UNIVERSITY OF SOUTHAMPTON

FACULTY OF NATURAL & ENVIRONMENTAL SCIENCES

SCHOOL OF CHEMISTRY

Enzymatic Modification of DNA and RNA 3′-Termini for Click Ligation

by

Xiong Chen

Thesis for the degree of Doctor of Philosophy

May 2014

This thesis is dedicated to the memory of my grandmas and to my parents and family in China.

UNIVERSITY OF SOUTHAMPTON
FACULTY OF NATURAL & ENVIRONMENTAL SCIENCES
SCHOOL OF CHEMISTRY
Doctor of Philosophy

ENZYMATIC MODIFICATION OF DNA AND RNA 3′-TERMINI FOR CLICK LIGATION

by Xiong Chen

ABSTRACT

Introducing alkyne or azide groups at the DNA/RNA 3′-end via enzymatic incorporation allows copper-catalysed alkyne-azide cycloaddition (CuAAC) or copper-free Strain Promoted Alkyne-Azide Cycloaddition (SPAAC) Click ligation reactions to be performed on PCR products and other natural DNA and RNA. The subsequent chemical ligation can be utilised for transformations which could be difficult to carry out using DNA/RNA ligases. To explore this approach, the enzymatic labelling of DNA and RNA at the 3′-end using commercially available or synthesised 2′ or 3′ alkyne/azide modified nucleoside triphosphates has been investigated. For DNA, 3′-*O*-propargylthymidine (PrdTTP) was synthesised from thymidine and triphosphorylation was performed in solution using $POCl_3$ (Method I). For RNA, 2′ and 3′-*O*-propargyluridine/ adenosine-5′-triphosphates were synthesised from commercially available nucleoside-functionalised resins using the salicyl chlorophosphite triphosphorylation method (Method II).

Both PrdTTP and commercially available 3′-azido-3′-deoxythymidine-5′-triphosphate (AzdTTP) were incorporated efficiently by 9°N$_m$ polymerase in a templated primer extension reaction. To demonstrate the efficiency of this enzymatic reaction, a linear end-sealed dsDNA was constructed using AzdTTP. Utilising the read-through capability of the triazole backbone generated in this study, a modified TA-cloning strategy has been proposed.

For RNA, only commercially available 2′-azido-2′-deoxyuridine/cytidine-5′-triphosphate (2′-AzdUTP/2′-AzdCTP) were successfully incorporated by Yeast Poly(A) Polymerase (Yeast PAP). RNA Click ligation was explored for 3′-adaptor ligation for use in the small RNA sequencing protocols to solve the sequence-bias problem caused by enzymatic ligation. The reverse-transcription of Click templates with a triazole link generated by CuAAC was found to be dependent on the triazole structure and the sequence at the triazole site. A triazole link generated by copper-free Click ligation was also found to be successfully read-through by M-MuLV Reverse Transcriptase (RNase H$^-$). This may enable copper-free Click ligation to be used for small RNA sequencing.

CONTENTS

Declaration

I, Xiong Chen, declare that the thesis entitled:

Enzymatic Modification of DNA and RNA 3´-Termini for Click ligation

and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as:
  Chen, X., El-Sagheer, A.H. & Brown, T. Reverse transcription through a bulky triazole linkage in RNA: implications for RNA sequencing. *Chem. Commun.* (2014).

Signed:

Date: June 12th, 2014

Acknowledgements

I would like to express my deep gratitude to Professor Tom Brown, my research supervisor, for his design of this project and for teaching me how to manage my project from simple to more complicated tasks, his appreciation of my own ideas regardless of the maturity of them, and most importantly his critique of the project and guidance that allowed me to develop my skills and ideas throughout the PhD programme.

I would also like to thank especially Dr Afaf H. El-Sagheer for the enormous help she gave me during these 4 years. Thank you for your patience and assistance which helped me to enter the nucleic acid field. The numerous discussions of my project were invaluable for the progress of this project and are greatly appreciated. The questions, corrections and suggestions I received from these discussions taught me how to think and solve problems in a logical and scientific manner. Thank you also for your support and encouragement when I was facing difficulties.

I would also like to extend my thanks to ATDBio® Ltd. for oligonucleotide synthesis and project funding, Dr Simon Gerrard for his help with reaction mechanisms, and other colleagues in our research group and department who provided help during the progress of the project.

I wish to thank my parents and family for their love, support, patience and understanding when I was away from home. I would also like to thank for all my friends for their willingness to listen when I was feeling lonely. Thanks for the opportunities to share our stories and our lives with each other. Finally, thank God who made all the things possible for every blessing he has given me.

ABBREVIATIONS

AAC	alkyne-azide cycloaddition
AMP	adenosine monophosphate
AzdTTP	3′-azido-3′-deoxythymidine triphosphate
AZT	3′-azido-3′-deoxythymidine
ATP	adenosine triphosphate
BARAC	biarylazacyclooctynone
BCN	bicyclo[6.1.0]nonyne
BTTAA	2-[4-{(bis[(1-tert-butyl-1*H*-1,2,3-triazol-4-yl)methyl]amino)methyl}-1*H*-1,2,3-triazol-1-yl]acetic acid
BTTES	2-[4-{(bis[(1-tert-butyl-1*H*-1,2,3-triazol-4-yl)methyl]amino)-methyl}-1*H*-1,2,3-triazol-1-yl]ethyl hydrogen sulphate
BSA	bovine serum albumin
cAMP	adenosine-3′,5′-cyclic monophosphate
$CDCl_3$	deuterated chloroform
CE	capillary electrophoresis
CPG	controlled pore glass
COSY	^{1}H-^{1}H correlation spectroscopy
CuAAC	copper-catalysed alkyne-azide cycloaddition
DCA	dichloroacetic acid
DCM	dichloromethane
ddNTPs	dideoxynucleoside triphosphates
DIBO	dibenzocyclooctyne
DIPEA	diisopropylethylamine
DMSO	dimethyl sulfoxide
DMT	4, 4′-dimethoxytrityl
dNTP	deoxyribonucleoside triphosphate
dsDNA	double-stranded DNA
DTT	dithiothreitol
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	ethylenediaminetetraacetic acid
ESI	electrospray ionisation
FAM	6-carboxyfluorescein
G-clamp	aminoethyl-phenoxazine-2′-deoxycytidine analogue (AP-dC)
HMQC	^{1}H-^{13}C correlation spectrometry
HPLC	High Performance Liquid Chromatography

HSBC	Single bond chemical shift correlation spectroscopy
LCAA	long chain alkylamino
LC-MS	Liquid Chromatography – Mass spectroscopy
MS	mass spectrometry
M.W.	molecular weight
MOFO	monofluorinated cyclooctyne
MSNT	1-(2-mesitylenesulfonyl)-3-nitro-1*H*-1,2,4-triazole
NAD^+	nicotinamide adenine dinucleotide
NHS	*N*-hydroxysuccinimide
NOFO	non-fluorinated cyclooctyne
NTP	ribonucleoside triphosphates
O.D.	optical density value
$O.D._{260\ nm}$	optical density value at 260 nm
ODN	oligodeoxyribonucleotide
ORN	oligoribonucleotide
PAGE	Polyacrylamide gel electrophoresis
PAP	Poly(A) polymerase (RNA)
PCR	Polymerase chain reaction
PPi	pyrophosphate
PrdTTP	3′-*O*-propargylthymidine triphosphate
PUP	Poly(U) polymerase (RNA)
RT	reverse-transcription
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SPAAC	strain-promoted alkyne-azide cycloaddition
ssDNA	single-stranded DNA
TBA	tri-*n*-butylamine
TBAF	tetra-*n*-butylammonium fluoride
TBDMS	*tert*-butyldimethylsilyl
TBDMSCl	*tert*-butyldimethylsilyl chloride
TBTA	tris((1-benzyl-1*H*-1,2,3,-triazol-4-yl)methyl)amine
TCA	trichloroacetic acid
TdT	Terminal deoxynucleotidyl Transferase
TEA	triethylamine
TEAB	triethylammonium bicarbonate
TLC	thin layer chromatography
THF	tetrahydrofuran
THPTA	tris[(1-hydroxypropyl-1*H*-1,2,3-triazol-4-yl)methyl]amine

T_m	Melting temperature
TPSCl	2,4,6-triisopropylbenzenesulfonyl chloride
tRNA	Transfer RNA
UTR	untranslated region
UV	Ultraviolet absorption

CHAPTER 1 Introduction

This introduction includes a concise description of the key topics that are relevant to my research project. This involved the incorporation of alkyne and azide modifications at the 3´-end of DNA and RNA and subsequent chemical (Click) ligation to introduce triazole linkages into nucleic acids for various applications.

1.1 Nucleic acid structure and physical properties

Nucleic acids are polymeric macromolecules that play a vital role in living organisms. Since the discovery of nucleic acids as the main storage media of genetic information, their importance has resulted in intense exploration and yielded a deep understanding of their biological functions [1]. At the same time, the development and improvement in oligonucleotide chemical synthesis has advanced their applications in biology, biotechnology and nanotechnology.

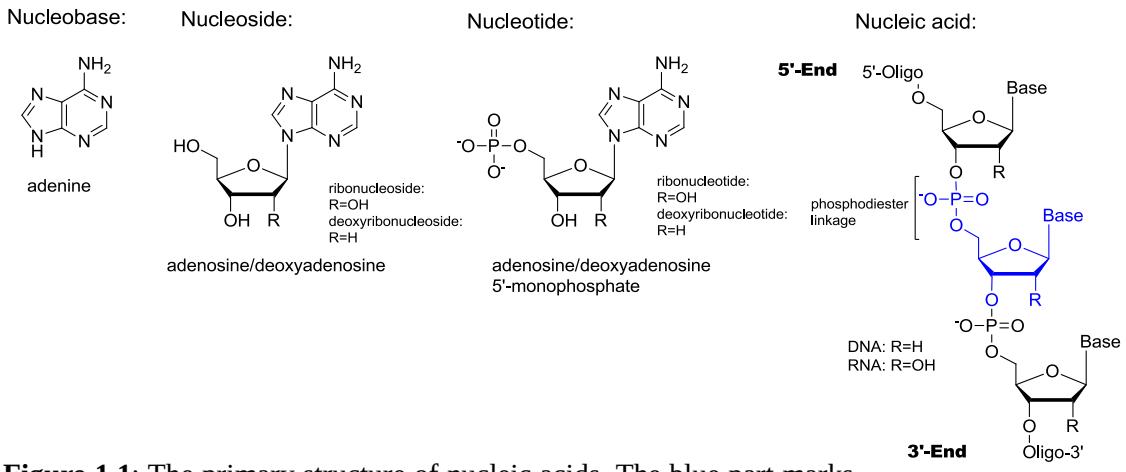


Figure 1.1: The primary structure of nucleic acids. The blue part marks a single nucleotide building block of the polymer chain.

The building blocks of nucleic acids are nucleotides which consist of a nucleobase, a pentose sugar ring and a phosphate group (Figure 1.1). The pentose is either a ribose or deoxyribose locked in a cyclic hemiacetal form by attaching of a nucleobase *via* a glycosidic bond. In the acyclic form, either of the hydroxyl groups at the C5´ and C4´ position can react with the aldehyde to form a hemiacetal, so the resultant cyclic sugar can either be pyranose or furanose (Figure 1.2). In natural nucleic acids, the nucleosides occur as the β-anomer (Figure 1.3). Based on the type of sugar, nucleic acids are divided to two categories: ribonucleic acid (RNA,

contains ribose) and deoxyribonucleic acid (DNA, contains deoxyribose). In most living organisms except certain viruses, DNA stores the genetic information, while RNAs such as messenger RNAs (mRNAs) and transfer RNAs (tRNAs) participate in important cellular functions. In nucleic acids, the 3′-OH of one nucleotide is attached to the 5′-phosphate of another nucleotide *via* a phosphodiester linkage. This makes the polymer chain directional and the two ends are referred as 5′-end and 3′-end (Figure 1.1). By default the sequence of nucleic acid is written from 5′ to 3′.

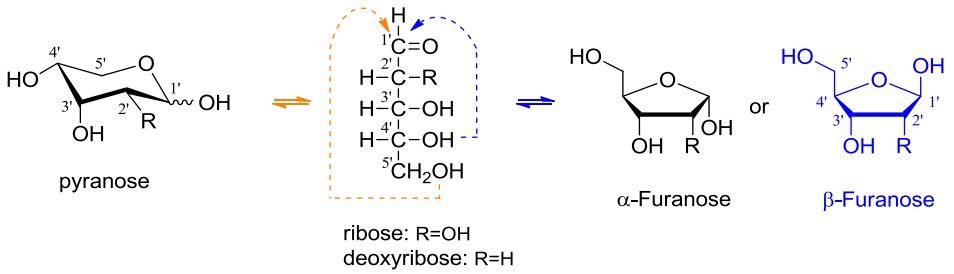


Figure 1.2: Formation of pyranose and furanose ribose/2′-deoxyribose sugar rings through cyclisation.

Four major nucleobases (A, G, C, T) provide the genetic code in DNA. In RNA, however, the major nucleobases are A, G, C and U. Adenine and guanine derive from purine, while cytosine, thymine and uracil derive from pyrimidine. Similarly to the sugar ring, the atoms on the purine or pyrimidine rings are numbered for easy reference. Due to the different aromatic structures of the nucleobases, nucleotides absorb UV light around 260 nm and different nucleotides have different absorption coefficients (ε) at 260 nm.

In addition to the standard structure mentioned above, nucleic acids in living organisms also undergo various modifications on nucleobases, sugar ring and phosphate [2]. The modified bases are abundant in certain nucleic acid molecules including eukaryotic mRNA 5′-cap and tRNAs (Figure 1.4). For mRNA, in addition to the 7-methylguanosine in the 5′-cap, the first and second nucleotide are often modified by 2′-*O*-methyl groups (2′-O-Me). One nucleobase modification: cytosine C5-methylation (Figure 1.4) in DNA was found to participate in regulation of eukaryotic gene expression. It is termed an epigenetic modification, because it functions as an addition to the general genetic code (A, G, C, T) and is inherited through a different pathway. Recently, other derivatives of 5-methylcytosine including 5-hydroxymethylcytosine (Figure 1.4) [3,4], 5-formylcytosine and 5-carboxylcytosine were also found, all of which are intermediates in cytosine demethylation [5,6].

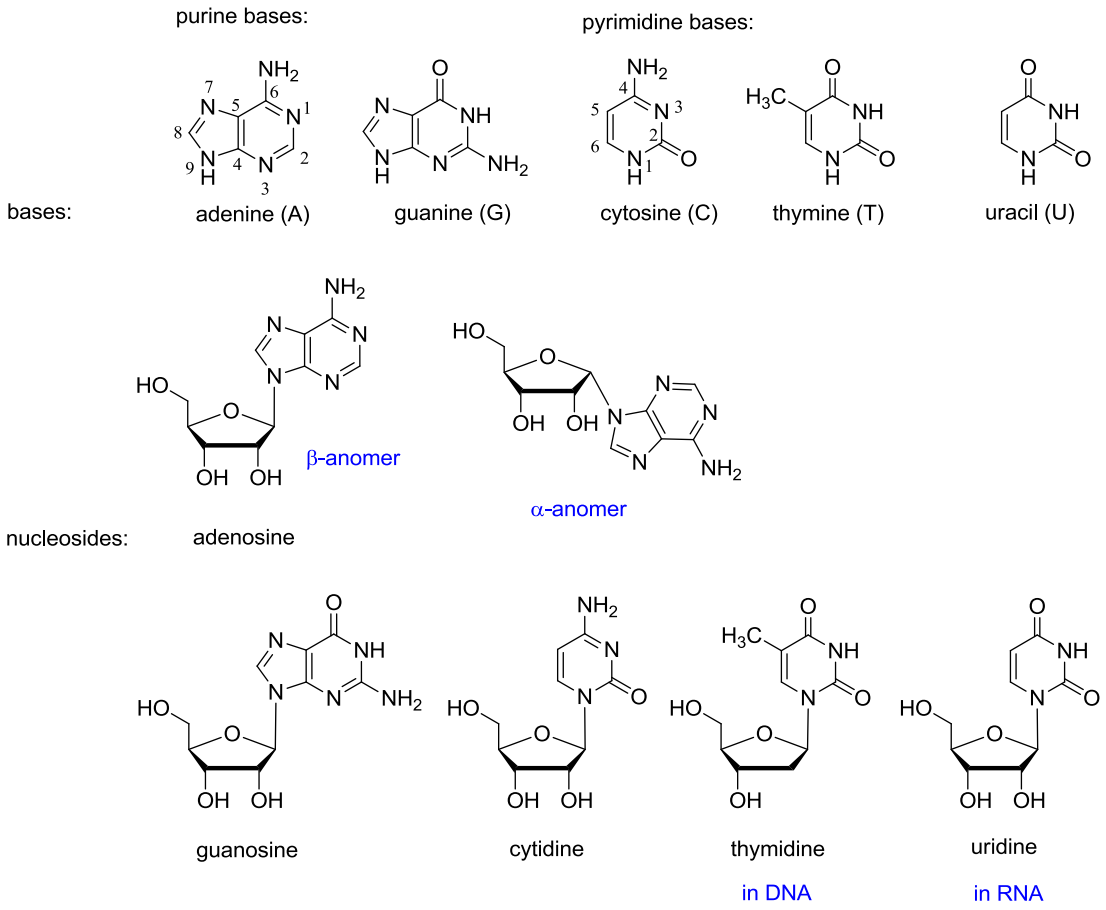


Figure 1.3: Nucleobases and their corresponding nucleoside names. In natural nucleic acids, the nucleosides occur as the β-anomer.

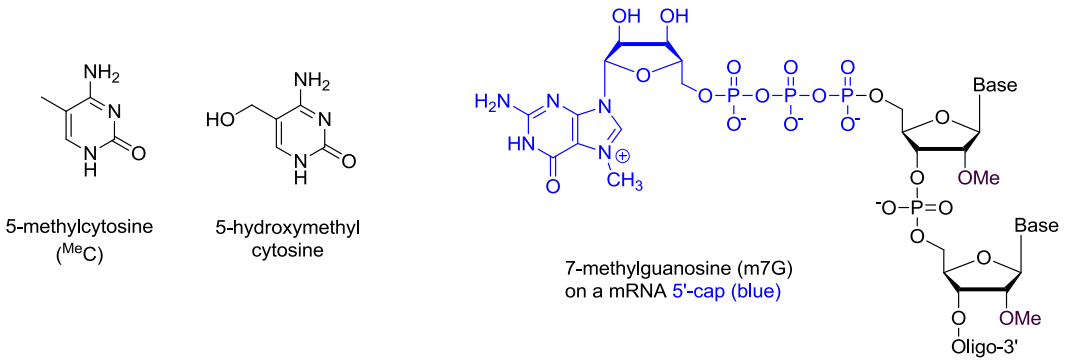


Figure 1.4: Examples of modified nucleotides in DNA and RNA. In mRNA the first and second nucleotide next to the 5´-cap (blue) also possess 2´-*O*-methyl modification.

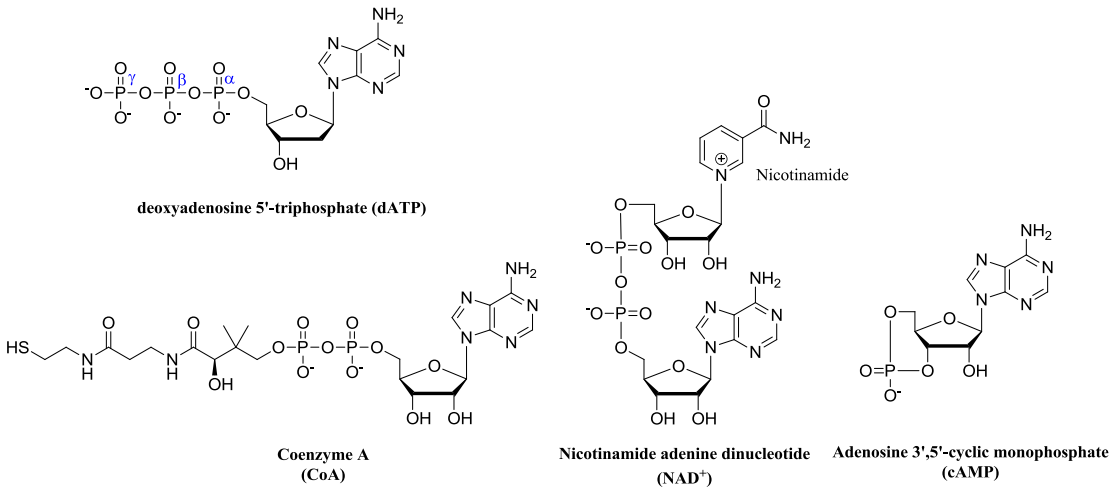


Figure 1.5: Small biomolecules containing nucleotides including nucleotide triphosphates, enzyme cofactor Coenzyme A, NAD^+, and regulatory nucleotide cAMP.

Nucleotides are not only found in macromolecules such as nucleic acids, they are also embedded in various small molecules that participate in important biological processes. The deoxynucleoside triphosphates such as dATP (Figure 1.5) participate in DNA replication which is discussed below. The similar adenosine 5´-triphosphate not only participates in RNA polymerization but also functions as an energy currency in cells. The three phosphates are labelled α, β and γ starting from the phosphate attached to the nucleoside. The energy released while hydrolysing the phosphate anhydride bonds can drive enzymatic reactions forward. Beside nucleoside triphosphates, the coenzyme A and nicotinamide adenine dinucleotide (NAD^+) are important enzyme cofactors, while adenosine-3',5'-cyclic monophosphate (cAMP) is one of the second messengers inside the cell that regulate downstream proteins in various signal pathways.

Based on the understanding that natural DNA is double-strand (dsDNA), and X-ray crystallography of DNA indicates a helical structure, Watson and Crick postulated the three-dimensional double-helix model. In this model, the hydrophilic phosphodiester backbones, including sugar rings, form the outside of the structure while the hydrophobic nucleobases form the core (Figure 1.6). More importantly, Watson and Crick also defined the base paring patterns between A and T (A:T), G and C (G:C). The G:C base pair is stronger than the A:T base pair because it is held together by three hydrogen bonds instead of two. The X-ray fibre diffraction data obtained by Rosalind Franklin also confirmed that the two phosphodiester backbones are anti-parallel as the 5´,3´-phosphodiester bonds in the two strand run in opposite directions. The anti-parallel model was also proved by studies with DNA polymerases [7].

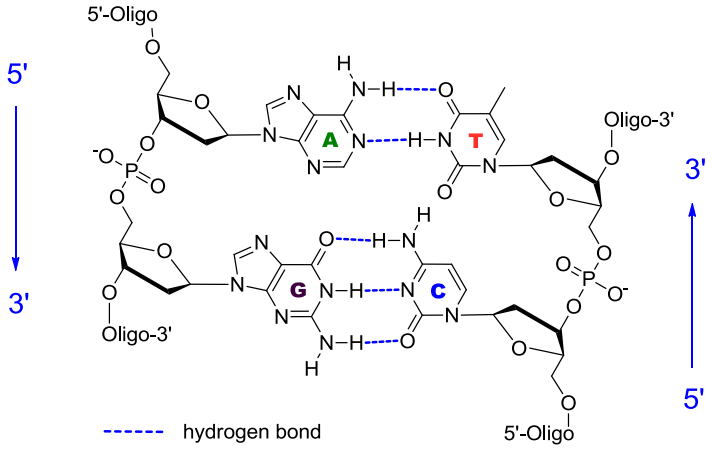


Figure 1.6: Watson-Crick DNA base pairing in a DNA duplex

In solutions containing salt, the negative charges of the phosphates of DNA are shielded by cations, so the two single strands can anneal without been repelled by their negative charges. The reverse process is called denaturing or melting of dsDNA and is often achieved by heating the solution. This process can be monitored experimentally by UV melting. This is because the bases in the DNA duplex (dsDNA) stack upon one another and consequently have lower UV absorption at 260 nm than the bases in single-stranded DNA (ssDNA). For a defined sequence, the temperature at which half of the double-strands are denatured in solution is termed the melting temperature (T_m). T_m is an indication of the DNA duplex stability and is determined by the length and sequence of the dsDNA.

1.2 DNA replication

DNA replication is catalysed by a cluster of enzymes. Among them the DNA polymerases (DNA pol.) catalyse the elongation of the new DNA chain (Figure 1.7). The first DNA polymerase the *E.coli* DNA polymerase I was discovered by Arthur Kornberg [7,8]. More recently, detailed studies showed that the elongation proceeds in 5′→3′ direction [7] and is templated (Figure 1.7). The basic mechanism of nucleic acid polymerization is universal to DNA polymerases and also to DNA-dependent RNA polymerases (discussed below). For DNA polymerases, a DNA or RNA primer is required as the enzyme cannot start polymerisation without it. The 3′-OH of the primer at the elongation site is essential as the polymerase catalyses its nucleophilic attack on the α-phosphate of the incoming deoxynucleoside triphosphate to form the new phosphodiester bond.

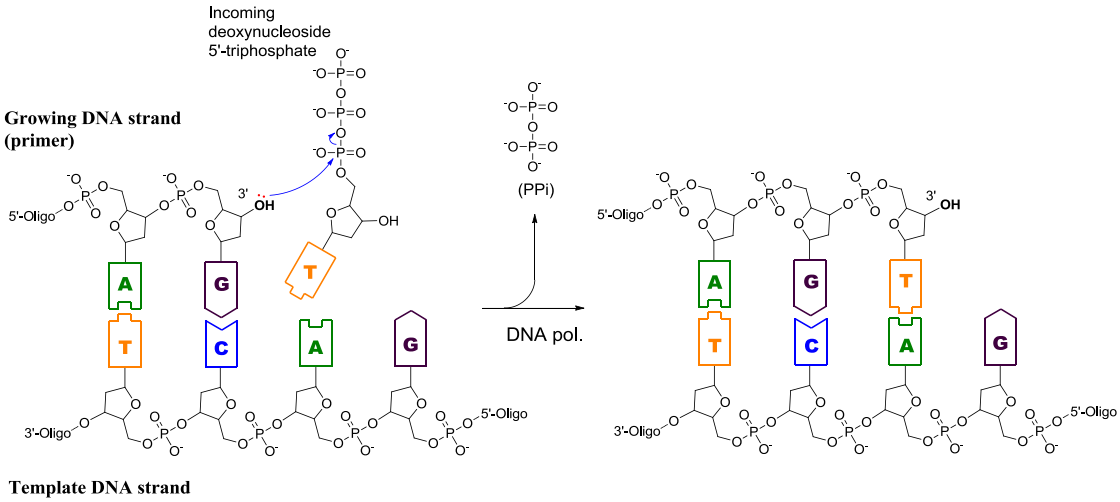


Figure 1.7: DNA polymerase catalysed templated DNA chain elongation

In *E.coli*, both strands of the DNA duplex are replicated in a semiconservative manner as the newly formed dsDNA contains one old ssDNA (Figure 1.8). The replication fork is extended as the DNA helicase separates the two strands of the duplex. The DNA topoisomerase releases the tension built up while the DNA is unwinding. The primer needed for DNA polymerase is synthesised by a DNA primase. As the DNA polymerase can only extend the chain in the 5′→3′ direction, one of the two new single strands has to be synthesised "backwards", initially in fragments (Okazaki fragment, Figure 1.8), and is named the lagging strand. After the removal of RNA primers by the 5′→3′ exonuclease activity of the DNA polymerase, the DNA fragments are ligated by a DNA ligase to form a continual lagging strand.

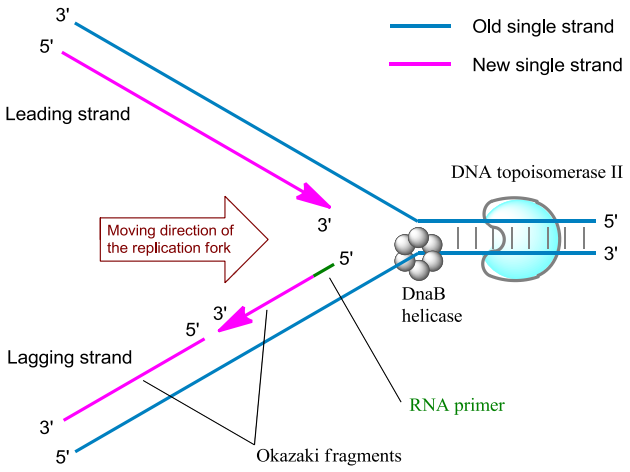
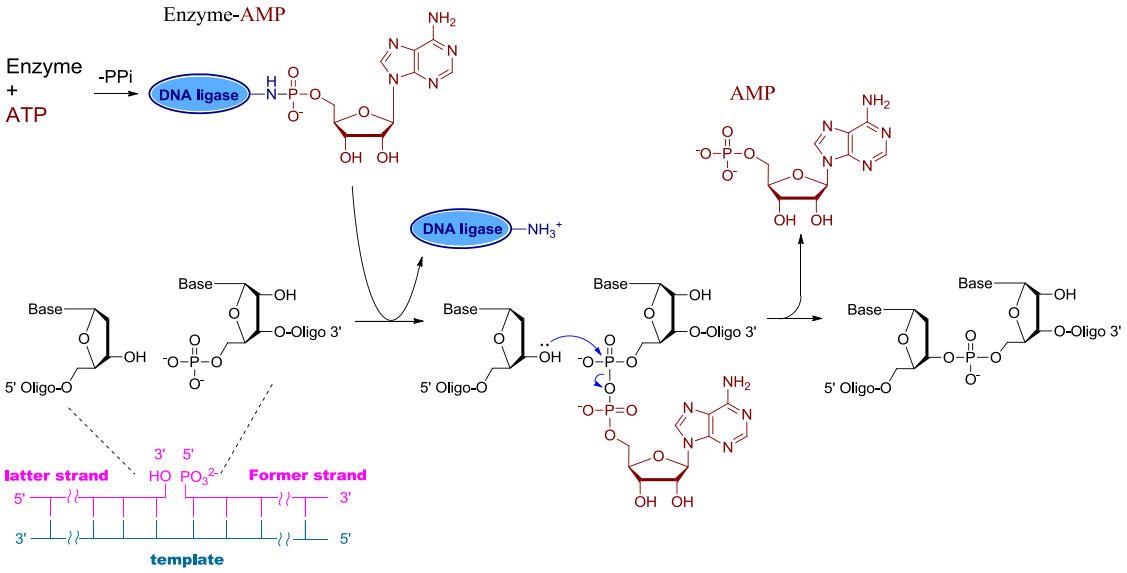


Figure 1.8: Leading strand and lagging strand at the DNA replication fork.



Scheme 1.1: Mechanism of DNA ligase catalysed strand ligation.

The mechanism of DNA ligation is illustrated in Scheme 1.1. DNA ligase activates the 5′-phosphate at the ligation site by transferring an adenine 5′-monophosphate (AMP) to it. The activated 5′-end is then coupled to the 3′-end of another strand with a 3′-OH group. Similarly, RNA ligases catalyse the ligation of two RNA strands or one RNA and one DNA strand. The RNA ligases are further discussed in Chapter 6.

A different kind of enzyme catalysing DNA polymerization is the terminal deoxynucleotidyl transferase (TdT) (NEB® #M0315, the cloned gene is from calf thymus). The primary function of mammalian TdT is adding non-templated nucleotides to the V, D and J exons during antibody gene recombination to increase their diversity [9]. Unlike DNA polymerases mentioned above, TdT is a template-independent DNA terminal transferase.

1.3 The flow of genetic information

With the finding of tRNA and the deciphering of the nucleotide code for 20 common amino acids in proteins, Francis Crick proposed the general flow of genetic information from DNA to RNA then to proteins and referred it as the central dogma of molecular biology in 1958 [10]. The process of copying sequence information from genomic DNA to RNA is called transcription, and process to transfer RNA sequence information to protein is called translation (Figure 1.9). In addition to this, the sequence of DNA and RNA can also be replicated in reactions catalysed by DNA polymerases and RNA-dependent RNA polymerases.

The central dogma was later amended according to the discovery that certain viruses such as the human immunodeficiency virus (HIV) can copy information from its RNA genome to DNA by a reverse transcriptase enzyme. Accordingly, this process is called reverse transcription. Although proteins can carry out DNA and RNA editing, their sequential information was never found to be copied back to RNA or DNA.

One big problem of the central dogma concept is that it suggests proteins to be the main mediators of biological function. However, recent studies have found that protein-coding genes cover only 2.94% of the genome [11], and the library of non-coding RNAs such as small RNAs that act as gene regulators is growing rapidly.

During transcription, the core enzymes catalysing RNA chain elongation are the DNA-dependent RNA polymerases (commonly referred as RNA polymerases but not abbreviated in this study). This enzyme utilises ribonucleoside triphosphates as the building block for the synthesis of new RNA chains following a mechanism similar to the DNA polymerization reaction discussed above. In addition, other proteins such as sigma factors in bacteria and transcription factors in eukaryotes are needed for the initiation of transcription [12].

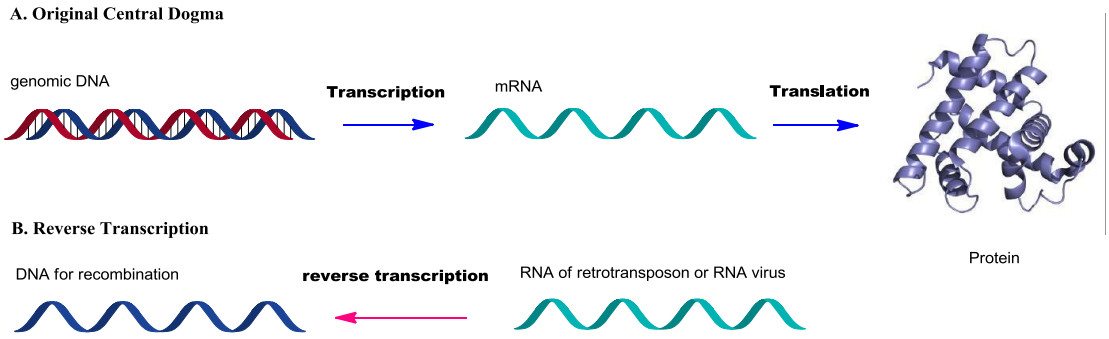


Figure 1.9: The flow of genetic information in the original central dogma and the reverse-flow during reverse transcription.

The RNAs initially synthesised by DNA-dependent RNA polymerases are called precursor mRNAs (pre-mRNAs). They are processed to mature mRNA post-transcriptionally through 5′-cap installation, poly-A tailing and splicing to remove the introns. The mature mRNAs (Figure 1.10) are then translated to proteins according to the coding sequences, while the untranslated regions (UTRs) often serve as post-transcriptional gene regulation sites [13]. The poly-A tailing process adds up to about 100 adenylate nucleotides [14] to the 3′-end of the

pre-miRNA and is catalysed by a RNA poly(A) polymerase (PAP). Like TdT for DNA, poly(A) polymerase is a template-independent ribonucleotidyl transferase [15]. Alternatively, certain RNAs such as histone mRNAs undergo polyuridylation, which is catalysed by a Poly(U) polymerase (PUP)[16].

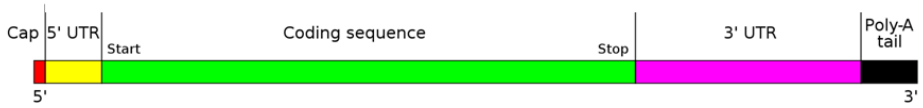


Figure 1.10: The structure of a typical human protein-coding mature mRNA including the untranslated regions (UTRs)

1.4 Polymerase chain reaction (PCR)

PCR has become a standard biotechnology tool to amplify a particular section of DNA using a DNA polymerase. The original protocol was developed by Kary Mullis [17]. The finding of a thermostable DNA polymerase greatly simplified the procedure [18]. A PCR program consists of several thermal cycles (Figure 1.11). During each cycle the dsDNA template is denatured, then annealed to primers (in excess). Each strand of the dsDNA then serves as the template for the elongation of the new synthesised strand. When the cycle is repeated, the template strands and newly synthesised strands in the last cycle serve as templates for the amplification in this cycle. By this method, the specific region of the template (usually genomic DNA) can be amplified exponentially. Also, the exponential amplification only occurs for the defined region between the two primers (orange section, Figure 1.11), so the product containing the sequence of this region only is enriched.

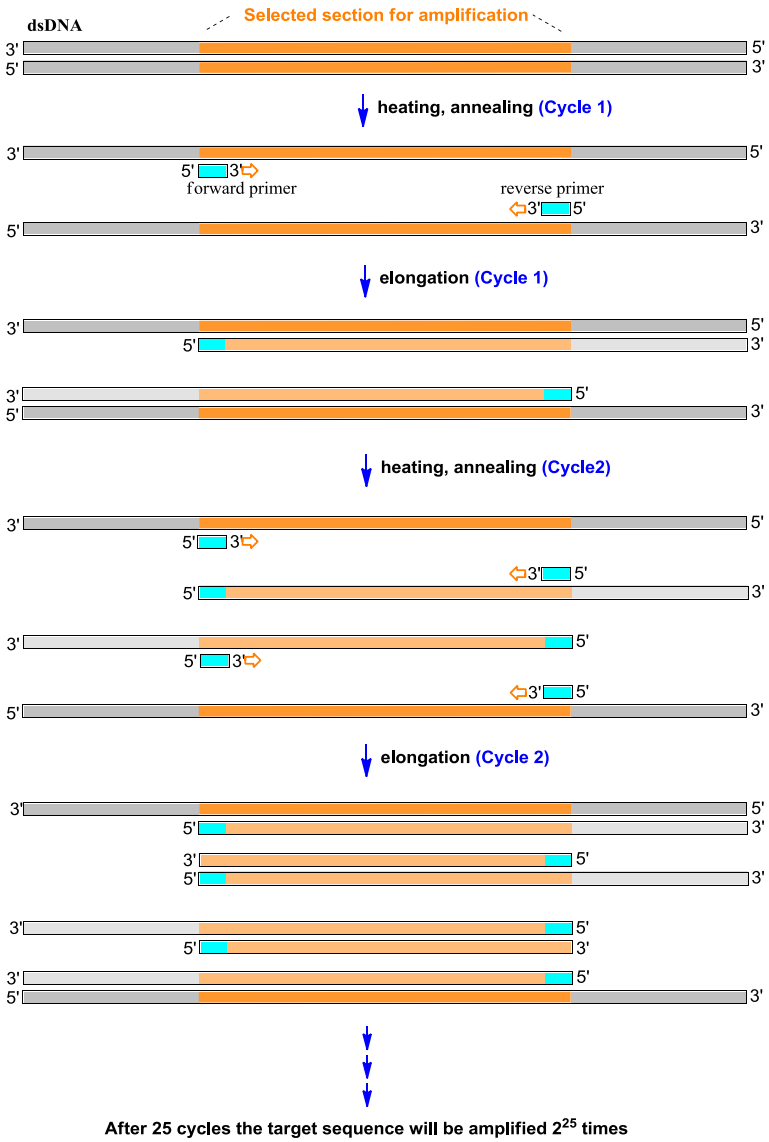


Figure 1.11: PCR amplification of a selected sequence in the genome.

1.5 DNA sequencing

The chain-termination sequencing method originally developed by Frederick Sanger was the most widely applied sequencing method before the commercialization of next generation sequencing, and was the main available technique used for the Human Genome Project [19,20]. The original Sanger sequencing method utilises four 2′,3′-dideoxynucleoside triphosphates (ddNTPs, lacking 3′-OH group, N = A, T, G or C) as PCR chain terminators in four parallel reactions (A, T, G, C). For each reaction, one ddNTP is mixed with normal four dNTPs during a linear extension reaction, resulting in random incorporation of a ddNTP and chain termination. This method was subsequently modified by using different dye-labelled primers in each of the four reactions. So the four PCR products corresponding to A, T, G, C with different colours can be mixed and analysed by a single run of capillary electrophoresis [21].

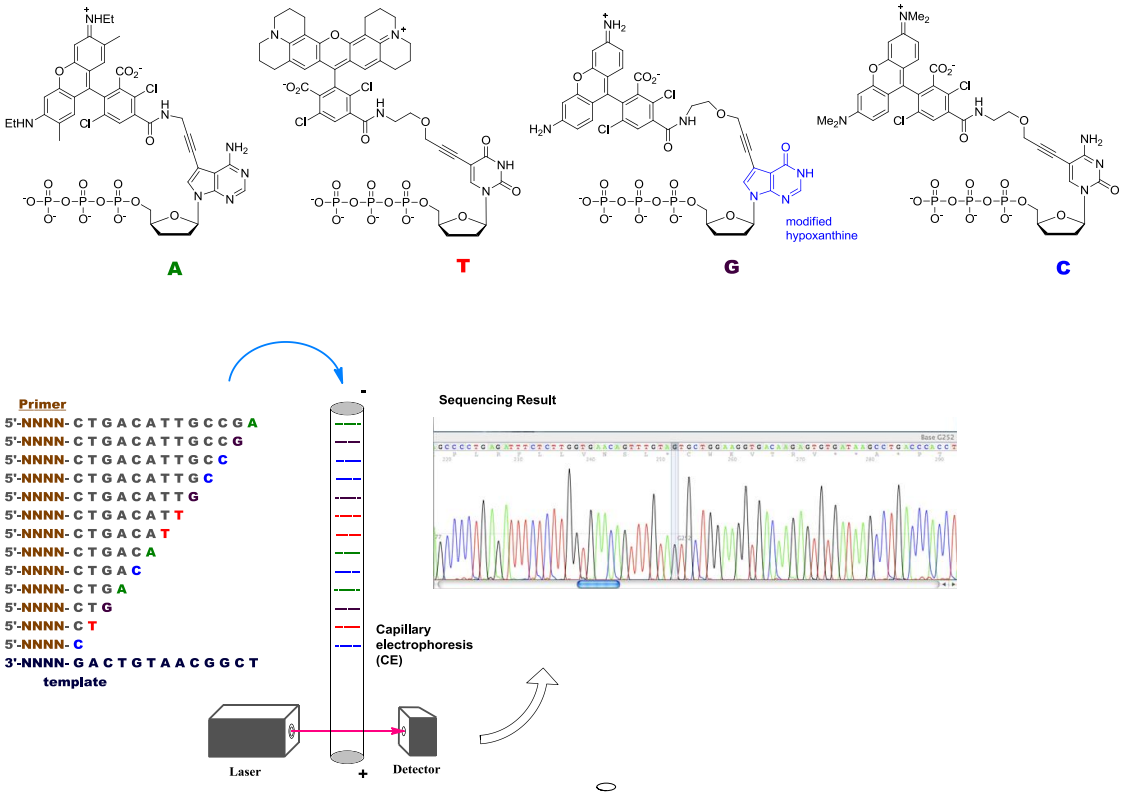


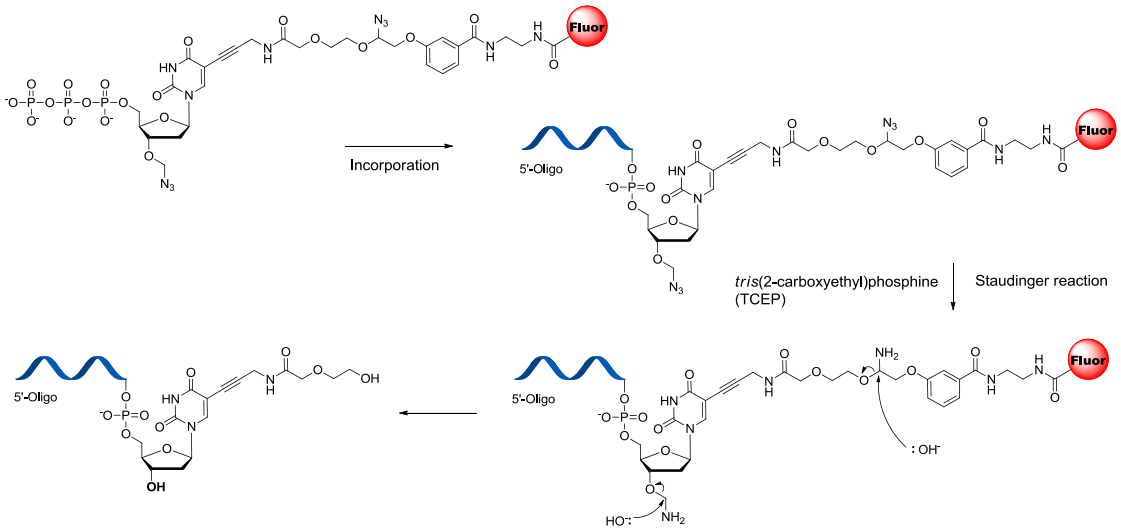
Figure 1.12: Four dye-labelled ddNTP chain terminators [22] and the modified Sanger sequencing workflow.

The development of fluorescent dye-terminator sequencing further advanced Sanger sequencing, in which the dyes were attached to the chain terminators instead of primers (Figure 1.12) [22]. It utilises four dideoxynucleoside triphosphates labelled with four fluorescent dyes with different colours corresponding to A, T, G and C. They are applied in a

single sequencing reaction (repeated primer extension by thermocycling) instead of four parallel reactions needed in original Sanger sequencing. The terminal sequences of extension products are determined by their colours. The development of engineered DNA polymerases further improved the incorporation of these modified nucleotide triphosphates [23].

The intrinsic limitation of Sanger sequencing is that each reaction needs a separated liquid chromatography or gel electrophoresis to resolve different bands. This makes the sequencing project lengthy and expensive for large genomes. For genomic sequencing, each section of the genome needs to be sub-cloned and PCR amplified to make enough DNA of the same sequence for one sequencing reaction. This process is time-consuming. The Next generation sequencing strategies were thus developed. Fundamentally, they are based on the Shotgun sequencing method, in which the whole sequence, longer than 1000 base pairs (Sanger sequencing limit [24]), is broken down into random fragments and sequenced individually. The whole sequence is then assembled computationally using the overlapping sections of different fragments.

Several next generation sequencing strategies including Illumina® dye sequencing [25] and Roche® 454 pyrosequencing have been developed. Both these methods mentioned above use the "sequencing by synthesis" principle. This does not require a chromatography procedure but involves direct monitoring of the triphosphate incorporation. The advantage of this approach is that multiple sequencing reactions can be performed and monitored simultaneously.



Scheme 1.2: The function of a reversible terminator 3´-*O*-azidomethyl-2´-deoxythymidine triphosphate (T) labelled with a removable fluorophore used in Illumina® dye sequencing [25].

The Illumina® dye sequencing uses four reversible terminators corresponding to A, T, G and C, each with a distinct and removable fluorophore and a removable 3´-blocker. An engineered 9 °N DNA polymerase was developed to ensure the efficient incorporation of these modified nucleotide triphosphates. (This reversible terminator sequencing method [26] and the engineered 9 °N DNA polymerase [26-28] were previously published by Jingyue Ju's research group in Columbia University.) After the detection of chain terminating nucleotides, both the fluorophore and 3´-blocker are removed by tris(2-carboxyethyl)phosphine (TCEP) treatment. This allows the incorporation of the next nucleotide to the 3´-end with a free 3´-OH group without the hindrance (or cumulative signal) of the fluorophore (Scheme 1.2, Figure 1.13).

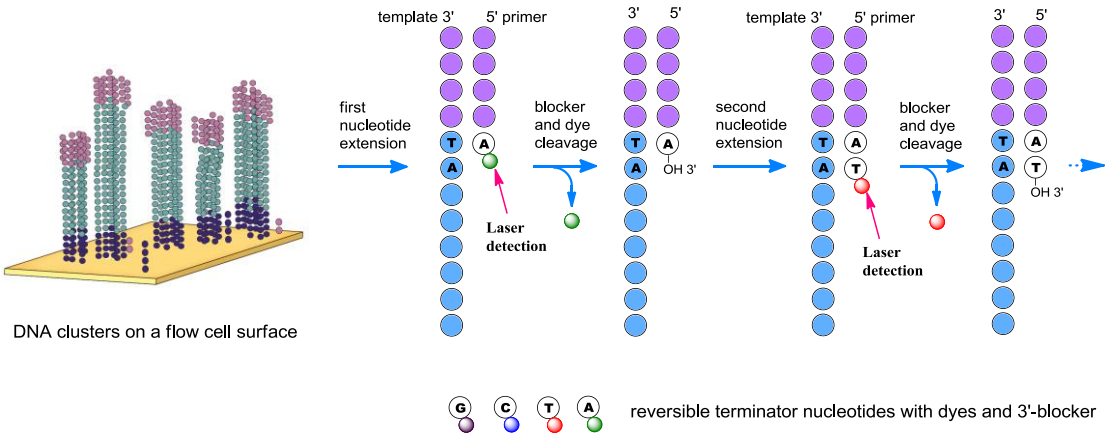


Figure 1.13: Illumina® dye sequencing of a DNA cluster in a flow cell. Reactions on different clusters are monitored simultaneously.

The pyrosequencing strategy differs from Illumina ® dye sequencing because it does not use modified nucleotide triphosphate terminator but detects the release of pyrophosphate (PPi) as a signal of nucleotide incorporation [29]. ATP sulfurylase converts PPi to ATP, which fuels luciferase mediated light emission. During each sequencing cycle, each one of the four dNTPs (including α-*S*-dATP) is added separately and single or multiple additions of the correct nucleotide are monitored by the amount of light that is generated.

Massively parallel automatic Illumina® sequencing is achieved because the sequencing cycles are performed simultaneous on different DNA clusters each with a single sequence. The colours of each of the dots (clusters) on the flow cell surface can then be monitored altogether. Similarly, the Roche® 454 pyrosequencing platform enables multiple sequencing reactions in different water droplets in an oil solution (emulsion) to be performed and monitored simultaneously.

One limitation of the next generation sequencing methods is that the length of each individual read is shorter than the chain-termination sequencing method. The Illumina® sequencing can read up to 150 nucleotides while the 454 pyrosequencing can achieve 300-500 nucleotides. Despite on-going improvements of these platforms, their read lengths are still significantly shorter than the chain termination sequencing method, which can read up to 1000 nucleotides [30-32]. This potentially makes the fragments hard to assemble if the section contains long repeating sequences.

The DNA sequencing methods mentioned above have also be applied in RNA sequencing. Due to the instability of RNA molecules, the RNAs needs to be reverse transcribed to DNAs using a reverse transcriptase before the actual sequencing. This process is part of the sequencing library preparation. The details are discussed in the RNA sequencing chapter below (Chapter 7).

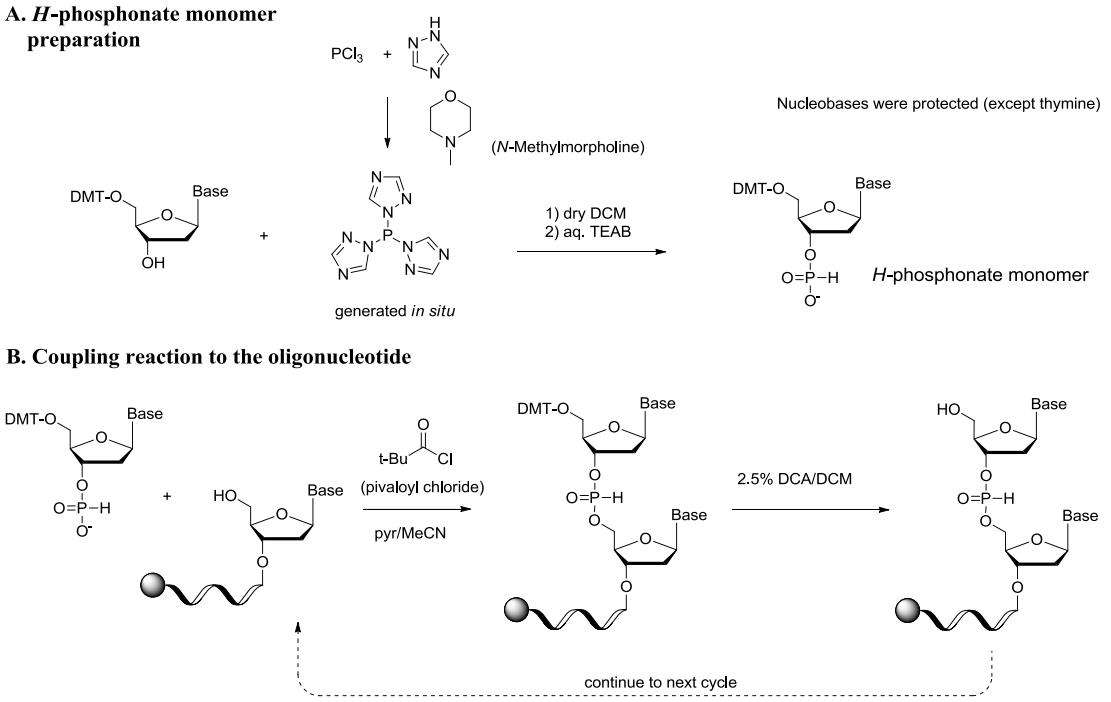
1.6 Solid-phase synthesis of oligonucleotides

Although the PCR reaction mentioned above is an efficient enzymatic method for synthesis of DNA of several kb in length, it requires a set of primers to initiate the elongation in each thermal cycle. These primers, typically 20 nucleotides in length, are made chemically using solid-phase oligonucleotide synthesis. Compared to PCR, the chemical synthesis can be performed on a much larger scale (μmole scale comparing to nmole scale for PCR). It also enables the introduction of various chemical modifications such as fluorophores.

