. In naïve stem cells, m^6A destabilises pluripotency-promoting transcripts, resulting in differentiation.[180] In contrast, in primed stem cells m^6A destabilises differentiation promoting transcripts, resulting in the maintenance of a pluripotent state.[78,128]

Zinc Finger Protein 217 (ZFP217), a transcription factor required for stem cell reprogramming, interacts with METTL3 and prevents methylation of genes involved in maintaining a pluripotent state including Nanog and Sox2. This METTL3 inhibition indirectly protects these transcripts from being destabilised and maintains the undifferentiated state of the stem cells.[181]

1.1.6.1 Cancer

Cancer refers to a group of diseases that are caused by uncontrolled cell division of abnormal cells.[182] For a cell to transform into a cancer cell, the genetic regulation of cell growth and differentiation must be altered. A number of studies have linked m^6A to cancer development, with abnormal levels of the m^6A modification promoting self-renewal and proliferation of cancer cells.

Hypoxia in breast cancer cells has been shown to upregulate the expression of the m^6A demethylase ALKBH5.[88] The increased levels of ALKBH5 result in demethylation of NANOG mRNA transcripts, stabilising them, and protecting them from degradation. NANOG is an oncogene (oncogenes promote the growth and differentiation of cancerous cells) that supresses differentiation and maintains the pluripotent state that is required for cancer progression.[88] Similarly, in glioblastoma (a type of brain cancer) ALKBH5 is upregulated,

resulting in demethylation of mRNA of the oncogene FOXM1, which protects the transcript from degradation.[86,87]

In hepatocellular carcinoma (liver cancer) the methyltransferase METTTL3 is upregulated. This leads to increased methylation of the tumor suppressor SOCS2 leading to its degradation via the m^{6}A reader YTHDF2.[183] METTL3 is also upregulated in acute myeloid leukemia (AML) (cancer of the blood and bone marrow), this increases methylation levels of a number of oncogenes including MYC, however, in this case the m^{6}A mark stabilises the transcripts and promotes their translation.[84,85]

m^{6}A has also been associated with many other cancers including lung cancer,[184] pancreatic cancer,[185] gastric cancer,[186] nasopharyngeal carcinoma,[187] colorectal cancer,[188] prostate cancer,[189] and renal cell carcinoma.[190] The strong association between m^{6}A and cancer development suggests that clinical applications involving the modification may be developed either for diagnosis or therapeutic treatment of these diseases.[11]

1.1.6.1 Xist m^{6}A mediated silencing

X chromosome inactivation is the random silencing of one of the X chromosomes in female mammals. This process is regulated by XIST, a nuclear non-coding RNA (ncRNA) that coats the entirety of one of the X chromosomes and recruits chromatin-modifying factors to facilitate the silencing of the chromosome.[191]

m^{6}A plays a fundamental role in this process.[145] The m^{6}A methyltransferase complex is targeted to XIST RNA by RBM15 and RBM15B, resulting in high

levels of methylation of the transcript.[145] A total of 76 m^{6}A residues have been identified within the XIST transcript and global m^{6}A depletion has been shown to inhibit X chromosome inactivation.[145] YTHDC1 binds to XIST m^{6}A and is also essential for X chromosome silencing, however, the exact mechanism by which YTHDC1 promotes X chromosome silencing is currently unknown.

1.1.6.2 Circadian clock

The circadian clock is an oscillation of approximately 24 hours that regulates transcription, translation, and biological activity in coordination with the day-night cycle.[192] A number of studies have linked m^{6}A with the regulation of the circadian clock. Many circadian clock genes and clock regulated transcripts contain the m^{6}A mark, and reducing global m^{6}A levels by METTL3 silencing has been shown to cause a delay in mRNA processing and an extension of the circadian period.[79,80]

1.1.6.3 The role of m^{6}A in viral infection and the immune response

m^{6}A in viral mRNA appears to enhance gene expression by promoting translation and enhances viral replication. HIV infection significantly increases the levels of m^{6}A in the mRNA of both virus and the host.[193,194] There is the potential to develop antiviral therapeutics which specifically inhibit viral m^{6}A.[194]

1.1.6.4 m^6A and miRNA

There is a link between the presence of m^6A and miRNA binding sites on mRNA transcripts. 67 % of transcripts with at least 1 m^6A residue in their 3' UTR contain an miRNA binding site.[13] This is particularly interesting as only ~30 % of genes have a miRNA binding site in their 3'UTR.[195] This is evidence to suggest that m^6A may play a role in the regulation of miRNA induced mRNA silencing. m^6A has recently been shown to play a key role in miRNA biogenesis.[196]

1.1.7 Concluding points

In recent years the study of the epitranscriptome and in particular m^6A has seen significant progress. It is likely that as new and improved methods of mapping RNA modifications become available, there will be a shift from locating these modifications to determining the exact biological roles they play. This will give us a greater understanding of how abnormalities in these modifications lead to disease. Replacing MeRIP-Seq and MiCLIP with more accurate and reliable methods of m^6A detection and quantification will be a major step forward in m^6A research. It is possible that an improved knowledge of RNA modifications will lead to the development of new therapeutic approaches to the treatment of human diseases, potentially through artificially manipulating the levels of these modifications.

The use of CRISPR-Cas9 for m^6A editing will likely become a common procedure in m^6A research. CRISPR-Cas9 can be used for the manipulation of m^6A writers, readers and erasers, and also for editing individual m^6A residues

by delivering m^6A modifying proteins to specific sequences in the mRNA.[156] However, this will first require an accurate and quantifiable method of m^6A detection.[92]

m^6A is the most clearly understood internal modification of mRNA though this does not necessarily mean m^6A is uniquely important. It largely reflects the fact m^6A it is the most prevalent internal modification of mRNA and therefore has been the primary focus of research.[92] The current knowledge of m^6A writers, erasers and readers allows researchers to easily modify the landscape of the m^6A modification which makes experimental design significantly easier. For these reasons research into m^6A has been considerably greater than other epitranscriptomic modifications. It is likely that in the coming years as methods of analysing RNA modifications continue to improve, we will learn more about the less well understood mRNA modifications such as pseudouridine, inosine, m^1A and m^5C and can hopefully learn from the struggles found in researching m^6A. First it will be important to identify the writers, readers and erasers involved in the maintenance of these modifications. This is a rapidly expanding field and it is likely that novel mRNA modifications will continue to be discovered as technology continues to improve the sensitivity and precision of detection.

1.2 Aims and objectives

i. Synthesis of modified oligonucleotides

The Hayes research group has developed a method for the single step synthesis of alkylated nucleoside phosphoramidites.[197] We aim to use this method to synthesise a range of biologically relevant modified nucleoside phosphoramidites to be incorporated into oligonucleotide chains. We will use these modified oligonucleotides for a range of collaborative research projects to study the biological activity of these modifications. These oligonucleotides will also be used to test out the novel methods of m^6A detection described in this thesis.

ii. Develop an improved version of the SCARLET method of m^6A detection

SCARLET[90] has been described as the 'gold standard' of m^6A detection and quantification.[91] Despite this there are very few examples of this method being used in the literature and there are examples of this method producing unusable and unreliable results.[91] We have identified a number of inefficient and unsuitable steps in the SCARLET protocol that we believe can be improved upon. We aim to develop a new method of site-specific detection and quantification based on the SCARLET method. We also aim to use these methods to detect and quantify m^6A at specific sites in the zipcode sequence of the β-actin mRNA in an attempt to elucidate the mechanism by which m^6A regulates the processing of this transcript.

iii. Determine whether the modified nucleoside (PhSeT) can be utilised for m^{6}A detection

Previous work in the Hayes and Fray research groups has shown that a modified nucleoside (PhSeT)[17] selectively crosslinks with adenosine and not *N*^{6}-methyladenosine under oxidative conditions. We aim to test whether this selective crosslink is maintained in a poly nucleotide sequence context and determine whether we can utilise this modified nucleoside for the site-specific detection and quantification of m^{6}A in mRNA.

iv. Develop reverse transcription-based methods of m^{6}A sequencing

The majority of reverse transcriptase enzymes are unable to differentiate between adenosine and m^{6}A.[16] We aim to identify a modified nucleoside triphosphate that stalls reverse transcription specifically at sites of m^{6}A in the RNA template.

Chapter 2: Synthesis of modified oligonucleotides

Overview: Oligonucleotides containing modified nucleotides have a range of biological applications. They can be used to study the functional role of the modification and as biological probes to identify proteins, and nucleic acid sequences that interact with the modification.

In this chapter we describe a one-step synthesis of alkylated phosphoramidites which can be incorporated into synthetic oligonucleotides. We synthesise oligonucleotides modified at their 5' end with either adenosine, A_m, or m^6A_m which will be used to identify proteins involved in the regulation of the 5' cap of mRNA. We synthesise m^6A modified, random sequence oligonucleotides which will be used in a SELEX based assay to identify the sequence specificity of m^6A binding proteins and anti-m^6A antibodies. Finally, we synthesise m^6A modified oligonucleotides corresponding to the zipcode sequence of chicken β-actin mRNA, to determine the effect of m^6A on ZBP protein binding.

2.1 Introduction

2.1.1 Modified oligonucleotides

Oligonucleotides containing modified nucleotides have a range of important biological applications. They can be used as biological probes to identify and study molecules such as nucleic acids and proteins that interact with the modification. m^6A modified oligonucleotides are valuable research tools that can be used for the study of m^6A writer, reader, and eraser proteins, and anti-m^6A antibodies. m^6A modified oligonucleotides can also be used to test newly developed methods of m^6A sequencing. Throughout this thesis a number of chemically modified oligonucleotides are utilised for the study of RNA modifications.

2.1.2 Synthesis of modified RNA oligonucleotides

There are three approaches to synthesising modified RNA oligonucleotides: solid phase chemical synthesis, post synthetic modification of oligonucleotides, and *in vitro* transcription.

2.1.2.1 Solid phase chemical synthesis

Nucleotides are not sufficiently reactive to be used directly in oligonucleotide synthesis and must therefore be modified to increase their reactivity.[21] Nucleoside phosphoramidites are relatively stable molecules which can undergo rapid nucleotide coupling. Nucleoside phosphoramidites contain a 3'-

[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite group that provides the reactivity required for nucleoside coupling. The nucleosides are also modified with protecting groups to eliminate undesirable side reactions. Nucleoside phosphoramidites are protected at their 5’ hydroxyl group with 4,4'-dimethoxytrityl (DMTr), the exocyclic amino groups of the bases are also protected, and RNA phosphoramidites are protected at their 2’ hydroxyl group with TBDMS (t-butyldimethylsilyl) (Figure 20 and Figure 21).[21]

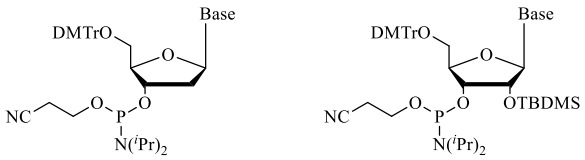
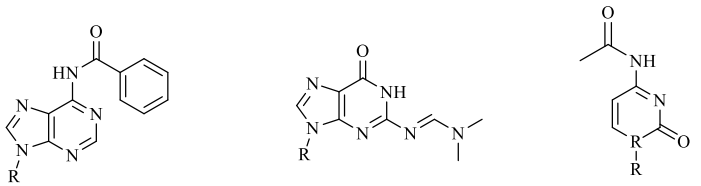


Figure 20: The structure of DNA (left) and RNA (right) phosphoramidites. The 5' hydroxyl group is protected with DMTr, the 3' position is modified with a cyanoethyl group, and the 2' of RNA phosphoramidites is protected with a TBDMS group.[21]



N-benzoyl A N-dimetylformamidyl G N-acetyl C

Figure 21: The protecting groups of A, G, and C RNA phosphoramidites.[21]

Oligonucleotides are synthesised by the sequential coupling of nucleoside phosphoramidites in the 3’ to 5’ direction whilst bound to a solid support at the 3’ end. Each coupling step involves 4 successive reactions, detritylation, coupling, capping, and oxidation (Figure 22).[21]

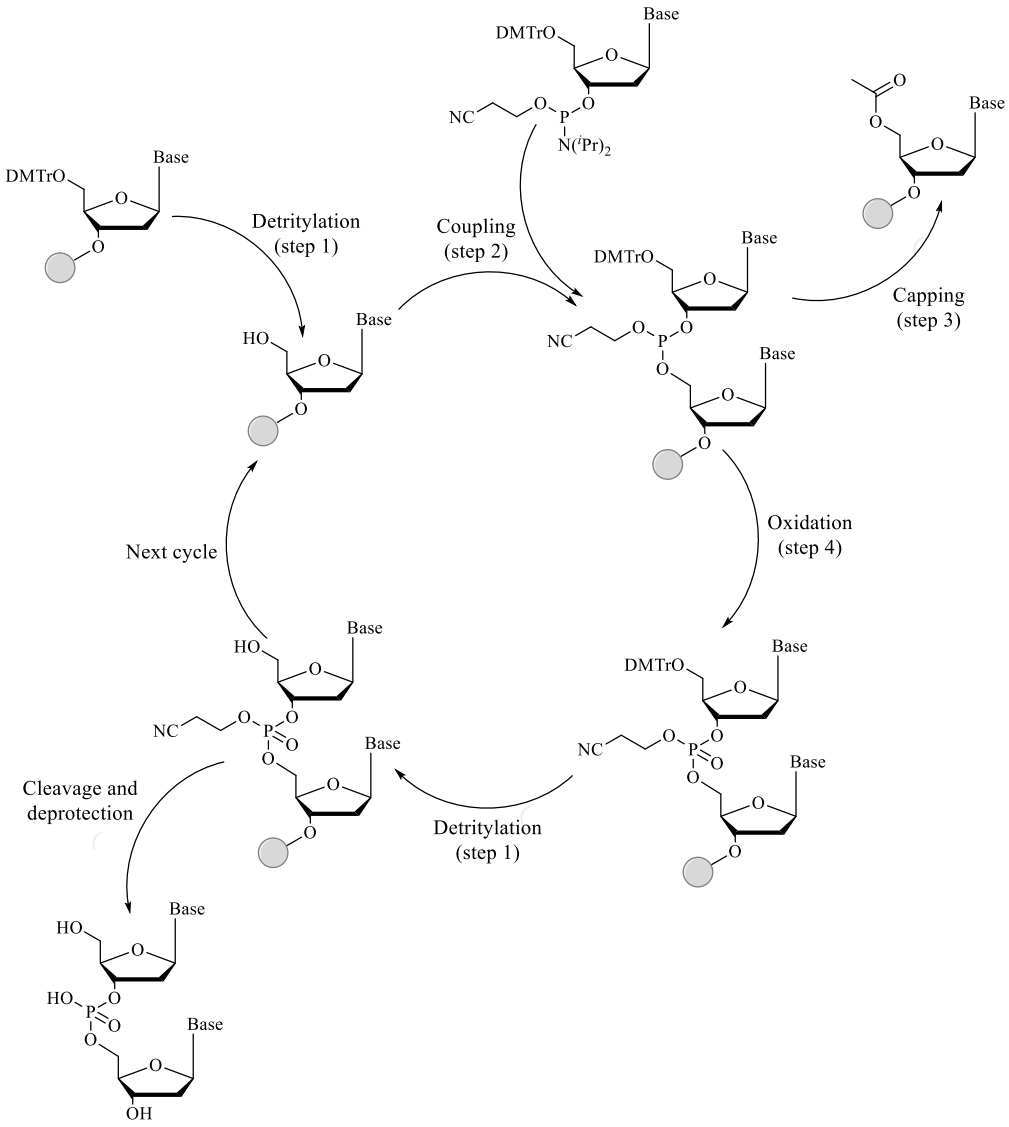


Figure 22: Oligonucleotide synthesis cycle.[21]

The first step is the removal of the 5’ DMTr group from the support bound phosphoramidite using trichloroacetic acid. This yields a 5'-terminal hydroxyl group that is ready to couple with the incoming phosphoramidite. The nucleoside phosphoramidite to be coupled to the oligonucleotide chain is activated by a tetrazole catalyst which protonates the diisopropylamino group which is displaced by the 5’ hydroxyl group of the support bound oligonucleotide.[21]

Yields of the coupling reaction are very high, however, they are rarely quantitative. This means that a small number of unreacted 5' hydroxyl groups are still available which must be capped to prevent them from taking part in subsequent coupling reactions. This is essential in order to eliminate the occurrence of deletion mutations in the oligonucleotide product. The capping step involves acetylation of the free 5' hydroxyl groups using acetic anhydride and *N*-methylimidazole. The coupling reaction produces an unstable phosphite triester (P(III)) which is converted to the more stable (P(v)) species by an iodine oxidation step.[21]

Modified phosphoramidites can be synthesised that are compatible with oligonucleotide synthesis chemistry.[21] m^6A modified oligonucleotides can be synthesised using an m^6A phosphoramidite (Figure 23).[198]

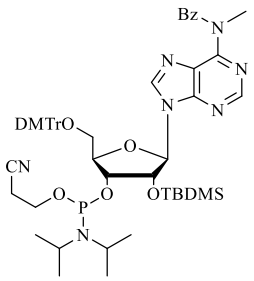
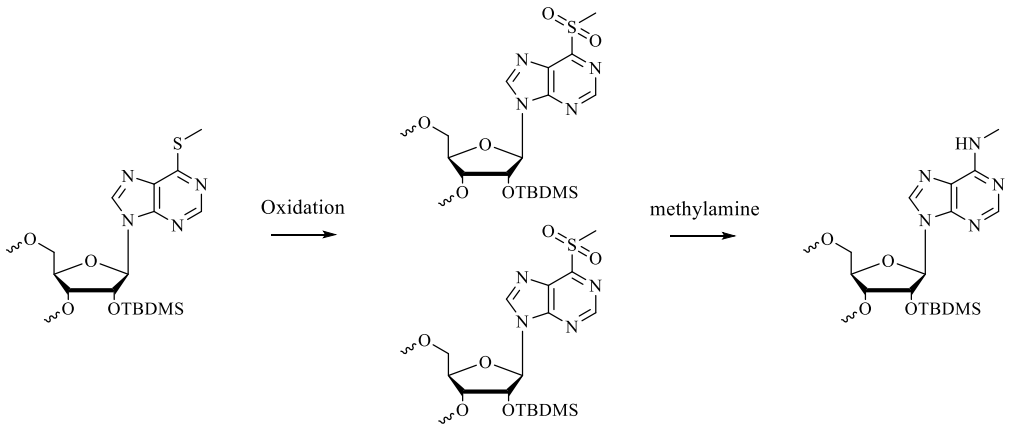


Figure 23: Structure of m^6A RNA phosphoramidite.[198]

2.1.2.2 Post synthetic oligonucleotide modifications

Certain nucleotide modifications are incompatible with solid phase oligonucleotide synthesis, and therefore the oligonucleotides must be modified after synthesis. It is possible to synthesise m^6A modified oligonucleotides by post synthetic modification.[199] The oligonucleotide is synthesised with a 6-

methylthiopurine nucleotide at the desired position which is oxidised using the magnesium salt of monoperoxyphtalic acid and subsequently reacted with methylamine to produce the m^{6}A modified oligonucleotides (Scheme 1).[199]



Scheme 1: Synthesis of m^{6}A modified oligonucleotides by post synthetic modification.[199]

2.1.2.3 *In vitro* transcription

In vitro transcription utilises a polymerase enzyme to synthesise an RNA molecule using a DNA template.[200] The DNA template contains a double stranded promoter region to which the RNA polymerase binds, and a single stranded region which determines the sequence of the RNA to be synthesised (Figure 24). The RNA polymerase uses ribonucleoside triphosphates (rNTPs) to synthesise the RNA molecule. Modified ribonucleoside triphosphates such as m^{6}ATP can be used in place of the unmodified nucleoside triphosphates and incorporated into the synthetic RNA.[200,201]

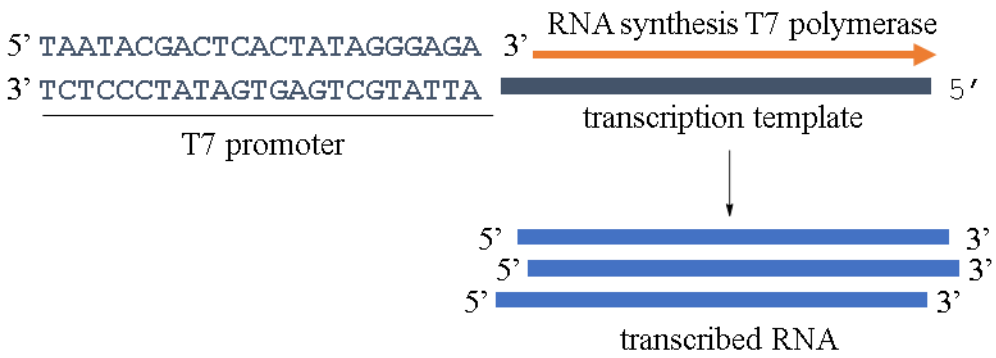


Figure 24: *In vitro* transcription. T7 RNA polymerase binds to a double stranded promoter region and transcribes and RNA template complementary to the single stranded region.[200]

2.1.2.4 Comparison of synthetic methods

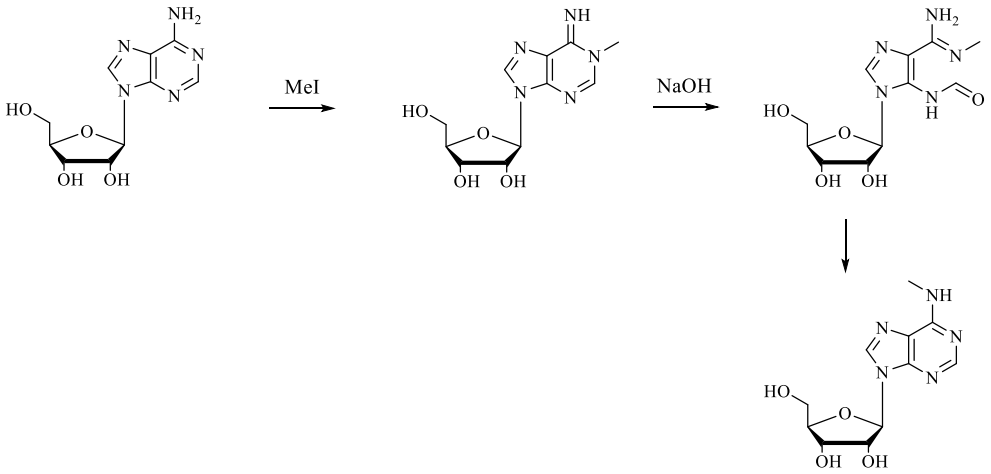
The advantages and disadvantages of methods to synthesise modified RNA oligonucleotides are summarised in Table 1. Solid phase chemical synthesis is the most suitable method for synthesising m^6A modified RNA as it allows the synthesis of oligonucleotides of a defined length and sequence, containing any number of m^6A nucleotides, which is not possible using other methods. However, this method is limited to short oligonucleotide sequences, and the synthesis is time consuming due to the slow coupling times of RNA phosphoramidites.

Table 1: Comparison of methods to synthesise modified RNA oligonucleotides.

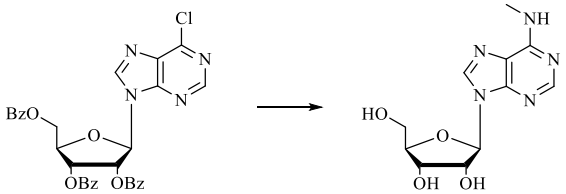
Method	**Advantages**	**Disadvantages**
Solid phase chemical synthesis	• Compatible with wide range of modifications • Low error rate	• Length < 50 nucleotides • Time consuming synthesis and purification
Post synthetic modification	• Difficult purification	• Time consuming • Not compatible with incorporating multiple different modifications
In vitro transcription	• Long sequences • Fast synthesis	• Few modifications are compatible • High error rate[202] • Single nucleotide positions cannot be modified

2.1.3 m^{6}A phosphoramidite synthesis

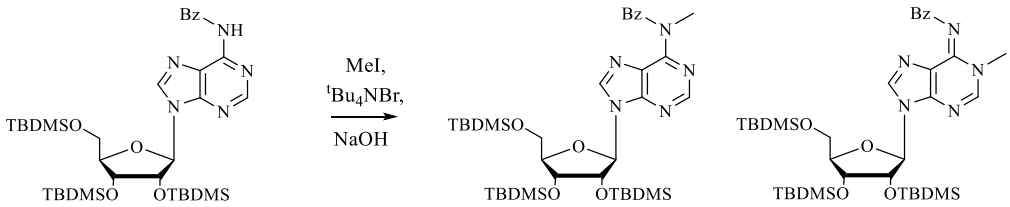
N^6-methyladenosine can be made using a number of synthetic approaches. One method involves methylation of the *N*^{1} position of the adenosine nucleoside followed by an alkali mediated Dimroth rearrangement to produce m^{6}A (Scheme 2).[203–205] Another involves the nucleophilic aromatic displacement of a 6-substituted adenosine derivative (Scheme 3).[198,206] A third option is to use Aritomo's phase transfer catalysis method to methylate a protected adenosine nucleoside (Scheme 4).[207]



Scheme 2: Synthesis of N^6-methyladenosine using the Dimroth rearrangement method.[203–205]

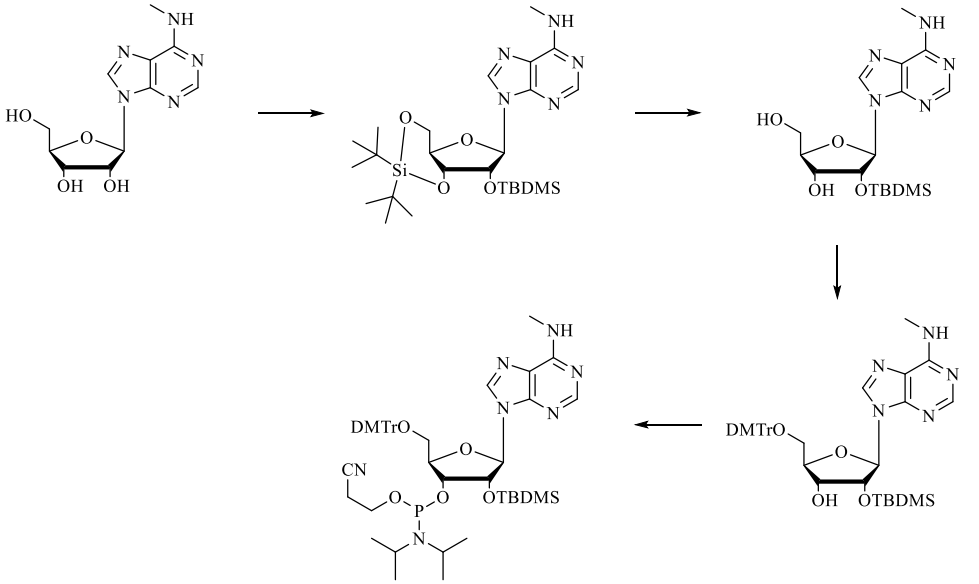


Scheme 3: Synthesis of N^6-methyladenosine by nucleophilic aromatic substitution.[198,206]



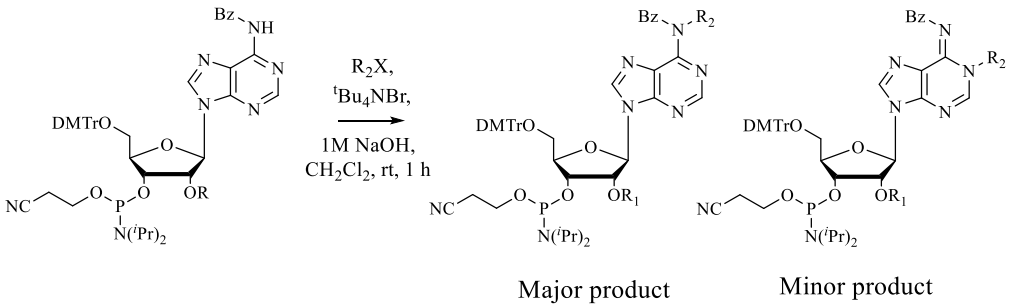
Scheme 4: Synthesis of N^6-methyladenosine using Aritomo's phase transfer catalysis method.[207]

In order to be compatible with solid phase chemical synthesis, the m^6A nucleoside must then be converted into a phosphoramidite, requiring four additional synthetic steps (Scheme 5).[198]



Scheme 5: Synthesis of the m^{6}A phosphoramidite from the m^{6}A nucleoside.[198]

Thankfully it is possible to synthesise the m^{6}A phosphoramidite directly from the unmodified adenosine phosphoramidite which is commercially available and relatively inexpensive. This is achieved by direct alkylation of the adenosine phosphoramidite using Aritomo's phase transfer catalysis method (Scheme 6).[197,207] This results in high yields of the phosphoramidite product as well as the minor product of the N^1-methyladenosine phosphoramidite. In this chapter we use this chemistry to synthesise the m^{6}A phosphoramidite along with a range of other alkylated RNA phosphoramidites.



Scheme 6: Adenosine alkylation using Aritomo's phase transfer catalysis.[207]

2.1.4 i^6A modified tRNA

tRNA molecules function in the decoding of mRNA. tRNA molecules recognise a three nucleotide codon sequences in mRNA using their complementary anticodon sequence. This determines which amino acid is added to the growing polypeptide chain. tRNA molecules have a distinctive three-dimensional structure containing three hairpin loops, referred to as its 'cloverleaf structure'. The 'anticodon loop' contains the three-nucleotide anticodon (Figure 25).[20]

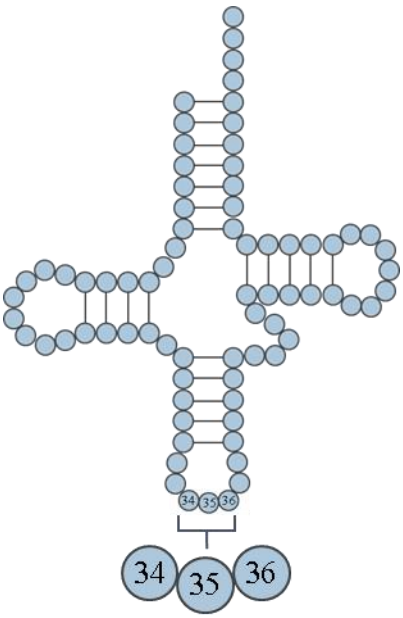
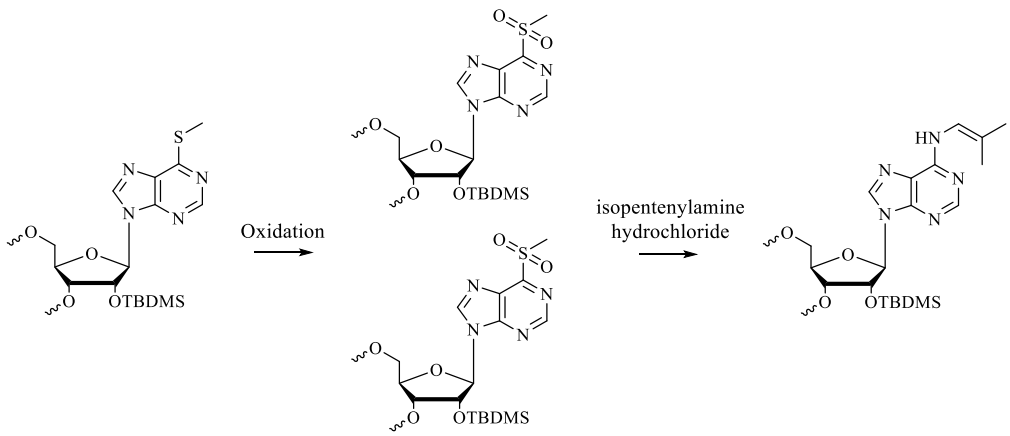


Figure 25: The tRNA cloverleaf structure. The anticodon loop consists of nucleotides 34, 35, and 36.[20]

The majority of RNA modifications reside in tRNA. These modifications are required for the specific three-dimensional structure of the RNA that is essential for its function.[22] One of the most common tRNA modifications is at position 37 which is the nucleotide immediately 3'- to the anticodon.

Modifications at position 37 stabilise base pairing between the codon and anticodon, preventing frameshift mutations.[208] The most common modifications at position 37 are t^6A and m^1G. i^6A is also commonly found at this position.[209] Mammalian tRNA-Sec-TCA contains the i^6A modification at position 37.[210]

The synthesis of i^6A modified oligonucleotides is challenging, however it is possible to synthesise these molecules by post synthesis modification. The oligonucleotide is synthesised with a 6-methylthiopurine nucleotide at the desired position and is oxidised using the magnesium salt of monoperoxyphtalic acid. The oligonucleotide is finally reacted with isopentenylamine hydrochloride to produce the i^6A modified product (Scheme 7).[199]



Scheme 7: Post synthetic modification method for the synthesis of i^6A modified oligonucleotides.[199]

A simpler method of synthesis would be to use an i^6A phosphoramidite, however, the double bond in the isopentyl group may be sensitive to the repeated oxidation cycles of oligonucleotide synthesis. In this chapter we

synthesise an i^6A nucleoside phosphoramidite which will be used to test whether this molecule is compatible with solid phase oligonucleotide synthesis chemistry.

2.1.5 Random sequence oligonucleotides

Systematic evolution of ligands by exponential enrichment (SELEX) is a method of identifying oligonucleotides that interact with a specific ligand, for example an RNA binding protein.[211] SELEX requires the generation of a random sequence oligonucleotide library. The oligonucleotides are composed of an internal region consisting of a random nucleotide sequence, flanked by conserved sequences at the 5'- and 3'- end to allow sequencing primers to bind (Figure 26). The RNA library is exposed to the ligand of interest. Sequences that bind to the ligand are enriched whilst non bound oligonucleotides are washed away. The oligonucleotides that bind to the ligand can be identified by RNA/DNA sequencing.[211–213]

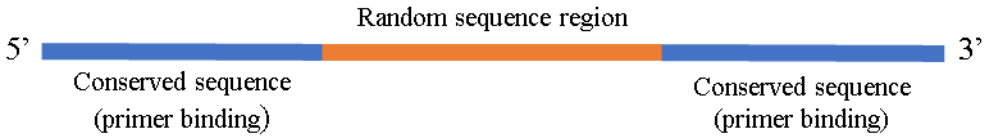


Figure 26: Structure of a random sequence oligonucleotide for SELEX assays.

The most common application of SELEX is to generate aptamers, which are single stranded oligonucleotides that bind to, and inhibit specific proteins for therapeutic applications.[211] Random sequence oligonucleotide libraries can also be used to identify the sequence binding motifs of RNA binding proteins.

Zhang, et al., (2010) used SELEX to determine the RNA sequence preference of the YT521-B domain. They found that the protein targeted a degenerate consensus sequence of 5′-NGANNN-3′.[214,215] Random sequence oligonucleotides containing the m^6A modification have the potential to be used to study the sequence preferences of m^6A binding proteins and anti-m^6A antibodies.

2.1.6 m^6A methylation of β-actin mRNA

2.1.6.1 Localised translation

mRNAs are typically translated into protein in the cytoplasm. The protein is then transported to wherever it is required in the cell. In certain cases, instead of transporting the protein, the mRNA transcript is targeted to the site at which the protein is required. This means the protein is located at its site of action immediately after synthesis. A single mRNA transcript can be the template for many copies of the protein, making it more energetically favourable to transport the mRNA molecule instead of a large number of proteins. Localised translation allows the cell to rapidly alter its protein output in response to its environment and reduces the risk of proteins accumulating in the wrong cellular compartment.[20]

2.1.6.2 β-actin mRNA undergoes localised translation

Actin is a major component of the cytoskeleton which is essential for the structural integrity of the cell along with numerous cellular processes.[216] β-actin mRNA is the best understood example of an mRNA that undergoes localised translation.[217]

Cell crawling is a mechanism of cell movement that is achieved by the continuous remodelling of the actin cytoskeleton. During cell crawling, β-actin mRNA is localised to the front of cells at the 'leading edge'. New actin filaments are then rapidly polymerised, extending the membrane outwards in the direction of travel, forming protrusions called filopodia (rod shaped) and lamellipodia (sheet like). Crawling cells require approximately 2500 β-actin proteins at their leading edge in order to facilitate movement.[217]

In situ hybridisation experiments have shown that in highly motile cells such as myoblasts and fibroblasts β-actin mRNAs accumulate at the cell periphery. This allows the cell to economically synthesise the high levels of β-actin required for cell motility.[218]

2.1.6.3 VICKZ proteins and the mRNA zipcode

Localised RNA molecules contain specific regions within their 3' UTR that target them for transport. These sequence motifs are known as 'zipcodes' and are specifically recognised by RNA binding proteins which function in the trafficking of mRNA transcripts. Zipcode motifs vary considerably in both sequence and length.[20] VICKZ are a highly conserved family of RNA binding

proteins that recognise the RNA zipcode and facilitate transport of the mRNA transcript.[219] Humans have three VICKZ proteins (IMP1, IMP2, IMP3),[220] while Mice (CRD-BP),[221] Chickens (ZBP),[222] and Xenopus (Vg1-RBP/VERA)[223,224] each have one. The VICKZ proteins interact with a large number of mRNA transcripts[225–229], which fulfil diverse roles in transport, stability and translation of the mRNA.[76,227,230–232]

The mechanism by which the VICKZ proteins interact with mRNA is poorly understood. The best studied example is β-actin mRNA. Two sequence elements within the chicken β-actin zipcode are essential for ZBP binding, GGACU, and ACA.[233] ZBP interacts with the β-actin transcript using its two C-terminal KH domains (KH3 and KH4). The two domains interact to form a pseudo-dimer with their RNA binding motifs located on the outer edge. The mRNA wraps around the outside of these protein domains with the KH3 domain binding to the CGGACU sequence, and the KH4 domain interacting with ACA (Figure 27).[232] The RRM domain of ZBP subsequently interacts with the molecular motor protein KIF11, which transports the ribonucleoprotein complex to its site of action.[234]

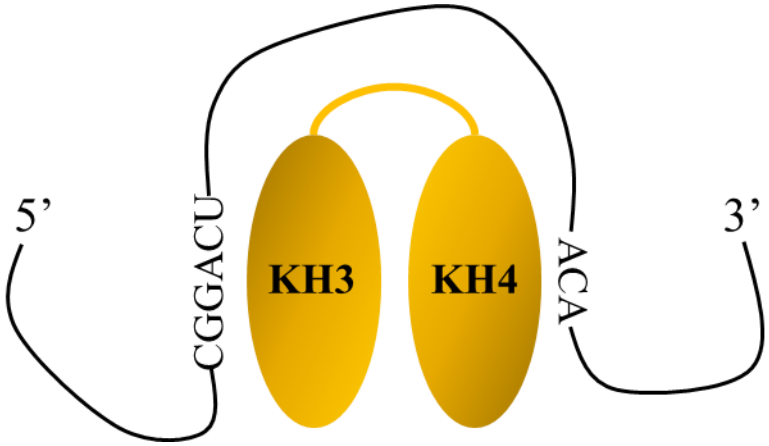


Figure 27: ZBP binding to β-actin mRNA. The KH3 domain binds to the CGGACU sequence, and the KH4 domain binds to ACA.[232]

A recent study has identified IMP1, IMP2, and IMP3 as a novel class of m^6A reader proteins in humans.[157] The IMP proteins increase the stability of m^6A containing transcripts, resulting in increased levels of transport and translation. Thousands of high confidence methylated targets of the IMP proteins have been identified, including MYC, β-actin and LIN28B11.[157]

A number of m^6A sequencing studies have identified m^6A within the β-actin zipcode,[12–14] however the exact location of the modification is unknown. We hypothesise that the β-actin zipcode is methylated within the VICKZ binding elements ‘GGACU’, and ‘ACA’ and the m^6A modification is regulating VICKZ protein binding. Using an *in vitro* pull-down assay, we demonstrate that m^6A methylation of the chicken zipcode ‘GGACU’ site promotes VICKZ protein binding.

2.2 Aims

i. Synthesis of alkylated phosphoramidites

To synthesise a range of alkylated adenosine phosphoramidites using the one step method described by Kruse et al., (2011).[197] These phosphoramidites will be incorporated into oligonucleotides for various biological applications.

ii. Synthesis of oligonucleotides containing modified 5' nucleotides

To synthesise 3'- biotinylated oligonucleotides containing either adenosine, A_m, or m^6A_m at their 5' end. These oligonucleotides will be used for the identification of proteins that interact with different cap adjacent modifications and will be used to test a new method of sequencing the 5' cap structure of mRNA molecules.

iii. Synthesis of random sequence oligonucleotides

To synthesise a set of random sequence oligonucleotides either containing or lacking m^6A. These will be used in a SELEX based assay for identifying the sequence preference of m^6A binding proteins and anti-m^6A antibodies.

iv. Synthesise m^6A modified chicken actin oligonucleotides for ZBP binding analysis

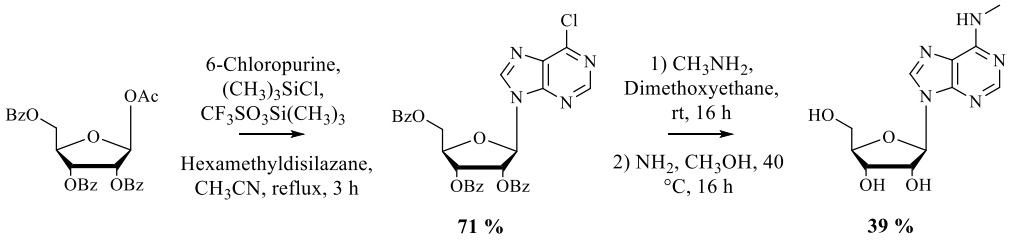
To synthesise oligonucleotides corresponding to the chicken β-actin zipcode sequence containing the m^6A modification at various positions. These

oligonucleotides would be used in a gel shift assay to determine the effect of the m^6A modification on ZBP binding *in vitro*.

2.3 Results and discussion

2.3.1 Synthesis of the adenosine nucleoside

We synthesised N^6-methyladenosine using the nucleophilic aromatic displacement method (described in section 2.1.3).[198,206] Firstly, (2R,3R,4R)-2-((benzoyloxy)methyl)-5-(6-chloro-9H-purin-9-yl)tetrahydrofuran-3,4-diyl dibenzoate was synthesised in a 71 % yield and subsequently reacted with aqueous methylamine to methylate the N^6 position. The nucleoside was then deprotected using methanolic ammonia to yield N^6-methyladenosine in a 39 % yield (Scheme 8).



Scheme 8: Synthesis of N^6-methyladenosine.[198]

2.3.2 Alkylation of adenosine phosphoramidite

Kruse et al., (2012) demonstrated that Aritomo's phase transfer catalysis method can be used to methylate the 2'-O-Me adenosine phosphoramidites to produce both m^6A_m and m^1A_m. [197,207] Using this method with methyl iodide as

the methyl donor, we synthesised m^6A, m^1A, m^6A_m and m^1A_m phosphoramidites (Figure 28).

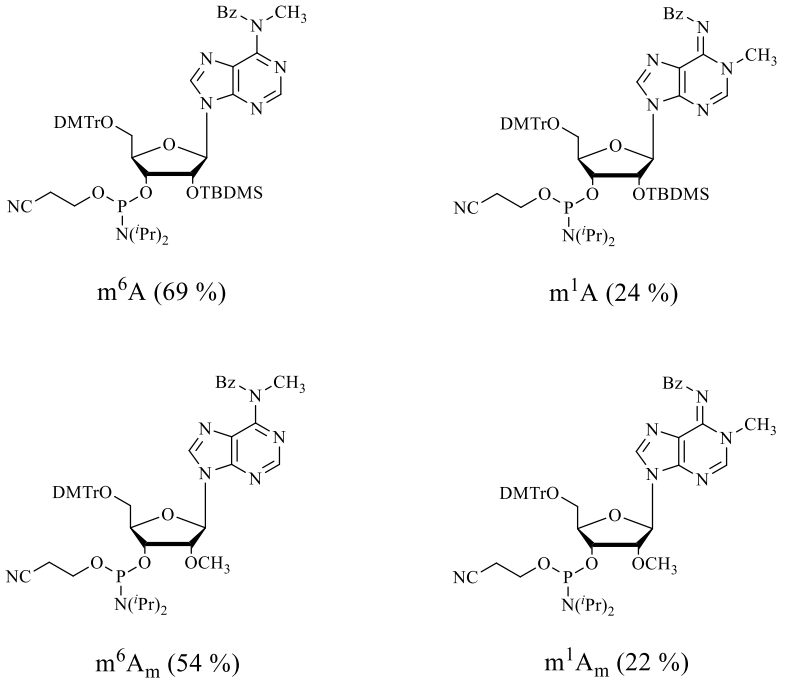


Figure 28: Synthesis of methylated phosphoramidites: m^6A, m^1A, m^6A_m, and m^1A_m.

2.3.3 Prenylation of the adenosine phosphoramidite

The phosphoramidite phase transfer catalysis method is also compatible with other alkylations. To demonstrate this, we synthesised i^6A, i^1A, i^6A_m and i^1A_m phosphoramidites (Figure 29). i^6A has only been identified in tRNA whist i^1A, i^6A_m, and i^1A_m have not yet been identified in RNA,[49] however new RNA modifications are regularly discovered as the technology to analyse RNA modifications improves. It is therefore important to be able to synthesise oligonucleotides containing these modifications. Future work will involve

testing whether the i^6A nucleoside phosphoramidite is compatible with solid phase oligonucleotide synthesis chemistry.

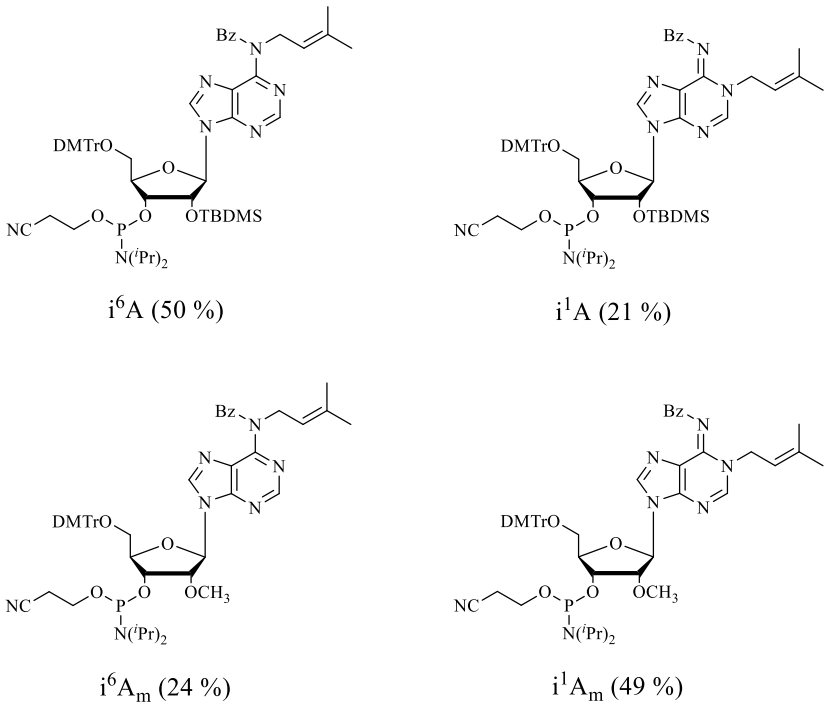


Figure 29: Prenylation of adenosine phosphoramidites: i^6A, i^1A, i^6A_m, and i^1A_m.

2.3.4 Synthesis of m^6A_m, and A_m modified oligonucleotides

As part of a collaborative research project with Eleanor Bellows (Fray research group) we synthesised oligonucleotides containing biotin at their 3' end and either adenosine, A_m, or m^6A_m at their 5' end (Table 2). Initially we aimed to use these oligonucleotides in a protein pull down experiment to identify the m^6A_m methyltransferase enzyme, however, soon after their synthesis, three research groups independently identified PCIF1 as the mammalian m^6A_m methyltransferase.[57–59]

Table 2: Oligonucleotides synthesised containing different 5' modified adenosine nucleosides.

Name	**DNA/RNA**	**Sequence (5'-3')**	**Length (nt)**
CAP-A	RNA	AACUGAUCAUCAUCACCA CAGAGCAGGU-biotin	28
CAP-A_m	RNA	(Am)ACUGAUCAUCAUCA CCACAGAGCAGGU-biotin	28
CAP-m^6A_m	RNA	(m^6A_m)ACUGAUCAUCAUC ACCACAGAGCAGGU-biotin	28

Instead, we will use these oligonucleotides to test our newly developed method for the site-specific detection and quantification of the first nucleotide adjacent to the 5' cap (Figure 30). This method involves de-capping the mRNA and radiolabelling the first nucleotide at the 5' end of the RNA. The RNA is then splint ligated to a single stranded DNA molecule (ssDNA). RNase treatment then removes any RNA that is not ligated to the ssDNA and thin layer chromatography is used to identify the nucleotide adjacent to the 5' cap.

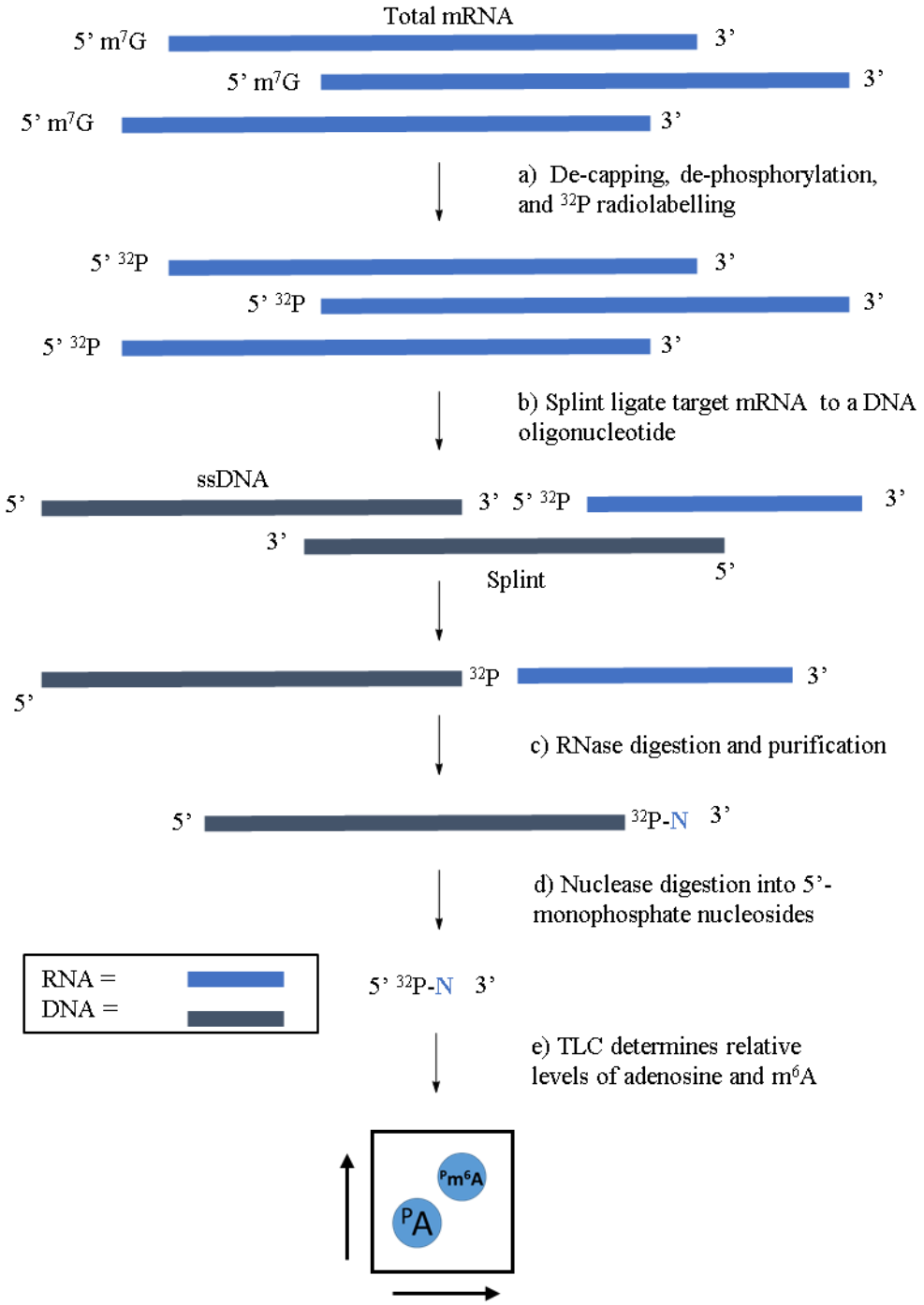


Figure 30: Method for the analysis of the first nucleotide at the 5' of specific mRNA molecules: a) The mRNA is de-capped, dephosphorylated and radiolabelled at its 5'- end with ^{32}P b) The mRNA is splint ligated at its 5'- end to a ssDNA molecule c) The nucleic acid is treated with RNase T1 and RNase A which degrades all of the RNA except for 5' most nucleotide which is

protected by the ligated ssDNA d) The DNA/RNA is digested into 5' nucleoside monophosphates e) and the identification of the 5' nucleotide is analysed by TLC.

For this method to work, the splint ligation reaction must be able to efficiently ligate the ssDNA to A_m and m^6A_m modified RNA. At the time of writing this thesis, the oligonucleotides in Table 2 are being used to explore this ligation reaction, in order to identify a ligase enzyme that can efficiently join single stranded DNA to modified RNA. These experiments are being carried out by Eleanor Bellows.

2.3.5 Synthesis of m^6A modified random sequence oligonucleotides

As part of a collaborative research project within the Fray research group, we synthesised a number of m^6A modified random sequence oligonucleotides (Table 3). These oligonucleotides are currently being used by Cameron Grundy (Fray research group) to determine the consensus binding motifs of m^6A binding proteins and anti-m^6A antibodies. In order to obtain an equal ratio of the four nucleotides at each position, an optimised ratio of the four nucleoside phosphoramidites must be used, based upon the varying coupling efficiencies of these molecules. We used a phosphoramidite ratio of (A/C/G/U = 0.30:0.25:0.23:0.20) to compensate for the slower coupling times of adenosine and cytosine.[235]

Table 3: Random sequence oligonucleotides for analysis of the sequence specificity of m^6A antibodies and m^6A binding proteins (N = random nucleotide of A/G/C/U).

Name	**DNA/RNA**	**Sequence (5'-3')**	**Length (nt)**
1agAgc	RNA	GUGAGUCUUCNNNNANNNNUCCUGCAGCG	28
2ga6cg	RNA	GUGGAUCUUCNNNN(m^6A)NNNNNUCCUCGAGCG	28
3ag6gc	RNA	GUGAGUCUUCNNNN(m^6A)NNNNNUCCUGCAGCG	28
4gaAcg	RNA	GUGGAUCUUCNNNNANNNNUCCUCGAGCG	28
5N9	RNA	GUGAGUCUUCNNNNNNNNNUCCUGCAGCG	28

These oligonucleotides contain a random sequence region consisting of nine nucleotides with a single adenosine, m^6A, or random nucleotide at the central position. Nine nucleotides will be sufficient for these experiments as m^6A binding proteins have been shown to target short sequence motifs. Nine random nucleotides will result in a pool of approximately (4^9) 262,000 unique oligonucleotides.

The random sequence region is flanked at both ends with a barcode region for the sequencing primers to bind to. Due to the possibility that the ligand may interact with the 5'- or 3'- primer regions the experiment will be repeated

using two different primer sequences. The 5'- primers are either 'gugagucuuc' or 'guggaucuuc' and the 3'- primer are either 'uccugcagcg' or 'uccucgagcg'.

The use of two different barcode sequences will allow us to undertake this *in vitro* binding assay using a mixture of m^6A and adenosine RNA oligonucleotide pools which will directly compete with each other. The presence of one barcode sequence will indicate binding to an unmodified oligonucleotide whilst the other barcode will indicate binding to an m^6A modified oligonucleotide.

In July 2019, Arguello and colleagues published their work on a similar study.[236] They synthesised random sequence m^6A modified oligonucleotides to study the sequence binding motifs of the YTH proteins. They found that YTHDC1 selectively binds to m^6A within the DRACH consensus motif whereas YTHDF1 and YTHDF2 bind to m^6A independently of its consensus motif.[236]

The study by Arguello et al., tested a limited subset of m^6A reader proteins and also did not test the binding sites of anti-m^6A antibodies. One limitation of their study that has been improved upon using our random sequence oligonucleotides is that they did not use different barcode sequence for the m^6A and unmodified oligonucleotides. Therefore, they were unable to evaluate the binding motifs of these proteins when in the presence of both m^6A modified and unmodified oligonucleotides.