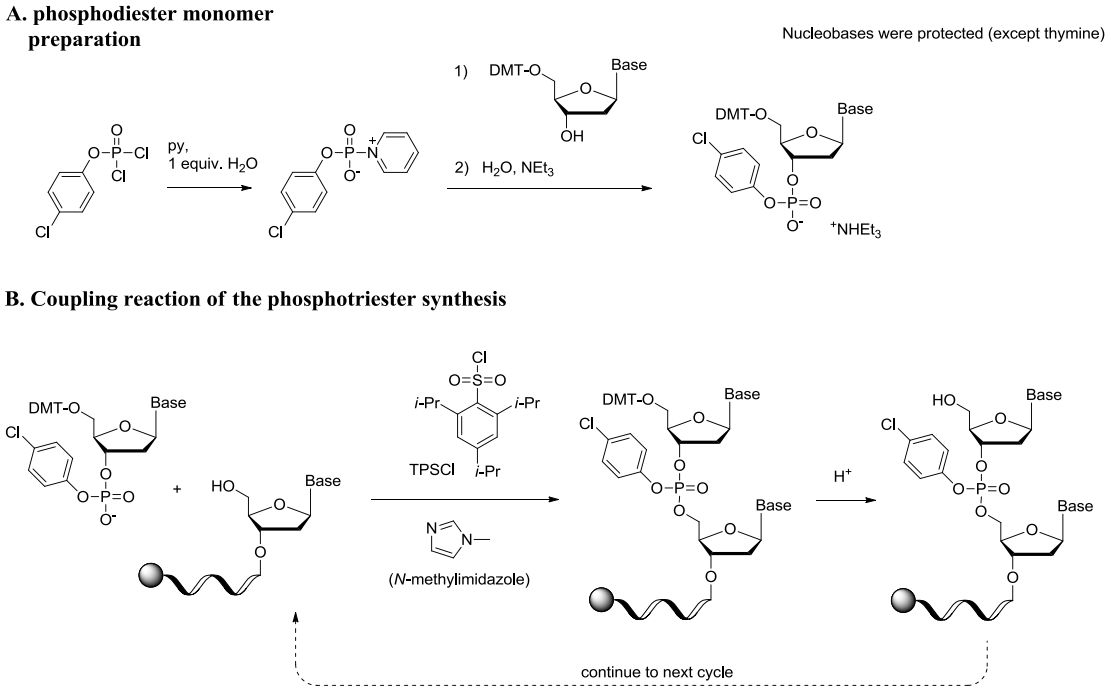
H-phosphonate synthesis was the first method adopted for oligonucleotide solid-phase assembly (Scheme 1.3) [33-35]. The *H*-phosphonate monomer is coupled to the 5′-OH of the extending chain using acid chlorides such as pivaloyl chloride as the coupling agent. After the complete assembly of the oligonucleotide chain, the *H*-phosphonates in the chain are then oxidised to phosphate by I_2. Although the oxidation state of *H*-phosphonate could be considered to be P^{III} (metaphosphate), it is regarded as P^{V} chemistry because the phosphorus has no electron lone pair.

Another early oligonucleotide synthesis method using P^{V} chemistry is phosphotriester synthesis (Scheme 1.4) [36-39]. Similar to *H*-phosphonate coupling, an acid chloride such as 2,4,6-triisopropylbenzenesulfonyl chloride (TPSCl) was commonly used as the coupling

agent. The chlorophenyl protecting group on the phosphate prevents multiple couplings on to the phosphate.

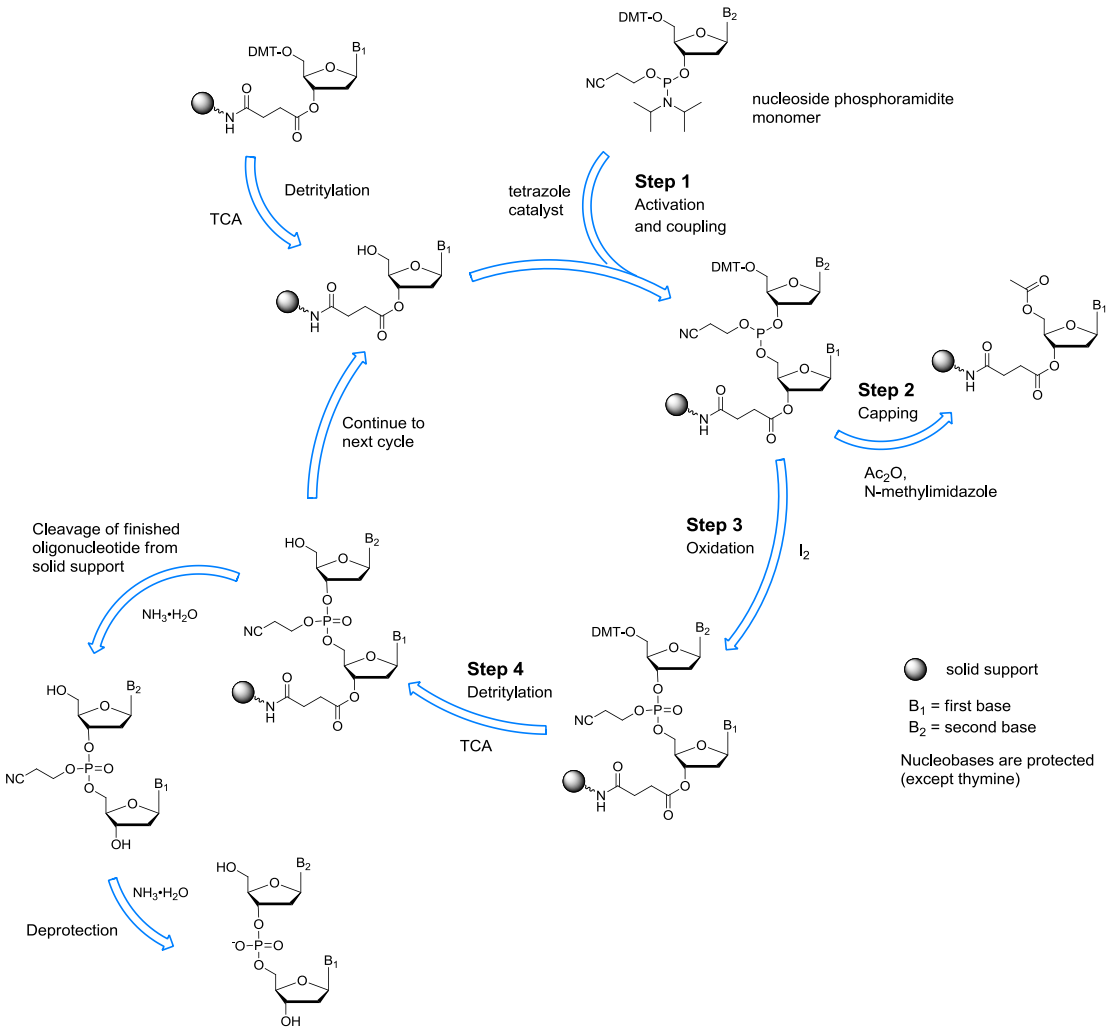


Scheme 1.3: *H*-phosphonate monomer preparation [33] and coupling during oligonucleotide solid-phase synthesis

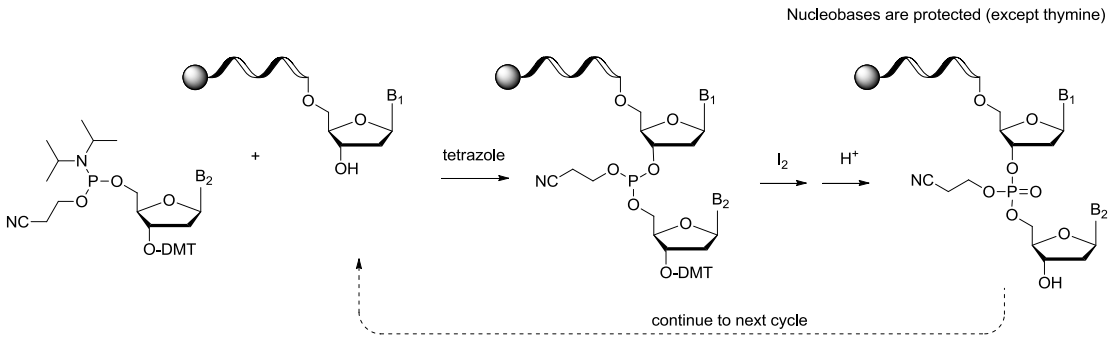


Scheme 1.4: Phosphodiester monomer preparation and phosphotriester oligonucleotide synthesis [38]

The phosphoramidite method has become the most reliable and widely applied method for oligonucleotide solid-phase synthesis. The monomer in this method is called a nucleoside phosphoramidite (Scheme 1.5). As the phosphoramidite is more reactive than *H*-phosphonate and phosphotriester, the coupling efficiency of this P^{III} chemistry is much higher than that of the above methods, which use P^{V} chemistry. The phosphoramidite DNA synthesis cycle is illustrated in Scheme 1.5. In step 1, the phosphoramidite monomer is activated by a weak acid (tetrazole or derivative), which protonates the diisopropylamino group. The activated monomer is then coupled to the free 5′-OH on the growing oligonucleotide chain. After coupling, any unreacted 5′-OH on the oligonucleotide chain is capped by an acetyl group during the capping step (step 2) to prevent the chain from further extension. The phosphite triester is then oxidised in step 3 to a phosphodiester linkage. After detritylation (step 4), the extended oligonucleotide chain with a new free 5′-OH is then ready to enter the next cycle.



Scheme 1.5: The phosphoramidite oligonucleotide synthesis cycle



Scheme 1.6: Reverse phosphoramidite oligonucleotide synthesis

The standard direction of phosphoramidite oligonucleotide synthesis is 3′→5′. This is because the 5′-OH as a primary alcohol is more reactive than the 3′-OH. However, reverse phosphoramidites are also commercially available (Link Technologies®) and their coupling is also efficient. They are utilised for 5′→3′ synthesis of oligonucleotide chains (Scheme 1.6) [40], and can also be used to synthesise DNA in which the direction of the chain is reversed internally, e.g. 5′→3′ followed by 3′→5′ or *vice versa*.

RNA can also be assembled using ribonucleoside phosphoramidites (Figure 1.14). The 2′-OH of ribonucleoside phosphoramidite is protected normally by a TBDMS protecting group. The coupling efficiency of ribonucleoside phosphoramidites is lower than that of the deoxyribonucleoside phosphoramidites, due to the steric hindrance caused by the additional TBDMS group, thus they require longer coupling times. After the cleavage and the base-deprotection of the oligonucleotide in aqueous ammonia/ethanol (3:1, v:v), the TBDMS protecting group can be removed by triethylamine trihydrofluoride.

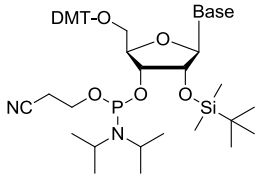
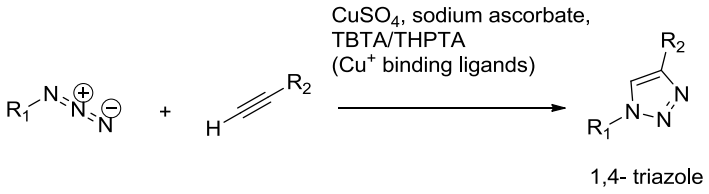


Figure 1.14: Ribonucleoside phosphoramidite with 2′-*O*-TBDMS protection. Nucleobases are protected (except uracil).

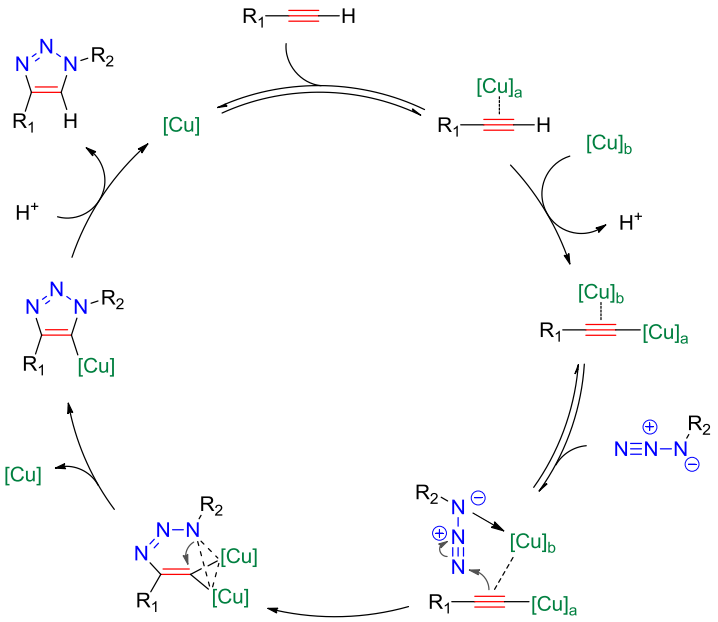
1.7 Copper-catalysed Alkyne-azide Cycloaddition reaction (CuAAC)

The copper-catalysed alkyne-azide cycloaddition (CuAAC) reaction (Scheme 1.7) is a variation of Huisgen's alkyne-azide cycloaddition (AAC) reaction [41]. It was discovered independently by M. Meldal [42] and K.B. Sharpless [43] in 2002. The normal uncatalysed AAC reaction generally requires elevated temperature to increase the reaction speed and it generates a mixture of 1,4-triazole and 1,5-triazole isomers. It was discovered that copper(I) catalysis greatly accelerates the AAC reaction, making it efficient at room temperature [43]. The CuAAC reaction is the most commonly used example of Click chemistry, which was defined by Sharpless as "modular, wide in scope, giving very high yields, generating only inoffensive by-products that can be removed by non-chromatographic methods and being stereospecific" [44]. The CuAAC reaction involves a terminal alkyne and an organic azide to form the 1,4-triazole product exclusively. Although azide is not the most active 1,3-dipole, the convenience of introduction and relatively inert nature of the azide group make it an excellent choice compared to others [43]. The reaction is marked by its high efficiency and regioselectivity. In a recent study of the CuAAC reaction mechanism, a dinuclear copper intermediate was proposed and evidenced (Scheme 1.8) [45].



Scheme 1.7: Copper-catalysed alkyne-azide cycloaddition (CuAAC) [42,43]

The copper(I) catalyst is commonly generated *in situ* from Cu^{2+} by a reductant such as sodium ascorbate. With the aid of the hydrophobic Cu^{I} binding ligand *tris*-(benzyltriazolylmethyl)-amine (TBTA) or the hydrophilic ligand *tris*-(hydroxypropyltriazolyl-methyl)amine (THPTA, **1**) [46], both of which further accelerate the reaction, it can be conveniently carried out in organic or aqueous solvents. In the nucleic acid field, the Cu^{I} binding ligands also help minimise DNA degradation caused by Cu^{I} [47].



Scheme 1.8: Proposed catalytic mechanism with two copper atoms in the recent literature [45].

One limitation of the CuAAC reaction in biology is the toxicity of copper [48]. This toxicity problem can be circumvented by the use of metal-free Click reaction systems such as the ring-strain promoted AAC (SPAAC) reaction. However, due to the complexity of synthesising the required strained alkyne reagents (normally cyclooctynes), the CuAAC reaction for *in vivo* labelling of biomolecules is still pursued. New Cu^{I} ligands such as BTTES (Figure 1.15) were found to minimise the toxicity of the copper catalyst. The water soluble BTTAA and BTTES ligands showed superior activity over THPTA thus allowing shorter Cu^{I} treatment times (3 min) thus minimising toxicity [49]. However, minor developmental defects of the zebrafish embryos after treatment with Cu^{I} and these new ligands have been reported [49]. A more recent article also suggests that L-histidine is an effective catalyst for the CuAAC reaction and minimises the toxic effects of copper *in vivo* [48].

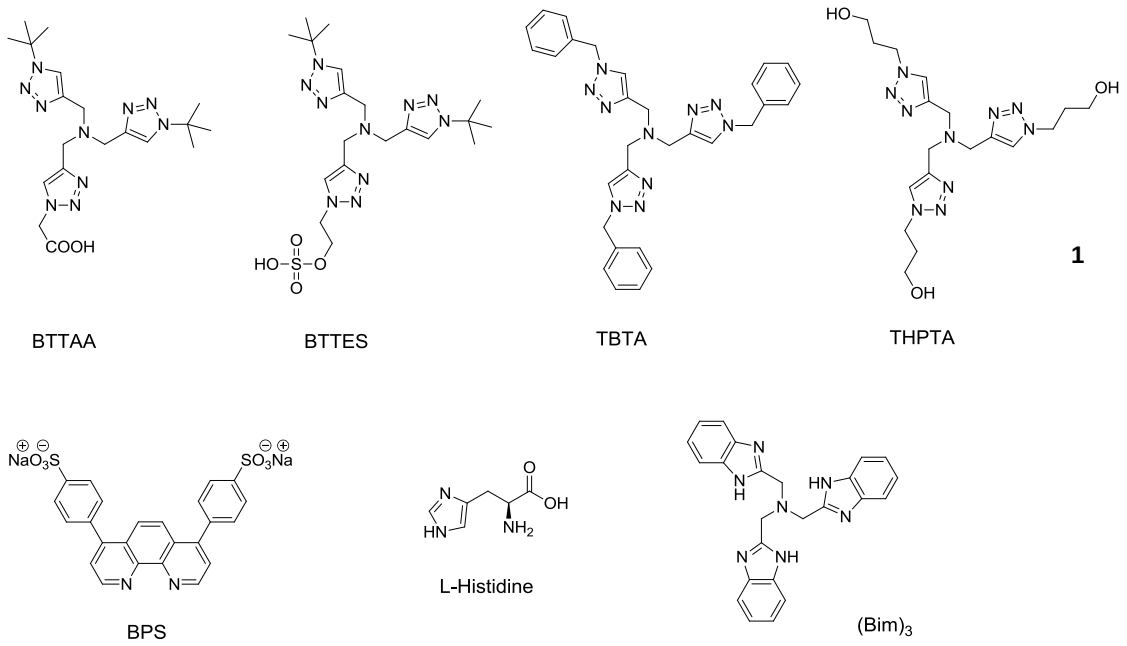


Figure 1.15: A selection of Cu^{I}-chelating ligands used in the CuAAC reaction [48].
BTTAA [49]: 2-[4-{(bis[(1-tert-butyl-1*H*-1,2,3-triazol-4-yl)methyl]amino)methyl}-1*H*-1,2,3-triazol-1-yl]acetic acid, BTTES [50]: 2-[4-{(bis[(1-tert-butyl-1*H*-1,2,3-triazol-4-yl)methyl]amino)-methyl}-1*H*-1,2,3-triazol-1-yl]ethyl hydrogen sulfate, TBTA: tris((1-benzyl-1*H*-1,2,3,-triazol-4-yl)methyl)amine, THPTA: tris[(1-hydroxypropyl-1*H*-1,2,3-triazol-4-yl)methyl]amine, BPS: bathophenanthroline disulfonate disodium salt, $(Bim)_3$ [51]: tris(2-benzimidazolyl-methyl)amine.

1.8 Strain-promoted Alkyne-azide Cycloaddition reaction (SPAAC)

Strain promoted Alkyne-Azide Cycloaddition (SPAAC) is the Huisgen 1,3-dipolar cycloaddition between a strained internal cyclic alkyne and an azide. It has been reported that the angular strain increases the dipolarophilicity of the alkene [52]. This phenomenon becomes more significant for cyclooctyne, the smallest stable cycloalkyne, exhibiting best activity among cycloalkynes towards phenyl azide. The reaction is likely driven by partially released ring-strain energy (~18 kcal/mol) [53]. SPAAC was first developed by Bertozzi's group as a substitute for the Staudinger ligation [54] during *in vivo* labelling of glycans, as the cyclooctyne possesses better bioorthogonality than the triaryl phosphine [55]. SPAAC is often referred to as a copper-free Click reaction. Unlike CuAAC, SPAAC retains the properties of the original AAC as it gives 1,2,3-triazole products as a mixture of regioisomers. It also differs from Staudinger ligation by generating no additional by-products. Examples of strained cycloalkyne systems [53] are listed in Figure 1.16.

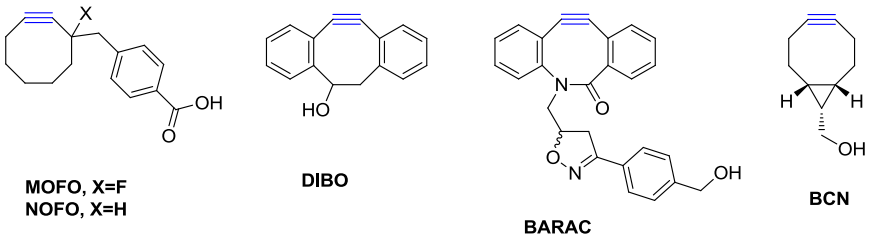


Figure 1.16: The family of strained cycloalkyne systems. MOFO, NOFO [56]; DIBO [57]; BARAC [58]; BCN [59].

Several strategies have been adopted to further improve these strained ring reagents. Conjugation of the dipolarophile with electron-withdrawing carboxylates can greatly enhance its activity [52,60]. However, this conjugation make the double/triple bond behave as a Michael addition site that is prone to react with nucleophiles like water. Alternatively, fluorine atoms have been attached to the α-position of the alkyne to withdraw electron-density through the σ-bond [56,61]. Other successful improvements were made by fusing the cyclooctyne ring with benzene rings. In 4-dibenzocyclooctynol (DIBO), the two benzene rings not only activate the alkyne through conjugation but also introduce double-bonds in the cyclooctyne ring to further increase the ring-strain. The recently developed BARAC contains an amide bond in the 8-membered ring to increase the strain further. Most recently, a simple bicyclo[6.1.0]nonyne (BCN) reagent was developed through a 4-step synthetic route [59]. It also possesses good kinetic activity as the fused three-member ring increases the strain of the

cyclooctyne. It has been utilised in our research group for copper-free DNA strand ligation [62].

1.9 Introduction of alkyne and azide groups into synthetic oligonucleotides

The efficiency and bioorthogonality of the CuAAC reaction makes it a convenient approach for labelling DNA with fluorescent dyes and other reporter groups [63], and for the construction of DNA, RNA or DNA/RNA chimeric structures *in vitro* [64]. The terminal alkyne group is generally stable to the I_2 oxidation step and other reagents encountered in solid-phase oligonucleotide synthesis. The inert nature of alkynes under these conditions enables them to be conveniently introduced into DNA as phosphoramidite monomers (Figure 1.17) and on modified solid supports used in oligonucleotide synthesis. Hence, a terminal alkyne can be located internally or terminally at any desired location in synthetic oligonucleotide chains.

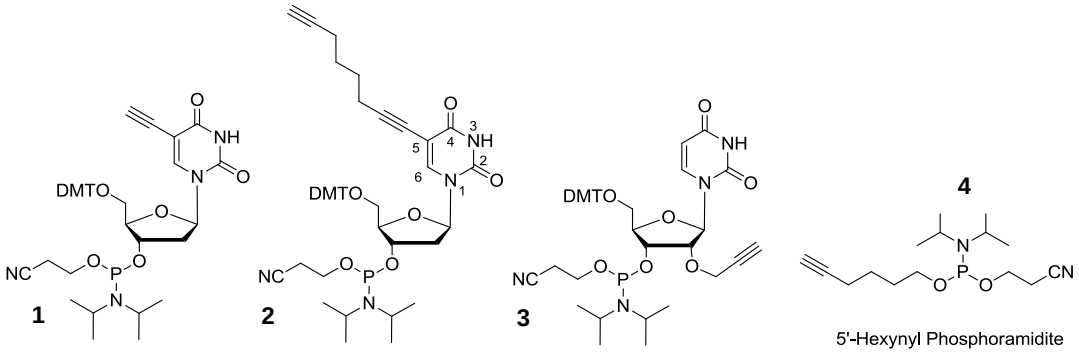


Figure 1.17: Alkyne-bearing phosphoramidites **1, 2, 3**: for DNA double strand crosslinking or labelling with azide modified reporter groups [65]; or for post-synthetic labelling [66]. **4**: 5′-alkyne linker.

Previously, the DIBO phosphoramidite (Figure 1.18) has been made by our research group and used for DNA strand ligation [67]. The BCN phosphoramidite is currently commercially available (Berry & Associates®) and utilised in this study. The Click reaction between a DIBO and an organic azide produces "*cis*" and "*trans*" regioisomers (Figure 1.18). Because the synthesised DIBO also has a chiral centre (marked by *) and is a mixture of enantiomers, each of the regioisomer produced by DNA Click ligation between DIBO and azide contains a pair of diastereoisomers. The commercially available BCN however is sterically pure, so the Click between BCN and organic azide produces a pair of enantiomers (Figure 1.18). In the chiral DNA chain, these enantiomers become diastereoisomers, however, because the chiral

centres of the DNA are far apart from the BCN, the HPLC and PAGE of BCN labelled oligonucleotides produce single peaks.

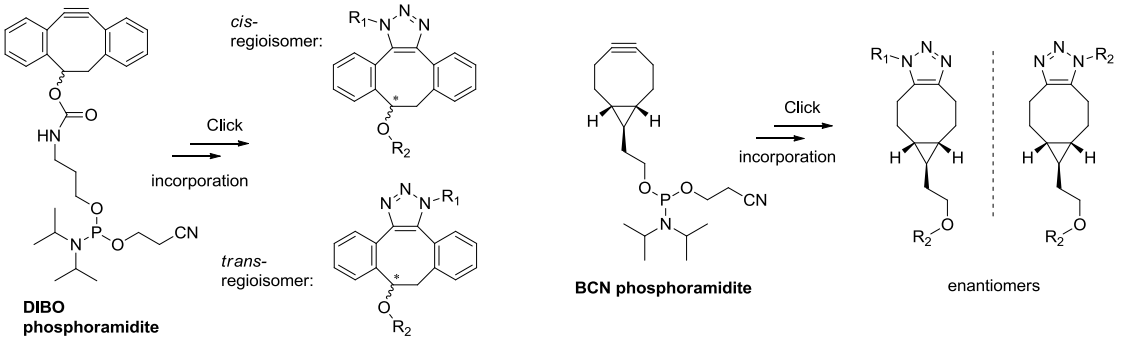


Figure 1.18: DIBO phosphoramidite [67], *cis* and *trans* regioisomer products and commercially available BCN phosphoramidite monomer (5′-Click-easy™ BCN CEP I, Berry & Associates®).

Introduction of the azide group is slightly more complicated as it is not compatible with P^{III} chemistry due to the Staudinger reaction of P^{III} with azides [68]. Hence, directly introducing an azide on to a phosphoramidite monomer causes severe problems. In order to use azide-bearing nucleotide monomers, P^{V} chemistry (phosphotriester DNA synthesis) can be used, which requires a phosphodiester monomer (**5**, Figure 1.19) and an activating agent (MSNT, **6**) to couple the monomer to the oligonucleotide (detailed in the next section) [69]. However, it is less efficient than phosphoramidite chemistry. More commonly, the azide is introduced after solid-phase oligonucleotide synthesis by labelling a pre-introduced primary amine residue with an azidoalkyl-carboxylic acid NHS ester (e.g. **7**) [70].

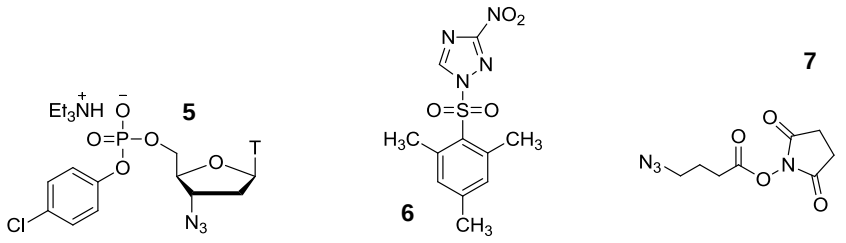


Figure 1.19: Thymidine phosphotriester monomer (**5**), 1-(2-mesitylenesulfonyl)-3-nitro-1*H*-1,2,4-triazole coupling agent (MSNT, **6**) [69] and azidobutyric acid NHS ester (**7**) [70] for azide incorporation into synthetic oligonucleotides.

In summary, both terminal alkyne and ring-strained alkyne may be readily introduced as phosphoramidites. Azide, however, cannot be introduced as a phosphoramidite due to P^{III}

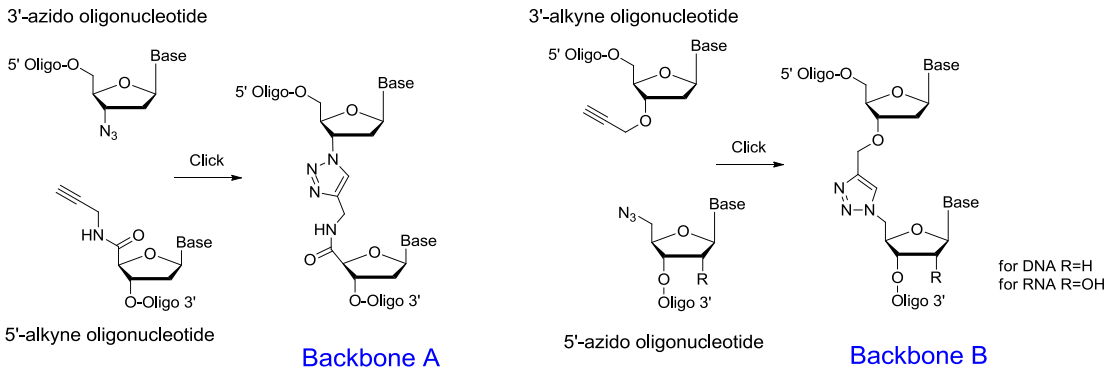
chemistry incompatibility, but can be introduced on to the oligonucleotide using P^V chemistry, or by post-synthetic labelling of a pre-installed terminal amine residue with an alkyne/azide bearing active ester. Azide groups can also be introduced at the 5′-end of the oligonucleotide chain post-synthetically by substitution of an iodo or bromo modification, which will be discussed in detail in the section below.

By these methods, DNA templated ligation was achieved using both CuAAC and SPAAC chemistry. However, one advantage of CuAAC over SPAAC is that it can generate a short sterically undemanding 1,4-triazole link rather than a bulky triazole fused to an 8-membered ring. The short linkages are more likely to be good mimics of the natural DNA/RNA phosphodiester backbone. The syntheses of these short linkages are described below.

1.10 Triazoles as phosphodiester backbone mimics

Short triazole linkages (triazole Backbone A and B, Figure 1.20) have great potential in biological applications. The triazole Backbone B in DNA was faithfully read through by various DNA polymerases such as GoTaq™, Taq and *Pfu* DNA polymerases, [69] while backbone A induced a single nucleotide deletion on read-through. An RNA ribozyme containing Backbone B maintained functionality and was able to cleave its substrate [64], indicating it is a good phosphodiester mimic. The related triazole RNA backbones and their bio-compatibility studies are summarized in the table below (Figure 1.20). Backbone A is made by "clicking" a 3′-azido oligonucleotide with a 5′-alkynyl oligonucleotide, while Backbone B is made by "clicking" a 3′-alkynyl oligonucleotide with a 5′-azido oligonucleotide.

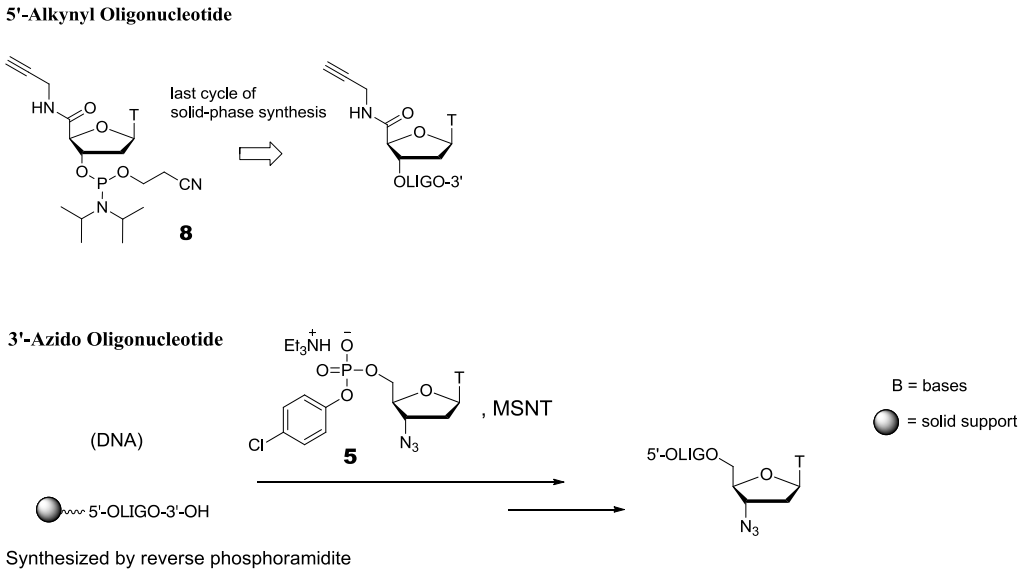
To synthesise triazole Backbone A, a 5′-alkyne was introduced on to DNA *via* an alkyne-bearing phosphoramidite such as 5′-propargylamido-dT (**8**, Scheme 1.9) in the final cycle of solid-phase synthesis [69]. In principle, this method should also be suitable for RNA. As mentioned above, the 3′-azido and 2′-azido nucleotides cannot be made as phosphoramidite monomers. 3′-Azido DNA was previously made in our research group using a 5′-phosphodiester monomer (4-chlorophenyl phosphodiester derivative, **5**, Scheme 1.9) and phosphotriester coupling (P^V chemistry) [69]. The oligonucleotide chain for this coupling was synthesised in the 5′→3′ direction using reverse phosphoramidites as discussed above [69], which allows 5′-phosphodiester monomer (**5**) to be added to the 3′-end of the chain.



Enzymes	Backbone A	Backbone B
DNA pol.	-1 [69]	✓ [71]
DNA dependent RNA pol.	?	✓ [72]
Reverse transcriptase	?	? (under evaluation)

"-1": one nucleotide omission. "✓" = correct product,
"?" have not been studied.

Figure 1.20: The biocompatibility of different triazole linkages as oligonucleotide backbone mimics.



Scheme 1.9: Synthesis of a 5′-alkynyl oligonucleotide and a 3′-azido oligonucleotide for use in the formation of triazole Backbone A in DNA [69].

Similarly, 2′-azido RNAs were made by using 2-chlorophenyl-3′-phosphodiester derivatives of 2′-azido-2′-deoxynucleotides (**9**) [73,74]. For 3′-phosphodiester (**9**), the RNA chain was synthesised in the 3′→5′ direction using standard phosphoramidites.

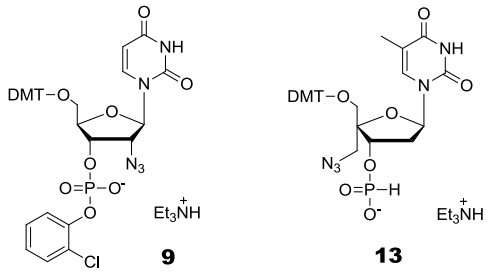
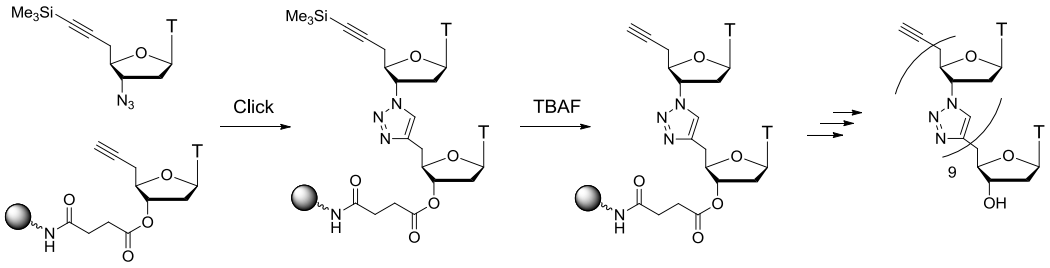


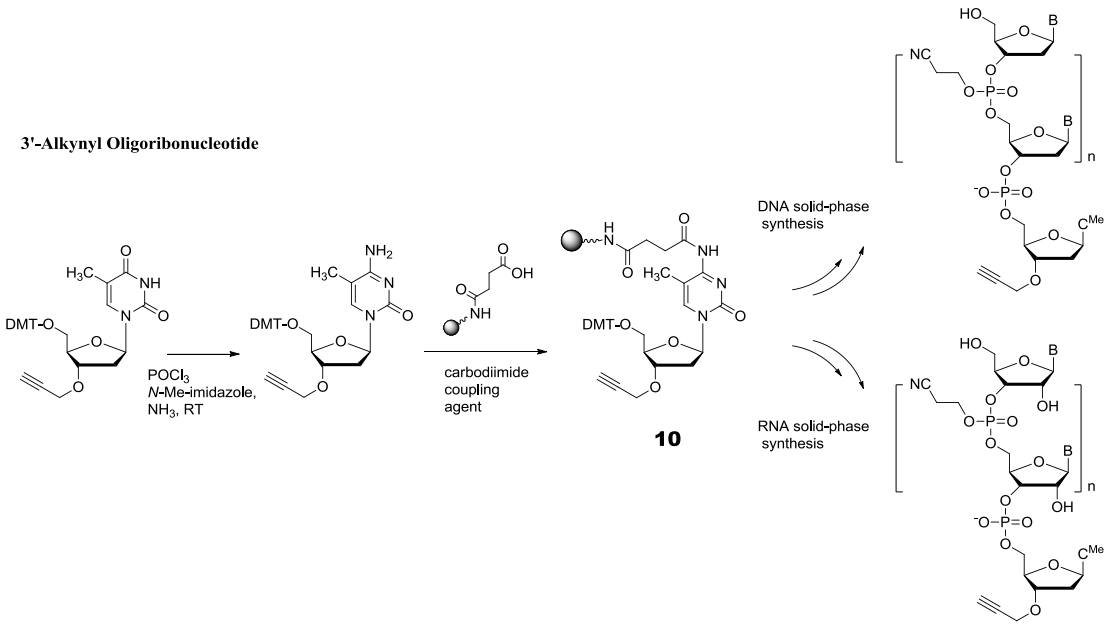
Figure 1.21: 2′-Azido-2′-deoxyuridine-3′-phosphodiester monomer (**9**) and 4′-*C*-azidomethyl-thymidine-3′-*H*-phosphonate (**13**) for solid-phase azide introduction

Additionally, trimethylsilyl protection of the 5′-alkyne has been reported to block the Click reaction. By subsequent removal of the trimethylsilyl group and Click coupling, triazole linked poly-T DNA analogue was synthesised (Scheme 1.10). This analogue is reported to form a stable duplex with a normal ssDNA [75].



Scheme 1.10: Synthesis of triazole-linked analogue of DNA (^{TL}DNA) [75]

The triazole Backbone A caused a single nucleotide deletion when the DNA polymerase read-through [69]. To improve the biocompatibility of triazole backbones, Backbone B was made by "clicking" a 3′-alkynyl oligonucleotide with a 5′-azido oligonucleotide. 3′-Alkynyl DNA [71] and RNA [64,76] (with a 2′-deoxyribonucleotide as the last nucleotide at the 3′-end of the RNA [64]) are made from alkyne-bearing nucleosides on solid support (**10**, Scheme 1.11). The commercially available 2′-*O*-propargyl and 3′-*O*-propargyl nucleoside resins (A, C, G, U, with appropriate nucleobase protection groups) (e.g. **11, 12**, Figure 1.22) further increase the capacity for the synthesis of 3′-alkynyl RNA.



Scheme 1.11: Synthesis of 3′-alkynyl oligonucleotides for use in the formation of triazole Backbone B in DNA [71] and RNA [64,76]

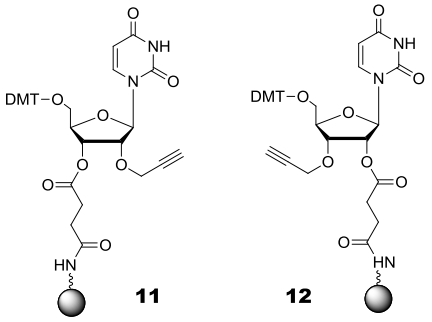
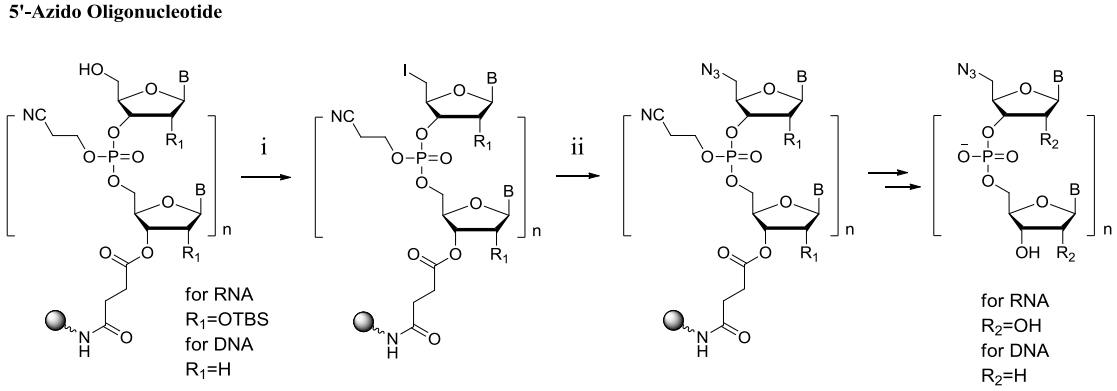


Figure 1.22: Commercially available 2′-*O*-propargyluridine 3′-LCAA CPG resin (**11**) and 3′-*O*-propargyluridine 2′-LCAA CPG resin (**12**).

To make Backbone B, the azide was introduced on to the 5′-position of the ribose sugar after solid-phase chain assembly using the method of Kool [77,78]. This proceeds *via* the 5′-iodo DNA and subsequent displacement of iodo by azide (Scheme 1.12). It has been used in our laboratory to introduce 5′-azide into DNA [71] and RNA [64] for various applications. 5′-Iodo-dT was also directly introduced into the oligonucleotide as the last phosphoramidite monomer during solid-phase synthesis [62], as an alternative to 5′-OH to 5′-I conversion on the oligonucleotide. Since halogens are stable to the oligonucleotide synthesis cycle and Click reaction condition, this method was extended by using a 5′-iodo-dT phosphoramidite to introduce iodine on to the 5′-end of the oligonucleotides with an alkyne at the 3′-end, and subsequently transforming the iodine to azide after "clicking" the 3′-end to another oligonucleotide. The newly converted azide was then used for another round of Click ligation [62].



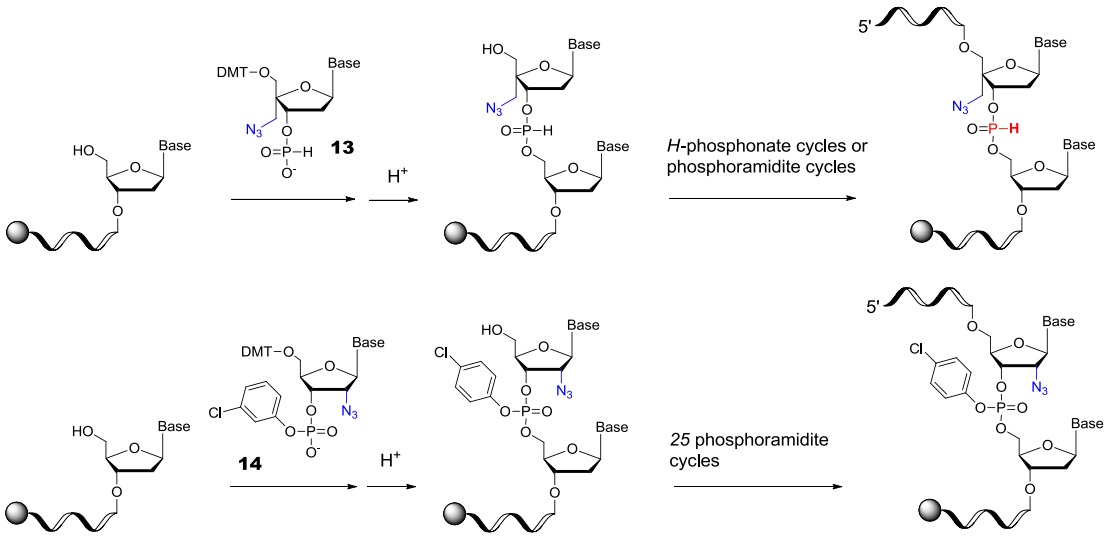
Scheme 1.12: Chemical methods for 5′-azide introduction for use in the formation of a short triazole linker in DNA: i) $(PhO)_3P^+CH_3I^-$/DMF, r.t.; ii) NaN_3/DMF, r.t.

A slightly different strategy for 5′-OH to 5′-azide conversion was used to synthesise a completely triazole linked DNA oligo-T analogue. In this solution-phase work, the 5′-OH of the triazole oligomer was mesylated by mesyl chloride (CH_3SO_2Cl) followed by nucleophilic substitution to produce the 5′-azide [79]. The 3′-azide and 2′-azide at the 3′-end of oligonucleotides cannot be introduced by this strategy as both 3′-OH and 2′-OH groups are more hindered and when activated, elimination competes with nucleophilic substitution.

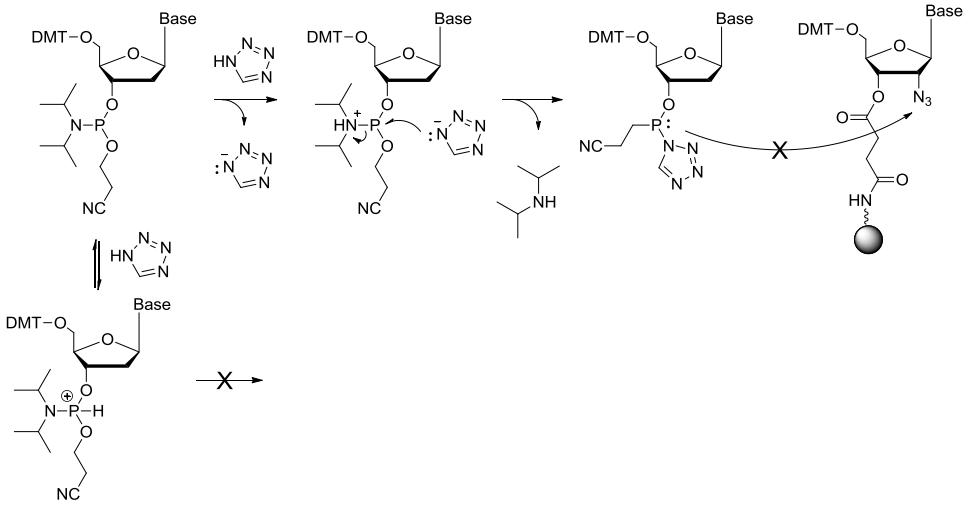
The azide on nucleotides was reported to be compatible with phosphoramidite chain assembly of a 15-mer modified DNA [80]. In this study, the azido nucleotides were introduced as *H*-phosphonates (**13**, e.g.) in the middle of the chain and the unmodified nucleotides were introduced as *H*-phosphonates or phosphoramidites (Scheme 1.13). A potential problem for this approach is that the *H*-phosphonates after the azide introduction (Scheme 1.13, red) can be oxidized during I_2 oxidation step in the subsequent phosphoramidite cycles. The free OH of the phosphate could then disrupt the next phosphoramidite coupling. This problem is not reported in the publication. In other studies, 2′-azide was introduced in the middle of the RNA oligonucleotide chain by using 2-chlorophenyl-3′-phosphodiesters (**14**) [73,74]. After the introduction of 2′-azido-2′-deoxynucleotide, up to 25 phosphoramidite cycles were performed and the 2′-azide was reported to be intact.

Introduction of an azide group directly on to the solid-support before chain-assembly was also shown to be compatible with phosphoramidite DNA chain assembly [81]. This provides another method for introducing azide at the oligonucleotide 3′-end. The authors suggested this compatibility is a result of low concentrations of phosphoramidites during solid-phase

chain assembly [74]. However, the compatibility is more likely to be because the phosphoramidite was partially converted to the tetrazole derivative or protonated by tetrazole (Scheme 1.14), so that the electron lone pair on P^{III} is not available for reaction with the azide functionality.



Scheme 1.13: Two approaches for introducing azide in the middle of the oligonucleotide chain



Scheme 1.14: Possible activated phosphoramidite intermediate that is inert to reaction with an azide group

1.11 Comparison of Click ligation and enzymatic ligation

Introduction of alkyne and azide groups by chemical means during oligonucleotide synthesis is generally robust. However, the length of the modified oligonucleotide is limited by the constraints of solid-phase synthesis. This restricts the potential of this chemical ligation system. Long DNA and RNA strands are normally assembled enzymatically. *In vitro* transcription is used to synthesise long RNA strands and the polymerase chain reaction (PCR) is a routine method to generate up to 20 kb lengths of dsDNA. As chemical methods of ligation are not readily applicable to naturally occurring nucleic acids such as mRNA, miRNA and siRNA, methods to introduce alkynes and azides at their termini are highly desirable. The alkyne or azide tags also allow subsequent Click labelling with fluorophores and other reporter groups [82].

These tags for chemical ligation will achieve similar goals as those conducted using enzymatic ligation. Enzymatic ligation is well-studied and widely applied; it has been carried out by T4 DNA ligase, T4 RNA ligase 1 (T4 Rnl1) and T4 RNA ligase 2 (T4 Rnl2 truncated) for sub-cloning and cDNA library construction [83]. In one study, T4 DNA ligase, T4 RNA ligase 1 and T4 RNA ligase 2 all produced more than 50% templated ligation of a RNA-3′-OH to a DNA-5′-phosphate at 37 ℃ in 20 min using a DNA template [84]. T4 RNA ligase 1 and 2 can also work in a non-templated fashion. In addition, T4 RNA ligase ligation of the 3′-hydroxyl group of 2′-azide-modified RNA to a 5′-phosphorylated downstream preadenylated DNA without templation proceeded in 80% yield while ligation of the same RNA 3′-end to a downstream RNA with the help of a DNA splint proceeded in the same yield [82]. These studies indicate the one-step enzymatic ligation is moderately efficient. To be undertaken at the same cost, the enzymatic methods typically need to be conducted at a much smaller scale compared to the chemical methods. However, if enzymes are used to introduce alkyne or azide into DNA for subsequent Click ligation, the amount of the final ligation product will be reduced to the same level as that obtained by DNA/RNA ligases. However, Click ligation is applicable to tasks that not only require high efficiency but also are difficult for ligases, rendering it potentially useful.

Although the Click ligation approach discussed below contains one enzymatic step and one additional chemical step (CuAAC), it provides several unique features over enzymatic ligation. Firstly, the separate enzymatic labelling and Click ligation steps can be conducted under different conditions and at different times, creating flexibility in the protocol. Secondly, and most importantly, the enzymes utilised for alkyne/azide labelling, such as DNA/RNA

polymerases, are different from ligases, so potentially could also solve problems caused by T4 RNA ligase. The problems of enzymatic ligation and the advantage of this two-step "label-then-Click" approach are further discussed in the RNA sequencing chapter (section 7.5).

1.12 Enzymatic oligonucleotide 3′-labelling for Click ligation

1.12.1 DNA 3′-labelling using modified nucleoside triphosphates

DNA is an excellent target for 3′-labelling for several reasons. First of all, DNA of several kb in length is conveniently produced by PCR. The 5′-alkyne or azide can be easily introduced *via* the PCR primer [85]. By combining this with 3′-alkyne or azide modified nucleoside triphosphates, both 5′ and 3′-labelling can be achieved on a single PCR strand. Secondly, there are several examples of modified deoxyribonucleoside triphosphate DNA chain terminators that are known to be substrates for DNA polymerases. The 3′-azido-3′-deoxythymidine triphosphate (AzdTTP, **15**, Figure 1.9) is a well-known chain-terminating inhibitor for certain DNA polymerases and reverse-transcriptases. Also, more than 10 different DNA polymerases from different species, and some in their mutated forms, are commercially available [86]. This increases the chances of finding a suitable enzyme from commercial sources for any modified nucleoside triphosphate. In addition, DNA is easier to synthesise chemically and to handle than RNA.

Enzymes that are used to synthesise DNA include various DNA polymerases from different species, reverse-transcriptases and terminal deoxynucleotidyl transferase, which uniquely works on single stranded DNA. To introduce alkyne or azide at the DNA 3′-end, *Taq* DNA polymerase is a potential candidate. In most cases, it adds a single 2′-deoxyadenosine to each of the two 3′-ends during PCR to form 3′-overhangs, and it lacks 3′ to 5′ proofreading activity that could excise the 3′-overhang [87]. *Taq* polymerase can even be used on DNA amplified by a proofreading polymerase after PCR, without changing buffer, to add a required 3′-dA overhanging. Generating double-stranded DNA (dsDNA) with propargyl-modified 3′-sticky ends potentially could be used to modify the traditional TA cloning procedure (discussed in Chapter 4).

The terminal deoxynucleotidyl transferase (TdT) could be a better candidate for alkyne/azide labelling because it is a template-independent polymerase. TdT is used in TA cloning, rapid amplification of cDNA ends (in RACE, TdT is used for homopolymeric tailing during 5′-

RACE [88]) and terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) [89]. It belongs to polymerase family X, which also includes eukaryotic polymerase β, polymerase σ, polymerase λ and polymerase μ [90]. TdT can add non-templated nucleotides to the 3′-end of the DNA chain in a separate step from PCR, so it enables more controlled manipulation. When using 3′-modified dNTPs such as 3′-propargyl and 3′-azido dNTPs, multiple-addition is prevented because the 3′-modification blocks further extension.

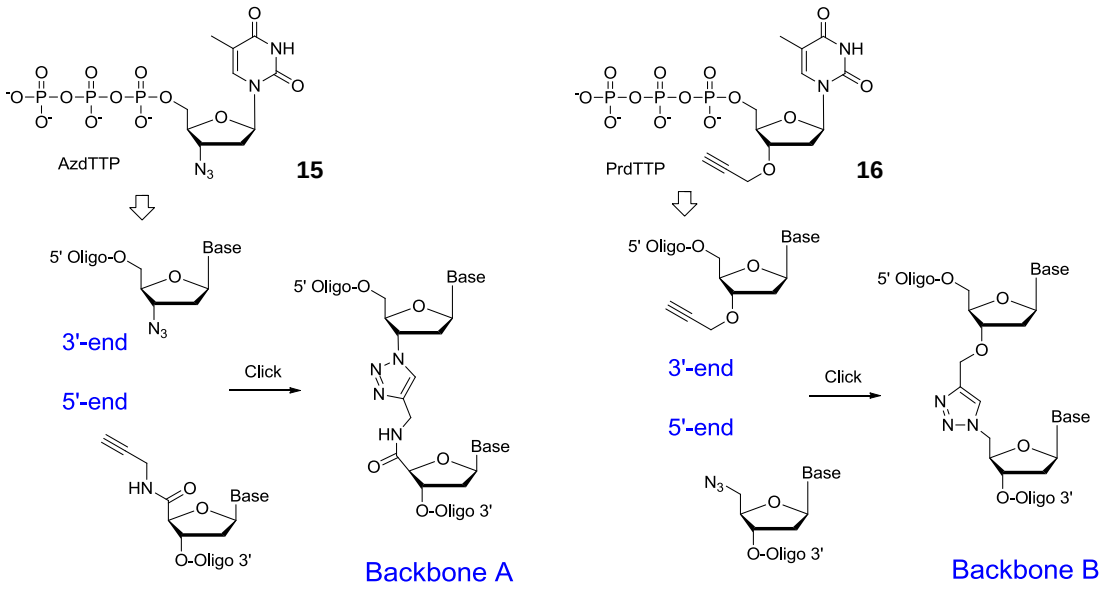


Figure 1.23: Two 3′-ends required for the formation of triazole backbones A and B on ssDNA formed by CuAAC reaction and the corresponding modified nucleoside triphosphate: 3′-azido-3′-deoxythymidine triphosphate (AzdTTP, **15**) and 3′-*O*-propargyl-2′-deoxythymidine triphosphate (**16**, PrdTTP).

To generate the 3′-ends for construction of DNA triazole Backbone A and B *via* click DNA ligation (Figure 1.23), a 3′-azido-2′,3′-dideoxynucleoside triphosphate and a 3′-*O*-propargyl-2′-deoxynucleoside triphosphate are required, and the corresponding nucleotide needs to be incorporated efficiently. The 3′-azido-3′-deoxythymidine triphosphate (**15**, AzdTTP) is commercially available. The 3′-*O*-propargyl-2′-deoxythymidine triphosphate (**16**, PrdTTP), however, must be synthesised.

The nucleoside triphosphates with alkyne or azide modification at the 2′or 3′-positions are much less studied compared to the ones with alkyne or azide modification on the nucleobases. Nucleoside triphosphates with modifications on pyrimidine C5 and deazapurine C7 on the nucleobases (Figure 1.24) are generally accepted by DNA/RNA polymerases [91,92].

However, for the 2′or 3′-modifications, although the terminal alkyne and azide groups are relatively small, they are still significantly larger than other common modifications on the sugar ring such as 3′-H, 2′-F, 2′-NH_2 and 2′-OMe. In the literature, certain 3′-modifications have been reported to prevent an NTP from being incorporated [93]. For example, a nucleoside triphosphate with an *O*-acyl protection of the 3′-OH (Figure 1.25) was not incorporated by T7 Sequenase v2.0, Klenow (exo$^-$) or AmpliTaq® DNA polymerase [93]. This indicates that DNA polymerases are sensitive to modifications at the 2′ and 3′-positions, and finding the enzymes that accept 3′-modifications could be problematic and not entirely predictable.

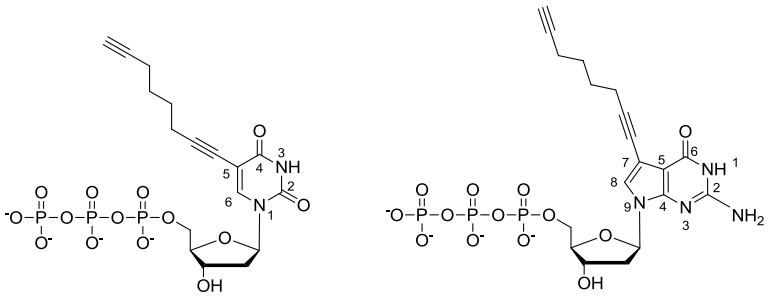


Figure 1.24: Examples of modified triphosphates with alkyne side-chains on the nucleobase that were incorporated by DNA polymerases [91].

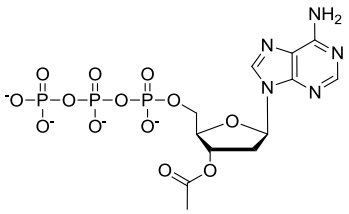


Figure 1.25: Examples of a modified triphosphates that was not incorporated by certain polymerases [93].

However, although the polymerase compatibility of 3′-*O*-propargyl deoxynucleoside triphosphates has not been reported, the structurally similar 3′-*O*-allyl-thymidine triphosphate is known to be incorporated by VentR(exo$^-$) DNA polymerase with low efficiency [93]. More recently, the 3′-*O*-allyl-thymidine triphosphate (Figure 1.26) was reported [26-28] to be efficiently utilised by a 9°N polymerase (exo$^-$) with mutations A485L and Y409V after screening several potential enzymes including Therminator™, Thermo Sequenase, Vent (exo$^-$), Deep Vent (exo$^-$), Tth, Tfl, Bst, Pfu (exo$^-$), Klenow (exo$^-$) fragment and Sequenase DNA polymerases, AMV, RAV2, M-MuLV, HIV reverse transcriptases, and this 9°N polymerase (exo$^-$, A485L/Y409V) [27]. In this series of studies, the 3′-*O*-allyl group served as a cleavable blocker for PCR capping during sequencing on chip. The gel-electrophoresis results were not included in the paper and only mass spectrometry data were shown. Hence, it is uncertain

how efficient the 3′-*O*-allyl-thymidine triphosphate incorporation was. However, as this triphosphate is reported for use in sequencing, the incorporation is likely to be efficient.

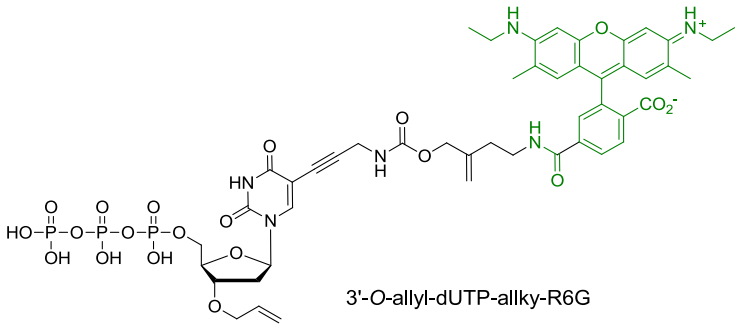


Figure 1.26: One of the 3′-*O*-allyl-deoxynucleoside triphosphates [27,28] efficiently utilised by a mutant 9°N$_m$ polymerase (exo$^-$, A485L Y409V).

1.12.2 RNA 3′-labelling using modified nucleoside triphosphates

Introducing Click chemistry to the RNA field could be particularly beneficial. The 3′-enzymatic labelling could be utilised to modify natural RNA. Only two recent studies [76,82] were found utilising commercially available 3′-azido-2′,3′-dideoxynucleotide triphosphate (AzdNTP) [76,82] and 2′-azido-2′-dNTP [82] to label RNA 3′ or 2′-ends for further dye labelling by CuAAC and SPAAC chemistry. In the most recent study (Figure 1.27) [82], azide-labelled 3′-dideoxynucleotide triphosphates (AzdNTP, N=A, T) and 2′-azido-2′-dNTP (N = A, C, G, U, Figure 1.28) and 8-azido-ATP were tested with Yeast Poly(A) polymerase (PAP), *E.coli* PAP, Cid 1 Poly(U) polymerase (PUP) and terminal deoxynucleotidyl transferase (TdT). The results showed that all four 2′-azido-2′-dNTPs (N = A, C, G, U) were incorporated by Yeast PAP (Figure 1.27). While the reaction with 2′-azido-2′-dGTPs was less efficient, the reaction with all other three triphosphates depleted the starting primer. Multiple additions were observed for all these 4 triphosphates. The 2′-azido-2′-dATP induced significant multiple additions and the others including 2′-azido-2′-dCTP terminated the extension to a large degree after one nucleotide incorporation (Figure 1.27). Also in this study, the reaction has been optimized for incorporation of a single 2′-azido-2′-dNTP.

Modified NTP	No. of residues added			
	Yeast PAP	*E.coli* PAP	Cid1 PUP	TdT
8-N_3-ATP	1-2	x	mostly 0	x
3´-N_3-2´,3´-ddATP	0-1	x	x	ND
3´-N_3-2´,3´-ddTTP	0-1	x	x	ND
2´-N_3-2´-dATP	multiple	1	mostly 0	x
2´-N_3-2´-dCTP	1-2	x	mostly 0	x
2´-N_3-2´-dGTP	1-2	x	mostly 0	mostly 0
2´-N_3-2´-dUTP	1-2	x	mostly 0	x

Reactions were carried out with commercial buffer, **500 µM NTP**, at 37 ℃ for 60 min (TdT: 90 min). Enzyme-specific conditions were the following: 0.2 µM RNA, 24 U/µL Yeast PAP; 0.25 µM RNA, 0.25 U/µL *E.coli* PAP; 0.2 µM RNA, 0.08 U/µL Cid1 PUP; 0.25 µM RNA, 1U/µL TdT.

x: not incorporated, ND: not determined.

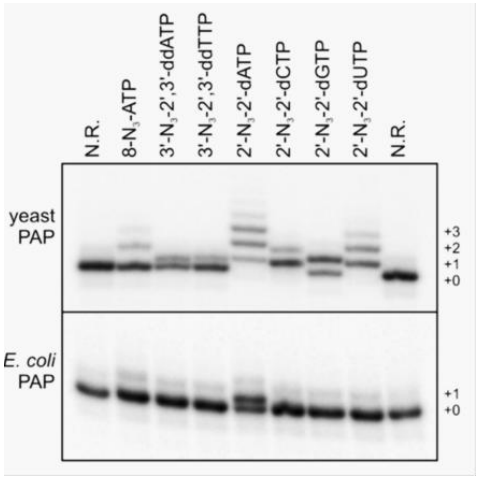


Figure 1.27: Results of the screening of nucleotidyl transferases for incorporation of azido nucleotides [82], reproduced with permission from Oxford University Press.

These results can be utilised to guide our work. However, in this paper no canonical triphosphates as positive controls were employed, making it difficult to evaluate the effect of the azide modification on the efficiency of primer extension or to compare different enzyme batches.

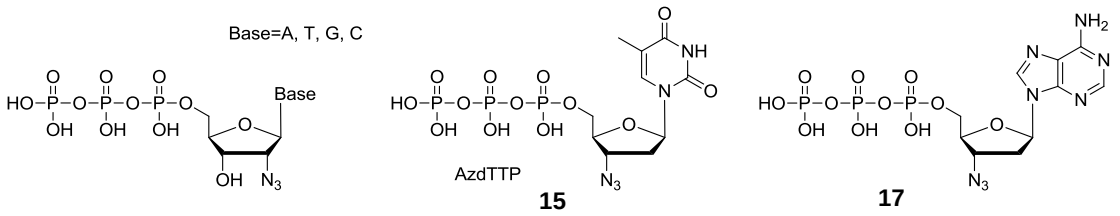


Figure 1.28: NTPs from left to right, 2´-azido-2´-dNTP (N = A, T, G, C); 3´-azido-2´,3´-ddTTP (AzdTTP, **15);** 3´-azido-2´,3´-ddATP (**17**). They were examined in the recent study [82].

Both studies [76,82] utilised Yeast PAP to incorporate 3´-azido-2´,3´-ddATP (Figure 1.28). In the above study [82], both 3´-azido-2´,3´-ddATP (**17**) and 3´-azido-3´-dTTP (AzdTTP, **15**) were examined using Yeast PAP but 3´-N_3-2´,3´-ddATP worked slightly better (Figure 1.27). However, the NTPs with 2´-azides are incorporated best, suggesting that the presence of a 3´-OH group in the NTP is important for enzyme activity. NTPs with 2´-azides and 3´-OH groups are not ideal for our work because they tend to be incorporated multiple times at the 3´-end of the target RNA. This will introduce multiple alkyne/azide tags and the Click ligation may generate a mixture of "branched" products.

Unlike enzymes for DNA 3′-labelling, the number of commercially available enzymes to label RNA is limited. Yeast Poly(A) polymerase (PAP), *E.coli* PAP, Cid 1 poly(U) polymerase (PUP), SP6 RNA polymerase and T7 RNA polymerase are currently the only RNA polymerases commercially available. The SP6 RNA polymerase and T7 RNA polymerase synthesise RNA without a primer but require a complementary DNA strand with the correct promoter sequence as the template [94]. T7 RNA polymerase is utilised for 5′-azide labelling [76] but not for 3′-labelling. Whether T7 RNA polymerase works on a DNA-templated RNA 3′-end is uncertain, but is unlikely because a GMP or a biotin-AMP conjugate [95] is utilised by this enzyme to initiate transcription. TdT is known to work well for DNA, but although TdT has been reported to label RNA using a biotin-dUTP conjugate [96], it is not commonly used for RNA labelling and the labelling efficiency on RNA is unclear.

In addition to RNA polymerases, T4 RNA ligase is another potential enzyme for RNA 3′-labelling. It is discussed in the RNA sequencing section below (Chapter 7.5).

1. Lehninger, A.L., Nelson, D.L. & Cox, M.M. *Lehninger principles of biochemistry (4th ed.)*, (Palgrave Macmillan, 2008).
2. Darzynkiewicz, E. et al. Beta-globin messenger-RNAs capped with M7G, M2(2),7G or M3(2)2,7G differ in intrinsic translation efficiency. *Nucleic Acids Res.* **16**, 8953-8962 (1988).
3. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain. *Science* **324**, 929-930 (2009).
4. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by mll partner tet1. *Science* **324**, 930-935 (2009).
5. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300-1303 (2011).
6. He, Y.F. et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-1307 (2011).
7. Adler, J., Lehman, I.R., Bessman, M.J., Simms, E.S. & Kornberg, A. Enzymatic synthesis of deoxyribonucleic acid. IV. Linkage of single deoxynucleotides to the deoxynucleoside ends of deoxyribonucleic acid. *Proc Natl Acad Sci* **44**, 641-647 (1958).

8. Lehman, I.R., Bessman, M.J., Simms, E.S. & Kornberg, A. Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from *Escherichia coli*. *Jour Biol Chem* **233**, 163-170 (1958).
9. Isobe, M. et al. Chromosome localization of the gene for human terminal deoxynucleotidyltransferase to region 10q23-q25. *Proc. Natl. Acad. Sci. USA* **82**, 5836-5840 (1985).
10. Crick, F. Central dogma of molecular biology. *Nature* **227**, 561-& (1970).
11. Dunham, I. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
12. Cramer, P., Bushnell, D.A. & Kornberg, R.D. Structural basis of transcription: RNA polymerase II at 2.8 angstrom ngstrom resolution. *Science* **292**, 1863-1876 (2001).
13. Barrett, L.W., Fletcher, S. & Wilton, S.D. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell. Mol. Life Sci.* **69**, 3613-3634 (2012).
14. Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H. & Bartel, D.P. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* (2014).
15. Martin, G. & Keller, W. RNA-specific ribonucleotidyl transferases. *RNA-A Publication of the RNA Society* **13**, 1834-1849 (2007).
16. Wilusz, C.J. & Wilusz, J. New ways to meet your (3') end-oligouridylation as a step on the path to destruction. *Genes Dev.* **22**, 1-7 (2008).
17. Mullis, K.B. & Faloona, F.A. Specific synthesis of DNA *in vitro via* a polymerase-catalyzed chain-reaction. *Methods Enzymol.* **155**, 335-350 (1987).
18. Saiki, R.K. et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA-polymerase. *Science* **239**, 487-491 (1988).
19. Schmutz, J. et al. Quality assessment of the human genome sequency. *Nature* **429**, 365-368 (2004).
20. Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
21. Smith, L.M. et al. Fluorescence detection in automated DNA-sequence analysis. *Nature* **321**, 674-679 (1986).
22. Rosenblum, B.B. et al. New dye-labeled terminators for improved DNA sequencing patterns. *Nucleic Acids Res.* **25**, 4500-4504 (1997).
23. Gardner, A.F. & Jack, W.E. Acyclic and dideoxy terminator preferences denote divergent sugar recognition by archaeon and Taq DNA polymerases. *Nucleic Acids Res.* **30**, 605-613 (2002).
24. Metzker, M.L. Applications of next-generation sequencing sequencing technologies - the next generation. *Nat. Rev. Genet.* **11**, 31-46 (2010).

25. Bentley, D.R. et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53-59 (2008).
26. Guo, J. et al. Four-color DNA sequencing with 3'-*O*-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proc. Natl. Acad. Sci. USA* **105**, 9145-9150 (2008).
27. Ruparel, H. et al. Design and synthesis of a 3'-*O*-allyl photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis. *Proc. Natl. Acad. Sci. USA* **102**, 5932-5937 (2005).
28. Ju, J.Y. et al. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc. Natl. Acad. Sci. USA* **103**, 19635-19640 (2006).
29. Ronaghi, M., Uhlen, M. & Nyren, P. A sequencing method based on real-time pyrophosphate. *Science* **281**, 363-+ (1998).
30. Quail, M.A. et al. A tale of three next generation sequencing platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**(2012).
31. Liu, L. et al. Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* (2012).
32. Luo, C.W., Tsementzi, D., Kyrpides, N., Read, T. & Konstantinidis, K.T. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS One* **7**(2012).
33. Froehler, B.C., Ng, P.G. & Matteucci, M.D. Synthesis of DNA via deoxynucleoside *H*-phosphonate intermediates. *Nucleic Acids Res.* **14**, 5399-5407 (1986).
34. Froehler, B.C. & Matteucci, M.D. Nucleoside *H*-phosphonates - valuable intermediates in the synthesis of deoxyoligonucleotides. *Tetrahedron Lett.* **27**, 469-472 (1986).
35. Garegg, P.J., Lindh, I., Regberg, T., Stawinski, J., Stromberg, R. & Henrichson, C. Nucleoside *H*-phosphonates .3. Chemical synthesis of oligodeoxyribonucleotides by the hydrogenphosphonate approach. *Tetrahedron Lett.* **27**, 4051-4054 (1986).
36. Letsinger, R.L. & Ogilvie, K.K. Nucleotide chemistry. XIII. Synthesis of oligothymidylates *via* phosphotriester intermediates. *J. Am. Chem. Soc.* **91**, 3350-3355 (1969).
37. Reese, C.B. Chemical synthesis of oligo-nucleotides and poly-nucleotides by phosphotriester approach. *Tetrahedron* **34**, 3143-3179 (1978).
38. Efimov, V.A., Buryakova, A.A., Reverdatto, S.V., Chakhmakhcheva, O.G. & Ovchinnikov, Y.A. Rapid synthesis of long-chain deoxyribooligonucleotides by the

N-methylimidazolide phosphotriester method. *Nucleic Acids Res.* **11**, 8369-8387 (1983).

39. Efimov, V.A., Molchanova, N.S. & Chakhmakhcheva, O.G. Approach to the synthesis of natural and modified oligonucleotides by the phosphotriester method using *O*-nucleophilic intramolecular catalysis. *Nucleosides Nucleotides & Nucleic Acids* **26**, 1087-1093 (2007).
40. Srivastava, S.C., Pandey, D., Srivastava, N.P. & Bajpai, S.P. RNA synthesis: Phosphoramidites for RNA synthesis in the reverse direction. Highly efficient synthesis and application to convenient introduction of ligands, chromophores and modifications of synthetic RNA at the 3'-end. *Nucleic acids symposium series (2004)*, 103-4 (2008).
41. Huisgen, R. 1,3-Dipolar cycloadditions past and future. *Angew. Chem.-Int. Edit.* **2**, 565-632 (1963).
42. Tornoe, C.W., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase: 1,2,3 -triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **67**, 3057-3064 (2002).
43. Rostovtsev, V.V., Green, L.G., Fokin, V.V. & Sharpless, K.B. A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Ed.* **41**, 2596-+ (2002).
44. Kolb, H.C., Finn, M.G. & Sharpless, K.B. Click chemistry: Diverse chemical function from a few good reactions. *Angew. Chem. Int. Ed.* **40**, 2004-+ (2001).
45. Worrell, B.T., Malik, J.A. & Fokin, V.V. Direct evidence of a dinuclear copper intermediate in Cu(I)-catalyzed azide-alkyne cycloadditions. *Science* **340**, 457-460 (2013).
46. Chan, T.R., Hilgraf, R., Sharpless, K.B. & Fokin, V.V. Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Org. Lett.* **6**, 2853-2855 (2004).
47. Kanan, M.W., Rozenman, M.M., Sakurai, K., Snyder, T.M. & Liu, D.R. Reaction discovery enabled by DNA-templated synthesis and *in vitro* selection. *Nature* **431**, 545-549 (2004).
48. Kennedy, D.C. et al. Cellular consequences of copper complexes used to catalyze bioorthogonal click reactions. *J. Am. Chem. Soc.* **133**, 17993-18001 (2011).
49. Besanceney-Webler, C. et al. Increasing the efficacy of bioorthogonal click reactions for bioconjugation: A comparative study. *Angew. Chem. Int. Ed.* **50**, 8051-8056 (2011).
50. del Amo, D.S. et al. Biocompatible copper(I) catalysts for *in vivo* imaging of glycans. *J. Am. Chem. Soc.* **132**, 16893-16899 (2010).

51. Presolski, S.I., Hong, V., Cho, S.H. & Finn, M.G. Tailored ligand acceleration of the Cu-catalyzed azide-alkyne cycloaddition reaction: Practical and mechanistic implications. *J. Am. Chem. Soc.* **132**, 14570-14576 (2010).
52. Huisgen, R. Kinetics and mechanism of 1,3-dipolar cycloadditions. *Angew. Chem.-Int. Edit.* **2**, 633-696 (1963).
53. Baskin, J.M. & Bertozzi, C.R. Copper-free click chemistry: Bioorthogonal reagents for tagging azides. *Aldrichimica Acta* **43**, 15-23 (2010).
54. Saxon, E. & Bertozzi, C.R. Cell surface engineering by a modified Staudinger reaction. *Science* **287**, 2007-2010 (2000).
55. Agard, N.J., Prescher, J.A. & Bertozzi, C.R. A strain-promoted 3+2 azide-alkyne cycloaddition for covalent modification of blomolecules in living systems. *J. Am. Chem. Soc.* **126**, 15046-15047 (2004).
56. Agard, N.J., Baskin, J.M., Prescher, J.A., Lo, A. & Bertozzi, C.R. A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol.* **1**, 644-648 (2006).
57. Ning, X.H., Guo, J., Wolfert, M.A. & Boons, G.J. Visualizing metabolically labeled glycoconjugates of living cells by copper-free and fast Huisgen cycloadditions. *Angew. Chem. Int. Ed.* **47**, 2253-2255 (2008).
58. Jewett, J.C., Sletten, E.M. & Bertozzi, C.R. Rapid Cu-free click chemistry with readily synthesized biarylazacyclooctynones. *J. Am. Chem. Soc.* **132**, 3688-+ (2010).
59. Dommerholt, J. et al. Readily accessible bicyclononynes for bioorthogonal labeling and three-dimensional imaging of living cells. *Angew. Chem. Int. Ed.* **49**, 9422-9425 (2010).
60. Laughlin, S.T., Baskin, J.M., Amacher, S.L. & Bertozzi, C.R. *In vivo* imaging of membrane-associated glycans in developing zebrafish. *Science* **320**, 664-667 (2008).
61. Baskin, J.M. et al. Copper-free click chemistry for dynamic *in vivo* imaging. *Proc. Natl. Acad. Sci. USA* **104**, 16793-16797 (2007).
62. Qiu, J.Q., El-Sagheer, A.H. & Brown, T. Solid phase click ligation for the synthesis of very long oligonucleotides. *Chem. Commun.* **49**, 6959-6961 (2013).
63. Gerowska, M., Hall, L., Richardson, J., Shelbourne, M. & Brown, T. Efficient reverse click labeling of azide oligonucleotides with multiple alkynyl Cy-dyes applied to the synthesis of hybeacon probes for genetic analysis. *Tetrahedron* **68**(2012).
64. El-Sagheer, A.H. & Brown, T. New strategy for the synthesis of chemically modified RNA constructs exemplified by hairpin and hammerhead ribozymes. *Proc. Natl. Acad. Sci. USA* **107**, 15329-15334 (2010).
65. Kocalka, P., El-Sagheer, A.H. & Brown, T. Rapid and efficient DNA strand cross-linking by click chemistry. *ChemBioChem* **9**, 1280-1285 (2008).

66. Berndl, S. et al. Comparison of a nucleosidic *vs* non-nucleosidic postsynthetic "click" modification of DNA with base-labile fluorescent probes. *Bioconjugate Chem.* **20**, 558-564 (2009).

67. Shelbourne, M., Chen, X., Brown, T. & El-Sagheer, A.H. Fast copper-free click DNA ligation by the ring-strain promoted alkyne-azide cycloaddition reaction. *Chem. Commun.* **47**, 6257-6259 (2011).

68. Gololobov, Y.G., Zhmurova, I.N. & Kasukhin, L.F. 60 Years of Staudinger reaction. *Tetrahedron* **37**, 437-472 (1981).

69. El-Sagheer, A.H. & Brown, T. Synthesis and polymerase chain reaction amplification of DNA strands containing an unnatural triazole linkage. *J. Am. Chem. Soc.* **131**, 3958-3964 (2009).

70. Kumar, R., El-Sagheer, A., Tumpane, J., Lincoln, P., Wilhelmsson, L.M. & Brown, T. Template-directed oligonucleotide strand ligation, covalent intramolecular DNA circularization and catenation using click chemistry. *J. Am. Chem. Soc.* **129**, 6859-6864 (2007).

71. El-Sagheer, A.H., Sanzone, A.P., Gao, R., Tavassoli, A. & Brown, T. Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **108**, 11338-11343 (2011).

72. El-Sagheer, A.H. & Brown, T. Efficient RNA synthesis by *in vitro* transcription of a triazole-modified DNA template. *Chem. Commun.* **47**, 12057-12058 (2011).

73. Aigner, M., Hartl, M., Fauster, K., Steger, J., Bister, K. & Micura, R. Chemical synthesis of site-specifically 2'-azido-modified RNA and potential applications for bioconjugation and RNA interference. *ChemBioChem* **12**, 47-51 (2011).

74. Fauster, K. et al. 2'-azido RNA, a versatile tool for chemical biology: Synthesis, X-ray structure, siRNA applications, click labeling. *ACS Chem. Biol.* **7**, 581-589 (2012).

75. Isobe, H., Fujino, T., Yamazaki, N., Guillot-Nieckowski, M. & Nakamura, E. Triazole-linked analogue of deoxyribonucleic acid ((TL)DNA): Design, synthesis, and double-strand formation with natural DNA. *Org. Lett.* **10**, 3729-3732 (2008).

76. Paredes, E. & Das, S.R. Click chemistry for rapid labeling and ligation of RNA. *ChemBioChem* **12**, 125-131 (2011).

77. Miller, G.P. & Kool, E.T. Versatile 5'-functionalization of oligonucleotides on solid support: Amines, azides, thiols, and thioethers *via* phosphorus chemistry. *J. Org. Chem.* **69**, 2404-2410 (2004).

78. Miller, G.P. & Kool, E.T. A simple method for electrophilic functionalization of DNA. *Org. Lett.* **4**, 3599-3601 (2002).

79. Fujino, T., Yamazaki, N. & Isobe, H. Convergent synthesis of oligomers of triazole-linked DNA analogue ((TL)DNA) in solution phase. *Tetrahedron Lett.* **50**, 4101-4103 (2009).

80. Kiviniemi, A., Virta, P. & Lonnberg, H. Utilization of intrachain 4'-C-azidomethylthymidine for preparation of oligodeoxyribonucleotide conjugates by click chemistry in solution and on a solid support. *Bioconjugate Chem.* **19**, 1726-1734 (2008).

81. Pourceau, G., Meyer, A., Vasseur, J.J. & Morvan, F. Azide solid support for 3'-conjugation of oligonucleotides and their circularization by click chemistry. *J. Org. Chem.* **74**, 6837-6842 (2009).

82. Winz, M.L., Samanta, A., Benzinger, D. & Jaschke, A. Site-specific terminal and internal labeling of RNA by poly(A) polymerase tailing and copper-catalyzed or copper-free strain-promoted click chemistry. *Nucleic Acids Res.* **40**, 13 (2012).

83. Sambrook, J. & Russell, D.W. Molecular cloning: A laboratory manual. *Molecular cloning: A laboratory manual* (2001).

84. Bullard, D.R. & Bowater, R.P. Direct comparison of nick-joining activity of the nucleic acid ligases from bacteriophage T4. *Biochem. J* **398**, 135-144 (2006).

85. Gierlich, J., Burley, G.A., Gramlich, P.M.E., Hammond, D.M. & Carell, T. Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA. *Org. Lett.* **8**, 3639-3642 (2006).

86. Schierwater, B. & Ender, A. Different thermostable DNA-polymerases may amplify different RAPD products. *Nucleic Acids Res.* **21**, 4647-4648 (1993).

87. Lawyer, F.C. et al. High-level expression, purification, and enzymatic characterization of full-length thermus aquaticus DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity. *PCR methods and applications* **2**, 275-87 (1993).

88. Strachan T, R.A. Figure 20.1, RACE-PCR facilitates the isolation of 5' and 3' end sequences from cDNA. *Human Molecular Genetics. 2nd edition.* (1999).

89. Gavrieli, Y., Sherman, Y. & Bensasson, S.A. Identification of programmed cell-death *in situ via* specific labeling of nuclear-DNA fragmentation. *J. Cell Biol.* **119**, 493-501 (1992).

90. Krayevsky, A.A., Victorova, L.S., Arzumanov, A.A. & Jasko, M.V. Terminal deoxynucleotidyl transferase: Catalysis of DNA (oligodeoxynucleotide) phosphorylation. *Pharmacol. Ther.* **85**, 165-173 (2000).

91. Gierlich, J., Gutsmiedl, K., Gramlich, P.M.E., Schmidt, A., Burley, G.A. & Carell, T. Synthesis of highly modified DNA by a combination of PCR with alkyne-bearing triphosphates and click chemistry. *Chem.-Eur. J.* **13**, 9486-9494 (2007).

92. Alexandrova, L.A., Skoblov, A.Y., Jasko, M.V., Victorova, L.S. & Krayevsky, A.A. 2'-deoxynucleoside 5'-triphosphates modified at alpha-, beta- and gamma-phosphates as substrates for DNA polymerases. *Nucleic Acids Res.* **26**, 778-786 (1998).

93. Metzker, M.L. et al. Termination of DNA-synthesis by novel 3'-modified-deoxyribonucleoside 5'-triphosphates. *Nucleic Acids Res.* **22**, 4259-4267 (1994).

94. Schenborn, E.T. & Mierendorf, R.C. A novel transcription property of SP6 and T7 RNA-polymerases - dependence on template structure. *Nucleic Acids Res.* **13**, 6223-6236 (1985).

95. Huang, F.Q., He, J., Zhang, Y.L. & Guo, Y.L. Synthesis of biotin-AMP conjugate for 5' biotin labeling of RNA through one-step *in vitro* transcription. *Nat. Protoc.* **3**, 1848-1861 (2008).

96. Rosemeyer, V., Laubrock, A. & Seibl, R. Nonradioactive 3'-end-labeling of RNA molecules of different lengths by terminal deoxynucleotidyltransferase. *Anal. Biochem.* **224**, 446-449 (1995).

CHAPTER 2 Synthesis of 2′ and 3′-*O*-Propargyl Modified Nucleotide Triphosphates

2.1. Introduction

In order to use Click chemistry to ligate natural DNA and RNA, the alkyne or azide tags need to be introduced at the DNA or RNA 3′-end by incorporation of modified nucleoside triphosphates. As the short triazole linkages (triazole backbones) generated by Click ligation are natural phosphodiester backbone mimics, the Click ligation involving these triazole backbones can generate functional products biologically. The triazole Backbone B, for example, can be read-through by a DNA polymerase [1] and a DNA dependent RNA polymerase [2]. Due to its excellent bio-compatibility, triazole Backbone B (Figure 2.1) was selected as our first target for Click ligation. This triazole backbone requires a 3′-*O*-propargyl-modified nucleoside triphosphate to be introduced to the 3′-end of an up-stream strand (Figure 2.1).

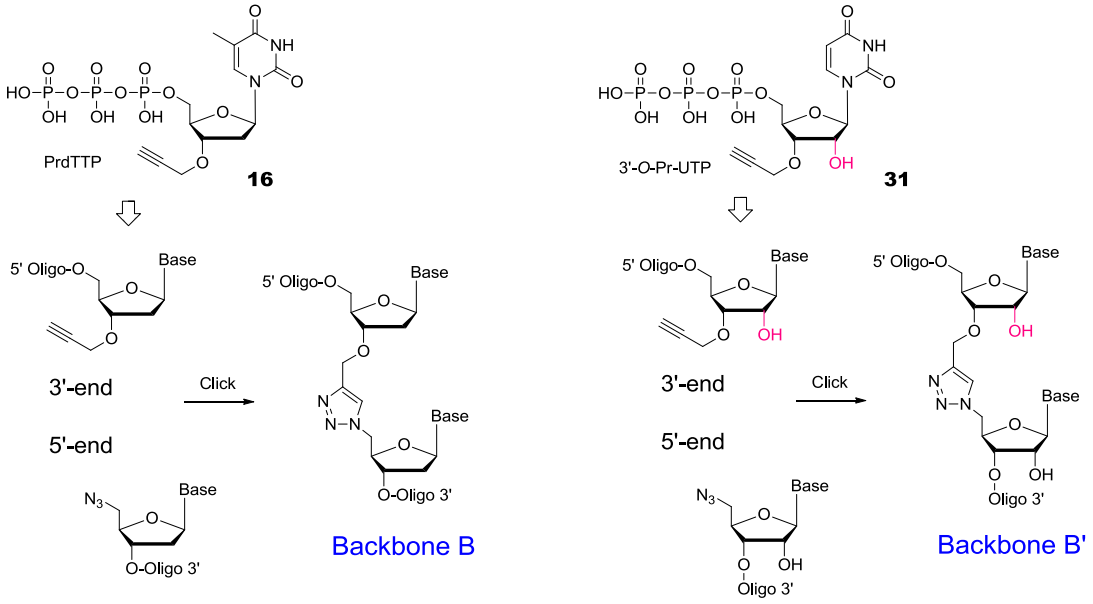


Figure 2.1: The Click-ligation approach for the construction of triazole Backbones B and B′

However, no 3′-*O*-propargyl modified nucleosides were commercially available as triphosphates. For DNA Click ligation, 3′-*O*-propargylthymidine triphosphate (PrdTTP) was selected for synthesis because thymidine is the most stable nucleoside chemically and requires no nucleobase protecting group during triphosphate synthesis. For RNA Click ligation, the 3′-*O*-propargyl modified nucleoside triphosphates could be synthesised from commercially available 3′-*O*-propargyl nucleoside functionalised resins (Figure 2.2) instead of the corresponding unmodified nucleosides. This could avoid the time-consuming

propargylation, protection and deprotection steps (discussed below). These ribonucleoside triphosphates with a 2′-OH will lead to triazole Backbone B′, where the 2′-OH is adjacent to the triazole site (Figure 2.1).

The 2′-*O*-propargyl nucleoside functionalised resins are also commercially available. As it is uncertain whether 2′-*O*-propargyl nucleoside triphosphates or 3′-*O*-propargyl nucleoside triphosphates are more efficiently incorporated by RNA polymerase, the 2′-*O*-propargyl nucleoside triphosphates were synthesised in parallel with 3′-*O*-propargyl modified nucleoside triphosphates for enzymatic incorporation. All triphosphates synthesised in this project are listed in Figure 2.3.

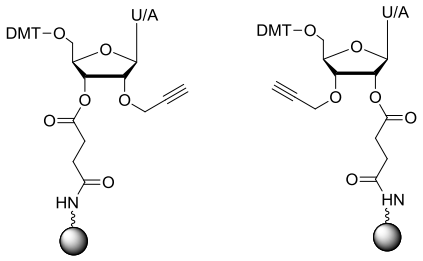


Figure 2.2: Commercially available 2′-*O*-propargyluridine/adenosine 3′-LCAA CPG resin and 3′-*O*-propargyluridine/adenosine 2′-LCAA CPG resin.

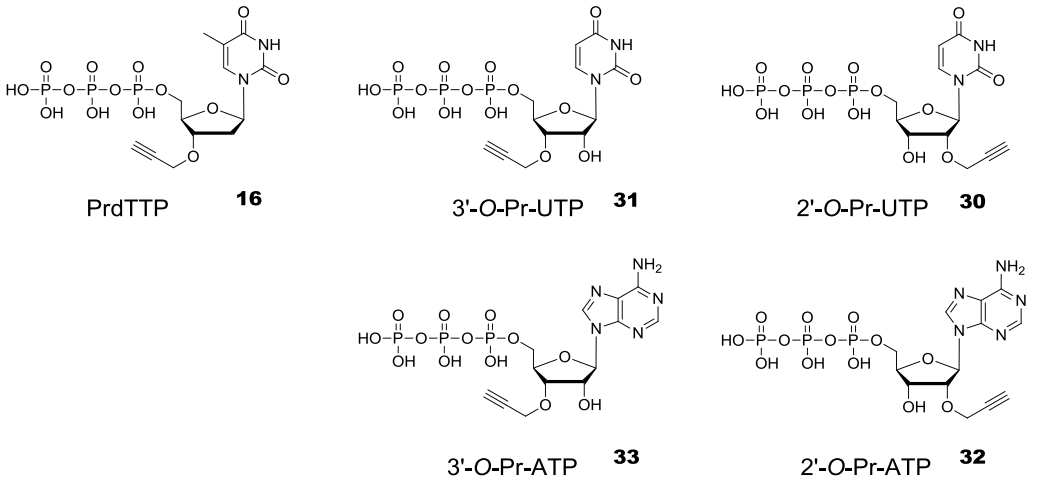
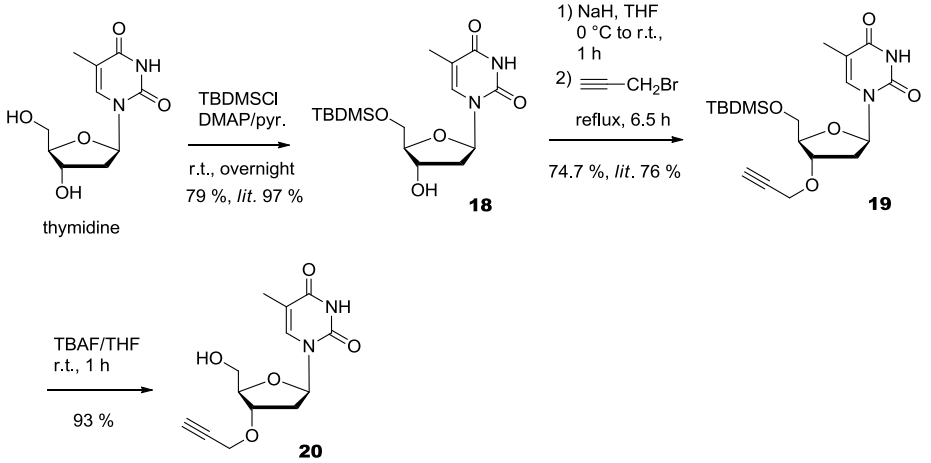


Figure 2.3: Nucleoside triphosphates synthesised

2.2 Synthesis of 3′-*O*-propargylthymidine triphosphate (PrdTTP)

2.2.1 Synthesis of 3′-O-propargylthymidine

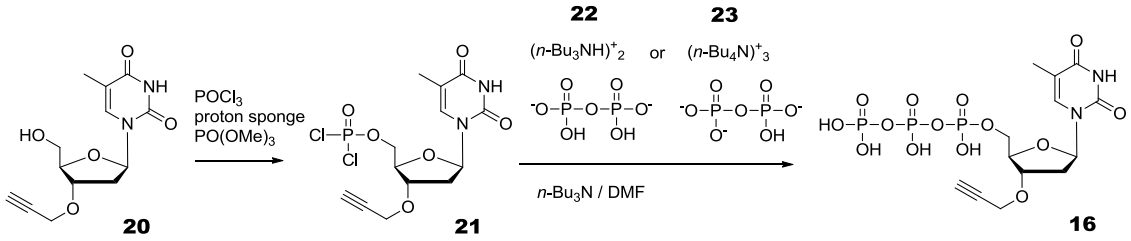
The synthesis of 3′-*O*-propargylthymidine from thymidine is straightforward (Scheme 2.1). The 5′-OH group of thymidine was selectively protected with a TBDMS group as it is more reactive than the 3′-OH. After propargylation, the TBDMS group was removed by TBAF (Step 3, Scheme 2.1). All the compounds synthesised at this stage are known compounds [3].



Scheme 2.1: Synthetic route to 3′-*O*-propargylthymidine (**20**) from thymidine.

2.2.2 Triphosphorylation of 3′-O-propargylthymidine in solution using phosphoryl chloride–Method I

Triphosphorylation of 3′-*O*-propargylthymidine was initially attempted using a simple one-pot protocol (Scheme 2.2) [4,5]. Phosphoryl chloride ($POCl_3$) was firstly added to react with the 5′-OH of the 3′-*O*-propargylthymidine. The phosphorodichloridate product (**21**) was then reacted *in-situ* with tri-*n*-butylammonium pyrophosphate to form the triphosphate. The protocol was reported to have a poor yield (2-22% reported yields) and required the use of expensive tri-*n*-butylammonium pyrophosphate (**22**). In our first experiments, the monophosphate formation went to completion when excess of $POCl_3$ was used. However, step 2, conversion of the monophosphorodichloridate to the triphosphate, proceeded very slowly and less than 50%, after stirring at 4 ℃ overnight. After workup, the more polar triphosphate bands were largely separated from the monophosphate band by ISOLUTE® Flash C18 reverse-phase column chromatography and elution with water.



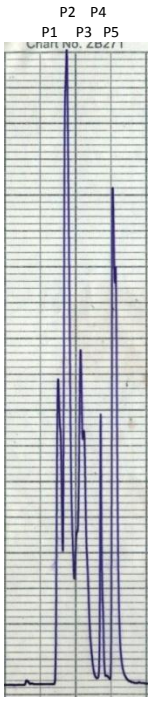
Scheme 2.2: One-pot triphosphorylation method I.

The expensive *bis*(tri-*n*-butylammonium) pyrophosphate (**22**) was replaced by the cheaper and more nucleophilic *tris*(tetra-*n*-butylammonium) hydrogen pyrophosphate (**23**). This hygroscopic pyrophosphate salt is soluble in DMF and showed similar reactivity to *bis*(tri-*n*-butylammonium) pyrophosphate towards the phosphorodichloridate (Scheme 2.2).

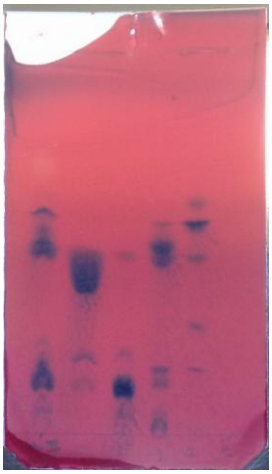
The crude product obtained using *tris*(tetra-*n*-butylammonium) hydrogen pyrophosphate contained both monophosphate and triphosphate as a mixture, and was similar to the crude product obtained by using *bis*(tri-*n*-butylammonium) pyrophosphate. The behaviour of the two crude products on C18 flash column chromatography, however, was completely different. When using *bis*(tri-*n*-butylammonium) pyrophosphate, the most polar triphosphate band quickly eluted out of the reverse-phase column followed by the monophosphate band using TEAB buffer or water, without MeCN as an eluent. This complied with the general order of elution of reverse-phase chromatography as the most polar compound eluted first. When using *tris*(tetra-*n*-butylammonium) hydrogen pyrophosphate however, the order was reversed. The monophosphate eluted first but very late. The triphosphate band eluted even later, with about 30% MeCN in TEAB buffer.

The flash column purified fractions were further purified by HPLC on a C8 reversed-phase column with a shallow gradient of MeCN in 0.1 M TEAB buffer (pH 7.5). The HPLC chromatogram (Figure 2.1) showed good separation. TLC and M.S. analysis of the HPLC-purified samples confirmed that the triphosphate triethylamine (TEA) salt was obtained in high purity. Suspecting that the TEA may interfere with polymerase enzymes, the dNTP was converted to a sodium salt by DOWEX® ion-exchange resin (Na^+-form) but the procedure caused partial degradation of the triphosphate. It was subsequently discovered that the triphosphate TEA salt is generally used in PCR [6-8]. The synthesised triphosphate TEA salt was tested by us and found to be accepted by Therminator™II (9 °N$_m$ pol.) as shown below.

This also helped to avoid sample degradation, because the triphosphate sodium salt is less stable than the TEA salt in solution. To our knowledge, a thorough comparison of the suitability of TEA and sodium salts of dNTPs with different polymerases has not yet been carried out.



HPLC trace (UV)

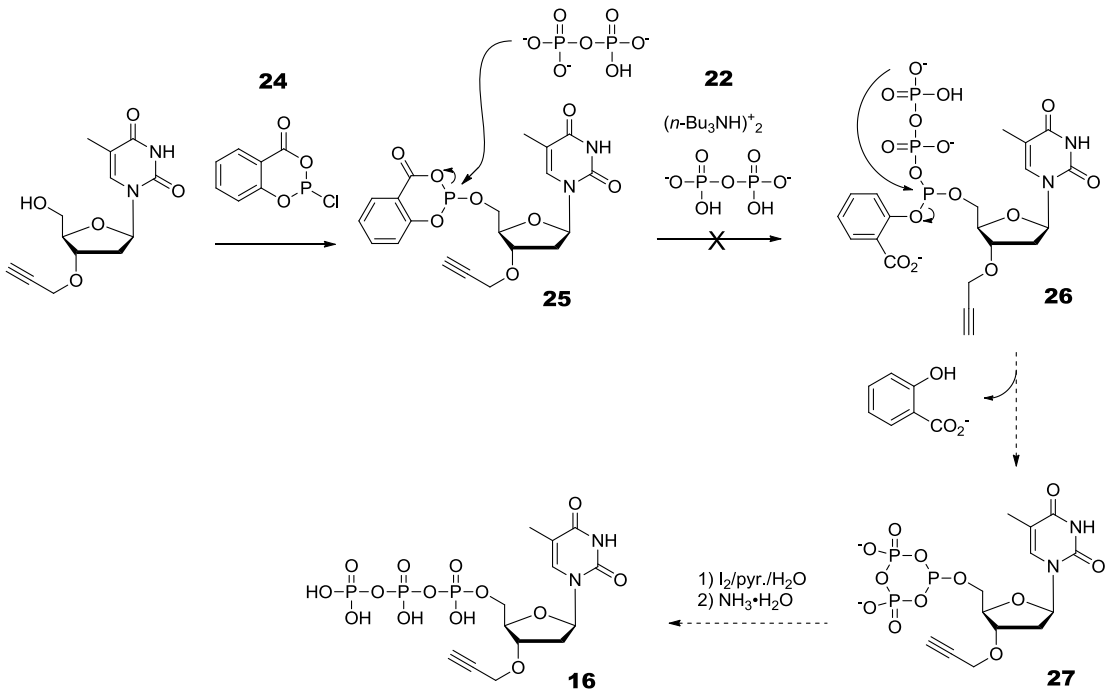


TLC

Figure 2.1: HPLC purification of the crude nucleoside triphosphate and the corresponding TLC (*n*-propanol : H_2O : *aq* NH_3, 3:1:1) result. Peak 2 (P2) corresponded to the correct 3′-*O*-propargylthymidine triphosphate.

2.2.3 Triphosphorylation of 3´-O-propargylthymidine in solution using Ludwig one-pot triphosphorylation – Method II-a

Another frequently adopted triphosphorylation method developed by Ludwig *et al.* [9] using salicyl chlorophosphite (**24**) was evaluated (Scheme 2.3). The *bis*(tri-*n*-butylammonium) pyrophosphate, which is available in our lab, was tested first. In several attempts, the first step, forming the salicyl phosphite, was largely completed but the second step did not proceed well and only a trace amount of triphosphate was generated. Further experiments in solution using more nucleophilic *tris*(tetra-*n*-butylammonium) hydrogen pyrophosphate were not performed as the triphosphorylation method I in solution using *tris*(tetra-*n*-butylammonium) hydrogen pyrophosphate was found to be efficient enough.