2.3.6 m^6A methylation of the β-actin transcript affects ZBP binding *in vitro*

Previous unpublished work in the Fray research group used MeRIP-Seq to analyse m^6A levels in the chicken embryo transcriptome. In this study, two m^6A peaks were identified in the β-actin mRNA, one of which was located in the 3' UTR close to the region of the zipcode sequence.[237] This is consistent with published m^6A sequencing results. [12–14] Two sequence elements in the Chicken β-actin zipcode are essential for ZBP binding, GGACU, and ACA.[233] The human homologs of ZBP (IGF1, IGF2, and IGF3) are m^6A readers[157] and the presence of an m^6A peak over the β-actin zipcode indicates that m^6A may regulate the interaction between ZBP and the β-actin transcript.

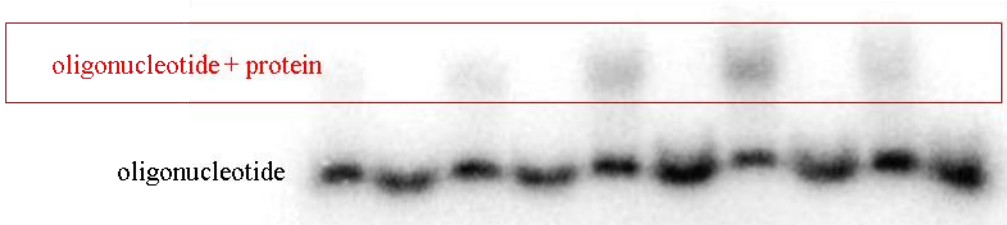
As part of a collaborative research project within the Fray research group we synthesised oligonucleotides corresponding to a 28 nt region of the β-actin zipcode containing m^6A at various nucleotide positions (Table 4). The same oligonucleotide sequence was used in a gel shift experiment by Chao and colleagues in 2012 when they identified the GGACU, and ACA sequences as essential for ZBP binding.[226]

Table 4: Chicken β-actin oligonucleotides containing m^6A at various positions. These oligonucleotides were used by Zsuzsanna Bodi in a gel shift assay to determine the effect of m^6A on VICKZ protein binding.

Name	**DNA/RNA**	**Sequence (5'-3')**	**Length (nt)**
Chicken β-actin zipcode unmodified	RNA	ACCGGACUGUUACCAACACCCACACCCC	28
Chicken β-actin zipcode m^6A (site 1202)	RNA	ACCGG(m^6A)CUGUUACCAACACCCACACCCC	28
Chicken β-actin zipcode m^6A (site 1211)	RNA	ACCGGACUGUUACCA(m^6A)CACCCACACCCC	28
Chicken β-actin zipcode m^6A (site 1212)	RNA	ACCGGACUGUUACC(m^6A)ACACCCACACCCC	28
Chicken β-actin zipcode m^6A (site 1218)	RNA	ACCGGACUGUUACCAACACCC(m^6A)CACCCC	28

A polyacrylamide gel electrophoretic mobility shift assay (EMSA) (undertaken by Zsuzsanna Bodi) was used to determine the affinity between the radiolabelled RNA and a recombinant ZBP protein (expressed and purified by Eleanor Bellows) (Figure 31). Consistent with results of Chao and colleagues in 2012, we found that ZBP binds to the unmodified transcript.[233] m^6A methylation of nucleotides 1211 and 1212 of the zipcode fragment had

no effect on ZBP binding. However, m^6A at position 1202 of the fragment significantly increased ZBP binding, whilst m^6A at position 1218 inhibited ZBP binding.



ZBP protein	+	-	+	-	+	-	+	-	+	-
RNA	m^6A 1218		m^6A 1212		m^6A 1211		m^6A 1202		A	
Protein bound oligonucleotide (%)	7	0	16	0	29	0	42	0	21	0

Figure 31: Gel shift assay of β-actin mRNA and ZBP. The RNA/protein complexes were resolved on a 5 % PAGE gel. The percentage of protein bound oligonucleotide was calculated using the ratio of protein bound oligonucleotide/ unbound oligonucleotide. m^6A at position 1202 promoted ZBP binding whilst m^6A at position 1218 inhibited m^6A binding. This experiment was undertaken by Zsuzsanna Bodi.

Positions 1202 and 1218 are located in the sequence motifs 'GGACU' and 'ACA' respectively that are essential for ZBP binding.[233] The 'GGACU' sequence fits perfectly with the m^6A DRACH motif, and the presence of an m^6A peak over this region makes this a likely site of m^6A methylation. As MeRIP-Seq peaks can account for multiple m^6A residues, it is possible that positions 1202 and 1218 of the β-actin zipcode are both m^6A methylated.

2.3.7 Concluding points

The aim of this chapter was to synthesise various modified nucleoside phosphoramidites and incorporate them into oligonucleotides. We have synthesised the phosphoramidites m^6A, m^1A, m^6A_m, m^1A_m, i^6A, i^1A, i^6A_m, and i^1A_m. We have also synthesised a number of modified oligonucleotides to be used in collaborative research projects.

Chapter 3: RedBaron detection of m^6A

Overview: To date, one of the most promising methods of m^6A sequencing is a method called SCARLET, which accurately detects and quantifies specific m^6A sites at a single nucleotide resolution.[90] SCARLET is an invaluable tool for m^6A research; however, we have identified a number of problems with the method that limit its accuracy, efficiency, and sensitivity.

In this chapter we present RedBaron, a novel method for the site-specific detection and quantification of m^6A. Both SCARLET and RedBaron involve radiolabelling of an adenosine/m^6A candidate site followed by analysis using thin layer chromatography (TLC). We tested the RedBaron method on synthetic oligonucleotides containing either adenosine or m^6A within the m^6A consensus sequence. This demonstrated that the method can be used to accurately quantify m^6A levels at a single nucleotide resolution.

We used the RedBaron and SCARLET methods to detect and quantify the m^6A levels of the mRNA from various organisms and tissues. Using the RedBaron method we observed m^6A stoichiometries that were significantly lower than expected. Unfortunately, we found that the presence of non-target mRNA causes the RedBaron method to underrepresent m^6A levels and for this reason, in its present format, it is unsuitable as a method of m^6A quantification.

3.1 Introduction

3.1.1 Targeted RNA cleavage using RNase H

Restriction endonuclease enzymes provide a means for the targeted, sequence dependent cleavage of DNA.[238,239] Analogous methods of targeted RNA cleavage have been historically limited, making research and manipulation of RNA particularly difficult. However, since 1987, the enzyme RNase H has been utilised for accurate, site specific cleavage of RNA.[240] RNase H is an endoribonuclease that specifically digests the RNA strand of a DNA/RNA hybrid.[241] The enzyme hydrolytically cleaves oxygen-phosphorus bonds in the RNA strand, yielding 5'- phosphate, and 3'- hydroxyl polynucleotide products (Figure 32).[241] Double stranded RNA, and RNA hybridised to 2'-OMe modified RNA are both resistant to the RNase H enzyme.[240]

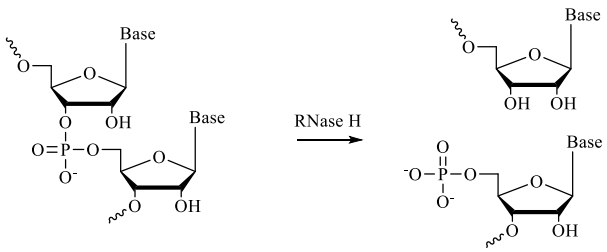


Figure 32: RNase H cleavage: RNase H hydrolytically cleaves the oxygen-phosphorus bonds of the RNA strand of a DNA/RNA hybrid, yielding 5'-phosphate and 3'- hydroxyl polynucleotide products.[241]

A minimum of 4 nucleotides of hybridised DNA/RNA are required for RNA cleavage.[242] Using a 2'-OMe-RNA/DNA chimeric oligonucleotide containing

exactly 4 DNA nucleotides as a guide, RNase H can be programmed to make highly specific cleavages at sites within the mRNA.[240,242] RNase H will cleave the RNA at the 3'- of the nucleotide that is base paired to the DNA nucleotide nearest to the 5'- end of the chimeric oligonucleotide (Figure 33).[242]

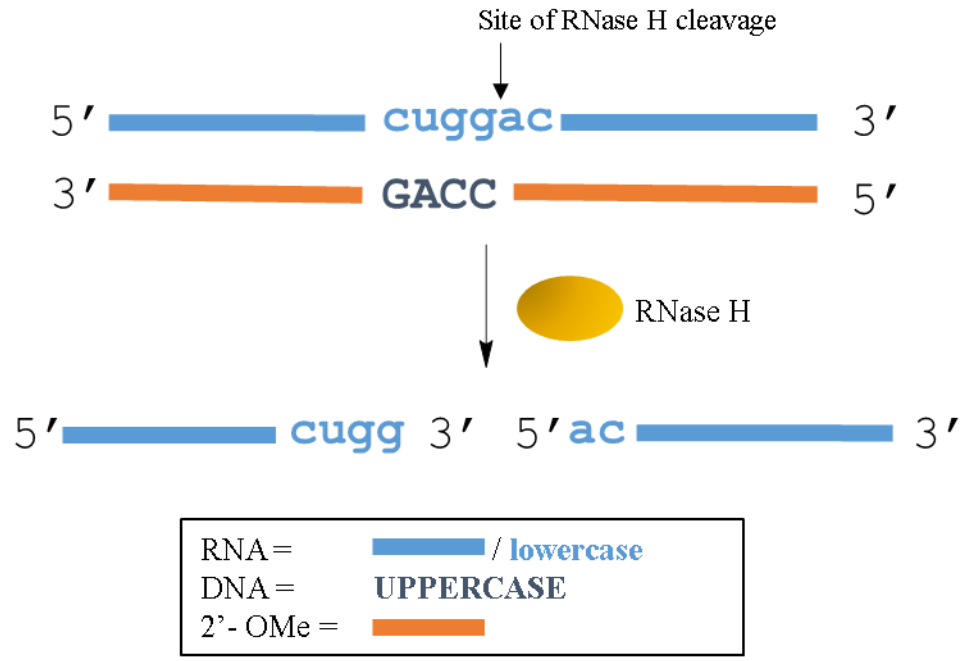


Figure 33: Targeted RNase H cleavage. In this example RNase H will cleave the oxygen-phosphorus bond between the g and a of the RNA.[242]

The specificity of this targeted RNase H cleavage has been utilised for the detection and quantification of a number of modified nucleotides in RNA; the earliest example of which is the detection of 2'-OMe modified nucleosides. This was achieved by identifying sites that are resistant to RNase H cleavage.[243] It is worth noting that different RNase H enzymes have been shown to possess different cleavage specificities. This has resulted in certain RNase H enzymes producing off target cleavages at the nucleotides adjacent to the target site.[244,245]

3.1.2 Detection of modified nucleotides by TLC

For many years, thin layer chromatography (TLC) has been used for the detection of modified nucleotides. Accurate nucleotide identification by TLC requires a buffer system that provides clear separation between a number of molecules. This was achieved in 2014 by Grosjean and colleagues who compiled TLC maps showing the migration patterns for 70 modified nucleotides.[246]

The SCARLET method (Site-specific Cleavage And Radioactive-labelling followed by Ligation-assisted Extraction and Thin-layer chromatography) is currently the most accurate and reliable method of m^6A detection and quantification available. This method was published in 2013 by Liu and colleagues,[90] however, an almost identical method had been developed many years previously for the detection of modified nucleotides in non-coding RNA such as pseudouridine.[244,247,248] Liu and colleagues demonstrated for the first time that this method can be used to identify m^6A within mRNA.[90]

The key step of the SCARLET method (outlined in Figure 34) is the RNase H directed cleavage of the RNA at the 5'- of the target m^6A/A candidate site. After the RNA cleavage, all of the 5'-terminal phosphates in the RNA are removed and subsequently replaced with a ^{32}P radiolabelled phosphate. The candidate m^6A/A site is radiolabelled, along with a large number of non-target nucleotides. The specific m^6A/A site of interest undergoes a series of rigorous purification steps during which the candidate site is splint ligated to a single

stranded DNA oligonucleotide. The sample is then treated with RNase T1 and RNase A which digest all of the RNA in the sample except for the single m^6A/A RNA nucleotide which is protected from the RNase degradation by the DNA oligonucleotide it is ligated to. The DNA-^{32}P-m^6A/A oligonucleotide is then PAGE purified and digested into 5'- nucleoside monophosphates using Nuclease P1. The sample is then analysed using TLC to quantify the relative levels of m^6A and adenosine.[90]

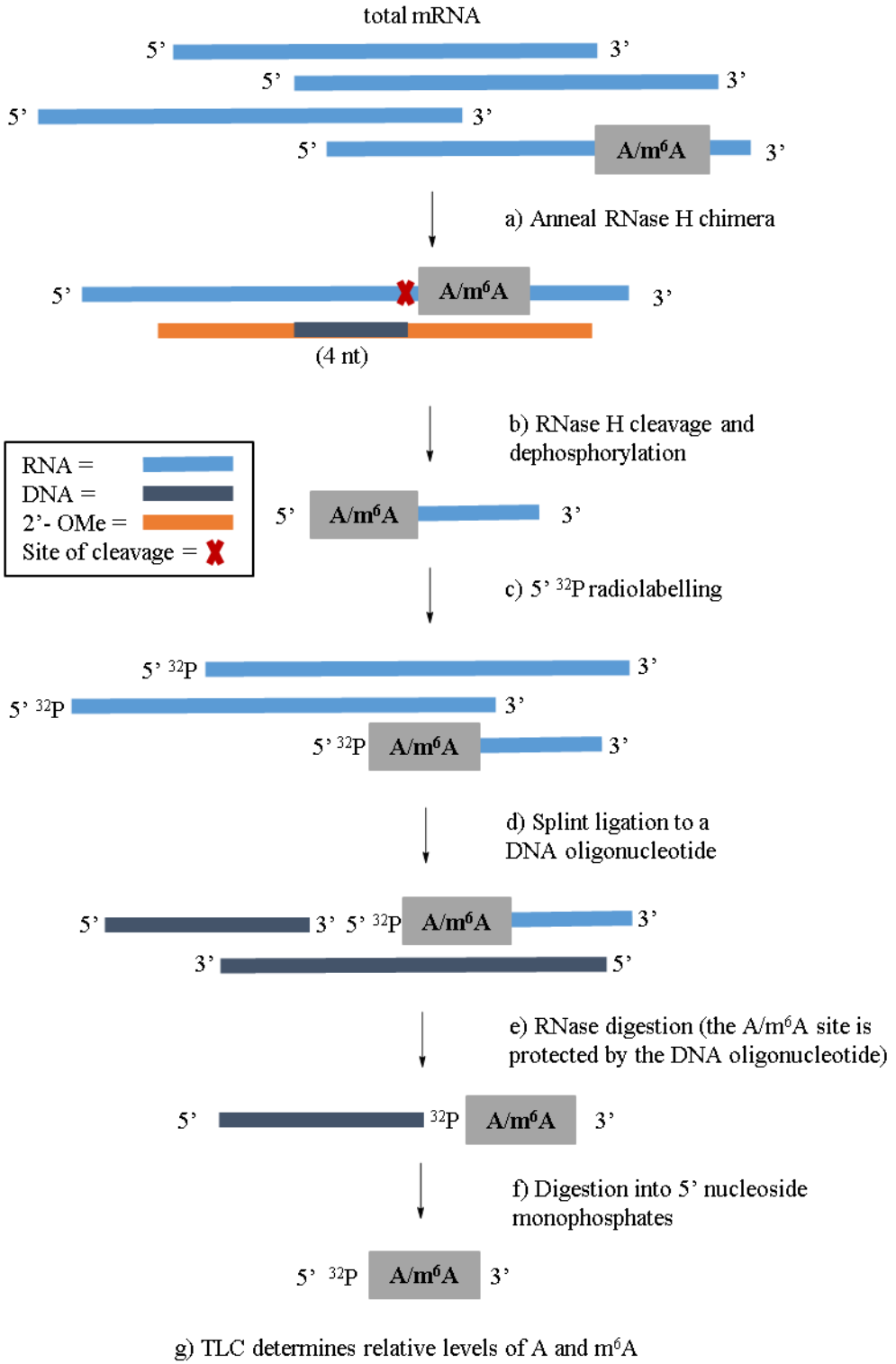


Figure 34: The SCARLET method of m^{6}A detection: a) A 2'-OMe/DNA chimera is used to target an RNase H cleavage of the oxygen-phosphorus bond immediately 5'- to the m^{6}A/A candidate site b) The RNA is cleaved using

RNase H and 5’ dephosphorylated c) The nucleic acid is re-phosphorylated with ^{32}P d) A single stranded DNA oligonucleotide is ligated to the 5’ of the m^6A/A site e) RNase enzymes degrade all of the RNA, except for the m^6A/A nucleotide which is protected by the ligated DNA f) The m^6A/A containing DNA/RNA is PAGE purified and digested into single mononucleotides g) The relative levels of m^6A and adenosine are quantified by TLC.[90]

3.1.3 Inefficiencies in the SCARLET method

Despite being referred to as the ‘gold standard’ of m^6A detection,[91] the SCARLET method is rarely used in the published literature. This is in contrast with the antibody-based methods that are regularly used in m^6A studies, despite the low accuracy and low specificity of the anti-m^6A antibody.[111] Recent work by Garcia-Campos et al., (2019) is one of the few examples of the SCARLET method being used in the literature. They reported that 4 out of the 23 sites that they analysed with SCARLET produced unusable results.[91]

The SCARLET method has a number of inefficient steps that we believe can be improved upon. The main problem with the SCARLET method is that the radiolabelling step is not specific to the m^6A/A candidate site, resulting in off-target labelling of every 5’ hydroxyl group in the sample. Prior to radiolabelling, the SCARLET method requires heat inactivation of the RNase H and alkaline phosphatase enzymes in the presence of the divalent cation $MgCl_2$. The Mg^{2+} ions will cause unintentional fragmentation of the mRNA yielding a large number of off-target 5’ hydroxyl groups which will be

radiolabelled.[249] This non-specific radiolabelling requires stringent purification of the target A/m^{6}A site including PAGE gel purification of the target site, which is hazardous, time consuming, and risks cross sample contamination if performing multiple SCARLET experiments in parallel.

The SCARLET method uses T4 DNA ligase to join a single stranded DNA molecule to the 5' candidate site. However, T4 DNA ligase is able to splint ligate nucleic acid with base mismatches and gaps of 1 or 2 nucleotides.[250] This will lead to off target ligation, resulting in additional contaminating nucleotides being present on the TLC plate. T4 DNA ligase has also been shown to exhibit poor binding to DNA/RNA hybrids, often resulting in abortive ligation.[251]

The SCARLET method quantifies the relative amounts of m^{6}A and adenosine using one-dimensional TLC. The adenosine and m^{6}A nucleotides can be clearly distinguished by TLC, however it is difficult to eliminate all of the contaminating radiolabelled material, particularly the free ^{32}P.[246] This radiolabelled material is often indistinguishable from the adenosine and m^{6}A spots using one-dimensional TLC. Additionally, m^{6}A and cytosine have similar migration patterns on the one-dimensional TLC plate, so off target ligations or RNase H digestions may result in false positive m^{6}A signals. For

this reason, we suggest that SCARLET experiments should always be analysed using two-dimensional TLC.

3.1.4 An alternative to the SCARLET method

A similar method to SCARLET devised by Wu et al., (2010) has been used to detect pseudouridine in small nucleolar RNAs (snRNA).[252] This method involves affinity purification of the target RNA using a biotinylated oligonucleotide probe followed by targeted RNase H digestion at the 3'- of the site of interest. The RNA was PAGE purified and ligated to a 5'- radiolabelled RNA oligonucleotide. The RNA was again PAGE purified and digested into 3'- labelled mononucleotides which were analysed by two dimensional TLC.[252] In this chapter we present RedBaron, a novel method of site-specific detection and quantification of m^6A, which is based on this method.

3.2 Aims

i. Develop a TLC based method for the site-specific detection of m^6A

Due to the limitations we have identified within the SCARLET method, we aimed to design a novel method for the detection and quantification of m^6A.

ii. Determine the migration patterns of 5'- and 3'-nucleoside monophosphates

In order to analyse the TLC results of the RedBaron method, we aimed to synthesise ^{32}P radiolabelled 5'- and 3'- nucleoside monophosphates and determine their relative migration patterns on the TLC plate.

iii. Determine whether the RedBaron method is able to accurately detect and quantify m^6A

To test the RedBaron method on synthetic oligonucleotides that either contain or lack the m^6A modification and to determine whether the method can be used to accurately quantify m^6A.

iv. Testing site specific detection and quantification of m^6A of real mRNA samples using SCARLET and RedBaron

To use the RedBaron and SCARLET methods to detect and quantify m^6A in RNA transcripts from a range of tissues, cells, and organisms.

3.3 Results and discussion

3.3.1 The RedBaron method of m^6A detection

We have developed a novel method for the detection and quantification of m^6A that will provide an alternative to SCARLET (outlined in Figure 36). The RedBaron method involves an RNase H cleavage in which the cut site is at the 3'- of the A/m^6A candidate site (as opposed to the 5'- in SCARLET) (Figure 35). A 5'- ^{32}P radiolabelled ssDNA oligonucleotide is splint ligated to the 3'- of the m^6A/A site. The DNA/RNA is then digested into 3'- nucleoside monophosphates, resulting in 3'- radiolabelling of the m^6A/A nucleotide. The relative levels of m^6A and adenosine are then quantified by two-dimensional TLC.

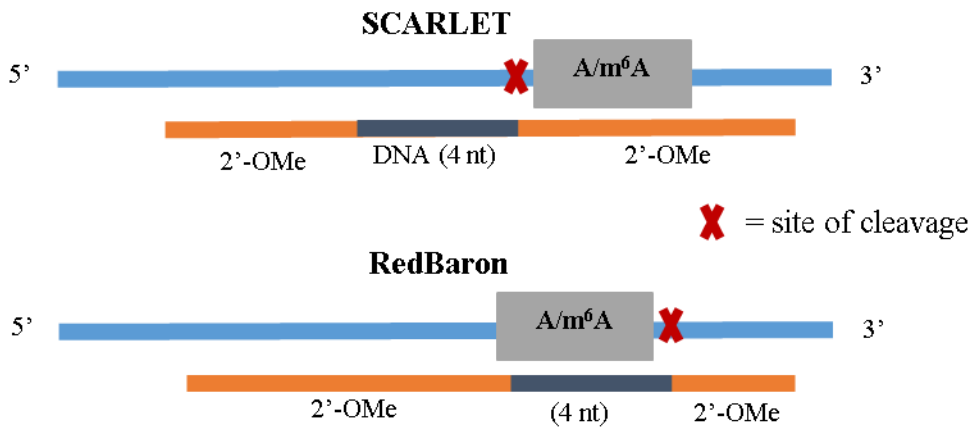


Figure 35: RNase H cleavage step of the SCARLET and RedBaron methods. RNase H cuts at the 5'- of the m^6A/A candidate sites in SCARLET,[90] and the 3'- in the RedBaron method.

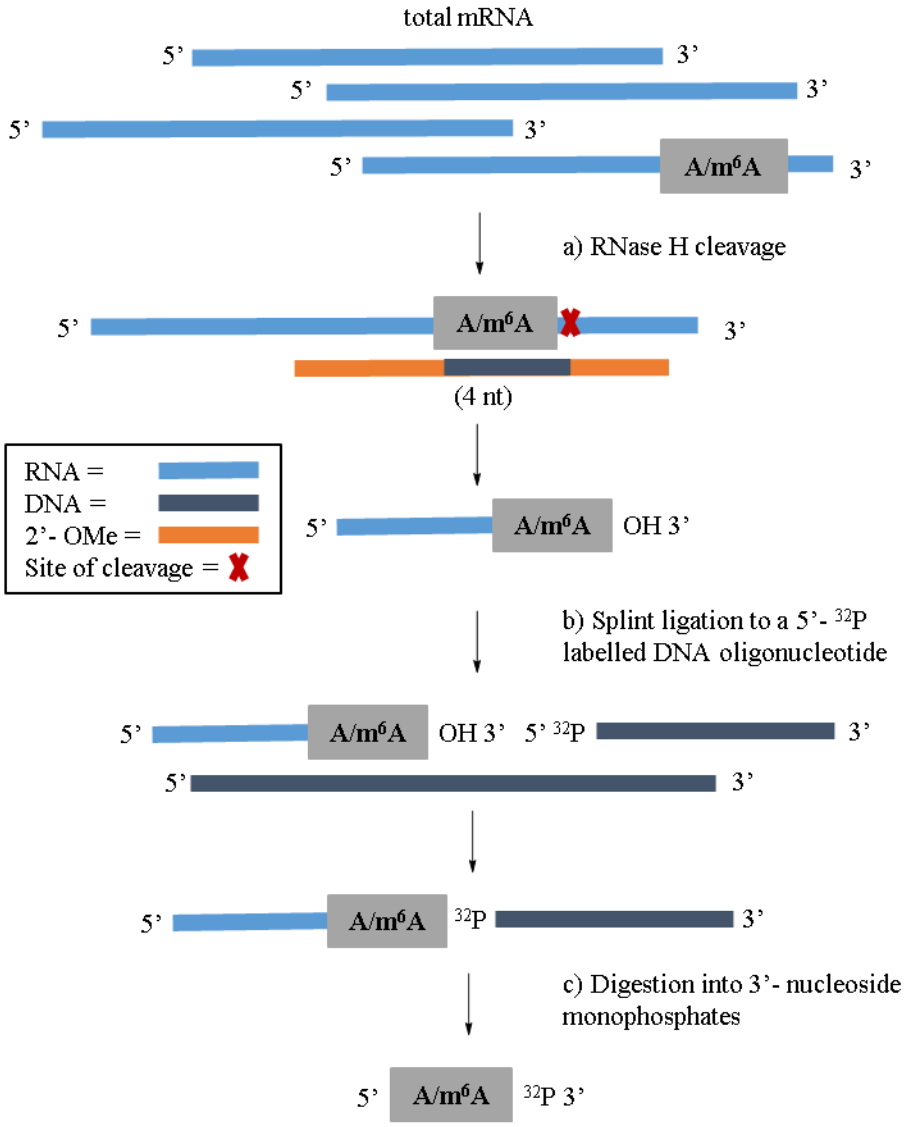


Figure 36: The RedBaron method of m^6A detection: a) A 2'-OMe/DNA chimera is used to target an RNase H cleavage of the oxygen-phosphorus bond immediately 3'- to the m^6A/A candidate site b) A ^{32}P radiolabelled single stranded DNA oligonucleotide is ligated to the 5'- of the m^6A/A candidate site c) The RNA is cartridge purified, digested into single mononucleotides e) The relative levels of m^6A and adenosine are quantified by TLC.

3.3.2 Selecting an enzyme for splint assisted ligation

For accurate detection of m^6A, the ligation step must be specific to the m^6A/A candidate site. This is achieved by utilising splint assisted ligation. Splint ligation is the joining of two adjacent single stranded nucleic acid sequences using a bridging/splint oligonucleotide.[253]

The RedBaron method required an enzyme that is capable of splint ligating a 5'- DNA oligonucleotide to a 3'- RNA molecule. Most splint ligase enzymes work optimally when using an RNA splint,[253] however due to the high cost of RNA oligonucleotide synthesis, a DNA splint is more economical. We identified a number of ligase enzymes capable of splint ligation (Table 5).

Table 5: Comparison of splint ligase enzymes.

Splint ligase enzyme	**Advantages/disadvantages**
T4 DNA ligase	Risk of off-target and abortive ligation[251]
T4 RNA ligase II	Reduced efficiency when ligating with DNA[250]
Splint R ligase	Faster turnover rate and a considerably lower Km than T4 DNA ligase[254]

As previously mentioned, the T4 DNA ligase used in the SCARLET method is prone to off-target and abortive splint ligation.[250,251] Therefore, an alternate ligase enzyme was required. We considered using RNA ligase II, however, this enzyme has a preference for double stranded RNA substrates meaning the

cost of synthetic RNA splints for each m^6A/A candidate site would be very high.[250]

We opted to use Splint R ligase which maintains a high ligation efficiency for RNA/DNA hybrids, allowing the low-cost synthesis of DNA splint and ssDNA oligonucleotides. Splint R ligase also has a faster turnover rate and a considerably lower Km than T4 DNA ligase.[254] Due to the problems we have described with T4 DNA ligase we would also recommend using the Splint R ligase when undertaking the SCARLET method.

3.3.3 Choice of nuclease enzyme

The standard SCARLET method uses Nuclease P1 to digest the nucleic acid into 5'- nucleoside monophosphates. As the RedBaron method requires digestion into 3'- nucleoside monophosphates an alternate nuclease enzyme was required. We identified a number of candidate nuclease enzymes (Table 6).

Table 6: Nuclease enzymes capable of digesting DNA/RNA into 3'-nucleoside monophosphates.

Nuclease enzyme	**Advantages/disadvantages**
RNase I	Unable to digest DNA[255] 2', 3'-cyclic monophosphate contamination[255]
Phosphodiesterase II	Minimal digestion observed in enzyme tests
Micrococcal nuclease	Dinucleotide contamination[256] Efficiently digests DNA and RNA[256]

We initially considered phosphodiesterase II, however when testing the enzyme, minimal digestion was observed, even with extended incubation times. RNase I was a suitable candidate; however, the enzyme is reported to produce 2'-, 3'- cyclic nucleoside monophosphates in addition to the required 3'- nucleoside monophosphates. As these molecules run very closely on the TLC plate it would make analysis difficult. Additionally, RNase I is unable to digest the ssDNA and the DNA splint[255]. We opted to use micrococcal nuclease for the digestion as it is an efficient nuclease of all forms of nucleic acid. One disadvantage of micrococcal nuclease is that it results in dinucleotide contamination[256], however this can be easily distinguished from m^6A and adenosine on the TLC plate.

3.3.4 Reference compounds for TLC analysis

The final step of the RedBaron method is to analyse the relative levels of m^6A and adenosine by TLC. In order to interpret the TLC results we needed to determine the migration patterns of the 5'- and 3'- nucleoside monophosphates (Figure 37). The TLC patterns of 5'- nucleoside monophosphates are known, and we hypothesised that the phosphate group at the 3'- end of the nucleotide would be sufficient to alter its position on the TLC plate. To test the migration patterns of these nucleosides we synthesised and radiolabelled a range of nucleoside monophosphates for comparison.

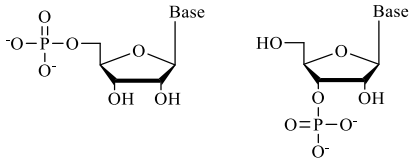


Figure 37: Structure of 5'- nucleoside monophosphates (left) and 3'- nucleoside monophosphates (right).

We first analysed 5'- nucleoside monophosphates to confirm that they had migration patterns consistent with the published literature.[246] 5'- nucleoside monophosphates from mouse testis mRNA were radiolabelled with ^{32}P. The mRNA was first digested into short fragments using RNase T1 (cuts at the 3' of every G). These fragments were then 5'- radiolabelled with ^{32}P, digested into 5'- nucleoside monophosphates using ribonuclease P1, and analysed by two-dimensional TLC (Figure 38). The observed migration patterns of ^{p}A, ^{p}G, ^{p}C, ^{p}U, and $^{p}m^{6}A$ were consistent with those previously reported.[246]

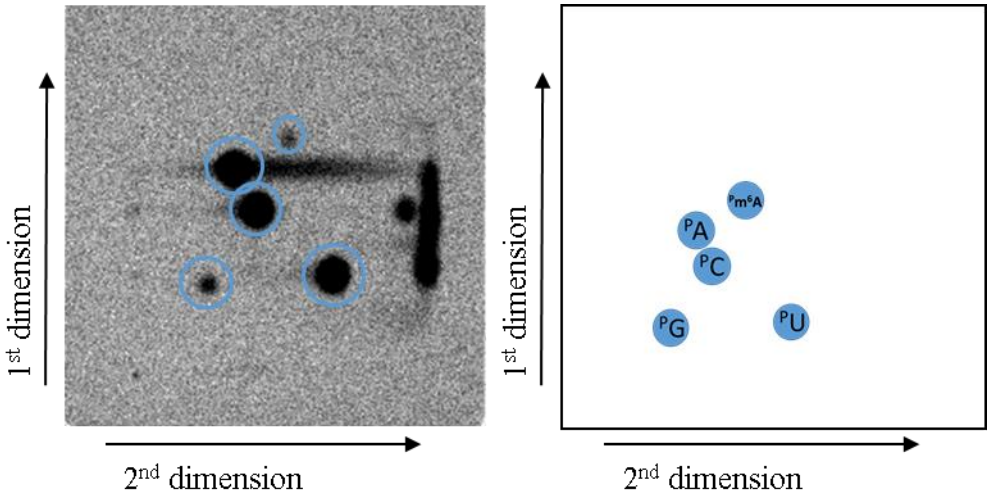


Figure 38: The TLC migration patterns of 5'- nucleoside monophosphates. Mouse testis mRNA has been radiolabelled and digested into

mononucleotides. Buffer 1 = isobutyric acid:0.5 M NH_4OH (5:3, v/v), buffer 2 = isopropanol:HCl:water (70:15:15, v/v/v).

To compare the relative migration patterns of the 5'- and 3- nucleoside monophosphates we radiolabelled the nucleosides from *in vitro* transcribed mRNA. The ^{32}P radiolabelled 3'- nucleoside monophosphates were synthesised by substituting [α-^{32}P]rCTP with rCTP in the *in vitro* transcription reaction. The RNA was then digested into 3'- nucleoside monophosphates resulting in radiolabelling of all nucleotides with an adjacent 3'- cytosine. This was repeated using rm^6ATP in place of rATP in the *in vitro* transcription reaction.

Analysis by two-dimensional TLC showed that the 3'- nucleoside monophosphates all ran slightly further in both the first and second dimensions compared with their respective 5'- nucleoside monophosphates (Figure 39).

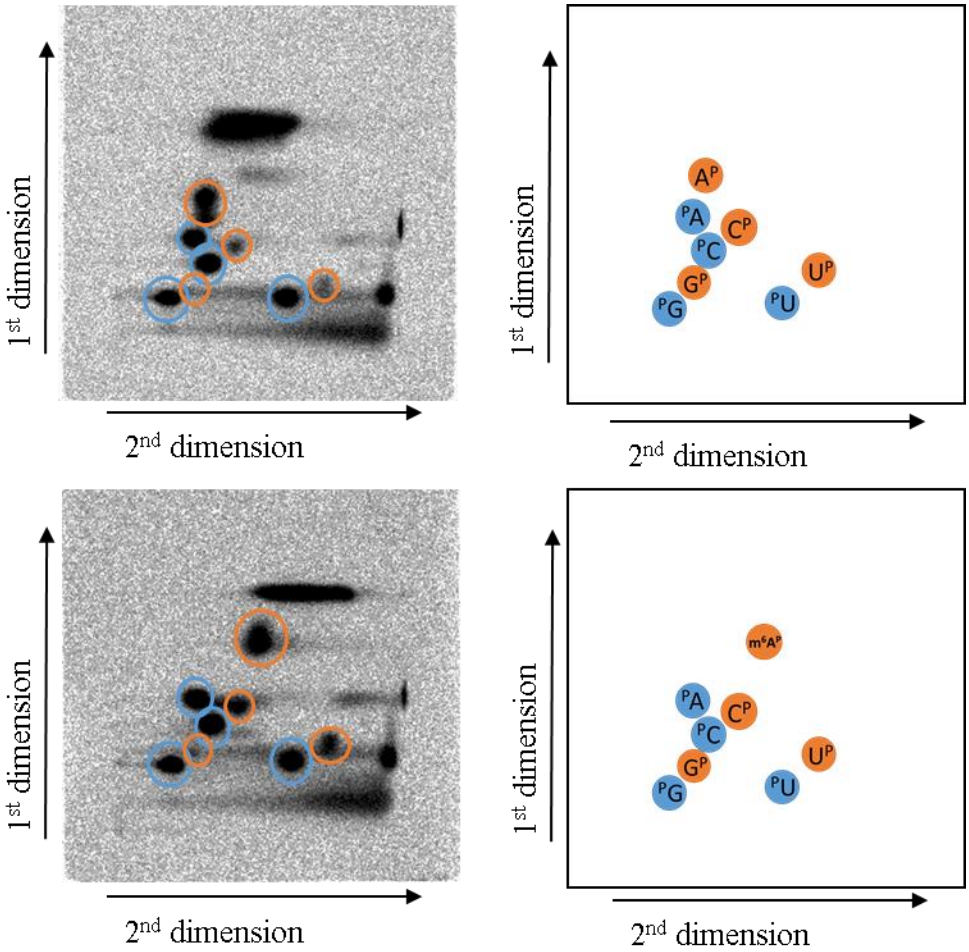


Figure 39: The TLC migration patterns of 5'- and 3'- nucleoside monophosphates. The 3'- nucleoside monophosphates run slightly further in both first and second dimension. Buffer 1 = isobutyric acid:0.5 M NH_4OH (5:3, v/v), buffer 2 = isopropanol:HCl:water (70:15:15, v/v/v).

We then synthesised radiolabelled 3'- nucleoside monophosphates of adenosine and m^6A in the absence of G^p, C^p, and U^p, using the method described in Figure 40.

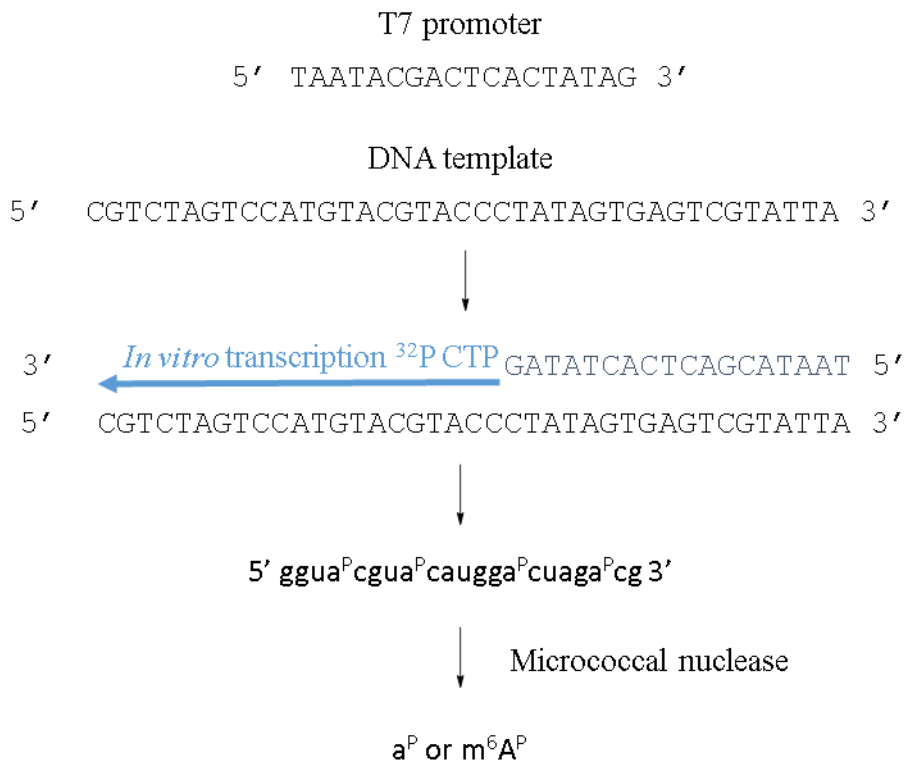


Figure 40: Synthesis of adenosine and m^6A ^{32}P radiolabelled 3'- nucleoside monophosphates. The *in vitro* transcription template was designed so that every transcribed cytosine (C) was preceded by a 5'- A. The *in vitro* transcription reaction used T7 polymerase, a primer containing the T7 promoter sequence and ^{32}P CTP. The transcribed RNA was then digested into 3'- nucleoside monophosphates, passing the 5'- ^{32}P of C to the 3'- of the A/m^6A.

Our TLC analysis (Figure 41) showed that A^p and m^6A^p are easily distinguishable from each other. The labelled 5'- nucleoside monophosphates ^pA, ^pG, ^pC, and ^pU make ideal reference molecules for the identification of A^p and m^6A^p by two-dimensional TLC when using our RedBaron method.

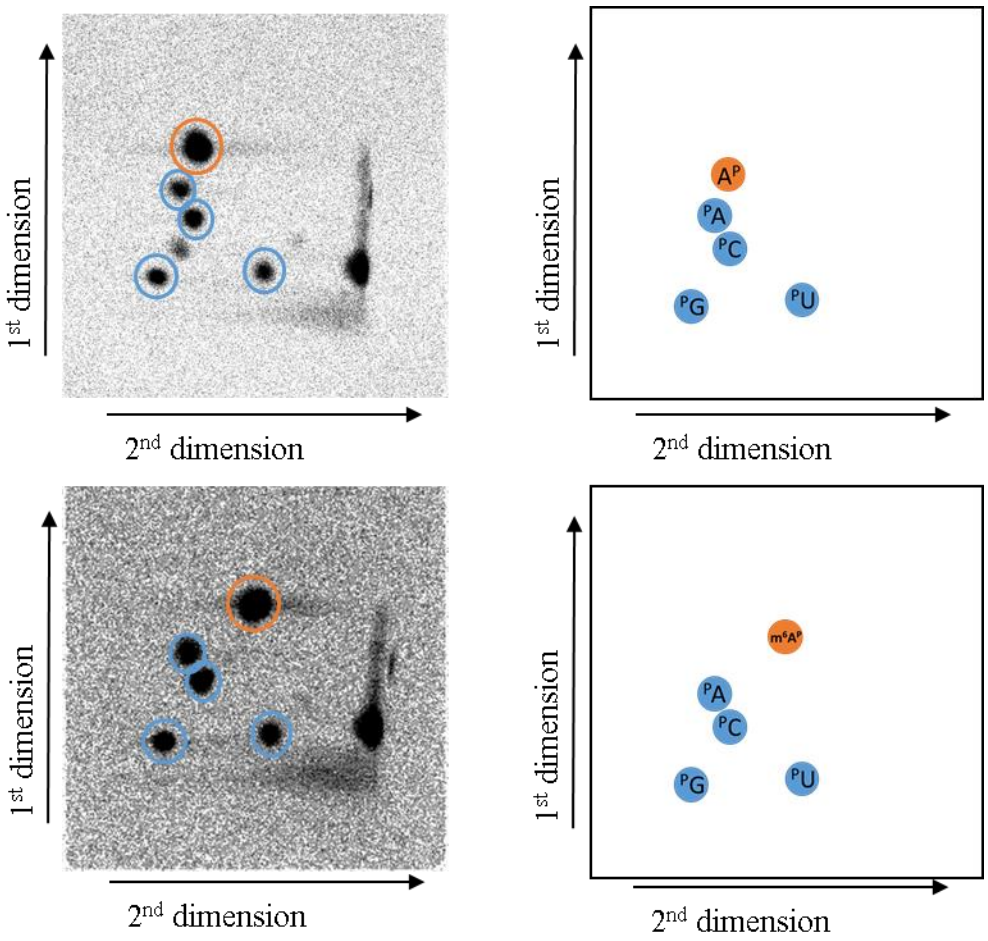


Figure 41: The 5'- nucleoside monophosphates are used as reference compounds for two-dimensional TLC analysis of RedBaron experiments. Buffer 1 = isobutyric acid:0.5 M NH_4OH (5:3, v/v), buffer 2 = isopropanol:HCl:water (70:15:15, v/v/v).

3.3.5 RedBaron can accurately detect and quantify m^6A in RNA oligonucleotides

In order to test the RedBaron method we synthesised oligonucleotides containing either adenosine or m^6A within the GGACU consensus motif

(Table 7). We found that our method was easily able to distinguish between the adenosine and m^6A containing oligonucleotides (Figure 42).

Table 7: Synthetic oligonucleotides for analysis of the RedBaron method.

Name	**DNA/RNA**	**Sequence (5'-3')**	**Length (nt)**
Scarlet test unmodified	RNA	GCAAGUGCUUCUAGGCG GACUGUUACUG	28
Scarlet test m^6A	RNA	GCAAGUGCUUCUAGGCG G(m^6A)CUGUUACUG	28

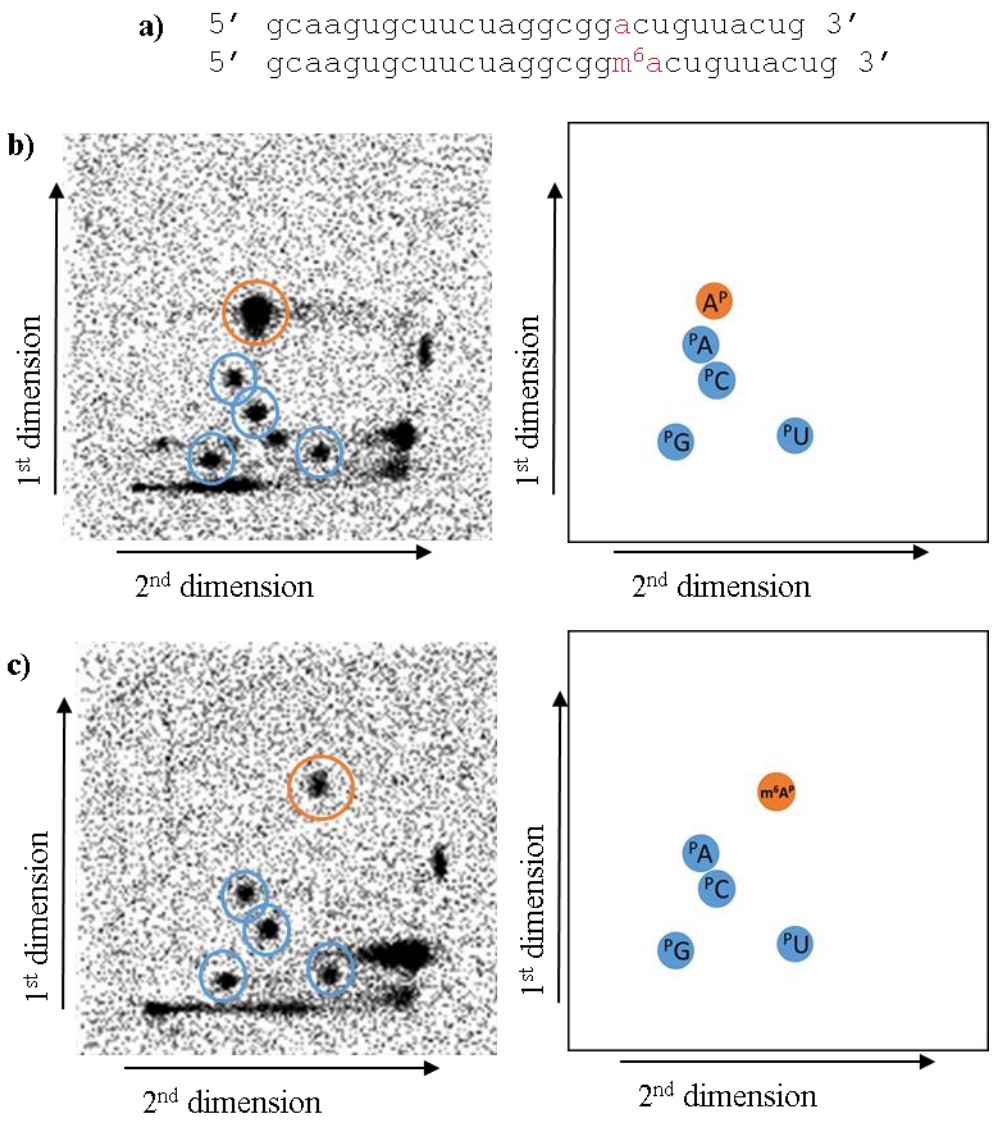
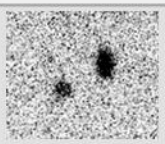
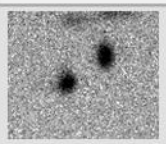
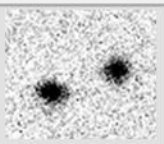
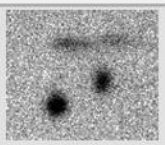
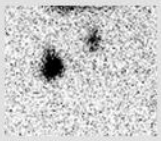
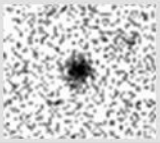
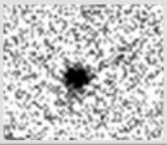


Figure 42: Using RedBaron to distinguish between adenosine and m^6A in synthetic RNA oligonucleotides. a) Synthetic RNA oligonucleotides used to test the RedBaron method containing either adenosine or m^6A at the GGACU consensus sequence. b) RedBaron experiment using unmodified adenosine oligonucleotide. c) RedBaron experiment using the m^6A modified oligonucleotide (reference molecules ^PA, ^PG, ^PC, ^PU (circled blue).

We then tested whether the method was able to quantify the relative levels of adenosine and m^6A. Varying ratios of the unmodified and m^6A synthetic RNA

oligonucleotides were analysed using the RedBaron method. The m^6A/adenosine ratio we observed accurately reflected that of the input synthetic oligonucleotides. This result indicated that the RedBaron method is able to accurately detect and quantify m^6A within synthetic RNA oligonucleotides (Figure 43). We have shown that the method is accurate when using RNA levels as low as 4 fmol. Further work is required to determine the feasibility of the method to detect m^6A in lower concentrations of RNA.

a)

Input $m^6A/(m^6A + A)$	0.80	0.60	0.50	0.40
				
Observed $m^6A/(m^6A + A)$	0.84	0.62	0.49	0.40
Input $m^6A/(m^6A + A)$	0.20	0.10	0.05	
				
Observed $m^6A/(m^6A + A)$	0.20	0.08	0.06	

b)

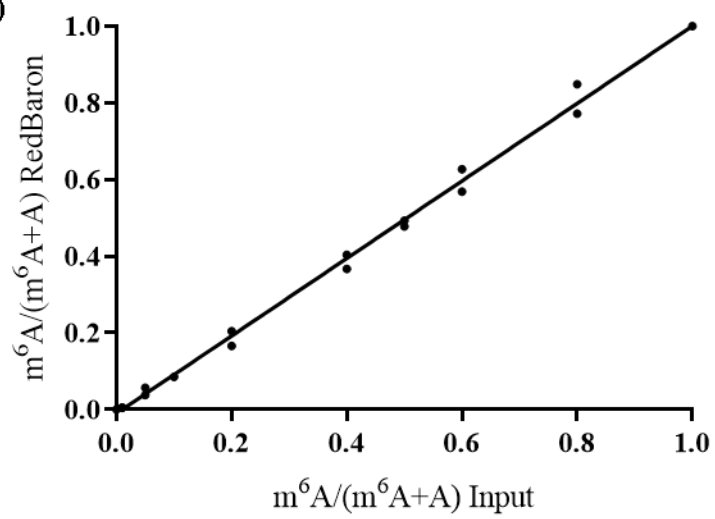
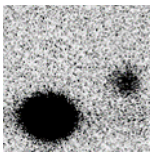
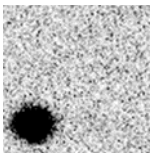
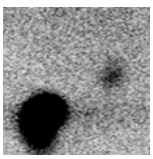
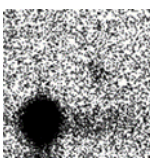
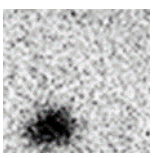


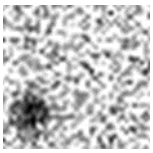
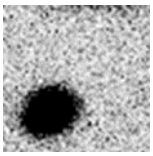
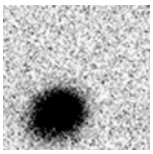
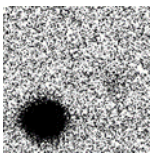
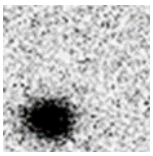
Figure 43: The RedBaron method is able to quantify levels of m^6A. a) RedBaron method using varying ratios of the adenosine and m^6A oligonucleotides (the full-size TLC images are in Supplementary Figures 1-7). b) Quantitative plot of the RedBaron experiments using varying ratios of the m^6A and adenosine oligonucleotides.

3.3.6 Testing the RedBaron method on real RNA samples

Having demonstrated that the RedBaron method can accurately detect and quantify m^6A and adenosine within synthetic oligonucleotides we tested the method on real mRNA samples from various tissues and organisms (Table 8).

Table 8: Quantification of m^6A in β-actin mRNA from various tissues, organisms and cells (the full-size TLC images are in Supplementary Figures 8-17).

mRNA	**Site**	**Organism and tissue**	**TLC**	**m^6A stoichiometry**
β-actin	1202	Chicken embryo		3 %
β-actin	1218	Chicken embryo		0 %
β-actin	1202	Chicken fibroblast cells		3 %
β-actin	1217	Human HeLa cells		3 %
β-actin	1242	Mouse heart		0 %

β-actin	1242	Mouse liver		0 %
β-actin	1242	Mouse Kidney		0 %
β-actin	1242	Mouse testis		0 %
β-actin	1242	Mouse brain		1 %
β-actin	1263	Mouse brain		0 %

In the Chicken β-actin experiments, no m^6A was observed at position 1218 ('ACA' site), however m^6A was identified at position 1202 ('GGACU' site) at approximately 3 % m^6A. The results from the ZBP gel shift assay in chapter 2 showed that 'm^6A-AC' at position 1218 inhibited ZBP binding. This is consistent with a previous study by Chao et al., (2012) which showed that substitution of this adenosine residue with either G, C, or U, also abolished ZBP binding.[233]

These results indicate that chicken ZBP selectively binds to m^6A at site 1202 of the β-actin zipcode using its KH3 domain. This may provide a mechanism to regulate which transcripts undergo localised translation. These results

suggest that the m^6A reader functionality of the human VICKZ proteins is conserved in chickens. As with the VICKZ proteins from other organisms, chicken ZBP targets a large number of mRNA transcripts in addition to β-actin, and therefore its role as an m^6A reader would likely play an important role in the regulation of gene expression.[219]

Localised translation of β-actin has been shown to play a role in cell motility.[230] As fibroblasts are highly motile cells, we hypothesised that they would have increased levels of m^6A within the β-actin zipcode. However, we found that in chicken fibroblast mRNA the m^6A levels at position 1202 of the β-actin zipcode were also at 3 % (Table 8).

No m^6A was identified in mouse heart, liver, testis, or kidney at the 'GGACU' site within the β-actin zipcode. However, in the mouse brain we identified m^6A at this site at a fraction of approximately 1 %, and at 3 % in HeLa (human) cells (Table 8). This is consistent with the published literature which reports elevated levels of m^6A in the brain compared with other tissues.[13]

We also used the standard SCARLET method to quantify m^6A levels of β-actin site 1202 in chicken embryo mRNA, and β-actin site 1242 in mouse brain mRNA (Figure 44). Using the SCARLET method, we observed very weak radioactive spots on the TLC plate and a clear m^6A spot was not visible. We analysed our results on a two-dimensional TLC and observed a number of spots in addition to adenosine and m^6A which may not have been observed using the one-dimensional TLCs described in the SCARLET protocol.[90] This reinforces the importance of two-dimensional analysis of TLC experiments.