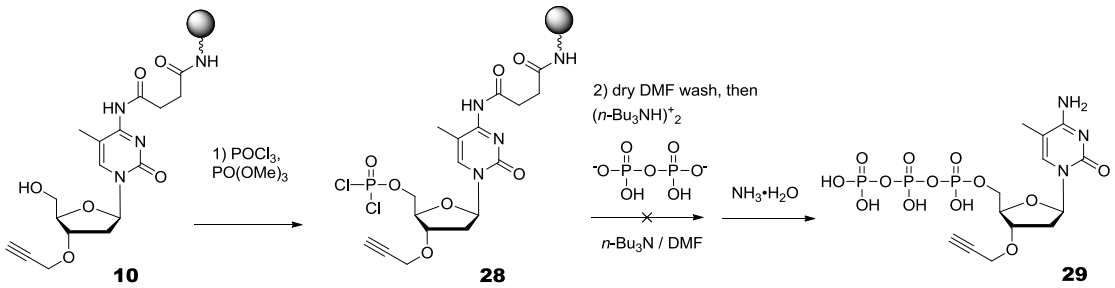


Scheme 2.3: Ludwig one-pot triphosphorylation method II-a.

2.3 Solid-phase triphosphorylation

2.3.1 Attempted triphosphorylation of 3′-O-propargyl-5-methyldeoxycytidine functionalised resin using method I

For triphosphorylation of nucleoside on solid-support, method I (phosphoryl chloride) was first attempted on 3′-*O*-propargyl-5-methyl-2′-deoxycytidine attached to the resin by a succinyl linkage at N4 (**10**, Scheme 2.4). It was shown that the succinyl linkage is stable during treatment with *tris*(tetra-*n*-butylammonium) hydrogen pyrophosphate solution and tributylamine, because no significant nucleoside compound was identified in the wash solution in this step. However, no triphosphate was obtained after cleavage of the product from the resin by concentrated aqueous ammonia treatment. The reaction generated two bands in the monophosphate region on TLC with uncertain identities (Mass spec. was not performed). A possible reason for this failure is that the 5′-phosphorodichloridate (**28**) may be too reactive and was thus transformed to 5′-phosphate during the DMF wash and thus lost its reactivity.



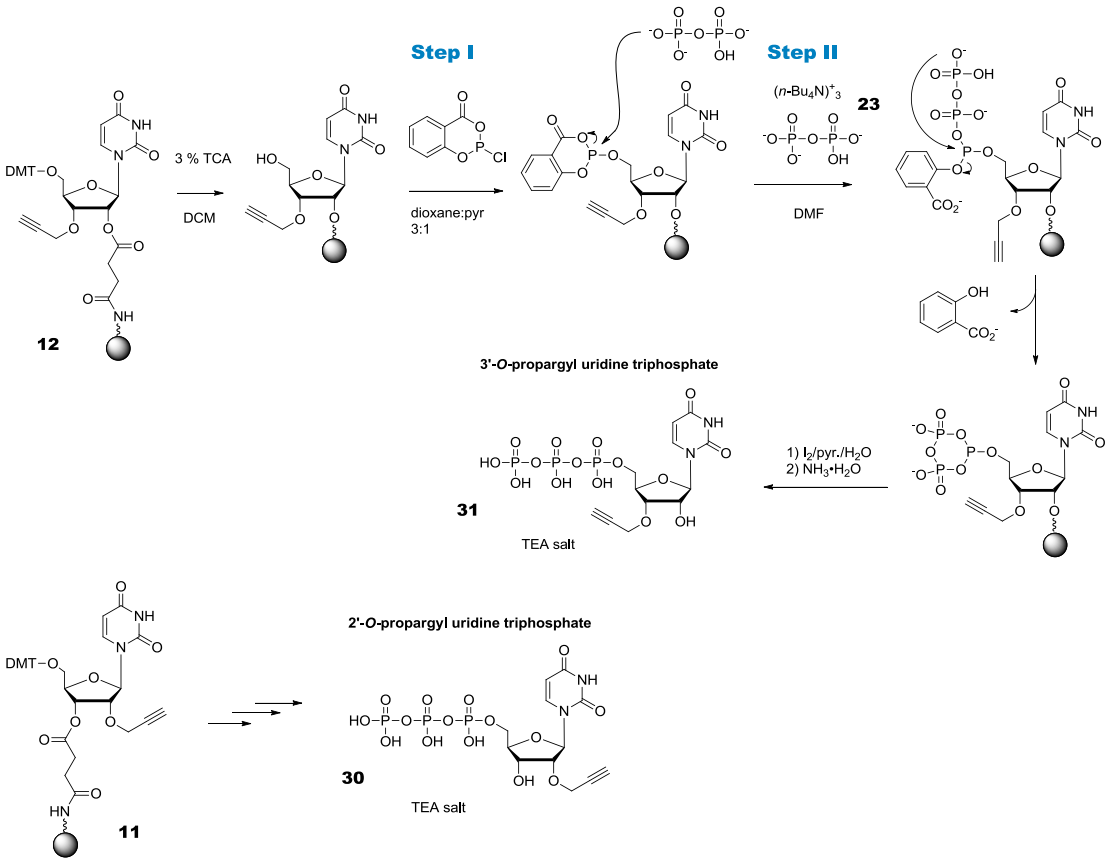
Scheme 2.4: Proposed phosphoryl chloride method I for solid-phase synthesis of 3′-*O*-propargyl-5-methyldeoxycytidine triphosphate (**29**). The *tris*(tetra-*n*-butyl-ammonium) hydrogen pyrophosphate was not tested.

2.3.2 Solid-phase synthesis of 2′-O-propargyluridine-5′-triphosphate and 3′-O-propargyluridine-5′-triphosphate using Ludwig triphosphorylation – Method II-a

The salicyl chlorophosphite method (Method II-a) mentioned above has one advantage over Method I: it is likely to be more compatible with solid-phase synthesis and success has been reported [10]. This is likely because the salicyl phosphite intermediate (compound **25**, Scheme 2.3) is more stable than the phosphorodichloridate (compound **28**, Scheme 2.4), thus excess salicyl chlorophosphite can be washed off the solid support while the salicyl phosphite intermediate remains intact. This strategy was successfully utilised by us for the synthesis of 2′-*O*-propargyluridine triphosphate (compound **30**) and 3′-*O*-propargyluridine triphosphate (compound **31**) from commercially available resins (compound **11**, **12**).

The two resins reacted differently with the salicyl chlorophosphite reagent (step I, Scheme 2.5). For the reaction on 3′-*O*-propargyluridine resin, step I was performed for 1 h, but for the reaction on 2′-*O*-propargyluridine resin step I was shortened to 15 min (as in the publication [29]) to generate the triphosphate (**30**).

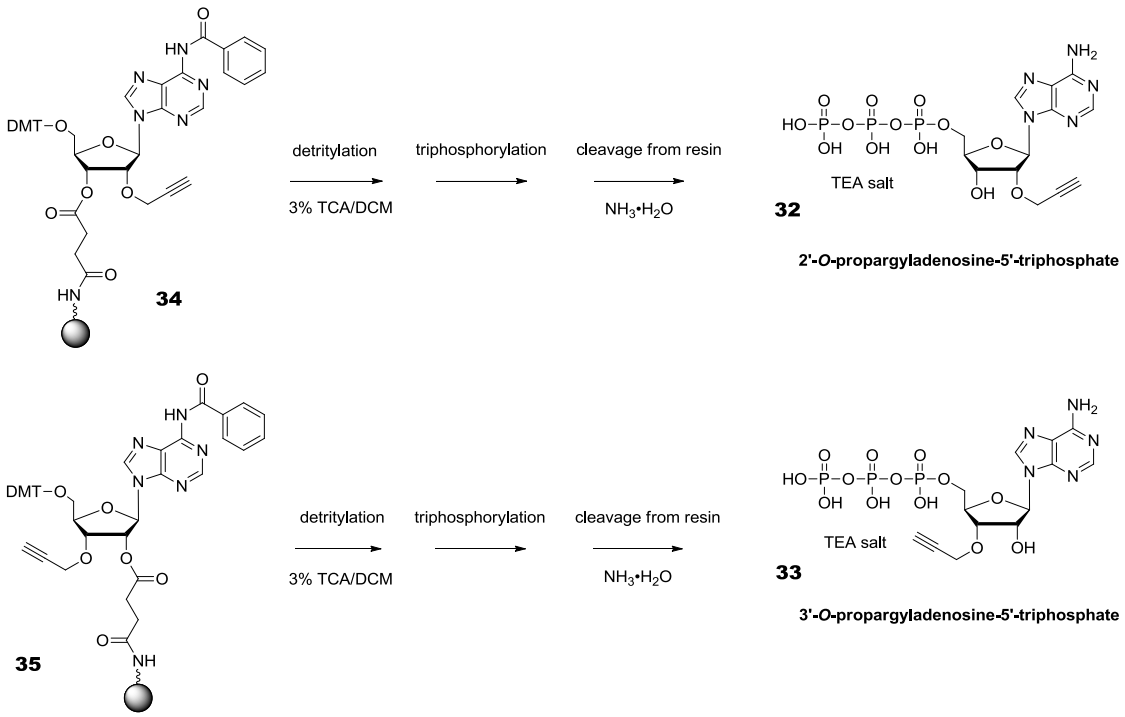
Upon cleavage from the solid support, the crude products were directly purified by HPLC using the procedure as previously described. Avoiding the flash-column purification of the oily crude product of solution-phase synthesis is another time-saving advantage of solid-phase synthesis.



Scheme 2.5: Ludwig triphosphorylation (method II-a) [9] on commercially available 3′-*O*-propargyl uridine 2′-lcaa CPG resin (**12**) and 2′-*O*-propargyl uridine 3′-lcaa CPG resin (**11**).

2.3.3 Solid-phase synthesis of 2′-O-propargyladenosine-5′-triphosphate and 3′-O-propargyladenosine-5′-triphosphate using Ludwig triphosphorylation – Method II-a

The 2′-*O*-propargyladenosine-5′-triphosphate and 3′-*O*-propargyladenosine-5′-triphosphate (**32** and **33**) were also synthesised from the corresponding commercially available resins (**34** and **35**, Scheme 2.6). The benzoyl protecting group on adenine needs to be deprotected after triphosphorylation. The stability of ATP to the deprotection conditions (ammonia, 55 ℃) was checked. TLC analysis showed that it is stable at 55 ℃ for 5 h with less than 10% degradation. To ensure minimal degradation, the crude 2′ and 3′-*O*-propargyladenosine-5′-triphosphate products were reacted with concentrated ammonia at 55 ℃ for only 2 h and the deprotected triphosphates were successfully obtained.

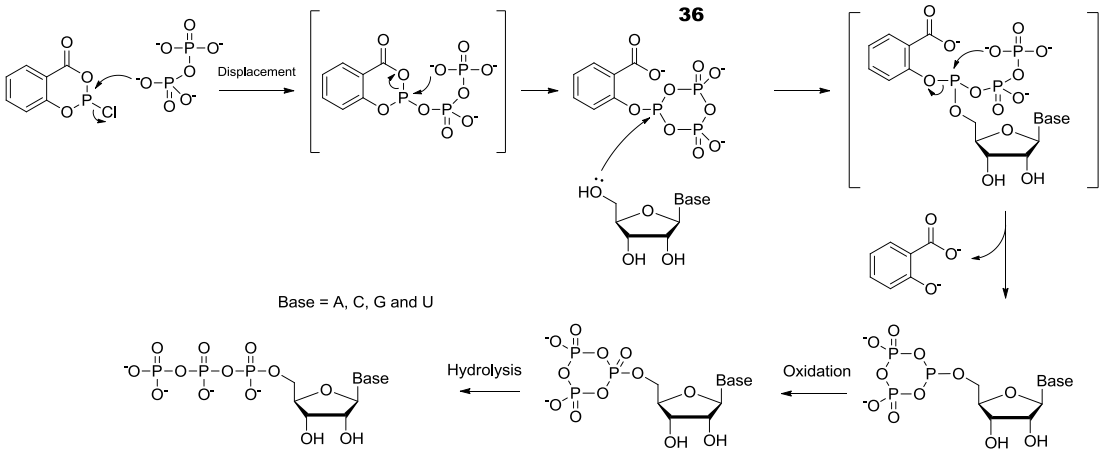


Scheme 2.6: Ludwig triphosphorylation (method II-a) on commercially available 5′-DMT-2′-*O*-propargyladenosine (*N*-Bz)-3′-lcaa CPG resin (**32**) and 5′-DMT-3′-*O*-propargyl adenosine (*N*-Bz)-2′-lcaa CPG resin (**33**).

In addition, the compatibility of the propargyl group with I_2 was also tested. It was shown by TLC that the side-reaction was very slow under conditions similar to those used on the DNA synthesiser. It was therefore concluded that iodine addition to the alkyne is negligible.

2.3.4 Triphosphorylation by active intermediate on solid-support – Method II-b

A recent protocol [11,12] reported a modification of the classic Ludwig method. By changing the sequence of addition of reagents: adding nucleoside after mixing salicyl chlorophosphite and *tris*(tetrabutylammonium) hydrogen pyrophosphate, the salicyl chlorophosphite is consumed by excessive pyrophosphate reagent to form an active intermediate (**36**, Scheme 2.7). This active intermediate is less reactive than salicyl chlorophosphite and selectively reacts with the 5′-OH of the nucleoside. As a result, only nucleoside triphosphate but little monophosphate is formed.



Scheme 2.7: Proposed mechanism in the literature [11].

Although use of this method has not been reported for solid-phase synthesis, it was tested by us on the 3′-*O*-propargyl-5-methyl-2′-deoxycytidine functionalised (N4) resin (**10**). However, the reaction did not proceed well, affording mainly the starting material after cleavage from resin. In one additional test, the active intermediate was separately prepared in two portions to be added twice to the resin and left for 1 h each time. The TLC of the cleaved crude products showed that although double-addition of the active intermediate increased the intensity of a potential triphosphate band, the reaction was very slow, and only a faint triphosphate band was observed when the active intermediate was left on the resin overnight, and a large amount of starting material was retrieved.

2.4 Conclusions

To synthesise alkyne/azide modified nucleoside triphosphates, two phosphorylation methods were established. The phosphoryl trichloride method was utilised to synthesise the novel 3′-*O*-propargylthymidine-5′-triphosphate (PrdTTP, **16**) for DNA primer extension. The salicyl chlorophosphite method was successfully used for solid-phase synthesis of novel 2′-*O*-propargyluridine/adenosine triphosphate (**30**/**32**) and 3′-*O*-propargyluridine/adenosine triphosphate (**31**/**33**) for subsequent RNA primer extension. The commercially-available alkyl nucleoside functionalised resins provide quick access to alkyne-bearing nucleoside triphosphates. This method potentially can also be applied to azido nucleoside functionalised resins.

1. El-Sagheer, A.H., Sanzone, A.P., Gao, R., Tavassoli, A. & Brown, T. Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in Escherichia coli. *Proc. Natl. Acad. Sci. USA* **108**, 11338-11343 (2011).
2. El-Sagheer, A.H. & Brown, T. Efficient RNA synthesis by *in vitro* transcription of a triazole-modified DNA template. *Chem. Commun.* **47**, 12057-12058 (2011).
3. James, D., Escudier, J.M., Amigues, E., Schulz, J., Vitry, C., Bordenave, T., Szlosek-Pinaud, M. & Fouquet, E. A 'click chemistry' approach to the efficient synthesis of modified nucleosides and oligonucleotides for PET imaging. *Tetrahedron Lett.* **51**, 1230-1232 (2010).
4. Wu, W., Stupi, B.P., Litosh, V.A., Mansouri, D., Farley, D., Morris, S., Metzker, S. & Metzker, M.L. Termination of DNA synthesis by *N*-6-alkylated, not 3-*O*-alkylated, photocleavable 2-deoxyadenosine triphosphates. *Nucleic Acids Res.* **35**, 6339-6349 (2007).
5. Litosh, V.A.L.V.A., Wu, W.D., Stupi, B.P., Wang, J.C., Morris, S.E., Hersh, M.N. & Metzker, M.L. Improved nucleotide selectivity and termination of 3'-OH unblocked reversible terminators by molecular tuning of 2-nitrobenzyl alkylated HOMedU triphosphates. *Nucleic Acids Res.* **39**, 13 (2011).
6. Ruparel, H., Bi, L.R., Li, Z.M., Bai, X.P., Kim, D.H., Turro, N.J. & Ju, J.Y. Design and synthesis of a 3'-*O*-allyl photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis. *Proc. Natl. Acad. Sci. USA* **102**, 5932-5937 (2005).
7. Ju, J.Y., Kim, D.H., Bi, L.R., Meng, Q.L., Bai, X.P., Li, Z.M., Li, X.X., Marma, M.S., Shi, S., Wu, J., Edwards, J.R., Romu, A. & Turro, N.J. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc. Natl. Acad. Sci. USA* **103**, 19635-19640 (2006).
8. Guo, J., Xu, N., Li, Z.M., Zhang, S.L., Wu, J., Kim, D.H., Marma, M.S., Meng, Q.L., Cao, H.Y., Li, X.X., Shi, S.D., Yu, L., Kalachikov, S., Russo, J.J., Turro, N.J. & Ju, J.Y. Four-color DNA sequencing with 3'-*O*-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proc. Natl. Acad. Sci. USA* **105**, 9145-9150 (2008).
9. Ludwig, J. & Eckstein, F. Rapid and efficient synthesis of nucleoside 5'-*O*-(1-thiotriphosphates), 5'-triphosphates and 2',3'-cyclophosphorothioates using 2-chloro-4*H*-1,3,2-benzodioxaphosphorin-4-one. *J. Org. Chem.* **54**, 631-635 (1989).
10. Gaur, R.K., Sproat, B.S. & Krupp, G. Novel solid-phase synthesis of 2'-*O*-methylribonucleoside 5'-triphosphates and their alpha-thio analogs. *Tetrahedron Lett.* **33**, 3301-3304 (1992).

11. Caton-Williams, J., Lin, L.N., Smith, M. & Huang, Z. Convenient synthesis of nucleoside 5'-triphosphates for RNA transcription. *Chem. Commun.* **47**, 8142-8144 (2011).
12. Caton-Williams, J., Smith, M., Carrasco, N. & Huang, Z. Protection-free one-pot synthesis of 2'-deoxynucleoside 5'-triphosphates and DNA polymerization. *Org. Lett.* **13**, 4156-4159 (2011).

CHAPTER 3 Enzymatic Incorporation of PrdTTP and AzdTTP into DNA

3.1. Introduction

There are two DNA Click ligation strategies, for which two modified deoxynucleoside triphosphate were ready for examination: 3′-*O*-propargylthymidine-5′-triphosphate (PrdTTP, synthesised) and 3′-azido-3′-deoxythymidine-5′-triphosphate (AzdTTP, commercially available). As the most bio-compatible triazole Backbone B (Figure 3.1) was desired, the corresponding PrdTTP incorporation is more valuable. A Click ligation method, which generates this backbone, has further biotechnology applications, as discussed below.

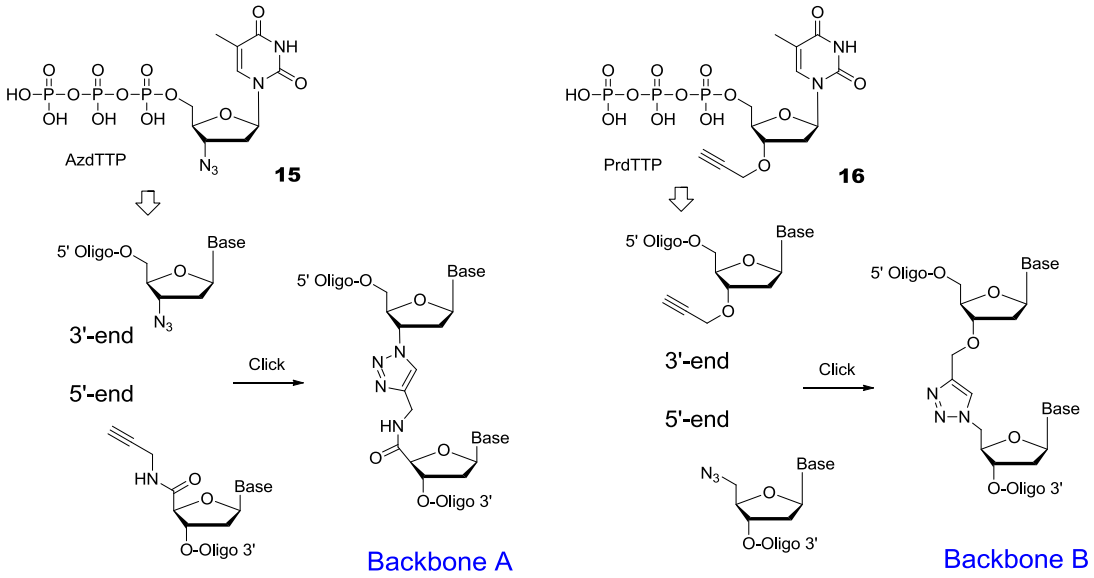


Figure 3.1: Two 3′-end modifications required for the formation of triazole backbones A and B on ssDNA formed by CuAAC reaction and the corresponding modified nucleoside triphosphate: 3′-azido-3′-deoxythymidine triphosphate (AzdTTP, **15**) and 3′-*O*-propargyl-2′-deoxythymidine triphosphate (**16**, PrdTTP).

Enzymes that utilise deoxynucleoside triphosphates include terminal deoxynucleotidyl transferase (TdT) [1], DNA-dependent DNA polymerases and RNA-dependent reverse-transcriptases [2]. Using TdT to introduce 3′-modifications is convenient, particularly in applications where the sequence of the target DNA is unknown, as it is a template-independent polymerase. Therefore, this was chosen for our initial examination. For dsDNA, the ideal substrate for TdT is a 3′-overhang, but it also works for blunt or recessed 3′ ends [3].

The most efficient 3´-terminus for TdT is deoxyguanosine [3], and the enzymatic reaction needs a divalent cation as a cofactor. If the dNTP used for 3´-polymerization is a purine, the best reaction rates are obtained by using Mg^{2+}. For pyrimidine dNTPs such as dTTP, the best choice is Co^{2+} [4].

3.2 Terminal deoxynucleotidyl transferase (TdT) experiments

3.2.1 Attempted incorporation of PrdTTP by TdT

The template-independent, terminal deoxynucleotidyl transferase enzyme (TdT, Promega® #M1871) was examined initially for incorporation of 3´-*O*-propargylthymidine-5´-triphosphate (PrdTTP). First TdT from Promega® was tested. In this case, synthetic oligodeoxynucleotides were used as substrates (ODN1 and ODN2, Table 8.2, Experimental section). The PAGE gel (Figure 3.2, Lane 1, 2) showed no difference between reactions with or without PrdTTP.

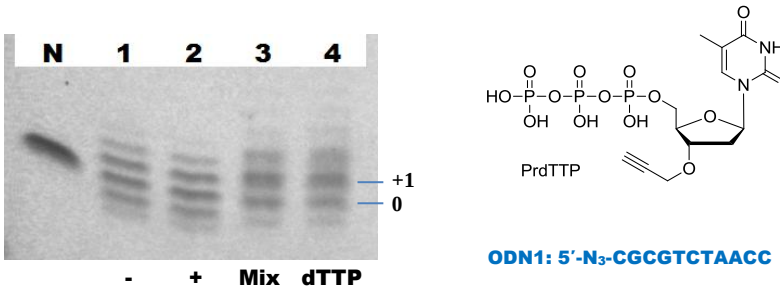


Figure 3.2: TdT(Promega®) 3´-elongation tests using PrdTTP.

Lane N: ODN1 (primer, 1.05 nmol) in water.

Lane 1: no triphosphate, Lane 2: 0.232 mM PrdTTP-Na^+, Lane 3: 0.232 mM PrdTTP-Na^+ and 0.320 mM dTTP,

Lane 4: 0.320 mM dTTP.

Composition: 42 µM primer (ODN1, 1.05 nmol), 1x Terminal Transferase Buffer (Promega®, 1 mM $CoCl_2$), 24 u TdT (Promega®, 30 u/µL) in total 25 µL, incubated at 37 °C for 4h.

An experiment using 5´-FAM-labelled oligonucleotide (ODN2) gave a similar result (Figure 3.3). It showed that the reaction using PrdTTP gave no significant difference to the negative control, while the positive control reaction using dTTP proceeded. PrdTTP sodium salt was prepared by passing the PrdTTP-TEA salt solution through DOWEX® 50WX8-200 (Na^+) ion-exchange resin. However, neither PrdTTP-TEA nor PrdTTP-Na^+ was incorporated by TdT from Promega®. Further tests using TdT from New England Biolabs® (NEB®

#M0315), incubating at 37 ℃ for 16 h (Figure 3.4), confirmed that TdT does not accept PrdTTP. The lower band of Lane 1 (Figure 3.4) was likely generated by pyrophosphorolytic dismutation, which is discussed below. This pyrophosphorolytic dismutation was inhibited by PrdTTP.

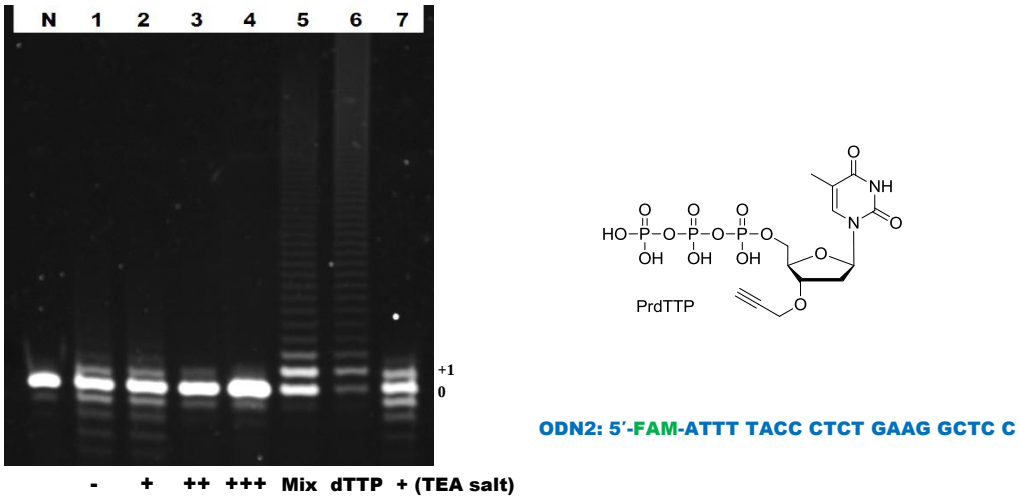


Figure 3.3: TdT(Promega®) 3′-elongation tests using PrdTTP.

Lane N: ODN2 (primer with 5′FAM, 47.0 pmol) in water.

Lane 1: no triphosphate, Lane 2: 0.288 mM PrdTTP-Na^+, Lane 3: 3.2 mM PrdTTP-Na^+, Lane 4: 16 mM PrdTTP-Na^+, Lane 5: 2.91 mM PrdTTP-Na^+ and 0.32 mM dTTP, Lane 6: 0.32 mM dTTP, Lane 7: 0.316 mM PrdTTP-TEA.

Composition: 1.88 µM primer (ODN2 with 5′-FAM, 47.0 pmol), 1x Terminal Transferase Buffer (Promega®, 1 mM $CoCl_2$), 15 u TdT (Promega®, 30 u/µL) in total 25 µL, incubated at 37 ℃ for 1 h.

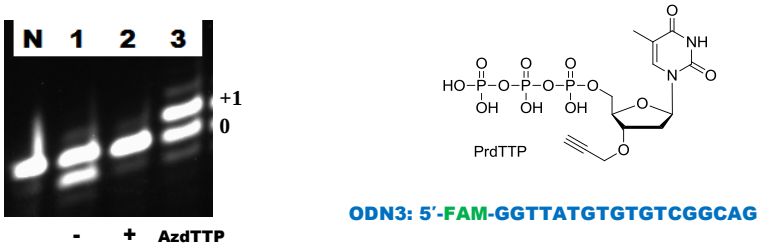


Figure 3.4: TdT(NEB®) elongation tests using PrdTTP and AzdTTP.

Lane N: ODN3 (primer with 5′-FAM, 115 pmol) in water.

Lane 1: no triphosphate, Lane 2: 0.4 mM PrdTTP-TEA^+ (8 nmol), Lane 3: 0.4 mM AzdTTP-Na^+ (8 nmol).

Composition: 5.75 µM primer (ODN3 with 5′-FAM, 115 pmol), 1x Terminal Transferase Reaction Buffer (NEB®, 10 mM $Mg(OAc)_2$), 0.25 mM $CoCl_2$, 15 u TdT (NEB®, 20 u/µL) in total 20 µL, incubated at 37 ℃ for 16 h.

3.2.2 Incorporation of AzdTTP by TdT

The 3′-azide labelling of DNA using TdT and commercially available AzdTTP was evaluated. The TdT from Promega® was tested first. Although this is an established general procedure [5], our result showed that the reaction proceeded slowly and the "+1" shift became significant only after 15 h (Figure 3.5, Lane 6). The mass spectrum (Figure 3.6) of the recovered "+1" bands (Figure 3.5, Lane 4-6) showed two major peaks on the HPLC chromatogram, one corresponding to the correct AZT addition product (+329) and another product suspected to be the result of 3′-phosphorylation of the oligonucleotide (+80). The kinase activity of TdT in the presence of triphosphates or dinucleoside-5′,5′-tetraphosphates has been reported [6,7]. Therefore, 3′-phosphorylation of the primer was likely to have occurred in our experiments.

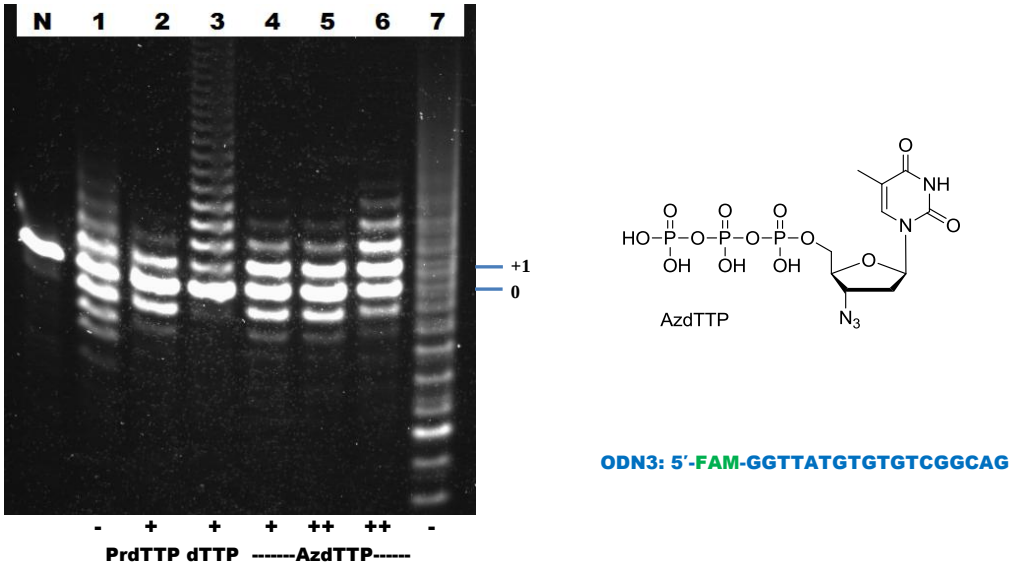


Figure 3.5: TdT(Promega®) elongation tests using PrdTTP (Lane 2) and AzdTTP (Lane 4-6).

Lane N: ODN3 (primer with 5′-FAM, 58 pmol) in water.

Lane 1 and 7: no triphosphate; Lane 2: 0.4 mM PrdTTP (8 nmol, TEA salt); Lane 3: 0.4 mM dTTP (8 nmol, Na^+ salt); Lane 4: 1 mM AzdTTP (Na^+ salt, 20 nmol); Lane 5 and 6: 2.5 mM AzdTTP (50 nmol).

Composition: 2.9 µM ODN3 (with 5′-FAM, 58 pmol), 1x Terminal Transferase Buffer (Promega®, 1 mM $CoCl_2$), 15 u TdT (Promega®) in total 20 µL.

Lane 1-5: incubated at 37 ℃ for 1 h; Lane 6 and 7: incubation at 37 ℃ for 15 h.

The overnight reaction with different concentrations of AzdTTP was also performed (Figure 3.7). However, the result suggested that varying the triphosphate concentration afforded no difference in the speed or efficiency of elongation, and a high concentration of AzdTTP did not significantly reduce the generation of ladders.

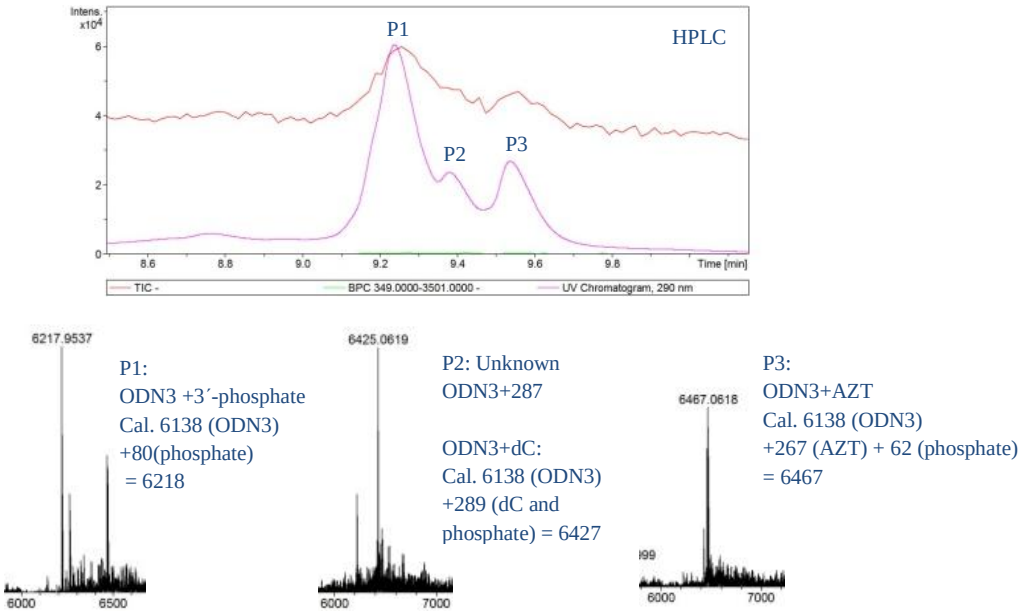


Figure 3.6 (Figure 9.12): MS of the TdT (Promega®) primer extension "+1" bands (Figure 3.5, Lane 4-6) using AzdTTP. The identity of the product "ODN3+287" is uncertain.

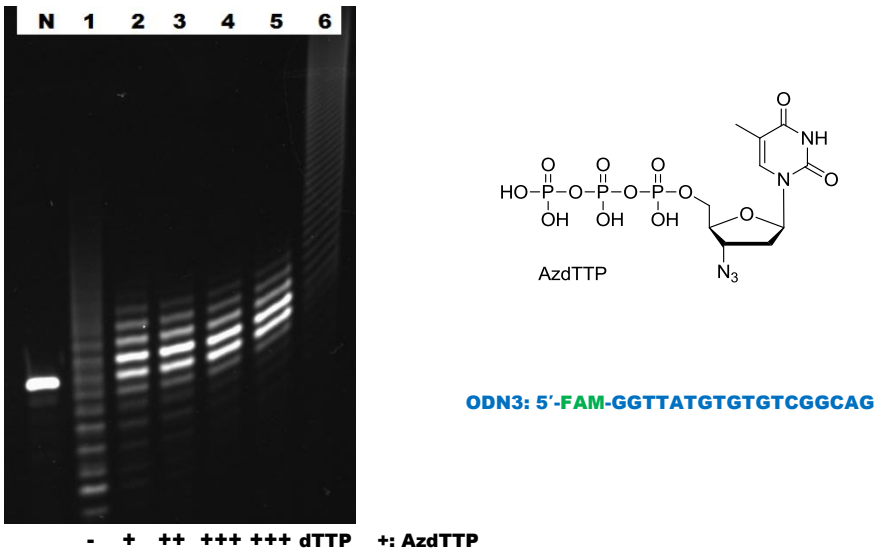


Figure 3.7: Overnight TdT(Promega®) elongation tests using different concentrations of AzdTTP.

Lane N: ODN3 (primer with 5′-FAM, 115 pmol) in water

Lane 1: no triphosphate; Lane 2: 0.2 mM AzdTTP (4 nmol); Lane 3: 0.4 mM AzdTTP (8 nmol); Lane 4 and 5 (same): 2.5 mM AzdTTP (50 nmol); Lane 6: 0.2 mM dTTP (4 nmol).

Composition: 5.75 µM ODN3 (with 5′-FAM, 115 pmol), 1x Terminal Transferase Buffer (Promega®, 1 mM $CoCl_2$), 15 u TdT (Promega®) in total 20 µL, incubated at 37 ℃ overnight.

It has been reported that TdT induces oligonucleotide pyrophosphorolytic dismutation [8]. During this process, TdT removes a 3′-nucleotide from one oligonucleotide and adds it to the

3´-end of another. This can occur because the elongation reaction is reversible, so one 3´-nucleotide can be removed, forming a triphosphate, which is subsequently incorporated into another strand. The ladders generated in our test resemble the ladders in the literature. It was also mentioned in this article that the commercially available yeast inorganic pyrophosphatase abolished the pyrophosphorolytic activity of TdT. Pyrophosphatase breaks down PPi, preventing it from transforming to triphosphate, while preserves the polymerase activity of TdT [8]. Yeast inorganic pyrophosphatase has not yet been tested by us.

The TdT from New England Biolabs (NEB®) was also evaluated under various conditions (Figure 3.8). Cobalt(II) ions were used, at various concentrations, because Co^{2+} changes the properties of the enzyme, enabling elongation of the blunt end and the 3´-recessive end of dsDNA by incorporation of dTTP (NEB® TdT manual). Our result shows that Co^{2+} accelerates the incorporation of AzdTTP (comparing Lane 2 and 4 of Figure 3.8). While the TdT from NEB® showed less pyrophosphorolytic dismutation, the efficiency for AzdTTP incorporation was still moderate. The "+1" bands of samples from lanes 4, 6, 8, 10 (Figure 3.8) were cut and combined for MS (Figure 3.9). This showed two major peaks in the HPLC chromatogram corresponding to the 3´-phosphorylation (+80) and +AZT products. The "+80" peak was the major one on the HPLC chromatogram comparing to the AZT-addition product peak. This result is similar to the one obtained using TdT from Promega® mentioned above. It indicates that the PAGE did not resolve the 3´-phosphorylation product and the +AZT product, which both showed as the "+1" band.

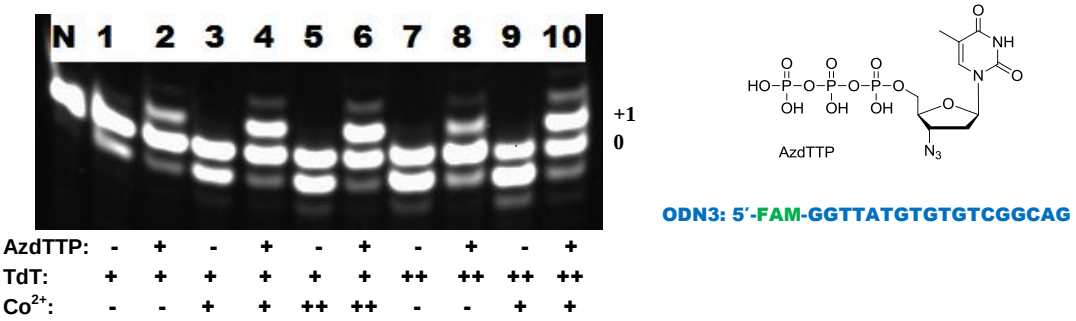


Figure 3.8: TdT(NEB®) elongation tests using AzdTTP.

Lane N: ODN3 (primer with 5´-FAM, 115 pmol) in water

Odd number: without triphosphate, even number: with 0.4 mM AzdTTP-Na^+ (8 nmol).

Lane 1, 2, 7, 8: 10 mM Mg^{2+};

Lane 3, 4, 9, 10: 10 mM Mg^{2+}; 0.25 mM Co^{2+};

Lane 5, 6: 10 mM Mg^{2+}, 1 mM Co^{2+};

Composition: 5.75 µM ODN3 (with 5´-FAM, 115 pmol); 1x Terminal Transferase Reaction Buffer (NEB®, 10 mM $Mg(OAc)_2$);

Lane 1-6: 15 u TdT (NEB®), Lane 7-10: 30 u TdT (NEB®) in total 20 µL; incubated at 37 ℃ for 16 h.

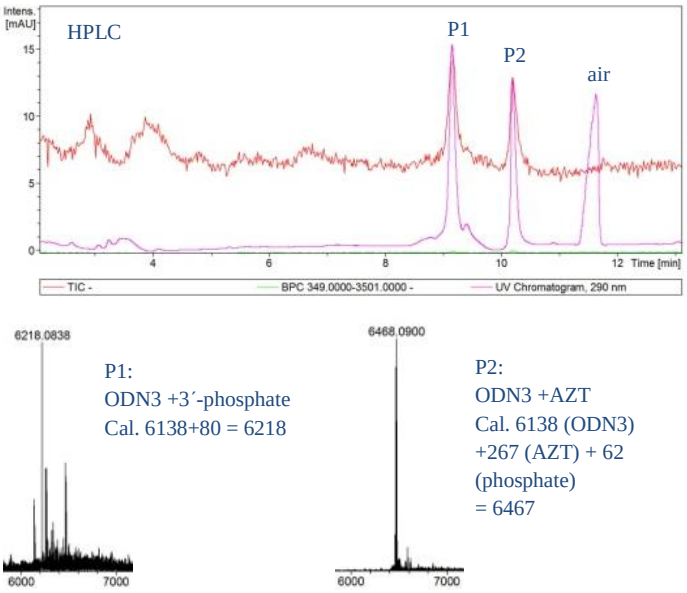


Figure 3.9 (Figure 9.13): MS of the TdT (NEB®) primer extension "+1" bands (Figure 3.8) using AzdTTP.

3.3 Screening template-dependent DNA polymerases for PrdTTP incorporation

To achieve PrdTTP incorporation into DNA, several available DNA polymerases including Klenow Fragment (DNA Polymerase I, Large (Klenow) Fragment, NEB® #M0210), T7 DNA polymerase (NEB® #M0274) and GoTaq™ polymerase (Promega® #M3171) were screened. The results (Figure 3.10) showed that both Klenow Fragment and GoTaq™ polymerase accept dTTP but not PrdTTP.

For the T7 polymerase reaction, we observed a strong degradation of the primer, possibly caused by the strong exonuclease activity of the T7 DNA polymerase [9]. DNA Polymerase I, Large (Klenow) Fragment and T7 DNA polymerase possess 3′→5′ exonuclease activity (Table 3.1), therefore they are not recommended by the manufacturers for labelling using triphosphates, because the exonuclease can remove the label, leading to poor labelling efficiency. They were examined by us, only because they are readily available in our lab.

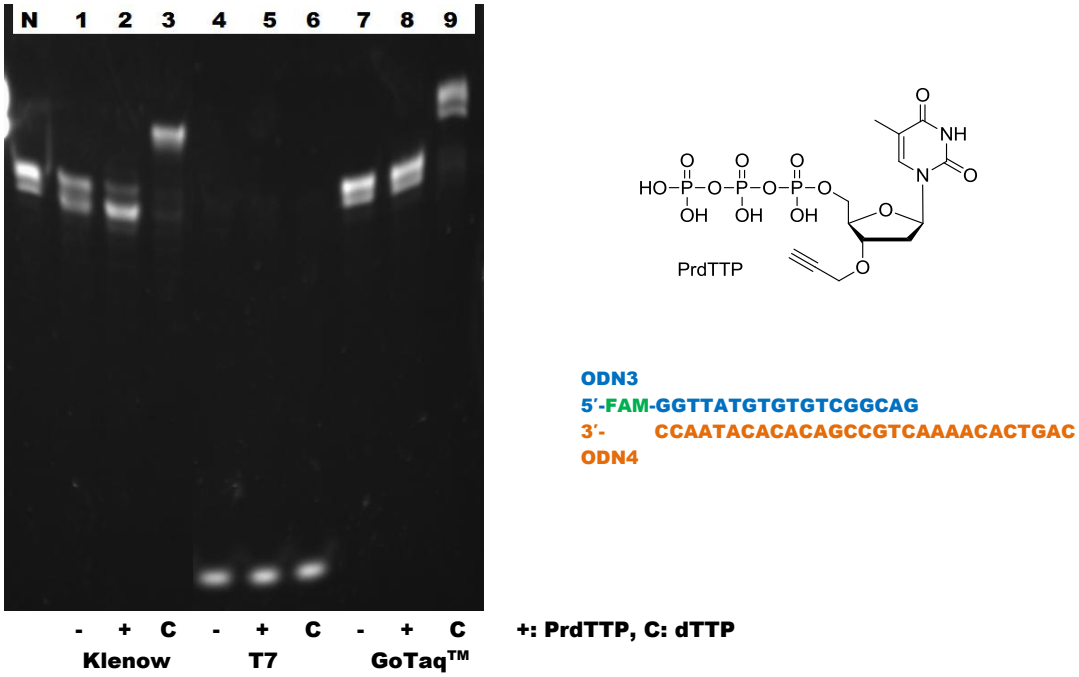


Figure 3.10: Primer extension tests using available DNA polymerases and PrdTTP.

Lane N: ODN3 (primer, 40 pmol), ODN4 (template, 40 pmol) in water.

Lane 1, 4, 7: Without triphosphate; Lane 2, 5, 8: with 1.0 mM PrdTTP-Na^+ (40 nmol); Lane 3, 6, 9: with 0.2 mM canonical dTTP (8 nmol).

Composition: 1.0 μM primer (ODN3 with 5′-FAM, 40 pmol), 1.0 μM template (ODN4, 40 pmol) in total 40 μL.

The detailed reaction conditions are described in the Experimental section.

	3′→5′ exonuclease	5′→3′ exonuclease	Note
TdT	No* [10]	No	*Catalyses pyrophosphorolytic dismutation
DNA Polymerase I, Large (Klenow) Fragment	Yes	No	The fragment lost 5′→3′ exonuclease activity (NEB®)
T7 DNA polymerase	Yes, Strong [9]	No [11]	Not recommended for labelling (NEB®)
GoTaq™ polymerase	No	No	Possess 5′→3′ flap endonuclease activity **
Therminator™ II DNA polymerase	No [12,13]	No	Discussed in the next section

Table 3.1: Summary of exonuclease properties of TdT and DNA polymerases as stated by the manufacturers.

**: the 5′→3′ flap endonuclease activity degrades secondary primers at the double-strand region the polymerase encountered during replication (NEB®).

3.4 9°N$_m$ DNA polymerase primer extension using PrdTTP

The thermophilic 9°N$_m$ DNA polymerase (Therminator™ II DNA polymerase, NEB® #M0266) was also tested. This enzyme was reported to incorporate 3´-*O*-allyl fluorescently-labelled dNTP, mentioned in the Introduction (Figure 1.27) [14,15]. The 3´-*O*-allyl modification is structurally similar to 3´-*O*-propargyl modification, suggesting that the 3´-*O*-propargyl modification might be tolerated by 9°N polymerase (exo$^-$, A485L/Y409V). It was discovered that it possesses excellent activity towards our modified nucleoside triphosphate TEA salt (Figure 3.11).

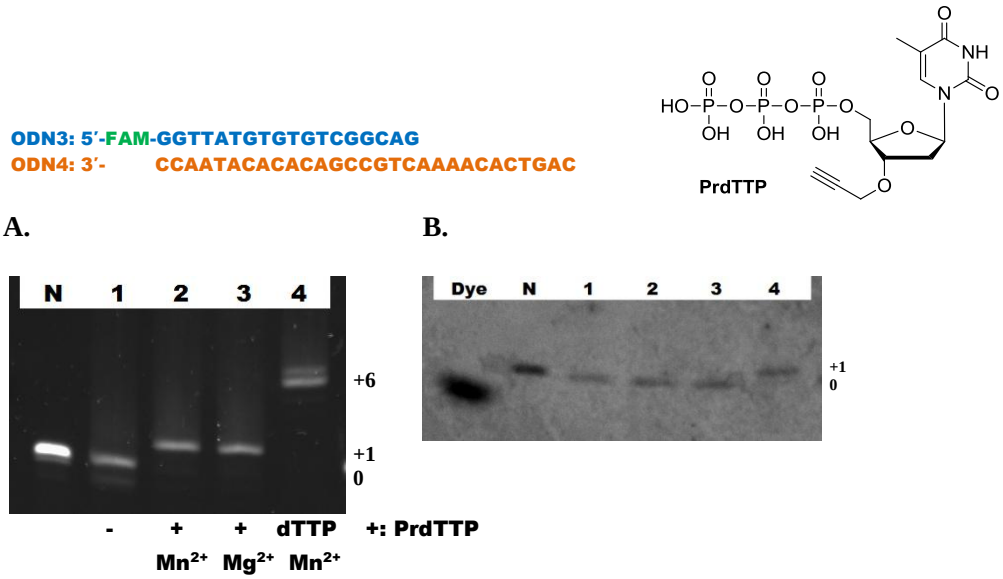


Figure 3.11: PCR using Therminator™ II DNA polymerase and PrdTTP.
A: 5´-FAM labelled primer on the gel visualized by transillumination. **B:** The same gel visualized by UV short wavelength absorption showing the template strand.

Lane N: ODN3 (primer, 33 pmol), ODN4 (template, 66 pmol) in water.
Lane 1: no triphosphate, 0.2 mM $MnCl_2$; Lane 2: 80 µM PrdTTP-TEA (1.6 nmol), 0.2 mM $MnCl_2$; Lane 3: 80 µM PrdTTP-TEA (1.6 nmol), 0.2 mM $MgCl_2$; Lane 4: 80 µM dTTP-Na^+ (1.6 nmol), 0.2 mM $MnCl_2$.
Reaction composition: 1.65 µM primer (ODN3, 33 pmol), 3.30 µM template (ODN4, 66 pmol), 1x ThermoPol® (0.2 mM $MgSO_4$, NEB®) buffer or 1x ThermolPol II®-Mn^{2+} (0.2 mM $MnCl_2$) buffer, 6 u Therminator™ II DNA pol. (NEB®) in total 20 µL. Thermocycling: 94 °C for 20 sec, 46 °C for 40 sec, 60 °C for 90 sec, 20 cycles [15]. T_m= 57 °C (calculated on NEB® website).

In the published protocols from a single research group [14-16], Mg^{2+} was interchanged with Mn^{2+} [15,16] in the buffer without any explanation. The reactivity of the enzyme in Mg^{2+}-containing buffer and Mn^{2+}-containing buffer (made by adding Mn^{2+} to the Mg^{2+}-free buffer)

was compared by us. However, no difference in the reaction was observed, and the primer extension in both buffers went to completion. The product was confirmed by MS to contain one 3′- propargylthymidine extension (Figure 3.12). The HPLC trace of this desalted PCR mixture (without gel purification) also showed that the reaction went to completion.

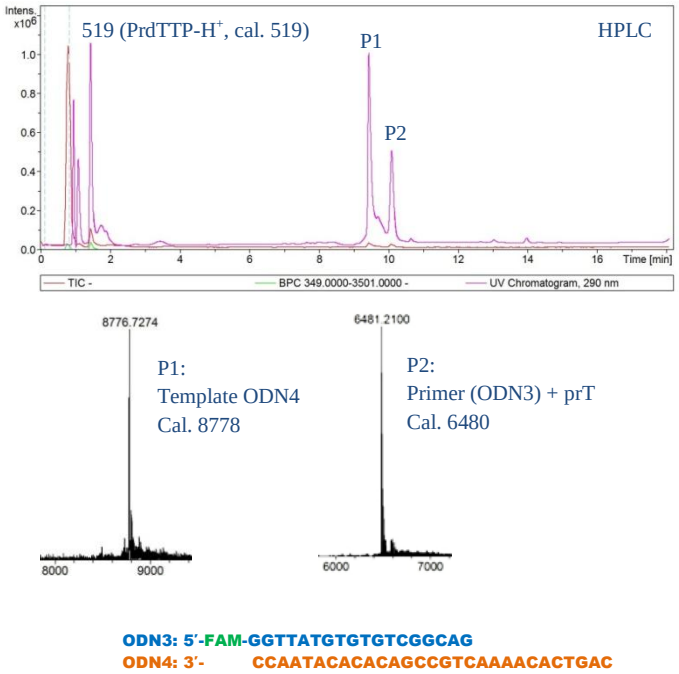


Figure 3.12 (Figure 9.14): MS analysis of the 9°N$_m$ DNA polymerase primer extension reaction mixture using 3′-*O*-propargylthymidine triphosphate.