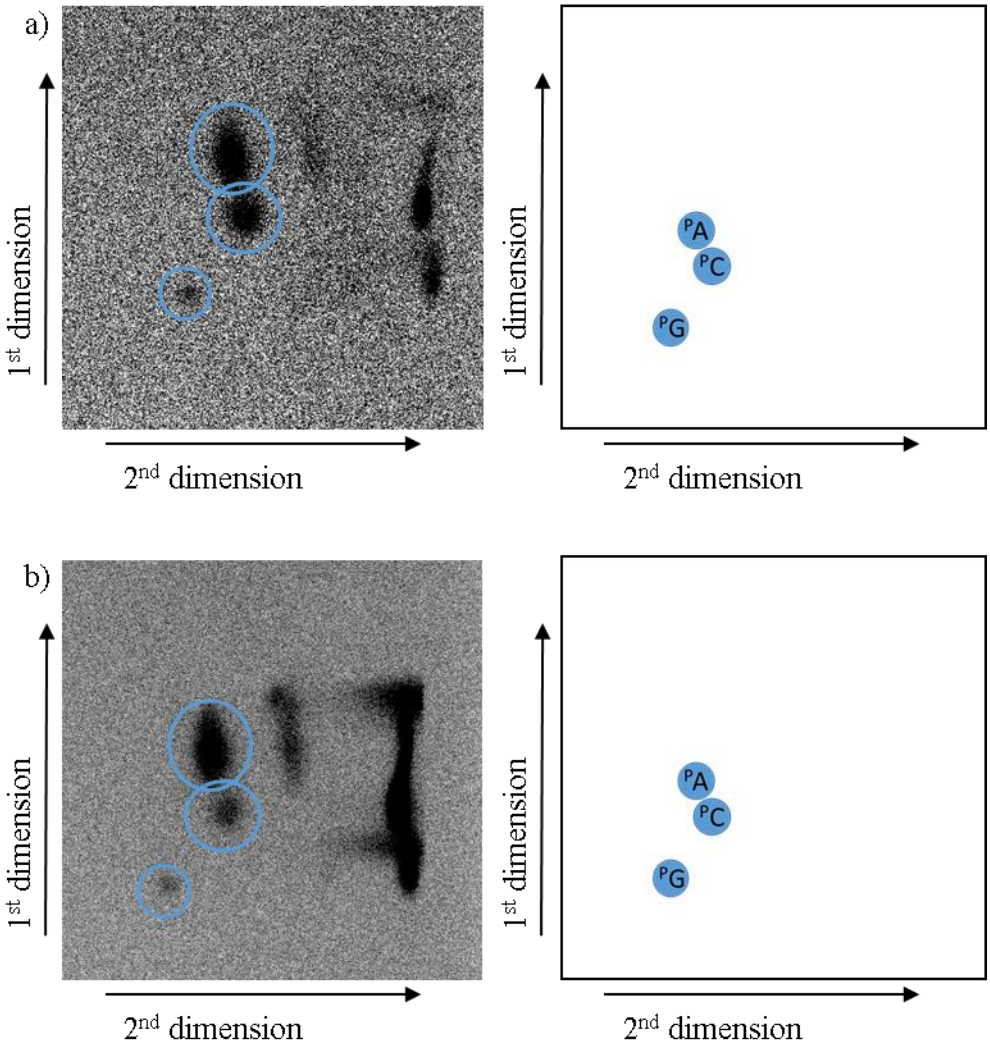


Figure 44: SCARLET analysis of β-actin mRNA. a) site 1202 in chicken embryo mRNA. b) site 1242 in mouse brain mRNA. No clear m^6A spot was observed for either sample.

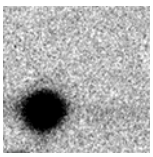
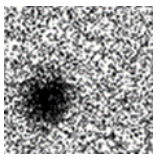
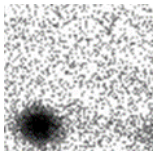
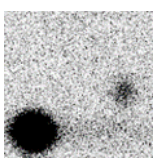
3.3.7 The RedBaron method underestimates m^6A stoichiometry

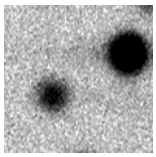
In all of the mRNA samples tested we failed to observe any sites with an m^6A stoichiometry above 5 %, this is in contrast to the published literature where reported m^6A stoichiometries are significantly higher.[90] We attempted to use

the SCARLET method to validate our results however we failed to observe sufficient m^6A and adenosine radioactive signals using this method.

Liu and colleagues used the SCARLET method to quantify m^6A in the 28S rRNA of HeLa cells at sites A4189 (5 % m^6A) and A4190 (96 % m^6A), and also in the lncRNA MALAT1 at site 2515 (61 % m^6A).[90] We used the RedBaron and SCARLET methods to reanalyse these m^6A sites to determine whether the RedBaron method is accurately quantifying m^6A levels (Table 9).

Table 9: Quantification of m^6A using RedBaron and SCARLET in MALAT1 and 28S rRNA (the full-size TLC images are in Supplementary Figures 18-22).

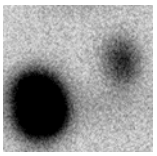
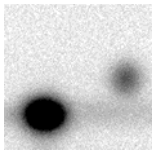
RNA (site)	**Method**	**TLC**	**m^6A stoichiometry (observed)**	**m^6A stoichiometry (literature)[90]**
MALAT1 (2515)	RedBaron		0 %	61 %
MALAT1 (2515)	SCARLET		0 %	61 %
28S rRNA (A4189)	RedBaron		0 %	5 %
28S rRNA (A4190)	RedBaron		4 %	96 %

28S rRNA (A4190)	SCARLET		88 %	96 %

Despite MALAT1 site 2515 being reported as containing 61 % m^6A in the literature,[90] using both the SCARLET and RedBaron methods we observed no m^6A. However, when testing site A4190 in the 28S rRNA we observed 88 % m^6A using the SCARLET method and just 4 % using the RedBaron method. This suggested that the RedBaron method may be under-representing the stoichiometry of m^6A when analysing real mRNA samples.

To test whether this is the case we repeated the quantification test using the synthetic oligonucleotides in the presence of 1 μg of non-target RNA (Table 10). We observed m^6A levels were significantly underrepresented upon addition of the non-target RNA indicating the RedBaron method is unable to quantify m^6A from real mRNA samples.

Table 10: Quantification of m^6A using synthetic oligonucleotides and 1 μg of non-target RNA (the full-size TLC images are in Supplementary Figures 23-24).

Oligonucleotide substrate	**Input RNA**	**TLC**	**m^6A observed**	**m^6A expected**
Chicken β-actin	Mouse embryo RNA (1 μg)		6 %	50 %
Mouse β-actin	Chicken brain RNA (1 μg)		7 %	50 %

3.3.7.1 Why does the RedBaron method underestimate m^6A levels?

Our results indicate that the RedBaron method underestimates m^6A levels when analysing real RNA samples but not when using the synthetic test oligonucleotides. In certain sequence contexts, both the SCARLET and RedBaron methods may potentially underestimate m^6A levels. This is because m^6A has been shown to slightly reduce the hybridisation efficiency of RNA and therefore the chimeric and splint oligonucleotides will have a slightly higher affinity for unmodified RNA sequences.[90,173,174] This does not account for the significantly different m^6A stoichiometries observed between the SCARLET and RedBaron methods as this phenomenon should affect both methods equally.

One explanation for the under representation of m^6A, observed with the RedBaron method, is that the radiolabelled DNA molecule may be non-specifically ligating to fragmented RNA with complementarity to the splint oligonucleotide. The RedBaron and SCARLET methods both involve heating of the RNA in the presence of divalent cations which is known to cause fragmentation of the RNA.[249] The ligation step should be specific to a single nucleotide position due to the use of a splint targeted ligation that is specific to the site of interest.[253] However, fragmentation of the RNA could potentially produce RNA fragments with complementarity to the splint which would result in radiolabelling of adenosines that are not located within the target site. An example of this is outlined in Figure 45.

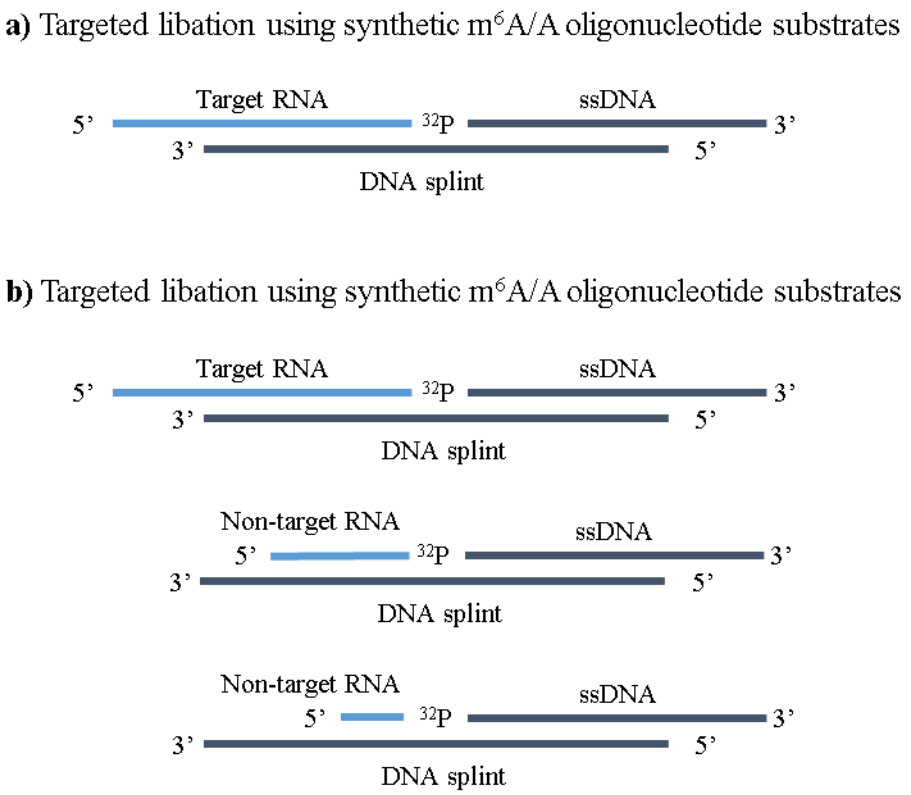
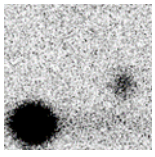
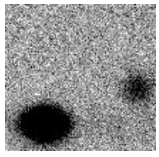


Figure 45: Off-target ligation caused by fragmentation of RNA. a) using the synthetic A/m^6A RNA substrate only site-specific ligation was observed. b)

When using real RNA substrates , RNA fragmentation resulted in RNA fragments with complementarity to the DNA splint resulting in off-target ligation and under representation of m^6A.

To test this hypothesis, we repeated the RedBaron analysis of the 28S rRNA site A1490, however prior to the heat inactivation steps we added EDTA to a final concentration of 20 mM in order to quench the divalent cations which should reduce the amount of RNA fragmentation and therefore the amount of off-target ligation. Using this modified version of the RedBaron method we observed an increase in the m^6A levels from 4 % to 7 % (Table 11). Although this value is considerably lower than that reported in the literature, this result indicates that there may be scope for the RedBaron method to be improved in the future.

Table 11: A modified version of the RedBaron method using EDTA to chelate divalent cations prior to sample heating. This increased the observed m^6A levels from 4 % to 7 % (the full-size TLC images are in Supplementary Figures 21, and 25).

mRNA site	**Method**	**TLC**	**m^6A stoichiometry**	**SCARLET m^6A stoichiometry**
HeLa 28S rRNA site A4190	RedBaron (standard)		4 %	88 %
HeLa 28S rRNA site A4190	RedBaron (EDTA before heating)		7 %	88 %

3.3.8 RedBaron using a biotinylated RNA probe

In an attempt to optimise the RedBaron method to prevent fragmentation of the mRNA we adapted the method to be undertaken on streptavidin magnetic beads. We *in vitro* transcribed a biotinylated RNA oligonucleotide probe, complementary to the mouse β -actin mRNA. This probe was used to purify the transcript using streptavidin magnetic beads. The RedBaron method was performed with the RNA bound to the magnetic beads allowing the reaction buffers to be easily replaced to suit the specific requirements of the various enzymes used. Unfortunately using this method, we failed to observe sufficient radioactive signals for either adenosine or m^6A as the streptavidin magnetic beads appeared to inhibit either the RNase H or ligation reactions.

3.3.9 Detecting 2'-OMe base modifications

The RedBaron method is unsuitable for the detection of nucleotides modified at their 2'- hydroxyl groups, for example A_m and m^6A_m. This is because RNase H is unable to make a cleavage at the 3'- of a 2'-OMe methylated nucleotide.[243] To date there is no evidence to suggest that SCARLET will work on 2'-OMe modified nucleotides. It is therefore important to consider that the m^6A fraction obtained from RedBaron and SCARLET results is not accounting for any 2'-OMe modified nucleotides including m^6A_m (if m^6A_m is located internally within mRNA).

3.3.10 Concluding points

The aim of this chapter was to develop a novel method of m^6A detection based upon the pre-existing SCARLET method. We developed the RedBaron method, which was able to accurately quantify m^6A levels in synthetic RNA oligonucleotides. We used this method to detect and quantify m^6A in RNA transcripts from a range of tissues, cells, and organisms, however, we observed m^6A levels that were lower than expected. Unfortunately, we observed that the RedBaron method significantly underestimates m^6A levels in real mRNA samples and is therefore unsuitable as a method of m^6A detection. In this chapter we made numerous attempts to use the SCARLET method for m^6A quantification, however, failed to observe clear and reliable spots for m^6A and adenosine on the TLC plate. This demonstrates the requirement for new, accurate methods of m^6A detection and quantification.

Chapter 4: PhSeT mediated nucleic acid crosslinking

Overview: Nucleic acid crosslinking can be artificially induced for a range of applications including the regulation of gene expression, the detection of specific nucleotide sequences, and the development of antitumor agents.[257,258]

5–phenyl selenide thymidine (PhSeT) is a modified nucleoside that forms a covalent crosslink with adenosine in nucleoside and oligonucleotide sequence contexts.[17,259] Unpublished work from the Hayes and Fray research groups has demonstrated that the PhSeT nucleoside is unable to crosslink with m^6A under oxidative conditions.[260] We aimed to determine whether this chemistry could be used as a biochemical method of m^6A detection.

We synthesised PhSeT modified oligonucleotide probes and tested the crosslinking reaction in a range of sequence contexts. The selective crosslinking with adenosine over m^6A was maintained in poly(A) sequence contexts. However, in more complex sequences, we found the crosslinking reaction to be inefficient and unreliable, and concluded that a crosslinking based method using this chemistry is not suitable for m^6A detection.

4.1 Introduction

4.1.1 Nucleic acid analogues

Nucleic acid analogues are structurally similar to naturally occurring nucleic acid, however, contain one or more modifications. The nitrogenous base, pentose sugar, and phosphate groups can all be modified (Figure 46).[21] The modification of nucleosides is an important avenue of research in the fields of medicinal chemistry and biochemistry with a range of applications, including the treatment of disease and use as diagnostic tools for the analysis of genetic information.[261,262]

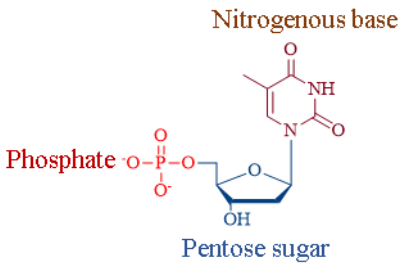


Figure 46: Structure of deoxythymidine. The nucleic acid can be modified at the nitrogenous base, pentose sugar, or phosphate group.

4.1.1.1 Examples of nucleic acid analogues

DNA and RNA molecules can be modified with fluorescent or radioactive tags to study their structure, localisation, and processing. Radiolabelling of DNA and RNA can be achieved by substituting a phosphate group with a radioactive isotope such as ^{32}P.[246] Fluorescent labelling can be accomplished by incorporating aminoallyl modified nucleotides into the nucleic acid chain.

Fluorescent dyes are reacted with the primary amine resulting in the generation of a fluorescent signal.[263–265] Nucleic acids can also be modified with a biotin tag which binds to a molecule called streptavidin with a high affinity under a wide range of conditions. This chemistry is commonly used for the isolation and purification of nucleic acid, and the identification of DNA and RNA binding proteins.[266]

4.1.2 Crosslinking of DNA and RNA

Nucleic acid crosslinking is the formation of a covalent bond between two nucleotides. Crosslinking prevents the separation of hybridised nucleic acid strands, which inhibits important cellular processes such as DNA replication and transcription, often resulting in cell death.[258] DNA crosslinking is a natural cellular process that is induced by endogenous cellular agents such as oxygen free radicals during oxidative stress,[267] and nitrous acids produced in the stomach.[268] Cells are able to repair these undesirable DNA crosslinks using a number of methods, including homologous recombination, and nucleotide excision repair.[269]

Nucleic acid crosslinking can be artificially induced using exogenous agents for a range of applications. These include the artificial regulation of gene expression, the detection of specific nucleotide sequences,[257] and the development of antitumor agents.[258]

4.1.3 Examples of artificial nucleic acid crosslinking

4.1.3.1 Mitomycin C

Mitomycin C (Figure 47) is an anti-cancer, antibiotic agent used in the treatment of breast, anal, bladder, and upper gastro-intestinal cancers. In the cell, mitomycin is converted into mitosene via reductive activation. The mitosene then *N*-alkylates two DNA bases resulting in crosslinking between two base paired nucleotides.[261] This crosslink inhibits DNA synthesis and leads to chromosomal breakage and genetic recombination.[261] A single crosslink per genome is sufficient to kill a bacterial cell.[270]

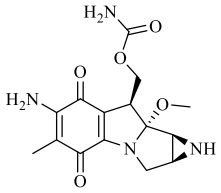


Figure 47: Mitomycin C is a DNA crosslinking agent used in the treatment of cancer.[261]

4.1.3.2 Detecting modified nucleotides using nucleic acid crosslinking

Nucleic acid crosslinking can be utilised for the detection of naturally occurring modified nucleotides. Dohno et al. (2010) developed a method for detecting the modified nucleotide 6mA in DNA. They synthesised a guanosine nucleotide with an electrophilic formyl group at its O^6 position (^fG) (Figure

48).[262] ^{f}G will crosslink with adenosine if it is located immediately to the 3'of the cytosine that it is base paired with. However, this crosslinking is suppressed if the adenosine is methylated to 6mA, allowing detection of the modified nucleotide.[262] Unfortunately, this method is limited to detecting 6mA in an 'AC' sequence context.

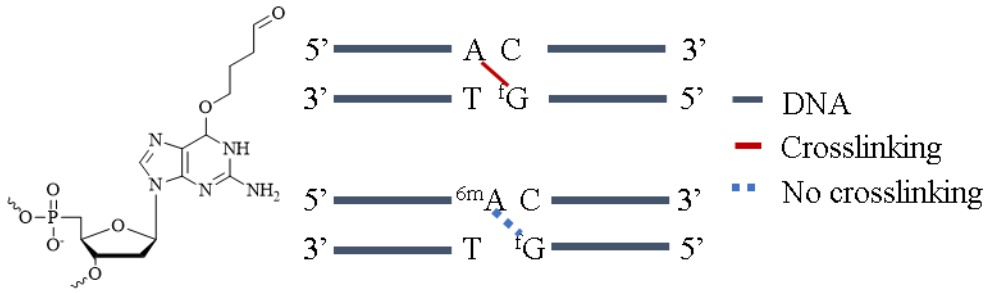
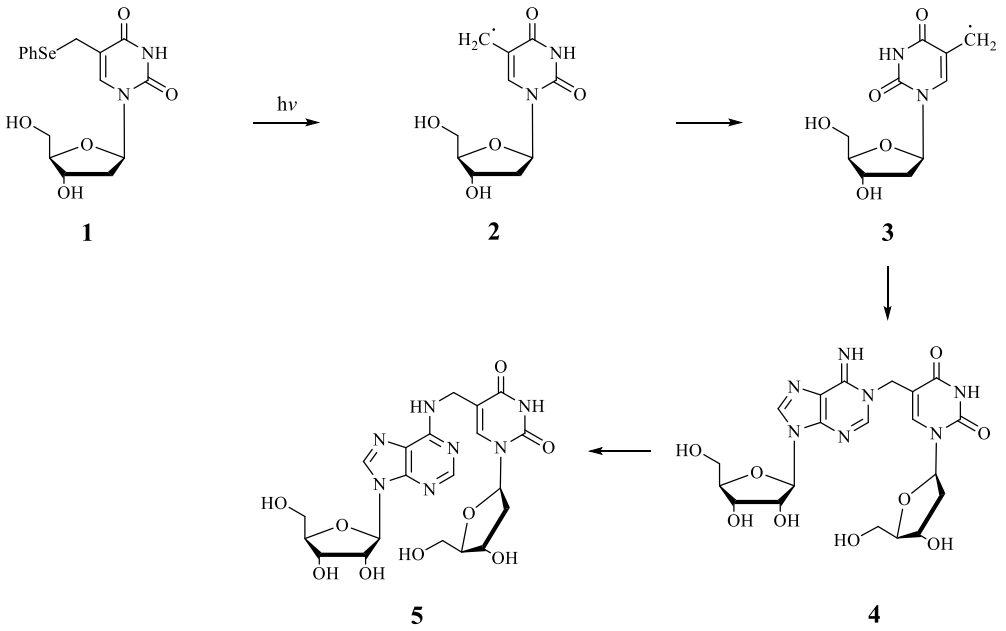


Figure 48: A formyl modified guanosine nucleotide (fG) used for the biochemical detection of 6mA in DNA. ^{f}G forms a crosslink with adenosine but not 6mA.[262]

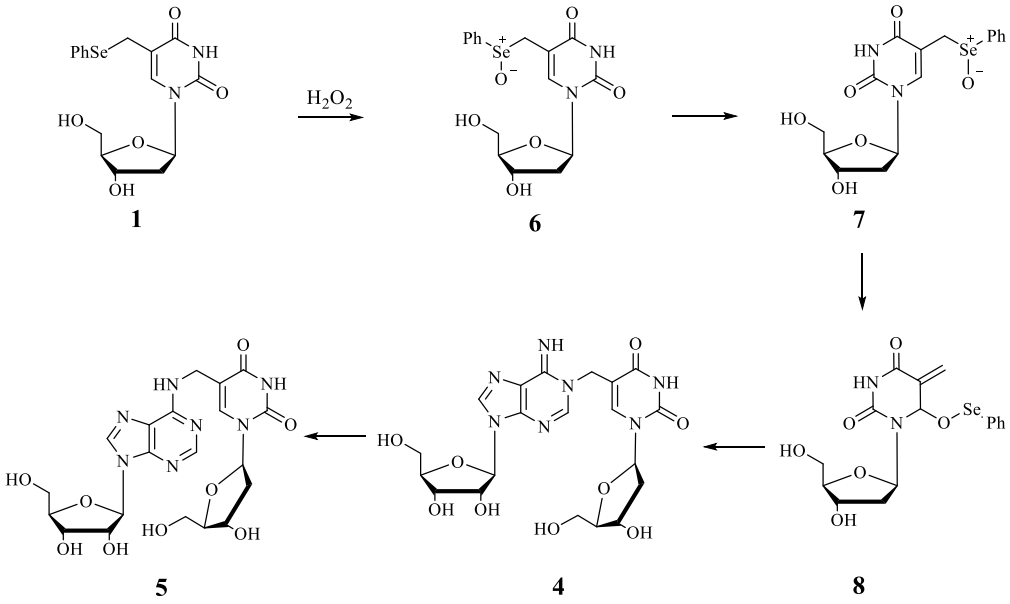
4.1.4 Phenyl Selenide modified thymidine

Over the last 15 years, Greenberg and colleagues have studied a phenyl selenide modified thymidine (PhSeT) nucleoside **1** that is able to form a selective and stable crosslink with adenosine.[17] Crosslinking between PhSeT and adenosine can be achieved by two distinct mechanisms. The first mechanism acts via a thymidine radical **2** that is formed by UV photolysis of PhSeT. The radical thymidine forms a crosslink with the complementary adenosine nucleoside **5** (Scheme 9).[271,272]



Scheme 9: Crosslinking between PhSeT and adenosine under radical conditions.[272]

The second mechanism of crosslinking between PhSeT and adenosine involves mild oxidation of PhSeT with either singlet oxygen, hydrogen peroxide, or sodium periodate. This converts PhSeT into a phenyl selenoxide **6** which rapidly forms an electrophilic allylic phenylselenate **8** via a sigmatrophic rearrangement. The electrophilic group then crosslinks with the N^1 position of adenosine to form **4**. On reaction with a nucleophile the crosslinked product then re-aromatizes via a Dimroth rearrangement to form the more stable crosslink at the N^6 position of adenosine **5** (Scheme 10).[272]

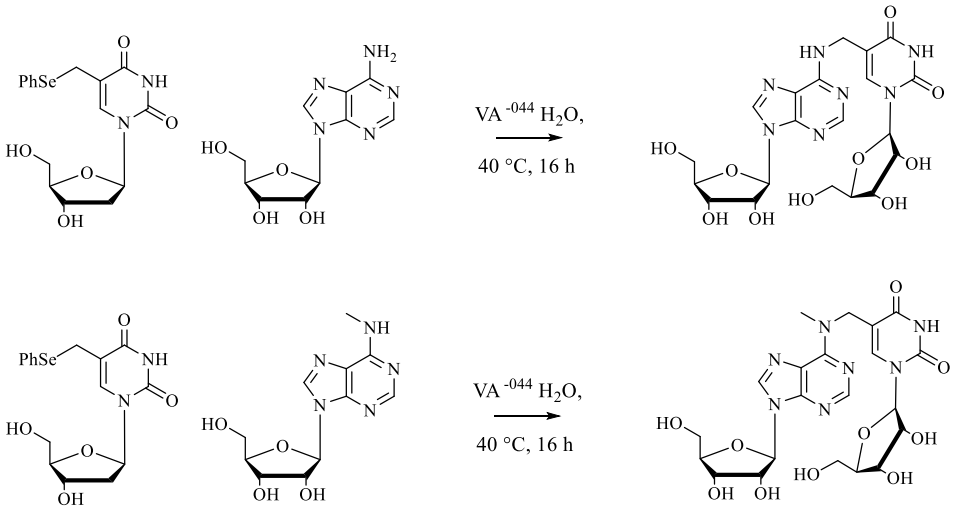


Scheme 10: Mechanism of PhSeT crosslinking with adenosine under oxidative conditions.[272]

4.1.5 PhSeT crosslinks with both adenosine and m^{6}A under UV photolysis

The ability of PhSeT to form a specific crosslink with adenosine led us to believe that the modified nucleotide could be utilised for m^{6}A detection. Unpublished work by the Hayes and Fray groups demonstrated that under radical conditions the PhSeT mononucleotide crosslinks with both adenosine and m^{6}A (Scheme 11).[260] We hypothesized that using these conditions, PhSeT modified oligonucleotides could be used as biological probes to detect m^{6}A. The PhSeT oligonucleotides could be targeted to crosslink with a specific adenosine residue, which could be isolated, digested into its single nucleoside

constituents, and the relative levels of adenosine and m^6A could be determined using mass spectrometry.

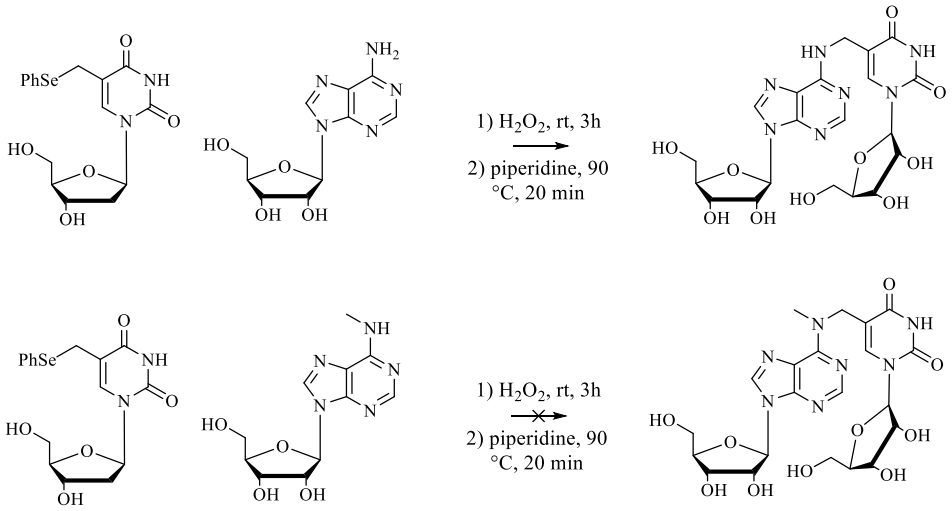


Scheme 11: PhSeT crosslinks with both adenosine and m^6A under radical conditions (unpublished work).[260]

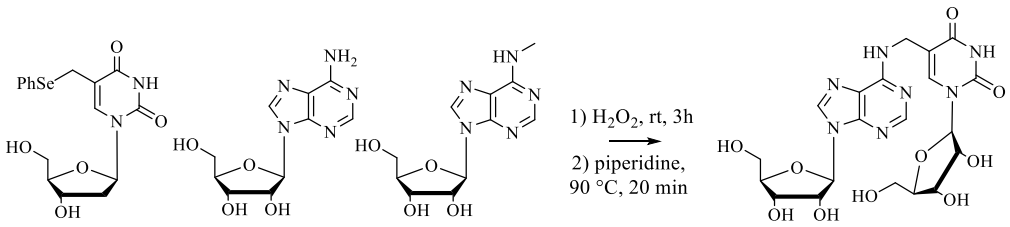
4.1.6 PhSeT selectively crosslinks with Adenosine and not m^6A under oxidative conditions

Further unpublished work has demonstrated that under the oxidative conditions of hydrogen peroxide or sodium periodate, crosslinking was uniquely identified between PhSeT and adenosine, with no crosslinking observed between PhSeT and m^6A (Scheme 12). Additionally, the presence of m^6A does not inhibit crosslinking between PhSeT and adenosine (Scheme 13).[260] This showed even greater potential for the PhSeT nucleoside to be utilised as an m^6A detection method. We hypothesised that a PhSeT modified oligonucleotide probe could be reacted with a candidate m^6A site and the

presence or absence of crosslinking could be used to distinguish between adenosine and m^6A.



Scheme 12: PhSeT crosslinks with adenosine but not m^6A under oxidative conditions (unpublished work).[260]



Scheme 13: Under oxidative conditions the presence of m^6A of does not inhibit crosslinking between PhSeT and adenosine (unpublished work).[260]

4.2 Aims

i. Synthesis of PhSeT modified oligonucleotides

To synthesise the PhSeT nucleoside and phosphoramidite and to synthesise PhSeT modified oligonucleotides containing the modification in a range of sequence contexts.

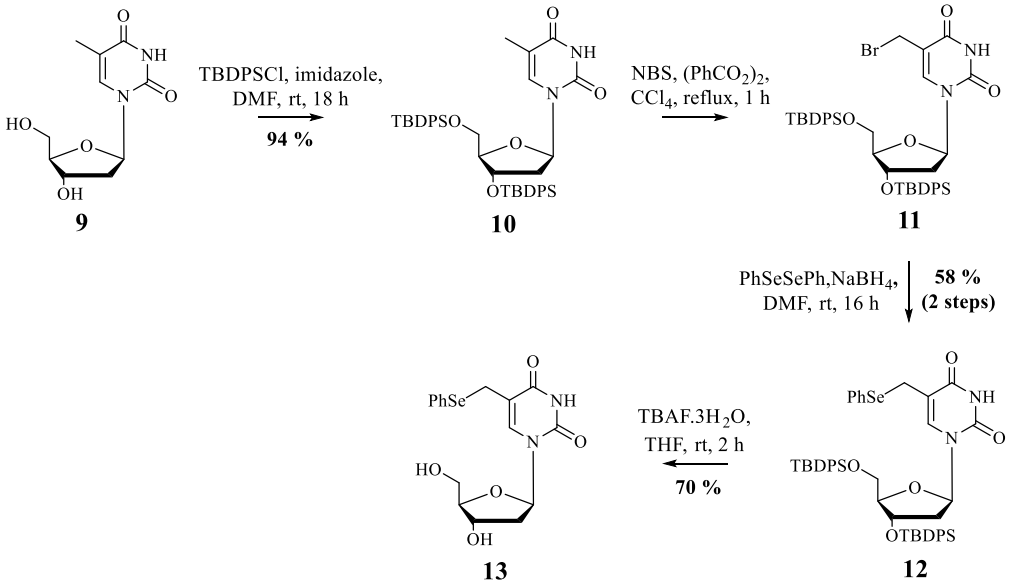
ii. Testing the selective PhSeT crosslinking in an interstrand context.

To test whether the selective crosslinking of PhSeT is maintained in an interstrand context to determine whether this chemistry could be utilised for the biochemical detection of m^6A.

4.3 Results and discussion

4.3.1 Nucleoside context crosslinking

We aimed to build on the unpublished work demonstrating that PhSeT crosslinks with adenosine but not m^6A in a nucleoside context.[260] We began by synthesising the PhSeT nucleoside **13** according to according to Hong and Greenberg (2004) (Scheme 14).[17] The 5'- and 3'- alcohols of thymidine **9** were protected with TBDPS groups to give **10**. The 5 - phenyl selenide group was then added by forming an allylic bromide intermediate **11** that was displaced with a phenyl selenide group to give **12**. The PhSeT nucleoside product **13** was produced by removal of the 5' and 3' TBDPS protecting groups using tetrabutylammonium fluoride trihydrate.

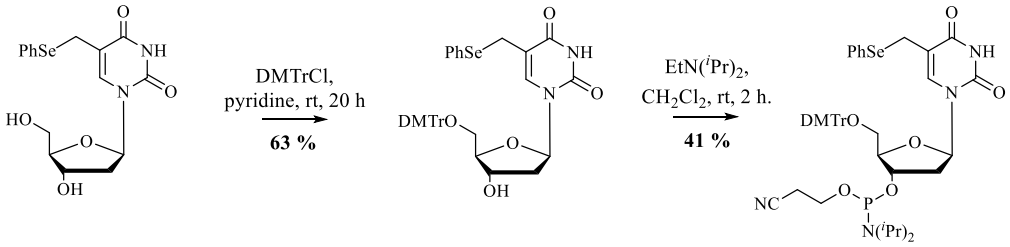


Scheme 14: Synthesis of 5- phenyl selenide thymidine nucleoside.[17]

In the unpublished work, crosslinking of PhSeT with adenosine and m^6A was analysed by Liquid chromatography mass spectrometry (LCMS). Under the oxidative conditions, no PhSeT- m^6A crosslinking was observed.[260] Instead of LCMS, we used high resolution mass spectrometry (HRMS) to determine whether even low levels of the PhSeT-m^6A crosslinked product can be observed. Using hydrogen peroxide as the oxidising agent we identified both PhSeT-adenosine and PhSeT-m^6A crosslinked products. From these results we concluded that PhSeT crosslinking is selective between m^6A and adenosine, however small amounts of the PhSeT-m^6A crosslink are formed in the reaction.

4.3.2 Synthesis of PhSeT oligonucleotides

We aimed to determine whether the selective crosslinking of PhSeT with adenosine and m^6A is maintained in an interstrand context. PhSeT and m^6A phosphoramidites were synthesised and incorporated into oligonucleotide chains. The PhSeT phosphoramidite **15** was synthesised according to Hong et al., 2005 (Scheme 15).[271] Firstly, the 5' alcohol was protected with a dimethyoxytrityl (DMTr) group **14**. The 3' alcohol was subsequently modified to form the cyanoethyl phosphoramidite **15**. The m^6A phosphoramidite was synthesised directly from the commercially available RNA BZ-A-CE-phosphoramidite (as described in section 2.3.2). We then synthesised simple oligonucleotide sequences of poly(A), poly(m^6A), and poly(T) with a central PhSeT nucleotide (Table 12).



13 **14** **15**

Scheme 15: Synthesis of 5-phenyl selenide-thymidine phosphoramidite.[271]

Table 12: Oligonucleotides for analysis of PhSeT interstrand crosslinking in a simple sequence context.

Name	**DNA/RNA**	**Sequence (5'-3')**	**Length (nt)**
Poly(A)	RNA	A_{18}	18
Poly(m^6A)	RNA	G-m^6A_{17}	18
T-PhSeT-T	DNA	T_9-PhSeT-T_8	18

4.3.2.1 Premature oxidation of the PhSeT oligonucleotide

Oligonucleotide synthesis (described in detail in section 2.1.2.1) involves an oxidation step which converts the unstable phosphite triester P(III) into a stabilised P(v) species.[21] This oxidation has the potential to prematurely oxidise the phenyl selenide group. HPLC analysis of the PhSeT modified oligonucleotides resulted in two peaks which we hypothesised were the oxidised and non-oxidised forms of the oligonucleotide (Figure 49 a). This was confirmed by reacting the oligonucleotide with H_2O_2 (10 mM), which resulted in full conversion of the oligonucleotide into the oxidised form, with

the lower retention time (Figure 49 b). The significantly different retention times allowed HPLC purification of the non-oxidised oligonucleotide (Figure 49 c).

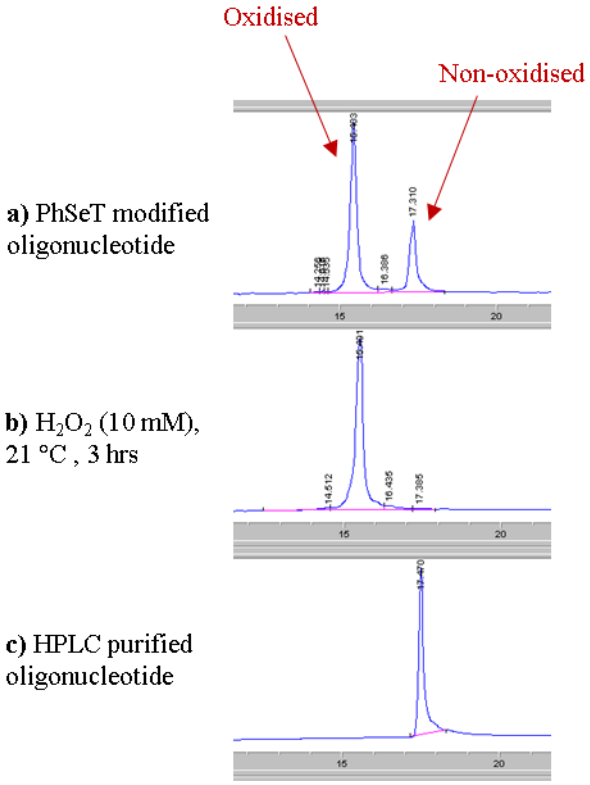
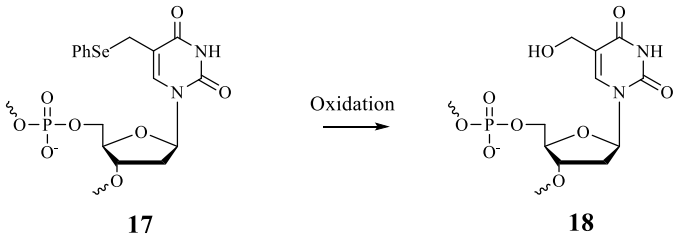
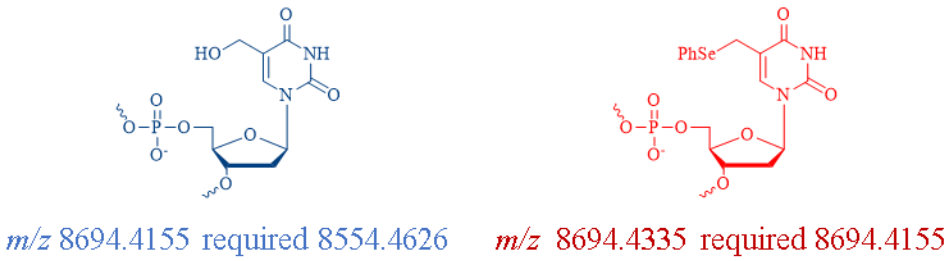


Figure 49: HPLC analysis and purification of PhSeT modified oligonucleotides. a) A PhSeT modified oligonucleotide is partially oxidised during synthesis. b) Treatment of the oligonucleotide with hydrogen peroxide confirmed that the oxidised oligonucleotide has a lower retention time. c) The non-oxidised oligonucleotide was HPLC purified.

Analysis of the oxidised product by mass spectrometry demonstrated that PhSeT is oxidised to form 5-hydroxymethyl-2‘-deoxyuridine **18** which is consistent with the published literature (Scheme 16 and Figure 50).[259]



Scheme 16: PhSeT is oxidised to form 5-hydroxymethyl-2‘-deoxyuridine.[271]



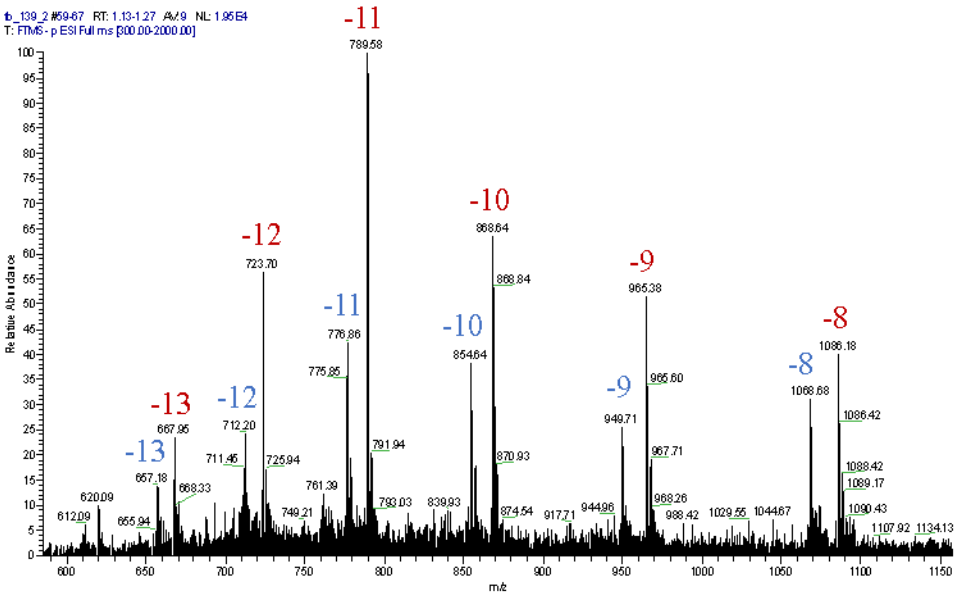


Figure 50: Mass spectrometry analysis of a PhSeT modified oligonucleotide before HPLC purification. Both the oxidised (5-hydroxymethyl-2‘-deoxyuridine) (blue) and non-oxidised (PhSeT) (red) products are observed.

We also observed complete oxidation of the oligonucleotide after treatment with ammonia hydroxide at 65 °C, for 72 hrs (Figure 51), demonstrating that

the deprotection conditions also cause premature oxidation of PhSeT. To reduce the levels of undesirable oxidation, the phenyl selenide modified oligonucleotides were deprotected by incubation with ammonia hydroxide at room temperature for 48 hrs, instead of the standard 55 °C overnight.

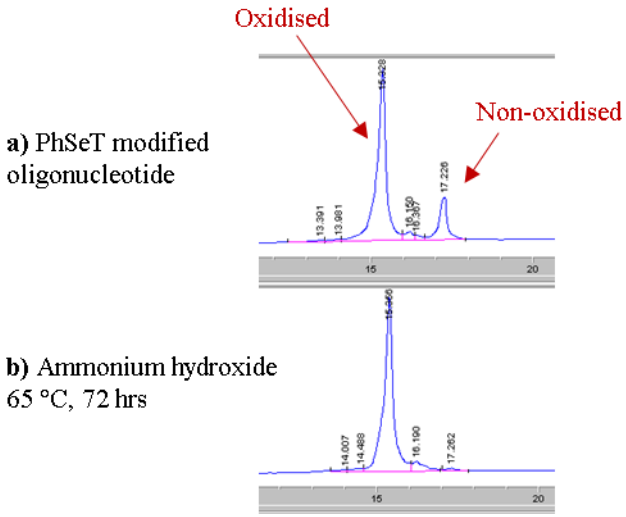


Figure 51: Oxidation of PhSeT modified oligonucleotides using ammonium hydroxide. a) A PhSeT modified oligonucleotide is partially oxidised during synthesis. b) Extended incubation in ammonium hydroxide at elevated temperatures results in complete oxidation of the oligonucleotide .

4.3.3 Selective crosslinking in simple oligonucleotide sequences

There are a number of potential oxidising agents suitable for the PhSeT crosslinking reaction. Of the oxidising agents used to crosslink double stranded DNA in the literature, sodium periodate resulted in the highest crosslinking yields.[272] Unfortunately, sodium periodate also functions in the

cleavage of vicinal diols that are present in RNA,[273] for this reason we used hydrogen peroxide as our oxidising agent.

We initially tested the crosslinking reaction in a simple poly(A)/ poly(m^6A) sequence context. The PhSeT modified oligonucleotide was radiolabelled at its 5'- end with ^{32}P and annealed to either poly(A) or poly(m^6A) oligonucleotides. The samples were treated with H_2O_2 (10 mM) for 1 hr, at 21 °C and analysed by denaturing polyacrylamide gel electrophoresis to observe the presence or absence of interstrand crosslinking. Consistent with our results from the nucleoside crosslinking experiment, crosslinking was only observed between PhSeT and adenosine (Figure 52). The lack of crosslinking between PhSeT and m^6A confirmed our hypothesis that the selective crosslinking of PhSeT is maintained in an interstrand context, indicating that it may be possible to utilise PhSeT as a method of m^6A detection.

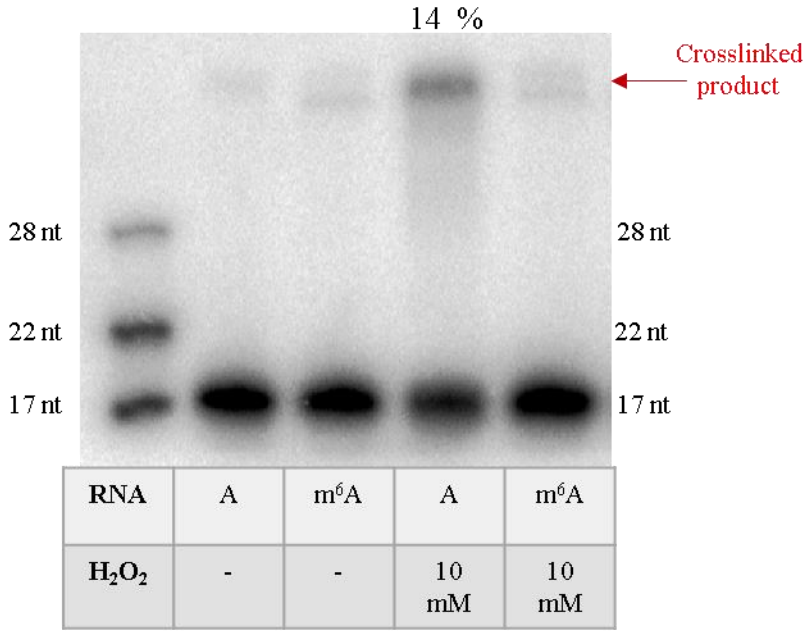
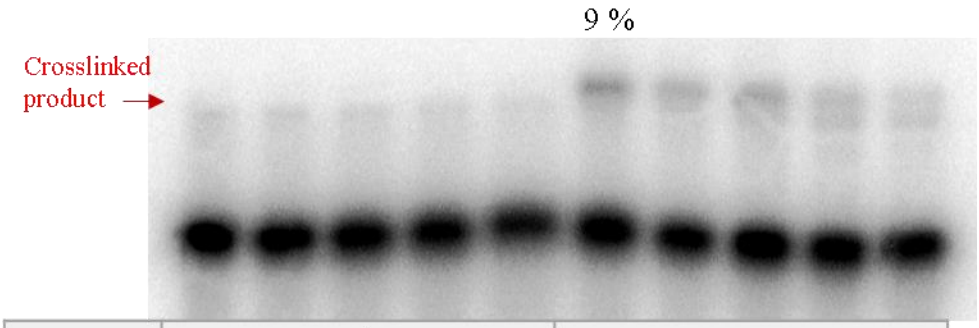


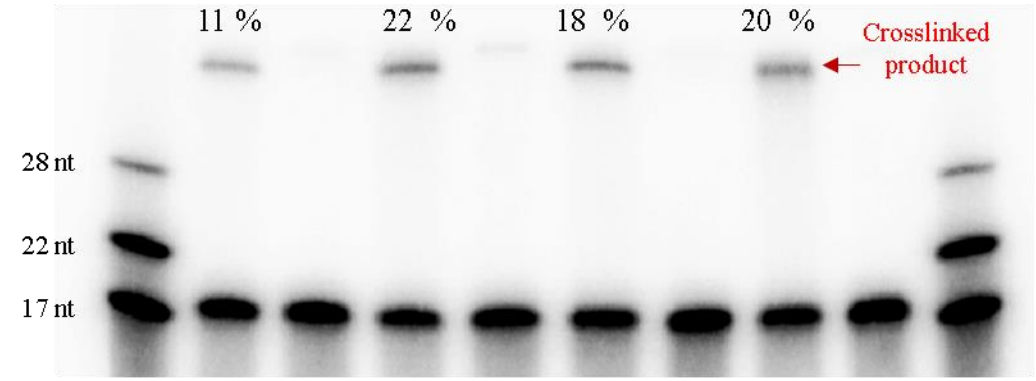
Figure 52: Interstrand crosslinking between PhSeT and poly(A)/(m^6A). Crosslinking was observed between PhSeT and adenosine in a 14 % yield. No crosslinking was observed between PhSeT and m^6A.

Only a small percentage of the PhSeT and adenosine containing oligonucleotides exhibited crosslinking. In an attempt optimise the yield of the crosslinked product the reaction was repeated using a range of hydrogen peroxide concentrations from 10 mM to 500 mM (Figure 53). The highest levels of crosslinking were observed in the 10 mM reactions and the amount of the crosslinked product decreased as the concentration of hydrogen peroxide was increased. We also tested a range of reaction temperatures and incubation times. The highest crosslinking levels of 22 % were observed after an incubation at 4 °C for 12 hrs (Figure 54).



RNA	m^6A					A				
H_2O_2 (mM)	10	20	50	100	500	10	20	50	100	500

Figure 53: Interstrand crosslinking between PhSeT and poly(A)/(m^6A). Testing the method using various concentrations of H_2O_2. The reaction was incubated at 21 °C for 1 hr. The highest crosslinking yield observed was 9 %, when 10 mM H_2O_2 was used.



RNA	A	m^6A	A	m^6A	A	m^6A	A	m^6A
Temp	4 °C				21 °C			
Time	1 hr		12 hr		1 hr		12 hr	

Figure 54: Interstrand crosslinking between PhSeT and poly(A)/(m^6A). Testing the method using various temperatures and incubation times with 10 mM H_2O_2. The highest crosslinking yield observed was 22 %.

4.3.4 Crosslinking in complex oligonucleotide sequences

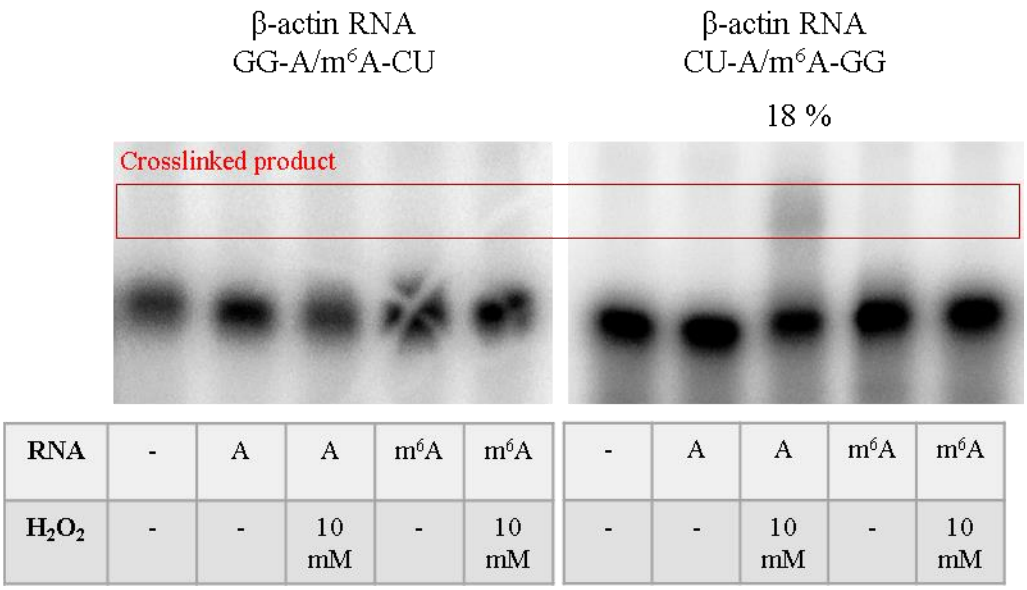
Confirmation that the selective reactivity of PhSeT is maintained in an interstrand context was highly promising. However, the oligonucleotide sequences used in this experiment were relatively simple and do not reflect natural m^6A modified mRNA. Therefore, it was important to test the crosslinking within the context of a known mRNA sequence. The zipcode of β-actin mouse mRNA was an ideal candidate to test the PhSeT crosslinking as β-actin is a highly abundant transcript with defined sites of m^6A methylation that have been mapped and quantified.[90]

We synthesised β-actin zipcode oligonucleotides that contained either adenosine or m^6A within the GG-A/m^6A-CU or CU-A/m^6A-GG sequence contexts. We also synthesised complementary DNA oligonucleotide probes with the PhSeT modification opposite the m^6A/A sites (Table 13).

Table 13: β-actin zipcode oligonucleotides modified with either PhSeT or m^6A for the analysis of PhSeT interstrand crosslinking.

Name	**DNA/RNA**	**Sequence (5'-3')**	**Length (nt)**
Mouse β-actin consensus PhSeT	DNA	TAACAG(PhseT)CCGCCT AGAAGC	28
Mouse β-actin control PhSeT	DNA	TAACAGTCCGCC(PhSeT) AGAAGC	28
Mouse β-actin unmodified	RNA	CUUCUAGGCGGACUGU UA	19
Mouse β-actin consensus m^6A	RNA	GCUUCUAGGCGG(m^6A) CUGUUA	19
Mouse β-actin control m^6A	RNA	GCUUCU(m^6A)GGCGGA CUGUUA	19

The corresponding oligonucleotides were annealed, treated with 10 mM H_2O_2 and incubated at 4 °C for 12 hrs. The selective crosslinking was maintained in the 'CU-A/m^6A-GG' sequence context with 18 % crosslinking observed between PhSeT and adenosine, and no crosslinking observed between PhSeT and m^6A. However, when analysing the m^6A consensus sequence 'GG-A/m^6A-CU' no crosslinking was observed with either m^6A or adenosine (Figure 55). This indicates that the PhSeT interstrand crosslinking is dependent on sequence context. Due to the unreliable nature of this crosslinking reaction we determined that a crosslinking based method using PhSeT will be unsuitable for m^6A detection.



RNA	-	A	A	m^{6}A	m^{6}A	-	A	A	m^{6}A	m^{6}A
H_2O_2	-	-	10 mM	-	10 mM	-	-	10 mM	-	10 mM

Figure 55: Interstrand crosslinking of PhSeT and m^{6}A modified β-actin zipcode oligonucleotides. The reactions were incubated at 4 °C for 12 hrs. Selective crosslinking was observed between adenosine and m^{6}A when in the CU-A/m^{6}A-GG sequence context. No crosslinking was observed when in the GG-A/m^{6}A-CU sequence context.

4.3.5 Discussion of the crosslinked product

There are a number of potential explanations for the inability of PhSeT to crosslink with m^{6}A. We hypothesise that the crosslinking is inhibited due to steric hindrance caused by the *N*^{6}-methyl group. Alternatively, the combination of the methyl group of m^{6}A and the phenyl selenide group of PhSeT may prevent normal base pairing between the two nucleotides, inhibiting the crosslinking reaction.

Although crosslinking was observed between PhSeT and adenosine, we have not yet determined the exact product that was formed. The crosslinking can occur at either the N^1 position **4** or the more stable N^6 position **5**.[272] In our nucleoside studies, the crosslinked product was treated with piperidine to ensure rearrangement to the more stable N^6 crosslinked product **5**. In this study piperidine was not used, in order to prevent undesirable RNA degradation,[274] therefore we have potentially formed the less stable crosslinked product **4**. In order to determine whether this is the case, NMR or HPLC analysis will be required.

4.3.6 PhSeT crosslinking as a method of mRNA purification

PhSeT crosslinking has proved unsuitable for the detection of m^6A, however, there are other potential applications for this chemistry. One possible application is to use PhSeT modified oligonucleotides for the poly(A) purification of mRNA.

The purification of mRNA from total RNA is a commonly used molecular biology technique.[19] The most common method involves the use of a poly(T) DNA probe bound to magnetic beads which base pair with to the poly(A) tail of the mRNA, allowing contaminating non-coding RNAs to be washed away.[19] A limitation to the poly(T) magnetic bead mRNA purification method is that the washing steps cannot be too severe as this will risk breaking the base pairing between the poly(dT) and poly(rA) sequences. This means that small quantities of the contaminating non-coding RNA will remain in the

sample. We hypothesise that a poly(PhSeT) probe could be used as an alternate method of mRNA purification. The covalent linkage formed between PhSeT and adenosine is considerably stronger than nucleotide base pairing. This will allow more stringent washing steps to be used which will ensure that all of the contaminating non-coding RNA is eliminated from the sample.

In order to be used as a method of mRNA purification, the PhSeT-adenosine crosslinking yields would need to be as close to 100 % as possible. This is considerably higher than the crosslinking yields we observed between the poly(A) and PhSeT containing oligonucleotides. In an attempt to improve the crosslinking yields we will synthesise a poly(T) oligonucleotide containing multiple PhSeT modifications. Efforts to develop this technology are in progress at the time of writing this thesis.

4.3.7 Concluding points

The aim of this chapter was to synthesise PhSeT modified oligonucleotides and to test whether they could be used as biological probes for the detection and quantification of m^6A. We observed selective crosslinking between PhSeT and adenosine in simple sequence contexts. Unfortunately, in more complex sequence contexts such GG-A/m^6A-CU we observed no crosslinking with either m^6A or adenosine indicating a crosslinking based method is unsuitable for m^6A detection.

Despite the lack of success using PhSeT for crosslinking based applications, there is potential for the PhSeT nucleoside to be utilised in reverse

transcription-based applications including the detection and quantification of m^6A. This possibility is explored in Chapter 5.

Chapter 5: PhSeT triphosphate based m^6A detection

Overview: Modified nucleoside triphosphates have a range of applications, including use as therapeutic drugs,[275] and in the detection of nucleotide modifications in DNA and RNA.[16,113] A 4-Se modified thymidine nucleoside triphosphate has previously been utilised for the detection of m^6A in RNA.[113] In chapter 4 we demonstrated that a PhSeT modified nucleoside selectively crosslinks with adenosine and not m^6A. We hypothesise that PhSeT and m^6A may also be incompatible in reverse transcription reactions which would allow site specific mapping of m^6A.

We have synthesised PhSeT triphosphate (PhSe-dTTP) and tested its ability to be incorporated in reverse transcription reactions opposite adenosine and m^6A in the RNA template, using a range of reverse transcriptase enzymes. We have found that the enzymes *Bst* DNA polymerase and *Bst* 3.0 DNA polymerase selectively incorporate PhSe-dTTP opposite adenosine and not m^6A. We have tested these enzymes under a range of conditions to optimise the efficiency and selectivity of this reaction. Further analysis and optimisation are required, however the selectivity of PhSe-dTTP has the potential to be utilised for transcriptome wide mapping of m^6A.

5.1 Introduction

5.1.1 Nucleoside triphosphates

Nucleoside triphosphates (NTPs) are nucleosides with three phosphate groups bonded to the 5'- of the sugar (Figure 56). These molecules are the building blocks of DNA and RNA and are the substrates of DNA replication and transcription.[21] Nucleoside triphosphates also fulfil a number of essential cellular functions. The most important of which is the provision of energy for the majority of chemical reactions in the cell as provided by ATP, a well-known example of a nucleoside triphosphate.[276]

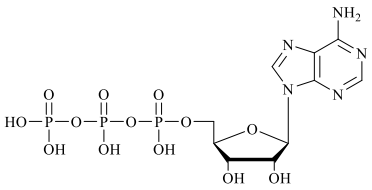


Figure 56: Structure of the nucleoside triphosphate ATP. Nucleoside triphosphates are composed of a nitrogenous base, and either a ribose or deoxyribose sugar bonded to three 5' phosphates.[21]

5.1.1.1 Nucleoside triphosphates in transcription and DNA replication

Transcription is the synthesis of single stranded RNA using a DNA template. A covalent bond is formed between the 3'- hydroxyl group of the growing RNA chain and the alpha phosphate of the incoming nucleoside triphosphate. The beta and gamma phosphate groups are released in the form of

pyrophosphate. This reaction is catalysed by an RNA polymerase enzyme (Figure 57). DNA replication occurs by a similar mechanism; however, the DNA template is used to synthesise a complementary DNA molecule, and the reaction is carried out by a DNA polymerase enzyme. Transcription uses the nucleoside triphosphates ATP, CTP, GTP, and UTP, whilst DNA replication uses the deoxynucleoside triphosphates dATP, dCTP, dGTP, and dUTP.[21]

Nucleoside triphosphates are essential for a number of molecular biology techniques including reverse transcription and polymerase chain reaction (PCR). These methods are used for the *in vitro* synthesis and amplification of RNA and DNA respectively.[277,278]

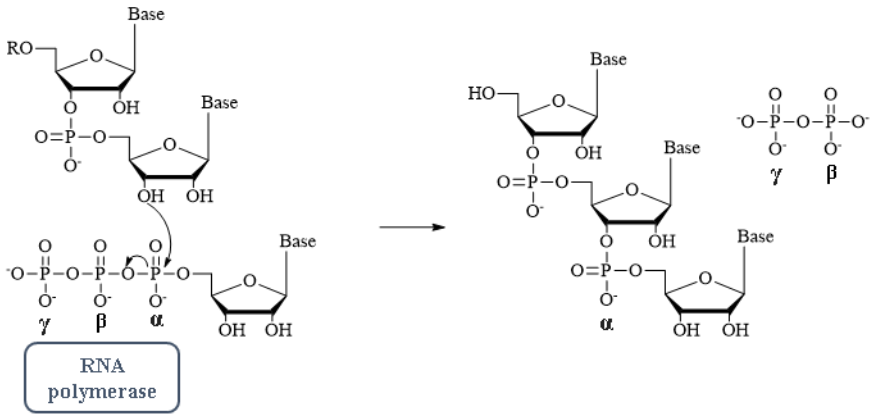


Figure 57: Mechanism of reverse transcription. RNA polymerase catalyses the formation of a covalent bond between the 3'- hydroxyl group of the growing RNA chain and the alpha phosphate of the incoming nucleoside triphosphate.[276]

5.1.2 Modified nucleoside triphosphates

Modified nucleoside triphosphates have a range of applications including use as therapeutic drugs,[275] and the sequencing of DNA and RNA.[279]. Modified

nucleoside triphosphates can be incorporated into nucleic acid using polymerase enzymes. Modified nucleoside triphosphates are incorporated less efficiently than unmodified molecules, however, the efficiency of incorporation varies greatly depending on the type of modification.[280]

5.1.2.1 Applications of modified nucleoside triphosphates

Modified nucleoside triphosphates can be used as antiviral agents. This is achieved when the nucleoside triphosphate acts as a chain terminator, which inhibits the polymerase enzymes and prevents the replication and expression of the virus's genetic information.[275] Aciclovir (Figure 58) is an example of a modified nucleoside used to treat herpes simplex virus, shingles, and chickenpox. Aciclovir mimics guanosine and is efficiently incorporated into the viral DNA. The lack of a 3' hydroxyl group in Aciclovir prevents further nucleotides from being incorporated into the growing DNA chain.[275]

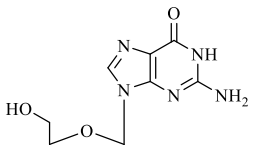


Figure 58: Aciclovir is a modified nucleoside used as an antiviral agent. Aciclovir acts as a chain terminator, inhibiting the activity of DNA polymerase enzymes.[275]

Modified nucleoside triphosphates are also used in the sequencing of DNA. An example of which is Sanger sequencing which utilises chain-terminating dideoxynucleotides (ddNTPs). These molecules lack the 3' hydroxyl group,

preventing the coupling of subsequent nucleotides to the growing chain (Figure 59). The sequence of nucleotides is mapped based on sites of stalling of each dideoxynucleotide (ddATP, ddGTP, ddCTP, and ddTTP).[279]

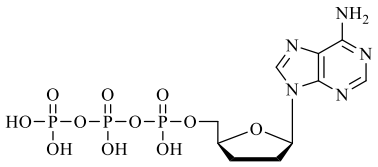


Figure 59: Structure of adenosine dideoxyadenosine triphosphate (ddATP) used in Sanger sequencing of DNA. This molecule lacks a 3' hydroxyl group, preventing the coupling of subsequent nucleotides which results in stalling of the polymerase enzyme.[279]

5.1.3 Using polymerase enzymes and modified nucleoside triphosphates for the detection of m^6A

Due to the numerous difficulties regarding the detection and quantification of m^6A, research attempting to utilise polymerase enzymes and modified nucleoside triphosphates for m^6A sequencing has been undertaken in recent years, this research is discussed below.

5.1.3.1 Polymerase dependent detection of modified nucleotides

Polymerase enzymes have the potential to be used for the detection of modified nucleotides. This is because they have sterically sensitive active sites which allow the enzyme to differentiate between modified and unmodified

nucleotides.[16] There are a number of mechanisms by which a polymerase enzyme could be used to detect a modified nucleotide:

1) **The polymerase may recognise the modified nucleotide differently from its unmodified counterpart.** For example, 8-oxoguanine (8oxoG) which is formed from guanosine by reactive oxygen species under anaerobic conditions. Certain polymerases enzymes have been shown to incorporate adenosine opposite 8oxoG instead of cytidine.[281]
2) **The polymerase enzyme may be stalled by the presence of a modified nucleoside.** An example of this is 2′-O-methylated nucleotides in ribosomal RNA which have been mapped by identifying sites of reverse transcription stalling.[282,283]
3) **The polymerase may take longer to incorporate the nucleoside triphosphate opposite the modified nucleotide.** This is how 6mA is detected in DNA using SMRT sequencing.[284]

Attempts have been made to identify polymerase enzymes that react differently with adenosine and m^6A. When using *Bst* DNA polymerase or *Tth* DNA polymerase, nucleoside triphosphates are incorporated less efficiently opposite m^6A compared with adenosine[16,285]. However, the reverse transcription efficiency is only slightly reduced and a difference in polymerase activity could only be observed in single nucleotide incorporation experiments with very short reaction times.[285]

5.1.3.2 Modified nucleoside triphosphates for the detection of m^6A

A series of 5-substituted 2’ deoxyuridine 5’ triphosphate analogues: 5-F-dUTP, 5-Br-dUTP, 5-Iodo-dUTP, 5-hmdUTP, and 5-Formyl-dUTP (Figure 60) have been tested for their ability to differentiate between adenosine and m^6A in a reverse transcription reaction. All of the triphosphates tested were incorporated opposite both adenosine and m^6A in reverse transcription reactions. A number of the triphosphates exhibited a reduced incorporation efficiency opposite m^6A. However, they only slightly improved upon the standard unmodified nucleoside triphosphates and the difference in incorporation efficiency was not sufficient for use as an m^6A detection method.[16]

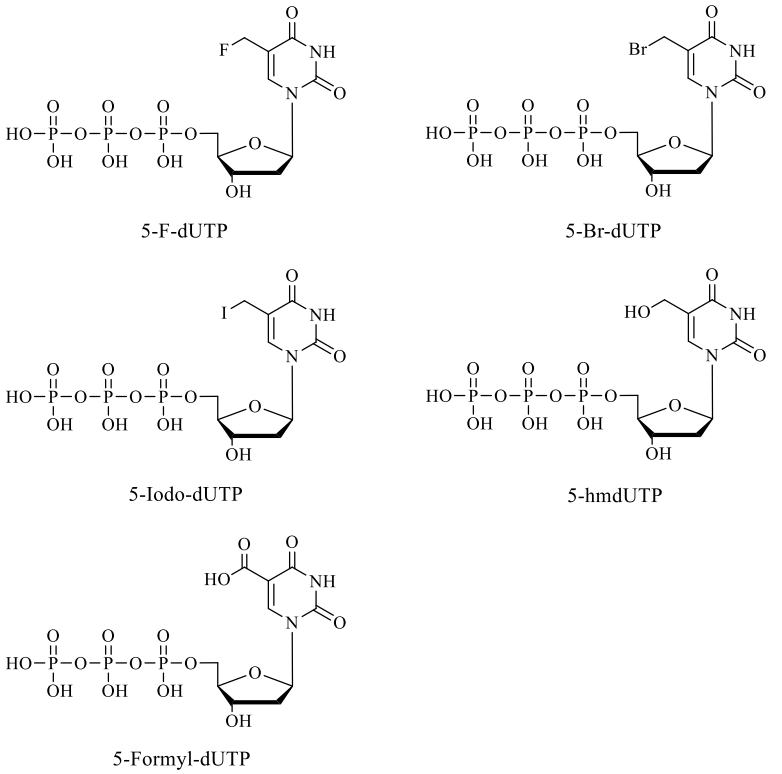


Figure 60: 5-substituted 2’ deoxyuridine 5’ triphosphate analogues: 5-F-dUTP, 5-Br-dUTP, 5-Iodo-dUTP, 5-hmdUTP, and 5-Formyl-dUTP.[16]

Hong and co-workers (2018) synthesised a modified nucleoside triphosphate (4Se-dTTP) which causes truncation of reverse transcription when opposite m^6A but not adenosine.[113] They substituted the oxygen at the 4-position of deoxythymidine triphosphate (dTTP) with a larger selenium atom (Figure 61).[113] The increased radius of the selenium atom resulted in significant truncation of reverse transcription at sites of m^6A in the RNA template, making it a promising tool for the detection of m^6A. However, the amount of truncation observed was only a fraction of the full-length product.[113] There is a requirement for further modified nucleoside triphosphates to be tested to evaluate their ability to differentiate between m^6A and adenosine.

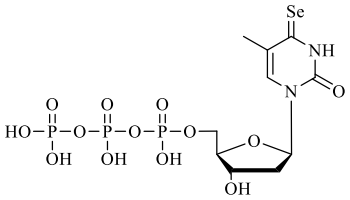


Figure 61: Structure of 4Se-dTTP, used for the detection of m^6A in RNA.[113]

5.1.4 PhSeT triphosphate may differentiate between m^6A and adenosine

In chapter 4, we demonstrated that the PhSeT nucleoside is able to selectively crosslink with adenosine but not m^6A in both mononucleotide and interstrand contexts. This chemistry is not suitable to be developed as an m^6A detection method. However, the inability of PhSeT to crosslink with m^6A led us to question whether a PhSeT modified nucleoside triphosphate (Figure 62) could be used to differentiate between m^6A and adenosine. We hypothesised that reverse transcription would terminate opposite m^6A when PhSe-dTTP is used in place of dTTP, which would provide an ideal method of m^6A detection that could potentially be adapted for transcriptome wide m^6A sequencing.

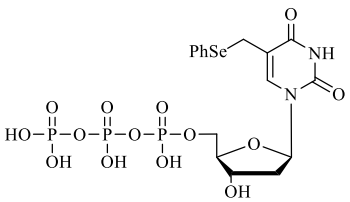


Figure 62: Structure of phenyl selenide modified thymidine triphosphate.[286]

5.2 Aims

i. Synthesis of the PhSeT nucleoside triphosphate

To synthesise and purify the PhSeT nucleoside triphosphate using the method described by Greenberg et al. (2006),[286] and to assess alternative methods of PhSe-dTTP synthesis in an attempt to increase the reaction yield.

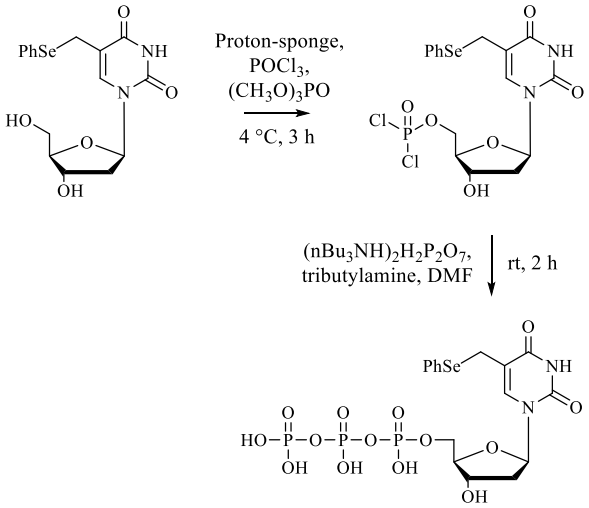
ii. Reverse transcription of PhSe-dTTP using m^6A and adenosine templates

To test the ability of PhSe-dTTP to be incorporated opposite m^6A and adenosine in the RNA template using a range of reverse transcriptase enzymes, and to determine whether this can be used for the detection and quantification of m^6A.

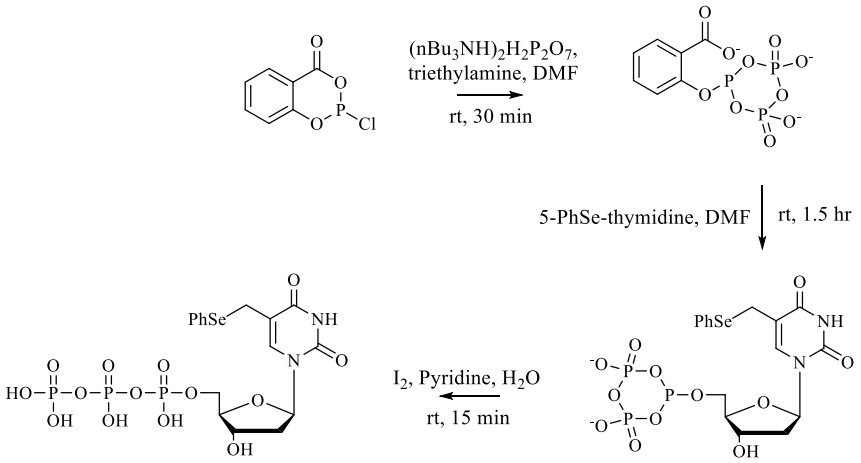
5.3 Results and discussion

5.3.1 Synthesis of PhSeT triphosphate

Synthesis of nucleoside triphosphates has historically proved to be very difficult.[287] Thankfully Hong et al. (2006) have successfully synthesised PhSeT triphosphate (PhSe-dTTP) using the Yoshikawa method,[288] in a low 3 % yield.[286] Using their methodology, we converted the PhSeT nucleoside into its triphosphate form (Scheme 17). A laborious, multistep purification was required in order to remove the numerous contaminants that are inhibitory to the reverse transcription reaction. The product was first purified by DEAE sephadex ion exchange chromatography, followed by purification using C-18 reverse phase HPLC. The identity and purity of the PhSe-dTTP was confirmed by HPLC and mass spectrometry. We observed low yields using this method and the two-step purification made synthesis of this molecule inefficient and time consuming. For this reason, we also synthesised PhSe-dTTP using a Ludwig Eckstein based method (Scheme 18)[113,289] which required a single HPLC purification step and resulted in significantly improved yields.



Scheme 17: Synthesis of 5-phenyl selenyl-thymidine-triphosphate (Yoshikawa method).[286,288]



Scheme 18: Synthesis of 5-phenyl selenyl-thymidine-triphosphate (Ludwig Eckstein method)[113,289].

5.3.2 PhSe-dTTP can be used as a substrate for primer extension reactions

For this experiment to be successful the reverse transcriptase must recognise PhSeT as its natural substrate thymidine. The Klenow exo-fragment of DNA polymerase I has previously been shown to incorporate PhSe-dTTP opposite adenosine in a DNA template, however, with an incorporation efficiency 150 times lower than unmodified dTTP.[286]

Our initial aim was to test whether PhSe-dTTP can be used as a substrate in reverse transcription reactions using an unmodified RNA template. Using the MuLV reverse transcriptase, we observed no difference in primer extension when using PhSe-dTTP and dTTP with the primer extending to the full length of the RNA template in both cases (Figure 63).

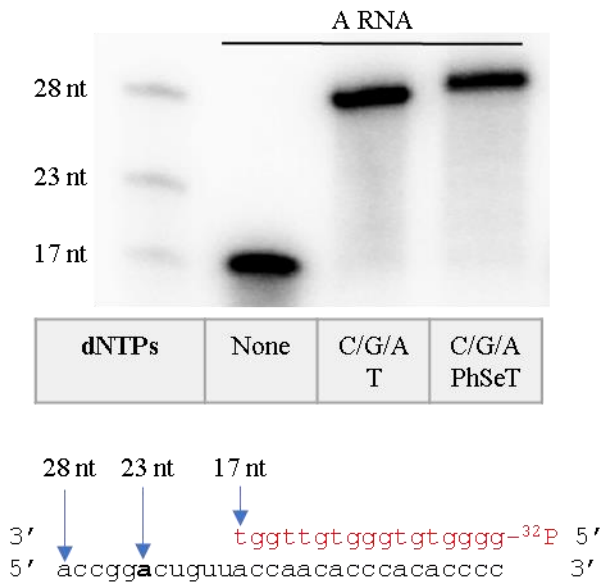


Figure 63: Testing PhSe-dTTP on an unmodified RNA template. Using MuLV reverse transcriptase, full length primer extension was observed when using both dTTP and PhSe-dTTP.

5.3.3 Determining the selectivity of PhSe-dTTP towards m^6A and adenosine

In order to determine whether PhSe-dTTP exhibits selectivity between m^6A and adenosine, we synthesised two identical RNA oligonucleotides, 28 nt in length, containing either m^6A or adenosine at position 23. Using a radiolabelled 17 nt primer we tested a series of reverse transcriptase enzymes to determine their ability to incorporate PhSe-dTTP opposite m^6A and adenosine. The majority of the reverse transcriptase enzymes showed no selectivity between m^6A and adenosine, with MuLV RT, AffinityScript Multiple Temperature RT, OPTIZYME, and ImProm-II all extending the full

length of the template when using the m^6A modified and unmodified templates (Figure 64). However, *Bst* DNA polymerase showed a high degree of selectivity between m^6A and adenosine, with the majority primer extension stalling opposite m^6A in the RNA template (Figure 65).

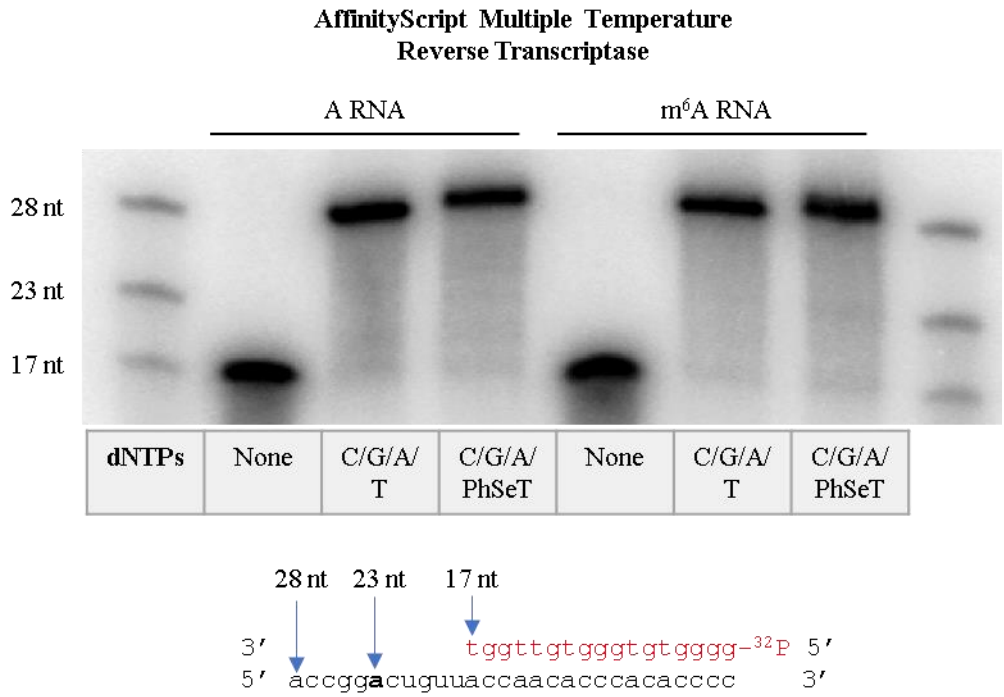


Figure 64: PhSe-dTTP was tested using a range of reverse transcriptase enzymes. AffinityScript Multiple Temperature Reverse Transcriptase was one the enzymes that exhibited no selectivity between m^6A and adenosine, with the primer extending the full length of the RNA template in both cases.

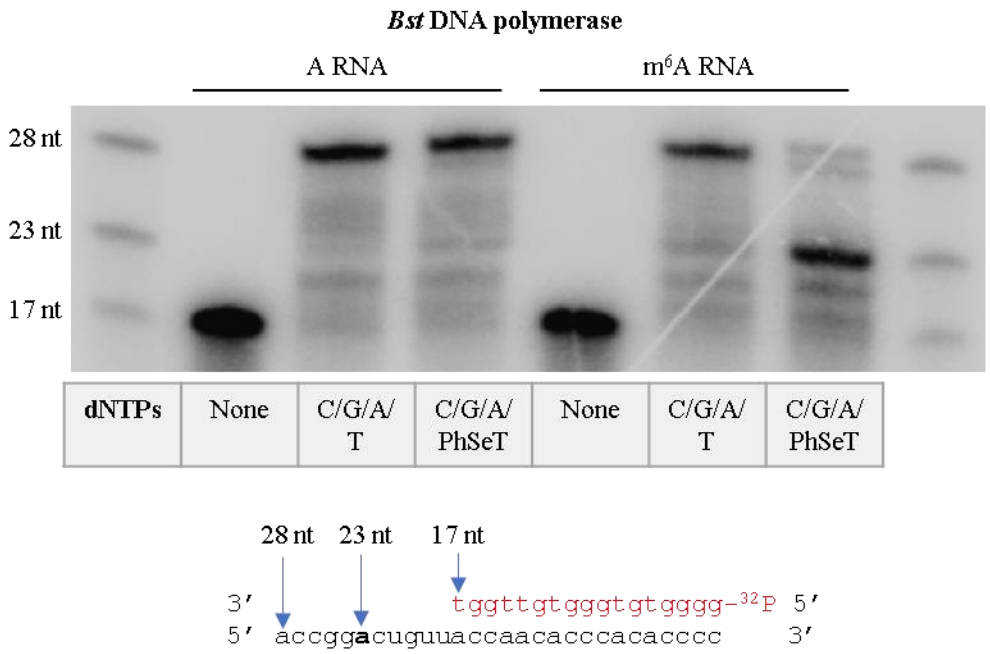


Figure 65: Using *Bst* DNA polymerase and PhSe-dTTP, the reverse transcription stalled opposite the m^{6}A but not adenosine.

5.3.4 Testing m^{6}A quantification using PhSe-dTTP and *Bst* DNA polymerase

We observed high levels of reverse transcription stalling opposite m^{6}A when using a combination of PhSe-dTTP and *Bst* DNA polymerase. To test whether this stalling could be used to quantitatively determine the relative levels of m^{6}A and adenosine at a single nucleotide position, we repeated the experiment using varying ratios of the m^{6}A modified and unmodified RNA templates (Figure 66). We observed a linear relationship between the amount of primer extension stalling and the ratio of m^{6}A modified and unmodified input RNA, however, at less than 10 % m^{6}A the amount of reverse transcription stalling did not accurately reflect m^{6}A levels.

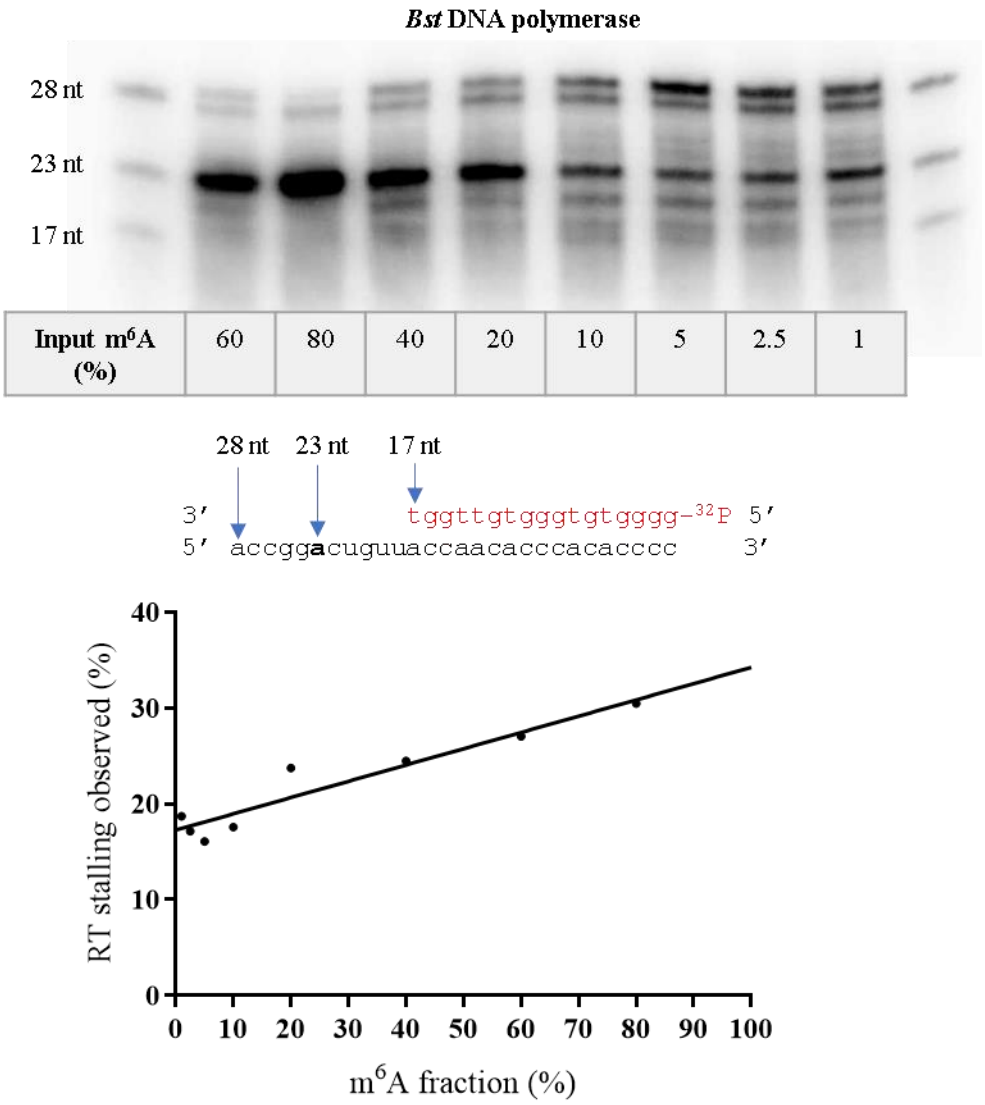


Figure 66: Primer extension using PhSe-dTTP and *Bst* DNA polymerase using different ratios of the m^{6}A modified and unmodified RNA template. We observed a linear relationship between the amount of primer extension stalling and the ratio of m^{6}A modified and unmodified input RNA.

5.3.5 *Bst* 3.0 DNA polymerase

Bst DNA polymerase is primarily a DNA dependent DNA polymerase, however it does possess some reverse transcriptase activity, as is observed in our primer extension experiments. For this reason, *Bst* DNA polymerase is

unsuitable for a transcriptome wide reverse transcription-based method. An alternative enzyme is *Bst* 3.0 DNA polymerase which is a genetically engineered version of *Bst* DNA polymerase with significantly improved reverse transcriptase activity.[290,291] We tested this enzyme using the PhSeT primer extension assay and observed high levels of stalling opposite the m^{6}A site in the RNA template (Figure 67). However, we also observed low intensity stalling bands opposite the unmodified adenosine when using both dTTP and PhSe-dTTP.

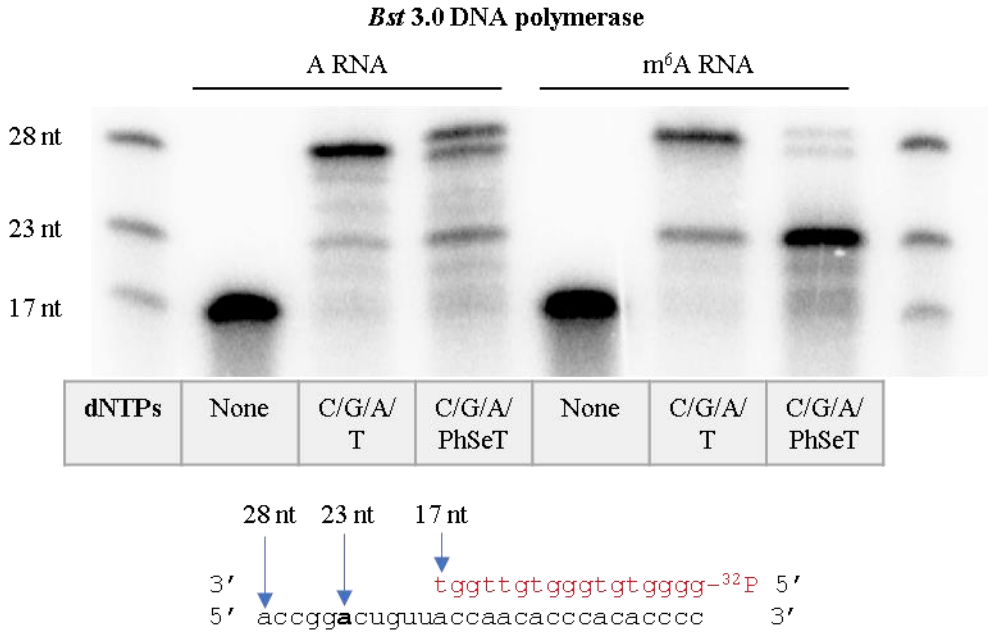


Figure 67: Using *Bst* 3.0 DNA polymerase, a large amount of stalling was observed opposite m^{6}A using PhSe-dTTP. Low intensity stalling bands were also observed opposite adenosine when using both dTTP and PhSe-dTTP.

We therefore attempted to optimise the reverse transcription reaction conditions to maximise stalling opposite m^{6}A residues whilst eliminating stalling opposite adenosine. We tested a range of reaction temperatures (Figure 68), and buffer components (Figure 69), and concluded that the

optimal conditions were using 4 mM $MgSO_4$, and a reaction temperature of 60 °C. under these conditions, we observed high levels of reverse transcription stalling opposite m^6A with minimal stalling opposite adenosine.

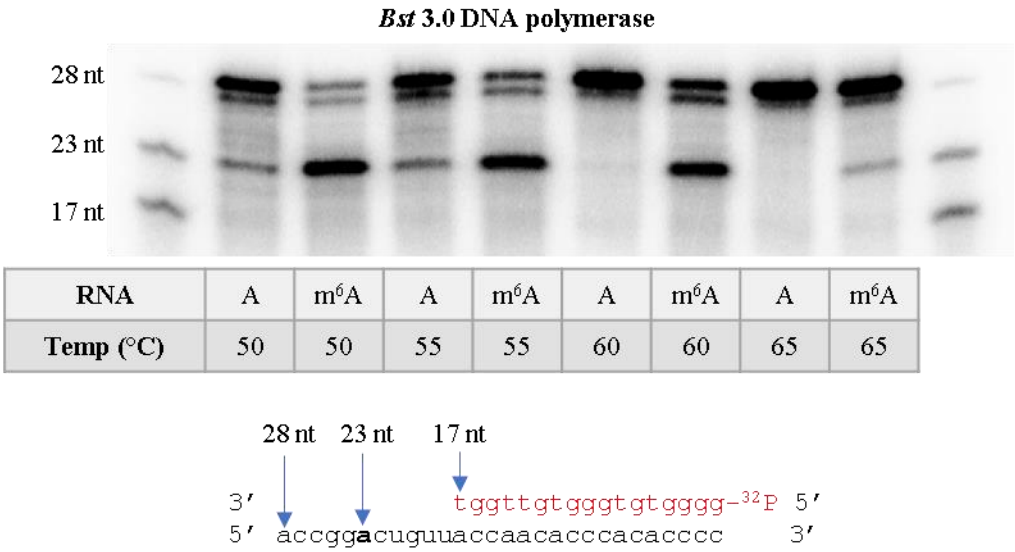


Figure 68: Temperature optimisation for the PhSe-dTTP *Bst* 3.0 polymerase primer extension reaction. We identified 60 °C as the optimal temperature with high levels of stalling opposite m^6A and full-length primer extension using the unmodified RNA template.

RNA	A	m^{6}A	A	m^{6}A	A	m^{6}A	A	m^{6}A
Temp (°C)	60	60	60	60	60	60	60	60
$MgSO_4$ (mM)	4	4	6	6	8	8	10	10

Figure 69: $MgSO_4$ optimisation for the PhSe-dTTP *Bst* 3.0 polymerase primer extension reaction. When using a reaction temperature of 60 °C, we identified 4 mM $MgSO_4$ as the optimal concentration for selective reverse transcription, with the high levels of stalling opposite m^{6}A and minimal stalling opposite adenosine.

Using these optimised conditions, we tested the reverse transcription reaction with a range of incubation times (Figure 70) and found that a 20 min incubation time was required to observe mostly full-length primer extension using the unmodified RNA template. Increasing the incubation time from 20 mins to 1 hr slightly reduced the amount of stalling at sites of m^{6}A.

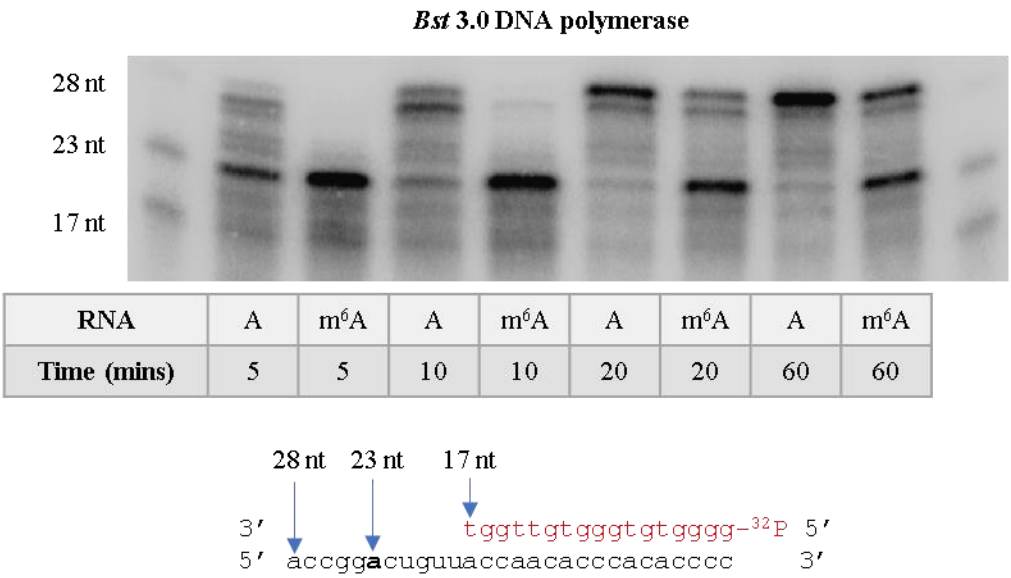


Figure 70: Reverse transcription incubation time. The amount of reverse transcription stalling opposite both m^{6}A and adenosine decreased as the reaction time was increased.

We tested the ability of *Bst* 3.0 DNA polymerase to quantify m^{6}A levels by using different ratios of the unmodified and m^{6}A modified RNA templates (Figure 71). As the amount of m^{6}A in the input RNA increased we observed higher levels of reverse transcription stalling, however, the amount of reverse transcription stalling did not accurately reflect the ratio of the input oligonucleotide. Additionally, the levels of reverse transcription stalling were not as high as in the quantification experiment using the standard *Bst* DNA polymerase (Figure 66, section 5.3.4).

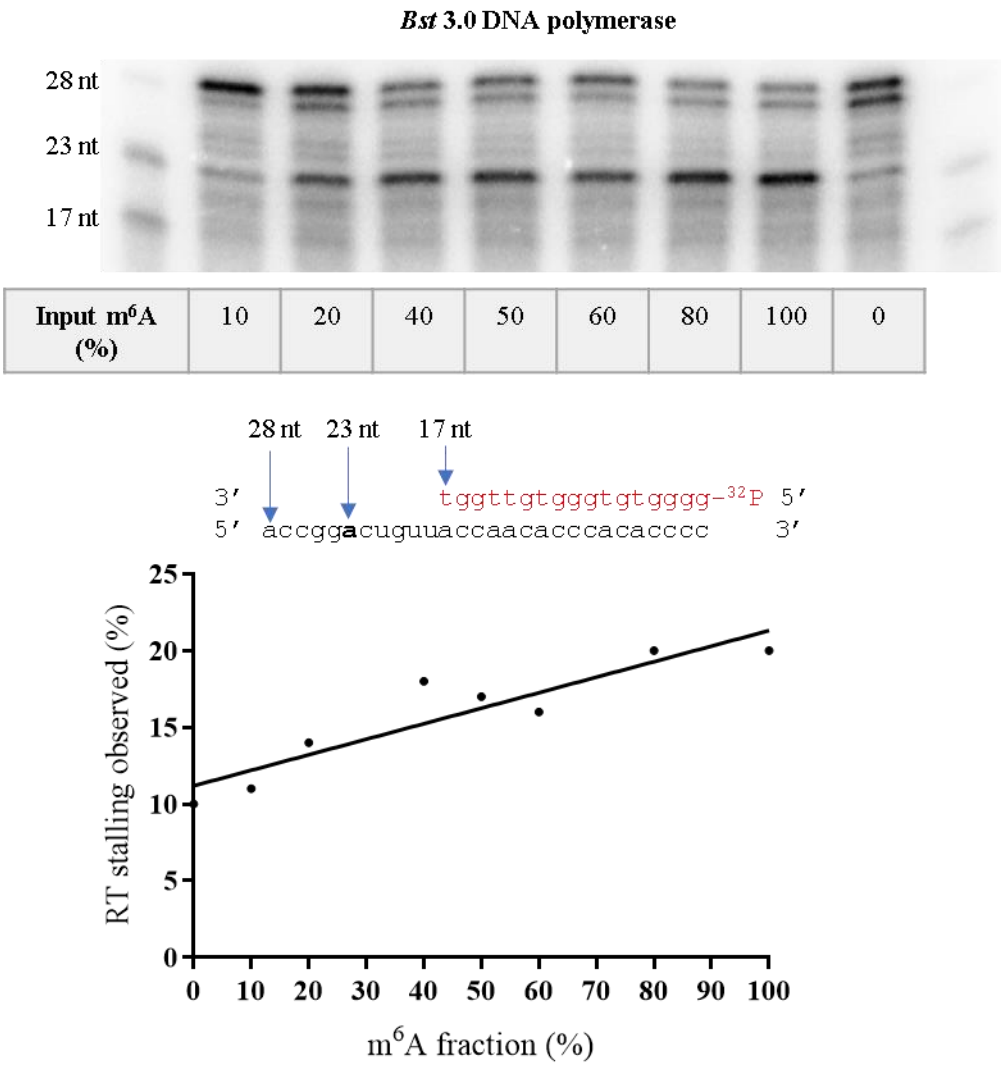


Figure 71: Testing quantification of m^{6}A levels using *Bst* 3.0 DNA polymerase. The amount of reverse transcription stalling did not accurately reflect the ratio of the input oligonucleotide.

One explanation for the low accuracy of this quantification experiment is the high reaction temperature of 60 °C. In this experiment, any DNA primer that stalls at an m^{6}A site will be 23 nucleotides in length and will have a melting temperature of approximately 58 °C. As the primer has a lower melting temperature than the reaction temperature, the partially extended primers will

be able to re-anneal with an unmodified RNA oligonucleotide and extend to the full length of the RNA template.

A solution to this problem would be to use longer template RNA oligonucleotides however synthesis of RNA oligonucleotides exceeding 40 nucleotides in length is challenging. Alternatively, the template RNA can be *in vitro* transcribed however using this method it is not possible to have both adenosine and m^6A nucleotides in the same RNA template.

5.3.6 Testing PhSe-dTTP on real mRNA samples

Having demonstrated that PhSe-dTTP selectively stalls reverse transcription at m^6A sites using synthetic RNA templates, we needed to test the molecule on a real mRNA sample. To determine at which nucleotides reverse transcription stalling was occurring, we designed the experiment outlined in Figure 72. Firstly a 5' biotinylated poly(dT) primer was annealed to poly(A)+ mRNA and the primer was extended in a reverse transcription reaction. The mRNA was then treated with RNase T1, which cleaves at the 3'- of every G nucleotide (however, any RNA that is annealed to the extended DNA strand would be protected from the RNase treatment). The DNA/RNA hybrids were then purified using streptavidin magnetic beads and 5' radiolabelled with ^{32}P. Finally, the RNA was digested into 5'-^{32}P-nucleoside monophosphates and analysed by two-dimensional TLC. This experiment should in theory determine the identity of the nucleotides at which reverse transcription stalling occurs.

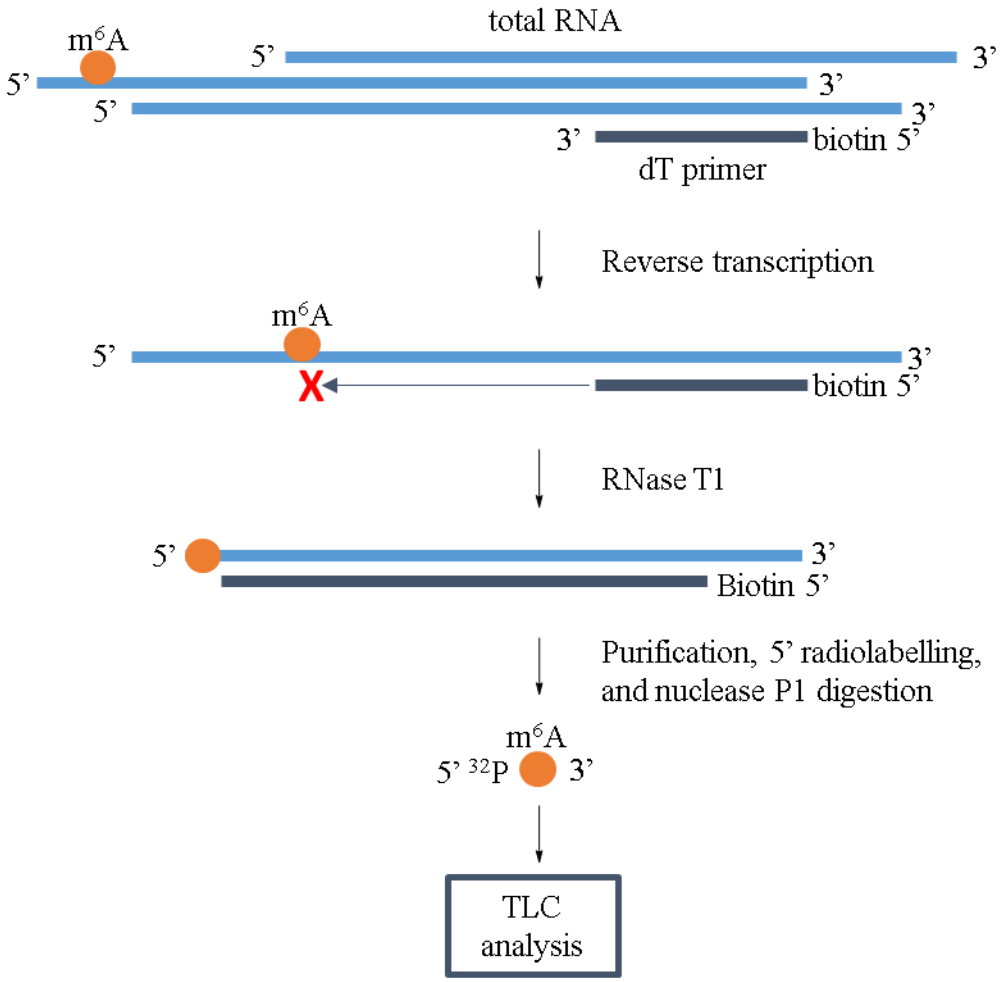


Figure 72: Determining the nucleoside identity of sites of reverse transcription stalling. A Biotinylated poly(dT) primer was annealed to an mRNA sample and extended in a reverse transcription reaction. The mRNA was then digested using RNase T1, purified using the biotin tag, radiolabelled, and digested into 5'- ^{32}P nucleoside monophosphates, and analysed by two-dimensional TLC.

The optimum temperature for m^{6}A detection using PhSe-dTTP and *Bst* 3.0 DNA polymerase is 60 °C. The melting temperature of a 17-nucleotide poly(dT) oligonucleotide primer is approximately 47 °C, which is too low for the reverse transcription reaction. To combat this problem, we used dT locked bases in the poly(dT) primer in order to increase the temperature of the 17-nucleotide primer to approximately 64 °C. Locked nucleic acid contains a 2'-

O, 4'-C methylene bridge that increases the rigidity of the ribofuranose ring which stabilises the base pairing and reduces the melting temperature of the oligonucleotide (Figure 73).[292]

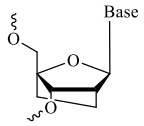


Figure 73: Structure of a locked base.[292]

We tested this method on chicken embryo total RNA using *Bst* 3.0 DNA polymerase and PhSe-dTTP in place of dTTP. Initially we only observed unincorporated phosphate (^{32}P) on the TLC plate. This was because the high melting temperature of the DNA/RNA hybrid prevented elution of the radiolabelled DNA from the RNA template when heating the sample to 95 °C. To overcome this problem, we eluted the DNA by treating the DNA/RNA hybrid with RNase H which digested the RNA strand into smaller fragments. After this digestion step, the DNA could then be eluted from the magnetic beads by incubating at 75 °C for 5 mins.

Using this method, we observed ^{P}A, ^{P}C, and ^{P}U 5' nucleoside monophosphates on the TLC plate, however, no spot for ^{P}m^{6}A was observed (Figure 74). This demonstrated that using the conditions tested, reverse transcription did not predominantly stall at sites of m^{6}A. Interestingly, the spots for A, C and U have similar intensities indicating that the reverse transcription stalling is not predominantly at sites of adenosine which appears to demonstrate that PhSe-dTTP is working efficiently in the reverse transcription reaction, as the primer must be extending beyond the poly(A) tail. Future work will involve repeating

this experiment using a greater range of reverse transcription conditions and with suitable experimental controls, both of which were not possible in this study due to time constraints.

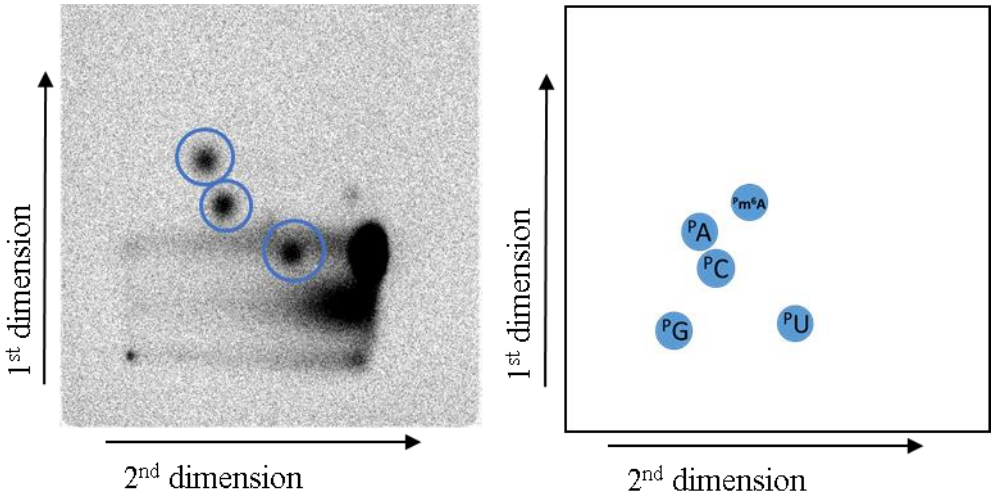


Figure 74: TLC analysis of PhSe-dTTP reverse transcription using *Bst* 3.0 DNA polymerase, and chicken embryo total RNA. Spots corresponding to ^{P}A, ^{P}C, and ^{P}U were present on the TLC plate with similar intensities. No spot for ^{P}m^{6}A was observed.

5.3.7 Comparison with other reverse transcription based m^{6}A detection methods

A number of studies have used reverse transcriptase enzymes and modified nucleoside triphosphates for the detection of m^{6}A[16,113,285]. The most successful of these methods is 4Se-dTTP which also stalls reverse transcription reactions specifically at m^{6}A sites in the RNA template.[113] For both 4Se-dTTP and PhSe-dTTP, the reverse transcription inhibition is not at 100 %, with a fraction

of the reverse transcription reactions extending past m^6A in the RNA template. This is potentially beneficial as it will allow the detection of multiple closely located m^6A sites, which would not be possible if the reverse transcription stalling was 100 %. However, quantification of the amount of methylation at a particular site will be more accurate with higher levels of reverse transcription stalling at sites of m^6A. There is room for further reaction optimisation in order to maximise the selective reverse transcription stalling at sites of m^6A.

It is likely that other modified nucleoside triphosphates will also selectively inhibit reverse transcription at m^6A sites. Future work should therefore involve testing a range of modified nucleoside triphosphates for the ability to differentiate between m^6A and adenosine. A combination of the 4-Se and 5-PhSe modifications may also improve the selectivity between m^6A and adenosine.

A further potential improvement would be to test Tth DNA polymerase with both 4Se-dTTP or PhSe-dTTP. *Tth* is a DNA polymerase from *Thermus thermophilus* that possess reverse transcriptase activity with an 18-fold selectivity for incorporating thymidine opposite adenosine in comparison with m^6A.[16]. The high cost of *Tth* DNA polymerase prevented us from testing this enzyme in our study.

5.3.8 Concluding points

The aims of this chapter were to synthesise PhSe-dTTP and to test its ability to differentiate between m^6A and adenosine in reverse transcription reactions. We successfully synthesised PhSe-dTTP and identified *Bst* DNA polymerase

and *Bst* 3.0 DNA polymerase which both exhibited selectivity between m^6A and adenosine when using PhSeTTP as a substrate. Bst 3.0 DNA polymerase is the most promising candidate due to its superior reverse transcriptase activity, however this technology will require significant optimisation in order to be a useful m^6A detection method.

Chapter 6: Conclusions

Overview: This chapter discusses the findings made in this thesis and the main conclusions reached. Future work that will follow this study is also discussed.

6.1.1 Conclusion

In order to conclude this thesis, it is important to review the aims we initially set out to meet. These were to synthesise various modified oligonucleotides as part of collaborative research projects and to develop new methods of m^6A detection and quantification. We have met these aims with varied levels of success, an evaluation of our respective outcomes and shortcomings are discussed in this chapter.

6.1.1.1 Synthesis of alkylated oligonucleotides

Chapter 2 describes the one step synthesis of m^6A phosphoramidite. This is considerably more efficient than the seven-step synthesis described in the literature.[198] The m^6A phosphoramidite is commercially available, however, at a cost of over £4,000/g,[293] this is considerably more expensive than our synthesis which costs well under £100/g. We believe our method of synthesis is suitable for researchers with little to no experience with synthetic organic chemistry and should improve the access of research groups to m^6A modified oligonucleotides.

This alkylation reaction can also be used for the synthesis of a range of other modified nucleoside phosphoramidites as demonstrated in unpublished work from the Hayes research group.[294] In recent years, many new modified nucleotides have been discovered in mRNA, and there are over a hundred modified nucleotides present in non-coding RNA with unknown functions.[49] The synthesis of oligonucleotides containing these modifications may prove to be invaluable research tools. Future work will involve determining whether

the i^6A nucleoside phosphoramidite is compatible with oligonucleotide synthesis chemistry.

6.1.1.2 RedBaron method of m^6A detection

We developed the RedBaron method as an alternative to SCARLET[90] for the site-specific detection and quantification of m^6A. This method was able to accurately detect and quantify m^6A within synthetic oligonucleotide sequences, however we found that the method significantly underrepresented m^6A levels when detecting m^6A within real mRNA samples.

The RedBaron and SCARLET methods are both time consuming and labour intensive, and their inability to detect more than one m^6A site per experiment makes them unsuitable for high throughput m^6A sequencing. The main benefit of the SCARLET method is the ability to accurately quantify m^6A which is not currently possible using other methods. As the RedBaron method is unable to quantify m^6A levels it will have very limited use as an m^6A detection method.

In this study we used the SCARLET method to detect and quantify a number of m^6A sites. Despite what was reported in the literature[90] we observed very low radioactive signals using this method and in the majority of attempts we were unable to distinguish clear spots for either adenosine or m^6A on the TLC plate. We were only able to observe clear signals when analysing highly abundant mRNA transcripts such as the 28S rRNA.

In the literature there are very few examples of the SCARLET method being used. Of the few examples, all were analysed using one-dimensional TLC.[91]

However, we ran our SCARLET TLCs in two-dimensions and a large number of spots were observed that would not be seen when using a one-dimensional TLC. For this reason, we suggest that SCARLET experiments should always analysed in two-dimensions, despite the fact that this will increase the time and labour intensity of SCARLET experiments.