The reaction mixture was desalted by NAP-10 then freeze-dried for M.S. PCR condition: 1.65 µM ODN3 (primer, 66.0 pmol), 3.30 µM ODN4 (template, 132 pmol), 80.0 µM PrdTTP-TEA$^+$ (3.20 nmol), 1x ThermoPol (0.2 mM Mg^{2+}) buffer, Therminator II (4 u, NEB®) in total 40 µL; Thermocycling: 94 ℃ for 20 sec, 46 ℃ for 40 sec, 60 ℃ for 90 sec, 20 cycles (method: PNAS2006) [15].

Interestingly, the control extension using canonical dTTP gave mainly two products: the primer with 6 additional "T" (instead of 4 consecutive "T" as templated) and the template with one additional "T" (MS, Figure 3.13). The template elongation was also observed by PAGE (Figure 3.11 B, Lane 4, Mn^{2+}-buffer used). It indicates 9°N$_m$ polymerase not only has the potential to introduce C·T mismatches under our conditions with 80 µM nucleoside triphosphate, but also possesses terminal transferase activity when using dTTP. This is consistent with the manufacturer's note (3′-terminal structure: blunt and 3′-A overhang, NEB®). However, 9°N$_m$ polymerase did not terminal-transfer the modified nucleotide (PrdTTP) as the template in the PCR mixture remained intact (Figure 3.12). The reason for this is uncertain.

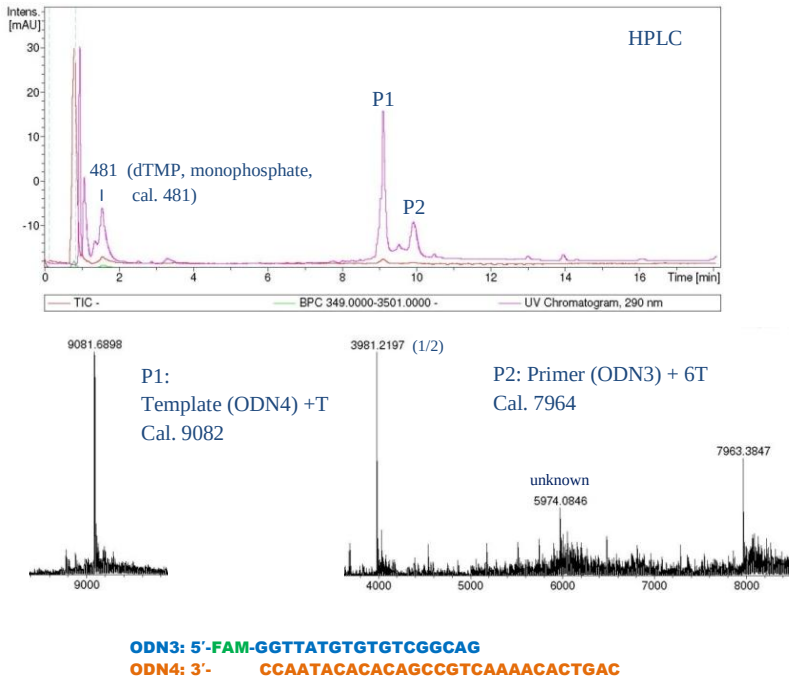


Figure 3.13 (Figure 9.15): MS analysis of the 9°N$_m$ DNA polymerase primer extension control products using dTTP.

The reaction mixture was desalted by NAP-10 then freeze-dried for M.S. PCR condition: ODN3 (primer, 66 pmol), ODN4 (template, 132 pmol), dTTP-Na^+ (3.2 nmol), 1x ThermoPol buffer (0.2 mM Mg^{2+}), Therminator II (4 u, NEB) in total 40 µL; Thermocycling: 94 ℃ for 20 sec, 46 ℃ for 40 sec, 60 ℃ for 90 sec, 20 cycles (method: PNAS2006) [15].

Additional optimization showed that 9°N$_m$ DNA polymerase elongation using PrdTTP went to completion with minimal amount of enzyme under PCR condition (Figure 3.14). It also works by simple incubation at 60 ℃ with lower efficiency (Figure 3.15). This enzyme solution smeared the oligonucleotide bands completely during PAGE when used in excess (10 u/10 µL reaction) (PAGE not shown).

As the 9°N$_m$ DNA polymerase works in a templated fashion, it can only generate dsDNA with blunt ends or 3′-recesses with 3′-modifications. The 3′-overhang is difficult to generate.

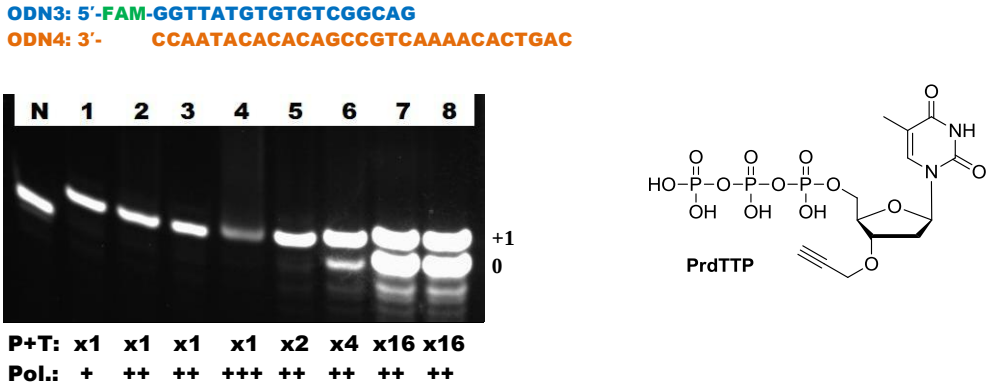


Figure 3.14: 9 ℕ$_m$ DNA polymerase evaluation using PrdTTP

Lane N: ODN3 (primer, 33 pmol), ODN4 (template, 66 pmol) in water.

Lane 1-4: 1.65 µM primer (33 pmol) and 3.30 µM template (66 pmol), 80 µM PrdTTP-TEA (1.6 nmol), 1x ThermoPol buffer (0.2 mM Mg^{2+}) in total 20 µL.

Lane 1: 1 u Therminator™ II DNA pol.(NEB®); Lane 2: 2 u DNA pol.; Lane 3: 2 u DNA pol., product was gel-filtered (NAP-10), Lane 4: 6 u DNA pol..

Lane 5-8: increasing oligonucleotide conc. with 2 u DNA pol. in total 20 µL.

Lane 5: (Primer+Template conc.) x 2 (3.30 µM primer and 6.60 µM template), 80 µM PrdTTP-TEA (1.6 nmol);

Lane 6: (Primer+Template conc.) x 4, 80 µM PrdTTP-TEA (1.6 nmol);

Lane 7: (Primer+Template conc.) x 16, 80 µM PrdTTP-TEA (1.6 nmol);

Lane 8: (Primer+Template conc.) x 16, 160 µM PrdTTP-TEA (3.2 nmol).

Thermocycling: 94 ℃ for 20 sec, 46 ℃ for 40 sec, 60 ℃ for 90 sec, 20 cycles (method: PNAS2006) [15].

Figure 3.15: 9 ℕ$_m$ DNA polymerase primer extension at 37 ℃ using PrdTTP.

Lane N: ODN3 (primer, 33 pmol), ODN4 (template, 66 pmol) in water.

Composition: 1.65 µM primer (33 pmol) and 3.30 µM template (66 pmol), 80 µM PrdTTP-TEA (1.6 nmol), 1x ThermoPol buffer (0.2 mM Mg^{2+}), 2 u Therminator™ II DNA pol. (NEB®) in total 20 µL.

Lane 1: extension at 60 ℃ for 1 min.

Lane 2-6: (Preheated) 94 ℃ for 20 sec, 46 ℃ for 40 sec, and 60 ℃ extension for the following time:

Lane 2: 1 min, lane 3: 5 min, lane 4: 10 min, lane 5: 20 min, Lane 6: 30 min.

3.5 9°N$_m$ DNA polymerase experiments using AzdTTP

AzdTTP is a common chain terminator that is incorporated by DNA polymerases and RNA polymerases [17]. To find a different enzyme that incorporates AzdTTP better than TdT can, 9°N$_m$ DNA polymerase (Therminator^{TM} II) was tested. Both PAGE (Figure 3.16) and MS analysis (Figure 3.17) of the desalted PCR mixture showed that the AZT was successfully incorporated and primer extension went to completion in a templated fashion.

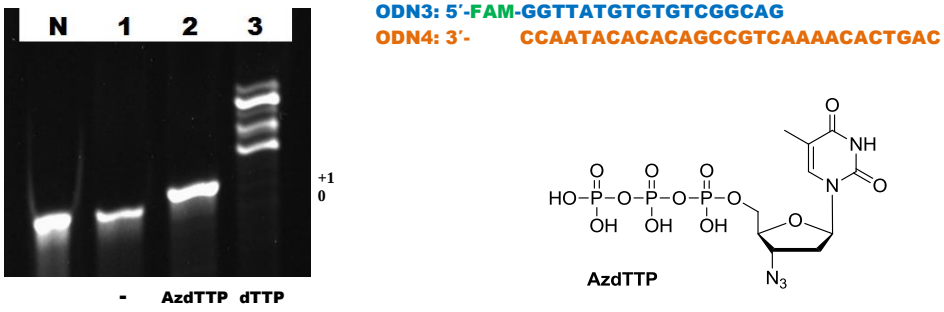


Figure 3.16: 9°N$_m$ DNA polymerase (Therminator^{TM} II) elongation using AzdTTP.

Lane N: ODN3 (primer, 66 pmol), ODN4 (template, 132 pmol) in water.

Lane 1: no triphosphate, Lane 2: 80 µM AzdTTP-Na$^+$ (1.6 nmol); Lane 3: 80 µM dTTP-Na$^+$ (1.6 nmol). Composition: 3.3 µM primer (66 pmol), 6.6 µM template (132 pmol), 1x ThermoPol buffer (0.2 mM Mg^{2+}), 2 u Therminator II (NEB®) in total 20 µL. Thermocycling (PNAS2006): 94 °C for 20 sec, 46 °C for 40 sec, 60 °C for 90 sec, 20 cycles [15]. T_m= 57 °C (calculated on NEB® website).

The terminal transferase activity of 9°N$_m$ DNA polymerase was further investigated. A similar situation to PrdTTP was observed (Figure 3.12), where HPLC-MS showed no terminal addition to the template strand when AzdTTP was used (Figure 3.17). A blunt end dsDNA sample, without 5′-FAM, was used to further confirm that the 9°N$_m$ DNA polymerase terminally transfers neither PrdTTP nor AzdTTP to form 3′-overhangs (both Mg^{2+} and Mn^{2+} buffer were tested, data not shown). The control reaction using dTTP showed that 9°N$_m$ DNA polymerase extended the template by 1 nucleotide by terminal transfer in Mg^{2+}-buffer as mentioned above (Figure 3.13).

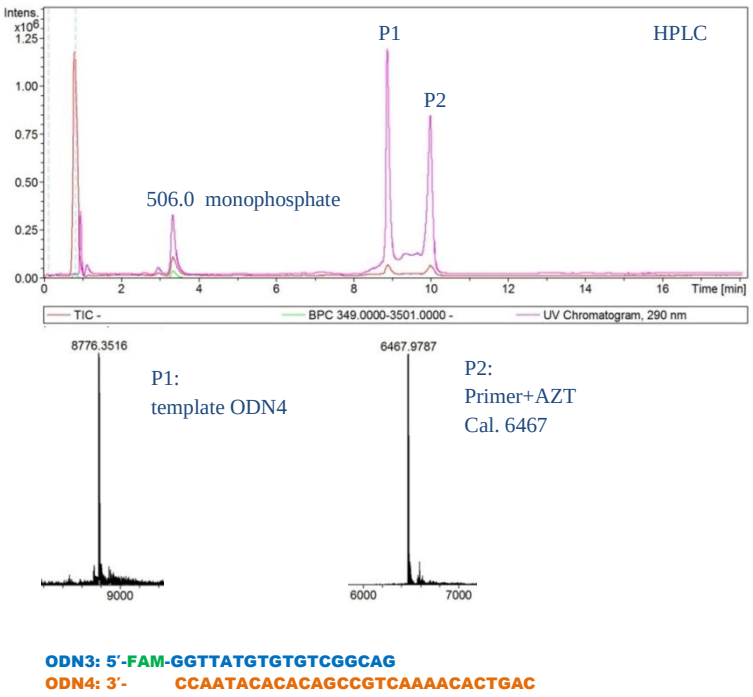


Figure 3.17: MS analysis of the 9°N$_m$ DNA polymerase (Therminator^{TM} II) elongation mixture using AzdTTP.

The reaction mixture was desalted by NAP-10 then freeze-dried for M.S. Composition: ODN3 (primer, 132 pmol), ODN4 (template, 264 pmol), AzdTTP-Na$^+$ (3.2 nmol), 1x ThermoPol buffer (0.2 mM Mg^{2+}), Therminator II (4 u, NEB) in total 40 µL; Thermocycling method: PNAS2006 [15].

3.6 Conclusions

For DNA 3′-labelling, TdT requires no template but was unable to incorporate PrdTTP. Although it can incorporate 3′-azido-3′-deoxythymidine-5′-triphosphate (AzdTTP) overnight, it also induced pyrophosphorolytic dismutation and generated a mixture of deletion and addition products. However, it was found that an engineered 9°N$_m$ polymerase (Therminator^{TM} II) was able to label a DNA primer with PrdTTP and AzdTTP in a templated manner. It was also discovered that Therminator^{TM} II catalysed "non-templated" terminal transferring of dTTP but not PrdTTP.

1. Isobe, M. et al. Chromosome localization of the gene for human terminal deoxynucleotidyltransferase to region 10q23-q25. *Proc. Natl. Acad. Sci. USA* **82**, 5836-5840 (1985).
2. Watson, J.D. *Molecular biology of the gene*, (2008).
3. Roychoudhury, R., Jay, E. & Wu, R. Terminal labeling and addition of homopolymer tracts to duplex DNA fragments by terminal deoxynucleotidyl transferase. *Nucleic Acids Res.* **3**, 101-116 (1976).
4. Chang, L.M.S. & Bollum, F.J. Molecular-biology of terminal transferase. *CRC Critical Reviews in Biochemistry* **21**, 27-52 (1986).
5. Nickel, W., Austermann, S., Bialek, G. & Grosse, F. Interactions of azidothymidine triphosphate with the cellular DNA-polymerases alpha, delta, and epsilon and with DNA primase. *J. Biol. Chem.* **267**, 848-854 (1992).
6. Krayevsky, A.A., Victorova, L.S., Arzumanov, A.A. & Jasko, M.V. Terminal deoxynucleotidyl transferase: Catalysis of DNA (oligodeoxynucleotide) phosphorylation. *Pharmacol. Ther.* **85**, 165-173 (2000).
7. Arzumanov, A.A., Victorova, L.S., Jasko, M.V., Yesipov, D.S. & Krayevsky, A.A. Terminal deoxynucleotidyl transferase catalyzes the reaction of DNA phosphorylation. *Nucleic Acids Res.* **28**, 1276-1281 (2000).
8. Anderson, R.S., Bollum, F.J. & Beattie, K.L. Pyrophosphorolytic dismutation of oligodeoxynucleotides by terminal deoxynucleotidyltransferase. *Nucleic Acids Res.* **27**, 3190-3196 (1999).
9. Hori, K., Mark, D.F. & Richardson, C.C. Bacteriophage-T7 deoxyribonucleic-acid replication invitro .15. Deoxyribonucleic-acid polymerase of bacteriophage-T7 - characterization of the exonuclease activities of the gene-5 protein and the reconstituted polymerase. *J. Biol. Chem.* **254**, 1598-1604 (1979).
10. Fowler, J.D. & Suo, Z. Biochemical, structural, and physiological characterization of terminal deoxynucleotidyl transferase. *Chem. Rev.* **106**, 2092-2110 (2006).
11. Sambrook, J. & Russell, D.W. *Molecular cloning: A laboratory manual*, (2001).
12. Gardner, A.F. & Jack, W.E. Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. *Nucleic Acids Res.* **27**, 2545-2553 (1999).
13. Gardner, A.F. & Jack, W.E. Acyclic and dideoxy terminator preferences denote divergent sugar recognition by archaeon and Taq DNA polymerases. *Nucleic Acids Res.* **30**, 605-613 (2002).
14. Ruparel, H. et al. Design and synthesis of a 3'-*O*-allyl photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis. *Proc. Natl. Acad. Sci. USA* **102**, 5932-5937 (2005).

15. Ju, J.Y. et al. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc. Natl. Acad. Sci. USA* **103**, 19635-19640 (2006).
16. Guo, J. et al. Four-color DNA sequencing with 3'-*O*-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proc. Natl. Acad. Sci. USA* **105**, 9145-9150 (2008).
17. Horwitz, J.P. et al. Substrates for cytochemical demonstration of enzyme activity. I. Some substituted 3-indolyl D-glycopyranosides. *J Med Chem* **7**, 574-575 (1964).

CHAPTER 4 Applications of DNA Click ligation

4.1 Introduction

In order to test the robustness of the enzymatic incorporation of AzdTTP, as mentioned in the previous chapter, the Click-ligation approach was utilised to construct DNA structures. A sealed dsDNA structure was previously made using synthetic dsDNA, with 5′-alkyne and 3′-azide on each strand [1]. The alkyne and azide on each end of the dsDNA were then "clicked" by CuAAC to seal the ends (Figure 4.1). As the Click ligation reaction on this structure can be conveniently monitored by gel electrophoresis, this structure was selected to evaluate our Click ligation approach utilising the efficiency of the AzdTTP incorporation by Yeast PAP.

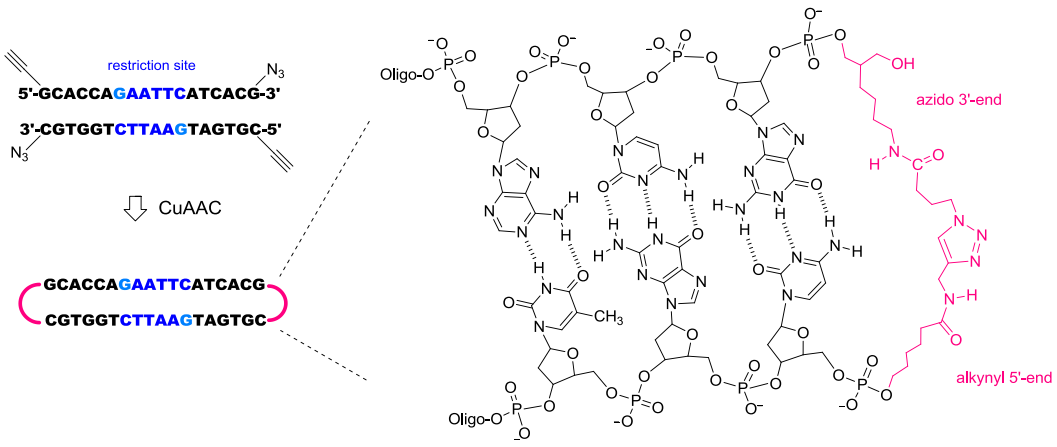


Figure 4.1: The sealed dsDNA structure made in-house, using synthetic oligonucleotides [1]

4.2 Introduction of AzdTTP into dsDNA

Two dsDNA structures (Figure 4.2) with a restriction site in the middle were designed with the 3′-recess for AzdTTP incorporation (ODN5/ODN6, ODN7/ODN8). One extra deoxyadenosine was added at the two 5′-overhangs in oligonucleotides ODN7/ODN8 to keep the C12 spacer and hexynol group further away from the amplification site to minimise their possible interference with the polymerase. Due to the short length of the two strands, the duplex was expected to form (T_m = 57 ℃, calculated on NEB® website) during an annealing step down to 46 ℃. The PAGE result (Figure 4.2) showed that the duplex of ODN7/ODN8 gave the better extension result (Lane 8). The two upper bands of lane 4 and 8 were cut and

the extension products, each containing two ssDNAs with 3´-azides, were confirmed by MS analysis (Figures 4.3, 4.4).

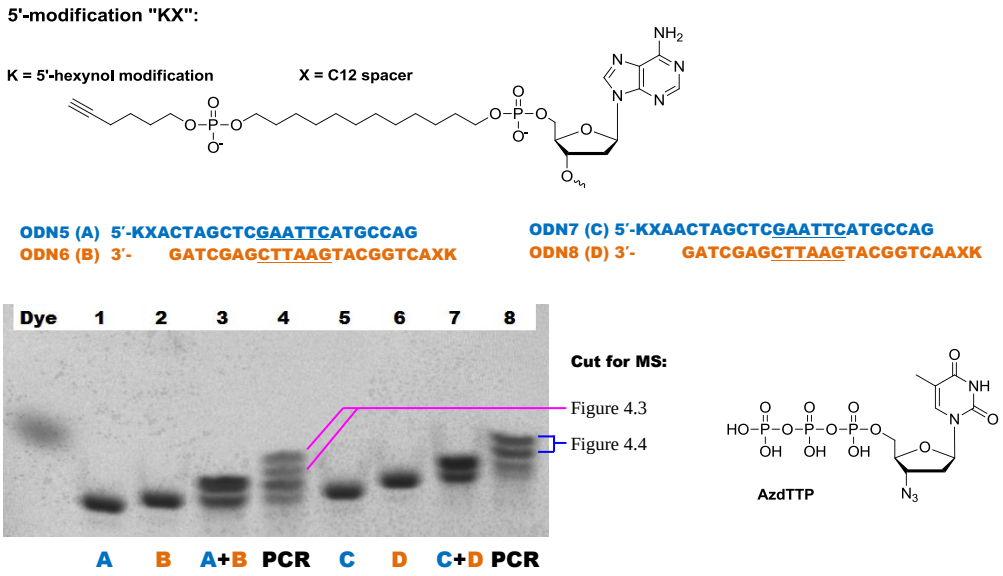


Figure 4.2: Incorporation of the AzdTTP into the dsDNA sample containing 5´-alkyne at both ends.

Lane 1-4: ODN5, ODN6 test (single A overhang)

Lane 1: ODN5 (0.4 nmol); Lane 2: ODN6 (0.4 nmol); Lane 3: ODN5 (0.4 nmol) + ODN6 (0.4 nmol); Lane 4: PCR products of ODN5/ODN6.

Lane 5-8: ODN7, ODN8 test (double A overhang)

Lane 5: ODN7 (0.4 nmol); Lane 6: ODN8 (0.4 nmol); Lane 7: ODN7 (0.4 nmol) + ODN8 (0.4 nmol); Lane 8: PCR products of ODN7/ODN8.

Composition: 10 µM each strand (0.4 nmol), 200 µM AzdTTP-Na^+ (8 nmol), 1x ThermoPol (Mg^{2+}) buffer, 6 u Therminator II (NEB®) in total 40 µL, Thermocycling: 94 ℃ for 20 sec, 46 ℃ for 40 sec, 60 ℃ for 90 sec, 20 cycles.

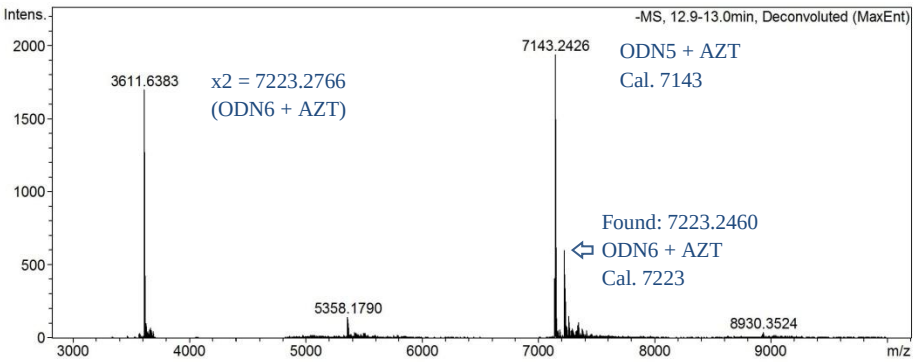


Figure 4.3 (Figure 9.17): MS analysis of the 9 ℕ$_m$ DNA polymerase primer extension of ODN5/ODN6 using AzdTTP (Figure 4.2, lane 4, two upper bands).

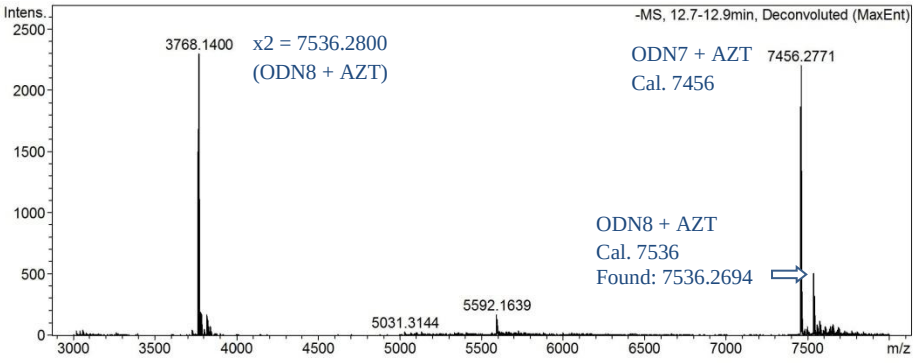


Figure 4.4 (Figure 9.18): MS analysis of the 9 ℃N_m DNA polymerase primer extension of ODN7/ODN8 using AzdTTP (Figure 4.2, lane 8, two upper bands).

4.3 End-sealing of the modified dsDNA structure

After AzdTTP incorporation, the dsDNA product from ODN7/ODN8 contained both alkyne and azide groups at each double-stranded end (Figure 4.5, Lane P). The two ends were sealed by a CuAAC reaction. The presence of sealed dsDNA was confirmed by PAGE (Figure 4.5, Lane C) and MS (Figure 4.6). This structure was further verified by EcoRI restriction enzyme cleavage (Figure 4.5, Lane C*) and also monitored by MS (Figure 4.8).

The HPLC-MS analysis of the CuAAC reaction crude products (Figure 4.6) showed two HPLC peaks (P1, P3) with the correct mass of the end-sealed product (i.e. plus two AZT, "+2"). It is likely that one of the two "+2" peaks migrating slower (closer to the middle small peak) represented the partially "clicked" product (one end "clicked" and the other end "unclicked") as the CuAAC reaction causes no change to the mass. This indicates that the CuAAC reaction of the double-AZT labelled duplex did not go to completion. The middle small peak (P2) appeared as a shoulder and had the mass of the double strand plus one AZT. It is expected that one end of the double-strand was sealed but the other end without AZT did not seal. This means the enzymatic incorporation did not go to completion.

A portion of the sealed dsDNA (CuAAC product) was used as a reference (Figure 4.5, Lane C), while the rest was purified by PAGE (Figure 4.7) for the restriction enzyme cleavage experiment. The EcoRI cleavage of the CuAAC sealed duplex generated two single strands, which migrated similarly to the original starting strands ODN7 and ODN8 (Figure 4.5, Lane C*). The remaining upper strand (Lane C*) was likely to be the result of incomplete

cleavage. The EcoRI cleavage gave 5′-phosphate and 3′-OH at the cleavage site. The cleavage products (two bands) were retrieved from the polyacrylamide gel and analysed by HPLC-MS (Figure 4.8). The HPLC trace showed only one peak, which means the HPLC did not separate the two bands. The reason for this could be that the two cleaved strands with the triazole in the middle have the same base composition. The two cleaved strands also gave the same correct mass.

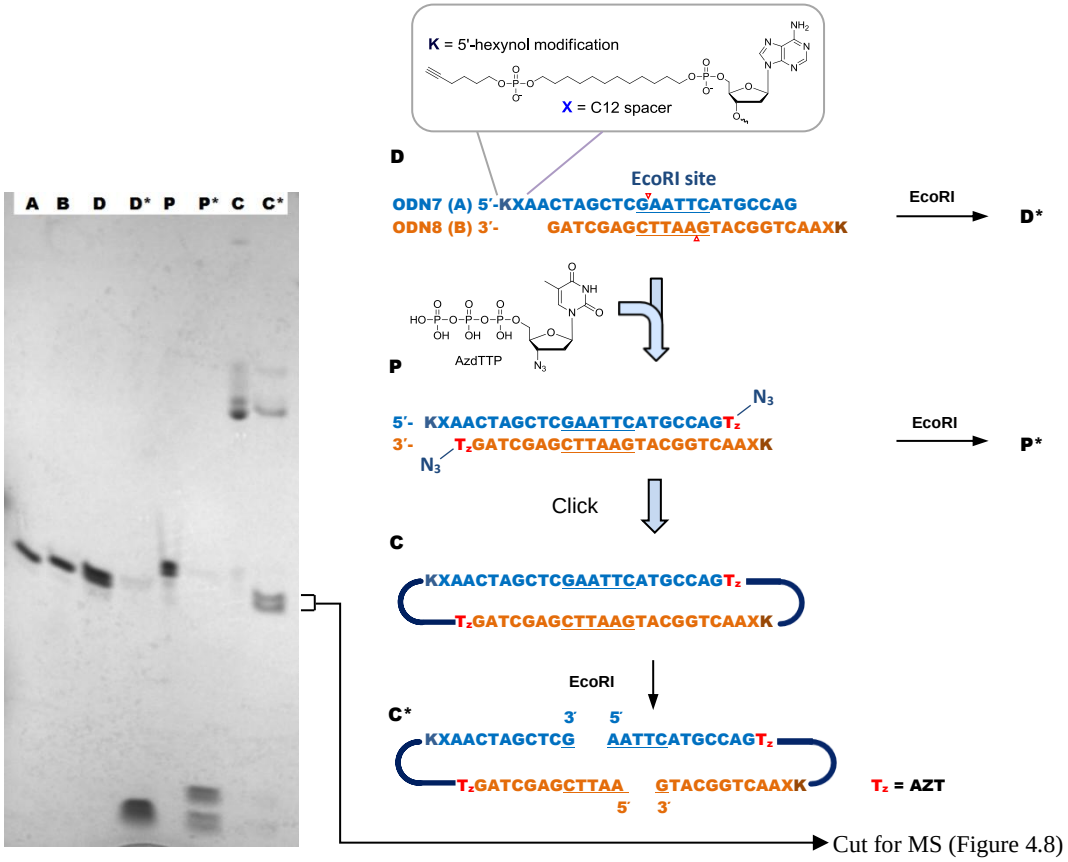


Figure 4.5: AzdTTP-PCR labelling and CuAAC end-seal products, followed by EcoRI cleavage. The EcoRI site is underlined.

Lane A: ODN7 (0.4 nmol) in water, Lane B : ODN8 (0.4 nmol) in water;

Lane D: dsDNA of ODN7 (0.4 nmol) and ODN8 (0.4 nmol) in 0.2 M NaCl (annealed by heating to 80 ℃ then cooling to r.t. but showed no significant difference to the directly loaded dsDNA sample in water - Figure 4.7, Lane D).

Lane P: Thermocycling products; composition: 10 µM (0.4 nmol) each strand, 200 µM (8 nmol AzdTTP-Na^{+}), 1x ThermoPol (Mg^{2+}) buffer, 6 u Therminator™ II (NEB®) in total 40 µL. Thermocycling: 94 ℃ for 20 sec, 46 ℃ for 40 sec, 60 ℃ for 90 sec, 20 cycles (PNAS2006).

Lane C: (Click) same divided PCR products, cyclised by CuAAC then desalted by NAP-10;

Lane D*, P*, C*: corresponding EcoRI (Promega® #R6011) cleavage products,

Lane D*: duplex cleaved by EcoRI, 0.4 nmol each, annealed before cleavage,

Lane P*: PCR products (0.4 nmol each) cleaved by EcoRI,

Lane C*: EcoRI cleavage of the Click product purified by PAGE (Figure 4.7).

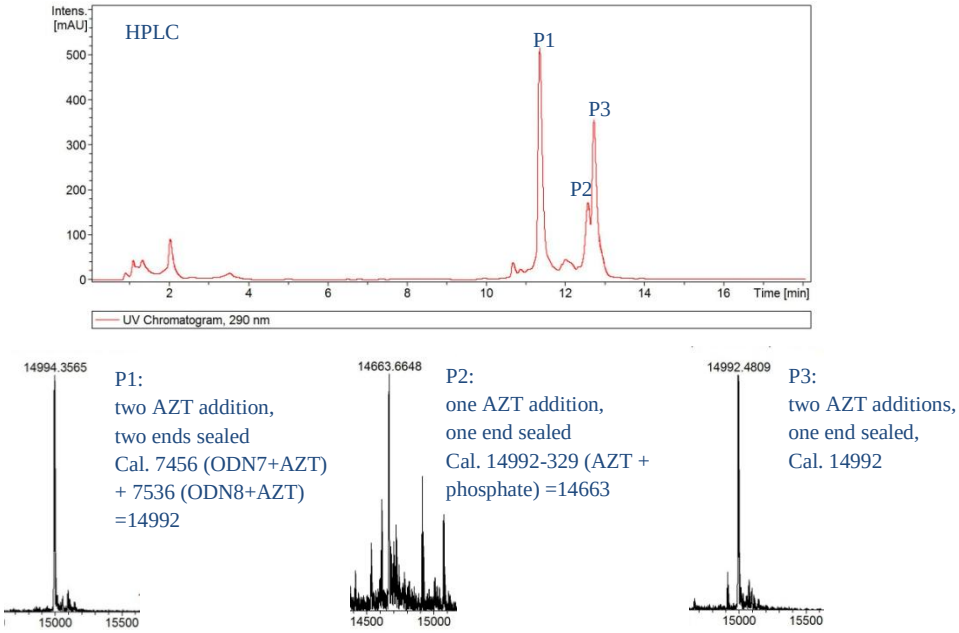
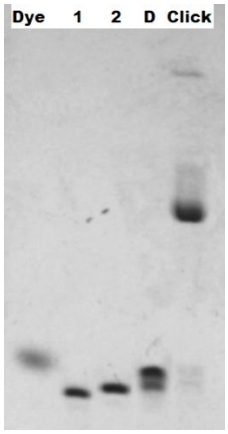


Figure 4.6 (Figure 9.19): MS analysis of the AzdTTP-PCR-CuAAC crude reaction product of ODN7/ODN8, desalted by NAP-10 without PAGE purification.



Band cut and retrieved for the EcoRI cleavage.

Figure 4.7: PAGE purification of CuAAC end-seal products for EcoRI cleavage (Lane C* in Figure 4.5).

Lane 1, 2, D: oligonucleotides in water. Lane 1: ODN7; Lane 2: ODN8; Lane D: ODN7 + ODN8;
Lane Click: 3/4 of the PCR product (x4 scale, 1.6 nmol each strand) was used for CuAAC (25 equiv. Cu^{I}, 1 h, r.t.), 2/3 of the CuAAC product was loaded for this lane. The rest was used for Lane C of Figure 4.5 (See experimental section 8.6 for details).

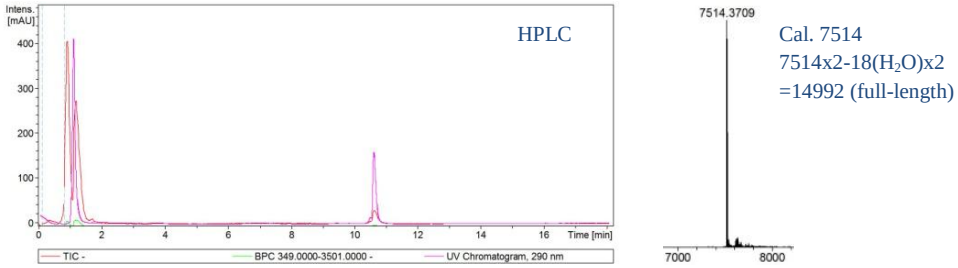


Figure 4.8 (Figure 9.20): HPLC-MS of the EcoRI cleavage product of the end-sealed duplex. Required mass: 7514, found mass: 7514.3709.

4.4 Conclusions

Introducing azide group on the 3′-end of the synthetic dsDNA was efficient. According to the results in the previous chapter, 3′-alkyne could also be introduced with similar efficiency. This suggests that the alkyne/azide group could be enzymatically introduced into long PCR products to produce end-sealed long dsDNAs, which are too long to be obtained by oligonucleotide synthesis. The alkyne or azide modification on the primer can be utilised to introduce the modifications at the 5′-end. Successful 3′-labelling of long PCR products enables the presence of both alkyne and azide groups at each end of a dsDNA PCR product. The alkyne and azide will then be "clicked" to seal each end of the dsDNA.

Typically, the dsDNA produced by PCR has blunt ends or single nucleotide 3′-overhangs. To facilitate the introduction of alkyne/azide to these ends, an additional ssDNA template, which is complementary to the 3′-end of target, is needed for this polymerization reaction. The end-sealing of the PCR product provides certain advantages over the unsealed dsDNA, because it will protect the linear dsDNA structure against exonuclease digestion *in vivo*. However, a similar property can also be achieved using a 2′, 3′-dideoxynucleotide triphosphate with an alkyne/azide modification on the base, as the base-modified nucleotide triphosphate is a much better substrate for DNA polymerases [2]. The uniqueness of the 2′- or 3′-modification is that it can generate triazole backbones that can be read-through by enzymes such as DNA polymerases [3]. A DNA Click ligation application utilising this property is described in the following section.

4.5 Future Work -I: Click ligation for sub-cloning

A strategy was designed, which uses Click ligation for TA-cloning (Figure 4.6). It uses Click chemistry instead of DNA ligase for the initial ligation between a PCR fragment and a plasmid vector. It generates triazole backbone B, which can be read-through by DNA polymerases and transcriptases, as discussed above. Two preconditions need to be fulfilled before CuAAC ligation can occur. Firstly, the single "A" 3′-overhangs need to be generated at both ends of the dsDNA insert with 5′-azide modifications. It is expected that this will be done by Taq polymerase as the normally observed single "A" addition is not expected to be prevented by the 5′-azide on the other strand. Secondly, the 3′-*O*-propargylthymidine needs to be introduced on to the two blunt ends of the vector as 3′-overhangs. This is more challenging, as the vector could be a plasmid from 1 to 1000 kb in length, and the introduction cannot be monitored by MS. It is possible to PCR amplify the whole plasmid, but the generation of a 3′-*O*-propargylthymidine overhang is still difficult, as the 9°N$_m$ DNA polymerase cannot terminally transfer PrdTTP, only dTTP. The first strategy involves using two ssDNA templates, which hybridise over each blunt end of the plasmid, *via* strand-invasion. The addition of PrdTTP will then be templated, and should be successful. After PCR, the template can be removed by gel electrophoresis or by simply attaching a modification on the template such as biotin to ensure its removal *via* treatment with streptavidin-labelled beads, for example.

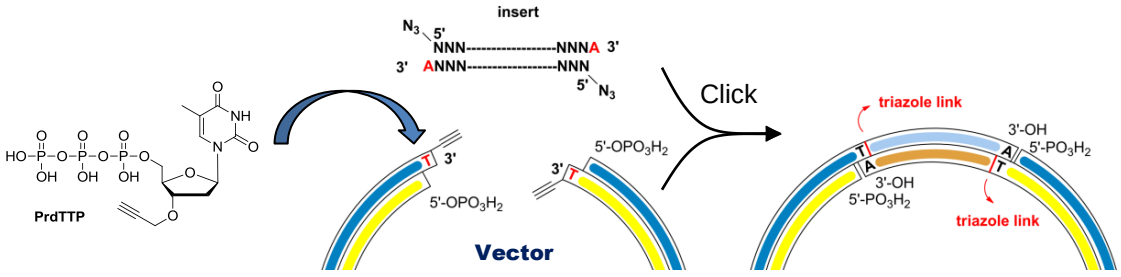


Figure 4.6: Modified TA-cloning procedure.

The second strategy (Figure 4.7) utilises restriction enzymes such as XhoI to generate two 5′-overhangs, which serves as the template for PrdTTP addition on the plasmid. The 3′-recess on

the insert is difficult to generate as the Mung Bean Nuclease can only remove a 5´-overhang and cannot digest further. The two single-stranded sections can be PCR amplified separately, but with a 5´-phosphorylated forward primer or reverse primer, so that the complementary strand can be digested by Lambda Exonuclease (digest the 5´-phosphorylated strand). By annealing the two separately prepared single strands, it is possible to generate the dsDNA with two 5´-overhangs. As the restriction sites are palindromic, the two ends of the vector cut by the restriction endonuclease have the same 5´-overhangs with the same sequence. This allows triphosphate incorporation on both sites at the same time.

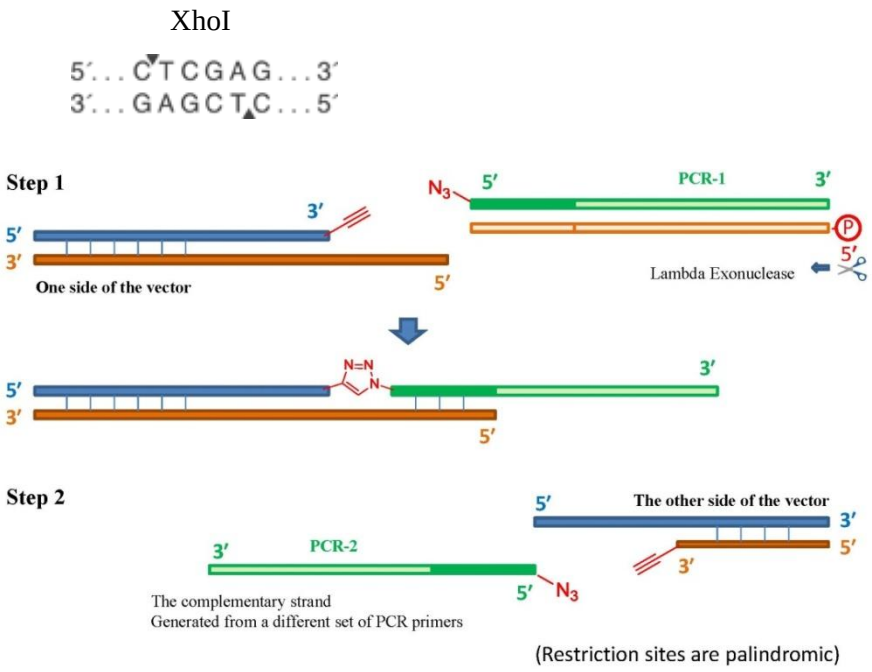


Figure 4.7: Strategy 2 of TA-cloning using 5´-overhangs instead of 3´-overhangs.

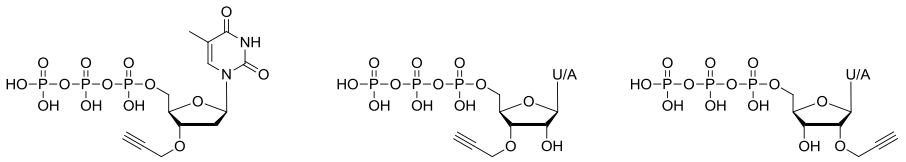
1. El-Sagheer, A.H. Very stable end-sealed double stranded DNA by click chemistry. *Nucleos. Nucleot. Nucl.* **28**, 315-323 (2009).
2. Gierlich, J., Gutsmiedl, K., Gramlich, P.M.E., Schmidt, A., Burley, G.A. & Carell, T. Synthesis of highly modified DNA by a combination of PCR with alkyne-bearing triphosphates and click chemistry. *Chem.-Eur. J.* **13**, 9486-9494 (2007).
3. El-Sagheer, A.H., Sanzone, A.P., Gao, R., Tavassoli, A. & Brown, T. Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **108**, 11338-11343 (2011).

CHAPTER 5 Introduction of Alkyne or Azide at the 3´-end of RNA by RNA Polymerases

5.1. Introduction

Similar to DNA labelling, alkyne or azide modified nucleoside triphosphates can be incorporated at the 3´-end of RNA by an RNA polymerase. RNA polymerases can be divided into two fundamentally different groups: DNA-dependent RNA polymerases and DNA-independent ribonucleotidyl transferases. DNA-dependent RNA polymerases catalyse the production of precursor mRNA (pre-miRNA) according to a DNA template during transcription in cells. DNA-independent ribonucleotidyl transferases [1] are involved in various cellular processes such as 3´-polyadenine tailing of pre-miRNA catalysed by poly(A) polymerases. As DNA-dependent RNA polymerases such as the commercially available T7 RNA Polymerase (NEB® #M0251) require no primer but a DNA template with a dsDNA promoter region, they are not suitable for RNA 3´-labelling. All RNA polymerases evaluated and referred in this study are DNA-independent ribonucleotidyl transferases, including Yeast Poly(A) Polymerase (Yeast PAP) and Cid1 Poly(U) Polymerase (Cid1 PUP).

Synthesised triphosphates



Commercially available triphosphates:

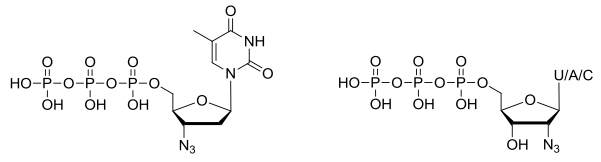


Figure 5.1: Alkyne/azide modified nucleotide triphosphates examined for RNA 3´-labelling by RNA polymerases (DNA-independent ribonucleotidyl transferases)

All synthesised triphosphates shown above were examined for enzymatic incorporation. In addition, commercially available AzdTTP and 2´-azido-2´-dUTP/dATP/dCTP (Figure 5.1)

were also examined according to the key publication [2], as cited in the introduction section (Figure 5.2).

Modified NTP	No. of residues added			
	Yeast PAP	*E.coli* PAP	Cid1 PUP	TdT
8-N_3-ATP	1-2	x	mostly 0	x
3´-N_3-2´,3´-ddATP	0-1	x	x	ND
3´-N_3-2´,3´-ddTTP	0-1	x	x	ND
2´-N_3-2´-dATP	multiple	1	mostly 0	x
2´-N_3-2´-dCTP	1-2	x	mostly 0	x
2´-N_3-2´-dGTP	1-2	x	mostly 0	mostly 0
2´-N_3-2´-dUTP	1-2	x	mostly 0	x

Reactions were carried out with commercial buffer, **500 µM NTP**, at 37 ℃ for 60 min (TdT: 90 min). Enzyme-specific conditions were the following: 0.2 µM RNA, 24 U/µL Yeast PAP; 0.25 µM RNA, 0.25 U/µL *E.coli* PAP; 0.2 µM RNA, 0.08 U/µL Cid1 PUP; 0.25 µM RNA, 1U/µL TdT.

x: not incorporated, ND: not determined.

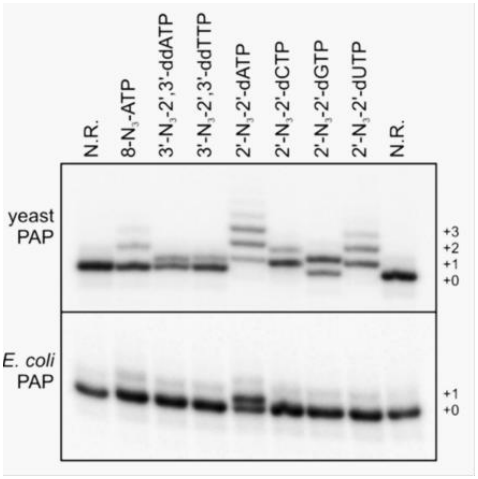


Figure 5.2: Results of the screening of nucleotidyl transferases for incorporation of azido nucleotides [2]**,** reproduced with permission from Oxford University Press.

5.2 RNA primer extension using commercially available azidonucleoside triphosphates

Following the methodology in the key publication in this field [2], the commercially available 3´-azido-3´-dTTP (AzdTTP), 2´-azido-2´-dUTP (2´-AzdUTP) and previously synthesised 3´-*O*-propargylthymidine triphosphate (PrdTTP) were tested for incorporation at the 3´-end of an RNA oligonucleotide using Yeast Poly(A) Polymerase (Yease PAP, USB® #74225) and Cid1 Poly(U) Polymerase (Cid1 PUP, NEB #M0337, fission yeast). *E.coli* PAP was not tested initially as the publication stated that Yeast PAP was superior (Figure 5.2). The PAGE gel (Figure 5.3) showed that 2´-AzdUTP was accepted by both yeast PAP and Cid1 PUP, which was consistent with the publication.

It is reported that AzdTTP is incorporated poorly by Yeast PAP (USB®) [2] but in our experiments it did not work at all. Although very unlikely, it is possible that the thymine base of the AzdTTP and PrdTTP (as opposed to adenine) is poorly accepted by Yeast PAP. This explains the low efficiency of dTTP incorporation comparing to that of ATP (Figure 5.4B). However, it is reported that both 2´ and 3´-dATP are poor substrates of Yeast PAP (about 50% labelling efficiency was reported for both substrates) [3]. The 3´-N_3-2´,3´-ddATP was also reported to be poorly incorporated by Yeast PAP (USB®) (Figure 5.2), indicating the Yeast PAP is selective for nucleobases as well as 3´-structures of the nucleoside triphosphates.

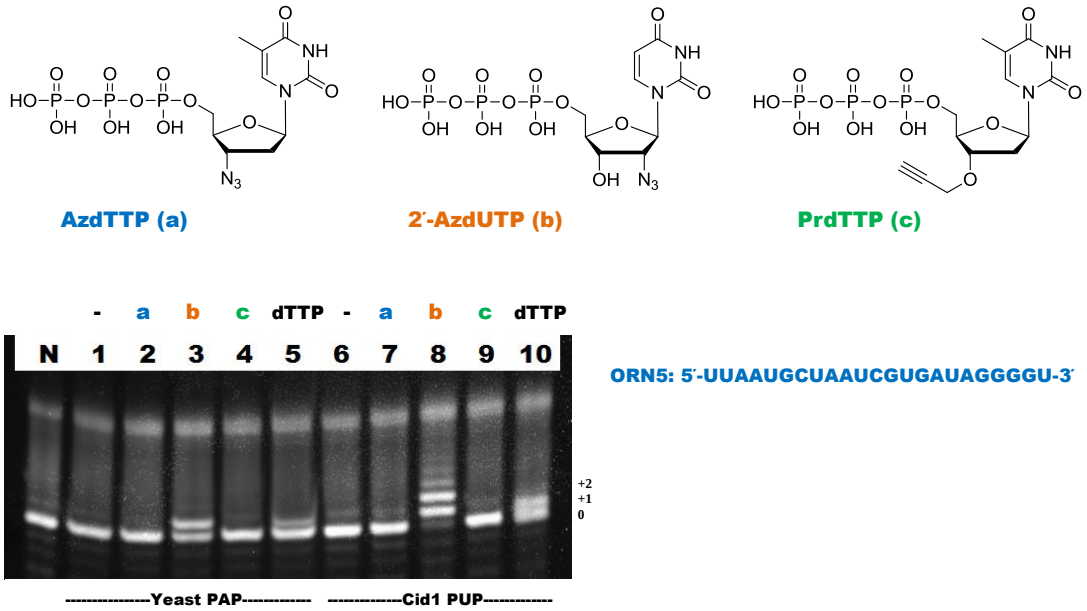


Figure 5.3: RNA non-templated primer extension using available modified nucleoside triphosphates.

Lane N: ORN5 (RNA primer, 76.7 pmol) in water.

Lane 1, 6: no triphosphate; Lane 2, 7: 160 µM AzdTTP-Na^+ (4 nmol); Lane 3, 8: 160 µM 2′-AzdUTP-Na^+ (4 nmol); Lane 4, 9: 160 µM PrdTTP-TEA (4 nmol); Lane 5, 10: 160 µM dTTP (4 nmol). (Triphosphate amount was calculated from $O.D._{260\ nm}$.)

Lane 1-5: Yeast PAP (USB®) test:

3.07 µM primer (76.7 pmol), 1x PAP buffer (USB®, 0.6 mM $MnCl_2$), 600 u Yeast PAP(USB®) in total 25 µL.