6.1.1.3 PhSeT crosslinking

The PhSeT nucleoside crosslinking method showed promising selectivity between adenosine and m^6A, however unfortunately we found the crosslinking reaction to be unreliable, with no crosslinking observed in certain sequence contexts. Importantly no crosslinking was observed in the 'GGACU' sequence context which is most common sequence motif of m^6A. For this reason, we decided that a crosslinking based method was unsuitable for m^6A detection, however, with increased time and resources it would be interesting to test this selective crosslinking method using a wider range of reaction conditions.

A further problem with the crosslinking method is that even with optimal crosslinking yields, PhSeT would only crosslink with the unmodified adenosine. This gives the method limited use for m^6A detection as a lack of crosslinking at a specific site could not be assumed to be caused by the m^6A modification, as other factors may also cause inhibition of the crosslinking reaction, as was observed with the sequence dependent crosslinking inhibition. Future work will involve testing the PhSeT crosslinking as a method of poly(A)+ mRNA purification.

6.1.1.4 PhSe-dTTP reverse transcription

The PhSe-dTTP reverse transcription based m^6A detection method shows great promise to be developed as a new technology for the transcriptome wide sequencing and quantification of m^6A. However, further improvements and tests are required to determine whether this method will be successful.

A number of research groups have attempted to develop reverse transcription based methods of m^6A detection by testing a range of reverse transcriptase enzymes and modified nucleoside triphosphates.[16,113,285] The majority of these methods exhibited limited ability to differentiate between adenosine and m^6A. However, 4Se-dTTP showed high levels of reverse transcription stalling specifically at sites of m^6A in the RNA template.[113] This molecule is currently the benchmark for reverse transcription based m^6A detection. Our initial studies appear to demonstrate that PhSe-dTTP results in greater selectivity between m^6A and adenosine than 4Se-dTTP. However, due to time limitations we were unable to test the reverse transcription in more than one sequence context.

Both the PhSe-dTTP and 4Se-dTTP molecules have not yet been tested on other modified adenosine nucleosides such as A_m, m^6A_m, and m^1A so it is possible that reverse transcription stalling will also occur at these nucleotides in the RNA template. If this is the case, then these methods will be unsuitable for differentiating between these modified nucleotides. As stalling is observed at m^6A it is likely that stalling will also be observed opposite m^6A_m due to the similar structures of these molecules.

At the time of submission of this thesis we are in the process of testing the PhSe-dTTP method on a range of mRNA samples and are in the process of trying to develop this method for a transcriptome wide sequencing of m^6A.

6.1.2 Final comments

There is a great demand for new methods of m^6A mapping and quantification with numerous attempts made in recent years with varied levels of success. Although many of the methods developed in this thesis have proved unsuccessful, we hope they will provide a foundation for the development of new methods of m^6A sequencing in the future. We are pleased that the PhSe-dTTP molecule shows promise, and hope that with improvements, this technology can be developed as a method of m^6A detection and quantification.

Chapter 7: Materials and methods

Overview: This chapter provides an account of the materials, and research methods used throughout this thesis.

7.1 Materials

7.1.1 Tissue and RNA samples

- Mouse tissues (heart, liver, testis, kidney, and brain) from five-month-old male mice (C57) were supplied by Catrin Rutland.
- Total RNA from 5-day old chicken embryos was supplied by Zsuzsanna Bodi.
- DF1 chicken fibroblasts cells were provided by Simon Welham.
- Human total RNA was purchased from Agilent Technologies: 'Total RNA, HeLa-S3 Cells, Human, 25 ug'

7.1.2 Kits, equipment, TLC, and PAGE

- QIAquick Nucleotide Removal Kit (QIAGEN)
- Riboprobe® System - T7 *in vitro* transcription kit (Promega)
- TLC Cellulose F Glass plates (20 x 20 cm) (Sigma Aldrich)
- Thin polythene sleeves to accommodate 20 × 20 cm glass TLC plates for imaging
- Kodak Imaging Screen-K
- Bio-Rad Molecular Imager FX Phospho scanner
- Bio-Rad "Quantity One" analysis software version 4.6.3.
- 15% Mini-PROTEAN® TBE-Urea Gel, 10 well, 30 µl (Bio-Rad)
- SYBR™ Gold Nucleic Acid Gel Stain (Invitrogen™)

- Mini-PROTEAN Tetra Cell (Bio-Rad)
- Mini Trans-Blot Module (Bio-Rad)
- GeneScreen Plus Hybridization Transfer Membrane
- Hydrophilic Streptavidin Magnetic Beads (NEB)
- Oligo d(T)25 Magnetic Beads (NEB)
- Phenomenex Clarity 3 µ Oligo-RP 50 × 4.6 mm column
- Agilent 1100 series HPLC
- Agilent Technologies, Chemstation for LC software
- Thermo Scientific: LTQ FT Ultra Mass Spectrometer

7.1.3 Enzymes, buffers and reagents

All enzymes, buffers, and reagents were purchased from either: New England Biolabs (NEB) Ltd (Hitchin, Thermo Fisher Scientific (Loughborough, Leicestershire, UK), Hertfordshire, UK), VWR International Ltd (Lutterworth, Leicestershire, UK), Sigma-Aldrich Company Ltd (Gillingham, Dorset, UK),

- ATP, [γ-32P]- 3000Ci/mmol 10mCi/ml Lead, 250 µCi, 25 µl (Perkin, Elmer)
- CTP, [α-32P]- 3000Ci/mmol 10mCi/ml , 250 µCi, 25 µl (Perkin, Elmer)
- T4 Polynucleotide Kinase (10,000 - units/ml) (NEB)
- FastAP Thermosensitive Alkaline Phosphatase (1 U/µL) (Thermo Fisher Scientific)
- RNase H (5 U/µL) (NEB)
- T4 DNA Ligase (400 U/µL) (NEB)

- SplintR® Ligase (25 U/µL) (NEB)
- RNase T1 (1000 U/µL) (Thermo Fisher Scientific)
- RNase A, DNase and protease-free (10 mg/mL) (Thermo Fisher Scientific)
- Nuclease P1 from Penicillium citrinum (≥ 200 units/mg, 0.1 mg/mL) (Sigma Aldrich)
- RNase I (10 U/µL) (Thermo Fisher Scientific)
- Phosphodiesterase II from bovine spleen (≥ 5.0 units/mg, 1 mg/mL) (Sigma Aldrich)
- Micrococcal Nuclease (2000 units/µL) (NEB)
- ATP (for ligation)
- Purified BSA (NEB)

7.1.4 Oligonucleotides purchased

Oligonucleotides were purchased from either Eurofins Genomics Ltd (Ebersberg, Germany) or Sigma-Aldrich Company Ltd (Gillingham, Dorset, UK) and were either HPSF, HPLC, or PAGE purified.

Table 14: Sequences of purchased oligonucleotides. Nm = 2′-OMe-modified nucleotide, UPPERCASE = DNA, lowercase = RNA

Name	**Sequence (5’ - 3’)**
ssDNA RedBaron (40 nt)	GGACTTGACGACTTACGGACCTTACGGA CTCACCCATCGG

ssDNA SCARLET (116 nt)	GGACTTGACGACTTACGGACCTTAAAAG ATTAATTTAAAATTTATCAAAAAGAGTA TTGACTTAAAGTCTAACCTATAGGATAC TTACAGCCATCGCTTACGGACTCACCCA TCGG
RedBaron - Chimera – Mouse - β-actin (GGm^{6}ACU)	CmGmCmAmGmCmUmCmAmGmUmAmAm CmAmGmTCCGCmCm
RedBaron - Splint - Mouse - β-actin (GGm^{6}ACU)	GTCCGTAAGTCGTCAAGTCCTCCGCCTA GAAGCACTTGCG
SCARLET - Chimera - Mouse - β-actin (GGm^{6}ACU)	GmUmCCGCCmUmAmGmAmAmGmCmAm CmUmUmGmCmGmGmUmGmCmAm
SCARLET - Splint - Mouse - β-actin (GGm^{6}ACU)	AAACGCAGCTCAGTAACAGTCCGATGGG TGAGTCCGTAAG
RedBaron - Chimera - Chicken - β-actin (GGm^{6}ACU)	GmUmGmGmGmUmGmUmUmGmGmUmA mAmCmAmGmUCCGGmUm
RedBaron - Splint - Chicken - β-actin (GGm^{6}ACU)	GTCCGTAAGTCGTCAAGTCCTCCGGTTT AGAAGCATTTGC
RedBaron - Chimera - Chicken - β-actin (CCm^{6}ACA)	GmTmTmTmCmAmTmCmAmCmAmGmGm GmGmTmGmTGGGTmGm
RedBaron - Splint - Chicken - β-actin (CCm^{6}ACA)	GTCCGTAAGTCGTCAAGTCCTGGGTGTT GGTAACAGTCCG
SCARLET - Splint - Chicken - β-actin (GGm^{6}ACU)	GTGTGGGTGTTGGTAACAGTCCGATGGG TGAGTCCGTAAG
RedBaron - Chimera - Human - β-actin (GGm^{6}ACU)	CmAmAmCmUmAmAmGmUmCmAmUmAm GTCCGmCmC

RedBaron - Splint - Human - β-actin (GGm^6ACU)	GTCCGTAAGTCGTCAAGTCCTCCGCCTAGAAGCATTTGCG
SCARLET - Chimera - Human - β-actin (GGm^6ACU)	CmAmUmAmGmUmCCGCCmUmAmGmAmAmGm
SCARLET - Splint - Human - β-actin (GGm^6ACU)	TAACGCAACTAAGTCATAGTCCGATGGGTGAGTCCGTAAG
RedBaron - Chimera - Human - MALAT1 site 2515	AmGmTCCTUmCmAmCmAmUmUmUmUmUmCmAmAmAm
RedBaron - Splint - Human - MALAT1 site 2515	GTCCGTAAGTCGTCAAGTCCTCCTTCACATTTTTCAAACT
SCARLET - Chimera - Human - MALAT1 site 2515	mAmCmGmAmAmAmGmUCCTTmCmAmCmAmUmU
SCARLET - Splint - Human - MALAT1 site 2515	ATTACTTCCGTTACGAAAGTCCGATGGGTGAGTCCGTAAG
RedBaron - Chimera - Human - 28S rRNA A4190	CmGmTTACCmGmTmTmTmGmAmCmAmGmGmTmGmTm
RedBaron - Splint - Human - 28S rRNA A4190	GTCCGTAAGTCGTCAAGTCCTTACCGTTTGACAGGTGTAC
RedBaron - Chimera - Human - 28S rRNA A4189	AmCmCmUmGmCmGmUmTACCGmUmUmUmGmAmCmAm
RedBaron - Splint - Human - 28S rRNA A4189	GTCCGTAAGTCGTCAAGTCCTACCGTTTGACAGGTGTACC

SCARLET - Chimera - Human - 28S rRNA A4190	Same oligonucleotide as ‘RedBaron - Chimera - Human - 28S rRNA A4189’
SCARLET - Splint - Human - 28S rRNA A4190	TCGCCTTAGGACACCTGCGTCCGATGGG TGAGTCCGTAAG
PCR primer Chicken β-actin zipcode T7 *in vitro* transcription template - Forward	TAATACGACTCACTATAGGAGTACGATG AGTCCGGCCCCTC
PCR primer Chicken β-actin zipcode T7 *in vitro* transcription template - Reverse	GCGCAAGTTAGGTTTTGTCAAAGAAAG
In vitro transcription template for 3’ radiolabelled A/m^6A	CGTCTAGTCCATGTACGTACCCTATAGT GAGTCGTATTA
T7 promoter primer	TAATACGACTCACTATAG

7.1.5 Oligonucleotide synthesis

7.1.5.1 Nucleoside phosphoramidites

- 5'-(4,4'-Dimethoxytrityl)-*N*-benzoyl-2'-deoxyadenosine, 3'-[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite
- 5'-(4,4'-Dimethoxytrityl)-*N*-acetyl-2'-deoxycytidine, 3'-[2-cyanoethyl)-*N*,*N*-(diisopropyl)]-phosphoramidite
- 5'-(4,4'-Dimethoxytrityl)-*N*-dimethylformamidine-2'-deoxyguanosine, 3'- [(2-cyanoethyl)-*N*,*N*-diisopropyl]-phosphoramidite
- 5'-(4,4'-Dimethoxytrityl)-thymidine, 3'-[(2-cyanoethyl)-*N*,*N*-diisopropyl]- phosphoramidite

- 5'-(4,4'-Dimethoxytrityl)-*N*-benzoyl-adenosine, 2'-O-TBDMS-3'- [(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite
- 5'-(4,4'-Dimethoxytrityl)-*N*-acetyl-cytidine, 2'-O-TBDMS-3'-[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite
- 5'-(4,4'-Dimethoxytrityl)-*N*-dimethylformamidine-guanosine, 2'- O-TBDMS-3'-[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite
- 5'-(4,4'-Dimethoxytrityl)-uridine, 2'-O-TBDMS-3'-[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite
- 5'-(4,4'-Dimethoxytrityl)-*N*-benzoyl-adenosine, 2'-O-methyl-3'- [(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite
- 5'-(4,4'-Dimethoxytrityl)-*N*-acetyl-cytidine, 2'-O-methyl-3'-[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite
- 5'-(4,4'-Dimethoxytrityl)-*N*-dimethylformamidine-guanosine, 2'- O-methyl-3'-[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite
- 5'-(4,4'-Dimethoxytrityl)-uridine, 2'-O-methyl-3'-[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite
- 1-(4,4'-Dimethoxytrityloxy)-2-(*N*-biotinyl-4-aminobutyl)-propyl- 3-O-[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite

7.1.5.2 Phosphoramidite solid supports

- Ac-dC SynBase™ CPG 1000/110 (0.2 µmol)
- dT SynBase™ CPG 1000/110 (0.2 µmol)
- Bz-A RNA SynBase™ CPG 1000/110 (0.2 µmol)
- Ac-C RNA SynBase™ CPG 1000/110 (0.2 µmol)
- dmf-G RNA SynBase™ CPG 1000/110 (0.2 µmol)

- 2'-OMe-A RNA SynBase™ CPG 1000/110 (0.2 µmol)
- 2'-OMe-dmf-G RNA SynBase™ CPG 1000/110 (0.2 µmol)
- 2'-OMe-U RNA SynBase™ CPG 1000/110 (0.2 µmol)
- 1-(4,4'-Dimethoxytrityloxy)-3-O-(*N*-biotinyl-3-aminopropyl)-triethyleneglycolyl- glyceryl-2-O-succinyl-lcaa-CPG 1000Å (0.2 µmol)

7.1.5.3 Solvents and liquid reagents

- Anhydrous acetonitrile
- Deblock Mix (3 %, w/v trichloroacetic acid in dichloromethane)
- BTT Activator (0.3M 5-benzylthio-1-H tetrazole in acetonitrile, anhydrous)
- Oxidiser (0.02M iodine in THF/pyridine/water = 7:2:1)
- Cap Mix A (THF/pyridine/acetic anhydride = 8:1:1)
- Cap Mix B (10 %, w/v methylimidazole in THF)

7.1.5.4 Oligonucleotides synthesised (sequences)

Table 15: Oligonucleotides synthesised. Nm = 2′-OMe-modified nucleotide, UPPERCASE = DNA, lowercase = RNA, m^6A = RNA, A_m = RNA, m^6A_m = RNA, n = random RNA nucleotide.

Oligonucleotide	**Sequence (5' – 3')**
Cap-1	aacugaucaucaucaccacagagcaggu-biotin
Cap-2	A_m-acugaucaucaucaccacagagcaggu-biotin

Cap-3	m^6A_m-acugaucaucaucaccacagagcaggu-biotin
1agAgc	gugagucuuc-nnnn-a-nnnn-uccugcagcg
2ga6cg	guggaucuuc-nnnn-m^6A-nnnn-uccucgagcg
3ag6gc	gugagucuuc-nnnn-m^6A-nnnn-uccugcagcg
4gaAcg	guggaucuuc-nnnn-a-nnnn-uccucgagcg
5N9	gugagucuuc-nnnnnnnnn-uccugcagcg
Sec-TCA-1-1-truncated	ugcaggcuucaaccugua
Chicken β-actin – unmodified	accggacuguuaccaacacccacacccc
Chicken β-actin – m^6A modified 1	accgg-m^6A-cuguuaccaacacccacacccc
Chicken β-actin – m^6A modified 2	accggacuguuacca-m^6A-cacccacacccc
Chicken β-actin – m^6A modified 3	accggacuguuacc-m^6A-acacccacacccc
Chicken β-actin – m^6A modified 4	accggacuguuaccaacaccc-m^6A-cacccc
Poly(A)	aaaaaaaaaaaaaaaaa
Poly(m^6A)	m^6A-m^6A-m^6A-m^6A-m^6A-m^6A-m^6A-m^6A-m^6A-m^6A-m^6A-m^6A-m^6A-m^6A-m^6A-m^6A-g
Poly(T/PhSeT)	TTTTTTTTT-PhSeT-TTTTTTTT
Poly(T)	TTTTTTTTTTTTTTTTT
Mouse β-actin - unmodified	gcuucuaggcggacuguua

Mouse β-actin – m^6A modified 1	gcuucuaggcgg-m^6A-cuguua
Mouse β-actin – m^6A modified 2	gcuucu-m^6A-ggcggacuguua
Mouse β-actin – PhSeT modified 1	CAGTAACAG-PhseT-CCGCCTAGAAGCACTTGC
Mouse β-actin – PhSeT modified 2	CAGTAACAGTCCGCC-PhSeT-AGAAGCACTTGC
SCARLET - Chimera – Chicken β-actin – A1	GmUmGmUmUmGmGmUmAmAmCmAmGmUm-CCGG-UmUm
SCARLET chimera – chicken β-actin – A4	GmGmGmUmGmUmGmGmGmUmGmUm-TGGT-AmAm
SCARLET chimera – chicken β-actin – A6	CmAmUmCmAmCmAmGmGmGmGmUmGmUm-GGGT-GmUm
RedBaron chimera – chicken β-actin – A4	GmGmGmGmUmGmUmGmGmGmUmGm-TTGG-UmAm
RedBaron chimera – chicken β-actin – A7	UmUmUmCmAmUmCmAmCmAmGmGmGmGm-TGTG-GmGm

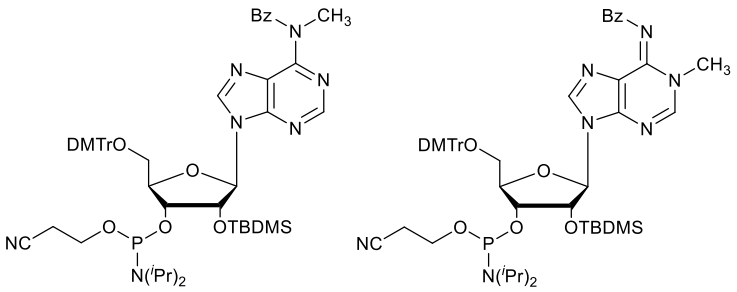
7.2 Methods

7.2.1 General methods for chemical synthesis

All starting materials and solvents were purchased from commercial suppliers. Monitoring of reactions was observed by TLC using TLC Silica gel 60 F_{254} aluminium sheets, visualized with ultraviolet light and subsequently stained with either a basic solution of potassium permanganate or an acidic solution of vanillin in methanol. All solvents were removed using a Büchi rotary evaporator. Fluorochem silica gel 60 Å was used as the stationary phase for column chromatography. NMR spectra were obtained using either a Bruker AV400, a Bruker AV3400, or a Bruker DPX300 spectrometer, and were ran using dilute sample solutions in $CDCl_3$, or CD_3OD. Coupling constants are reported in Hertz (Hz). Multiplicities have been labelled s (singlet), d (doublet), t (triplet), q (quartet), sep (septet), m (multiplet), br (broad), app (apparent). Hydrogen atoms of the sugar ring are designated by the superscript prime mark. Hydrogen atoms of the base lack the superscript prime mark. The labels H_A and H_B have been used where there are two hydrogen atoms attached to a sugar ring carbon. NMR spectra assignments were made based on two-dimensional NMR spectroscopy; COSY, and HSQC, and HMBC. Mass spectra were recorded on a Bruker MicroTOF spectrometer.

7.2.2 Chemical synthesis

7.2.2.1 5′-(4,4′-Dimethoxytrityl)-6-*N*-benzoyl-6-*N*-methyl-adenosine,2′-O-TBDMS-3′- [(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite and 5′-(4,4′-Dimethoxytrityl)-6-*N*-benzoyl-1-methyl-adenosine,2′-O-TBDMS-3′- [(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite



5'-(4,4'-Dimethoxytrityl)-*N*-benzoyl-adenosine, 2'-O-TBDMS-3'- [(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite (148 mg, 150 μmol) was dissolved in anhydrous dichloromethane (1.3 mL). Iodomethane (38 μL, 606 μmol) was added under an argon atmosphere. The solution was stirred for 5 mins at room temperature. Tetrabutylammonium bromide (50 mg, 156 μmol) and aq. NaOH (1 M, 1.3 mL) were added. The solution was stirred vigorously for 30 mins at room temperature. Water (50 mL) and diethyl ether (50 mL) were added and the layers were separated. The aqueous layer was extracted with diethyl ether (3 x 25 mL). The organic phase was dried over anhydrous

Na_2SO_4 and the solvent was evaporated *in vacuo*. The crude material was purified by silica gel column chromatography (1:3 ethyl acetate/ petrol to 9:1 ethyl acetate/methanol) to yield product A (104 mg, 69 %) and product B (36 mg, 24 %) as white powders.

Product A

δ_H (400 MHz, $CDCl_3$) 8.47 (0.5H, s, H-2), 8.45 (0.5H, s, H-2), 8.16 (0.5H, s, H-8), 8.14 (0.5H, s, H-8), 7.51 – 7.40 (4H, m, Ar-H), 7.41 – 7.31 (4H, m, Ar-H), 7.33 – 7.19 (4H, m, Ar-H), 7.20 – 7.12 (2H, m Ar-H), 6.87 – 6.78 (4H, m, Ar-H), 6.06 (0.5H, d, *J* 6.7 Hz, H-1'), 6.00 (0.5H, d, *J* 6.6 Hz, H-1'), 5.06 – 4.95 (1H, m, H-2'), 4.47 – 4.33 (2H, m, H-3', H-4'), 4.02 – 3.86 (1H, m, CH_2CH_2CN), 3.85 – 3.75 (9H, m, NCH_3, 2 x $ArOCH_3$), 3.72 – 3.48 (4H, m, H_A-5', CH_2CH_2CN, $CH(CH_3)_2$), 3.36-3.28 (1H, m, H_B-5′), 2.75 – 2.57, (1H, m, CH_2CH_2CN), 2.40 – 2.22, (1H, m, CH_2CH_2CN), 1.25 – 1.15 (9H, m, $CH(CH_3)_2$), 1.07 (3H, d, *J* 6.6 Hz, $CH(CH_3)_2$), 0.75 (9H, s, $SiC(CH_3)_3$), -0.04 (1.5H, s, $Si(CH_3)_2$), -0.07 (1.5H, s, $Si(CH_3)_2$), -0.29 (1.5H, s, $Si(CH_3)_2$), -0.30 (1.5H, s, $Si(CH_3)_2$).

δ_C (101 MHz, $CDCl_3$) 172.1 (CO), 158.6 (ArC), 155.0 (C-6), 155.0 (C-6), 152.8 (C-4), 151.8 (C-2), 144.6 (ArC), 144.4(ArC), 142.5(CH-8), 142.5 (CH-8), 136.2 (ArC), 135.7 (ArC), 135.7(ArC), 135.5 (ArC), 135.5 (ArC), 130.5 (ArC), 130.2 (ArC), 130.1 (ArC), 130.1 (ArC), 130.0 (ArC), 128.7 (ArC), 128.2 (ArC), 128.1 (ArC), 127.9 (ArC), 127.9 (ArC), 127.9 (ArC), 127.0 (ArC), 126.8 (C-5), 126.7 (C-5), 117.6 (CN), 117.3 (CN), 113.2 (ArC), 113.2 (ArC), 88.1 (CH-1′), 87.9 (CH-1′), 86.8 (CAr_3), 86.6 (CAr_3), 84.4 (CH-4′), 84.04 (d, *J* 4.0 Hz, CH-4'), 75.07 (d, *J* 2.9 Hz, CH-2'), 74.40 (d, *J* 6.0 Hz, CH-

2’), 73.46 (d, *J* 10.2 Hz, CH-3’), 72.65 (d, *J* 14.6 Hz, CH-3’) , 63.26 (CH2-5′), 63.12 (CH2-5′), 58.84 (d, *J* 16.8 Hz, CH_2CH_2CN), 57.56 (d, *J* 20.6 Hz, CH_2CH_2CN), 55.3 ($ArOCH_3$), 55.2 ($ArOCH_3$), 43.43 (d, *J* 13.1 Hz, $CH(CH_3)2$), 42.95 (d, *J* 12.5 Hz, $CH(CH_3)_2$), 35.9 (NCH_3), 25.6 ($SiC(CH_3)_3$), 25.6 ($SiC(CH_3)_3$), 24.8 ($CH(CH_3)_2$), 24.7 ($CH(CH_3)_2$), 24.6 ($CH(CH_3)_2$), 24.6 ($CH(CH_3)_2$), 20.46 (d, *J* 6.5 Hz, CH_2CH_2CN), 20.09 (d, *J* 7.0 Hz, CH_2CH_2CN), 17.9 ($SiC(CH_3)_3$), 17.8 ($SiC(CH_3)_3$), -4.6 ($Si(CH_3)_2$), -4.7 ($Si(CH_3)_2$), -5.2 ($Si(CH_3)_2$).

δ_P (162 MHz, $CDCl_3$) 151.20, 149.02.

HRMS (ESI+) calculated for $C_{54}H_{70}N_7O_8P1Si$ (M+H)+ 1002.4709, found 1002.4688.

Product B

δ_H (400 MHz, $CDCl_3$) 8.21 – 8.51 (2H, m, ArH), 7.87 (0.5H, s, H-8), 7.85 (0.5H, s, H-8), 7.84 (0.5H, s, H-2), 7.82 (0.5H, s, H-2), 7.54 – 7.39 (4H, m, ArH), 7.38 – 7.30 (4H, m, ArH), 7.30 – 7.13 (4H, m, ArH), 6.86 – 6.76 (4H, m, ArH), 5.93 (0.5H, d, *J* 6.1 Hz, H-1’), 5.86 (0.5H, d, *J* 6.0 Hz, H-1’), 4.91-4.86 (0.5H, dd, *J* 4.9, 6.0 Hz, H-2’), 4.83 (0.5H, dd, *J* 4.3, 6.2 Hz, H-2’), 4.40 (0.5H, app q, *J* 3.6 Hz, H-4’), 4.36 – 4.23 (1.5H, m, H-3’, H-4’), 4.02 – 3.82 (1H, m, CH_2CH_2CN), 3.79 (3H, m, $ArOCH_3$), 3.78 (3H, m, $ArOCH_3$) 3.73 (3H, s, NCH_3), 3.68 – 3.50 (4H, m, CH_2CH_2CN, H_A-5′, $CH(CH_3)_2$), 3.43 (0.5H, dd, J 10.5, 3.3, H_A-5’), 3.24 (1H, 3.30-3.19, H_B-5’), 2.65 (1H, t, *J* 6.6 Hz, CH_2CH_2CN), 2.34-2.20 (1H, m, CH_2CH_2CN), 1.22 – 1.14 (9H, m, $CH(CH_3)_2$), 1.02 (3H, d, *J* 6.7 Hz, $CH(CH_3)_2$), 0.85 – 0.79 (9H, s, $SiC(CH_3)_3$),

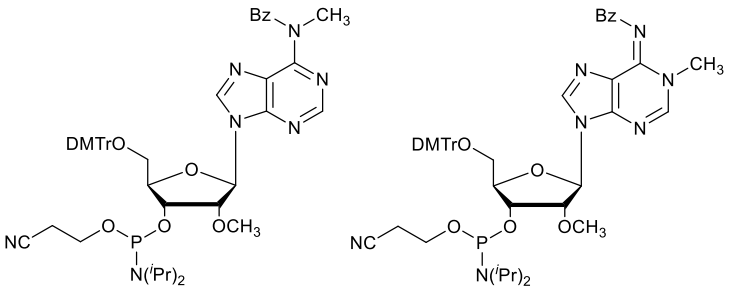
0.02 (1.5H, s, $Si(CH_3)_2$), 0.00 (1.5H, s, $Si(CH_3)_2$), -0.11 (1.5H, s, $Si(CH_3)_2$), -0.12 (1.5H, s, $Si(CH_3)_2$).

δ_C (101 MHz, $CDCl_3$) 177.05 (CO), 177.00 (CO), 158.52 (4 x ArC), 147.63 (C-6), 147.55 (C-6), 146.68 (C-2), 146.55 (C-2), 145.61 (2 x C-4), 144.69 (ArC), 144.55 (ArC), 138.81 (CH-8), 138.57 (CH-8), 135.88 (ArC), 135.86 (ArC), 135.81 (ArC), 135.71 (ArC), 135.55 (ArC), 135.54 (ArC), 131.80 (2 x ArCH), 130.13 (ArCH), 130.11 (ArCH), 130.07 (2 x ArCH), 129.74 (4 x ArCH), 128.18 (ArCH), 128.04 (ArCH), 128.00 (ArCH), 127.89 (ArCH), 127.85 (ArCH), 126.87 (ArCH), 122.72 (C-5), 122.57 (C-5), 117.53 (CN), 117.22 (CN), 113.21 (ArCH), 113.15 (ArCH), 88.35 (CH-1′), 87.94 (CH-1′), 86.59 (CAr_3), 86.42 (CAr_3), 84.00 (CH-4′), 83.53 (CH-4’), 83.51 (d, *J* 4.1 Hz, CH-4’), 75.44 (d, *J* 1.9 Hz, CH-2’), 74.62 (d, *J* 5.8 Hz, CH-2’), 73.14 (d, *J* 9.6 Hz, CH-3’), 72.62 (d, *J* 15.2 Hz, CH-3’), 63.41 (CH2-5′), 63.31 (CH2-5′), 58.78 (d, *J* 16.8 Hz, CH_2CH_2CN), 57.63 (d, *J* 20.7 Hz, CH_2CH_2CN), 55.27 ($ArOCH_3$), 55.24 ($ArOCH_3$), 43.40 (d, *J* 12.8 Hz, $CH(CH_3)_2$), 42.94 (d, *J* 12.7 Hz, $CH(CH_3)_2$), 36.88 (2 x NCH_3), 25.70 (3 x $SiC(CH_3)_3$), 25.66 (3 x $SiC(CH_3)_3$), 24.78 ($CH(CH_3)_2$), 24.68 ($CH(CH_3)_2$), 24.62 ($CH(CH_3)_2$), 24.60 ($CH(CH_3)_2$), 24.52 ($CH(CH_3)_2$), 20.37 (d, *J* 6.2 Hz, CH_2CH_2CN), 20.01 (d, *J* 6.9 Hz, CH_2CH_2CN), 17.97 ($Si(CH_3)_2$), 17.89 ($Si(CH_3)_2$), -4.62 ($Si(CH_3)_2$), -4.66 ($Si(CH_3)_2$), -4.69 ($Si(CH_3)_2$), -4.93 ($Si(CH_3)_2$), -5.00 ($Si(CH_3)_2$).

δ_P (162 MHz, $CDCl_3$) 150.90, 149.25.

HRMS (ESI+) calculated for $C_{54}H_{70}N_7O_8P_1Si$ $(M+H)^+$ 1002.4709, found 1002.4700.

7.2.2.2 5′-(4,4′-Dimethoxytrityl)-6-*N*-benzoyl-6-*N*-methyl-adenosine,2′-O-methyl-3′- [(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite and 5′-(4,4′-Dimethoxytrityl)-6-*N*-benzoyl-1-methyl-adenosine,2′-O-methyl-3′- [(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite[197]



5'-(4,4'-Dimethoxytrityl)-*N*-benzoyl-adenosine, 2'-O-methyl-3'- [(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite (133 mg, 150 µmol) was dissolved in anhydrous dichloromethane (1.3 mL). Iodomethane (38 µL, 606 µmol) was added under an argon atmosphere. The solution was stirred for 5 mins at room temperature. Tetrabutylammonium bromide (50 mg, 156 µmol) and aq. NaOH (1 M, 3.76 mL) were added. The solution was stirred vigorously for 30 mins at room temperature. Water (50 mL) and diethyl ether (50 mL) were added and the layers were separated. The aqueous layer was extracted with diethyl ether (3 x 25 mL). The organic phase was dried over anhydrous Na_2SO_4 and the solvent was evaporated *in vacuo*. The crude

material was purified by silica gel column chromatography (3:2 ethyl acetate/ petrol to 9:1 ethyl acetate/ methanol) to yield product A (73 mg, 54 %) and product B (30 mg, 22 %) as white powders.

Product A

δ_H (400 MHz, $CDCl_3$) 8.49 (1H, s, H-2), 8.15 (0.5H, s, H-8), 8.09 (0.5H, s, H-8), 7.50 – 7.38 (4H, m, ArH), 7.37 – 7.21 (8H, m, ArH), 7.16 (2H, m, ArH), 6.87-6.77 (4H, m, ArH), 6.14 (0.5H, d, *J* 5.2 Hz, H-1’), 6.12 (0.5H, d, *J* 5.3 Hz, H-1’), 4.62 (2H, m, H-2′, H-3′), 4.41 (0.5H, app q, *J* 3.9 Hz, H-4′), 4.35 (0.5H, app q, *J* 3.9 Hz, H-4′), 4.01 – 3.85 (1H, m, CH_2CH_2CN), 3.82 (6H, s, $ArOCH_3$), 3.81 (3H, s, NCH_3), 3.75 – 3.48 (4H, m, CH_2CH_2CN, HA-5′, 2 x $CH(CH_3)_2$), 3.46 (3H, s, 2′-OCH_3), 3.40 – 3.31 (1H, m, H_B-5'), 2.66 (1H, t, *J* 6.4 Hz, OCH_2CH_2CN), 2.40 (1H, t, *J* 6.4 Hz, OCH_2CH_2CN), 1.24 – 1.18 (9H, m, $CH(CH_3)_2$), 1.09 (3H, d, *J* 6.8 Hz, $CH(CH_3)_2$).

δ_C (101 MHz, $CDCl_3$) 172.21 (2 x CO), 158.60 (4 x ArC), 154.87 (C-6), 154.82 (C-6), 152.59 (C-4), 152.54 (C-4), 151.86 (C-2), 151.83 (C-2), 144.45 (ArC), 144.38 (ArC), 142.22 (2 x CH-8), 136.08 (ArC), 136.06 (ArC), 135.59 (2 x ArC), 135.52 (ArC), 135.50 (ArC), 130.59 (ArCH), 130.16 (ArCH), 130.14 (ArCH), 130.09 (ArCH), 130.07 (ArCH), 128.64 (4 x ArCH), 128.22 (2 x ArCH), 128.14 (2 x ArCH), 127.87 (8 x ArCH), 126.99 (ArCH), 126.97 (ArCH), 126.61 (2 x C-5), 117.65 (CN), 117.37 (CN), 113.17 (8 x ArCH), 86.68 (CAr_3), 86.64 (CH-1′), 86.60 (CAr_3), 86.55 (CH-1′), 83.87 (CH-4’), 83.84 (CH-4’), 82.28 (d, *J* 4.0 Hz, CH-2’), 81.86 (d, *J* 5.0 Hz, CH-2’), 71.19 (d, *J* 15.5 Hz, CH-3’), 70.63 (d, *J* 17.6 Hz, CH-3’), 63.02 (CH2-5'), 62.59 (CH2-5'), 58.95 (d, *J* 22.6 Hz, CH_2CH_2CN), 58.77 (d, *J* 1.5 Hz, 2’-OCH_3),

58.34 (d, *J* 2.9 Hz, 2'-OCH_3), 57.94 (d, *J* 19.8 Hz, CH_2CH_2CN), 55.26 (2 x $ArOCH_3$), 55.24 (2 x $ArOCH_3$), 43.45 (d, *J* 12.2 Hz, $CH(CH_3)_2$), 43.15 (d, *J* 12.0 Hz, $CH(CH_3)_2$), 35.87 (2 x NCH_3), 24.66 (2 x $CH(CH_3)_2$), 24.63 (2 x $CH(CH_3)_2$), 24.59 (2 x $CH(CH_3)_2$), 24.56 (2 x $CH(CH_3)_2$), 20.37 (d, *J* 6.2 Hz, CH_2CH_2CN), 20.18 (d, *J* 7.0 Hz, CH_2CH_2CN).

δ_P (162 MHz, $CDCl_3$) 151.09, 150.32.

HRMS (ESI+) calculated for $C_{49}H_{56}N_7O_8P_1$ $(M+Na)^+$ 924.3820, found 924.3812.

Product B

δ_H (400 MHz, $CDCl_3$) 8.26 – 8.16 (2H, app d, *J* 7.5 Hz, ArH), 7.90 – 7.85 (1.5H, m, H-2, H-8), 7.84 (0.5H, s, H-8) 7.55 – 7.47 (1H, m, AH), 7.47 – 7.38 (4H, m, ArH), 7.36 – 7.13 (7H, m, ArH), 6.87 – 6.73 (4H, m, ArH), 6.01 (0.5H, d, *J* 6.2 Hz, H-1'), 5.98 (0.5H, d, *J* 6.0 Hz, H-1'), 4.51 (1H, m, H-3'), 4.39 (1.5H, m, H-2', H-4'), 4.34 – 4.29 (0.5H, m, H-4'), 3.99 – 3.83 (1H, m, OCH_2CH_2CN), 3.80 (3H, s, $ArOCH_3$), 3.79 – 3.78 (3H, s, $ArOCH_3$), 3.74 (3H, s, NCH_3), 3.70 – 3.46 (4H, m, CH_2CH_2CN, HA-5′, 2 x $CH(CH_3)_2$)), 3.44 (1.5H, s, 2′-OCH_3), 3.41 (1.5H, s, 2′-OCH_3), 3.33-3.26 (1H, m, 2.67 (1H, t, *J* 6.4 Hz, CH_2CH_2CN), 2.37 (1H, t, *J* 6.5 Hz, CH_2CH_2CN), 1.24 – 1.15 (9H, m, $CH(CH_3)_2$), 1.06 (3H, d, *J* 6.8 Hz, $CH(CH_3)_2$).

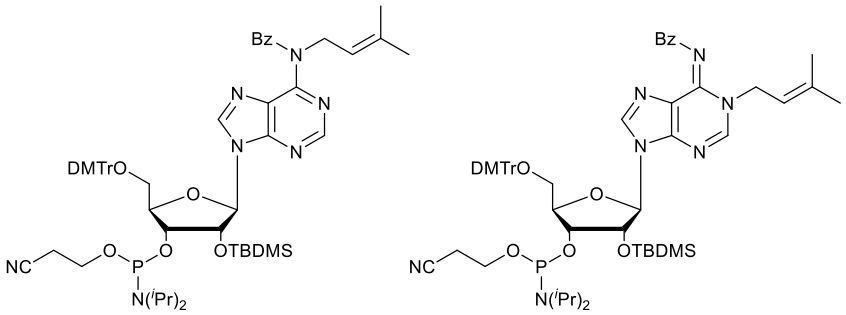
δ_C (101 MHz, $CDCl_3$) 176.97 (CO), 158.55 (4 x ArC), 147.84 (C-6), 147.80 (C-6) 146.83 (C-2), 146.77 (CH-2), 145.64 (CH-4), 145.59 (CH-4), 144.54 (ArC), 144.47 (ArC), 138.55 (CH-8), 138.42 (CH-8), 135.87 (ArC), 135.86 (ArC), 135.66 (ArC), 135.62 (ArC), 135.53 (2 x ArC), 131.87 (2 x ArC),

130.13 (4 x ArCH), 130.06 (4 x ArCH), 129.78 (4 x ArCH), 128.20 (2 x ArCH), 128.09 (2 x ArCH), 128.02, (4 x ArCH), 127.86 (4 x ArCH), 126.91 (2 x ArCH), 122.64 (C-5), 122.61 (C-5), 117.68 (CN), 117.33 (CN), 113.18 (4 x ArCH), 113.16 (8 x ArCH), 86.59 (CAr_3), 86.47 (CAr_3), 86.27 (CH-1'), 86.09 (CH-1'), 84.09 (d, *J* 2.1 Hz, CH-4'), 83.95 (d, *J* 3.5, CH-4') 82.83 (d, *J* 3.0 Hz, CH-2'), 82.35 (d, *J* 5.3 Hz, CH-2'), 71.25 (d, *J* 15.8 Hz, CH-3'), 70.45 (d, *J* 19.3 Hz, CH-3'), 63.28 (CH2-5'), 62.92 (CH2-5'), 58.94 (d, *J* 16.7 Hz, CH2CH2CN), 58.79(d, *J* 4.5, 2′-OCH_3) 58.32 (d, *J* 2.8 Hz, 2′-OCH3), 57.93 (d, *J* 19.2 Hz, CH2CH2CN), 55.27 (2 x $ArOCH_3$), 55.25 (2 x $ArOCH_3$), 43.39 (d, *J* 12.1 Hz, 2 x $CH(CH_3)_2$), 43.18 (d, *J* 12.5 Hz, 2 x $CH(CH_3)_2$), 36.88 (2 x NCH_3), 24.66 (2 x $CH(CH_3)_2$), 24.63 (2 x $CH(CH_3)_2$), 24.57 (2 x $CH(CH_3)_2$), 24.54 (2 x $CH(CH_3)_2$), 20.37 (d, *J* 5.9 Hz, CH_2CH_2CN), 20.10 (d, *J* 7.2 Hz, $CH2CH_2CN$).

δ_P (162 MHz, $CDCl_3$) 151.13, 150.48.

HRMS (ESI+) calculated for $C_{49}H_{56}N_7O_8P_1$ $(M+Na)^+$ 924.3820, found 924.3788.

7.2.2.3 5′-(4,4′-Dimethoxytrityl)-6-*N*-benzoyl-6-*N*-prenyl-adenosine,2′-O-TBDMS-3′- [(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite and 5′-(4,4′-Dimethoxytrityl)-6-*N*-benzoyl-1-prenyl-adenosine,2′-O-TBDMS-3′- [(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite



5'-(4,4'-Dimethoxytrityl)-*N*-benzoyl-adenosine, 2'-O-TBDMS-3'- [(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite (148 mg, 150 µmol) was dissolved in anhydrous dichloromethane (1.3 mL). Prenyl bromide (70 µL, 606 µmol) was added under an argon atmosphere. The solution was stirred for 5 mins at room temperature. Tetrabutylammonium bromide (50 mg, 156 µmol) and aq. NaOH (1 M, 3.76 mL) were added. The solution was stirred vigorously for 30 mins at room temperature. Water (50 mL) and diethyl ether (50 mL) were added and the layers were separated. The aqueous layer was extracted with diethyl ether (3 x 25 mL). The organic phase was dried on anhydrous Na_2SO_4 and the solvent was evaporated *in vacuo*. The crude

material was purified by silica gel column chromatography (1:3 ethyl acetate/ petrol to 9:1 ethyl acetate/methanol) to yield product A (79 mg, 50 %) and product B (33 mg, 21 %) as white powders.

Product A

δ_H (400 MHz, $CDCl_3$) 8.49 (0.5H, s, H-2), 8.47 (0.5H, s, H-2), 8.12 (0.5H, s, H-8), 8.10 (0.5H, s, H-8), 7.50 – 7.42 (4H, m, ArH), 7.40 – 7.31 (4H, m, ArH), 7.31 – 7.22 (4H, m, ArH), 7.20 – 7.11 (2H, m, ArH), 6.87 – 6.78 (4H, m, ArH), 6.05 (0.5H, d, J = 7.0 Hz, CH-1'), 5.98 (0.5H, d, J = 6.9 Hz, CH-1'), 5.46 – 5.36 (1H, m, $NCH_2CH=C(CH_3)_2$), 5.04 – 4.93 (3H, m, $NCH_2CH=C(CH_3)_2$, CH-2'), 4.47 – 4.30 (2H, m, H-3', H-4'), 4.06 – 3.85 (1H, m, OCH_2CH_2CN), 3.82 – 3.80 (6H, m, $ArOCH_3$), 3.75 – 3.48 (4H, m, OCH_2CH_2CN, 2x $CH(CH_3)2$), H_A-5', 3.32 (1H, m, H_B-5'), 2.77 – 2.60 (1H, m, OCH_2CH_2CN), 2.31 (1H, td, *J* 3.8, 6.5 Hz, OCH_2CH_2CN), 1.65 (3H, s, $NCH_2CH=C(CH_3)_2$), 1.63 (3H, s, $NCH_2CH=C(CH_3)_2$), 1.25 – 1.15 (9H, m, $CH(CH_3)_2$), 1.07 (3H, d, *J* 6.7 Hz, $CH(CH_3)_2$), 0.73 (9H, s, $Si(CH_3)_3$), -0.07 (1.5H, s, $Si(CH_3)_2$), -0.10 (1.5H, s, $Si(CH_3)_2$), -0.35 (1.5H, s, $Si(CH_3)_2$), -0.37 (1.5H, s, $Si(CH_3)_2$).

δ_C (101 MHz, $CDCl_3$) 171.87 (CO), 171.86 (CO), 158.59 7 (4 x ArC), 154.47 (C-6), 154.45 (C-6), 152.64 (C-4), 151.94 (CH-2), 144.54 (ArC), 144.41 (ArC), 142.46 (CH-8), 142.38 (CH-8), 136.54 (4 x ArC), 136.14 (2 x ArC), 135.73 ($NCH_2CH=C(CH_3)_2$), 135.65 ($NCH_2CH=C(CH_3)_2$), 135.52 ($NCH_2CH=C(CH_3)_2$), 135.48 ($NCH_2CH=C(CH_3)_2$), 130.51 (2 x ArCH), 130.15 (2 x ArCH), 130.10 (2 x ArCH), 130.07 (2 x ArCH), 130.03 (2 x ArCH), 128.75 (4 x ArCH), 128.19 (2 x ArCH), 128.07 (2 x ArCH), 127.93 (2

x ArCH), 127.90 (2 x ArCH), 127.81(4 x ArCH), 127.28 (2 x C-5), 126.95 (2 x ArCH), 119.93 (2 x $NCH_2CH=C(CH_3)_2$), 117.56 (CN), 117.24 (CN), 113.24 (4 x ArCH), 113.20 (4 x ArCH), 87.90 (CH-1'), 87.67 (CH-1'), 86.77 (CAr_3), 86.62 (CAr_3), 84.57 (CH-4'), 84.18 (d, *J* 3.7 Hz, CH-4'), 75.03 (d, *J* 3.1 Hz, CH-3'), 74.42 (d, *J* 5.9 Hz, CH-3'), 73.61 (d, *J* 10.3 Hz, CH-2'), 72.66 (d, *J* 14.4 Hz, CH-2'), 63.35 (CH_2-5'), 63.19 (CH_2-5'), 58.91 (d, *J* 16.7 Hz, OCH_2CH_2CN), 57.55 (d, *J* 21.0 Hz, OCH_2CH_2CN), 55.26 (2 x $ArOCH_3$), 55.23 (2 x $ArOCH_3$), 46.69 (2 x $NCH_2CH=C(CH_3)_2$), 43.44 (d, *J* 12.8 Hz, $CH(CH_3)_2$), 42.94 (d, *J* 12.5 Hz, $CH(CH_3)_2$), 25.70 (2 x $NCH_2CH=C(CH_3)_2$), 25.59 ($SiC(CH_3)$, 25.55 ($SiC(CH_3)$, 24.80 (2 x $CH(CH_3)_2$), 24.72 (2 x $CH(CH_3)_2$), 24.63 (2 x $CH(CH_3)_2$), 24.55 (2 x $CH(CH_3)_2$), 20.46 (d, *J* 6.4 Hz, OCH_2CH_2CN), 20.08 (d, *J* 6.7 Hz, OCH_2CH_2CN), 17.95 ($NCH_2CH=C(CH_3)_2$), 17.86 ($NCH_2CH=C(CH_3)_2$), 17.79 (2 x $SiC(CH_3)_3$), -4.65 ($Si(CH_3)_2$), -4.68 ($Si(CH_3)_2$), -5.26 ($Si(CH_3)_2$), -5.29 ($Si(CH_3)_2$).

δ_P (162 MHz, $CDCl_3$) 151.38, 148.93

HRMS (ESI+) calculated for $C_{58}H_{74}N_7O_8P_1Si_1$ $(M+Na)^+$ 1078.4998, found 1078.4968.

Product B

δ_H (400 MHz, $CDCl_3$) 8.21 – 8.12 (2H, m, ArH), 7.86 (1H, d, *J* 2.4 Hz, H-8), 7.84 (1H, d, *J* 1.1 Hz, H-2), 7.53 – 7.38 (6H, m, ArH), 7.38 – 7.30 (4H, m, ArH), 7.30 – 7.16 (4H, m, ArH), 6.81 (4H, dd, *J* 6.4, 8.7 Hz, ArH), 5.93 (0.7H, d, *J* 6.1 Hz, H-1'), 5.87 (0.3H, d, *J* 6.0 Hz, H-1'), 5.47 – 5.38 (1H, m, $NCH_2CH=C(CH_3)_2$), 4.89 – 4.79 (1H, m, H-2'), 4.77 (2H, d, *J* 7.2 Hz, $NCH_2CH=C(CH_3)_2$), 4.39 (0.3H, q, *J* 3.6 Hz, H-4'), 4.29 (1.7H, dt, *J* 3.3, 14.8

Hz, H-3’ and H-4’), 4.00 – 3.83 (1.3H, m, OCH_2CH_2CN), 3.79 (2H, s, $ArOCH_3$), 3.78 (4H, s, $ArOCH_3$), 3.67 – 3.49 (2H, m, $CH(CH_3)_2$ and H_A-5’), 3.45 (0.7H, dd, *J* 3.4, 10.5 Hz, OCH_2CH_2CN), 3.23 (1H, ddd, *J* 4.2, 10.5, 15.7 Hz, H_B-5'), 2.65 (0.7H, t, *J* 6.6 Hz, OCH_2CH_2CN), 2.37 – 2.21 (1.3H, m, OCH_2CH_2CN), 1.86 – 1.77 (6H, m, $NCH_2CH=C(CH_3)_2$), 1.18 (8H, dt, *J* 3.8, 7.2 Hz, m), 1.02 (4H, d, *J* 6.7 Hz, $CH(CH_3)_2$), 0.84 – 0.79 (9H, m, $SiC(CH_3)_3$), 0.02 (1H, s, $Si(CH_3)_2$), 0.01 (2H, s, $Si(CH_3)_2$), -0.11 (3H, s, $Si(CH_3)_2$).

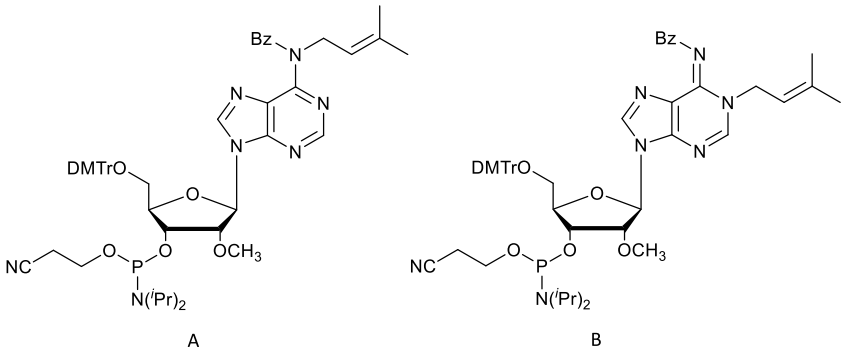
δ_C (101 MHz, $CDCl_3$) 176.85 (CO), 176.79 (CO), 158.53 (4 x ArC), 147.02 (C-6), 146.92 (C-6), 146.25 (CH-2), 146.15 (CH-2), 145.50 (C-4), 145.47 (C-4), 144.68 (ArC), 144.54 (ArC), 139.35 (4 x ArC), 138.63 (CH-8), 138.42 (CH-8), 136.04 (ArC), 135.97 (ArC), 135.79 (ArC), 135.70 (ArC), 135.56 ($NCH_2CH=C(CH_3)_2$), 135.53 ($NCH_2CH=C(CH_3)_2$), 131.68 (2 x ArCH), 130.14 (2 x ArCH), 130.10 (2 x ArCH), 130.06 (4 x ArCH), 129.72 (4 x ArCH), 128.19 (2 x ArCH), 128.06 (2 x ArCH), 127.96 (4 x ArCH), 127.88 (4 x ArCH), 127.84 (2 x ArCH), 126.85 (2 x ArCH), 122.79 (C-5), 122.66 (C-5), 118.08 ($NCH_2CH=C(CH_3)_2$), 118.01 ($NCH_2CH=C(CH_3)_2$), 117.53 (CN), 117.20 (CN), 113.20 (4 x ArCH), 113.16 (4 x ArCH), 88.23 (CH-1'), 87.87 (CH-1’), 86.60 (CAr_3), 86.44 (CAr_3), 83.95 (CH-4’), 83.47 (d, *J* 4.2 Hz, CH-4’), 75.47 (d, *J* 3.0 Hz, CH-3’), 74.67 (d, *J* 5.4 Hz, CH-3’), 73.13 (d, *J* 10.1 Hz, CH-2’), 72.59 (d, *J* 15.2 Hz, CH-2’), 63.42 (CH2-5’), 63.30 (CH2-5’), 58.78 (d, *J* 16.7 Hz, OCH_2CH_2CN), 57.63 (d, *J* 21.0 Hz, OCH_2CH_2CN), 55.25 (2 x $ArOCH_3$), 55.23 (2 x $ArOCH_3$), 46.63 (2 x $NCH_2CH=C(CH_3)_2$), 46.60 (2 x $NCH_2CH=C(CH_3)_2$), 43.40 (d, *J* 13.0 Hz, $CH(CH_3)_2$), 42.94 (d, *J* 12.6 Hz, $CH(CH_3)_2$), 25.82 (2 x $NCH_2CH=C(CH_3)_2$), 25.70 (2 x $SiC(CH_3)_3$), 25.66 (2 x $SiC(CH_3)_3$), 24.78 (2 x $CH(CH_3)_2$), 24.68 (2 x $CH(CH_3)_2$), 24.59 (2 x

$CH(CH_3)_2$), 24.52 (2 x $CH(CH_3)_2$), 20.38 (d, *J* 6.1 Hz, OCH_2CH_2CN), 19.99 (d, *J* 7.1 Hz, OCH_2CH_2CN), 18.29 ($NCH_2CH=C(CH_3)_2$), 17.96 ($SiC(CH_3)_3$), 17.89 ($SiC(CH_3)_3$), 1.03, -4.62 ($Si(CH_3)_2$), -4.65 ($Si(CH_3)_2$), -4.96 ($Si(CH_3)_2$), -5.01 ($Si(CH_3)_2$).

δ_P (162 MHz, $CDCl_3$) 150.89, 149.21.

HRMS (ESI+) calculated for $C_{58}H_{74}N_7O_8P_1Si_1$ $(M+Na)^+$ 1078.4998, found 1078.5060.

7.2.2.4 5′-(4,4′-Dimethoxytrityl)-6-*N*-benzoyl-6-*N*-prenyl-adenosine,2′-O-methyl-3′- [(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite and 5′-(4,4′-Dimethoxytrityl)-6-*N*-benzoyl-1-prenyl-adenosine,2′-O-methyl-3′- [(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite



5'-(4,4'-Dimethoxytrityl)-*N*-benzoyl-adenosine, 2'-O-methyl-3'- [(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite (Link Technologies, 133 mg, 150 μmol) was dissolved in anhydrous dichloromethane (1.3 mL). Prenyl bromide (70 μL, 606 μmol) was added under an argon atmosphere. The solution was stirred for 5 mins at room temperature. Tetrabutylammonium bromide (50 mg, 156 μmol) and aq. NaOH (1 M, 3.76 mL) were added. The solution was stirred vigorously for 30 mins at room temperature. Water (50 mL) and diethyl ether (50 mL) were added and the layers were separated. The aqueous layer was extracted with diethyl ether (3 x 25 mL). The organic phase was dried on anhydrous Na_2SO_4 and the solvent was evaporated *in vacuo*. The crude material was purified by silica gel column chromatography (3:2 ethyl

acetate/ petrol to 9:1 ethyl acetate/ methanol) to yield product A (34 mg, 24 %) and product B (70 mg, 49 %) as white powders.