Lane 6-10: Cid1 PUP (NEB®) test:

3.07 µM primer (76.7 pmol), 1x NEBuffer 2 (NEB®, 10 mM $MnCl_2$), 2 u Cid1 PUP (NEB®) in total 25 µL.

Primer extension at 37 ℃ for 1 h. After PAGE, the gel is stained by SYBR® Gold.

To determine the base-selectivity of the Yeast PAP (USB®) and Cid1 PUP (NEB®), canonical UTP, dTTP and ATP were used as the positive controls (Figure 5.4). It was shown that UTP is not accepted by Yeast PAP (USB®) but is more efficiently accepted by Cid1 PUP than dTTP.

It was repeatedly found that Yeast PAP extension using 2′-azido-2′-dUTP gave only one nucleotide extension (Figure 5.3, Lane 3; Figure 5.4-A, Lane 2 and Figure 5.4-B, Lane 2). This differs from the publication [2], which showed multiple additions (Figure 5.2). This is likely because of the use of 160 µM triphosphates in our study instead of 500 µM in the literature. The "+1" addition of 2′-azido-2′-dUTP by Yeast PAP and "+1, +2" additions of 2′-azido-2′-dUTP by Cid1 PUP were confirmed by HPLC-MS of the desalted (by NAP column) reaction mixtures (Appendix, Figure 9.21, 9.22).

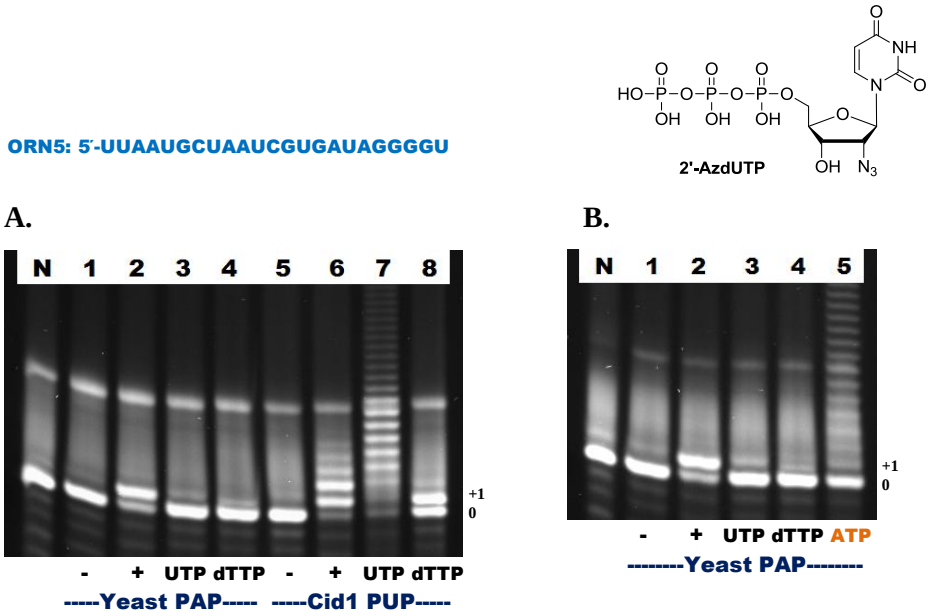


Figure 5.4: RNA non-templated primer extension using 2′-azido-2′-dUTP and canonical nucleoside triphosphates as controls.

Lane N: ORN5 (RNA primer, 76.7 pmol) in water.

A. Test of Yeast PAP (USB®) and Cid1 PUP using UTP and dTTP controls.

Lane 1, 5: no triphosphate, Lane 2, 6: 160 µM 2′-AzdUTP-Na^+ (4 nmol), Lane 3, 7: 160 µM UTP (4 nmol), Lane 4, 8: 160 µM dTTP (4 nmol).

Lane 1-4: Yeast PAP (USB®) test:

3.07 µM primer (76.7 pmol), 1xYeast PAP buffer (USB®), 600 u Yeast PAP (USB®) in total 25 µL.

Lane 5-8: Cid1 PUP test:

3.07 µM primer (76.7 pmol), 1x NEBuffer 2, 2 u Cid1 PUP (NEB®) in total 25 µL.

B. Repeated Yeast PAP (USB®) test with ATP control.

Lane 1: no triphosphate, Lane 2: 160 µM 2′-AzdUTP-Na^+ (4 nmol), Lane 3: 160 µM UTP (4 nmol), Lane 4: 160 µM dTTP (4 nmol), Lane 5: 160 µM ATP (4 nmol).

Composition: 0.0767 nmol primer, 1x Yeast PAP buffer (USB®), 600 u Yeast PAP (USB®) in total 25 µL.

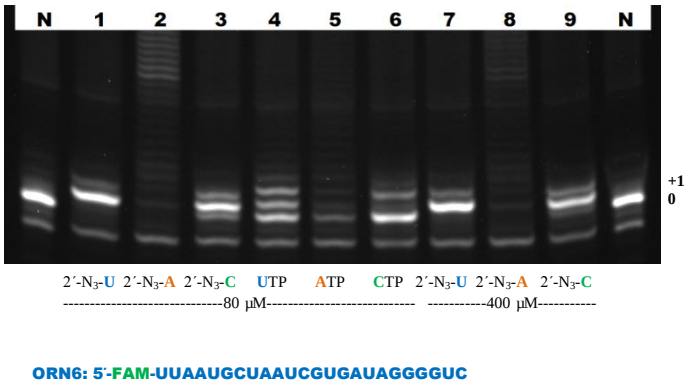
Primer extension: 37 ℃ for 1 h. After PAGE, the gel is stained by SYBR® Gold.

Interestingly, 2′-azido-2′-dUTP and ATP were accepted by Yeast PAP (USB®) but UTP was not accepted (Figure 5.4). The 2′-azide modification changed the enzyme's preference. This is also observed in the key publication [2] that among four 2′-azido-2′-dNTPs (N = A, U, C, G) tested for Yeast PAP (USB®) incorporation, 2′-azido-2′-dUTP worked. Unfortunately, no ATP and UTP controls were applied in this literature for comparison. However, it was reported in 1996 that *E.coli* Poly(A) polymerase is more base-selective than Cid1 PUP and selects only ATP but not UTP, CTP and GTP [4]. This is similar to our discovery using Yeast PAP and raises the question why 2′-azido-2′-dUTP was accepted.

5.3 Incorporation of 2′-azido-2′-dUTP and 2′-azido-2′-dCTP using Yeast PAP from MCLAB®

Yeast PAP from MCLAB® was found to have superior activity compared to the same enzyme from Affymetrix-USB®. The incorporation of 2′-azido-2′-dUTP was complete in 1 h and the incorporation of 2′-azido-2′-dCTP was nearly complete but also generated "+2 bands" (Figure 5.5). The "+1" and "+2" products were confirmed by MS to be single-addition product and double-addition product (Figure 5.6, 5.7). 2′-Azido-2′-dUTP was shown as a better chain terminator comparing to 2′-Azido-2′-dCTP as it generated less "+2" product. The time-course of the addition showed that the reaction proceeded quickly and the majority of the product formed in 10 min. Extended incubation generated insignificant differences (Figure 5.8). This higher activity may be the result of the presence of Mn^{2+} in the supplied reaction buffer (MCLAB®).

It was found that Yeast PAP from MCLAB® also accepted 2′-Azido-2′-dUTP and 2′-Azido-2′-dCTP better than their control UTP and CTP respectively. This was similar to that found using Yeast PAP from USB® mentioned above. The crystal structure of Yeast PAP may help to explain this phenomenon.



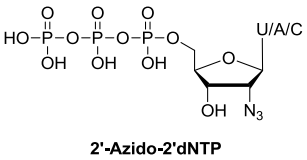


Figure 5.5: Incorporation of 2′-azido-2′-dNTP (N=U,A,C) by Yeast PAP (MCLAB®).

Lane N: primer ORN6 (5′-FAM) in water;

Lane 1, 7: 2′-N3-2′-dUTP-Na^+; Lane 2, 8: 2′-N_3-2′-dATP-Li^+; Lane 3, 9: 2′-N_3-2′-dCTP-Li^+; Lane 4: UTP-Na^+; Lane 5: ATP-Na^+; Lane 6: CTP-Na^+. Triphosphate concentrations were calculated according to O.D.260 nm; UTP, ε = 10000 $M^{-1}\cdot cm^{-1}$, ATP, ε = 15400 $M^{-1}\cdot cm^{-1}$, CTP, ε = 9000 $M^{-1}\cdot cm^{-1}$, (Thermo Scientific®).

Lane 1-6: 1.5 µM primer (30 pmol), 80 µM triphosphate (1.6 nmol), 1x Yeast PAP buffer (MCLAB®, 0.6 mM $MnCl_2$), 10 mM additional DTT, 5 u Yeast PAP (MCLAB®) in total 20 µL.
Lane 7-9: 400 µM triphosphate (8 nmol)
Primer extension: 37 ℃, 1 h

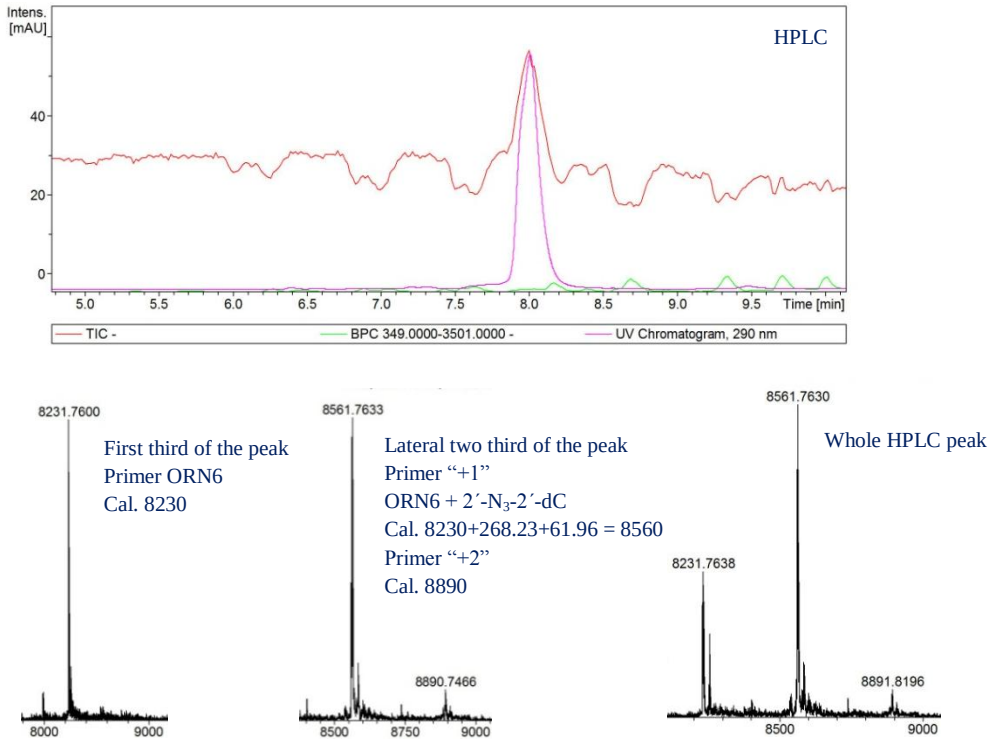


Figure 5.6: Yeast PAP (MCLAB®) reaction mixture for incorporation of 2´-N_3-2´-dCTP using RNA primer ORN6. The sample was incubated at 37 ℃ for 1 h (see also Figure 9.23).

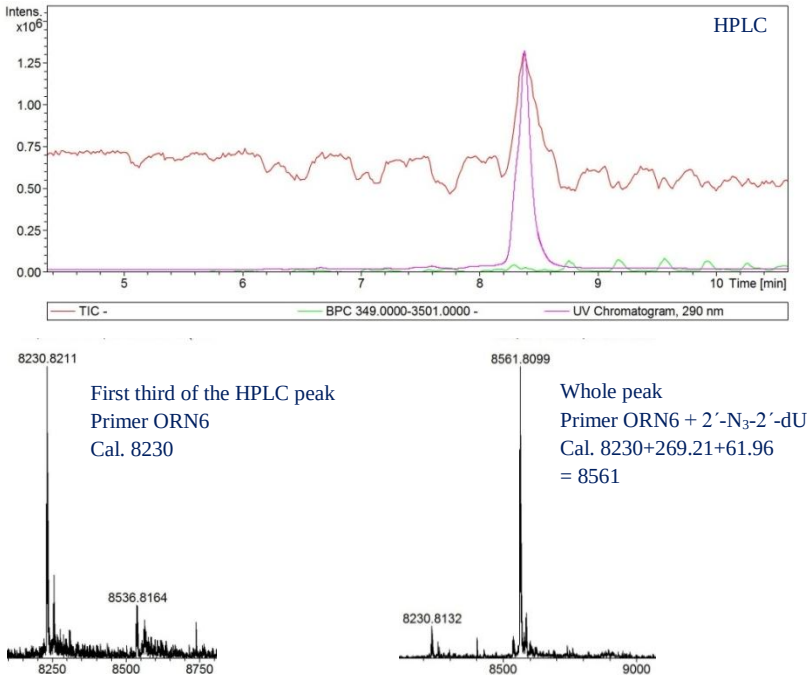
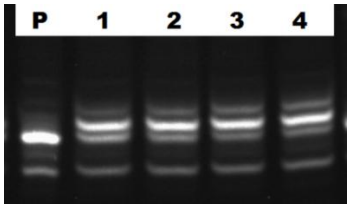
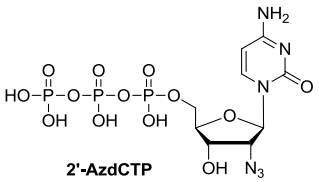


Figure 5.7: Yeast PAP (MCLAB®) reaction mixture for incorporation of 2´-N_3-2´-dUTP using RNA primer ORN6. The sample was incubated at 37 ℃ for 1 h (see also Figure 9.24).





ORN6: 5´-FAM-UUAAUGCUAAUCGUGAUAGGGGUC

Figure 5.8: Time-course for incorporation of 2´-azido-2´-dCTP by Yeast PAP (MCLAB®).

Oligonucleotide ORN6, 30 pmol each lane.

N: primer ORN6 (5´-FAM) in water;

Lane 1-4: 30 pmol primer (final 1.5 µM), 1.6 nmol 2´-N_3-2´-dCTP- Li^+ (final triphosphate conc. 80 µM, calculated according to $O.D._{260\ nm}$, CTP, ε = 9000 $M^{-1}{\cdot}cm^{-1}$-Thermo Scientific®), 1x Yeast PAP buffer (MCLAB, 0.6 mM $MnCl_2$), 10 mM additional DTT (NEB), 5 u Yeast PAP (MCLAB®) in total 20 µL.

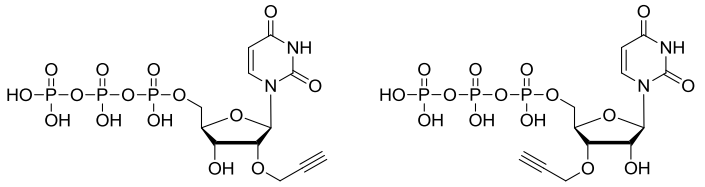
Primer extension: 37 ℃, 10 min/20 min/1 h/2 h

After labelling, the small RNAs will possess a 2´-azide at the 3´-end and in some cases (miRNA and siRNA) a 5´-phosphate group at the 5´-end. Performing 3´-adaptor Click ligation either before or after the 5´-adaptor ligation is discussed in detail in the sequencing section below.

5.4 RNA primer extension using 2´-*O*-propargyl-NTPs and 3´-*O*-propargyl-NTPs

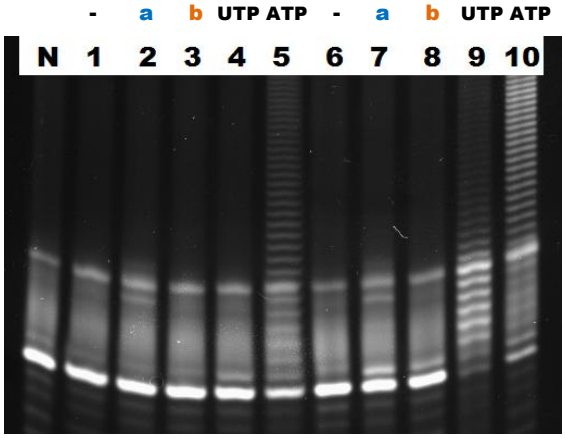
5.4.1 RNA primer extension using 2´-O-propargyl-UTP (2´-O-Pr-UTP) and 3´-O-propargyl-UTP (3´-O-Pr-UTP)

Primer extension using 3´-*O*-propargyl-NTPs was initially considered more important than incorporation of azido-NTPs because it leads to a possibly bio-compatible triazole linkage (Figure 5.11, Backbone B´). Unfortunately, PAGE analysis showed that Yeast PAP (USB®) does not accept 2´-*O*-Pr-UTP or 3´-*O*-Pr-UTP (Figure 5.9). Cid1 PUP incorporated 2´-*O*-Pr-UTP and 3´-*O*-Pr-UTP to a limited degree, and 2´-*O*-Pr-UTP was incorporated slightly more. The known preference of Yeast PAP for adenine rather than uracil [4] may cause 2´-*O*-Pr-UTP and 3´-*O*-Pr-UTP to be very poor substrates.



2′-*O*-Pr-UTP (a)

3′-*O*-Pr-UTP (b)



ORN5: 5′-UUAAUGCUAAUCGUGAUAGGGGU

Figure 5.9: The incorporation of 2′-*O*-Pr-UTP and 3′-*O*-Pr-UTP by Yeast PAP (USB®) and Cid1 PUP. The canonical triphosphates were used as controls.

Lane N: ORN5 (RNA primer, 76.7 pmol) in water.

Lane 1, 6: no triphosphate; Lane 2, 7: 160 μM 2′-*O*-Pr-UTP-TEA (4 nmol); Lane 3, 8: 160 μM 3′-*O*-Pr-UTP-TEA (4 nmol); Lane 4, 9: 160 μM UTP (4 nmol); Lane 5, 10: 160 μM dTTP (4 nmol). (Triphosphate amount was calculated from $O.D._{260\ nm}$.)

Lane 1-5: Yeast PAP test:

3.07 μM primer (76.7 pmol), 1X PAP buffer, 600 u Yeast PAP(USB®) in total 25 μL.

Lane 6-10: Cid1 PUP test:

3.07 μM primer (76.7 pmol), 1X NEBuffer 2, 2 u Cid1 PUP (NEB) in total 25 μL.

Primer extension at 37 ℃ for 1 h. After PAGE, the gel is stained by SYBR® Gold.

5.4.2 RNA primer extension using 2′-O-propargyl-ATP (2′-O-Pr-ATP) and 3′-O-propargyl-ATP (3′-O-Pr-ATP)

As discussed above, Yeast PAP utilises ATP much more efficiently than UTP (Figure 5.4 B). As 2′-*O*-Pr-UTP and 3′-*O*-Pr-UTP were not accepted by Yeast PAP, modified ATPs were considered as better choices compared to modified UTPs. Consequently, 2′-*O*-Pr-ATP (**32**) and 3′-*O*-Pr-ATP (**33**) were synthesised from commercially available resins (Scheme 2.6).

Enzymatic reactions were conducted using Yeast PAP from MCLAB®. PAGE analysis (Figure 5.10) showed that 2′-*O*-Pr-ATP was utilised by Yeast PAP (MCLAB®). The activity was further enhanced by additional DTT. However, because 2′-*O*-Pr-ATP possesses a 3′-OH, the extension did not stop at "+1" state but gave multiple-addition products. The positive control reactions using ATP (Lane 4 and 8) gave faint extended products at the upper region of the gel not shown in the figure.

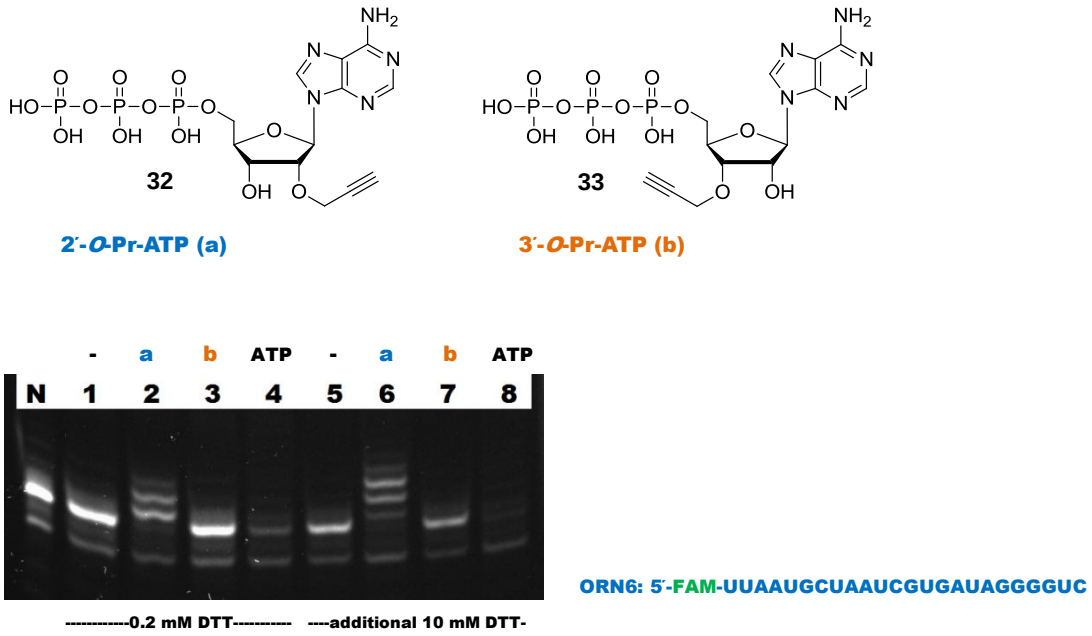


Figure 5.10: Yeast PAP (MCLAB®) primer extension test of 2′-*O*-Pr-ATP and 2′-*O*-Pr-ATP.

Lane N: ORN6 (RNA primer with 5′-FAM, 34.4 pmol) in water.

Lane 1, 5: no triphosphate; Lane 2, 6: 80 μM 2′-*O*-Pr-ATP-TEA (1.6 nmol); Lane 3, 7: 80 μM 3′-*O*-Pr-ATP-TEA (1.6 nmol); Lane 4, 8: 80 μM ATP (1.6 nmol). (Triphosphate concentration was calculated from $O.D._{260\ nm}$, $\varepsilon = 15400\ M^{-1}\cdot cm^{-1}$ for adenosine.)

Composition: 1.72 μM primer (0.0344 nmol), 1X Yeast PAP buffer (with 0.2 mM DTT), 5 u Yeast PAP (MCLAB®) in total 20 μL.

Lane 5-8: 10.2 mM DTT (with additional 2 μL 0.1 M DTT, 10x DTT from NEB®)

Primer extension: 37 ℃ for 1 h.

The incorporation results described above are summarized below:

1. 3′-*O*-Pr-ATP was not accepted by Yeast PAP (MCLAB®)
2. Cid1 PUP accepts 2′-*O*-Pr-UTP but not 3′-*O*-Pr-UTP
3. Yeast PAP and Cid1 PUP accept 2′-azido-2′-dUTP but not 3′-azido-3′-dTTP (AzdTTP).

The combination of these results indicates that available RNA polymerases may be very selective to the 3′-substituent on the triphosphates and only accept small groups such as 3′-H and 3′-OH. MS of the 2′-*O*-Pr-ATP incorporation by Yeast PAP (MCLAB®) was not performed.

In addition, *E.coli* PAP (NEB®) was also tested. It accepts neither 2′-*O*-Pr-ATP nor 3′-*O*-Pr-ATP (PAGE not shown).

5.5 Conclusions and further discussion

In summary (Table 5.1), among the modified triphosphates, only 2′-azido-2′-dUTP and 2′-azido-2′-dCTP can be used for further RNA ligation applications. They were only added once at each 3′-end by Yeast PAP although they possess 3′-OH groups. This makes them potentially useful tools for further applications. Consequently, potential triazole backbones were limited to the ones generated by 2′-azide modified triphosphate (Backbone C, Figure 5.11) (see also Figure 7.4 for BCN backbones), because the RNA polymerases did not incorporate 3′-modified triphosphate very well.

	Cid1 PUP (NEB®)	Yeast PAP (USB®)	Yeast PAP (MCLAB®)
AzdTTP	x	x	-
PrdTTP	x	x	-
2′-AzdUTP	1-3	0-1	1
2′-AzdCTP	-	-	Mostly 1
2′-AzdATP	-	-	Multiple additions
2′-*O*-Pr-UTP	x (trace of “+1”)	x	-
3′-*O*-Pr-UTP	x (trace of “+1”)	x	-
2′-*O*-Pr-ATP	-	-	1-3
3′-*O*-Pr-ATP	-	-	x

-: Not determined, x: not incorporated.

Table 5.1: Summary of modified nucleoside triphosphate incorporation results.

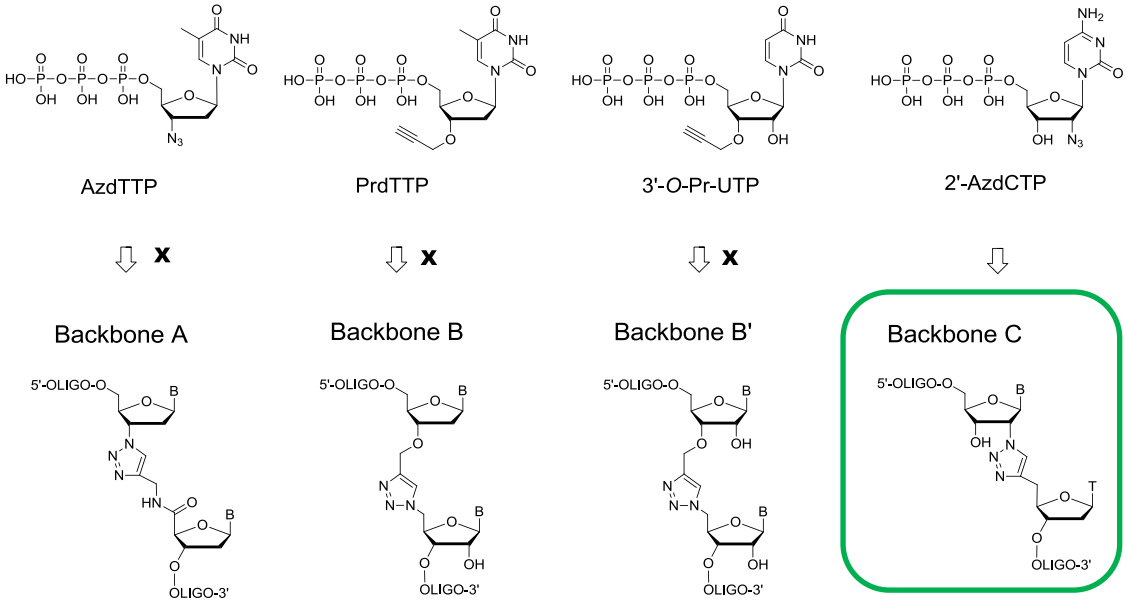


Figure 5.11: Triazole backbones generated by incorporation of nucleoside triphosphates into RNA.

2′-Azido-2′-dUTP was accepted efficiently by Cid1 PUP but induced multiple additions. In contrast, 2′-azido-2′,3′-dideoxyuridine triphosphate (2′-azido-2′,3′-ddUTP, **37**) has the potential to generate clean addition because of the presence of a 3′-H instead of OH. For Yeast PAP, 3′-Azido-3′-dATP (**38**) is another candidate because the 2′-OH may render it acceptable by the enzyme, but it is currently not commercially available.

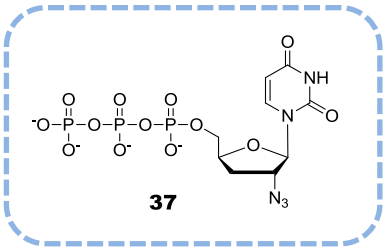


Figure 5.12: Potential triphosphate substrate for Cid1 PUP (NEB®): 2′-azido-2′,3′-dideoxy-uridine triphosphate (2′-azido-2′,3′-ddUTP, **37**).

More importantly, 3′-deoxyadenosine-5′-triphosphate (3′-dATP, Figure 5.13) was reported to be very efficiently incorporated by Yeast PAP [5]. This indicates that 3′-OH is not essential for recognition by the enzyme. Hence, 2′-azido-2′,3′-dideoxyadenosine triphosphate (2′-azido-2′,3′-ddATP, **39**) could potentially be incorporated by Yeast PAP, resulting in single addition. When taking the reverse-transcription results below (Chapter 7) in to consideration, it may be more useful to incorporate a cytosine base, rather than adenine or uracil, because

the corresponding triazole-containing template can achieve efficient reverse transcription (section 7.8 below). As Yeast PAP successfully incorporated 2′-azido-2′-dCTP, 2′-azido-2′,3′-dideoxycytidine triphosphate (2′-azido-2′,3′-ddCTP, **40**) is very likely to be incorporated by Yeast PAP. It may also be a better choice than 2′-azido-2′-dCTP, because it can only generate a single addition, however, this triphosphate needs to be synthesised.

For introduction of the propargyl group, only 2′-*O*-Pr-ATP was significantly incorporated by Yeast PAP (MCLAB®). However, the reaction did not stop after one nucleotide addition. 2′-*O*-Propargyl-3′-deoxyadenosine triphosphate (with 3′-H, **41**) may solve this problem if it is accepted by Yeast PAP. However, these compounds are not commercially available so chemical synthesis will also be needed.

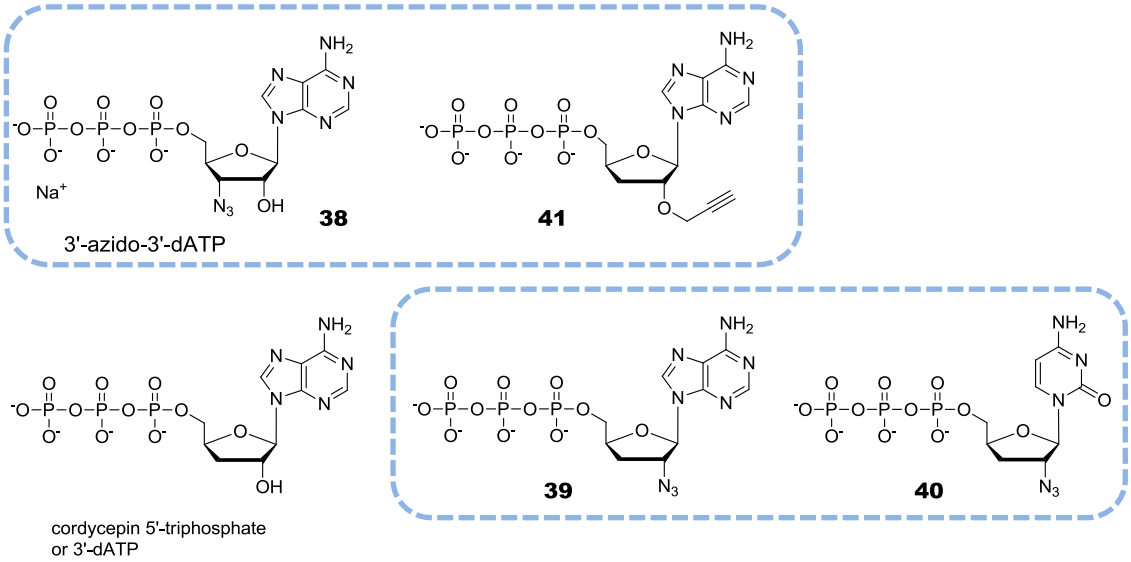
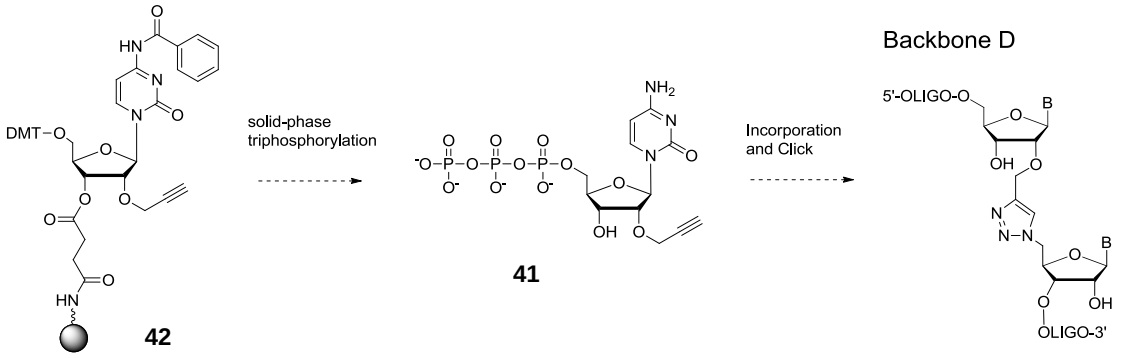


Figure 5.13: Triphosphate substrates and potential substrates for Yeast PAP: 3′-azido-3′-dATP (compound **21**, IBA Lifesciences® #5-0611-053, not currently been sold), 2′-*O*-propargyl-3′-deoxyadenosine triphosphate (**41**), 3′-deoxy-adenosine triphosphate (3′-dATP), 2′-azido-2′,3′-dideoxyadenosine triphosphate (2′-azido-2′,3′-ddATP, **39**), 2′-azido-2′,3′-dideoxycytidine triphosphate (2′-azido-2′,3′-ddCTP, **40**).

Based on the finding that Yeast PAP (MCLAB®) induces multiple additions of 2′-azido-2′-dATP but only 1-2 additions 2′-azido-2′-dCTP, Yeast PAP (MCLAB®) may incorporate 2′-*O*-propargyl-cytidine triphosphate (**41**) only once while incorporating 2′-*O*-propargyl-adenosine (2′-*O*-Pr-ATP) triphosphate multiple times (Figure 5.10). 2′-*O*-Propargyl-cytidine triphosphate can be readily synthesised from commercially available 5′-DMT-2′-*O*-propargylcytidine (*N*-Bz)-3′-lcaa CPG resin (**42**) using the solid-phase triphosphorylation method mentioned above (Scheme 5.1).

However, if incorporation of this triphosphate is possible, it generates triazole Backbone D (Scheme 5.1), which, like triazole backbone C, has not been studied (Figure 5.11). The reverse transcription of triazole Backbone D has also not been investigated.



Scheme 5.1: Potential synthesis of 2´-*O*-propargyl-cytidine triphosphate (**41**) from commercially available 5´-DMT-2´-*O*-propargylcytidine (*N*-Bz)-3´-lcaa CPG resin (**42**).

Optimisation of the primer extension has not covered all aspects. Possible parameters to be changed in the future include extension time, type of cation (triphosphate salt) and divalent cation concentration in the reaction buffer. Extension of the incubation time from 1 h to overnight may improve the results in case incorporation speed is slow. The incorporation of nucleotide triphosphates could be improved by testing triphosphate sodium salts rather than the TEA salts. For 9 °N$_m$ DNA polymerase (Therminator^{TM} II) in the DNA extension section (Section 3.4), the TEA cation did not affect the incorporation of PrdTTP. However, for RNA polymerases this effect is still uncertain. Finally, the Yeast PAP buffer from MCLAB® has Mn^{2+} already added but other concentrations beside the default one have not been tested.

1. Martin, G. & Keller, W. RNA-specific ribonucleotidyl transferases. *RNA* **13**, 1834-1849 (2007).
2. Winz, M.L., Samanta, A., Benzinger, D. & Jaschke, A. Site-specific terminal and internal labeling of RNA by poly(A) polymerase tailing and copper-catalyzed or copper-free strain-promoted click chemistry. *Nucleic Acids Res.* **40**, 13 (2012).
3. Chen, L.S., Plunkett, W. & Gandhi, V. Polyadenylation inhibition by the triphosphates of deoxyadenosine analogues. *Leukemia Res.* **32**, 1573-1581 (2008).
4. Zaug, A.J., Lingner, J. & Cech, T.R. Method for determining RNA 3' ends and application to human telomerase RNA. *Nucleic Acids Res.* **24**, 532-533 (1996).
5. Martin, G. & Keller, W. Tailing and 3'-end labeling of RNA with yeast poly(A) polymerase and various nucleotides. *RNA* **4**, 226-230 (1998).

CHAPTER 6 Introduction of 3′-alkyne to RNA by T4 RNA ligase 1

6.1 Introduction

Bacterial phage T4 RNA ligase catalyses the formation of a phosphodiester bond between a terminal 5′-phosphate (phosphate donor) and a terminal 3′-OH (phosphate acceptor) on RNA with the consumption of a molecule of ATP. T4 RNA ligases are commonly utilised for intramolecular cyclization, and more usefully, oligonucleotide intermolecular 3′-labelling such as C5-biotin labelling (Figure 6.1) [1]. T4 RNA ligase 1 has been utilised to introduce the alkyne of a 5-alkynylated uridine-3′,5′-*bis*phosphate on to the 3′-end of DNA [2]. The potential of this enzyme for 3′-propargyl introduction is explored below.

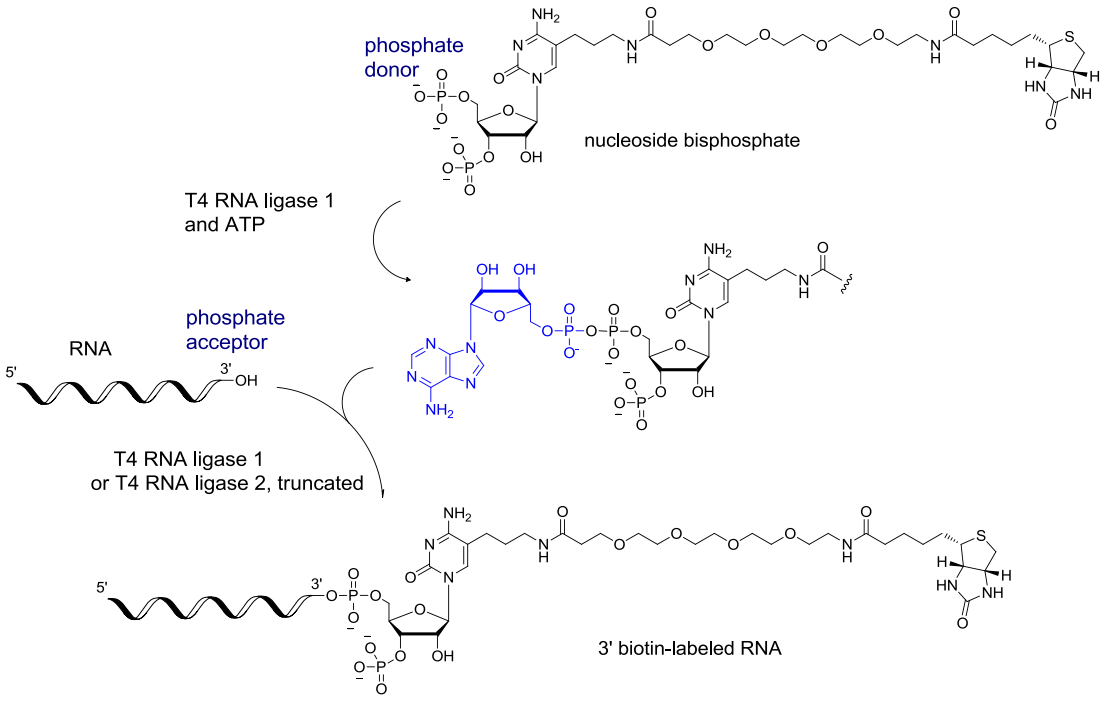


Figure 6.1: Biotin labelling of RNA 3′-end using T4 RNA ligase and a biotinylated cytidine 3′,5′-*bis*-phosphate analogue [1] (the synthesis of *bis*-phosphates is described in a patent document [3]). Adenylation portion is marked in blue.

T4 RNA ligase utilises ATP to adenylate the donor substrate, and the activated intermediate reacts with the acceptor oligonucleotide to form the phosphodiester linkage as well as concomitantly releasing AMP [4]. Pre-adenylated substrates are used in ATP-independent reactions catalysed by "T4 RNA ligase 2, truncated" to minimize undesired ligation.

6.2 Substrate requirement of T4 RNA ligase

For intermolecular reactions instead of intramolecular cyclization, the minimal acceptor for T4 RNA ligase is a trinucleoside diphosphate with a 3′-OH. The donor can either be a dinucleoside-5′,5′-pyrophosphate (e.g. **43**, Figure 6.2) (ATP-independent) or mononucleoside -3′,5′-*bis*-phosphates (ATP-dependent) [5]. It is reported that pre-adenylated donors such as diadenosine pyrophosphates (**43**) give greater incorporation efficiency compared to nucleoside 3′,5′-*bis*-phosphates [4,6]. However, synthesising the pre-adenylated substrates can be laborious as it requires another step in addition to the nucleoside-5′-monophosphate synthesis procedure [7]. One recently developed enzymatic adenylation method using *Methanobacterium thermoautotrophicum* RNA ligase 1 (MthRn1, NEB® 5′ DNA Adenylation Kit) was reported to quantitatively transfer the ssDNA or ssRNA 5′-phosphate end to the adenylated end (5′-App) [8]. It may greatly facilitate the ATP-free labelling method. The ATP-dependent ligation is catalysed by T4 RNA ligase 1 (NEB® #M0204), while the ATP-independent ligation is catalysed by T4 RNA ligase 2, truncated (NEB® #M0242). T4 RNA ligase 2, truncated (NEB® #M0242) lacks adenylation activity.

Among nucleoside 3′,5′-*bis*-phosphates "pCp" is the best donor and "pGp" is the poorest [4]. The 2′-deoxyribonucleoside-3′,5′-*bis*-phosphates can also serve as donors, and DNA rather than RNA can be used as a acceptor but with a reduced reaction rate [9]. Nucleoside 5′-monophosphates and nucleoside 5′,2′-diphosphates are not accepted as donors, demonstrating the essential role of the 3′-phosphate group [10,11]. This suggests that 3′-*O*-propargylnucleoside 5′-monophosphates are unlikely to be accepted (the 1-mer in Figure 6.4 below).

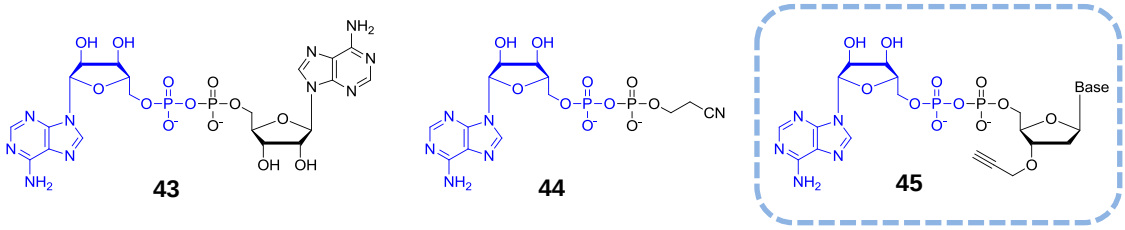
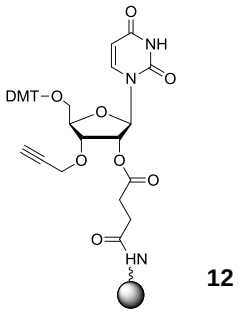


Figure 6.2: T4 RNA ligase 2 substrates include diadenosine pyrophosphate (**43** [7]) and adenylated cyanoethylphosphate (**44**). Potential substrates include adenylated 3′-propargyl-nucleoside-5′-monophosphates (**45**). AMP portions are marked in blue.

Among dinucleoside pyrophosphates, T4 RNA ligase showed a stringent selectivity for the AMP portion of the substrate but no selectivity towards the other part [12] (Figure 6.2). A cyanoethylphosphate can be ligated to the acceptor by T4 RNA ligase using an adenylated substrate (**44**) [12], suggesting that adenylated 3′-propargylnucleoside-5′-monophosphate (**45**) may be a substrate for T4 RNA ligase.

6.3 ATP-dependent ligation of alkyne-bearing short RNAs by T4 RNA ligase

To avoid synthesis of adenylated substrates, T4 RNA ligase 1 was selected for ATP-dependent reactions. The 3′-propargyl group was carried on a short RNA (1mer–4mer), which is conveniently obtained through RNA solid-phase synthesis using 3′-*O*-propargyl uridine 2′-lcaa CPG resin (**12**, Figure 6.3). The triazole backbone that we intended to generate in this study is Backbone B′ (Figure 6.3). The oligonucleotide chains were assembled through standard RNA synthesis using 2′-*O*-TBDMS protected ribonucleoside phosphoramidites. The short 1- to 4-mers also contain a 5′-phosphate, which is introduced through chemical phosphorylation (see oligonucleotide synthesis section 8.2 in Experimental chapter 8). In order to maintain consistency in the experimental procedure, TBDMS deprotection was performed during the synthesis of all four short RNAs. However, it is unnecessary for the 1-mer because the nucleotide originates from the resin and does not contain a TBDMS protecting group. The HPLC purified 1- to 4-mers were characterised by MS.



12

Backbone B'

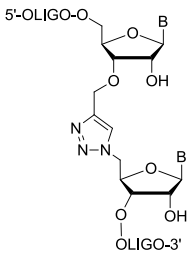


Figure 6.3: 3′-*O*-Propargyl uridine 2′-lcaa CPG resin (**12**) and triazole Backbone B′.

The results (Figure 6.4) show that 2-mer, 3-mer and 4-mer were efficiently ligated to the ssRNA strand but the 1-mer was not, which is consistent with our predictions. The three ligation products were confirmed by MS (Appendix, Figure 9.25, 9.26, 9.27).

ORN6: 5′-FAM-UUAAUGCUAAUCGUGAUAGGGGUC

Short RNAs with 5′-phosphate and 3′-propargyl:

1-mer: 5′-P-**C**-3′-*O*-propargyl (ORN1)
2-mer: 5′-P-**CC**-3′-*O*-propargyl (ORN2)
3-mer: 5′-P-**CUC**-3′-*O*-propargyl (ORN3)
4-mer: 5′-P-**CUAC**-3′-*O*-propargyl (ORN4)

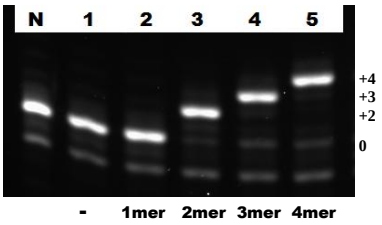


Figure 6.4: T4-RNA ligase 1 labelling of ssRNA (ORN6, with 5′-FAM).

Lane 1-5: 1.15 µM RNA primer (ORN6 with 5′-FAM, 34.4 pmol), 10% DMSO, 33 µM monomer-tetramer (1 nmol except Lane 1), 1X T4 RNA ligase buffer, 1 mM ATP, 10 u T4 RNA ligase 1 (NEB®) in total 30 µL. Ligation: 16 °C, 15.5 h (overnight).

The ligation capacity of all 2- to 4-mers were explored. The labelling of the 10x concentrated ssRNA target went to completion (Figure 6.5) with the help of doubled concentrations of enzyme and short RNAs (compared to the conditions in Figure 6.4), demonstrating the high-efficiency of this reaction.

ORN6: 5′-FAM-UUAAUGCUAAUCGUGAUAGGGGUC

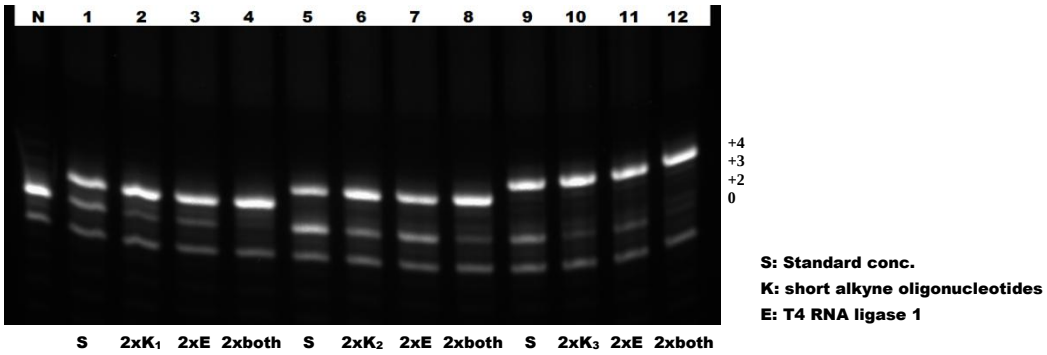


Figure 6.5: T4-RNA ligase 1 labelling of ssRNA (concentrated, ORN6, with 5′-FAM). Lane 1-4: ORN2 (2-mer), Lane 5-8: ORN3 (3-mer), Lane 9-12: ORN4 (4-mer).

Lane 1, 5, 9: Standard, 33.3 µM dimer-tetramer (1 nmol), 10 u T4 RNA ligase 1 (NEB®)

Lane 2, 6, 10: 66.7 µM dimer-tetramer (2 nmol), standard enzyme conc.

Lane 3, 7, 11: 20 u T4 RNA ligase 1 (NEB®), standard dimer-tetramer conc.

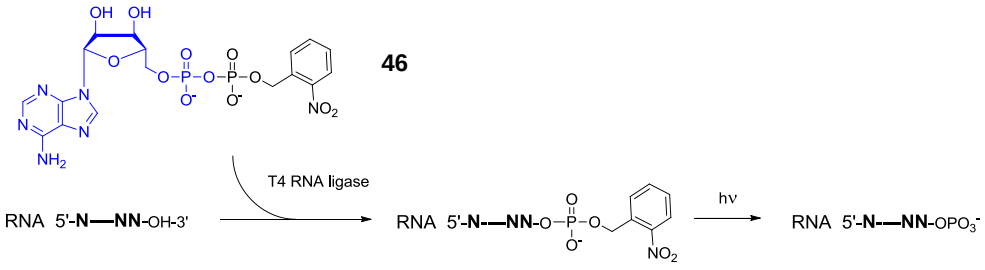
Lane 4, 8, 12: 66.7 µM dimer-tetramer (2 nmol) and 20 u T4 RNA ligase 1 (NEB®)

Lane 1-12: 11.5 µM RNA primer (ORN6 with 5′-FAM, 0.344 nmol), 10% DMSO, 1X T4 RNA ligase buffer, 1 mM ATP in total 30 µL. Ligation: 16 °C, 15.5 h (overnight).

6.4 Advantages and disadvantages of T4 RNA ligase labelling

It has been reported that T4 RNA ligase is less sensitive to a 2′-*O*-Me modification at the ligation site compared to poly(A) polymerases (EPICENTRE Biotechnologies website). T4 RNA ligase and diadenosyl pyrophosphate (AppA, **43**, Figure 6.2) were developed to polyadenylate 2′-*O*-Me modified small RNAs such as siRNAs (Small interfering RNA), piRNAs (Piwi-interacting RNA), and plant miRNAs (microRNA) (EPICENTRE Biotechnologies). The comparative study of the ligation of 2′-*O*-Me modified RNAs is discussed in the next section.

In addition, T4 RNA ligase was also utilised for 3′-phosphorylation of oligonucleotides. As mentioned previously, the T4 RNA ligase is not selective towards the non-AMP portion of the pyrophosphate substrate. P1-adenosine-5′-P2-*O*-nitrobenzylphosphate (**46**) was developed for 3′-phosphorylation by photo-cleavage of the *O*-nitrobenzyl-phosphate group (Scheme 6.1) [13].



Scheme 6.1: RNA 3′-phosphorylation using P1-adenosine-5′-P2-*O*-nitrobenzylphosphate (**46**).

Although introducing alkyne or azide into RNA by T4 RNA ligase is convenient, it has an intrinsic limitation when applied to RNA sequencing. This is because the enzyme used for the labelling is the same enzyme catalysing conventional RNA enzymatic ligation. Using the same enzyme may result in similar substrate selectivity so it may inherit the problems generated by enzymatic ligation. In the case of RNA sequencing, it is the sequencing bias problem that is discussed in detail in the next chapter.