Product A

δ_H (400 MHz, $CDCl_3$) 8.52 (0.6H, s, H-2), 8.51 (0.4H, s, H-2), 8.12 (0.4H, s, H-8), 8.07 (0.6H, s, H-8), 7.50 – 7.37 (4H, m, ArH), 7.36 – 7.20 (8H, m, ArH), 7.14 (2H, m, ArH), 6.88 – 6.76 (4H, m, ArH), 6.13 (0.6H, d, *J* 5.4 Hz, H-1'), 6.11 (0.4H, *J* 5.4 Hz, H-1'), 5.43 (1H, m, $NCH_2CH=C(CH_3)_2$), 5.01 – 4.91 (2H, m, $NCH_2CH=C(CH_3)_2$), 4.70 – 4.50 (2H, m, H-2', H-3'), 4.40 (0.4H, app q, *J* 3.8 Hz, H-4'), 4.34 (0.6H, app q, *J* 3.7 Hz, H-4'), 4.0-3.84 (1H, m, CH_2CH_2CN), 3.82-3.80 (6H, s, $ArOCH_3$), 3.76 – 3.46 (4H, m, OCH_2CH_2CN, 2 x CH(CH3)2, H_A-5'), 3.45 (3H, s, 2'-OCH_3), 3.37-3.29 (1H, m, H_B-5'), 2.66 (1H, t, *J* 6.4 Hz, CH_2CH_2CN), 2.40 (1H, t, *J* 6.4 Hz, CH_2CH_2CN), 1.64 (6H, m, $NCH_2CH=C(CH_3)_2$), 1.23-1.17 (9H, m, $CH(CH_3)_2$), 1.08 (3H, d, *J* 6.8 Hz, $CH(CH_3)_2$).

δ_C (101 MHz, $CDCl_3$) 171.94 (CO), 171.15 (CO), 158.60 (ArC), 154.40 (C-6), 154.36 (C-6), 152.46 (C-4), 152.41 (C-4), 151.97 (CH-2'), 151.95 (CH-2'), 144.45 (ArC), 144.38 (ArC), 142.07 (CH-8), 136.39 (ArC), 136.37 (ArC), 136.14 ($NCH_2CH=C(CH_3)_2$), 136.12 (NCH2CH=$C(CH_3)_2$), 135.62 (ArC), 135.58 (ArC), 135.52 (ArC), 135.32 (ArC), 130.55 (ArCH), 130.16 (ArCH), 130.14 (ArCH), 130.08 (ArCH), 130.07 (ArCH), 129.66 (ArCH), 128.70 (ArCH), 128.22 (ArCH), 128.13 (ArCH), 127.88 (ArCH), 127.81 (ArCH), 127.02 (C-5), 126.99 (ArCH), 126.96 (ArCH), 120.03 (NCH2CH=C(CH3)2), 120.02 ($NCH_2CH=C(CH_3)_2$), 117.67 (CN), 117.38 (CN), 113.17 (ArC), 66.67 (CH-1'), 86.59 (CH-1'), 86.51 (CAr_3), 86.42 (CAr_3), 83.86 (CH-4'), 83.82

(CH-4’), 82.27 (d, *J* 3.4 Hz, CH-3’), 81.84 (d, *J* 4.7 Hz, CH-2’), 71.20 (d, *J* 15.6 Hz, CH-3’), 70.62 (d, *J* 18.0 Hz, CH-2’), 63.04 (CH2-5’), 62.60 (CH2-5’), 58.90 (d, *J* 17.1 Hz, OCH_2CH_2CN), 58.73 (d, *J* 3.6 Hz, 2’-OCH_3), 58.34 (d, *J* 2.7 Hz, 2’-OCH_3), 57.96 (d, *J* 19.8 Hz OCH_2CH_2CN), 55.26 ($ArOCH_3$), 55.23 ($ArOCH_3$), 46.78 ($NCH_2CH=C(CH_3)_2$), 43.38 (d, *J* 12.2 Hz, $CH(CH_3)_2$), 43.21 (d, *J* 12.6 Hz, $CH(CH_3)_2$), 25.71 (NCH2CH=$C(CH_3)_2$), 24.66 ($CH(CH_3)_2$), 24.63 ($CH(CH_3)_2$), 24.60 ($CH(CH_3)_2$), 24.57 ($CH(CH_3)_2$), 20.37 (d, *J* 6.2 Hz OCH_2CH_2CN), 20.18 (d, *J* 6.8 Hz, OCH_2CH_2CN), 17.94 ($NCH_2CH=C(CH_3)_2$).

δ_P (162 MHz, $CDCl_3$) 151.12, 150.30

HRMS (ESI+) calculated for $C_{53}H_{62}N_7O_8P_1$ $(M+Na)^+$ 978.4290, found 978.4278.

Product B

δ_H (400 MHz, $CDCl_3$) 8.23 – 8.20 (1H, m, H-2), 8.20 – 8.17 (1H, m, H-8), 7.92 – 7.81 (2H, m, ArH), 7.55 – 7.45 (1H, m, ArH), 7.48 – 7.36 (4H, m, ArH), 7.37 – 7.10 (7H, m, ArH), 6.87 – 6.76 (4H, m, ArH), 6.01 (0.4H, d, *J* 6.2 Hz, H-1’), 5.98 (0.6H, d, *J* 5.9 Hz, H-1’), 5.44 (1H, m, $NCH_2CH=C(CH_3)_2$), 4.78 (2H, d, *J* 7.2 Hz, $NCH_2CH=C(CH_3)_2$), 4.54 (0.6H, ddd, *J* 3.6, 5.1, 10.0 Hz, H-3’), 4.48 (0.4H, ddd, *J* 3.2, 4.9, 11.1 Hz, H-3’),4.43 – 4.34 (1.6H, m, H-2’ and H-4’), 4.34 – 4.29 (0.4H, m, H-4’), 3.98 – 3.81 (1H, m, CH_2CH_2CN), 3.79 (6H, d, m, $ArOCH_3$), 3.72 – 3.38 (7H, m, OCH_2CH_2CN, 2 x $CH(CH_3)_2$, H_A-5’, 2’-OCH_3), 3.28 (1H, m, H_B-5’), 2.66 (1H, t, *J* 6.4 Hz, CH_2CH_2CN), 2.36 (1H, t, *J* 6.5 Hz, CH_2CH_2CN), 1.83 (3H, s,

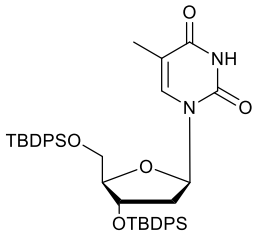
$NCH_2CH=C(CH_3)_2$), 1.82 (3H, s, $NCH_2CH=C(CH_3)_2$), 1.22 – 1.16 (9H, m, $CH(CH_3)_2$), 1.05 (3H, d, J 6.8 Hz, $CH(CH_3)_2$)

δ_C (101 MHz, $CDCl_3$) 176.76 (CO), 176.74 (CO), 158.65 (ArC), 158.55 (ArC), 147.29 (C-6), 147.24 (C-6), 146.43 (C-2), 146.38 (C-2), 145.52 (C-5), 145.48 (C-5), 144.53 (ArC), 144.46 (ArC), 139.43 (ArC), 138.38 (CH-8), 138.26 (CH-8), 136.04 (ArC), 136.02 (ArC), 135.65 (ArC), 135.61, 135.52 ($NCH_2CH=C(CH_3)_2$), 131.76 (ArCH), 130.12 (ArCH), 130.06 (ArCH), 129.77 (ArCH), 128.21 (ArCH), 128.11 (ArCH), 127.98 (ArCH), 127.86 (ArCH), 126.91 (ArCH), 122.69 (C-5), 122.68 (C-5), 118.02 ($NCH_2CH=C(CH_3)_2$), 117.70 (CN), 117.34 (CN), 113.23 (ArCH), 113.18 (ArCH), 113.16 (ArCH), 86.60 (CH-1'), 86.48 (CH-1'), 86.16 (CAr3), 86.00 (CAr3), 84.07 (d, *J* 2.7 Hz, CH-4'), 83.91 (d, *J* 4.1 Hz, CH-4'), 82.87 (d, *J* 3.4 Hz, CH-2'), 82.38 (d, *J* 5.4 Hz, CH-2'), 71.24 (d, *J* 16.3 Hz, H-3'), 70.52 (d, *J* 18.6 Hz, H-3'), 63.29 (CH2- 5'), 62.90 (CH2-5), 58.94 (d, *J* 16.4 Hz, OCH_2CH_2CN), 58.80 (d, *J* 3.8 Hz, 2'-OCH_3), 58.30 (d, *J* 3.2 Hz, 2'-OCH_3), 57.92 (d, *J* 19.2 Hz, OCH_2CH_2CN), 55.26 ($ArOCH_3$), 55.24 ($ArOCH_3$), 46.65 ($NCH_2CH=C(CH_3)_2$), 46.63 ($NCH_2CH=C(CH_3)_2$), 43.39 (d, *J* 12.1 Hz, $CH(CH_3)_2$), 43.18 (d, *J* 12.1 Hz, $CH(CH_3)_2$), 25.84 ($NCH_2CH=C(CH_3)_2$), 24.66 ($CH(CH_3)_2$), 24.63 ($CH(CH_3)_2$), 24.60 ($CH(CH_3)_2$), 24.56 ($CH(CH_3)_2$), 24.54 ($CH(CH_3)_2$), 20.34 (d, *J* 6.1 Hz, OCH_2CH_2CN), 20.09 (d, *J* 7.3 Hz, OCH_2CH_2CN), 18.28($NCH_2CH=C(CH_3)_2$).

δ_P (162 MHz, $CDCl_3$) 151.15, 150.44.

HRMS (ESI+) calculated for $C_{53}H_{62}N_7O_8P_1$ $(M+Na)^+$ 978.4290, found 978.4290.

7.2.2.5 3',5'-bis-O-[(1,1-Dimethylethyl)diphenylsilyl]-thymidine **10**[17]



Thymidine (5.0 g, 20 mmol) was dissolved in anhydrous dimethylformamide (41 mL) under an argon atmosphere and stirred for 30 minutes. *tert*-Butyl(chloro)diphenylsilane (11.9 mL, 82.6 mmol) was then added to the solution followed by 1*H*-imidazole (5.61 g, 82.64 mmol). The reaction was stirred at room temperature for 24 h. The mixture was diluted with ethyl acetate (200 mL) and extracted with water (3 x 100 mL), dried over anhydrous $MgSO_4$, filtered and concentrated *in vacuo* to yield a crude white foam. The crude material was purified by silica gel column chromatography (2:1 petrol/ diethyl ether) to yield the title compound as a white foam (14.05 g, 94%).

δ_H (400 MHz, $CDCl_3$) 8.59 (1H, br. s., NH), 7.66 (2H, d, *J* 7.3 Hz, 2H, Ar-H), 7.61 (2H, d, *J* 7.4 Hz, Ar-H), 7.55 (2H, d, *J* 7.3 Hz, Ar-H), 7.51 – 7.38 (9H, m, H-6 and Ar-H), 7.38 – 7.27 (6H, m, Ar-H), 6.54 (1H, dd, *J* 9.1, 5.2 Hz, H-1'), 4.58 (1H, d, J 5.5 Hz, H-3'), 4.02 (1H, d, J 2.7 Hz, H-4'), 3.77 (1H, dd, J 2.3, 11.6 Hz, HA-5'), 3.33 (1H, dd, *J* 1.8, 11.7 Hz, H**B**-5'), 2.36 (1H, ddd, *J* 13.1, 5.3, 1.5 Hz, H_A-2'), 1.99 (1H, ddd, *J* 13.9, 9.0, 5.7 Hz, H_B-2'), 1.52 (3H, d, *J* 1.2 Hz, H-7), 1.11 (9H, s, $SiC(CH_3)_3$), 0.96 (9H, s, $SiC(CH_3)_3$).

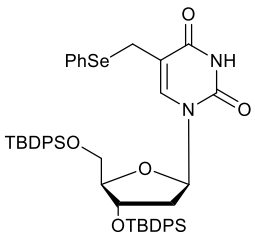
δ_C (101 MHz, $CDCl_3$) 163.49 (CO-4), 150.18 (CO-2), 135.72, 135.67, 135.44, 135.40, 135.15, 133.21, 133.12, 132.99, 132.15, 130.04, 130.00, 129.97, 127.92, 127.89 (ArCH, CH-6), 111.07 (C-5), 87.78 (CH-4'), 84.78 (CH-1'),

74.01 (CH-3’), 64.02 (CH_2-5’), 41.30 (CH_2-2’), 26.90 ($SiC(CH_3)_3$), 19.31 ($SiC(CH_3)_3$)), 19.02 ($SiC(CH_3)_3$)), 11.88 (CH_3, C-7).

HRMS (ESI+) calculated for $C_{42}H_{50}N_2O_5Si_2Na$ $(M+Na)^+$ 741.3150, found 741.3173.

Data in accordance with literature.

7.2.2.6 3', 5'-Bis-O-(*tert*-butyldiphenylsilyl)-5-phenylselenylmethyl-2'-deoxyuridine **12**[17]



3', 5’-Bis-O-(*tert*-butyldiphenylsilyl)-thymidine (1.5 g, 2.1 mmol) was flushed 3 times with argon using a Schlenk line. Carbon tetrachloride (29 mL) was added, and the solution was stirred for 30 minutes at reflux. *N*-bromo-succinimide (728 mg, 4.09 mmol), and benzoyl peroxide (54 mg, 0.2 mmol) were then added to the reaction which was stirred at reflux for a further 1h. The solution changed from colourless to orange. An equal volume of toluene was added to the reaction (29 mL) The mixture was then filtered to remove succinimide and concentrated *in vacuo* to yield the crude bromide intermediate as an orange oil.

Diphenyldiselenide (729 mg, 2.34 mmol) was dissolved in anhydrous dimethylformamide (18 mL) under an argon atmosphere. Sodium borohydride (202 mg, 5.34 mmol) was added and the solution stirred for 30 minutes at room temperature. The bromide intermediate was dissolved in anhydrous dimethylformamide (37 mL) and was added dropwise to the reaction. The solution was stirred room temperature overnight. The mixture was diluted with

ethyl acetate (200 mL) and extracted with water (3 x 100 mL). The solution was then washed with saturated $NaHCO_3$ (3 x, 50 mL), followed by brine (2 x 50 mL) and then dried over anhydrous Na_2SO_4. The solution was filtered and concentrated *in vacuo* to yield a yellow foam. The crude material was purified by silica gel column chromatography (2:1 petrol/ diethyl ether) to yield the title compound as a white foam (717 mg, 58%).

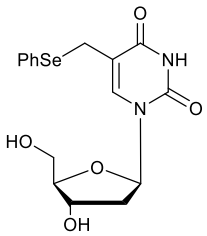
δ_H (400 MHz, $CDCl_3$) 8.31 (1H, br. s., NH), 7.66 (1H, q, *J* 1.6 Hz, Ar-H), 7.64 (1H, d, *J* 1.5 Hz, Ar-H), 7.60 (1H, q, *J* 1.6 Hz, Ar-H), 7.58 (1H, d, *J* 1.5 Hz, Ar-H), 7.53 (1H, q, *J* 1.4 Hz, Ar-H), 7.51 (2H, dt, *J* 2.8, 1.7 Hz, Ar-H), 7.50 – 7.47 (2H, m, Ar-H), 7.46 – 7.39 (6H, m, Ar-H), 7.37 – 7.29 (9H, m, Ar-H), 7.11 – 7.04 (3H, m, Ar-H), 7.03 (1H, s, H-6), 6.38 (1H, dd, *J* 9.0, 5.2 Hz, H-3'), 4.42 (1H, dt, *J* 5.8, 1.7 Hz, H-4'), 4.02 (1H, td, *J* 3.2, 1.8 Hz, H-4'), 3.57 (1H, dd, *J* 11.4, 3.2 Hz, H_A-5'), 3.51 – 3.35 (2H, m, H-7), 3.31 (1H, dd, *J* 11.4, 3.4 Hz, H_B-5'), 2.20 (1H, ddd, *J* 13.2, 5.3, 1.5 Hz, H_A-2'), 1.61-1.54 (1H, m, H_B-2') 1.10 (9H, s, $SiC(CH_3)_3$), 0.94 (9H, s, $SiC(CH_3)_3$).

δ_C (101 MHz, $CDCl_3$) 161.74 (CO-4), 149.75 (CO-2), 135.85, 135.73, 135.67, 135.57, 135.51, 135.29, 134.91, 134.82, 134.61, 133.14, 133.12, 132.93, 132.32, 130.07, 130.02, 130.00, 129.97, 129.65, 129.09, 129.05, 128.96, 127.91, 127.88, 127.58 (ArCH and CH-6), 112.13 (C-5), 87.69 (CH-4'), 84.80 (CH-1'), 73.64 (CH-3'), 63.91 (CH_2-5'), 40.86 (CH_2-2'), 26.90 ($SiC(CH_3)_3$), 26.88 ($SiC(CH_3)_3$), 23.16 (CH_2, C-7), 19.22 ($SiC(CH_3)_3$), 19.02 ($SiC(CH_3)_3$).

HRMS (ESI+) calculated for $C_{48}H_{55}N_2O_5Si_2Se$ $(M+H)^+$ 875.2809, found 875.2816.

Data in accordance with literature.

7.2.2.7 5-phenylselenylmethyl-2'-deoxyuridine **13**[17]



3',5'-Bis-O-(*tert*-butyldiphenylsilyl)-5-phenylselenylmethyl-2'-deoxyuridine (800 mg, 0.92 mmol) was dissolved in tetrahydrofuran (6.5 mL) under an argon atmosphere. Tetrabutylammonium fluoride•$3H_2O$ (1.5 g, 4.6 mmol) was added and the solution was stirred at room temperature for 1.5 h. The solution was concentrated *in vacuo*. The crude material was purified via silica gel column chromatography (100:1 ethyl acetate/methanol) to yield the title compound as a white solid (253 mg, 70 %).

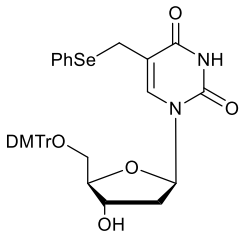
δ_H (400 MHz, CD_3OD) δ 7.57 – 7.46 (2H, m, Ar-H), 7.44 (1H, s, H-6), 7.38 – 7.25 (3H, m, Ar-H), 6.18 (1H, dd, *J* 7.4, 6.1 Hz, H-1'), 4.22 (1H, dt, *J* 6.5, 3.3 Hz, H-3'), 3.84 (1H, q, *J* 3.4 Hz, H-4'), 3.81 – 3.71 (2H, m, H-7), 3.66 (1H, dd, *J* 12.0, 3.3 Hz, H_A-5'), 3.61 (1H, dd, *J* 11.9, 3.7 Hz, H_B-5'), 2.09 (1H, ddd, *J* 13.6, 6.2, 3.3 Hz, H_A-2'), 1.75 (1H, ddd, *J* 13.7, 7.4, 6.3 Hz, H_B-2').

δ_C (101 MHz, CD_3OD) 163.2 (CO-4), 150.5 (CO-2), 136.8, 134.7, 129.0, 127.5 (ArCH and CH-6), 111.7 (C-5), 87.4 (C-4'), 84.6 (C-1'), 70.7 (CH-3'), 61.4 (CH_2-5'), 39.8 (CH_2-2'), 23.4 (CH_2-7).

HRMS (ESI+) calculated for $C_{16}H_{18}N_2O_5Se_1Na$ $(M+Na)^+$ 421.0273, found 421.0279.

Data in accordance with literature.

7.2.2.8 5’-O-Dimethoxytrityl-5-phenylselenylmethyl-2’-deoxyuridine **14**[271]



5-phenylselenylmethyl-2'-deoxyuridine (1.26 g, 3.17 mmol) was dissolved in anhydrous pyridine (25 mL). The solution was stirred at 0^{o}C for 15 minutes under an argon atmosphere. Dimethoxytrityl chloride (1.08 g, 3.17 mmol) was dissolved in anhydrous pyridine (7.5 mL) under an argon atmosphere and was added drop wise to the above solution. The reaction was stirred for 1 hr at 0 °C followed by 12 hrs at room temperature. The solvent was evaporated *in vacuo* to give a crude purple oil. The oil was diluted with ethyl acetate (50 mL) and washed with saturated sodium hydrogen carbonate (3× 15 mL), followed by brine (2 x 15 mL), dried over anhydrous Na_2SO_4 and concentrated *in v*acuo. The crude material was purified via silica gel column chromatography using oven dried silica gel (2:1 ethyl acetate/Hexane with 1 % triethylamine) to yield the title compound as a white foam (1.40 g, 63 %).

δ_H (400 MHz, $CDCl_3$) 8.64 (1H, s, NH), 7.44 – 7.35 (4H, m, Ar-H), 7.35 – 7.29 (7H, m, Ar-H, H-6), 7.28 – 7.25 (1H, m, Ar-H), 7.21 – 7.11 (3H, m, Ar-H), 6.91 – 6.81 (4H, m, Ar-H), 6.26 (1H, t, *J* 6.6 Hz, H-1’), 4.34 (1H, dt, *J* 7.2, 3.8 Hz, H-3’), 3.99 (1H, q, *J* 4.1 Hz, H-4’), 3.80 (6H ,d, *J* 2.5 Hz, OCH_3), 3.51 (1H, dd, *J* 12.1, 0.7 Hz, H_A-7), 3.42 – 3.35 (2H, m, H_A-5’, H_B-7), 3.29 (1H, dd, *J* 10.3, 4.6 Hz, H_B-5’), 2.32 (1H, ddd, *J* 13.7, 6.2, 3.9 Hz, H_A-2’), 1.95 (1H, dt, *J* 13.6, 6.8 Hz, H_B-2’).

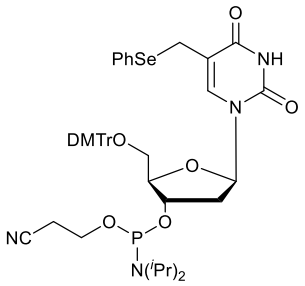
δ_C (101 MHz, CDCl3) 171.19 (2 x BzCO), 161.85 (CO-4), 158.72 (ArC), 149.76 (CO-2), 144.34 (ArC), 136.09, 135.49, 135.35, 135.00, 134.41, 130.08,

130.05, 129.81, 129.14, 129.03, 128.09, 128.05, 127.77, 127.58, 127.17 (ArCH and CH-6), 113.33 (ArC), 112.21 (C-5), 86.89 (CAr3), 85.45 (CH-4'), 84.60 (CH-1'), 72.20 (CH-3'), 63.58 (CH2-5'), 55.27 (OCH3) , 40.46 (CH2-2'), 23.02 (CH2-7).

HRMS (ESI+) calculated for $C_{37}H_{36}N_2O_7Se_1Na$ $(M+Na)^+$ 723.1580, found 723.1608.

Data in accordance with literature.

7.2.2.9 5’-O-Dimethoxytrityl-5-phenylselenylmethyl-2’-deoxyuridine-cyanoethyl phosphoramidite **15**[271]



5’-O-Dimethoxytrityl-5-phenylselenylmethyl-2’-deoxyuridine (700 mg, 1.00 mmol) was dissolved in anhydrous dichloromethane (2.5 mL). Anhydrous diisopropylethylamine (703 µL, 4.05 mmol) was added to the solution which was stirred at 0^{o}C for 30 minutes. 2-cyanoethyl diisopropylchlorophosphoramidite (250 mg, 1.06 mmol) was diluted with anhydrous dichloromethane (2.5 mL) and was added dropwise to the above solution. The reaction was stirred for 2 hrs at room temperature, diluted with ethyl acetate (50 mL), washed with saturated sodium bicarbonate (3 x 15 mL), brine (2 x 15 mL), dried over anhydrous Na_2SO_4 and concentrated *in vacuo*. The crude product was purified via silica gel flash chromatography using oven-dried silica gel (1:1 ethyl acetate/petrol ether) to yield the title compound as a white foam (370 mg, 41 %).

δ_H (400 MHz, $CDCl_3$) 7.42 (2H, ddt, *J* 5.3, 3.7, 2.1 Hz, Ar-H), 7.34 (8H, dddd, *J* 15.6, 7.0, 4.0, 2.2 Hz, Ar-H, H-6), 7.28 – 7.12 (4H, m, Ar-H), 6.85 (4H, ddd, *J* 9.0, 4.3, 2.6 Hz, Ar-H), 6.27 (1H, dt, *J* 7.6, 6.1 Hz, H-1’), 4.59 – 4.51 (1H, m, H-4’), 4.16 (1H, dd, *J* 11.2, 3.5 Hz, H-3’), 3.92 – 3.72 (7H, m, OCH_3, CH_2CH_2CN), 3.73 – 3.54 (2H, m, CH_2CH_2CN), 3.50 (1H, dd, *J* 11.9, 1.7 Hz, H_A-7), 3.41 – 3.23 (3H, m, H_A-5’, H_B-5’ H_B-7), 2.64 (1H, t, *J* 6.2 Hz, CH_2CH_2CN), 2.52 – 2.34 (2H, m, CH_2CH_2CN, H_A-2’), 2.16 – 1.95 (1H, m,

H_B-2’), 1.20 (9H, dd, *J* 6.8, 4.5 Hz, $CH(CH_3)_2$), 1.11 (3H, d, *J* 6.8 Hz, $CH(CH_3)_2$).

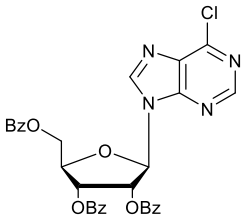
δ_C (101 MHz, $CDCl_3$) 171.15 (2 x CO), 161.81 (CO-4), 158.70 (ArC), 149.74 (CO-2), 149.69 (CO-2), 144.31 (ArC), 136.35, 136.29, 135.50, 135.35, 135.32, 134.17, 130.19, 130.14, 130.10, 129.99, 128.98, 128.24, 128.17, 128.01, 127.50, 127.16, 127.12 (ArCH and CH-6), 117.53 (CN), 117.39 (CN), 113.29 (ArC), 112.23 (C-5), 112.18 (C-5), 86.79 (CAr_3), 85.56 (d, *J* 3.4 Hz, CH-3’), 85.30 (d, *J* 6.0 Hz, CH-3’), 84.87 (CH-1’), 73.81 (d, *J* 17.3 Hz, CH-4’), 73.30 (d, *J* 17.4 Hz, CH-4’), 63.35 (CH_2-5’), 63.14 (CH_2-5’), 58.31 (d, *J* 4.5 Hz, OCH_2CH_2CN), 58.12 (d, *J* 4.6 Hz, OCH_2CH_2CN), 55.28 (OCH_3), 55.25 (OCH_3), 43.34 (d, *J* 7.0 Hz, $CH(CH_3)_2$), 43.22 (d, *J* 6.9 Hz, $CH(CH_3)_2$), 39.77 (CH_2-2’), 39.72 (CH_2-2’), 24.66 ($CH(CH_3)_2$), 24.59 ($CH(CH_3)_2$), 24.56 ($CH(CH_3)_2$), 24.52 ($CH(CH_3)_2$), 22.89 (CH_2, C-7), 21.07, 20.43 (d, *J* 7.3 Hz, OCH_2CH_2CN), 20.21 (d, *J* 7.2 Hz, OCH_2CH_2CN).

δ_P (162 MHz, $CDCl_3$) 149.05, 148.58.

HRMS (ESI+) calculated for $C_{46}H_{54}N_4O_8P_1Se$ $(M+H)^+$ 901.2839, found 901.2863.

Data in accordance with literature.

7.2.2.1 (2R,3R,4R)-2-((benzoyloxy)methyl)-5-(6-chloro-9H-purin-9-yl)tetrahydrofuran-3,4-diyl dibenzoate[198]



1-O-acetyl-2,3,5-tri-O-benzoyl-D-ribofuranose (2.10 g, 4 mmol) and 6-chloropurine (700 mg, 4.5 mmol) were dissolved in acetonitrile (25 mL). Hexamethyldisilazane (816 µL, 4 mmol), trimethylsilyl chloride (795 µL, 6.25 mmol) and trimethylsilyl triflate (1.13 mL, 6.25 mmol) were then added. The solution was heated at reflux for 2.5 hrs, followed by room temperature for 15 mins. Dichloromethane (50 mL) was added, and the solution was extracted three times with saturated sodium bicarbonate solution (50 mL). The organic phase was dried over sodium sulfate and evaporated under vacuum. The crude material was purified by column chromatography (4:1 petrol/ethyl acetate) to yield the title compound as a colourless foam (1.72 g, 71 %).

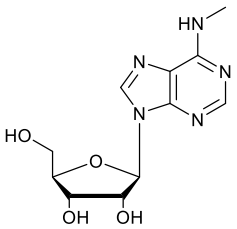
δ_H (400 MHz, $CDCl_3$) 8.63 (1H, s, H-8), 8.30 (1H, s, H-2)), 8.13 – 8.06 (2H, m, ArH), 8.06 – 8.01 (2H, m, ArH), 7.98 – 7.90 (2H, m, ArH), 7.66 – 7.53 (3H, m, ArH), 7.52 – 7.33 (6H, m, ArH), 6.51 – 6.39 (2H, m, H-1', H-2'), 6.27 (1H, dd, *J* 5.0, 5.6 Hz, H-3'), 4.96 (1H, dd, *J* 3.2, 12.3 Hz, HA-5'), 4.88 (1H, ddd, *J* 3.2, 4.2, 5.1 Hz, H-4'), 4.72 (1H, dd, *J* 4.2, 12.2 Hz, HB-5').

δ_C (400 MHz, $CDCl_3$) 166.09 (ArH), 165.31 (ArH), 165.13 (ArH), 152.32 (CH-8), 151.64 (ArH), 151.27 (ArH), 143.95 (CH-2), 133.94 (ArH), 133.84 (ArH), 133.56 (ArH), 132.40 (ArH), 129.84 (ArH), 129.71 (ArH), 129.24 (ArH), 128.62 (ArH), 128.58 (ArH), 128.23 (ArH), 87.53 (CH-1'), 81.05 (CH-4'), 73.86 (H-2'), 71.38 (CH-3'), 63.28 (H2-5').

HRMS (ESI+) calculated for $C_{31}H_{23}Cl_1N_4O_7$ $(M+Na)^+$ 621.1147, found 621.1145

Data in accordance with literature

7.2.2.2 N^6-methyladenosine[198]



(2R,3R,4R)-2-((benzoyloxy)methyl)-5-(6-chloro-9H-purin-9-yl)tetrahydrofuran-3,4-diyl dibenzoate (1.0 g, 1.66 mmol) was dissolved in anhydrous 1,2-dimethoxyethane (40 mL) . Methylamine (443 µL, 10 mmol) was added and the solution was stirred for 16 hrs at room temperature. The mixture was filtered and concentrated in vacuo. The product was purified silica gel column chromatography (100 % ethyl acetate) to yield a colourless foam. This was dissolved in methanolic ammonia (8 mL, 7M) and the solution was stirred at 40 °C for 16 hrs. The solution was again concentrated *in vacuo* and the crude material was purified by silica gel column chromatography (4:1 ethyl acetate/ ethanol) to yield the title compound as a white solid. (185 mg, 39 %).

δ_H (400 MHz, MeOD) 8.25 (1H, s, H-8), 8.25 (1H, s, H-2) 5.97 (1H, d, *J* 6.4 Hz, H-1'), 4.76 (1H, dd, *J* 5.1, 6.5 Hz, H-2'), 4.34 (1H, dd, *J* 2.5, 5.1 Hz, H-3'), 4.19 (1H, q, *J* 2.5 Hz, H-4'), 3.91 (1H, dd, *J* 2.5, 12.5 Hz, HA-5'), 3.76 (1H, dd, *J* 2.7, 12.6 Hz, HB-5'), 3.33 (3H, p, *J* 1.6 Hz, 3 x OH), 3.13 (3H, s, CH_3).

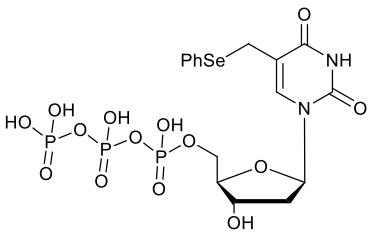
δ_C (400 MHz, MeOD) 155.50 (C-6), 152.07 (C-4), 140.06 (CH-8), 89.89 (CH-1'), 86.82 (CH-4'), 74.04 (CH-2'), 71.31 (CH-3'), 62.13 (CH_2-5').

HRMS (ESI+) calculated for $C_{11}H_{15}N_5O_4$ $(M+H)^+$ 282.1197, found 282.1195

Data in accordance with literature

7.2.2.3 5-Phenylselenylmethyl-2´-Deoxyuridine 5´-Triphosphate[286]

7.2.2.3.1 Method 1[286,288]



5-Phenylselenylmethyl-2´-deoxyuridine (52 mg, 0.13 mmol) and 1,8-bis-(dimethylamino)naphthalene (Proton Sponge®, 42 mg, 0.19 mmol) were dissolved in trimethyl phosphate (0.6 mL). The solution was stirred at 0 °C for 30 minutes. Phosphorous oxytrichloride (27 μL, 0.29 mmol) was then added and the reaction was stirred at 4 °C for 3 hrs.

Tributylammonium pyrophosphate (308 mg, 0.65 mmol) was dissolved in anhydrous dimethylformamide (1.3 mL) and tributylamine (130 μL, 0.55 mmol) was added. This solution was added dropwise to the reaction over the course of 5 mins. The reaction was stirred for 2 hrs at room temperature and was then quenched by the addition of 1 M triethylammonium bicarbonate buffer (12 mL, pH 8.5), and stirred for 10 mins at room temperature. The solution was lyophilised, dissolved in methanol (200 μL) and mixed with a solution of 0.75 M NaClO4 in acetone (1.2 mL). The solution was incubated at room temperature for 10 mins, followed by -80 °C for 10 mins. The solution was centrifuged at 13,000 x g for 10 mins and the supernatant was discarded. The pellet was then washed 3 times with acetone (1 mL) to remove excess $NaClO_4$.

The product was then purified by DEAE Sephadex A-25 column chromatography. The Sephadex resin was swelled in NaCl (200 mM) at room temperature for 3 days and was packed into an empty column (2 x 10 cm) using a peristaltic pump. The column washed with 3 bed volumes of H_2O. The 2 μL reaction was loaded onto the column. Chromatography was performed using a gradient of 100 % H_2O to 100 % 1 M triethylammonium bicarbonate buffer (pH 8.5) at a flow rate of 6 ml/min. Seventy 15 ml fractions were collected and fractions eluted with between 400 mM and 800 mM triethylammonium bicarbonate buffer were lyophilised. Fractions containing the product were identified by mass spectrometry, and combined. The product was then purified by reverse phase HPLC using a BETASIL C18 250 x 10 mm column and the conditions outlined in Table 16.

Table 16: PhSe-dTTP HPLC purification elution gradient. BETASIL C18 250 x 10 mm column. 3 mL/min. Buffer A: TEAA (50 mM) in H_2O, Buffer B: HPLC grade acetonitrile.

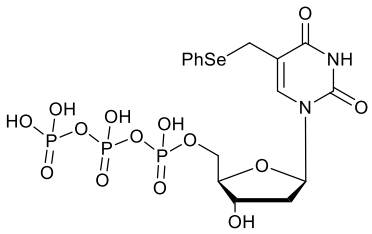
Time (mins)	Buffer A (%)	Buffer B (%)
0	95	5
5	95	5
40	80	20
50	20	80
55	95	5
60	95	5

Fractions containing the triphosphate product were dissolved in methanol (200 μL) and mixed with a solution of 0.75 M $NaClO_4$ in acetone (1.2 m). The solution was incubated at room temperature for 10 mins, followed by -80 °C for 10 mins. The solution was centrifuged at 13,000 x g for 10 mins and the supernatant was discarded. The pellet was then washed 3 times with acetone (1 mL). The precipitate was dissolved in H_2O, and the concentration of the nucleoside triphosphate was calculated using the extinction coefficient at 270 nm (10,100 M-1cm-1) (Hong, Ding, Greenberg, 2007).

HRMS (ESI-) calculated for $C_{16}H_{20}N_2O_{14}P_3Se$ (M-H)- 636.9298, found 636.9306

Data in accordance with literature.

7.2.2.3.2 Method 2[113,289]



Tributylammonium pyrophosphate (308 mg, 0.65 mmol) was dissolved in anhydrous dimethylformamide (250 μL) and triethylamine (750 μL). The solution was stirred at room temperature for 5 mins. 2-chloro-4-H-1,3,2- benzodioxaphosphorin-4-one (20 mg, 0.1 mmol) dissolved in DMF (1 mL) was then added and the solution was stirred at room temperature for 30 mins. 5-Phenylselenylmethyl-2´-deoxyuridine (20 mg, 0.05 mmol) was then added and the reaction was stirred at room temperature for 1.5 hours. A solution of iodine (3 %) dissolved in pyridine/ water (9:1) was slowly added until a brown colour was maintained. The solution was stirred at room temperature for 5 mins. H_2O (5 mL) was added and the solution was stirred at room temperature for 1.5 hours.

The solution was transferred to two 50 mL centrifuge tubes. NaCl (3M, 0.6 mL) and EtOH (19 mL) were then added and the solution was vortexed. The solution was incubated at -80°C for 1 h. The sample was centrifuged (3200 x g, 20 mins) and the supernatant discarded. The nucleoside triphosphate was then HPLC purified and precipitated with $NaClO_4$ in acetone (as described in method 1).

7.2.3 Oligonucleotide synthesis

An Applied Biosystems 3400, or 394 DNA Synthesizer was used for oligonucleotide synthesis.

7.2.3.1 DNA oligonucleotide synthesis and purification

The standard 0.2 µM DNA synthesis cycle was used with 0.2 µM solid CPG supports. The coupling time was set to 20 seconds for unmodified phosphoramidites and 200 seconds for modified phosphoramidites. Following the final coupling the 5' most DMTr group was not cleaved. The columns were vacuum dried for 20 minutes. The support bound oligonucleotides were then incubated in 2 mL concentrated ammonium hydroxide, at 55 °C, overnight to facilitate cleavage from the solid support and to remove the base-labile protecting groups. PhSeT containing oligonucleotides were deprotected at room temperature for 48 hrs. The Oligonucleotides were then purified and detritylated using an OPC cartridge. The oligonucleotides were then desalted using NapTM-5 columns and eluted in H_2O (1 mL).

7.2.3.2 RNA oligonucleotide synthesis and purification

The standard 0.2 µM RNA synthesis cycle was used with 0.2 µM solid CPG supports. The coupling time was set to 15 mins. The final coupling was followed by cleavage of the 5'-DMTr group on the synthesiser. Once synthesised, the oligonucleotides were deprotected and purified using standard

conditions[197]. Firstly, the support bound oligonucleotides were resuspended in $MeNH_2$ (1 mL) and incubated for 10 mins at 65 °C to cleave the base protecting and cyanoethyl groups. The solution was then incubated on ice for 15 mins and centrifuged (1 min, 10,000 x g). The supernatant was removed, and the beads were washed with RNase free H_2O (0.25 mL) and centrifuged. This washing step was repeated, and the supernatants were combined and lyophilized under vacuum. The 2′-O-TBDMS groups were then cleaved by resuspending the beads in 250 µL of an anhydrous solution of NEt3·3HF/NEt3/NMP (1.5 mL NMP, 750 µL NEt3 and 1.0 mL NEt3·3HF). The solution was incubated for 1.5 h at 65 °C, cooled to room temperature, briefly centrifuged, and quenched by the addition of 3M NaOAc (25 µL). The oligonucleotides were precipitated from the solution by the addition of *n*-BuOH (1 mL). The solution was briefly vortexed and incubated at -70 °C for 2 h. The solution was centrifuged for 30 mins (13,000 x g) and the supernatant discarded. The pellet was then washed twice with 70 %, v/v EtOH (500 µL) and dried under vacuum (1 hr). The pellet was resuspended in 1 mL RNase free H_2O and desalted using a NapTM-5 column.

7.2.3.1 Oligonucleotide HPLC purification

Phenyl selenide modified thymidine oligos were subject to purification by HPLC. A BETASIL C18 250 x 10 mm column was used with the conditions described in Table 17.

Table 17: HPLC oligonucleotide purification elution gradient. BETASIL C18 250 x 10 mm column. 3 mL/min. Buffer A: TEAA (50 mM) in H_2O, Buffer B: HPLC grade acetonitrile.

Time (mins)	**Buffer A (%)**	**Buffer B (%)**
0	92.5	7.5
5	92.5	7.5
35	80	20
40	80	20
41	92.5	7.5
45	92.5	7.5

7.2.3.1 Oligonucleotide PAGE purification

Certain oligonucleotides were purified using a 15 % Mini-PROTEAN® TBE-Urea Gel pre-cast gel. The oligonucleotide was visualised by soaking the gel in a solution of 'SYBR™ Gold Nucleic Acid Gel Stain' (5 µL, 50,000 x) and TBE buffer (50 mL, 1 x) for 30 mins. The corresponding band was excised and added to a 1.5 ml Eppendorf tube containing Crush and Soak buffer (400 µL, KCl (200 mM), Potassium Acetate (50 mM), pH 7), the tube was rotated for 4 hrs at room temperature. The DNA/RNA was recovered by ethanol precipitation (See section 7.2.4.2) and the pellet was dissolved in an appropriate volume of H_2O

7.2.3.2 Oligonucleotide quantification.

The concentration of the oligonucleotide was calculated using a Thermo Scientific NanoDrop™ 1000 Spectrophotometer.

7.2.3.3 Oligonucleotide HPLC analysis

The purity of the synthesised DNA and RNA oligonucleotides was analysed by analytical HPLC. A Phenomenex Clarity 3 µ Oligo-RP 50 × 4.6 mm column was used with the conditions described in Table 18.

Table 18: HPLC elution gradient for oligonucleotide analysis. Phenomenex Clarity 3 µ Oligo-RP 50 × 4.6 mm column. 1 mL/min. Buffer A: TEAA (100 mM) in H_2O, Buffer B: HPLC grade acetonitrile.

Time (mins)	**Buffer A (%)**	**Buffer B (%)**
0	95	5
5	95	5
25	80	20
30	80	20
31	95	5
35	95	5

Table 19: Results of oligonucleotide HPLC analysis

Oligonucleotide	**DNA/RNA**	**Length (nt)**	**Retention time (mins)**
Cap-1	RNA	28	20.27
Cap-2	RNA	28	19.14
Cap-3	RNA	28	19.17
1agAgc	RNA	28	13.43
2ga6cg	RNA	28	13.68
3ag6gc	RNA	28	13.62
4gaAcg	RNA	28	13.51
5N9	RNA	28	13.40
Chicken β-actin zipcode – unmodified	RNA	28	13.73
Chicken β-actin zipcode – m^6A (site 1202)	RNA	28	14.03
Chicken β-actin zipcode – m^6A (site 1211)	RNA	28	12.64
Chicken β-actin zipcode – m^6A (site 1212)	RNA	28	14.26
Chicken β-actin zipcode – m^6A (site 1218)	RNA	28	13.94
A^{18}	RNA	18	10.23
$G^1m^6A^{17}$	RNA	18	19.75
$T^9PhSeT^1T^8$	DNA	18	11.01

T^{18}	DNA	18	17.72
Mouse β-actin – unmodified	RNA	19	13.61
Mouse β-actin – m^6A modified 1	RNA	19	13.41
Mouse β-actin – m^6A modified 2	RNA	19	15.30
Mouse β-actin – PhSeT modified 1	DNA	28	17.47
Mouse β-actin – PhSeT modified 2	DNA	28	17.32
SCARLET chimera – chicken β-actin – A1	2'-O-Me/DNA	20	17.12
SCARLET chimera – chicken β-actin – A4	2'-O-Me/DNA	18	16.79
SCARLET chimera – chicken β-actin – A6	2'-O-Me/DNA	20	15.91
RedBaron chimera – chicken β-actin – A4	2'-O-Me/DNA	18	17.64
RedBaron chimera – chicken β-actin – A7	2'-O-Me/DNA	20	17.38

7.2.3.4 Oligonucleotide mass spectrometry analysis

The identity of the oligonucleotide was confirmed by mass spectrometry using a Thermo Scientific: LTQ FT Ultra Mass Spectrometer in negative mode. Mass ions were calculated using the equation $p_1z_1 = Mr - M_az_1 = Mr - 1.0079$ z_1 (p_1 is the peaks m/z value, z_1 is the peaks charge, Mr is the molecular weight of the oligonucleotide, and M_a is the molecular weight of the charge-carrying

species (which is a proton)[295]. The peak charge used for the calculation was either: -7, -8, -9, -10 -11, -12, -13, or -14, depending on the oligonucleotide analysed.

Table 20: Oligonucleotide mass spectrometry analysis

Oligonucleotide	**Expected mass**	**Adduct**	**Observed mass**
Cap-1 (A)	9493.4806	M-12H	9493.4952
Cap-2 (A_m)	9507.4963	M-11H	9507.5442
Cap-3 (m^6A_m)	9521.5119	M-12H	9521.5296
Chicken β-actin – unmodified	8794.2606	M-13H	8794.2634
Chicken β-actin – m^6A modified 1	8808.2763	M-13H	8808.2852
Chicken β-actin – m^6A modified 2	8808.2763	M-12H	8808.2676
Chicken β-actin – m^6A modified 3	8808.2763	M-11H	8808.2940
Chicken β-actin – m^6A modified 4	8808.2763	M-11H	8808.2863
A^{18}	5860.9896	M-7H	5861.0139
$G^1m^6A^{17}$	6115.2506	M-7H	6115.2455
$T^9PhSeT^1T^8$	5566.8207	M-6H	5566.8444
T^{18}	5410.8729	M-6H	5410.8924
Mouse β-actin – unmodified	6051.8034	M-8H	6051.8224

Mouse β-actin – m^6A modified 1	6065.8190	M-7H	6065.8220
Mouse β-actin – m^6A modified 2	6065.8190	M-7H	6065.8220
Mouse β-actin – PhSeT modified 1 (control)	8694.4155	M-11H	8694.4473
Mouse β-actin – PhSeT modified 2 (consensus)	8694.4155	M-11	8694.4418
SCARLET chimera – chicken β-actin – A1	6558.1055	M-8H	6558.1000
SCARLET chimera – chicken β-actin – A4	6043.0462	M-7H	6043.0580
SCARLET chimera – chicken β-actin – A6	6650.1655	M-8H	6650.1792
RedBaron chimera – chicken β-actin – A4	6059.0411	M-7H	6059.0467
RedBaron chimera – chicken β-actin – A7	6625.1590	M-8H	6625.1752

7.2.4 General molecular biology methods

7.2.4.1 Phenol chloroform extraction of RNA

The RNA solution was mixed with an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1). The solution was centrifuged (5 mins at 14,000 x g) and the upper aqueous phase was transferred to a fresh Eppendorf tube. The RNA solution was mixed with chloroform: isoamyl alcohol (24:1) (200 μL)

and the solution was centrifuged (5 mins at 13,000 x g). The upper aqueous phase was transferred to a fresh Eppendorf tube and the RNA was purified by ethanol precipitation.

7.2.4.2 Ethanol precipitation of DNA and RNA

The RNA solution was mixed with 3 M NaOAc ($1/10^{th}$ of the volume, pH 5.2), ethanol (3 volumes), and glycogen (1 μL, 20 μg). The solution was incubated at -20 °C for a minimum of 1 hr. The solution was centrifuged (30 mins at 13,000 x g), and the supernatant discarded. The pellet was then washed three times with ethanol (500 μL, 75 % v/v). The Eppendorf tube was incubated at room temperature for 20 mins with the lid open to allow residual ethanol to evaporate. The pellet was dissolved H_2O, and the concentration of the RNA was analysed using a Nanodrop spectrophotometer.

7.2.4.3 5’ radiolabelling of oligonucleotides

Oligonucleotides were radiolabelled with ^{32}P at their 5’- end. The oligonucleotide (1 μL, 5 pmol) was mixed with ATP [γ-^{32}P] (2 μL, 6.6 pmol, 20 μCi), T4 PNK buffer A (2 μL 10 x), H_2O (14 μL), and T4 Polynucleotide kinase (1 μL, 10 units). The solution was then incubated for 1 hr at 37 °C. The labelled oligonucleotides were purified using a QIAquick Nucleotide Removal Kit to remove unincorporated ATP and eluted in H_2O (30 μL).

7.2.4.4 Denaturing PAGE analysis of radioactively labelled oligonucleotides

Radiolabelled oligonucleotides were analysed using a 15 % (w/v) Mini-PROTEAN® TBE-Urea denaturing gel. Approximately 100 cps in a volume of 10 μL was loaded into each well. The gel was run in TBE buffer (1 x) at 180 volts for approximately 1 hr.

The gel was prepared for visualization using one of two methods:

1) The gel was soaked in a solution of TBE (1 %)/ methanol (4:1) for 20 mins. The gel was then dried on a gel drier for (3 hrs, 70 ^{o}C).
2) The RNA was transferred from the gel to a 'GeneScreen Plus Hybridization Transfer Membrane' using a 'Mini Trans-Blot Module' ran at 100 volts for 30 mins in a buffer of TBE (0.5 x).

The dried gel/ membrane was then exposed to a phosphor imager screen and visualised using a phosphor imager.

7.2.4.5 Two-dimensional TLC analysis

Two-dimensional thin layer chromatography was used for the analysis of 3'- and 5'- nucleoside monophosphates. The RNA/DNA solution (~1 μL, ~20 counts per second) was spotted onto the bottom left corner of a glass backed cellulose F TLC plate (20 x 20 cm). The plate was resolved in two dimensions. The first-dimension buffer was composed of isobutyric acid: NH_4OH (0.5 M) (5:3, v/v). The second-dimension buffer was composed of isopropanol:HCl:water (70:15:15, v/v/v). Each dimension was run for

approximately 16 hours at room temperature. The TLC plate was then dried at room temperature for 3 hrs, exposed to a phosphorimager screen and visualised using a phosphor imager.

7.2.4.6 One-dimensional TLC analysis

One dimensional TLC was achieved using a buffer composed of isopropanol: HCl: H_2O (70:15:15, v/v/v). The TLC plate was dried at room temperature for 3 hrs, exposed to a phosphor imager screen and visualised using a phosphor imager.

7.2.5 RNA purification

7.2.5.1 total RNA purification from tissue samples

A pestle and mortar were chilled using liquid nitrogen and used to grind approximately 1 cm^3 of tissue sample into a fine powder. Liquid nitrogen was used to ensure the sample remained frozen during this process. The powder was added to an Eppendorf tube containing SDS (40 µL, 10 % w/v) and AG buffer (400 µL) (NaOAc (50 mM, pH 5.2), EDTA (10 mM)). The samples were vortexed, mixed with phenol (400 µL) and vortexed again. The solution was incubated at 65 °C for 5 mins. The samples were briefly dipped in liquid nitrogen, centrifuged (5 mins at 13,000 x g) and the supernatant was transferred to a fresh 1.5 mL Eppendorf. The RNA was purified by a phenol chloroform extraction (See section 7.2.4.1) followed by ethanol precipitation (See section 7.2.4.2). The RNA pellet was dissolved in H_2O, and the

concentration was analysed using a Nanodrop spectrophotometer. The purity of the extracted RNA was analysed on an agarose gel (1 % w/v).

7.2.5.2 Total RNA purification from cell samples

The cell suspension was transferred to 1.5 mL Eppendorf tube (5-6 million cells per tube). The cells were centrifuged (15 mins, 12,000 × g) and the supernatant was discarded. The pellet was mixed with TRIzol™ (750 µL) by pipetting the solution up and down several times. The solution was incubated at room temperature for 5 mins. The solution was mixed with chloroform (150 µL) and incubated at room temperature for 3 mins. The solution was centrifuged (15 mins, 13,000 x g) and the upper aqueous phase was transferred to a new Eppendorf tube. Isopropanol (375 µL) and glycogen (1 µL, 20 µg) were mixed with the supernatant and the solution was incubated at room temperature for 10 mins. The solution was centrifuged (15 mins, 12,000 × g) and the supernatant discarded. The pellet was resuspended in ethanol (750 µL, 75 % v/v), briefly vortexed, and centrifuged (5 mins, 7500 x g). The supernatant was discarded, and the Eppendorf tube was incubated at room temperature for 20 mins with the lid open to allow residual isopropanol to evaporate. The pellet was dissolved H_2O, and the concentration of the RNA was analysed using a Nanodrop spectrophotometer.

7.2.5.3 poly(A)+ RNA purification

NEBNext Magnetic Oligo d(T)25 Beads (200 µL) were added to a 1.5 mL Eppendorf tube. The supernatant was separated from the magnetic beads using a magnetic rack and discarded. The magnetic beads were washed 2 times with

NEBNext RNA binding buffer (2 x, 100 μL) (Table 21). The magnetic beads were resuspended in NEBNext RNA binding buffer (2 x, 50 μL) and total RNA in H_2O (50 μL, 25 μg). The solution was incubated at 65 °C for 5 mins, followed by room temperature for 10 mins to anneal the RNA to the d(T)25 magnetic beads. The supernatant was discarded using the magnetic rack. The beads were then washed two times with NEBNext wash buffer (1 x, 200 μL) (Table 21). The dry beads were then resuspended in H_2O (50 μL). The solution was incubated at 80 °C for 2 mins, followed room temperature for 5 mins to denature the RNA. The solution was mixed with NEBNext binding buffer (2 x, 50 μL) and incubated at room temperature for 5 mins to anneal the RNA d(T)25 magnetic beads. The supernatant was discarded, and the beads were washed with NEBNext wash buffer (1 x, 200 μL). The dry beads were resuspended in H_2O (20 μL), incubated at 80 °C for 2 mins to denature the RNA and the supernatant was removed from the magnetic beads and transferred to a fresh Eppendorf tube. The concentration of the mRNA was analysed using a Nanodrop spectrophotometer.

Table 21: Buffers used for the purification of poly(A)+ mRNA

Buffer	**Components**
NEBNext RNA binding buffer (2 x)	LiCl (2 M), Tris-HCL (40 mM, pH 7.5), EDTA (2 mM), NP40 (0.1 %, v/v)
NEBNext wash buffer (1 x)	LiCl (150 mM), Tris-HCL (20 mM, pH 7.5), EDTA (1 mM), NP40 (0.01 %, v/v)

7.2.6 SCARLET and RedBaron experiments

7.2.6.1 Synthesising 5’ labelled mononucleotide reference molecules

A 4 kb RNA molecule was synthesised using a Riboprobe System T7 *in vitro* transcription kit. pGEM® Express Positive Control Template (1 µL, 1 µg), was mixed with: rATP (1 µL, 10 mM), rGTP (1 µL, 10 mM), rCTP (1 µL, 10 mM), rUTP (1 µL, 10 mM), Transcription Optimized (5 x) Buffer, Recombinant RNasin® Ribonuclease Inhibitor (2 µL, 40 units), DTT (2 µL, 100 mM), and T7 RNA Polymerase (1 µL, 20 units), in a total volume of 20 µL. The solution was incubated at 37 °C for 1 hr. RQ1 RNase-Free DNase (1 µL, 1 unit) of was then added and the solution was incubated at 37 °C for 15 mins.

The RNA was purified by a phenol chloroform extraction (See section 7.2.4.1) followed by ethanol precipitation (See section 7.2.4.2). The RNA pellet was resuspended in H_2O (26 µL) and mixed with PNK buffer (3 µL, 10 x) , and T1 ribonuclease (1 µL, 2000 units). The solution was incubated at 37 °C for 1 hr. PNK buffer (2 µL, 10 x), ATP [γ-^{32}P] (3.3 pmol, 10 µCi), and T4 PNK (1 µL, 10 units) were then added. The solution was incubated for 1 hr at 37 °C. The enzyme was inactivated by incubating at 65 °C for 20 mins. The RNA was purified by ethanol precipitation and the RNA pellet was resuspended in a solution of H_2O (17 µL), nuclease P1 (2 µL, 1 unit), and NaOAc (1 µL, 1 M,

pH 5.3). The solution was incubated at 37 °C for 2 hrs to digest the RNA into 5'- nucleoside monophosphates.

7.2.6.2 Synthesising 3' labelled mononucleotide reference molecules

A 4 kb RNA molecule was synthesised using a Riboprobe System T7 *in vitro* transcription kit. pGEM® Express Positive Control Template (1 µL, 1 µg), was mixed with: rCTP [α-^{32}P] (1 µL, 10 µCi, 3.3 µM), rCTP (1 µL, 10 mM), rATP (1 µL, 10 mM), rGTP (1 µL, 10 mM), rCTP (1 µL, 10 mM), rUTP (1 µL, 10 mM), Transcription Optimized (5 x) Buffer, Recombinant RNasin® Ribonuclease Inhibitor (2 µL, 40 units), DTT (2 µL, 100 mM), and T7 RNA Polymerase (1 µL, 20 units), in a total volume of 20 µL. The solution was incubated a 37 °C for 1 hr. RQ1 RNase-Free DNase (1 µL, 1 unit) and Calf Intestinal Alkaline Phosphatase (1 µL, 20 units) were then added, and the solution was incubated at 37 °C for 15 mins.

The RNA was purified by a phenol chloroform extraction (See section 7.2.4.1) followed by ethanol precipitation (See section 7.2.4.2). The RNA was then purified using a QIAquick nucleotide removal kit. The RNA was eluted in nuclease free H_2O (30 µL). NEB buffer 3 (4 µL, 10 x), RNase I (1 µL, 10 units), and H_2O (6 µL) were then added, and the solution was incubated at 37 °C for 3 hrs to digest the RNA into 3' nucleoside monophosphates.

Synthesis of adenosine and m^{6}A radiolabelled 3'- nucleoside monophosphates that did not contain any ^{32}P-G/C/A was achieved using the same method however the oligonucleotides 5'-

CGTCTAGTCCATGTACGTACCCTATAGTGAGTCGTATTA-3', and 5'-TAATACGACTCACTATAGG-3' were used in place of the pGEM® Express Positive Control Template.

7.2.6.3 SCARLET protocol[90]

7.2.6.3.1 RNase H digestion and dephosphorylation

Poly(A)+ RNA (1 μg) was mixed with the chimeric oligonucleotide (3 pmol) in Tris-HCl (3 μL, 30 mM, pH 7.5). The oligonucleotide was annealed to the RNA by incubating the solution at 95 °C for 1 min followed by incubation at 20 °C for 10 mins. FastAP Thermosensitive Alkaline Phosphatase (0.4 μL, 0.4 units), RNase H (0.4 μL, 2 units), and T4 PNK buffer (0.2 μL, 10 x) were then added and the solution was incubated at 44 °C for 1 hr. The solution was then incubated at 75 °C for 5 mins to inactivate the RNase H and alkaline phosphatase enzymes.

7.2.6.3.2 ^{32}P Radiolabelling

Radiolabelling of 5' hydroxyl groups was achieved by adding T4 PNK (0.8 μL, 8 units), ATP [γ-^{32}P] (2.8 μL, 28 μCi, 9.24 pmol), T4 PNK buffer (0.4 μL, 8 x) and incubating at 37 °C for 1 hr. The PNK enzyme was then inactivated by incubating the solution at 75 °C for 5 mins.

7.2.6.3.3 Splint assisted ligation

The splint oligonucleotide (0.5 μL, 4 pmol) and the ssDNA oligonucleotide (0.5 μL, 5 pmol) were added to the reaction. The oligonucleotides were annealed to the mRNA by incubating the solution at 75 °C for 5 mins,

followed by 20 °C for 10 mins. T4 PNK buffer (0.4 µL, 10 x), ATP (0.2 µL, 60 mM), DMSO (1.2 µL), and T4 DNA ligase (1 µL, 400 units) were then added, and the solution was incubated at 37 °C for 3.5 hrs.

7.2.6.3.4 Dephosphorylation, and RNase digestion

The ligase enzyme was inactivated by adding RNA loading buffer (14 µL, 2 x, Urea (9 M) and EDTA (100 mM)). RNase T1 (0.5 µL, 500 units), and RNase A (0.5 µL, 5 ng) were then added and the solution was incubated at 37 °C overnight to digest all of the RNA in the sample except for the A/m^6A nucleotide that was ligated to the ssDNA.

7.2.6.3.5 Purification, RNase Digestion, and TLC (two-dimensional TLC)

The DNA/RNA was purified using a QIAquick Nucleotide Removal Kit and eluted in H_2O (30 µL). 8 µL of this DNA/RNA solution was mixed with Nuclease P1 (1 µL, 100 units), and Nuclease P1 buffer (1 µL, 10 x). The solution was incubated at 37 °C for 2 hrs to digest the DNA/RNA into 5'-nucleoside monophosphates. 1 µL of this solution was spotted onto a TLC plate and ran in two dimensions (See section 7.2.4.5).

7.2.6.3.6 Purification, RNase Digestion, and TLC (1D TLC)

The DNA/RNA was purified using a 15 % Mini-PROTEAN® TBE-Urea Gel pre-cast gel. The DNA/RNA was visualised by soaking the gel in a solution of 'SYBR™ Gold Nucleic Acid Gel Stain' (5 µL, 50,000 x) and TBE buffer (50 mL, 1 x). The corresponding band was excised and added to a 1.5 mL Eppendorf tube containing Crush and Soak buffer (400 µL, KCl (200 mM),

Potassium Acetate (50 mM), pH 7), the tube was rotated for 4 hrs at room temperature. The DNA/RNA was recovered by ethanol precipitation (See section 7.2.4.2). The DNA/RNA pellet was resuspended in H_2O (3 µL), Nuclease P1 (1 µL, 100 units), and Nuclease P1 buffer (1 µL, 5 x). The solution was incubated at 37 °C for 2 hrs to digest the DNA/RNA into 5' nucleoside monophosphates. 1 µL of the reaction mixture was spotted onto a TLC plate and ran in a single dimension (See section 7.2.4.6).

7.2.6.4 RedBaron protocol

7.2.6.4.1 RNase H digestion and dephosphorylation

Poly(A)+ RNA (1 µg) was mixed with the chimeric oligonucleotide (3 pmol) in Tris-HCl (3 µL, 30 mM, pH 7.5). The oligonucleotide was annealed to the RNA by incubating the solution at 95 °C for 1 min followed by incubation at 20 °C for 10 mins. RNase H (1 µL, 5 units), and RNase H buffer (1 µL, 5 x) were then added and the solution was incubated at 37 °C for 1 hr. The solution was then incubated at 75 °C for 5 mins to inactivate the enzyme.

7.2.6.4.2 ^{32}P Radiolabelling of ssDNA oligonucleotide

In a separate tube, the ssDNA oligonucleotide (1 µL, 5 pmol) was mixed with ATP [γ-^{32}P] (2 µL, 6.6 pmol, 20 µCi), and T4 Polynucleotide kinase (1 µL, 10 units) in a final volume of 10 µL T4 PNK buffer A (1 x). The reaction was incubated at 37 °C for 1 hr, followed by 65°C for 20 mins to inactivate the enzyme.

7.2.6.4.3 Splint assisted ligation and dephosphorylation

The ssDNA PNK reaction (10 μL) was combined with the RNase H reaction (5 μL) and the splint oligonucleotide (1 μL, 4 pmol). The oligonucleotides were annealed to the mRNA by incubating the solution at 75 °C for 5 mins, followed by 20 °C for 10 mins. SplintR® Ligase buffer (1.5 μL, 10 x), ATP (1.5 μL, 10 mM), PEG (3 μL, 50 % v/v), H_2O (6 μL), and T4 SplintR® Ligase (1 μL, 25 units) were then added and the solution was incubated at 37 °C for 3.5 hrs.

FastAP Thermosensitive Alkaline Phosphatase (1 μL, 1 unit) was added and the solution was incubated at 37 °C for 30 mins. EDTA (3 μL, 500 mM) was added and the solution was incubated at 75 °C for 5 mins to inactivate the ligase and alkaline phosphatase enzymes.

7.2.6.4.4 Purification, RNase Digestion, and TLC (two-dimensional TLC)

The DNA/RNA was purified using a QIAquick Nucleotide Removal Kit and eluted in H_2O (30 μL). 6 μL of this solution was mixed with Micrococcal Nuclease Reaction Buffer (1 μL, 10 x), Purified BSA (1 μL, 2 μg), and Micrococcal Nuclease (1 μL, 2000 units). The solution was incubated at 37 °C for 3 hrs to digest the DNA/RNA into 3’ nucleoside monophosphates. 1 μL of the reaction mixture was spotted onto a TLC plate and ran in two dimensions (See section 7.2.4.5).

7.2.6.5 Synthesis of a biotinylated RNA probe for RedBaron experiments

7.2.6.5.1 Mouse cDNA synthesis

Mouse total RNA (5 μg) was mixed a Poly(T)$_{23}$ DNA oligonucleotide primer (200 pmol), dATP (10 nmol), dGTP (10 nmol), dCTP (10 nmol), and dTTP (10 nmol), in a volume of 10 μL H_2O. The solution was incubated at 65 °C for 5 mins and placed immediately onto ice to anneal the primer to the mRNA. H_2O (6 μL), AMV buffer (2 μL, 10 x), AMV reverse transcriptase (1 μL, 10 units), and RNase Inhibitor (1 μL, 40 units) were then added. The solution was incubated at 42 °C for 1 hr. The enzymes were then inactivated by incubating the solution at 80 °C for 5 mins.

7.2.6.5.2 PCR amplification of *in vitro* transcription template

PCR was used to synthesise a template for the T7 *in vitro* transcription of a RNA probe with complementarity to the mouse β-actin zipcode. The forward primer contains the T7 promoter at its 5'- end. The mouse cDNA (1 μL) was mixed with H_2O (12 μL), Phusion HF buffer (4 μL, 5 x), dNTP mix (1 μL, (dATP, dGTP, dCTP, dTTP), 10 nmol each), Forward primer (1 μL, 10 pmol), Reverse primer (1 μL, 10 pmol), Phusion Hot Start II DNA Polymerase (1 μL, 2 units). The solution was incubated at 98 °C for 30 secs, followed by 30 cycles of incubation at 98 °C for 10 secs, 64 °C for 30 secs, and 72 °C for 30 secs. This was followed by a final incubation at 72 °C for 10 mins. The cDNA product was visualised using agarose gel electrophoresis (1 %).

Table 22: Primers used for the synthesis of a template for the *in vitro* transcription of an RNA probe with complementarity to the mouse β-actin zipcode. The Forward primer contains the T7 promoter sequence at its 5'- end (highlighted yellow).

Primer	**Sequence (5'-3')**
Forward primer	TAATACGACTCACTATAGGAGTACGATGAGTCC GGCCCCTC
Reverse primer	GCGCAAGTTAGGTTTTGTCAAAGAAAG

7.2.6.5.3 *In vitro* transcription of biotinylated RNA

A Riboprobe System T7 *in vitro* transcription kit was used to synthesise the biotinylated RNA probe. The mouse β-actin zipcode T7 *in vitro* transcription template (21 µL, 1 µg) was mixed with Transcription optimised buffer (8 µL, 5 x), DTT (2 µL 100 mM) Recombinant RNasin® Ribonuclease Inhibitor (2 µL, 40 units), rATP (1 µL, 10 mM), rCTP (1 µL, 10 mM), and rGTP (1 µL, 10 mM), rUTP (1 µL, 3.5 mM), biotin-16-UTP (1 µL, 6.5 mM), and T7 RNA Polymerase (1 µL, 20 units). The solution was incubated a 37 °C for 1 hr. RQ1 RNase-Free DNase (1 µL, 1 unit) was then added and the solution was incubated at 37 °C for 15 mins. The RNA was then purified by a phenol chloroform extraction (See section 7.2.4.1) followed by ethanol precipitation (See section 7.2.4.2). The concentration of the RNA was calculated using a NanoDrop spectrophotometer.

7.2.6.6 RedBaron method using streptavidin magnetic beads

7.2.6.6.1 Purification of β-actin mRNA using biotinylated RNA probe

The biotinylated RNA probe (400 ng) was mixed with hydrophilic streptavidin magnetic beads (25 µL, and mouse total RNA (10 µg) in Binding buffer (50 µL, 1 x) (0.5 M NaCl, 20 mM Tris HCl (pH 7.5), 1 mM EDTA). The solution was incubated at room temperature for 5 mins to bind the biotin to the streptavidin magnetic beads. The solution was then incubated at 90 °C for 1 min followed by room temperature for 10 mins to anneal the mRNA to the biotinylated RNA probe. The magnetic beads were washed twice with Tris HCl (200 µL, 30 mM, pH 7.5) to remove unbound RNA.

7.2.6.6.2 RNase H digestion and dephosphorylation

The dry magnetic beads were mixed with the chimeric oligonucleotide (3 pmol) in Tris-HCl (10 µL 30 mM, pH 7.5). The solution was incubated at 95 °C for 1 min followed by room temperature for 10 mins to anneal the chimera. The supernatant was discarded, and the magnetic beads were resuspended in H_2O (17 µL). RNase H Reaction Buffer (2 µL, 10 x) and RNase H (1 µL, 5 units) were added and the solution was incubated at 37 °C for 1 hr. EDTA (2 µL, 500 mM) was added and the solution was incubated at 75 °C for 5 mins to inactivate the enzyme. The supernatant was then discarded.

7.2.6.6.3 ^{32}P Radiolabelling

In a separate tube, the ssDNA oligonucleotide (1 µL, 5 pmol) was mixed with ATP [γ-^{32}P] (2 µL, 6.6 pmol, 20 µCi), and T4 Polynucleotide kinase (1 µL, 10 units) in a final volume of 10 µL T4 PNK buffer A (1 x). The reaction was incubated at 37 °C for 1 hour, followed by 65°C for 20 mins to inactivate the enzyme.

7.2.6.6.4 Splint assisted ligation and dephosphorylation

The dry magnetic beads were mixed with the splint oligonucleotide (1 µL, 4 pmol), the radiolabelled ssDNA (30 µL, 5 pmol) and Tris HCl (5 µL, 360 mM). The solution was incubated at 75 °C for 3 mins, followed by room temperature for 10 mins to anneal the DNA and RNA. The magnetic beads were then washed with SplintR® Ligase buffer (100 µL, 1 x). The dry magnetic beads were mixed with H_2O (15 µL), SplintR® Ligase buffer (2 µL, 10 x), PEG (2 µL, 50 % v/v), and T4 SplintR® Ligase (1 µL, 25 units) and the solution was incubated at 37 °C for 3.5 hrs.

FastAP Thermosensitive Alkaline Phosphatase (1 µL, 1 unit) was added and the solution was incubated at 37 °C for 30 mins. EDTA (3 µL, 500 mM) was added and the solution was incubated at 75 °C for 5 mins to inactivate the ligase and alkaline phosphatase enzymes.

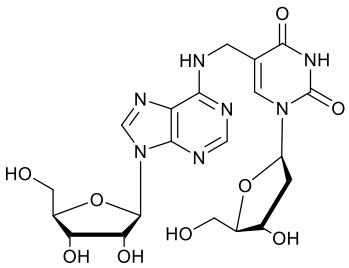
7.2.6.6.5 Purification, RNase Digestion, and TLC (two-dimensional TLC)

The supernatant was separated from the magnetic beads and discarded. The magnetic beads were mixed with H_2O (7 µL), Micrococcal Nuclease Reaction

Buffer (1 µL, 10 x), Purified BSA (1 µL, 2 µg), and Micrococcal Nuclease (1 µL, 2000 units). The solution was incubated at 37 °C for 3 hrs to digest the DNA/RNA into 3’ nucleoside monophosphates. 1 µL of the reaction mixture was spotted onto a TLC plate and ran in two dimensions.

7.2.7 Phenyl selenide thymidine crosslinking

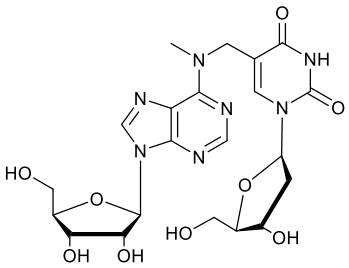
7.2.7.1 Nucleoside crosslinking of PhSeT and adenosine[17,259,272]



5-phenylselenylmethyl-2'-deoxyuridine (6 mg, 15 µmol) and adenosine (4 mg, 15 µmol) were dissolved in H_2O (125 µL). Hydrogen peroxide (25 µL, 0.075 mmol) was then added and the solution was stirred at room temperature for 3 hrs. The solution was concentrated *in vacuo* and aqueous piperidine (1M, 125 µL, 0.125 mmol) was added. The solution was then incubated at 90 °C for 20 mins. The mixture was concentrated *in vacuo* (9.6 mg, 69%), dissolved in 125 µL H_2O, and analysed by high resolution mass spectrometry.

HRMS (ESI+) calculated for $C_{20}H_{26}N_7O_9$ (M+H)+ 508.1787, found 508.1792.

7.2.7.2 Nucleoside crosslinking PhSeT and N^6-methyladenosine[17,259,272]



5-phenylselenylmethyl-2'-deoxyuridine (6 mg, 15 µmol) and N^6-methyladenosine (4.2 mg, 15 µmol) were dissolved in H_2O (125 µL). Hydrogen peroxide (25 µL, 0.075 mmol) was then added and the reaction was stirred at room temperature for 3 hrs. The solution was concentrated *in vacuo* and aqueous piperidine (1M, 125 µL, 0.125 mmol) was added. The solution was then incubated at 90 °C for 20 mins. The mixture was concentrated *in vacuo* (9.6 mg, 69%), dissolved in 125 µL H_2O, and analysed by high resolution mass spectrometry.

HRMS (ESI+) calculated for $C_{21}H_{28}N_7O_9$ (M+H)+ 522.1943, found 522.1962.

7.2.7.3 Interstrand crosslinking reaction

The ^{32P} radiolabelled PhSeT modified DNA oligonucleotide (1 µL, 0.3 pmol) was mixed with the unlabelled complementary RNA oligonucleotide (1 µL, 1.5 pmol), sodium phosphate (1 µL, 100 mM, pH 7.4), NaCl (1 µL, 1M), and H_2O (11 µL). The oligonucleotides were annealed by incubating the solution at 65 °C for 10 mins, followed by incubation at 21 °C for 10 mins.

Sodium phosphate (1 µL, 100 mM, pH 7.2), NaCl (1 µL, 1 M), H_2O_2 (1 µL, 300 mM), and H_2O (12 µL) were then added. The solution was incubated at 21 °C for 1 hr, and then mixed with 95 % formamide loading buffer (30 µL) (deionised formamide (95 %), EDTA (5 mM, pH 8.0), SDS (0.025 %)). The

reaction was analysed using a 15 % Mini-PROTEAN® TBE-Urea Gel pre-cast gel.

7.2.8 PhSe-dTTP reverse transcription reactions[113]

7.2.8.1 PhSeT primer extension

Either A-RNA or m^6A-RNA (2 pmol) was mixed with ^{32}P radiolabelled RT primer (1 pmol) in H_2O (6 µL) (Table 23). The oligonucleotides were annealed by incubating the solution at 75°C for 5 mins followed by room temperature for 10 mins. Either dTTP (1 µL, 30 pmol) or PhSe-dTTP (1 µL, 120 pmol) was then added along with the reverse transcription buffer (1 µL, 10 x) (Table 24), dATP/dGTP/dCTP mix (1 µL, 30 pmol each), and the reverse transcriptase enzyme (1 µL) (Table 24). The solution was incubated at 42 °C for 1 hr. RNase H (0.5 µL, 2 units) was added and the solution was incubated at 37 °C for 1 hr. The enzymes were then inactivated by incubating at 90 °C for 5 mins. An equal volume of 95 % formamide loading buffer (10.5 µL) was then added and the solution was incubated at 90 °C for 5 mins, followed by incubation on ice for 5 mins. The reaction was then analysed using a 15 % Mini-PROTEAN® TBE-Urea Gel pre-cast gel.

Table 23: Oligonucleotides used in primer extension reactions. Nm = 2′-OMe-modified nucleotide, UPPERCASE = DNA, lowercase = RNA, m^6A = RNA

Oligonucleotides	**Sequence (5'-3')**
A-RNA	accggacuguuaccaacacccacacccc
m^6A-RNA	accgg-m^6A-cuguuaccaacacccacacccc
RT primer (DNA)	GGGGTGTGGGTGTTGGT

Table 24: Reverse transcriptase enzymes and reaction buffers used in PhSe-dTTP primer extension reactions.

RT enzyme	**RT buffer**
M-MuLV Reverse Transcriptase (200 units/µL)	M-MuLV Reaction Buffer
Optizyme™ Reverse Transcriptase (200 units/µL)	OPTIZYME™ M-MLV RT Buffer (1 x)
ImProm-II™ Reverse Transcriptase (1 unit/µL)	ImProm-II™ Reaction Buffer (1 x)
Bst DNA polymerase, Large Fragment (8 units/µL)	ThermoPol® Reaction Buffer (1 x)
Bst 3.0 DNA polymerase (8 units/µL)	Isothermal Amplification Buffer II (1 x)

7.2.8.2 PhSe-dTTP primer extension reaction optimisation

The primer extension reactions using *Bst* 3.0 DNA polymerase were repeated with varying reaction temperatures, buffer compositions, and enzyme concentrations in order to optimise the reaction conditions.

7.2.8.3 PhSe-dTTP primer extension test quantification

The PhSe-dTTP primer extension reactions were repeated using varying ratios of the A-RNA and m^6A-RNA oligonucleotides.

7.2.8.4 TLC analysis of PhSe-dTTP reverse transcription of total RNA

7.2.8.4.1 Reverse transcription

5' biotinylated poly(dT)$_{17}$ oligonucleotide containing 9 locked bases (100 pmol) was mixed with mouse total RNA (1 µg) in H_2O (6 µL). The solution was incubated at 90 °C for 1 min, followed by room temperature for 10 mins to anneal the oligonucleotide to the RNA. The solution was mixed with $MgSO_4$ (2 µL, 20 mM), Isothermal Amplification Buffer II (2 µL, 10 x), dATP/dGTP/dCTP mix (1 µL, 10 nmol each), PhSe-dTTP (1 µL, 12.5 nmol), and *Bst* 3.0 DNA polymerase (1 µL, 8 units) and incubated at 60 °C for 1 hr.

7.2.8.4.2 RNase T1 digestion

Hydrophilic Streptavidin Magnetic Beads (100 µL, 400 ng) were added. The solution was vortexed and incubated at room temperature for 10 mins to bind to the biotinylated oligonucleotide to the streptavidin magnetic beads. A magnetic rack was used to remove the supernatant which was discarded. The magnetic beads were washed in RNase T1 buffer (100 µL, 1 x) (Tris HCl (50 mM, pH 7.5), and EDTA (2 mM)) and resuspended in H_2O (17 µL). RNase T1 buffer (2 µL, 10 x) and RNase T1 (1 µL, 1 unit) were added, and the solution was incubated at 37 °C for 1 hr.

7.2.8.4.3 PNK radiolabelling

The supernatant was discarded using the magnetic rack and the magnetic beads were washed in T4 Polynucleotide Kinase Buffer (100 µL, 1 x) and resuspended in H_2O (15 µL). ATP [γ-^{32}P] (2 µL, 6.6 pmol, 20 µCi), T4 PNK buffer A (2 µL 10 x), and T4 Polynucleotide Kinase (1 µL, 10 units) were added and the solution was incubated at 37 °C for 1 hr. EDTA (1 µL, 500 mM) was added and the solution was incubated at 65 °C for 20 mins to inactivate the enzyme.

7.2.8.4.4 Digestion into 5’ nucleoside monophosphates

The magnetic beads were washed with RNase H reaction buffer (100 µL, 1 x), and resuspended in H_2O (17 µL). RNase H reaction buffer (2 µL, 10 x), and RNase H (1 µL, 5 units) were added, and the solution was incubated at 37 °C for 1 hr, followed by 65 °C for 20 mins to inactivate the enzyme. The RNA/DNA was purified by ethanol precipitation (See section 7.2.4.2). The

pellet was resuspended in H_2O (8 μL), Nuclease P1 buffer (1 μL, 10 x) and Nuclease P1 (1 μL, 100 units) and the solution was incubated at 37 °C for 2 hours to digest the DNA/RNA into 5'- nucleoside monophosphates. 1 μL of this solution was spotted onto a TLC plate and ran in two dimensions (See section 7.2.4.5).

References

1. Xiao, W. *et al.* Nuclear m6A reader YTHDC1 regulates mRNA splicing. *Mol. Cell* **61**, 507–519 (2016).

2. Ke, S. *et al.* A majority of m6A residues are in the last exons, allowing the potential for 3 ′ UTR regulation. *Genes Dev.* **29**, 2037–2053 (2015).

3. Roundtree, I. A. *et al.* YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs. *eLife* **6**, e31311 (2017).

4. Lesbirel, S. *et al.* The m6A-methylase complex recruits TREX and regulates mRNA export. *Sci. Rep.* **8**, 13827 (2018).

5. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117–120 (2014).

6. Wang, X. *et al.* N6-methyladenosine modulates messenger RNA translation efficiency. *Cell* **161**, 1388–1399 (2015).

7. Zhou, J. *et al.* Dynamic m6A mRNA methylation directs translational control of heat shock response. *Nature* **526**, 591–594 (2015).

8. Meyer, K. D. *et al.* 5’ UTR m6A promotes cap-independent translation. *Cell* **163**, 999–1010 (2015).

9. Du, H. *et al.* YTHDF2 destabilizes m6A-containing RNA through direct recruitment of the CCR4–NOT deadenylase complex. *Nat. Commun.* **7**, 12626 (2016).

10. Batista, P. J. The RNA Modification N6-methyladenosine and its implications in human disease. *Genomics. Proteomics Bioinformatics* **15**, 154–163 (2017).

11. Hodgson, J. RNA epigenetics spurs investor interest, but uncertainties linger. *Nat. Biotechnol.* **36**, 1123–1124 (2018).

12. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201–206 (2012).

13. Meyer, K., Saletore, Y., Zumbo, P., Elemento, O. & Mason, C. Comprehensive analysis of mRNA methylation reveals enrichment in 3’ UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).

14. Linder, B. *et al.* Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* **12**, 767–772 (2015).

15. Schwartz, S. *et al.* High-Resolution mapping reveals a conserved,

widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* **155**, 1409–1421 (2013).

16. Harcourt, E. M., Ehrenschwender, T., Batista, P. J., Chang, H. Y. & Kool, E. T. Identification of a selective polymerase enables detection of N6-methyladenosine in RNA. *J. Am. Chem. Soc.* **135**, 19079–19082 (2013).

17. Hong, I. S. & Greenberg, M. M. Mild generation of 5-(2'-deoxyuridinyl) methyl radical from a phenyl selenide precursor. *Org. Lett.* **6**, 511–513 (2004).

18. Crick, F. H. C. On protein synthesis. *Symp. Soc. Exp. Biol.* **12**, 139–163 (1958).

19. David, C., Pazdernik, N. & McGehee, M. *Molecular biology*. (Academic Press, 2018).

20. Elliott, D. & Ladomery, M. *Molecular biology of RNA*. (Oxford University Press, 2011).

21. Blackburn, G., Gait, M., Loakes, D. & Williams, D. *Nucleic acids in chemistry and biology*. (RSC Publishing, 2006).

22. Pan, T. Modifications and functional genomics of human transfer RNA. *Cell Res.* **28**, 395–404 (2018).

23. Shatkin, A. J. Capping of eukaryotic mRNAs. *Cell* **9**, 645–653 (1976).

24. Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3171–3175 (1977).

25. Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**, 291–336 (2003).

26. Edmonds, M., Vaughan, M. H. & Nakazato, H. Polyadenylic acid sequences in the heterogeneous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: possible evidence for a precursor relationship. *Proc. Natl. Acad. Sci. U.S.A.* **68**, 1336–1340 (1971).

27. Guhaniyogi, J. & Brewer, G. Regulation of mRNA stability in mammalian cells. *Gene* **265**, 11–23 (2001).

28. Gu, M. & Lima, C. D. Processing the message: structural insights into capping and decapping mRNA. *Curr. Opin. Struct. Biol.* **15**, 99–106 (2005).

29. Muthukrishnan, S., Both, G., Furuichi, Y. & Shatkin, A. 5'-Terminal 7-methylguanosine in eukaryotic mRNA is required for translation. *Nature* **225**, 33–37 (1975).

30. Piovesan, A. *et al.* Human protein-coding genes and gene feature statistics in 2019. *BMC Res. Notes* **12**, 315 (2019).

31. Dupont, C., Armant, R. & Brenner, C. Epigenetics: definition, mechanisms and clinical perspective. *Semin. Reprod. Med.* **27**, 351–357 (2009).

32. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).

33. Moore, L. D., Le, T. & Fan, G. DNA methylation and its basic function. *Neuropsychopharmacology* **38**, 23–38 (2013).

34. Saletore, Y. *et al.* The birth of the epitranscriptome: deciphering the function of RNA modifications. *Genome Biol.* 175 (2012).

35. Wang, X. & He, C. Dynamic RNA modifications in posttranscriptional regulation. *Mol. Cell* **56**, 5–12 (2014).

36. Wyatt, G. Recognition and estimation of 5-methylcytosine in nucleic acids. *Biochemistry* **48**, 581–584 (1951).

37. Meehan, R., Lewis, J., Mckay, S., Kleiner, E. & Bird, A. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* **58**, 499–507 (1989).

38. Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* **18**, 6538–6547 (1998).

39. Deaton, M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).

40. Chodavarapu, R. *et al.* Relationship between nucleosome positioning and DNA methylation. *Nature* **466**, 388–392 (2010).

41. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129– 1151 (2010).

42. Nesterova, T. B. *et al.* Dicer regulates Xist promoter methylation in ES cells indirectly through transcriptional control of Dnmt3a. *Epigenetics Chromatin* **1**, 2 (2008).

43. Li, E., Beard, C. & Rudolf, B. Role for DNA methylation in genomic imprinting. *Nature* **366**, 1991–1994 (1993).

44. Murray, N. Immigration control of DNA in bacteria: self versus non-self. *Microbiology* **148**, 3–20 (2002).

45. Fu, Y., Luo, G. & Chen, K. N6-methyldeoxyadenosine marks active transcription start sites in chlamydomonas. *Cell* **161**, 879–892 (2015).

46. Greer, E., Blanco, M., Mario, A. & Gu, L. DNA methylation on N6-adenine in C. elegans. *Cell* **161**, 868– 878 (2015).

47. Ehrlich, M., Wilson, G., Kuo, K. & Gehrke, C. N4-methylcytosine as a minor base in bacterial DNA. *J. Bacteriol.* **169**, 939–943 (1987).

48. He, Y., Li, B., Li, Z., Liu, P. & Wang, Y. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).

49. Machnicka, M. A. *et al.* MODOMICS: A database of RNA modification pathways - 2013 update. *Nucleic Acids Res.* **41**, 262–267 (2013).

50. Chan, P. & Lowe, T. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* **44**, 184–189 (2016).

51. Desrosiers, R., Friderici, K. & Rottman, F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3971–3975 (1974).

52. Lanberg, S. & Moss, B. Post-transcriptional modifications of mRNA. Purification and characterization of cap I and cap II RNA (nucleoside-2'-)-methyltransferases from HeLa cells. *J. Biol. Chem.* **256**, 54–60 (1981).

53. Stepinski, J., Darzynkiewicz, E. & Pelletier, J. Characterization of hMTr1, a human Cap1 2'-O-ribose methyltransferase. *J. Biol. Chem.* **285**, 33037–33044 (2010).

54. Werner, M. *et al.* 2'-O-ribose methylation of cap2 in human: function and evolution in a horizontally mobile family. *Nucleic Acids Res.* **39**, 4756–4768 (2011).

55. Wei, C. & Gershowitz, A. Methylated nucleotides block 5 ' terminus of HeLa cell messenger RNA. *Cell* **4**, 379–386 (1975).

56. Shatkin, A. *et al.* 5'-Terminal caps, cap-binding proteins and eukaryotic mRNA function. *Biochem. Soc. Symp.* **47**, 129–143 (1982).

57. Akichika, S., Hirano, S., Shichino, Y. & Suzuki, T. Cap-specific terminal N6-methylation of RNA by an RNA polymerase II – associated methyltransferase. *Science* **363**, 10.1126/science.aav0080 (2019).

58. Sendinc, E. *et al.* PCIF1 catalyzes m6Am mRNA methylation to regulate gene expression. *Mol. Cell* **75**, 620–630 (2019).

59. Boulias, K. *et al.* Identification of the m6Am methyltransferase PCIF1

reveals the location and functions of m6Am in the transcriptome. *Mol. Cell* **75**, 631–643 (2019).

60. Daffis, S., Szretter, K. & Schriewer, J. 2′-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature* **468**, 452–456 (2010).

61. Züst, R. *et al.* Ribose 2′-O -methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. *Nat. Publ. Gr.* **12**, 137–143 (2011).

62. Mouaikel, J., Bertrand, E., Tazi, J. & Fischer, U. Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Mol. Cell* **9**, 891–901 (2002).

63. Li, X. *et al.* Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* **11**, 592–597 (2015).

64. Carlile, T. M. *et al.* Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**, 143–146 (2014).

65. Huang, G., Chen, W., Liu, J. & Gu, N. Genome-wide identification of mRNA 5-methylcytosine in mammals. *Nat. Struct. Mol. Biol.* **26**, 380–388 (2019).

66. Huber, S. M. *et al.* Formation and abundance of 5-hydroxymethylcytosine in RNA. *ChemBioChem* **16**, 752–755 (2015).

67. Arango, D. *et al.* Acetylation of cytidine in mRNA promotes translation efficiency. *Cell* **175**, 1872–1886 (2018).

68. Paul, M. S. & Bass, B. L. Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *EMBO J.* **17**, 1120–1127 (1998).

69. Dominissini, D. *et al.* The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530**, 441–446 (2016).

70. Ayadi, L., Galvanin, A., Pichot, F., Marchand, V. & Motorin, Y. RNA ribose methylation (2′-O-methylation): Occurrence, biosynthesis and biological functions. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1862**, 253–269 (2019).

71. Grozhik, A. V & Jaffrey, S. R. Distinguishing RNA modifications from noise in epitranscriptome maps. *Nat. Publ. Gr.* **14**, 215–225 (2018).

72. Bokar, J., Shambaugh, M., Polayes, D., Matera, A. & Rottman, F. Purification and cDNA cloning of the AdoMet-binding subunit of the

human mRNA (N6-adenosine)-methyltransferase. *RNA* **3**, 1233–1247 (1996).

73. Liu, J. *et al.* A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* **10**, 93–95 (2014).

74. Zhong, S. *et al.* MTA Is an arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* **20**, 1278–1288 (2008).

75. Xu, C. *et al.* Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat. Chem. Biol.* **10**, 927–929 (2014).

76. Weidensdorfer, D. *et al.* Control of c-myc mRNA stability by IGF2BP1-associated cytoplasmic RNPs. *RNA* **15**, 104–115 (2009).

77. Zheng, G. *et al.* ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **49**, 18–29 (2013).

78. Batista, P. J. *et al.* m6A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **15**, 707–719 (2014).

79. Fustin, J. M. *et al.* XRNA-methylation-dependent RNA processing controls the speed of the circadian clock. *Cell* **155**, 793–806 (2013).

80. Hastings, M. H. m6A mRNA methylation: A new circadian pacesetter. *Cell* **155**, 740–741 (2013).

81. Li, H. *et al.* m6A mRNA methylation controls T cell homeostasis by targeting the IL-7/STAT5/SOCS pathways. *Nature* **548**, 338–342 (2017).

82. Tong, J., Cao, G., Zhang, T. & Sefik, E. m6A mRNA methylation sustains Treg suppressive functions. *Nat. Cell Res.* **28**, 253–256 (2018).

83. Winkler, R. *et al.* m6A modification controls the innate immune response to infection by targeting type I interferons. *Nat. Immunol.* **20**, 173–182 (2019).

84. Vu, L. P. *et al.* The N6-methyladenosine (m6A) forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. *Nat. Med.* **23**, 1369–1376 (2017).

85. Barbieri, I. *et al.* Promoter bound METTL3 maintains myeloid leukemia by m6A dependent translational control. *Nature* **552**, 126–131 (2017).

86. Zhang, S. *et al.* m6A demethylase ALKBH5 maintains tumorigenicity of glioblastoma stem-like cells by sustaining FOXM1 expression and cell proliferation program. *Cancer Cell* **31**, 591–606 (2017).

87. Cui, Q. *et al.* m6A RNA methylation regulates the self-renewal and

tumorigenesis of glioblastoma stem cells. *Cell Rep.* **18**, 2622–2634 (2017).

88. Zhang, C. *et al.* Hypoxia induces the breast cancer stem cell phenotype by HIF-dependent and ALKBH5-mediated m6A-demethylation of NANOG mRNA. *Proc. Natl. Acad. Sci. U.S.A* **113**, 2047–2056 (2016).

89. Carroll, S. M., Narayan, P. & Rottman, F. M. N6-methyladenosine residues in an intron-specific region of prolactin pre-mRNA. *Mol. Cell. Biol.* **10**, 4456–4465 (1990).

90. Liu, N. *et al.* Probing N6 -methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* **19**, 1848–1856 (2013).

91. Garcia-campos, M. A. *et al.* Deciphering the “m6A code” via antibody-independent quantitative profiling. *Cell* **178**, 731–747 (2019).

92. Frye, M., Jaffrey, S. R., Pan, T., Rechavi, G. & Suzuki, T. RNA modifications: what have we learned and where are we headed? *Nat. Rev. Genet.* **17**, 365–372 (2016).

93. Perry, R. P. & Kelley, D. E. Existence of methylated messenger RNA in mouse L cells. *Cell* **1**, 37–42 (1974).

94. Lavi, S. & Shatkin, A. J. Methylated simian virus 40-specific RNA from nuclei and cytoplasm of infected BSC-1 cells. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2012–2016 (1975).

95. Wei, C. & Moss, B. Nucleotide sequences at the N6-methyladenosine sites of HeLa cell messenger ribonucleic acid. *Biochemistry* **16**, 1672–1676 (1977).

96. Nichols, J. L. Cap structures in maize poly(A) containing RNA. *Biochem. Biophys. Acta* **563**, 490–495 (1979).

97. Kennedy, T. & Lane, B. Wheat embryo ribonucleates. XIII. Methyl-substituted nucleoside constituents and 5’-terminal dinucleotide sequences in bulk poly(AR)-rich RNA from imbibing wheat embryos. *Can. J. Biochem.* **57**, 927–931 (1979).

98. Nichols, J. & Welder, L. Nucleotides adjacent to N6-methyladenosine in maize poly (A)-containing RNA. *Plant Sci. Lett.* **21**, 75–81 (1981).

99. Csepanys, T., Lint, A. & Baldick, C. J. Sequence specificity of mRNA N6-adenosine methyltransferase. *Biochemistry* **265**, 20117–20122 (1990).

100. Narayan, P. & Rottman, F. An in vitro system for accurate methylation of internal adenosine residues in messenger RNA. *Science* **242**, 1159–

1162 (1998).

101. Narayan, P., Ludwiczak, R. L., Goodwin, E. C. & Rottman, F. M. Context effects on N6-adenosine methylation sites in prolactin mRNA. *Nucleic Acids Res.* **22**, 419–426 (1994).

102. Meyer, K. D. & Jaffrey, S. R. The dynamic epitranscriptome: N6-methyladenosine and gene expression control. *Nat. Rev. Mol. Cell Biol.* **15**, 313–326 (2014).

103. Bodi, Z. *et al.* Adenosine methylation in Arabidopsis mRNA is associated with the 3' end and reduced levels cause developmental defects. *Front. Plant Sci.* **3**, 48 (2012).

104. Grippo, P., Iaccarino, M., Rossi, M. & Scarano, E. Thin-layer chromatography of nucleotides, nucleosides and nucleic acid bases. *Biochim. Biophys. Acta* **95**, 1–7 (1965).

105. Beemon, K. & Keith, J. Localization of IV-methyladenosine in the Rous sarcoma virus genome. *J. Mol. Biol.* **113**, 165–179 (1977).

106. Hrdlickova, R., Toloue, M. & Tian, B. RNA-Seq methods for transcriptome analysis. *Wiley Interdiscip. Rev. RNA* **8**, (2017).

107. Suzuki, T., Ueda, H., Okada, S. & Sakurai, M. Transcriptome-wide identification of adenosine-to-inosine editing using the ICE-seq method. *Nat. Protoc.* **10**, 715–732 (2015).

108. Fraga, M. & Esteller, M. DNA methylation: A profile of methods and applications. *Biotechniques* **33**, 632–649 (2002).

109. Bazak, L. *et al.* A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res.* **24**, 365–376 (2014).

110. Hussain, S., Aleksic, J., Blanco, S., Dietmann, S. & Frye, M. Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome Biol.* **14**, 215 (2013).

111. Schwartz, S. *et al.* Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148–162 (2014).

112. Schwartz, S. Cracking the epitranscriptome. *RNA* **22**, 169–174 (2016).

113. Hong, T., Yuan, Y. & Chen, Z. Precise antibody-independent m6A identification via 4SedTTP-involved and FTO-assisted strategy at single-nucleotide resolution. *J. Am. Chem. Soc.* **140**, 5886–5889 (2018).

114. Zhang, Y., Zhang, J. & Hara, H. Insights into the mRNA cleavage

mechanism by MazF, an mRNA interferase. *J. Biol. Chem.* **280**, 3143–3150 (2005).

115. Imanishi, M., Tsuji, S., Suda, A. & Futaki, S. Detection of N6-methyladenosine based on the methyl-sensitivity of MazF RNA endonuclease. *Chem. Commun.* **50**, 12930–12933 (2017).

116. Zhang, Z. *et al.* Single-base mapping of m6A by an antibody-independent method. *Sci. Adv.* **5**, 10.1126/sciadv.aax0250 (2019).

117. Kellner, S. *et al.* Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. *Nucleic Acids Res.* **42**, e142 (2014).

118. Shu, X., Dai, Q., Wu, T. & Bothwell, I. N6-allyladenosine: a new small molecule for RNA labeling identified by mutation assay. *J. Am. Chem. Soc.* **139**, 17213–17216 (2017).

119. Willnow, S., Martin, M., Lüscher, B. & Weinhold, E. A selenium-based click AdoMet analogue for versatile substrate labeling with wild-type protein methyltransferases. *ChemBioChem* **13**, 1167–1173 (2012).

120. Hartstock, A. K., Nilges, B., Ovcharenko, A., Puellen, N. & Leidel, S. Enzymatic or in vivo installation of propargyl groups in combination with click chemistry for the enrichment and detection of methyltransferase target sites in RNA. *Angew. Chem.* **57**, 6342–6346 (2018).

121. Dijk, E. L. Van, Jaszczyszyn, Y., Naquin, D. & Thermes, C. The third revolution in sequencing technology. *Trends Genet.* **34**, 666–681 (2018).

122. Levy, S. E. & Myers, R. M. Advancements in next-generation sequencing. *Annu. Rev. Genomics Hum. Genet.* **17**, 95–115 (2016).

123. Workman, R. E. *et al.* Nanopore native RNA sequencing of a human poly (A) transcriptome. *bioRxiv* 10.1101/459529 (2018).

124. Vilfan, I. D. *et al.* Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. *J. Nanobiotechnology* **11**, 10.1186/1477-3155-11–8 (2013).

125. Tuck, M. The formation of internal 6-methyladenine residues in eucaryotic messenger RNA. *Int. J. Biochem.* **24**, 379–386 (1992).

126. Hongay, C. F. & Orr-weaver, T. L. Drosophila inducer of meiosis 4 (IME4) is required for notch signaling during oogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14855–14860 (2011).

127. Clancy, M. J., Shambaugh, M. E., Timpte, C. S. & Bokar, J. A. Induction of sporulation in Saccharomyces cerevisiae leads to the formation of N6-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. *Nucleic Acids Res.* **30**, 4509–4518 (2002).

128. Geula, S., Moshitch-Moshkovitz, S., Dominissini, D. & Mansour, A. Stem cells. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* **347**, 1002–1006 (2015).

129. Ping, X. *et al.* Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Nat. Publ. Gr.* **24**, 177–189 (2014).

130. Haussmann, I. U. *et al.* m6A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. *Nature* **540**, 301–304 (2016).

131. Lence, T. *et al.* m6A modulates neuronal functions and sex determination in Drosophila. *Nat. Publ. Gr.* **540**, 242–247 (2016).

132. Bujnicki, J. M., Feder, M., Radlinska, M. & Blumenthal, R. M. Structure prediction and phylogenetic analysis of a functionally diverse family of proteins homologous to the MT-A70 subunit of the human mRNA: m6A Methyltransferase. *J. Mol. Evol.* **55**, 431–444 (2002).

133. Bodi, Z., Button, J. D., Grierson, D. & Fray, R. G. Yeast targets for mRNA methylation. *Nucleic Acids Res.* **38**, 5327–5335 (2010).

134. Sledz, P. & Jinek, M. Structural insights into the molecular mechanism of the m(6)A writer complex. *eLife* **14**, e18434 (2016).

135. Wang, X. *et al.* Structural basis of N6-adenosine methylation by the METTL3–METTL14 complex. *Nature* **534**, 575–578 (2016).

136. Wang, P., Doxtader, K. & Nam, Y. Structural basis for cooperative function of Mettl3 and Mettl14 methyltransferases. *Mol. Cell* **63**, 306–317 (2016).

137. Huang, H. *et al.* Histone H3 trimethylation at lysine 36 guides m6A RNA modification co-transcriptionally. *Nature* **567**, 414–419 (2019).

138. Ortega, A. *et al.* Biochemical function of female-lethal (2)D/Wilms' tumor suppressor-1-associated proteins in alternative pre-mRNA splicing. *J. Biol. Chem.* **278**, 3040–3047 (2003).

139. Vespa, L. *et al.* The immunophilin-interacting protein AtFIP37 from Arabidopsis is essential for plant development and is involved in trichome endoreduplication. *Plant Physiol.* **134**, 1283–1292 (2004).

140. Schwartz, S. *et al.* Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5’ sites. *Cell Rep.* **8**, 284–296 (2014).

141. Ruzicka, K. *et al.* Identification of factors required for m6A mRNA methylation in Arabidopsis reveals a role for the conserved E3 ubiquitin ligase HAKAI. *New Phytol.* **215**, 157–172 (2017).

142. Wen, J. *et al.* Zc3h13 regulates nuclear RNA m6A methylation and mouse embryonic stem cell self-renewal. *Mol. Cell* **69**, 1028–1038 (2018).

143. Knuckles, P. *et al.* Zc3h13/Flacc is required for adenosine methylation by bridging the mRNA-binding factor RbM15/spenito to the m6A machinery component Wtap/Fl(2)d. *Genes Dev.* **32**, 415–429 (2018).

144. Yue, Y. *et al.* VIRMA mediates preferential m6A mRNA methylation in 3 ’ UTR and near stop codon and associates with alternative polyadenylation. *Cell Discov.* **4**, 10.1038/s41421-018-0019–0 (2018).

145. Patil, D. P. *et al.* m6A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537**, 369–373 (2016).

146. Raffel, G. D. *et al.* Ott1 (Rbm15) is essential for placental vascular branching morphogenesis and embryonic development of the heart and spleen. *Proc. Natl. Acad. Sci. U.S.A* **29**, 333–341 (2009).

147. Raffel, G. D. *et al.* Ott1(Rbm15) has pleiotropic roles in hematopoietic development. *Proc. Natl. Acad. Sci. U.S.A* **104**, 6001–6006 (2007).

148. Jia, G. *et al.* N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7**, 885–887 (2012).

149. Fu, Y. *et al.* FTO-mediated formation of N6-hydroxymethyladenosine and N6 -formyladenosine in mammalian RNA. *Nat. Commun.* **4**, 2–9 (2013).

150. Mauer, J. & Jaffrey, S. R. FTO, m6Am, and the hypothesis of reversible epitranscriptomic mRNA modifications. *FEBS Lett.* **592**, 2012–2022 (2018).

151. Bartosovic, M. *et al.* N6-methyladenosine demethylase FTO targets pre-mRNAs and regulates alternative splicing and 3 -end processing. *Nucleic Acids Res.* **45**, 11356–11370 (2018).

152. Mauer, J. *et al.* Reversible methylation of m6Am in the 5 ’ cap controls mRNA stability. *Nat. Publ. Gr.* **541**, 371–375 (2017).

153. Zhao, X. *et al.* FTO-dependent demethylation of N6-methyladenosine

regulates mRNA splicing and is required for adipogenesis. *Cell Res.* **24**, 1403–1419 (2014).

154. Zou, S. *et al.* N6-Methyladenosine: a conformational marker that regulates the substrate specificity of human demethylases FTO and ALKBH5. *Sci. Rep.* **6**, 25677 (2016).

155. Hess, M. E. *et al.* The fat mass and obesity associated gene (Fto) regulates activity of the dopaminergic midbrain circuitry. *Nat. Neurosci.* **16**, 1042–1048 (2013).

156. Liu, X., Zhou, J., Mao, Y., Ji, Q. & Qian, S. Programmable RNA N6-methyladenosine editing by CRISPR-Cas9 conjugates. *Nat. Chem. Biol.* **15**, 865–871 (2019).

157. Huang, H., Weng, H., Qin, X. & Zhao, B. S. Recognition of RNA N6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat. Cell Biol.* **20**, 285–295 (2018).

158. Alarcón, C. R. *et al.* HNRNPA2B1 is a mediator of m(6)A-dependent nuclear RNA processing events. *Cell* **162**, 1299–308 (2015).

159. Patil, D. P., Pickering, B. F. & Jaffrey, S. R. Reading m6A in the transcriptome: m6A-binding proteins. *Trends Cell Biol.* **28**, 113–127 (2017).

160. Lee, M., Bonneau, A. & Giraldez, A. Zygotic genome activation during the maternal-to-zygotic transition. *Annu. Rev. Cell Dev. Biol.* **30**, 581–613 (2014).

161. Zhao, B., Wang, X. & Beadell, A. m6A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature* **542**, 475–478 (2017).

162. Shi, H. *et al.* YTHDF3 facilitates translation and decay of N6-methyladenosine modified RNA. *Nat. Publ. Gr.* **27**, 315–328 (2017).

163. Li, A., Chen, Y. & Ping, X. Cytoplasmic m6A reader YTHDF3 promotes mRNA translation. *Cell Res.* **27**, 444–447 (2017).

164. Salditt-Georgieff, M., Jelinek, W. & Darnell, J. E. Methyl labeling of HeLa cell hnRNA: a comparison with mRNA. *Cell* **7**, 227–237 (1976).

165. Ajiro, M., Jia, R., Yang, Y., Zhu, J. & Zheng, Z. A genome landscape of SRSF3-regulated splicing events and gene expression in human osteosarcoma U2OS cells. *Nucleic Acids Res.* **44**, 1854–1870 (2018).

166. Hinnebusch, A. G. & Lorsch, J. R. The mechanism of eukaryotic translation initiation: New insights and challenges. *Cold Spring Harb.*

Perspect. Biol. **4**, 10.1101/cshperspect.a011544 (2012).

167. Stoneley, M. & Willis, A. E. Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. *Oncogene* **23**, 3200–3207 (2004).

168. López-lastra, M., Rivas, A. & Barría, M. I. Protein synthesis in eukaryotes: The growing biological relevance of cap- independent translation initiation. *Biol. Res.* **38**, 121–146 (2005).

169. Schumann, U., Shafik, A. & Preiss, T. METTL3 gains R/W access to the epitranscriptome. *Mol. Cell* **62**, 323–324 (2016).

170. Roost, C., Lynch, S. R., Batista, P. J., Qu, K. & Chang, H. Y. Structure and thermodynamics of N6-methyladenosine in RNA: A spring-loaded base modification. *J. Am. Chem. Soc.* **137**, 2107–2115 (2015).

171. Kierzek, E. & Kierzek, R. The thermodynamic stability of RNA duplexes and hairpins containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines. *Nucleic Acids Res.* **31**, 4472–4480 (2003).

172. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**, 701–705 (2014).

173. Liu, N. *et al.* N6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518**, 560–564 (2015).

174. Liu, N. & Pan, T. N6-methyladenosine–encoded epitranscriptomics. *Nat. Struct. Mol. Biol.* **23**, 98–102 (2016).

175. König, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* **17**, 909–915 (2010).

176. Gorlach, M., Burd, C. G. & Dreyfuss, G. The determinants of RNA-binding specificity of the heterogeneous nuclear ribonucleoprotein C proteins. *J. Biol. Chem.* **16**, 23074–23078 (1994).

177. Cienikova, Z., Damberger, F. F., Hall, J., Allain, F. & Maris, C. Structural and mechanistic insights into poly(uridine) tract recognition by the hnRNP C RNA recognition motif. *J. Am. Chem. Soc.* **136**, 14536–14544 (2014).

178. Edupuganti, R. R. *et al.* N6-methyladenosine (m6A) recruits and repels proteins to regulate mRNA homeostasis. *Nat. Struct. Mol. Biol.* **24**, 870–878 (2017).

179. Zakrzewski, W., Dobrzyński, M., Szymonowicz, M. & Rybak, Z. Stem

cells: past, present, and future. *Stem Cell Res. Ther.* **10**, 10.1186/s13287-019-1165–5 (2019).

180. Wang, Y. *et al.* N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell Biol.* **16**, 191–198 (2014).

181. Aguilo, F. *et al.* Coordination of m6A mRNA methylation and gene transcription by ZFP217 regulates pluripotency and reprogramming. *Cell Stem Cell* **17**, 689–704 (2015).

182. Bradbury, R. & Angibaud, P. *Cancer*. (Springer, 2007).

183. Chen, M. *et al.* RNA N6-methyladenosine methyltransferase-like 3 promotes liver cancer progression through YTHDF2-dependent posttranscriptional silencing of SOCS2. *Hepatology* **67**, 2254–2270 (2018).

184. Lu, Y. *et al.* Methylated DNA/RNA in body fluids as biomarkers for lung cancer. *Biol. Proced. Online* **19**, 10.1186/s12575-017-0051–8 (2017).

185. Taketo, K., Konno, M., Asai, A., Koseki, J. U. N. & Toratani, M. The epitranscriptome m6A writer METTL3 promotes chemo- and radioresistance in pancreatic cancer cells. *Int. J. Oncol.* **52**, 621–629 (2018).

186. You, Y. *et al.* Genomic characterization of a Helicobacter pylori isolate from a patient with gastric cancer in China. *Gut Pathog.* **6**, 10.1186/1757-4749-6–5 (2014).

187. Lai, W. *et al.* Baicalin hydrate inhibits cancer progression in nasopharyngeal carcinoma by affecting genome instability and splicing. *Oncotarget* **9**, 901–914 (2018).

188. Nishizawa, Y. *et al.* Oncogene c-Myc promotes epitranscriptome m6A reader YTHDF1 expression in colorectal cancer. *Oncotarget* **9**, 7476–7486 (2018).

189. Li, J. *et al.* Downregulation of N6-methyladenosine binding YTHDF2 protein mediated by miR-493-3p suppresses prostate cancer by elevating N6-methyladenosine levels. *Oncotarget* **9**, 3752–3764 (2018).

190. Li, X. *et al.* The m6A methyltransferase METTL3 : acting as a tumor suppressor in renal cell carcinoma. *Oncotarget* **8**, 96103–96116 (2017).

191. Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X-chromosome. *Science* **341**, 10.1126/science.1237973 (2014).

192. Panda, S., Hogenesch, J. B. & Kay, S. A. Circadian rhythms from flies to human. *Nature* **417**, 329–335 (2002).

193. Lichinchi, G., Gao, S., Saletore, Y., Gonzalez, G. & Bansal, V. Dynamics of the human and viral m6A RNA methylomes during HIV-1 infection of T cells. *Nat. Microbiol.* **1**, doi:10.1038/nmicrobiol.2016.11 (2016).

194. Kennedy, E. M., Courtney, D. G., Tsai, K. & Cullen, B. R. Viral epitranscriptomics. *J. Virol.* **91**, 10.1128/JVI.02263-16 (2017).

195. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).

196. Alarcón, C. R., Lee, H., Goodarzi, H., Halberg, N. & Tavazoie, S. F. N6-methyladenosine marks primary microRNAs for processing. *Nature* **519**, 482–485 (2015).

197. Kruse, S. *et al.* A novel synthesis and detection method for cap-associated adenosine modifications in mouse mRNA. *Sci. Rep.* **1**, 10.1038/srep00126 (2011).

198. Dai, Q. *et al.* Identification of recognition residues for ligation-based detection and quantitation of pseudouridine and N6-methyladenosine. *Nucleic Acids Res.* **35**, 6322–6329 (2007).

199. Kierzek, E. & Kierzek, R. The synthesis of oligoribonucleotides containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines via post-synthetic modification of precursor oligomers. *Nucleic Acids Res.* **31**, 10.1093/nar/gkg632 (2003).