6.5 3′-Labelling of RNA with 2′-*O*-Me modification at 3′-end

Some natural small RNA species, including mammalian piRNAs, possess 2′-*O*-Me modifications (summarised in Table 7.1). This modification was reported to severely hinder *E.coli* PAP and Cid1 PUP catalysed polyadenylation and decrease the efficiency of T4 RNA ligase mediated adapter ligation [14]. This may limit the use of Click ligation in sequencing of certain small RNAs (The details of the small RNA sequencing protocol is described in Chapter 7). According to the results described above, the terminal alkyne or azide was introduced either by Yeast PAP (using modified triphosphates) or T4 RNA ligase 1 (using modified short 2- to 4-mer oligonucleotides). Both these two methods need to be evaluated for labelling 2′-*O*-Me modified RNA 3′-end.

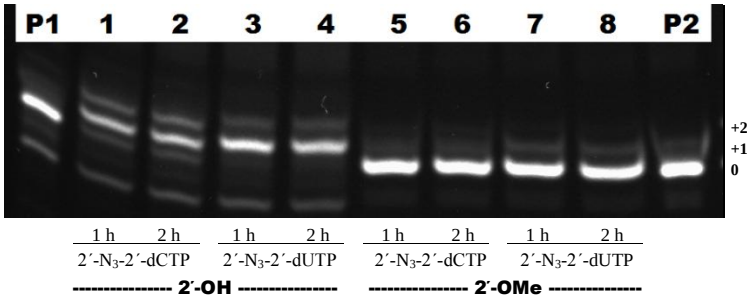


Figure 6.6: Yeast PAP Labelling of RNA 3′-ends with or without 2′-*O*-Me modification.

Oligonucleotide: ORN6 (2′-OH), 30 pmol for Lane P1, 1, 2, 3, 4;
ORN7 (2′-*O*-Me), 30 pmol for Lane 5, 6, 7, 8, P2.
P1: primer ORN6 (5′-FAM) in water; P2: primer ORN7 (5′-FAM) in water;
Lane 1, 2, 5, 6: 2′-N_3-2′-dCTP; Lane 3, 4, 7, 8: 2′-N_3-2′-dUTP.

Lane 1-8: 1.5 µM primer (30 pmol), 80 µM (1.6 nmol) triphosphate, 1x Yeast PAP buffer (MCLAB®, 0.6 mM $MnCl_2$), 10 mM additional DTT, 5 u (1 µL) Yeast PAP (MCLAB®) in total 20 µL.

Primer extension: 37 ℃, 1 h/2 h.

Using a synthesised RNA with a 2′-*O*-methyl modification, the RNA polymerase *in vitro* labelling efficiency at the modified 3′-end was compared with that of the unmodified 3′-end. The result showed that 2′-*O*-Me modification blocked the addition of 2′-azido-2′-dUTP and 2′-azido-2′-dCTP by Yeast PAP in 2 h (Figure 6.6). This presents a problem for our small

RNA sequencing strategy as the RNA with 2′-*O*-Me modification cannot be labelled, and consequently their sequences will be lost. Overnight labelling may improve the result but has not been tested.

The overnight reaction also has the potential to select the 2′-*O*-Me modified RNA for the sequencing library. If the overnight reaction generates a single addition for 2′-*O*-Me modified RNA, under the same conditions, the unmodified RNA may generate multiple-addition products. The multiple additions of the modified triphosphate will introduce multiple 2′-azide labels which can "click" to several 5′-BCN labelled adapters. This branched structure will likely prevent the reverse-transcriptase from reading-through, thus RNAs with no 2′-*O*-Me modification could be excluded in the sequencing library.

The 2′-*O*-Me moiety also significantly hindered T4 RNA Ligase labelling with the 3′-*O*-propargyl modified oligonucleotide dimer, mentioned in the T4 RNA Ligase labelling section above (Figure 6.7). As the 3′-*O*-propargyl is unlikely to affect the ligation efficiency at the other end, this result indicates the T4 RNA Ligase 1 ligation efficiency is generally hindered by 2′-*O*-Me modification. This confirms the literature report [15] that 2′-*O*-Me modified RNA is under-represented in the sequencing result if the library preparation involves T4 RNA Ligase ligation (discussed in detail in the Chapter 7).

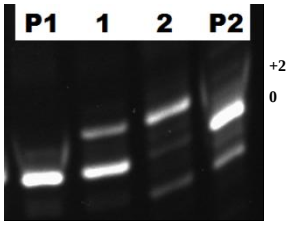


Figure 6.7: T4 RNA ligase 1 Labelling of RNA 3′-ends with or without 2′-*O*-Me modification

Oligonucleotide: ORN7 (2′-*O*-Me), 30 pmol for Lane P1, 1;
ORN6 (2′-OH), 30 pmol for Lane 2, P2.
P1: primer ORN7 (5′-FAM) in water; P2: primer ORN6 (5′-FAM) in water;

Lane 1, 2: 1.5 µM primer (30 pmol), 10% DMSO (2 µL), 50 µM (1 nmol) dimer ORN2 (3′-*O*-Propargyl), 1x T4 RNA ligase buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.5 at 25 ℃), 1 mM ATP, 10 u (1 µL) T4 RNA ligase 1 (NEB®) in total 20 µL. Ligation: 16 ℃, 16 h (overnight).

6.6 Conclusions

RNA 3´-labelling using T4 RNA ligase 1 and modified short oligonucleotides (2-mer, 3-mer and 4-mer) was highly efficient for the unmodified RNA 3´-end. This enzyme did not accept the 1-mer with a 3´-propargyl modification. It provides a very convenient strategy to introduce alkyne/azide modifications because it does not require synthesis of modified nucleotide triphosphates but uses ATP.

For 3´-labelling at the 2´-*O*-Me modified RNA 3´-end, T4 RNA ligase 1 reaction worked partially, while Yeast PAP reaction did not work at all. This indicates that T4 RNA ligase 1 may be a better choice for labelling certain small RNAs that may contain 2´-*O*-Me modifications.

1. Barone, A.D., Chen, C., McGall, G.H., Rafii, K., Buzby, P.R. & Dimeo, J.J. Novel nucleoside triphosphate analogs for the enzymatic labeling of nucleic acids. *Nucleosides Nucleotides & Nucleic Acids* **20**, 1141-1145 (2001).
2. Dojahn, C.M., Hesse, M. & Arenz, C. A chemo-enzymatic approach to specifically click-modified RNA. *Chem. Commun.* **49**, 3128-3130 (2013).
3. Opperman, K., Kaboord, B.J., Schultz, J.-S., Etienne, C.L. & Hermanson, G. Modified nucleotides. *Official Gazette of the United States Patent and Trademark Office Patents* (2013).
4. McLaughlin, L.W., Piel, N. & Graeser, E. Donor activation in the T4 RNA ligase reaction. *Biochemistry* **24**, 267-273 (1985).
5. England, T.E. & Uhlenbeck, O.C. 3'-terminal labeling of RNA with T4 RNA ligase. *Nature* **275**, 560-561 (1978).
6. Hoffmann, P.U. & McLaughlin, L.W. Synthesis and reactivity of intermediates formed in the T4 RNA ligase reaction. *Nucleic Acids Res.* **15**, 5289-5303 (1987).
7. Moffatt, J.G. & Khorana, H.G. Nucleoside polyphosphates. X.1 the synthesis and some reactions of nucleoside-5' phosphoromorpholidates and related compounds. Improved methods for the preparation of nucleoside-5' polyphosphates. *J. Am. Chem. Soc.* **83**, 649-658 (1961).
8. Zhelkovsky, A.M. & McReynolds, L.A. Simple and efficient synthesis of 5' pre-adenylated DNA using thermostable RNA ligase. *Nucleic Acids Res.* **39**, E117-U71 (2011).

9. Hinton, D.M., Baez, J.A. & Gumport, R.I. T4-RNA ligase joins 2'-deoxyribonucleoside 3',5'-bisphosphates to oligodeoxyribonucleotides. *Biochemistry* **17**, 5091-5097 (1978).
10. Kikuchi, Y., Hishinuma, F. & Sakaguchi, K. Addition of mono-nucleotides to oligoribonucleotide acceptors with T4-RNA ligase. *Proc. Natl. Acad. Sci. USA* **75**, 1270-1273 (1978).
11. England, T.E. & Uhlenbeck, O.C. Enzymatic oligoribonucleotide synthesis with T4 RNA ligase. *Biochemistry* **17**, 2069-2076 (1978).
12. England, T.E., Gumport, R.I. & Uhlenbeck, O.C. Dinucleoside pyrophosphates are substrates for T4-induced RNA ligase. *Proc. Natl. Acad. Sci. USA* **74**, 4839-4842 (1977).
13. Ohtsuka, E., Uemura, H., Doi, T., Miyake, T., Nishikawa, S. & Ikehara, M. Transfer ribonucleic-acids and related compounds .28. New method for 3'-labeling of polyribonucleotides by phosphorylation with RNA ligase and its application to the 3'-modification for joining reactions. *Nucleic Acids Res.* **6**, 443-454 (1979).
14. Munafo, D.B. & Robb, G.B. Optimization of enzymatic reaction conditions for generating representative pools of cDNA from small RNA. *RNA* **16**, 2537-2552 (2010).
15. Zhuang, F., Fuchs, R.T. & Robb, G.B. Small RNA expression profiling by high-throughput sequencing: Implications of enzymatic manipulation. *J. Nucleic Acids.* **2012**, 360358-360358 (2012).

CHAPTER 7 Click ligation for small RNA sequencing

7.1 Introduction

The two-step 'label-then-Click' approach of RNA Click ligation is further explored for small RNA sequencing. Both standard and modified protocols for small RNA sequencing library preparation are described in detail. In the modified protocol, Click ligation is designed to substitute for the enzymatic 3′-adapter ligation to solve the sequence bias problem generated by using T4 RNA ligase. Although Click ligation can generate various triazole linkages with different adjacent sequences, two equally important requirements for the triazole linkage need to be met to allow Click ligation to be used for RNA sequencing. Firstly, the 3′-alkyne/azide labelling, corresponding to the triazole linkage with a suitable sequence, needs to be achievable. Secondly, the triazole linkage in the template (triazole backbone) needs to be read-through efficiently by a reverse transcriptase. An *in vitro* reverse transcription assay was performed to find the triazole backbone with the optimal nucleotide sequence, adjacent to the triazole linkage, that can be read-through by various reverse transcriptases.

7.2 Small RNAs

RNAs extracted from eukaryotic cells are classified into five different categories: ribosomal RNAs (rRNA), transfer RNAs (tRNA), messenger RNAs (mRNAs), long noncoding RNAs (lncRNAs, including circular RNAs [1]), and small RNAs [2]. Over 90% of the total RNA molecules in the extract are rRNA and tRNA. Although small RNAs account for only about 1% of the total RNA population [2], they have become a major focus, because of their important roles in regulation of chromatin structure, chromosome segregation, transcription, RNA processing, RNA stability, and translation [3]. Three major classes of small RNAs found only in eukaryotes: microRNAs (miRNAs), small interfering RNAs (siRNAs) and piwi-interacting RNAs (piRNAs) were defined mainly according to their distinctive precursors, generation pathways and associated effector proteins [3].

In animals, the generation of miRNAs starts with the transcribed primary miRNAs (pri-miRNAs). They are cleaved by the RNase III enzyme, Drosha and Dicer into precursor miRNAs (pre-miRNAs), and then transported to the cytoplasm to be further cleaved by Dicer [2]. Both pri-miRNAs and pre-miRNAs have stem-looped structures. In animals, the mature microRNAs have lengths of 21 to 24 nt and possess a monophosphate at the 5′-terminus and a 3′-OH at the 3′-terminus [2]. They regulate gene expression in the transcriptional and post-

transcriptional levels. Due to of their conserved functions, miRNAs were used as biomarkers for cell differentiation states and diseases, including cancer [2].

The term siRNA was first used to describe the 21 to 23 nt length double-stranded RNA (dsRNA) derived from exogenous dsRNA, that plays a vital role in RNA interference (RNAi) [2]. However, it was found that siRNA can also be endogenic [3,4]. An siRNA has a 19-21 base-pair duplex and a two-nucleotide 3′-overhang on each strand. The endogenic siRNAs originate from endogenous dsRNA, through a Dicer-dependent pathway, and possess 5′-monophosphate groups. Derivatives of this family called secondary siRNAs are generated through a Dicer-independent pathway and possess 5′-polyphosphate groups.

The third category is the piRNAs, which are 26-30 nt long and are predominantly expressed in germline cells [2]. Unlike miRNA and siRNA, which associate with members of the Ago clade of Argonaute proteins, piRNA associates with members of the Piwi clade [3]. The terminal modifications of these three classes of small RNAs are summarised in the literature [2] (Table 7.1).

Class	Organism	5′-end modification	3′-end modification
miRNA	Mammals	Monophosphate	2′-OH
	Nematodes	Monophosphate	2′-OH
	Insects	Monophosphate	2′-OH
	Plants	Monophosphate	2′-*O*-methyl
siRNA	Mammals	Monophosphate	2′-OH
	Nematodes	Monophosphate	2′-OH
	Insects	Monophosphate	2′-*O*-methyl
	Plants	Monophosphate	2′-*O*-methyl
Secondary siRNA	Nematodes	Polyphosphate	2′-OH
	Plants	Monophosphate	2′-*O*-methyl
piRNA	Mammals	Monophosphate	2′-*O*-methyl
	Nematodes	Monophosphate	2′-*O*-methyl
	Insects	Monophosphate	2′-*O*-methyl

Table 7.1: 5′ and 3′ modifications of different small RNA classes [2].

7.3 Next generation sequencing for RNA

7.3.1 Current RNA sequencing library preparation work-flow

Current RNA profiling techniques are generally divided into two categories: hybridization-based techniques (such as cDNA Microarrays) and high-throughput sequencing (HTS, or next generation sequencing) [5]. As HTS is not based on known sequences on the chip, it has certain advantages, such as the capabilities of novel small RNA detection, allelic expression detection and miRNA editing detection [6]. Compared to traditional cDNA cloning and conventional sequencing [7], HTS offers a significant time-saving advantage. Two HTS techniques most widely applied currently are Illumina® dye sequencing (Figure 7.1) and Roche® 454 pyrosequencing.

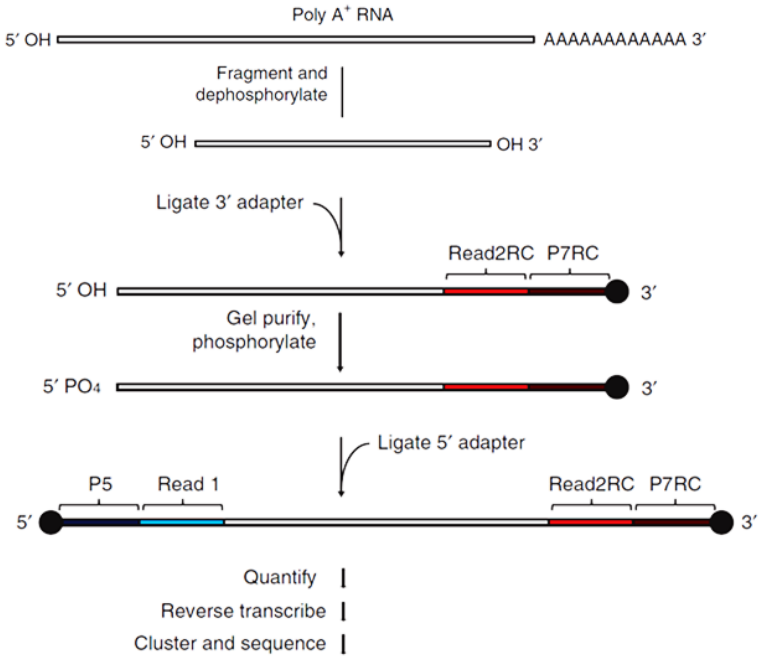


Figure 7.1: mRNA Sequencing library preparation workflow for Illumina® on-flowcell reverse transcription sequencing [6], reproduced with permission from Nature Publishing Group.

In order to generate RNA sequencing libraries, biologically-derived RNAs need to be copied into DNAs and then PCR-amplified. To reverse-transcribe and PCR amplify RNAs with unknown sequences, the RNA 3′-end needs to be extended with a known sequence often

referred as a “handle”. This is often done either by poly(A) tailing or by 3′-adapter ligation catalysed by a T4 RNA ligase. As small RNAs are just over 20 bases long, poly(A) tailing will generate a great percentage of poly-A signal. It also introduces the uncertainty as to whether or not there is a terminal A in the original small RNA. In this case, an additional poly C tailing reaction needs to be run in parallel, which further complicates the procedure. So for small RNA sequencing, poly(A) tailing is not suitable.

Both, conventional sequencing of cDNA clones [7] and HTS sequencing including Illumina®, SOLiD® (Sequencing by Oligonucleotide Ligation and Detection) [6], and also Roche® 454 platforms [8] mentioned above require the ligation of a 3′- and 5′-adapter. The adapter sequences provide the handles for reverse transcription and subsequent PCR amplification. In addition, they also provide specified sequences for each sequencing strategy and platform. In the case of the cloning method, the adapter sequence provides the restriction sites for the assembly of concatamers and clones [7]. For the next generation sequencing methods, it provides primer recognition sites for bridge amplification and cluster generation on a flow cell (Illumina®) surface [6] or amplification on a magnetic bead (Roche® 454, SOLiD®) [8].

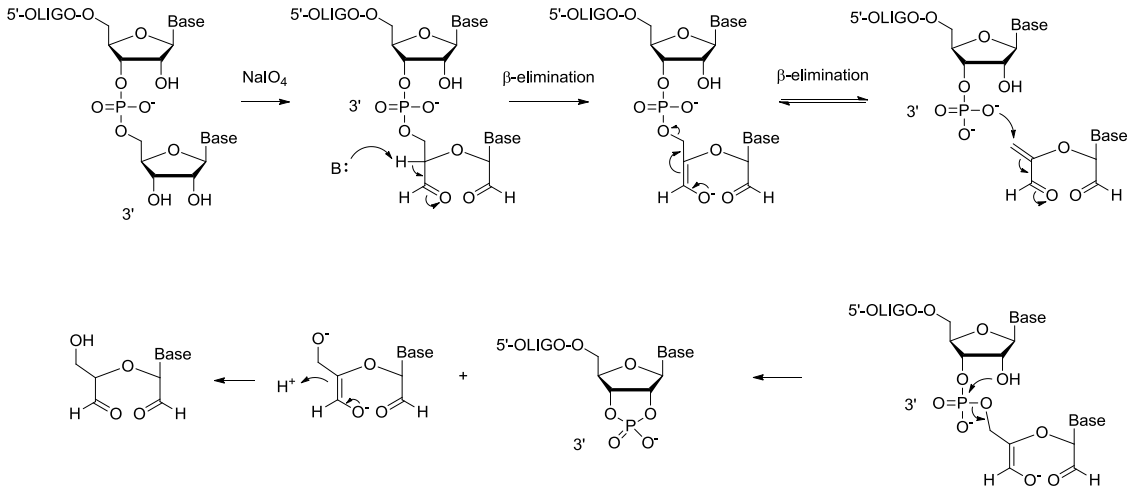
Both DNA [7] and RNA [9,10] have been used as 3′-adapters. Although T4 RNA ligase joins the RNA 3′-end and the adapter DNA 5′-end slightly faster than RNA-RNA ligation [11], and the M-MuLV reverse transcriptase (NEB® #M0253) can read through DNA/RNA junctions (NEB® M-MuLV reverse transcriptase, FAQ), the RNA adapter as the reverse transcription template retains the best reverse transcription efficiency [12], rendering it a better choice. However, direct comparative studies have not been found. In addition, DNA/RNA chimera-based 5′- and 3′-adapters were also used, which have RNA parts at the ends that are to be ligated to targeting RNAs [6].

Various strategies are applied to minimise ligation side-products. The 3′-adapter needs a 5′-phosphate for ligation and a 3′-blocker to prevent self-ligation, which generates an adapter dimer. The 3′-adapter ligation and the 5′-adapter ligation need to be separated by a gel purification step, otherwise the excess of 3′-adapter will ligate to the 5′-adapter. This 5′,3′-adapter dimer needs to be avoided because it is PCR amplifiable. Additional modifications were also made to the protocol to minimise 5′,3′-adapter dimer formation, by digesting excess pre-adenylated 3′-adapter before 5′-adapter ligation (ScriptMiner™, Illumina®). The exact enzyme type used to selectively degrade the 3′-adapter (“Degradase”, EPICENTRE Biotechnologies®) is not certain, but could possibly be an enzyme that recognises and degrades the pre-adenylated RNA.

7.3.2 Sequencing of RNA containing 5′ and 3′ modifications

As mentioned above, small RNAs may possess a 5′-triphosphate and 2′-*O*-Methyl (2′-*O*-Me) modification at the 3′-end. In addition, the viral RNAs, small nuclear RNAs (snRNAs), small nucleaolar RNAs (snoRNAs) and heterogeneous nuclear RNAs (hnRNAs, or pre-mRNA) possess methylated 5′-cap structures (7-methylguanosine cap or hypermethylated 2,2,7-trimethylguanosine cap [13]) [2]. All these modifications may block or impair 5′ or 3′ adapter ligation catalysed by the T4 RNA ligase family and consequently cause sequence loss or bias problem [2].

For 5′-triphosphate or 5′-cap modified RNA, a decapping step involving treatment with tobacco acid pyrophosphatase (TAP) can be performed to transfer the 5′ structure to 5′-monophosphate before adapter ligation [2]. Solving the problem caused by 2′-*O*-Me modification is more challenging as this modification is reported to; severely hinder *E.coli* PAP and Cid1 PUP catalysed polyadenylation; decrease the efficiency of T4 RNA ligase mediated adapter ligation; and moderately impair M-MuLV Reverse Transcriptase (wild type) from reading-through this modification, which can be improved by increasing enzyme concentration. For the third point, the study was done on an RNA-DNA chimera with the 3′-adapter section as DNA. The effect of 2′-*O*-Me on an RNA-only RT template has not been reported [14]. Although it was mentioned that AMV Reverse Transcriptase is less affected by 2′-*O*-Me modification [14], no method has so far been documented to remove the 2′-*O*-Me modification.



Scheme 7.1: Mechanism of 2′,3′-cyclic phosphate end formation through oxidation and β-elimination.

However, in the same study mentioned above [14], optimization of the ligation of 2′-*O*-Me modified 3′-ends was achieved utilizing T4 RNA ligase 2 truncated, and a pre-adenylated adapter. The use of a pre-adenylated adapter evened the ligation efficiency towards 2′-OH end and 2′-*O*-Me end. The recently developed enzymatic preparation method mentioned above (MthRn1, NEB® 5′ DNA Adenylation Kit) was reported to quantitatively transfer ssDNA or ssRNA 5′-phosphate to adenylated end (5′-App) [15], which facilitates the preparation of 5′-adenylated 3′-adapter. Alternatively, methods were developed to selectively build a 2′-*O*-Me modified sequencing library, which converts the 2′,3′-OH end to an unligatable 2′,3′-cyclic phosphate end through $NaIO_4$ oxidation followed by β-elimination (Scheme 7.1) [2,16].

7.4 Modified small RNA sequencing workflow

Sequencing bias of the small RNA profiling results generated by HTS has been reported and systematically studied [17-20]. The bias is generated as target RNAs with a variety of sequences in the sample, possess varying adapter ligation efficiencies. This results in an uneven representation of the RNA population in the sequencing library. The bias affects the reliability of the studies based on the expression level of miRNAs. It is mentioned above that 2′-*O*-Me modification can cause sequencing bias. More profoundly, the bias of miRNA sequencing (which does not have 2′-*O*-Me modifications) is generated by sequence-specific adapter ligation during library preparation. The problem is mainly caused by the secondary structure preference of T4 RNA ligase [20,21]. If the target RNA, as well as the adapter, forms a secondary structure with the ligation site buried inside the structure, ligation of this particular sequence is likely to be poor. As all sequencing platforms use the T4 RNA ligase-mediated adapter ligation procedure, the bias is reproducible and consistent between different platforms [17]. The presence of 3′-end secondary structures of small RNAs, the intramolecular secondary structure of small RNAs and the co-folding of small RNAs and adapters, all reduce T4 RNA ligase-mediated ligation efficiency [18,20].

A current solution to this problem is to use a pool of different adapters so the adapter that will not co-fold with the small RNAs can be ligated efficiently [19-21]. In order to find a better solution, a thermostable 5′ AppDNA/RNA Ligase from the species *Methanobacterium thermoautotrophicum* has recently been made available commercially from NEB® [22,23]. It ligates a 5′ pre-adenylated linker to the 3′-OH end of either RNA or single stranded DNA (ssDNA) at 60-65 ℃ (optimal temperature). This condition minimises the formation of

secondary structures. Surprisingly, this enzyme also ligates RNA with 2´-*O*-methylated 3´-end to 5´-adenylated linkers, with much improved efficiency compared to that of T4 RNA ligase 1 and T4 RNA ligase 2, truncated [22].

In order to find a fundamental solution to this intrinsic problem, Click ligation was hypothesised as a substitute for the enzymatic 3´-adapter ligation in the whole protocol (Figure 7.2). The 3´-adapter ligation step conducted by T4 RNA ligase is the cause of major RNA sequencing bias [20]. Because the copper and the copper-free Click ligations are robust chemical ligation reactions that can be performed in water without salt, it is expected that the chemical ligation itself will generate much less bias. The new sequencing library preparation protocols using Click ligation will, therefore, change the bias-generating point from the ligation step to the enzymatic labelling step. RNA polymerases are expected to be less sensitive to the secondary structures of small RNA and to have different substrate preferences compared to T4 RNA ligases. It may, therefore, lead to sequencing libraries with different preferences and potentially less bias.

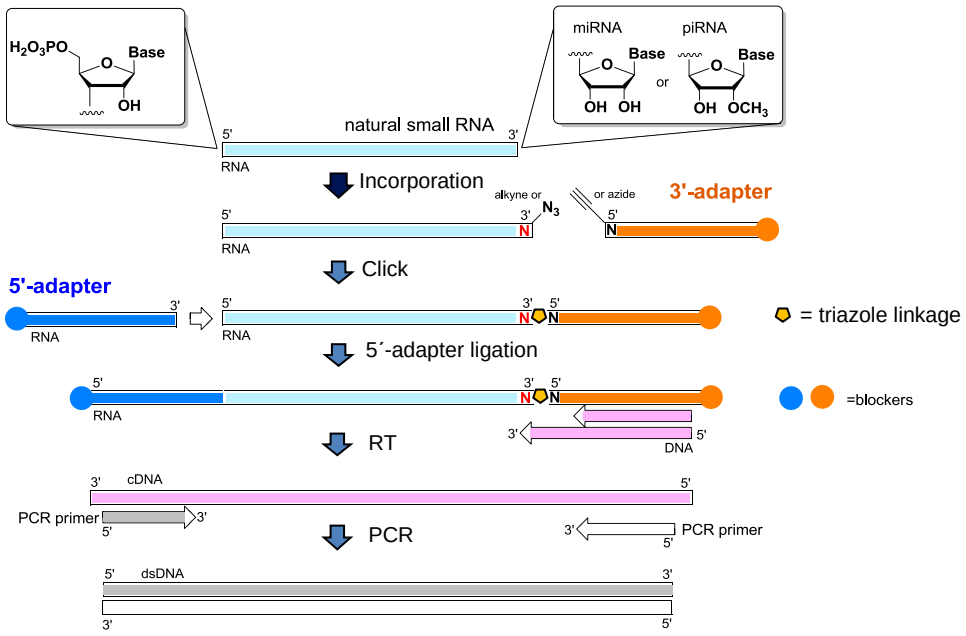


Figure 7.2: The workflow in the modified small RNA sequencing protocol.

As the enzymatic 5´-adapter ligation and 3´-adapter Click ligation are orthogonal to each other, it is possible to design a single adapter to cyclise the small RNAs by ligating both its 3´-end and 5´-end (Figure 7.3). If the 5´-adapter ligation is performed first using an adapter

with 5′-alkyne at the 5′-end, the product will possess a 5′-alkyne at the 5′-end and a 2′-azide at the 3′-end on the same oligonucleotide strain. This may facilitate its cyclisation by Click ligation without the help of a splint (a complementary strand that brings the adapter 5′-end and labelled small RNA 3′-end close to each other). The 5′-alkyne also serves as a blocker to prevent ligase-mediated cyclisation and multiple additions of this 5′-adapter. As a 2′-N_3 is neither a T4 RNA ligase nor a DNA ligase blocker [24], it does not prevent the self-cyclisation and dimerisation of the labelled small RNA themselves. However, because of their short length, cyclisation is unlikely.

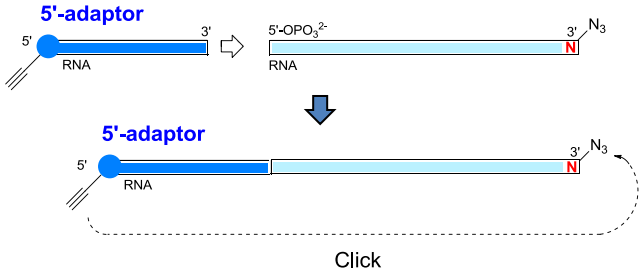


Figure 7.3: A single adapter approach.

7.5 RNA polymerase labelling *vs.* T4 RNA ligase labelling

The RNA polymerases studied above are likely to have different RNA target preferences compared to T4 RNA ligase. Our RNA Click ligation approach may be used to tackle tasks that are difficult for T4 RNA ligase, for example, ligation at the 3′-uridine end [25]. The biggest problem in using T4 RNA ligase to introduce 3′-alkyne or azide for RNA sequencing is that the same enzyme is utilised for enzymatic 3′-adapter ligation. If T4 RNA ligase was chosen for RNA 3′-labelling, the RNA strands that perform poorly in T4 RNA ligase-catalysed conventional adapter ligation, may also be poor substrates for 3′-labelling catalysed by the same enzyme.

However, ligation of short chemically-modified RNA fragments bearing alkyne or azide tags by T4 RNA ligase is expected to be more efficient than the ligation of two long RNA strands. This is because a short RNA is unlikely to co-fold with the target RNA compared to common 3′-adapters containing normally more than 20 nucleotides. The whole workflow, as a consequence, is still expected to generate less bias. A comparative labelling study was also reported using Yeast PAP and T4 RNA ligase 1. The PAGE showed that long RNAs (over 500 nt) were labelled drastically more efficiently by Yeast PAP, while short RNAs (less than 100

nt) were clearly labelled more efficiently by T4 RNA ligase 1 [26]. It indicates that T4 RNA ligase 1 may be more suitable for our RNA sequencing application.

7.6 Triazole backbones generated for reverse transcription studies

The modified small RNA sequencing protocol requires the template containing a triazole linkage (triazole backbone) to be successfully reverse transcribed. A DNA mimic containing triazole-linkages exclusively (without phosphodiester linkages) as a primer was reported to initiate reverse transcription successfully [27]. However, the reading-through of one triazole linkage on an RNA template by a reverse transcriptase has not been reported. In order to study the reverse transcription of triazole backbones thoroughly, a large quantity of the particular backbone is needed. The alkyne or azide labelled RNAs can be made enzymatically as described above. However, to make the approach economic for our reverse transcription studies, the alkyne or azide labelled RNA precursors for making triazole backbones are directly synthesised. The sequence of miR-155 (similar to the RNA sequence in the labelling studies) was selected for the 5′-half of the template with an additional nucleotide next to the triazole resembling triphosphate incorporation. The 3′-half of the template came from the RNA part of the Illumina® 3′-adapter [6]. The alkyne and azide modifications on the 5′-half and 3′-half of the templates were introduced by solid-phase synthesis using methods discussed in the Introduction (Chapter 1) and Experimental sections (Chapter 8, oligonucleotide synthesis section 8.2). The 5′-half of the template with a 3′-end modification was "clicked" with the 3′-half of the template with a 5′-end modification by CuAAC or SPAAC. The Click products (triazole backbones) were purified by PAGE. The details are included in the Experimental section.

The triazole backbones initially studied are Backbone B and its derivative Backbone B′ (Figure 7.4). Backbone B is the most studied backbone with the greatest biocompatibility discussed above. Although introducing 3′-propargyl by RNA polymerase was not achieved, it was successfully introduced to the RNA 3′-end by T4 RNA ligase 1. As described above, 2′-azido-2′-dUTP and 2′-azido-2′-dCTP were incorporated by Yeast PAP and the chain terminated after one addition. Consequently, the capacity of triazole backbone C (Figure 7.4) as a reverse transcription template needs to be investigated.

In addition, the 2′-azide can also ligate to a strained alkyne through copper-free Click ligation. Copper-free Click ligation or SPAAC is an efficient reaction between a strained alkyne and an azide. Although SPAAC ligation generates a bulky BCN-triazole linkage, it does not require

copper catalyst so it can simplify our Click ligation procedure. The use of SPAAC also helps circumvent the copper-catalysed degradation of RNA, which is discussed in section 7.11. The biocompatibility of these new triazole-BCN backbones (Figure 7.4) in this context is currently unknown.

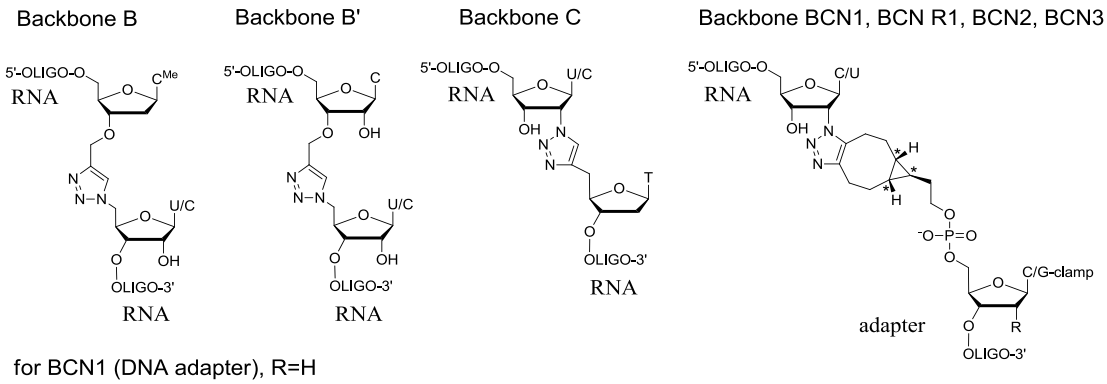


Figure 7.4: Triazole backbones generated in small RNA sequencing study. For BCN Backbones, the centres marked by * represent a mixture of isomers (the bonds facing up can also be facing down), as the copper-free Click generates a mixture of two enantiomers.

7.7 Reverse transcription of triazole backbones generated by CuAAC

Triazole Backbone B´ was selected for reverse transcription evaluation initially as it resembles triazole Backbone B, which showed good bio-compatibility, as discussed above. The commercially available AMV Reverse Transcriptase (NEB® #M0277), Superscript® III (Life Technologies® #18080-044), M-MuLV reverse transcriptase (wild type, NEB® #M0253) as well as M-MuLV reverse transcriptase (RNase H^-) (NEB® #M0368) were selected for the initial tests on Backbone B´ (CzU).

7.7.1 Reverse transcription of triazole Backbone B´ with the sequence "CzU" using M-MuLV Reverse Transcriptase (RNase H^-) and M-MuLV Reverse Transcriptase (wild type)

The mutant M-MuLV Reverse Transcriptase (RNase H^-) is the NEB® version of SuperScript® II (Invitrogen®), an enzyme frequently used for RNA sequencing applications [9,10,28]. It is an engineered enzyme with reduced RNase H^- activity compared to the wild type enzyme. The mutation minimises its degradative effect to the RNA pool [29]. As the unnatural triazole site might hinder reverse transcription, four DNA primers of different lengths were designed. When the primers annealed to the template, they generate reverse transcription starting points (its 3´-end) far before (-3), just before (+0), 1 nucleotide after (+1) and 4 nucleotides after (+4) the triazole point (5´ to 3´ direction, Figure 7.5). In the last two cases, the triazole is bridged by the primers.

The result (Figure 7.5) showed that M-MuLV Reverse Transcriptase (RNase H^-) (NEB®) could not read-through the triazole template but likely stopped at the triazole point (MS analysis of the reaction under this condition was not performed, see below for MS of the wild type enzyme reaction). Only the primers which bridged the triazole point were further extended. In this experiment, the supplied enzyme buffer was used, which contains $MgCl_2$ as the divalent cation.

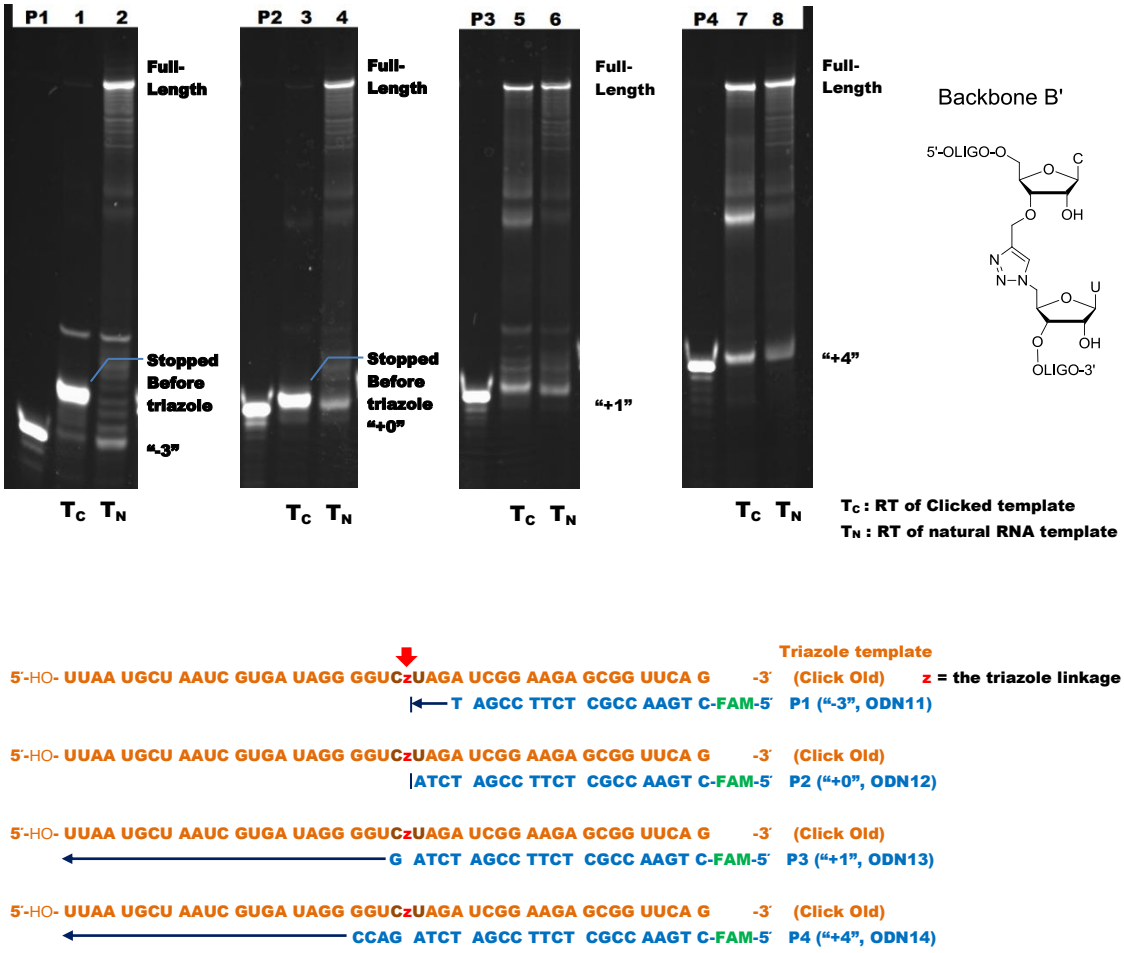


Figure 7.5: Reverse Transcription of the triazole Backbone B´ (CzU) by M-MuLV Reverse Transcriptase (RNase H$^-$) (NEB®) in Mg^{2+}-buffer.

RT template: Click Old (PAGE purification), 60 pmol for lane 1, 3, 5, 7 (T_C).
Control: canonical RNA template ORN13 (CU), 60 pmol for lane 2, 4, 6, 8 (T_N).
T_C: reverse transcription of Click template, T_N: reverse transcription of natural RNA template
Lane P1-P4: primers (with 5´-FAM) in water
Lane P1, 1, 2: 60 pmol primer ODN11; Lane P2, 3, 4: 60 pmol primer ODN12;
Lane P3, 5, 6: 60 pmol primer ODN13; Lane P4, 7, 8: 60 pmol primer ODN14;

Lane 1-8: 3 µM primer (60 pmol), 3 µM template (60 pmol), 1x supplied M-MuLV Reverse Transcriptase (RNase H$^-$) Reaction Buffer (50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$, pH 8.3 at r.t.), 10 mM DTT (NEB®), 0.5 mM dNTP, 200 u (1 µL) M-MuLV Reverse Transcriptase (RNase H$^-$) (NEB®) in total 20 µL. Reverse transcription: 37 °C, 2 h. RNase H digestion was not performed.

To find conditions for the reverse transcriptase to read-through the triazole linkage, $MnCl_2$ was added to the Mg^{2+} buffer. The effect of Mn^{2+} on M-MuLV Reverse Transcriptase (wild type) was studied in 1970, which showed that under certain circumstances, manganese acetate significantly enhanced the reverse transcription activity compared to magnesium acetate [30]. Our result (Figure 7.6) using the mutant reverse transcriptase showed that by applying a

reaction buffer containing Mn^{2+}, the M-MuLV Reverse Transcriptase (RNase H^-) was able to read-pass the triazole site and generate reverse transcription products at the full-length region.

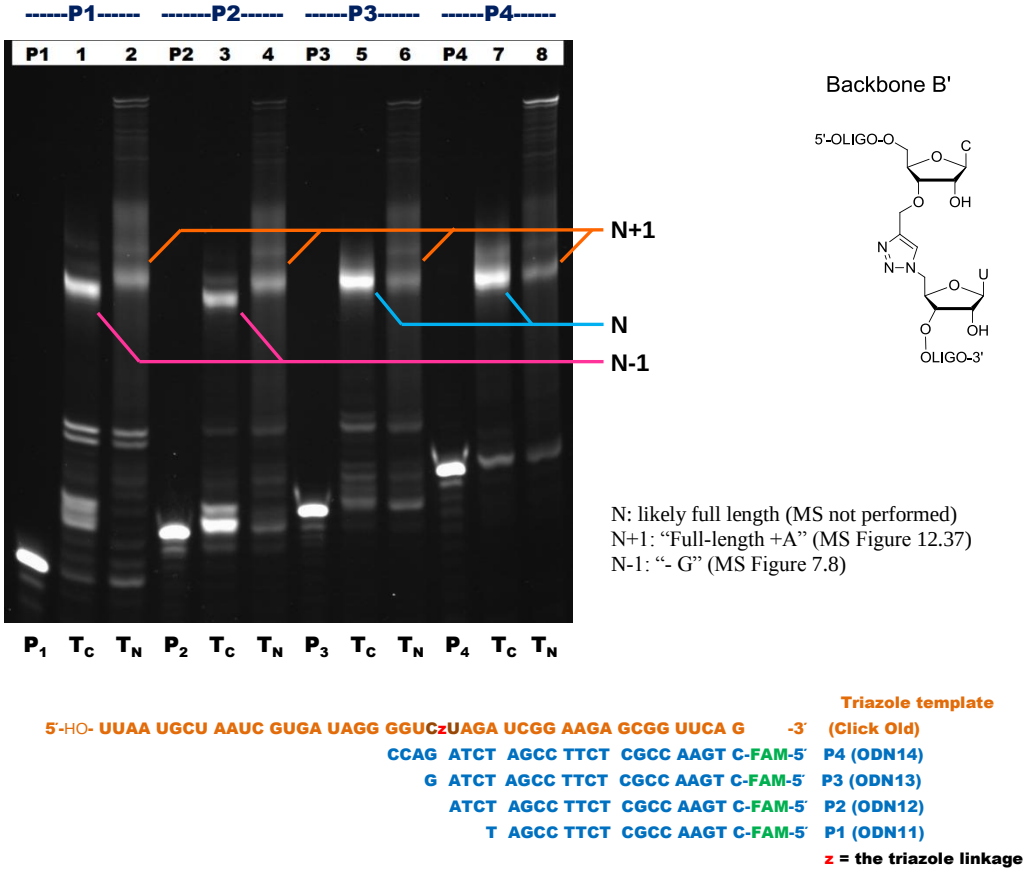


Figure 7.6: Reverse Transcription of the triazole backbone B´ (CzU) by M-MuLV reverse transcriptase (RNase H^-) (NEB®) in Mn^{2+}-buffer.

RT template: Click Old (ORN9 click ORN12, PAGE purification), 60 pmol for Lane 1, 3, 5, 7, T_C.
Control: canonical RNA template ORN13 (CU), 60 pmol for Lane 2, 4, 6, 8, T_N.
T_C: reverse transcription of Click template, T_N: reverse transcription of nature RNA template
Lane P1-P4: primers (with 5´-FAM) in water
Lane P1, 1, 2: 60 pmol primer ODN11; Lane P2, 3, 4: 60 pmol primer ODN12;
Lane P3, 5, 6: 60 pmol primer ODN13; Lane P4, 7, 8: 60 pmol primer ODN14.

Lane 1-8: 3 µM primer (60 pmol), 3 µM template (60 pmol), 1x M-MuLV Reverse Transcriptase (RNase H^-) Reaction Buffer (Mg^{2+}-free, 50 mM Tris-HCl, 75 mM KCl, pH 8.3 at r.t.), 10 mM DTT (NEB), 3 mM $MnCl_2$, 0.5 mM dNTP (Promega), 200 u (1 µL) M-MuLV Reverse Transcriptase (RNase H^-) (NEB) in total 20 µL. Reverse transcription: 37 ℃, 2 h. RNase H digestion was not performed.

The reverse transcription products of the shortest primer ODN10 ("-7", shorter than P1-ODN11) were analysed by MS. The results (Figure 7.7, 7.8) showed that both PAGE purified bands ("N-1") and the desalted PCR products contain a mixture of DNAs representing mainly

a “single G deletion” product (with highest MS intensity) and a “single G deletion +A” product. The “+A” fragment was likely resulted from the terminal transferase activity of the reverse transcriptase. The “no deletion +A” product was also observed with low MS intensity (about 10 % of RT products). It can be assumed that the ratio of the intensity of different MS peaks represents the ratio of different DNA fragments in the mixture. This is because the fragments have similar physical properties, differing typically by only 1 nucleotide.

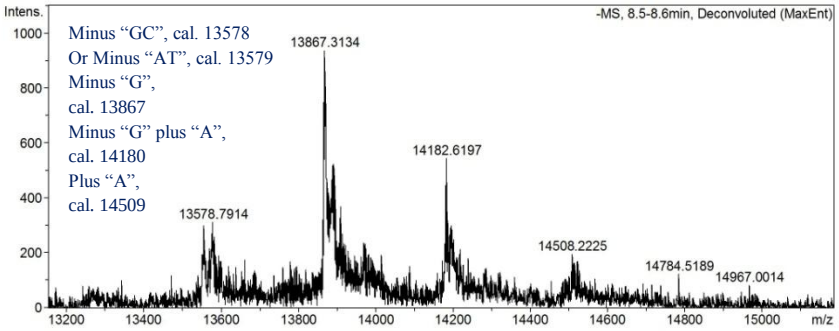


Figure 7.7: M-MuLV Reverse Transcriptase (RNase H^-) RT products (6 mM Mn^{2+}, primer ODN10- “-7”) of the triazole template Click Old (Tc, Backbone B´, CzU) retrieved from the gel (PAGE not shown) (see also Figure 9.35).

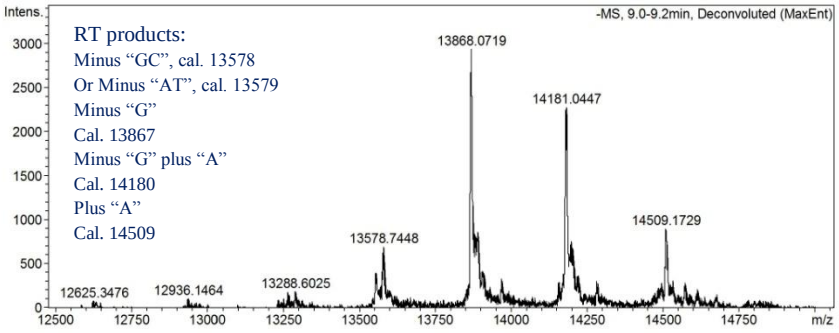


Figure 7.8: M-MuLV Reverse Transcriptase (RNase H^-) RT reaction mixture (3 mM Mn^{2+}, primer ODN10- “-7”) of the triazole template (Click Old, Backbone B´, CzU) desalted by gel-filtration (NAP-10) (see also Figure 9.36).

In the PAGE results above, the bands at the full-length region were significantly smeared. This was more significant when Mg^{2+}-containing reaction buffer was used, in which case a higher band was generated (Figure 7.5). A separate test was conducted using the canonical control RNA template (CC) and the upper bands were cut (PAGE not shown) and analysed by

MS. It was confirmed to be a mixture of RNA template and reverse transcription product (Appendix, Figure 9.40). Therefore, the product probably co-migrated with the template as a duplex on the gel, resulting in smeared bands. The co-migration was also observed during HPLC of the MS samples (see Figure 9.40).

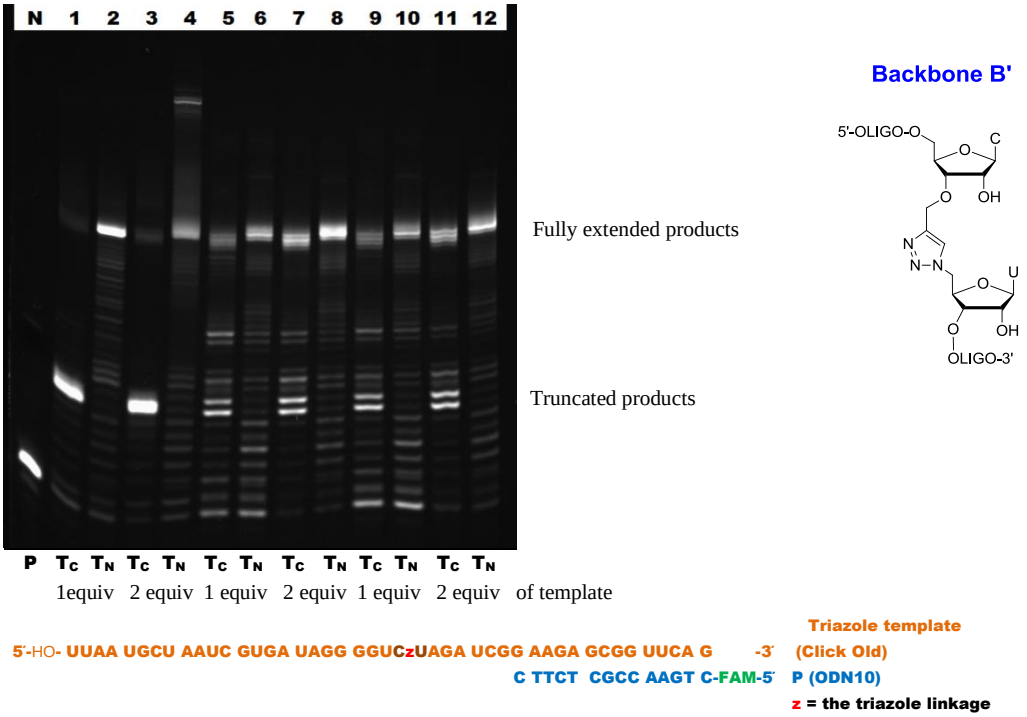


Figure 7.9: M-MuLV Reverse Transcriptase (wild type) reading Backbone B´-CzU.