200. Nielsen, H. *RNA methods and protocols*. (Humana Press, 2011).

201. Liu, H. *et al.* Accurate detection of m6A RNA modifications in native RNA sequences. *bioRxiv* 10.1101/525741 (2019).

202. Helm, M., Brulé, H., Giegé, R. & Florentz, C. More mistakes by T7 RNA polymerase at the 5’ ends of in vitro-transcribed RNAs. *RNA* **5**, 618–621 (1999).

203. Engel, J. Mechanism of the Dimroth rearrangement in adenine. *Biochem. Biophys. Res. Commun.* **64**, 581–586 (1975).

204. Fujii, T., Saito, T., Itaya, T. & Kizu, K. Purines. XXIX. Syntheses of 9-alkyl-2-deuterio-N6-methoxyadenines and 2-deuterio-N6,9-dimethyladenine: Tautomerism in 9-substituted N6-alkoxyadenines. *Pharm. Bull.* **35**, 4482–4493 (1987).

205. Itaya, T., Tanaka, F. & Fujii, T. Purines—VI: Rate study and mechanism

of the dimroth rearrangement of 1-alkoxy-9-alkyladenines and 1-alkyl-9-methyladenines. *Tetrahedron* **28**, 535–547 (1972).

206. Jones, J. W. & Robins, R. K. Methylation studies of certain naturally occurring purine nucleosides. *J. Am. Chem. Soc.* **85**, 193–201 (1962).

207. Aritomo, K., Wada, T. & Sekine, M. Alkylation of 6-N-acylated adenosine derivatives by the use of phase transfer catalysis. *J. Chem. Soc. Perkin Trans. 1* **1**, 1837–1844 (1995).

208. Agris, P., Vendeix, F. & Graham, W. tRNA's wobble decoding of the genome: 40 years of modification. *J. Mol. Biol.* **366**, 1–13 (2007).

209. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A. & Steinberg, S. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **26**, 148–153 (1998).

210. Warner, G., Rusconi, C., White, I. & Faust, J. Identification and sequencing of two isopentenyladenosine-modified transfer RNAs from Chinese hamster ovary cells. *Nucleic Acids Res* **26**, 5533–5535 (1998).

211. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).

212. Oliphant, A. R., Brandl, C. J. & Struhl, K. Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein. *Mol. Cell. Biol.* **9**, 2944–2949 (1989).

213. Ellington, A. & Szostak, J. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).

214. Theler, D., Dominguez, C., Blatter, M., Boudet, J. & Allain, H. Solution structure of the YTH domain in complex with N6-methyladenosine RNA: a reader of methylated RNA. *Nucleic Acids Res.* **42**, 13911–13919 (2014).

215. Zhang, Z. *et al.* The YTH domain is a novel RNA binding domain. *J. Biol. Chem.* **285**, 14701–14710 (2010).

216. Wegner, A. Head to tail polymerization of actin. *J. Mol. Biol.* **108**, 139–150 (1976).

217. Condeelis, J. & Singer, R. H. How and why does β -actin mRNA target? *Biol. cell* **97**, 97–110 (2005).

218. Lawrence, J. & Singer, R. Intracellular localization of messenger RNAs for cytoskeletal proteins. *Cell* **45**, 407–415 (1986).

219. Yisraeli, J. K. VICKZ proteins: a multi-talented family of regulatory RNA-binding proteins. *Biol. Cell* **97**, 87–96 (2005).

220. Nielsen, J. *et al.* A family of insulin-like growth factor II mRNA-binding proteins represses translation in late development. *Mol. Cell. Biol.* **19**, 1262–1270 (1999).

221. Doyle, G. A. R. *et al.* The c-myc coding region determinant-binding protein: a member of a family of KH domain RNA-binding proteins. *Nucleic Acids Res.* **26**, 5036–5044 (1998).

222. Ross, A. F., Oleynikov, Y., Kislauskis, E. H., Taneja, K. L. & Singer, R. H. Characterization of a B-actin mRNA zipcode-binding protein. *Mol. Cell. Biol.* **17**, 2158–2165 (1997).

223. Havin, L. *et al.* RNA-binding protein conserved in both microtubule- and RNA localization. *Genes Dev.* **12**, 1593–1598 (1998).

224. Deshler, J. O., Highett, M. I. & Schnapp, B. J. Localization of Xenopus Vg1 mRNA by Vera protein and the endoplasmic reticulum. *Science* **276**, 1128–1131 (1997).

225. Jønson, L. *et al.* Molecular Composition of IMP1 Ribonucleoprotein Granules. *Mol. Cell. Proteomics* **6**, 798–811 (2007).

226. Patel, V. L. *et al.* Spatial arrangement of an RNA zipcode identifies mRNAs under post-transcriptional control. *Genes Dev.* **26**, 43–53 (2012).

227. Conway, A. E. *et al.* Enhanced CLIP uncovers IMP protein-RNA targets in human pluripotent stem cells important for cell adhesion and survival. *Cell Rep.* **15**, 666–679 (2016).

228. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).

229. Hansen, H. T. *et al.* Drosophila Imp iCLIP identifies an RNA assemblage coordinating F-actin formation. *Genome Biol.* **16**, doi: 10.1186/s13059-015-0687-0 (2015).

230. Leung, K., Horck, F. P. G. Van, Lin, A. C., Allison, R. & Holt, C. E. Asymmetrical β -actin mRNA translation in growth cones mediates attractive turning to netrin-1. *Nat. Neurosci.* **9**, 1247–1256 (2006).

231. Huttelmaier, S. *et al.* Spatial regulation of b-actin translation by Src-dependent phosphorylation of ZBP1. *Nature* **438**, 512–515 (2005).

232. Nicastro, G. *et al.* Mechanism of B-actin mRNA recognition by ZBP1.

Cell Rep. **18**, 1187–1199 (2017).

233. Chao, J. A. *et al.* ZBP1 recognition of β-actin zipcode induces RNA looping. *Genes Dev.* **24**, 148–158 (2010).

234. Song, T. *et al.* Specific interaction of KIF11 with ZBP1 regulates the transport of B-actin mRNA and cell motility. *J. Cell Sci.* **128**, 1001–1010 (2015).

235. Hall, B. *et al.* Design, synthesis, and amplification of DNA pools for in vitro selection. in *Current protocols in molecular biology* 1–27 (2009). doi:10.1002/0471142727.mb2402s88

236. Arguello, A., Leach, R. & Kleiner, R. In vitro selection with a site-specifically modified RNA library reveals the binding preferences of N6 -methyladenosine (m6A) reader proteins. *Biochemistry* **58**, 3386–3395 (2019).

237. Bodi, Z. & Fray, R. The role of m6A in ZBP bindng to B-actin mRNA. *Unpublished work*

238. Arber, W. & Linn, S. DNA modification and restriction. *Annu. Rev. Biochem.* **38**, 467–500 (1969).

239. Brownlee, C. Danna and Nathans: Restriction enzymes and the boon to modern molecular biology. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 5909 (2005).

240. Inoue, H., Hayase, Y., Iwai, S. & Ohtsuka, E. Sequence-dependent hydrolysis of RNA using modified oligonucleotide splints and RNase H. *Nucleic Acids Symp. Ser.* **215**, 327–330 (1987).

241. Berkower, I., Leis, J. & J, H. Isolation and characterization of an endonuclease from Escherichia coli specific for ribonucleic acid in ribonucleic acid-deoxyribonucleic acid hybrid structures. *J. Biol. Chem.* **248**, 5914–5921 (1973).

242. Donis-Keller, H. Site specific enzymatic cleavage of RNA. *Nucleic Acids Res.* **7**, 179–192 (1979).

243. Yu, Y., Shu, M. & Steitz, J. A new method for detecting sites of 2-O-methylation in RNA molecules. *RNA* **3**, 324–331 (1997).

244. Zhao, X. & Yu, Y. Detection and quantitation of RNA base modifications. *RNA* **10**, 996–1002 (2004).

245. Lapham, J. & Crothers, D. RNase H cleavage for processing of in vitro transcribed RNA for NMR studies and RNA ligation. *RNA* **2**, 289–296 (1996).

246. Grosjean, H., Keith, G. & Droogmans, L. Detection and quantification of modified nucleotides in RNA using thin-layer chromatography. *Methods Mol. Biol.* **256**, 357–391 (2004).

247. Wu, G., Xiao, M. & Yang, C. U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. *EMBO J.* **30**, 79–89 (2011).

248. Ma, X., Yang, C. & Alexandrov, A. Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism. *EMBO J.* **24**, 2403–2413 (2005).

249. Abouhaidar, M. G. & Ivanovb, I. G. Non-enzymatic RNA hydrolysis promoted by the combined catalytic activity of buffers and magnesium ions. *Zeitschrift für Naturforsch. C* **54**, 542–548 (1999).

250. Properties of DNA and RNA Ligases. (2015). Available at: https://www.neb.com/tools-and-resources/selection-charts/properties-of-dna-and-rna-ligases. (Accessed: 12th April 2019)

251. Lohman, G. J. S., Zhang, Y., Zhelkovsky, A. M., Cantor, E. J. & Evans, T. C. Efficient DNA ligation in DNA – RNA hybrid helices by Chlorella virus DNA ligase. *Nucleic Acids Res.* **42**, 1831–1844 (2014).

252. Wu, G., Xiao, M., Yang, C. & Yu, Y. U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. *EMBO J.* **30**, 79–89 (2010).

253. Moore, M. & Query, C. Splinted ligation method. *Methods Enzymol.* **317**, 109–123 (2000).

254. Jin, J., Vaud, S., Zhelkovsky, A. M., Posfai, J. & Mcreynolds, A. Sensitive and specific miRNA detection method using SplintR Ligase. *Nucleic Acids Res.* **44**, e116 (2016).

255. Shen, V. & Schlessinger, D. RNases, I, II, and IV of Escherichia coli. *Enzym.* **15**, 501–515 (1982).

256. Cuatrecasas, P. Catalytic properties and specificity of the extracellular nuclease of Staphylococcus aureus. *J. Biol. Chem.* **242**, 1541–1547 (1967).

257. Grant, K. B. & Dervan, P. B. Sequence-specific alkylation and cleavage of DNA mediated by purine motif triple helix formation. *Biochemistry* **35**, 12313–12319 (1996).

258. Deans, A. J. & West, S. C. DNA interstrand crosslink repair and cancer. *Nat. Rev. Cancer* **11**, 467–480 (2013).

259. Hong, I. S. & Greenberg, M. M. DNA interstrand cross-link formation

initiated by reaction between singlet oxygen and a modified nucleotide. **4**, 10510–10511 (2005).

260. Willcox, D., Fray, R. & Hayes, C. PhSeT selective crosslinking between adenosine and m6A. *Unpublished work*

261. Tomasz, M. Mitomycin C: small, fast and deadly (but very selective). *Chem. Biol.* **2**, 575–579 (1995).

262. Dohno, C., Shibata, T. & Nakatani, K. Discrimination of N6-methyl adenine in a specific DNA sequence w. *Chem. Commun.* **46**, 5530–5532 (2010).

263. Randolph, J. B. & Waggoner, A. S. Stability, specificity and fluorescence brightness of multiply-labeled fluorescent DNA probes. *Nucleic Acids Res.* **25**, 2923–2929 (1997).

264. Hogan, D. J., Riordan, D. P. & Gerber, A. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol.* **6**, e255 (2008).

265. Kricka, L. J. & Fortina, P. Analytical ancestry: “firsts” in fluorescent labeling of nucleosides, nucleotides, and nucleic acids. *Clin. Chem.* **55**, 670–683 (2009).

266. Wilchek, M. & Bayer, E. A. Applications of avidin-biotin technology: literature survey. in *Methods in Enzymology* **184**, 14–45 (Academic Press, New York, 1990).

267. Evans, M. D., Dizdaroglu, M. & Cooke, M. S. Oxidative DNA damage and disease: induction, repair and significance. *Mutat. Res.* **567**, 1–61 (2004).

268. Kirchner, J., Sigurdsson, S. & Hopkins, P. Interstrand cross-linking of duplex DNA by nitrous acid: covalent structure of the dG-to-dG cross-link at the sequence 5’-CG. *J. Am. Chem. Soc.* **114**, 4021–4027 (1992).

269. Douwel, D. K. *et al.* XPF-ERCC1 acts in unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4. *Mol. Cell* **54**, 460–471 (2014).

270. Szybalski, W. & Iyer, V. Crosslinking of DNA by enzymatically or chemically activated mitomycins and porfiromycins, bifunctionally ‘alkylating’ antibiotics. *Fed. Proc.* **23**, 946–947 (1964).

271. Hong, I. S. & Greenberg, M. M. Efficient DNA interstrand cross-link formation from a nucleotide radical. *J. Am. Chem. Soc.* **127**, 3692–3693 (2005).

272. Ding, H., Majumdar, A., Tolman, J. R. & Greenberg, M. M. Multinuclear NMR and kinetic analysis of DNA interstrand cross-link formation. *J. Am. Chem. Soc.* **130**, 17981–17987 (2008).

273. Proudnikov, D. & Mirzabekov, A. Chemical methods of DNA and RNA fluorescent labeling. *Nucleic Acids Res.* **24**, 4535–4542 (1996).

274. Fleming, A. M., Alshykhly, O., Zhu, J., Muller, J. G. & Burrows, C. J. Rates of chemical cleavage of DNA and RNA oligomers containing guanine oxidation products. *Chem. Res. Toxicol.* **28**, 1292–1300 (2015).

275. Clercq, E. De & Field, H. J. Antiviral prodrugs – the development of successful prodrug strategies for antiviral chemotherapy. *Br. J. Pharmacol.* **147**, 1–11 (2006).

276. Lodish, H., Berk, A. & Zipursky, S. *Molecular cell biology*. (W.H.Freeman, 2000).

277. Freeman, W., Walker, S. & Vrana, K. Quantitative RT-PCR: pitfalls and potential. *Biotechniques* **26**, 112–125 (1999).

278. Bartlett, J. & Stirling, D. A short history of the polymerase chain reaction. in *PCR Protocols. Methods in Molecular Biology* 3–6 (2003).

279. Sanger, F., Nicklen, S. & Coulson, A. DNA sequencing with chain-terminating inhibitors. *Biochemistry* **74**, 5463–5467 (1977).

280. Seo, Y. J., Malyshev, D. A., Lavergne, T., Ordoukhanian, P. & Romesberg, F. E. Site-specific labeling of DNA and RNA using an efficiently replicated and transcribed class of unnatural base pairs. *J. Am. Chem. Soc.* **133**, 19878–19888 (2011).

281. Hsu, G. W., Ober, M., Carell, T. & Beese, L. S. Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase. *Nature* **431**, 217–221 (2002).

282. Maden, B. E. H. Mapping 2'-O-methyl groups in ribosomal RNA. *Methods* **25**, 374–382 (2001).

283. Dong, Z. *et al.* RTL-P: a sensitive approach for detecting sites of 2'-O-methylation in RNA molecules. *Nucleic Acids Res.* **40**, e157 (2012).

284. Flusberg, B., Webster, D. & Lee, J. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **7**, 461–465 (2015).

285. Wang, S. *et al.* N6-methyladenine hinders RNA -and DNA-directed DNA synthesis: application in human rRNA methylation analysis of clinical specimens. *Chem. Sci.* **7**, 1440–1446 (2016).

286. Hong, I. S., Ding, H. & Greenberg, M. M. Radiosensitization by a modified nucleotide that produces DNA interstrand cross-links under hypoxic conditions. *J. Am. Chem. Soc.* **5**, 2230–2231 (2006).

287. Burgess, K. & Cook, D. Syntheses of nucleoside triphosphates. *Chem. Rev.* **100**, 2047−2059 (2000).

288. Yoshikawa, M., Kato, T. & Takenishi, T. Studies of phosphorylation. III. selective phosphorylation of unprotected nucleosides. *Bull. Chem. Soc. Jpn.* **42**, 3505–3508 (1969).

289. Ludwig, J. & Eckstein, F. Rapid and efficient synthesis of nucleoside 5'-0-(1-thiotriphosphates), 5'-triphosphates and 2',3'-cyclophosphorothioates using 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one. *Am. Chem. Soc.* **54**, 631–635 (1989).

290. Lee, D., Shin, Y., Chung, S., Hwang, K. S. & Yoon, D. S. Simple and highly sensitive molecular diagnosis of Zika virus by lateral flow assays. *Anal. Chem.* **88**, 12272−12278 (2016).

291. Bst 3.0 DNA Polymerase. (2015). Available at: https://www.neb.com/products/m0374-bst-3-0-dna-polymerase#Product Information_Properties & Usage. (Accessed: 18th April 2019)

292. Grünweller, A. & Hartmann, R. Locked nucleic acid oligonucleotides: the next generation of antisense agents? *BioDrugs* **21**, 235–243 (2007).

293. N6-Methyl-A-CE Phosphoramidite. (2019). Available at: https://www.glenresearch.com/n6-methyl-a-ce-phosphoramidite-1.html. (Accessed: 10th June 2019)

294. Palframan, M. *et al.* Synthesis of alkylated nucleoside phosphoramidites. *Unpublished work*

295. Polo, L. & Limbach, P. Analysis of oligonucleotides by electrospray ionization mass spectrometry. *Curr. Protoc. Nucleic Acid Chem.* **1**, 10.2.1-10.2.20 (2001).

Appendix

The Professional Internship for PhD Students (PIPS)

Note to examiners: This statement is included as an appendix to the thesis in order that the thesis accurately captures the PhD training experienced by the candidate as a BBSRC Doctoral Training Partnership student.

The Professional Internship for PhD Students is a compulsory 3-month placement which must be undertaken by DTP students. It is usually centred on a specific project and must not be related to the PhD project. This reflective statement is designed to capture the skills development which has taken place during the student's placement and the impact on their career plans it has had.

PIPS Reflective Statement

I spent three months completing my Professional Internship for PhD Students (PIPS) through the Net4FS, as part of my BBSRC-funded, Doctoral Training Programme (DTP) PhD. I worked with the research group of Professor Zheng Yuan at Shanghai Jiao Tong University (SJTU) in China.

The aim of this placement was to undertake research into the network of proteins that regulate Jasmonic acid (JA) dependent flower development in Rice. A previous student had completed a number of yeast 2 hybrid assays to identify key protein-protein interactions in the Rice JA signalling pathway. The aim of my research was to validate these protein-protein

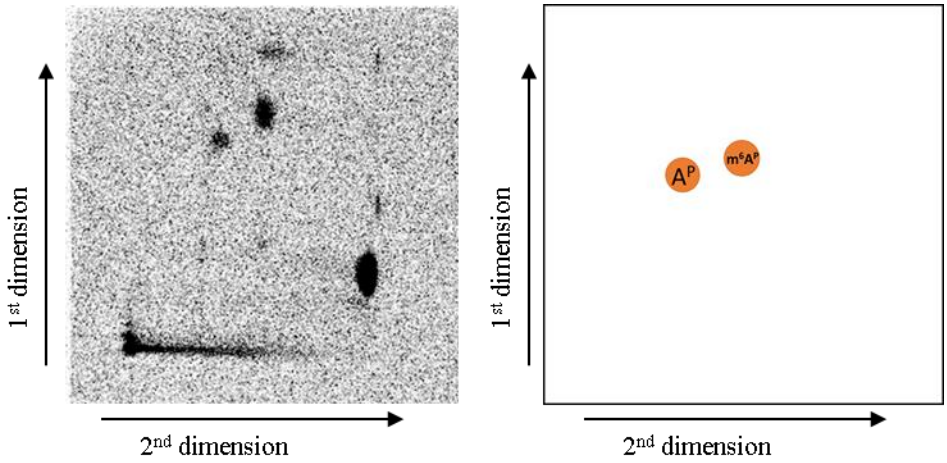
interaction results using biomolecular inflorescence complementation (BIFC) assays yeast 2 hybrid. We identified 79 BIFC experimental combinations that needed to be analysed. Over the 3 months we completed the majority of the molecular cloning and experimentation for the majority of these combinations.

In a separate project we used yeast 2 hybrid analysis to identify novel interacting partners of the key JA signalling proteins JAZ1 and MADS32. We managed to complete all of the cloning for this experiment however due to time constraints we were unable to undertake the yeast 2 hybrid assays during my placement. During my placement I presented both my literature and research findings in research group meetings.

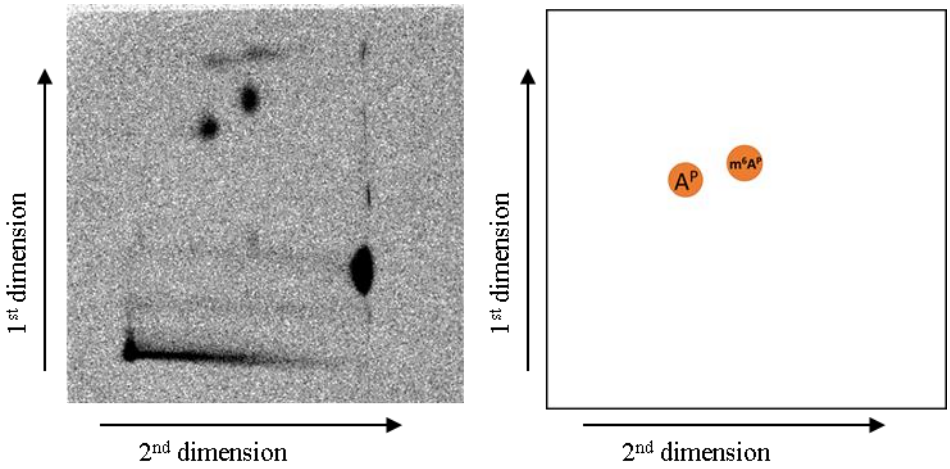
My placement gave me lots of new experiences in molecular biological techniques and plant research methods. I found it very interesting to work in a completely different research environment and enjoyed the challenge of living and working abroad. During my placement I also enjoyed the opportunity to travel around Shanghai and its neighbouring cities.

I thoroughly enjoyed my time working in Shanghai. I am very grateful to Miriam Colombi and Professor Zoe Wilson for arranging my placement. I would like to thank Professor Zheng Yuan for supervising my project and allowing me to be part of his research group. I would also like to thank all of the members of my research group for being so friendly and accommodating and for all of their help during my placement.

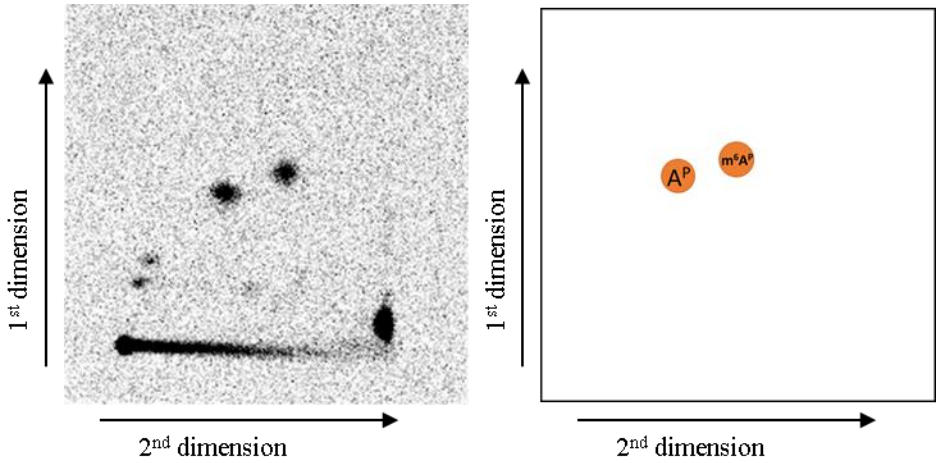
Supplementary data



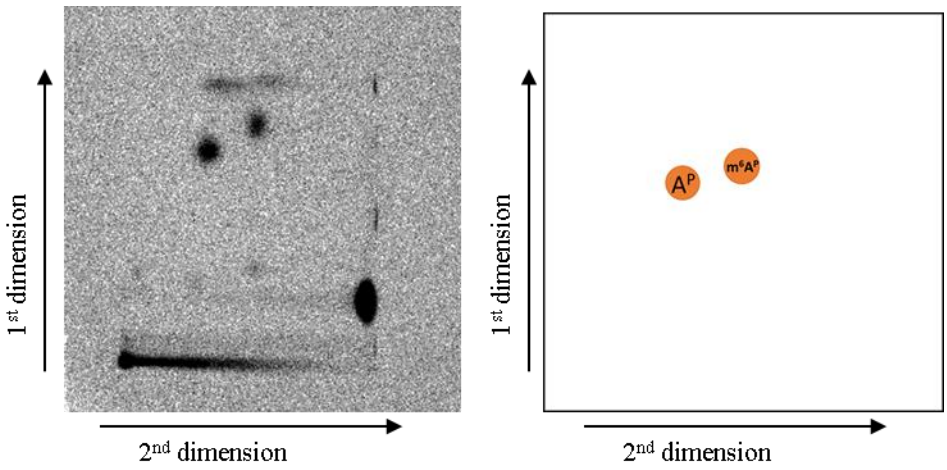
Supplementary Figure 1: RedBaron test quantification 80 % m^6A.



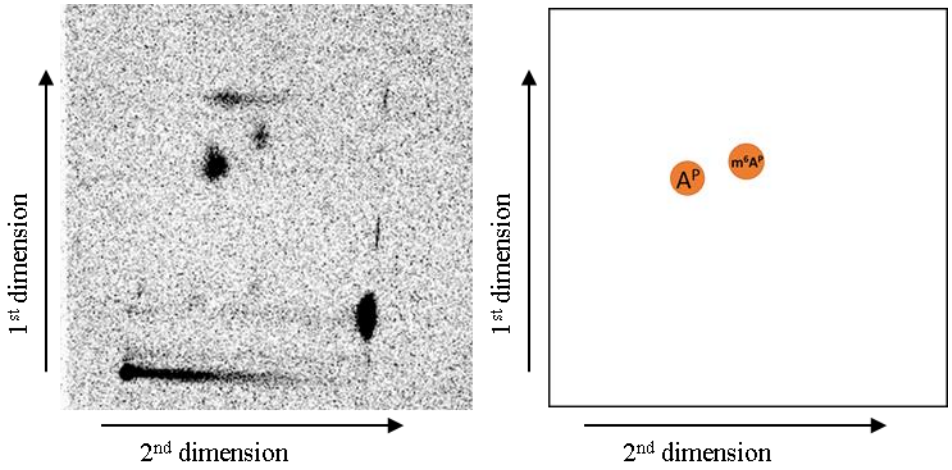
Supplementary Figure 2: RedBaron test quantification 60 % m^6A.



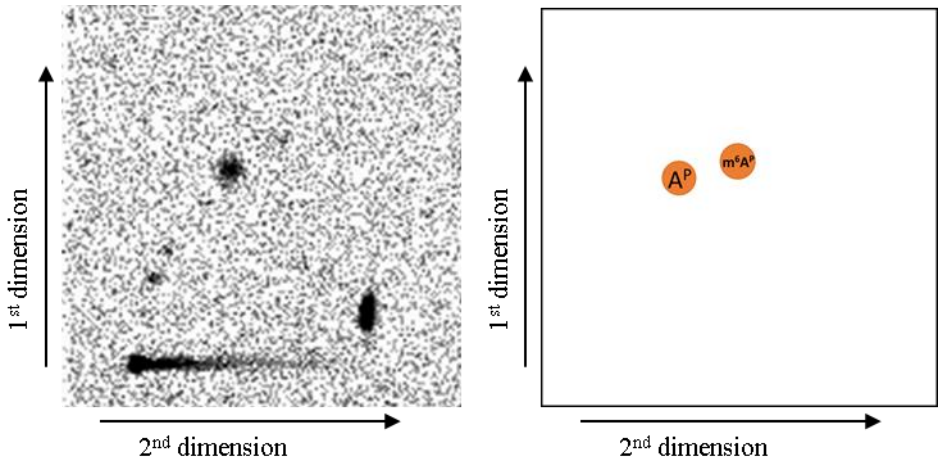
Supplementary Figure 3: RedBaron test quantification 50 % m^6A.



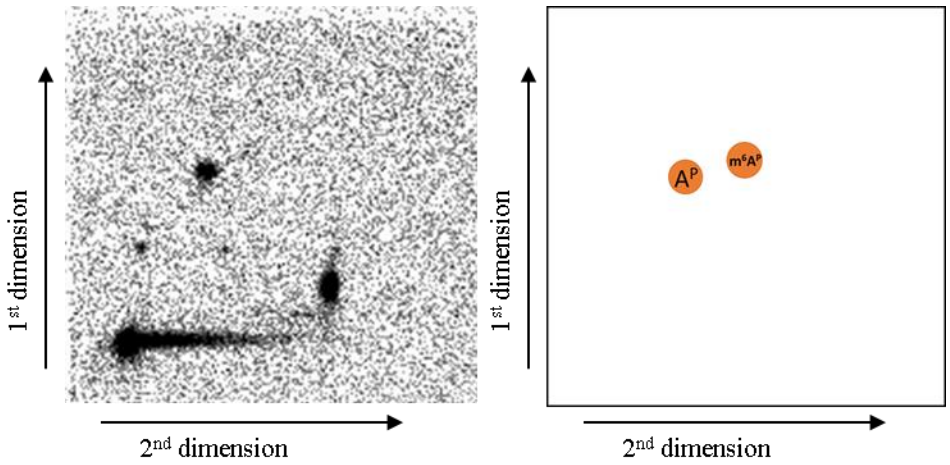
Supplementary Figure 4: RedBaron test quantification 40 % m^6A.



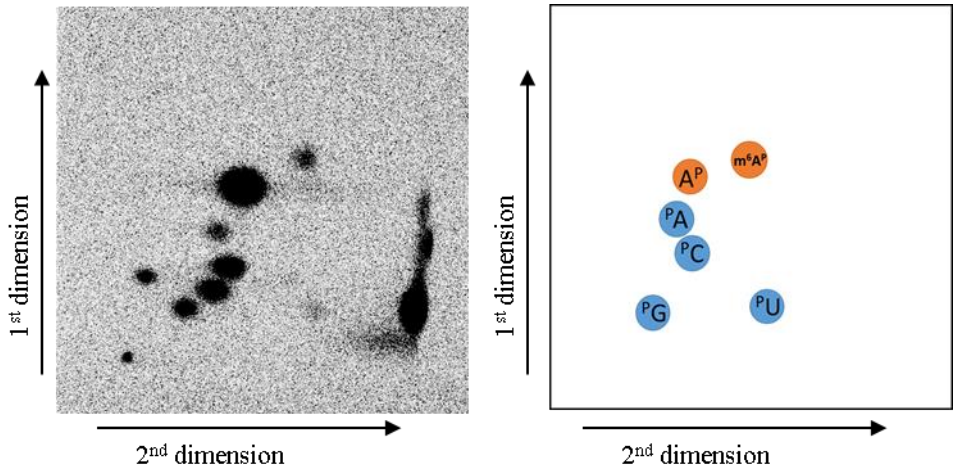
Supplementary Figure 5: RedBaron test quantification 20 % m^6A.



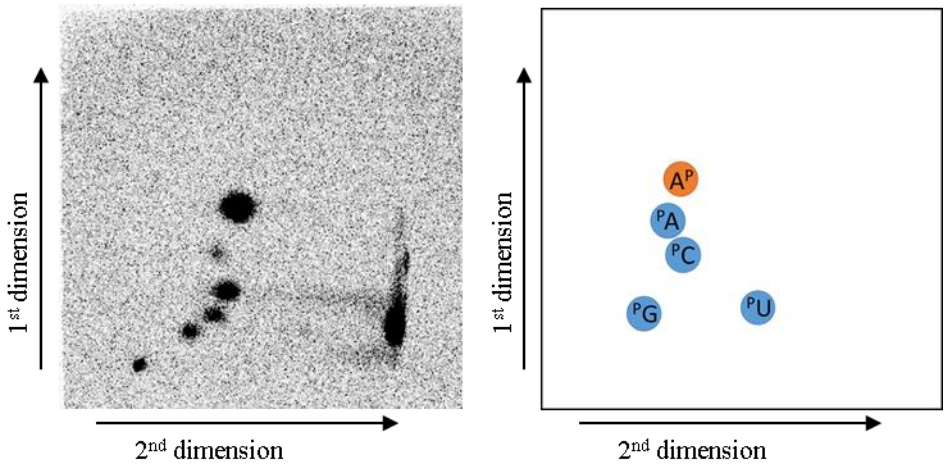
Supplementary Figure 6: RedBaron test quantification 10 % m^6A.



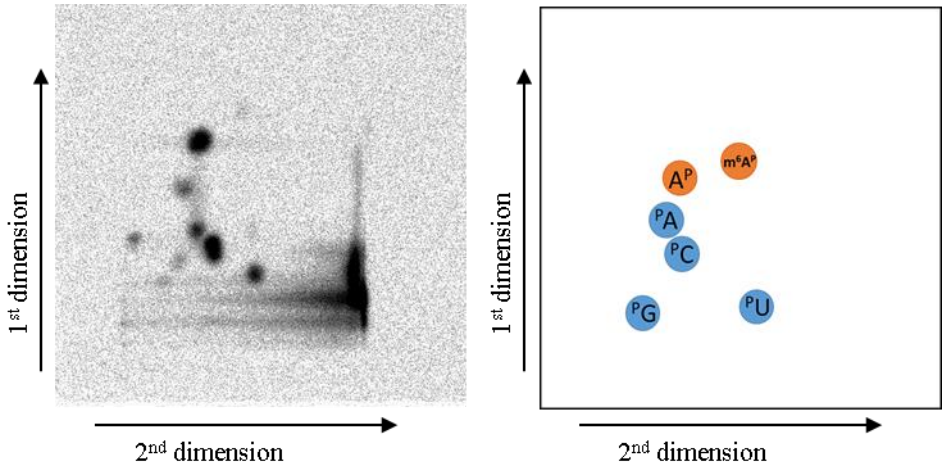
Supplementary Figure 7: RedBaron test quantification 5 % m^6A.



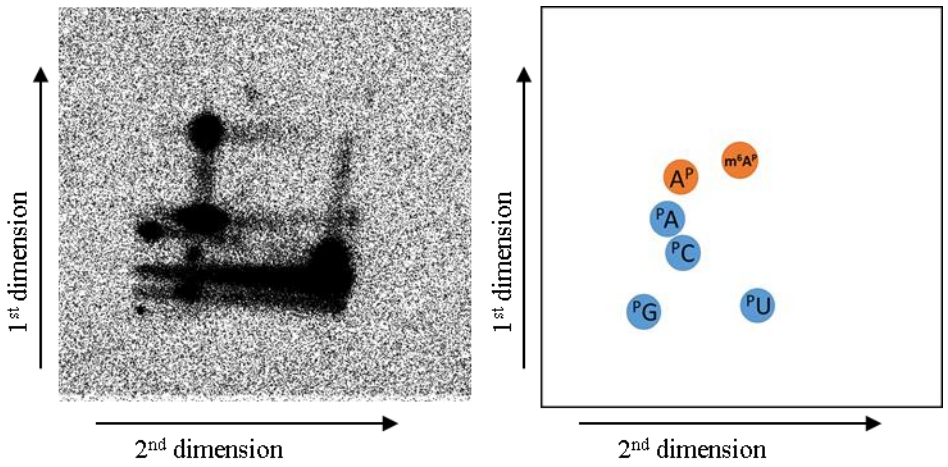
Supplementary Figure 8: RedBaron Chicken embryo β-actin (site 1202).



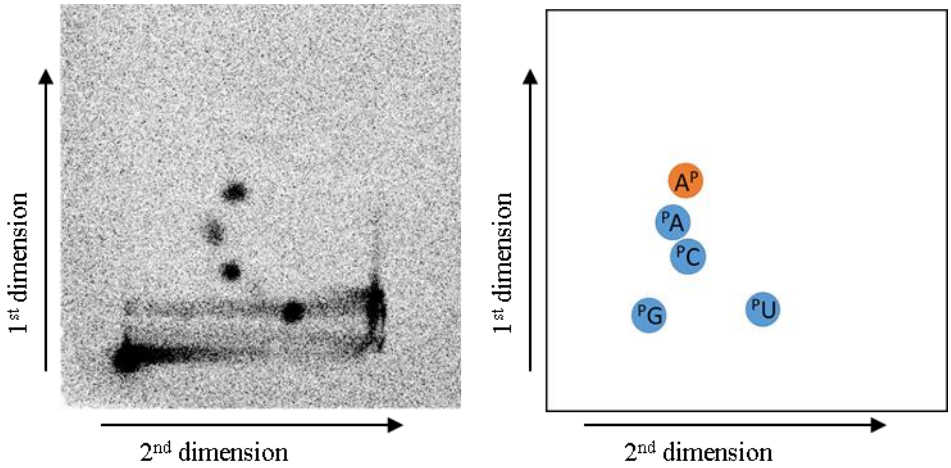
Supplementary Figure 9: RedBaron Chicken embryo β-actin (site 1218).



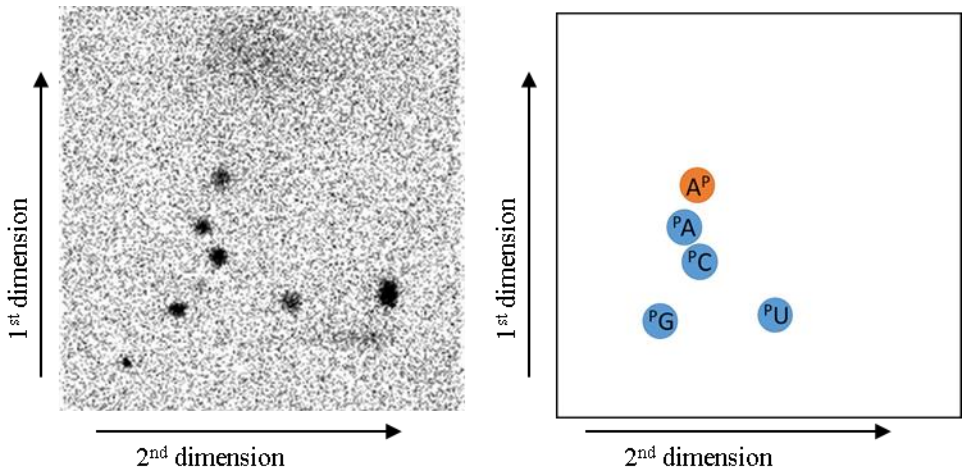
Supplementary Figure 10: RedBaron fibroblast β-actin (site 1202).



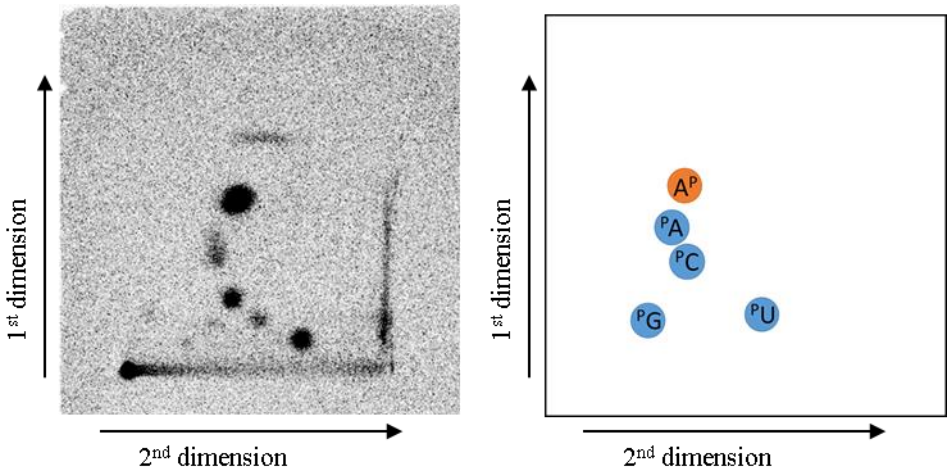
Supplementary Figure 11: RedBaron Human HeLa cells β-actin (site 1217).



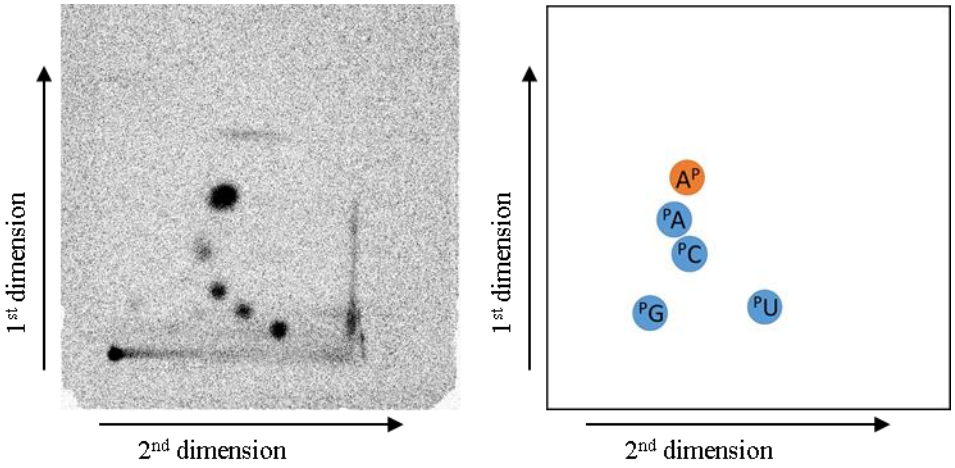
Supplementary Figure 12: RedBaron Mouse heart β-actin (site 1242).



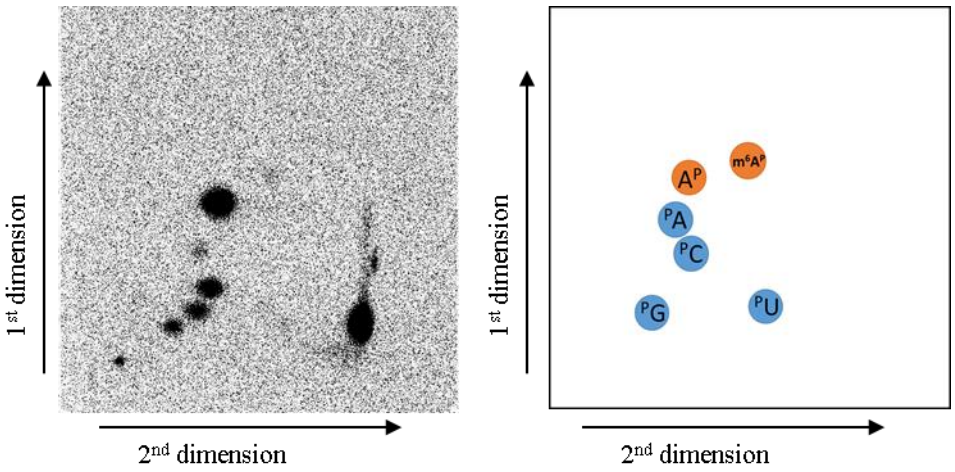
Supplementary Figure 13: RedBaron Mouse liver β-actin (site 1242).



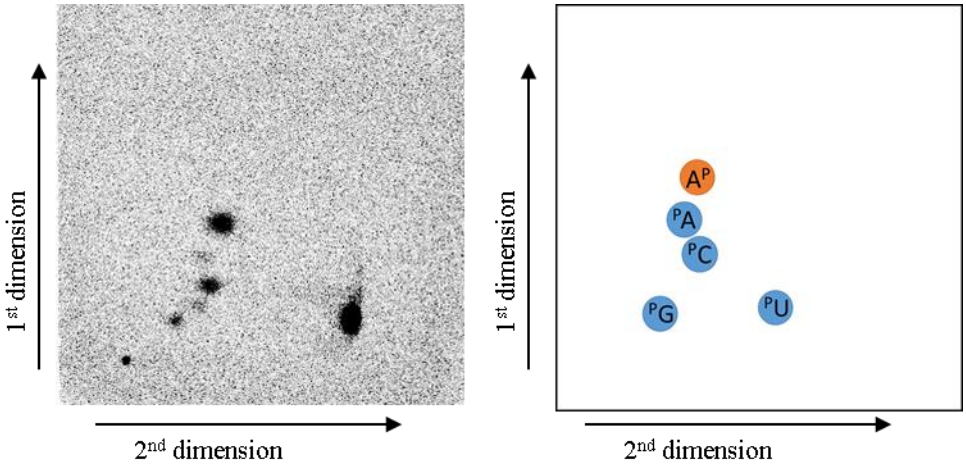
Supplementary Figure 14: RedBaron Mouse kidney β-actin (site 1242).



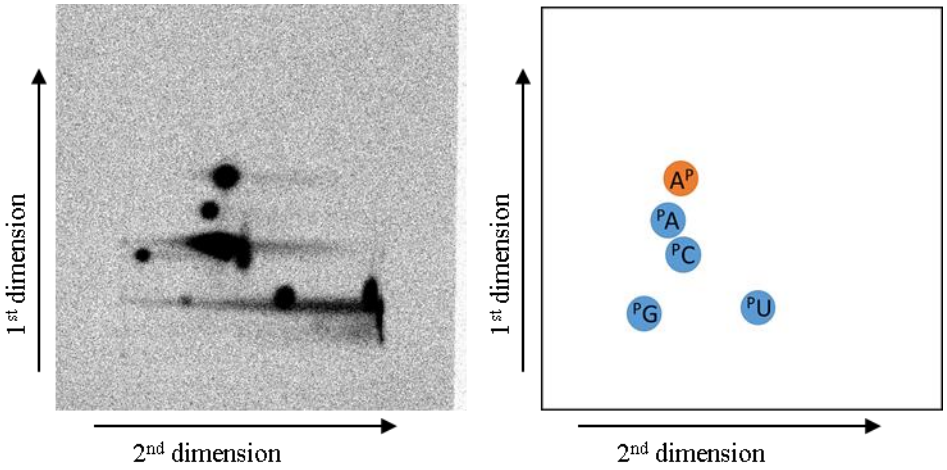
Supplementary Figure 15: RedBaron Mouse testis β-actin (site 1242).



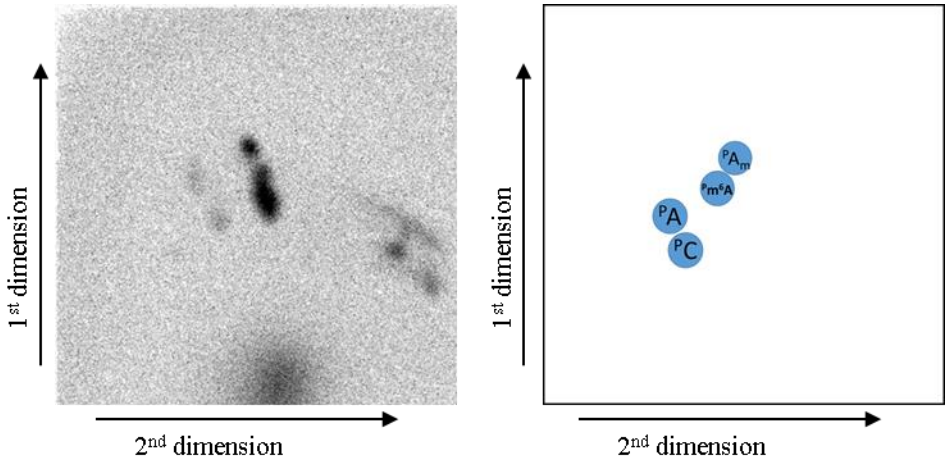
Supplementary Figure 16: RedBaron Mouse brain β-actin (site 1242).



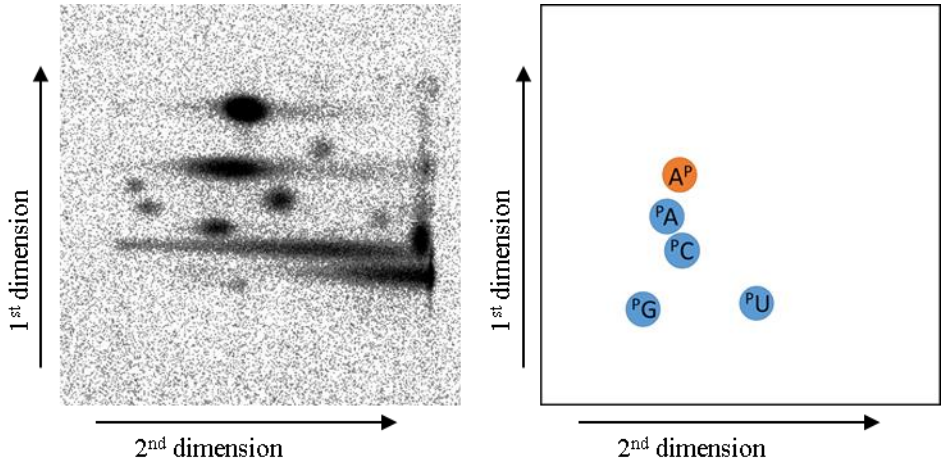
Supplementary Figure 17: RedBaron Mouse brain β-actin (site 1263).



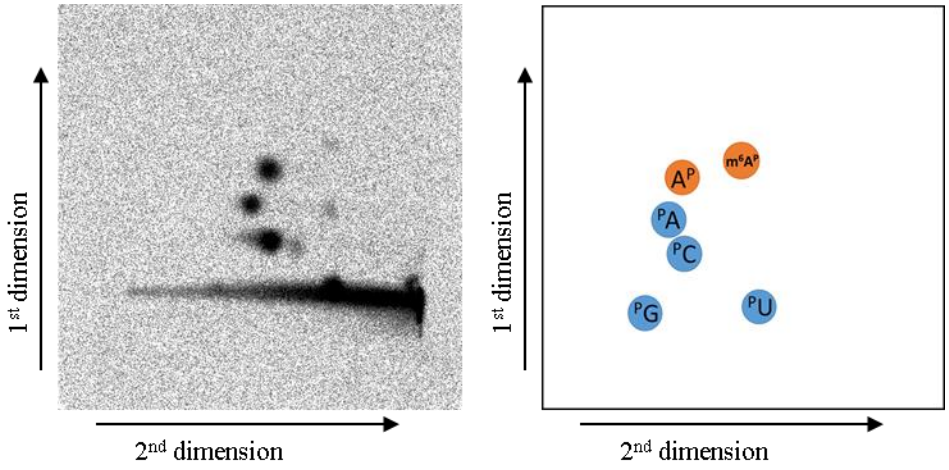
Supplementary Figure 18: RedBaron MALAT1 (site 2515).



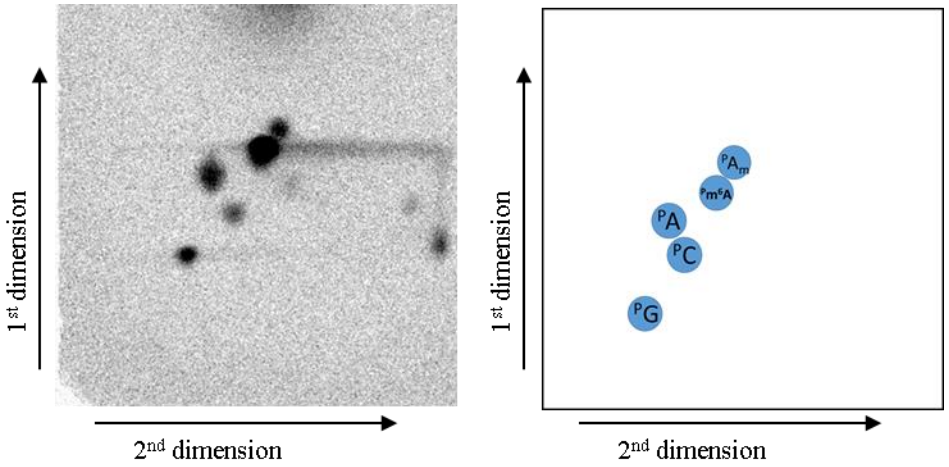
Supplementary Figure 19: SCARLET MALAT1 (site 2515).



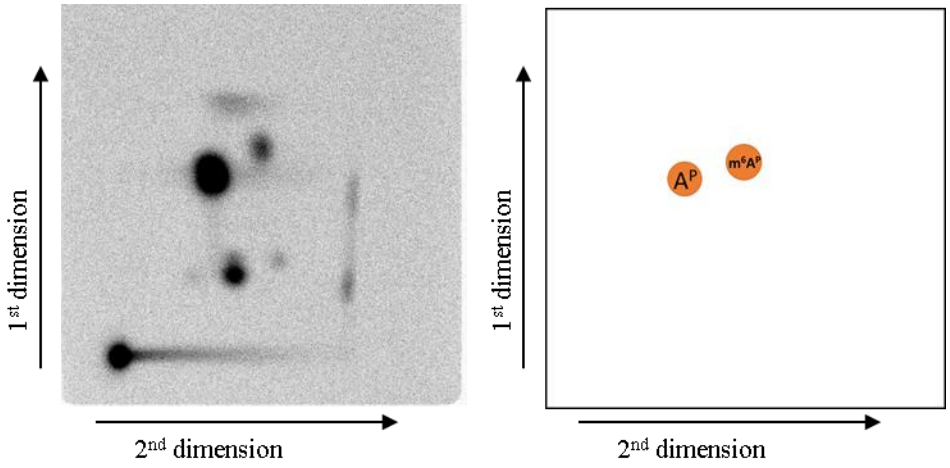
Supplementary Figure 20: RedBaron 28S rRNA (site A4189).



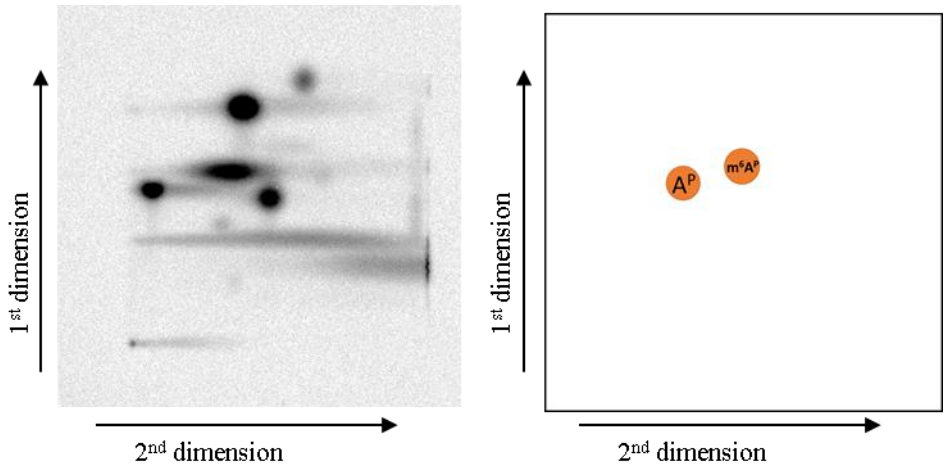
Supplementary Figure 21: RedBaron 28S rRNA (site A4190).



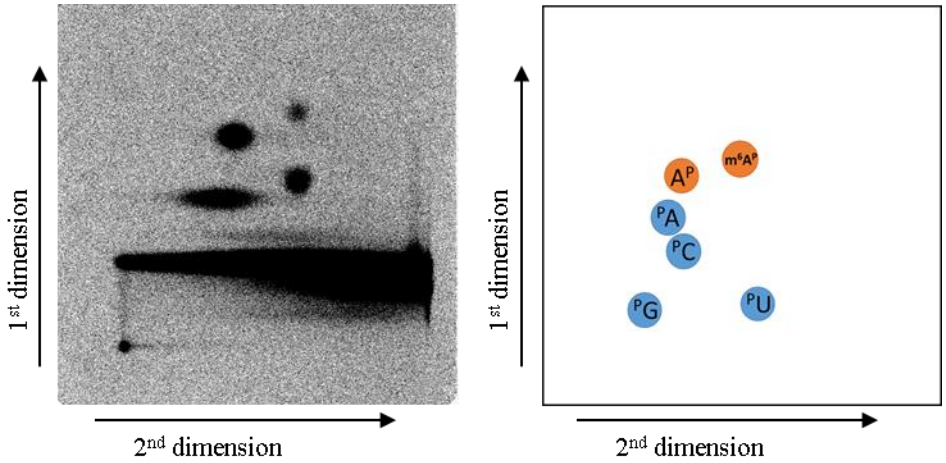
Supplementary Figure 22: SCARLET 28S rRNA (site A4190).



Supplementary Figure 23: RedBaron (test) chicken β-actin oligonucleotide 50 % m^6A (with mouse embryo input RNA).



Supplementary Figure 24: RedBaron (test) mouse β-actin oligonucleotide 50 % m^6A (with chicken embryo input RNA).



Supplementary Figure 25: RedBaron (EDTA before heating) HeLa 28S rRNA (site A4190).