Lane N: 60 pmol DNA primer P (ODN10 with 5´-FAM) in water

RT template:
Lane 1, 5, 9: 3 µM Click Old (ORN9 click ORN12, 60 pmol, 1 equiv. to primer),
Lane 2, 6, 10: 3 µM ORN13 (canonical RNA template-CU, 60 pmol, 1 equiv. to primer),
Lane 3, 7, 11: 6 µM Click Old (ORN9 click ORN12, 120 pmol, 2 equiv. to primer),
Lane 4, 8, 12: 6 µM ORN13 (canonical RNA template, 120 pmol, 2 equiv. to primer).
T_C: reverse transcription of Click template, T_N: reverse transcription of nature RNA template

Mg^{2+}/Mn^{2+} concentrations:
Lane 1-4: 3 mM $MgCl_2$
Lane 5-8: 3 mM $MgCl_2$ + 3 mM $MnCl_2$
Lane 9-12: 3 mM $MnCl_2$

RT composition: 3 µM primer P, 3/6 µM template, 0.5 mM dNTP, 1x M-MuLV reverse transcriptase reaction buffer (NEB®, $MgCl_2$ and $MnCl_2$ concentration adjusted), 10 mM DTT, 200 u M-MuLV Reverse Transcriptase (NEB®) in total 20 µL.
Reverse transcription: 37 ℃ for 1 h.

M-MuLV Reverse Transcriptase (wild type) (NEB® #M0253)was tested for all triazole backbones generated by CuAAC. The wild type enzyme possesses a higher level of endonuclease activity [31]. The RNase H domain was active in Mg^{2+}-buffer as well as Mn^{2+}-buffer and it significantly degraded both natural RNA and the triazole RNA templates, which was confirmed by MS (Mg^{2+}: Figure 9.41 - Click Old template, Figure 9.60 - control "UU" template; Mn^{2+}: Figure 7.10 - Click Old template, Figure 9.39 - control "CU" template). As the degraded template no longer interfered with the DNA products during PAGE, clearer product bands representing the mixture were observed (Figure 7.9). Starting from the shortest "-7"-primer (ODN10), M-MuLV Reverse Transcriptase (wild type) in Mn^{2+} buffer generated a mixture of reverse transcription products (MS, Figure 7.10) with similar composition as that obtained using M-MuLV Reverse Transcriptase (RNase H⁻) (Figure 7.8).

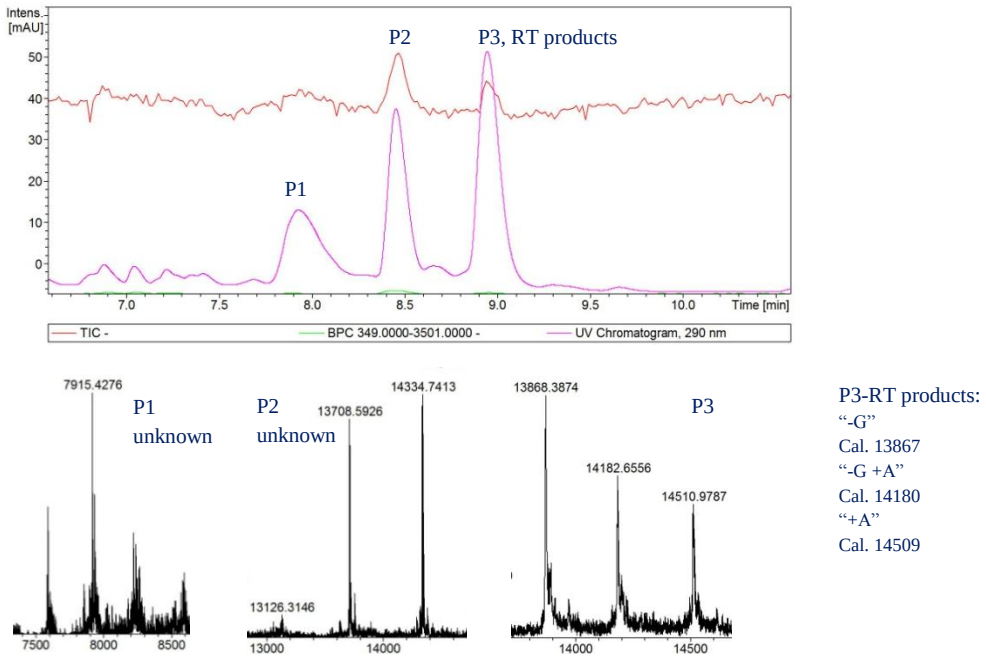


Figure 7.10: M-MuLV Reverse Transcriptase (wild type) RT reaction mixture (3 mM Mn^{2+}, primer ODN10) of the triazole template Click Old (Tc, Backbone B´, CzU) desalted by gel-filtration (NAP-10) (see also Figure 9.38). The unknown peaks are likely derived from the template.

As mentioned in the introduction section above, triazole Backbone B can be read-through by DNA polymerases (CzC, GzG) [32] and T7 RNA polymerase (CzT) [33]. The blocking of this backbone (CzU) to reverse transcriptase is possibly due to the insufficient duplex stability at

the triazole site. Different backbones with the sequence "CzC" instead of "CzU" on each side of the triazole linkage were synthesised and their reverse transcription results are discussed below.

For our RNA sequencing application, precise read-through of the triazole linkage is in fact not required. The enzymatic reaction to introduce alkyne/azide adds at least one known nucleotide. Thus the reverse transcription primer can be designed to bridge the triazole linkage by extending over it for at least one nucleotide. If this bridging primer can be extended, the omission is also likely to be circumvented. It indicates the primer P3 (+1) in the above experiments (Figure 7.5, 7.6) is suitable for our RNA sequencing application because it successfully initiated reverse transcription at the "+1" site. Evaluation of these reverse transcription primers are described further in the following sections below.

7.7.2 Reverse transcription of triazole Backbone B´ with the sequence "CzU" using AMV Reverse Transcriptase and Superscript® III Reverse Transcriptase

AMV Reverse Transcriptase and Superscript® III Reverse Transcriptase were also tested for reverse transcription of triazole Backbone B´ (CzU). However, they did not read-through the triazole linkage (Figure 7.11). The result was not improved by using buffer containing Mn^{2+}. These two enzymes were not subsequently tested for other backbones mentioned below.

In addition, commercially available HIV1 Reverse Transcriptase (Active, Abcam® #ab63979) was also tested to read this backbone. However, it showed strong DNase activity that partially degraded the FAM labelled primer (PAGE not shown), so it was not further tested.

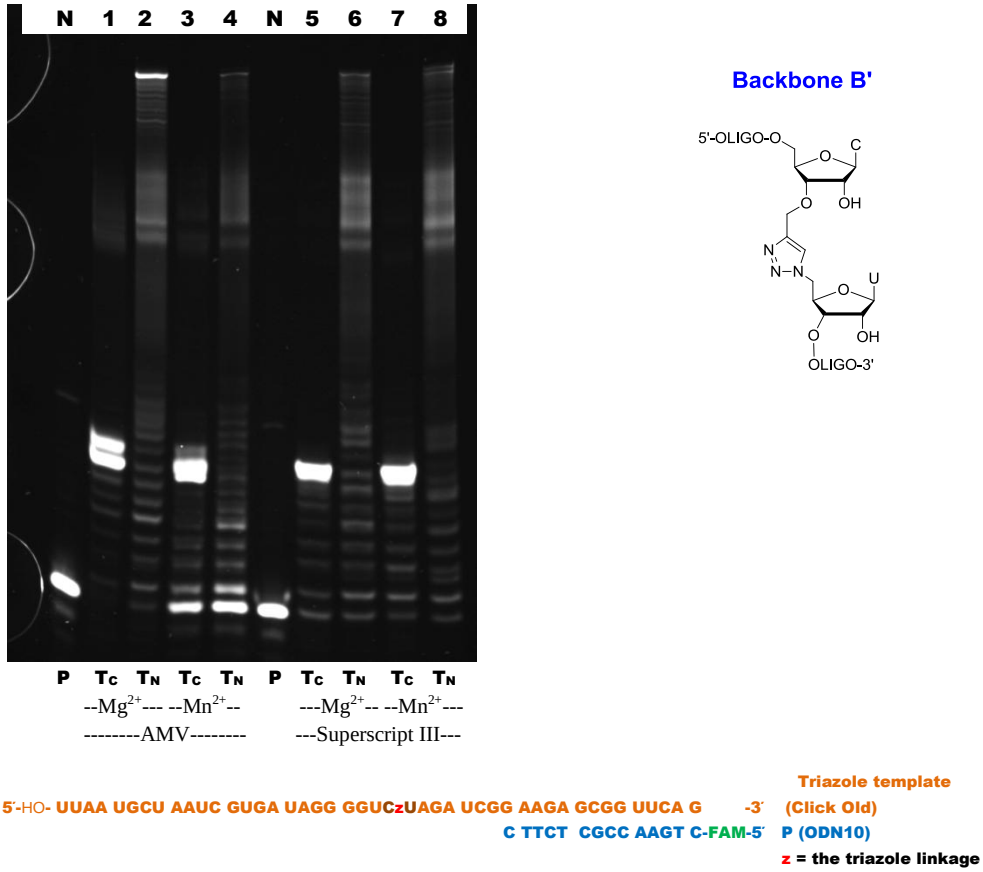


Figure 7.11: AMV Reverse Transcriptase and Superscript® III Reverse Transcriptase test on Backbone B´ (CzU)

RT primer: ODN10 (5´-FAM, 60 pmol)

RT template: Click Old (ORN9 click ORN12, CzU), 120 pmol for Lane 1, 3, 5, 7, 9 (2 equiv. to the primer).

Control: canonical RNA template ORN13, 120 pmol for Lane 2, 4, 6, 8, 10 (2 equiv. to the primer).

Lane N: 5´-FAM labelled primers ODN10 in water

T_C: reverse transcription of Click template, T_N: reverse transcription of nature RNA template

Lane N, 1-4: AMV reverse transcriptase overnight test

Lane 1-4: 3 µM primer (60 pmol), 6 µM template (120 pmol), 0.5 mM dNTP, 1x Mg^{2+}-free AMV RT reaction buffer (50 mM Tris-HCl, 75 mM KOAc, pH 8.3, DTT-free), 10 mM DTT, 8 mM $MgCl_2$ (Lane 1,2) or 8 mM $MnCl_2$ (Lane 3,4), 10 u (1 µL) AMV Reverse Transcriptase (NEB®) in total 20 µL.

Reverse transcription time at 37 ℃ overnight.

Lane N, 5-8: Repeat Superscript® III reverse transcriptase test

For lane 5,6: 1x supplied First-Strand Buffer (50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$, pH 8.3 at r.t.),

For lane 7,8: 1x home-made Mg^{2+}-free M-MuLV RT reaction buffer (50 mM Tris-HCL, 75 mM KCl, pH 8.3 at r.t.), 3 mM $MnCl_2$,

lane 5-8: 3 µM primer (60 pmol), 6 µM template (120 pmol), 0.5 mM dNTP, 10 mM DTT, 200 u (1 µL) Superscript™ III (Invitrogen®) in total 20 µL.

Reverse transcription at 37 ℃ for 1 h.

7.7.3 M-MuLV Reverse Trascriptase (wild type) reverse transcription comparison of Backbone B´ and Backbone B with the sequences "CzU" and "CzC"

In order to study the effect of the sequences on each side of the triazole, backbones with the sequence "CzU" and "CzC" were tested together. Triazole Backbone B without the 2´-OH at the triazole site is the published backbone for read-through by DNA polymerases [32] and T7 RNA polymerase [33], as mentioned above. The new triazole Backbone B´, as a variation of the published backbone, results from T4 RNA ligase labeling with 3´-*O*-propargyl labelled 2-4mers (described above). The effect of 2´-OH on triazole Backbone B´ is thus, important to know. The result (Figure 7.12-7.16) showed no read-through difference between Backbone B and Backbone B´. However, the sequences at the triazole site determined whether the triazole template can be read-through by M-MuLV reverse transcriptase in Mg^{2+}-buffer. The triazole Backbone B or B´ with the sequencing "CzC" were read-through by M-MuLV Reverse Transcriptase (wild type) while the "CzU" caused blockage.

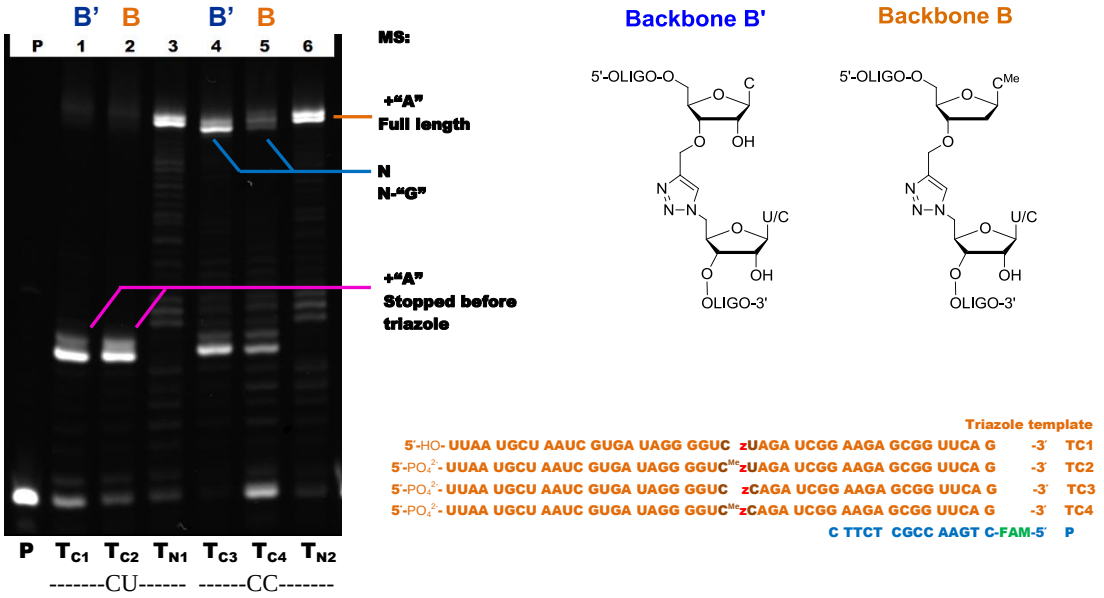


Figure 7.12: M-MuLV Reverse Transcriptase (Wild type) test in Mg^{2+} buffer for read-through of various triazole backbones.

RT primers (with 5´-FAM): ODN10, 60 pmol for each reaction. Lane P: 60 pmol primer in water

Lane P: 60 pmol primer in water

RT template, 60 pmol for each reaction:

Lane 1: Click Old (CzU, Triazole backbone B´); Lane 2: Click B (CzU, Triazole backbone B, impure, Figure 12.30);

Lane 3: control (ORN13, CU);

Lane 4: Click D (CzC, Triazole backbone B´); Lane 5: Click E (CzC, Triazole backbone B);

Lane 6: control (ORN15, CC).

Lane 1-6: 3 µM primer (60 pmol), 3 µM template (60 pmol), 1x supplied M-MuLV Reverse Transcriptase (RNase H^-) Reaction Buffer (50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$, pH 8.3 at r.t.), 10 mM DTT, 0.5 mM dNTP, 200 u (1 µL) M-MuLV Reverse Transcriptase (wild type, NEB®) in total 20 µL. Reverse transcription: 37 ℃, 2 h.

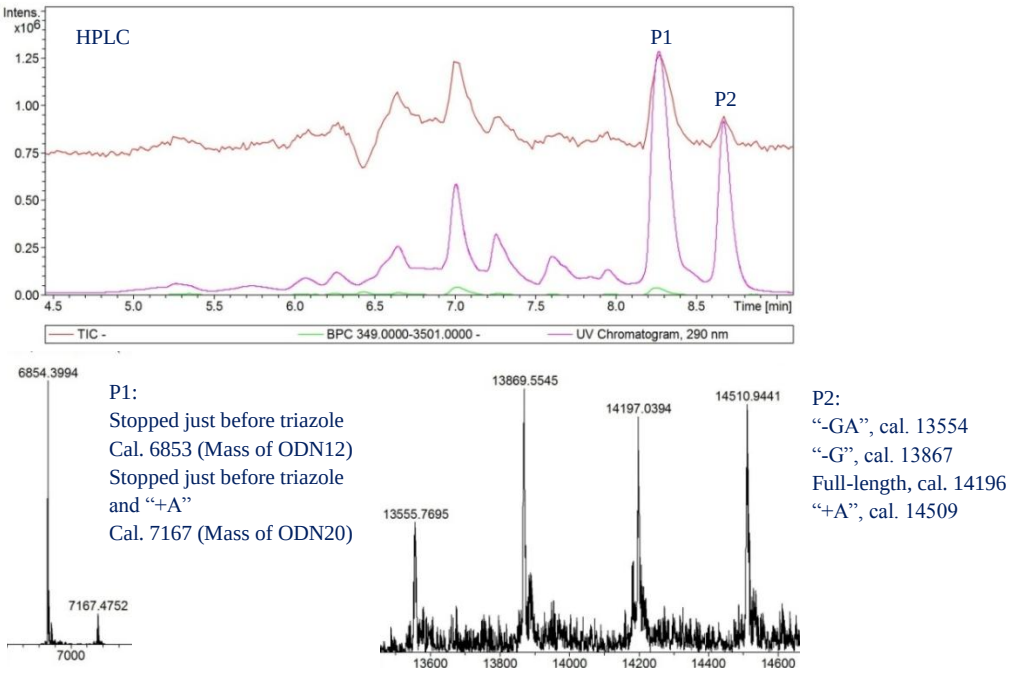


Figure 7.13: M-MuLV reverse transcriptase (wild type) RT reaction mixture (3 mM Mg^{2+}, primer ODN10-"-7") of the triazole template Click Old (Backbone B´, CzU) after incubation at 37 ℃ for 2 h then desalting by gel-filtration (NAP-10) (see also Figure 9.41).

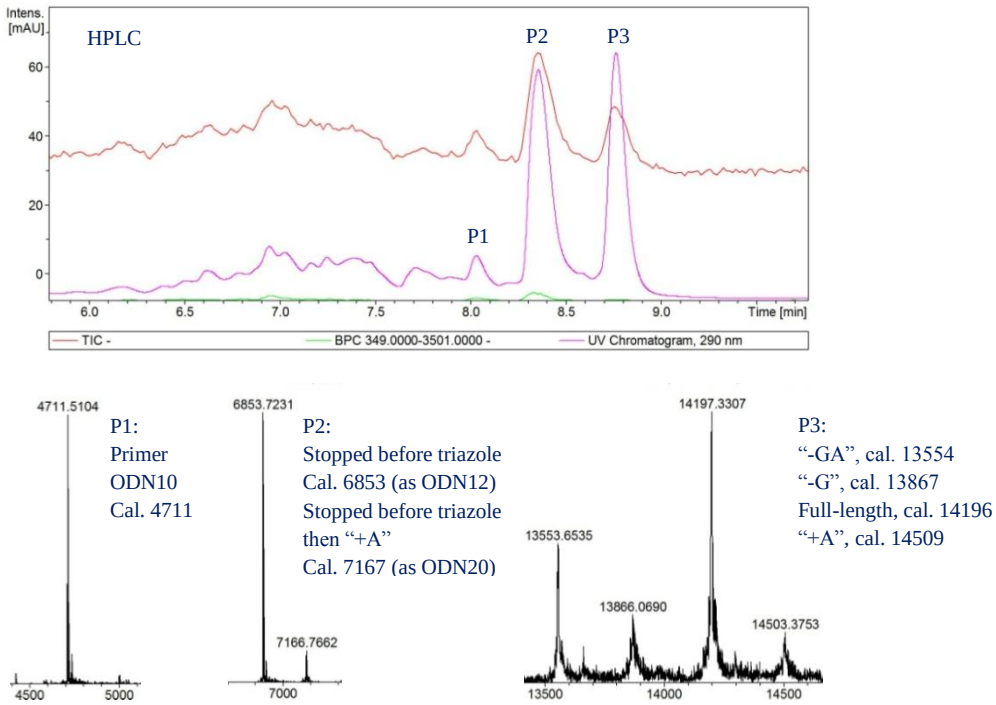


Figure 7.14: M-MuLV reverse transcriptase (wild type) RT reaction mixture (3 mM Mg^{2+}, primer ODN10) of the triazole template Click B (Backbone B, CzU) after incubation at 37 ℃ for 2 h then desalting by gel-filtration (NAP-10) (see also Figure 9.45).

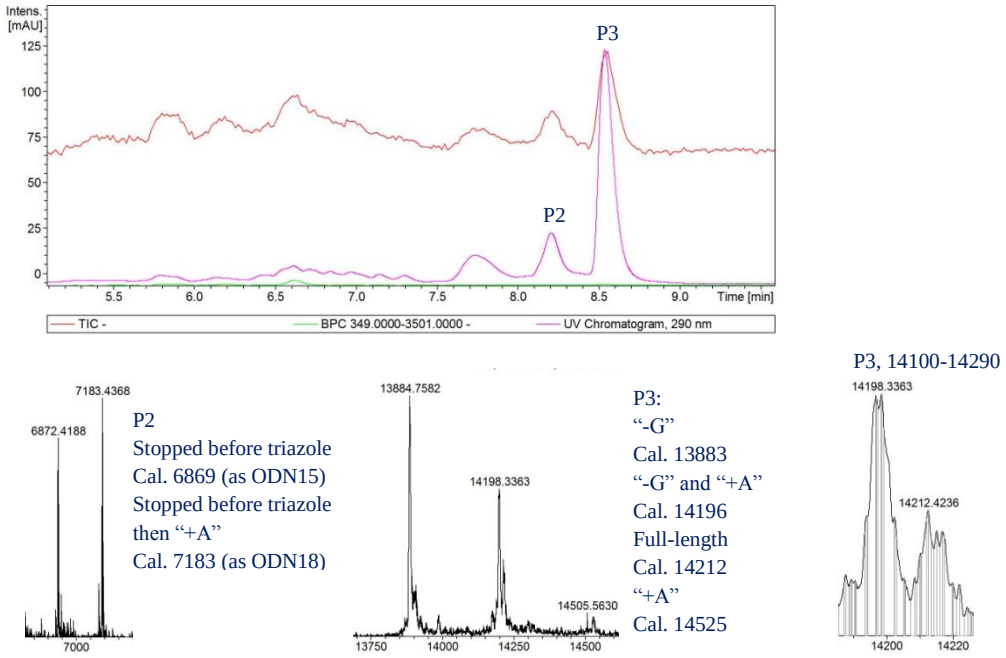


Figure 7.15: M-MuLV reverse transcriptase (wild type) RT reaction mixture (3 mM Mg^{2+}, primer ODN10) of the triazole template Click D (Backbone B´, CzC) after incubation at 37 ℃ for 2 h then desalting by gel-filtration (NAP-10) (see also Figure 9.46).

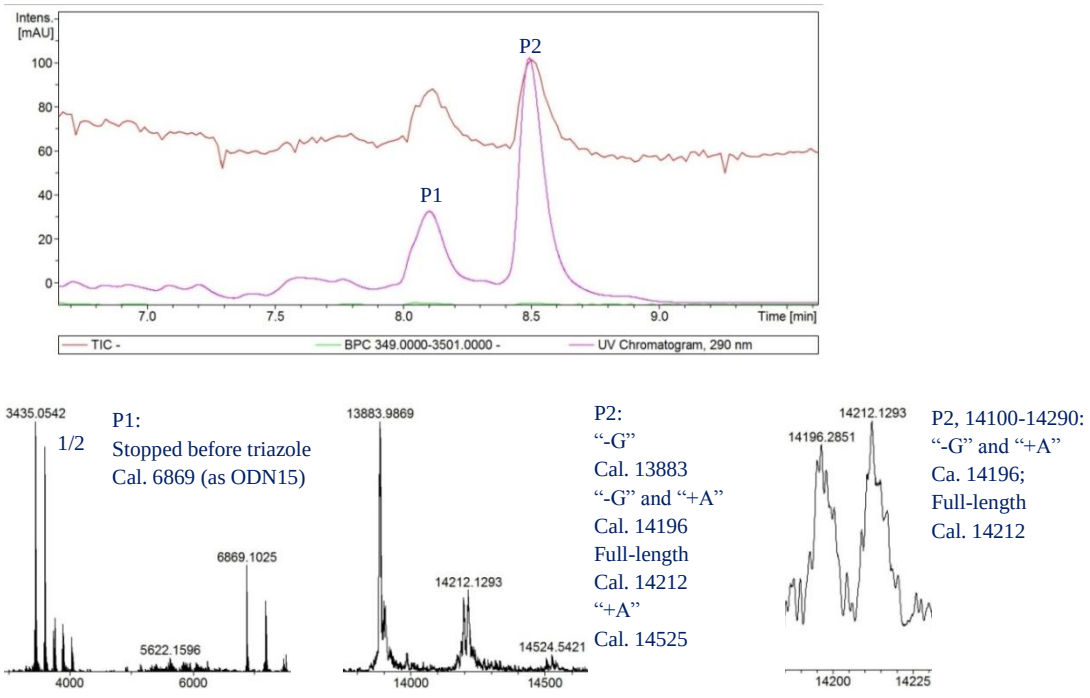


Figure 7.16: M-MuLV reverse transcriptase (wild type) RT reaction mixture (3 mM Mg^{2+}, primer ODN10) of the triazole template Click E (Backbone B, CzC) after incubation at 37 ℃ for 2 h then desalting by gel-filtration (NAP-10) (see Figure 9.50).

It was also found that using the longer primer, with the 3′-end 3 nucleotides before the triazole site (“-3”-primer, Figure 7.17) instead of 7 nucleotides before the triazole site (“-7”-primer, Figure 7.12), the primer was extended past the triazole site of “CzU” but still prematurely stopped before reaching the end (Figure 7.17). This difference between “-3”-primer and “-7”-primer was confirmed by another experiment displayed on a single gel (PAGE not shown). The reverse transcription of “CzC” backbone, however, was comparable to that of the canonical RNA control.

By using the primer with the 3′-end one nucleotide after the triazole (“+1”-primer, Figure 7.18), the reverse transcription was restored and full-length product and its “plus A” derivative were observed by PAGE and identified by MS (Figure 9.44, Appendix).

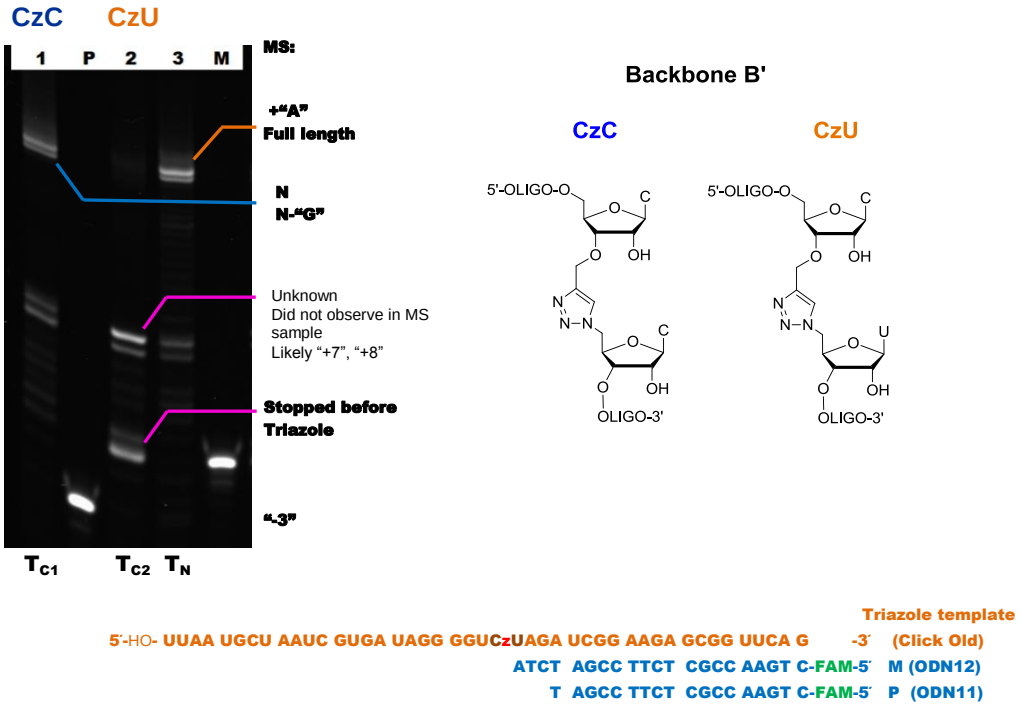


Figure 7.17: M-MuLV Reverse Transcriptase (Wild type) test in Mg^{2+} buffer for read-through of triazole backbone B′ (CzC and CzU)

RT primers (with 5′-FAM): ODN11, 60 pmol for each reaction.

RT template: T_{C1}: Click D (triazole B′, CzC), 60 pmol for lane 1; T_{C2}: Click Old (triazole B′, CzU), 60 pmol for lane 2;
T_N: Control: canonical RNA template ORN15 (CC), 60 pmol for lane 3.

Lane P: 60 pmol primer ODN11 in water;

Lane M: marker (60 pmol ODN12).

The reverse transcription condition is the same as that in Figure 7.12.

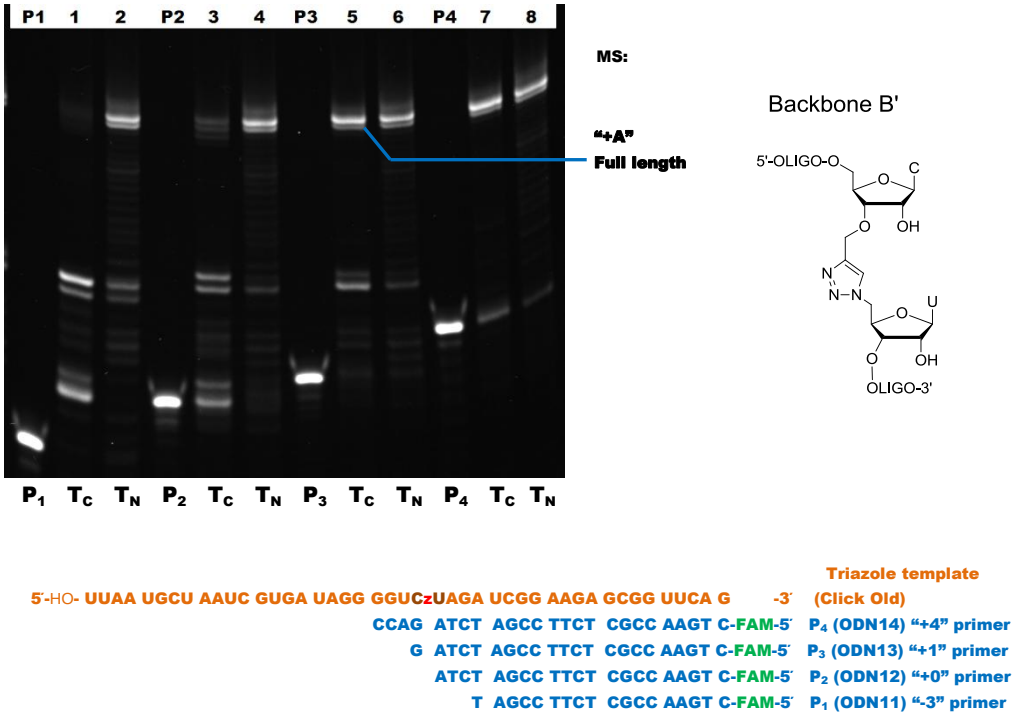


Figure 7.18: M-MuLV Reverse Transcriptase (wild type) test in Mg^{2+} buffer for read-through of triazole backbone B´ (CzU) with 4 primers.

RT template: Click Old (ORN9 click ORN12, CzU), 60 pmol for Lane 1, 3, 5, 7.
Control: canonical RNA template ORN15 (CC), 60 pmol for Lane 2, 4, 6, 8.

Lane P1-P4: primers (with 5´-FAM) in water
Lane P1, 1, 2: 60 pmol primer ODN11; Lane P2, 3, 4: 60 pmol primer ODN12;
Lane P3, 5, 6: 60 pmol primer ODN13; Lane P4, 7, 8: 60 pmol primer ODN14.
T_C: reverse transcription of Click template (Click Old), T_N: reverse transcription of nature RNA template.

The reverse transcription condition is the same as that in Figure 7.12:
Lane 1-8: 3 μM primer (60 pmol), 3 μM template (60 pmol), 1x supplied M-MuLV Reverse Transcriptase (RNase H^-) Reaction Buffer (50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$, pH 8.3 at r.t.), 10 mM DTT, 0.5 mM dNTP, 200 u (1 μL) M-MuLV Reverse Transcriptase (wild type, NEB®) in total 20 μL. Reverse transcription: 37 ℃, 2 h.

The reverse transcription of Backbone B (CzC) and Backbone B´ (CzC) was further investigated (Figure 7.19, 7.21). For these two backbones, although the primers with the 3´-end before the triazole can extend pass the triazole site, it caused partial one "G" deletion, which is confirmed by MS (Backbone B´- CzC: Figure 7.15, Backbone B- CzC: Figure 7.16). The primer extension after the triazole of Backbone B´ (CzC) did not cause deletion (Figure 7.20), which is the same as that observed for Backbone B (CzC) (Figure 7.22). Because the only "C" within 4 nucleotides on either side of the triazole is the "C" right after the triazole linkage, the reverse transcriptase likely omitted this "C" but continually extended

the primer to the end of the template. The deletion at the "+1" site was also confirmed by the observation that "+0" primer caused deletion and the "+1" primer (which has the nucleotide at the "+1" site) caused no deletion. The single base omission was not observed for triazole backbone B during PCR [32] and transcription [33]. So this deletion is unique for reverse transcription and was observed for reverse transcription of both Backbone B and Backbone B´.

The triazole backbones generated much less "+A" (terminal transfer) product compared to canonical RNA template. This is possibly because the reverse transcription on these backbones is slower and the terminal addition of A is also a slower process compared to polymerisation. Therefore, 2 h incubation time may not be enough for the reverse transcriptase to add an additional "A" to the products of the triazole backbones.

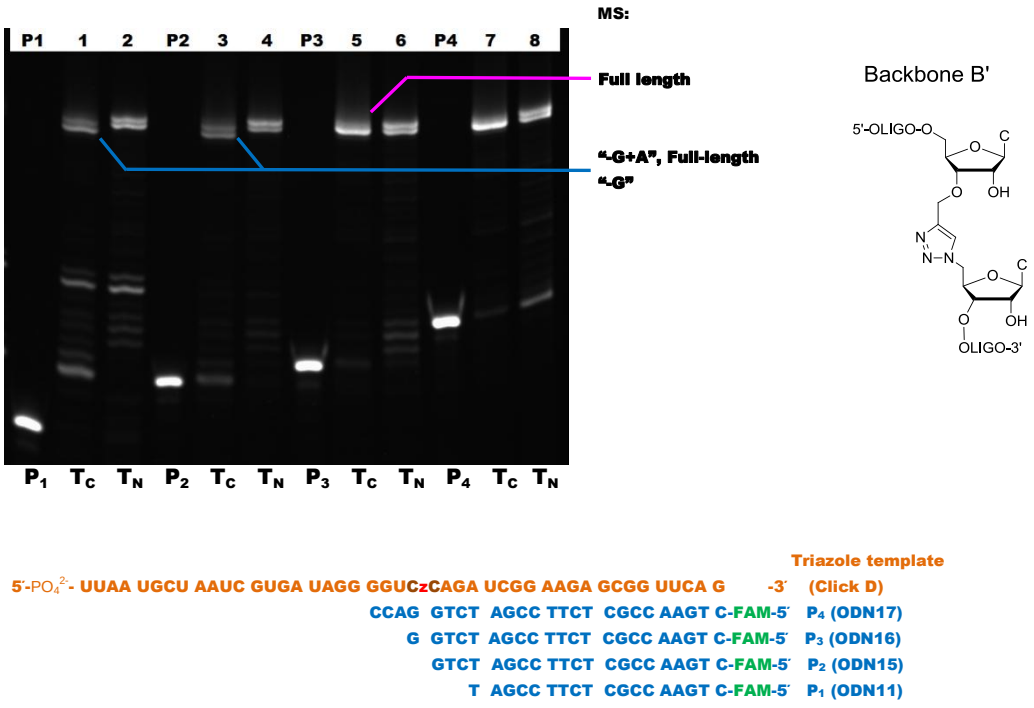


Figure 7.19: M-MuLV Reverse Transcriptase (Wild type) test in Mg^{2+}-buffer for read-through of triazole backbone B´ (CzC) with 4 primers.

RT template: Click D (ORN10 click ORN14), 60 pmol for Lane 1, 3, 5, 7;
Control: canonical RNA template ORN15 (CC), 60 pmol for Lane 2, 4, 6, 8.

Lane P1-P4: primers (with 5´-FAM) in water
Lane P1, 1, 2: 60 pmol primer ODN11; Lane P2, 3, 4: 60 pmol primer ODN15;
Lane P3, 5, 6: 60 pmol primer ODN16; Lane P4, 7, 8: 60 pmol primer ODN17.
T_C: reverse transcription of Click template, T_N: reverse transcription of nature RNA template.

Lane 1-8: same reverse transcription condition as that in Figure 7.12.

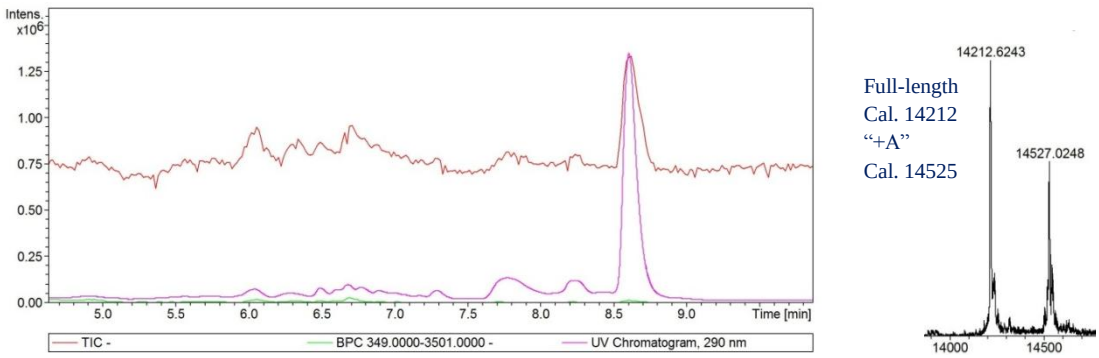


Figure 7.20: M-MuLV reverse transcriptase (wild type) RT reaction mixture (3 mM Mg^{2+}, primer ODN16-"+1") of the triazole template Click D (Backbone B´, CzC) after incubation at 37 ℃ for 2 h then desalting by gel-filtration (NAP-10) (see also Figure 9.49).

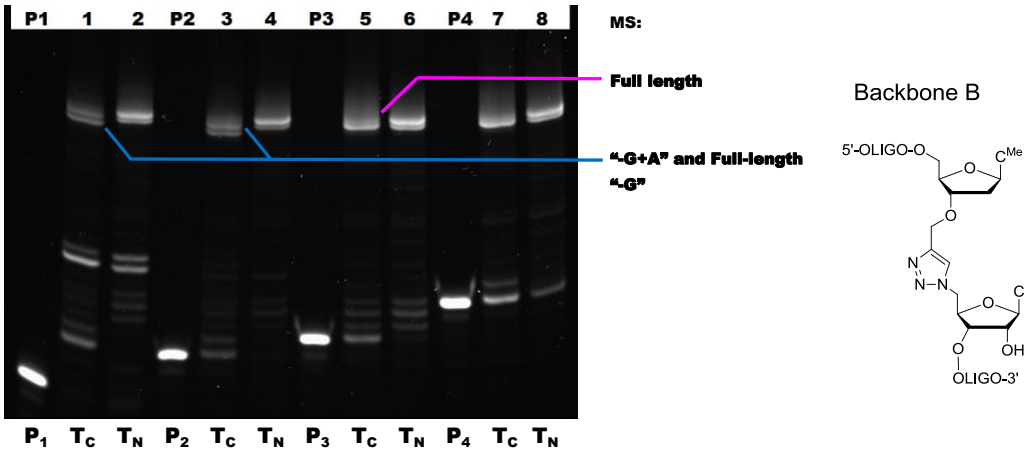


Figure 7.21: M-MuLV Reverse Transcriptase (Wild type) test in Mg^{2+}-buffer for read-through of triazole backbone B (CzC) with 4 primers

RT template: Click E (ORN11 click ORN14), 60 pmol for Lane 1, 3, 5, 7, T_C.
Control: canonical RNA template ORN15 (CC), 60 pmol for Lane 2, 4, 6, 8, T_N.

Lane P1-P4: primers (with 5´-FAM) in water
Lane P1, 1, 2: 60 pmol primer ODN11; Lane P2, 3, 4: 60 pmol primer ODN15;
Lane P3, 5, 6: 60 pmol primer ODN16; Lane P4, 7, 8: 60 pmol primer ODN17.
T_C: reverse transcription of Click template, T_N: reverse transcription of nature RNA template.

Lane 1-8: same reverse transcription condition as that in Figure 7.12.

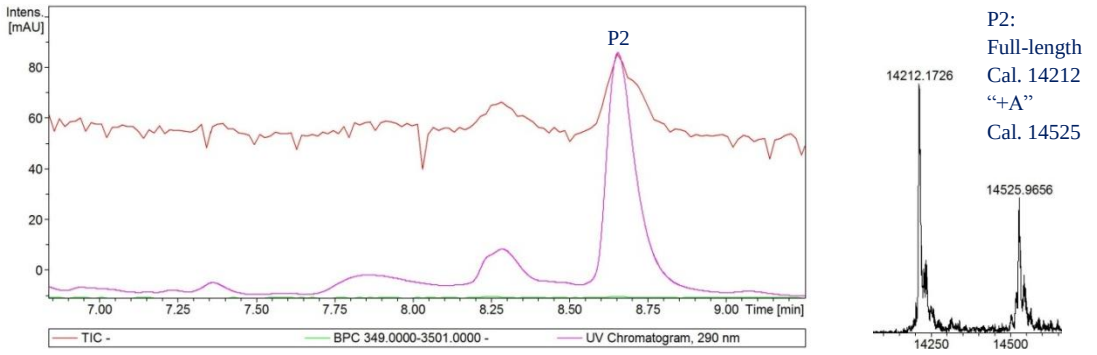


Figure 7.22: M-MuLV reverse transcriptase (wild type) RT reaction mixture (3 mM Mg^{2+}, primer ODN16-"+1") of the triazole template Click E (Backbone B, CzC) after incubation at 37 ℃ for 2 h then desalting by gel-filtration (NAP-10) (see also Figure 9.53).

7.7.4 Reverse transcription of Triazole Backbone C

Triazole Backbone C can be made through enzymatic incorporation of 2′-azido-2′-dUTP/dCTP by Yeast PAP. Successful reverse transcription of this backbone will allow RNA polymerase labelled natural RNA sample to be used for sequencing library preparation.

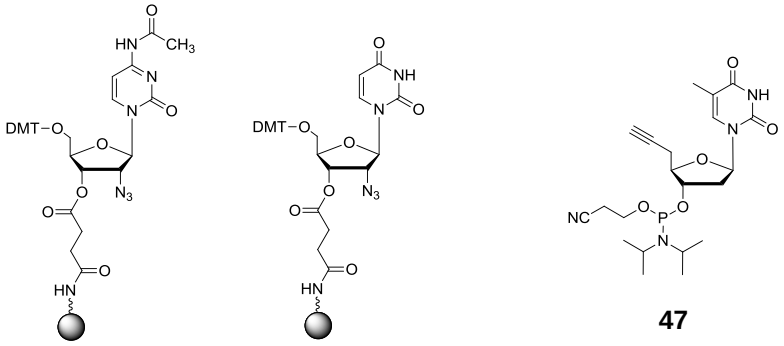


Figure 7.23: 5′-*O*-(4,4′-Dimethoxytrityl)-2′-azido-2′-deoxycytidine (*N*-Ac) [34] and 5′-*O*-(4,4′-dimethoxytrityl)-2′-azido-2′-deoxyuridine [35] functionalised resins for the synthesis of 2′-azido RNAs, and unpublished 5′-alkyne T phosphoramidite (**47**).

The 2′-azido-RNAs were synthesised from 2′-azido-2′-deoxynucleotide functionalised resins (Figure 7.23**).** The oligonucleotide chains were assembled from 3′ to 5′ *via* standard phosphoramidite chemistry. During the synthesis of one of the four 2′-azido RNAs of the same length, a side product was identified by MS of the crude product as a minor HPLC shoulder. It has a mass of "-26" of the M.W. of the azido-RNA (Appendix, Figure 9.28) and

was expected to be the 2′-amino RNA, which was a reduction product from 2′-azido RNA *via* Staudinger reaction with the P^{III} species. This side product was separated and removed by HPLC.

The 5′-alkyne nucleotide phosphoramidites are currently not commercially available, and the only monomer available in our group is the 5′-alkyne T phosphoramidite (unpublished, **47**, Figure 7.23). This limited the sequence on the 3′-adapter side of the triazole to be "T" in this study of triazole Backbone C. The backbone with the sequence "UzT" was tested first. The result showed that the extension from the point before the triazole gave two clean bands (Figure 7.24) and two HPLC peaks (1:1) corresponding to one "A" deletion product and its "+A" derivative ("-A+A") (Figure 7.25).

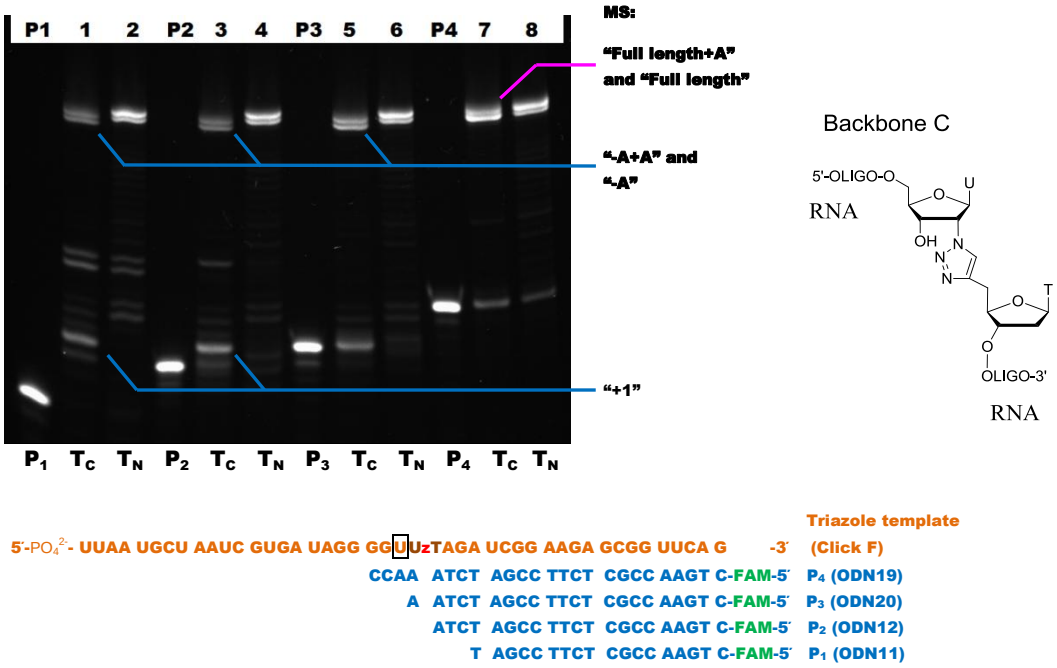


Figure 7.24: M-MuLV Reverse Transcriptase (wild type) test in Mg^{2+} buffer for read-through of triazole backbone C (UzT) with 4 primers

RT template: Click F (ORN16 click ORN19, UzT), 60 pmol for Lane 1, 3, 5, 7, T_C.
Control: canonical RNA template ORN20 (UU), 60 pmol for Lane 2, 4, 6, 8, T_N.

Lane P1-P4: primers (with 5′-FAM) in water
Lane P1, 1, 2: 60 pmol primer ODN11; Lane P2, 3, 4: 60 pmol primer ODN12;
Lane P3, 5, 6: 60 pmol primer ODN20; Lane P4, 7, 8: 60 pmol primer ODN19.
T_C: reverse transcription of Click template, T_N: reverse transcription of nature RNA template.

Lane 1-8: same reverse transcription condition as that in Figure 7.12.

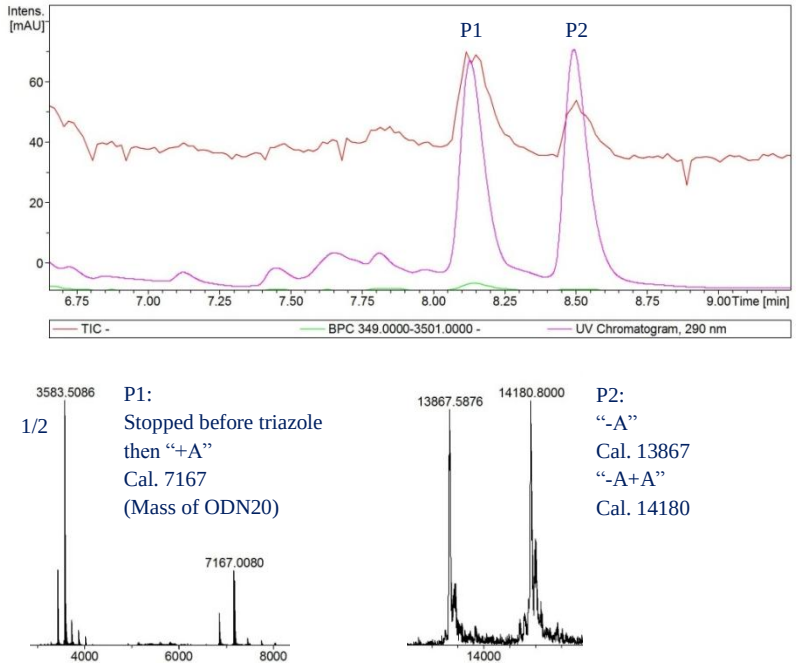


Figure 7.25: M-MuLV reverse transcriptase (wild type) RT reaction mixture (3 mM Mg^{2+}, primer ODN10-"-7") of the triazole template Click F (Backbone C, UzT) after incubation at 37 ℃ for 2 h then desalting by gel-filtration (NAP-10) (see also Figure 9.57).

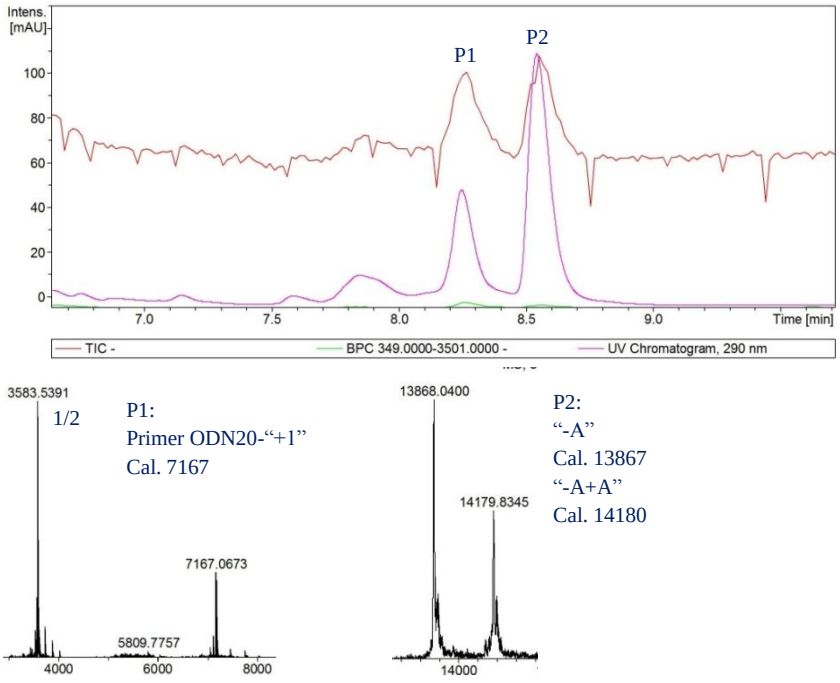


Figure 7.26: M-MuLV reverse transcriptase (wild type) RT reaction mixture (3 mM Mg^{2+}, primer ODN20-"+1") of the triazole template Click F (Backbone C, UzT) after incubation at 37 ℃ for 2 h then desalting by gel-filtration (NAP-10) (see also Figure 9.58).

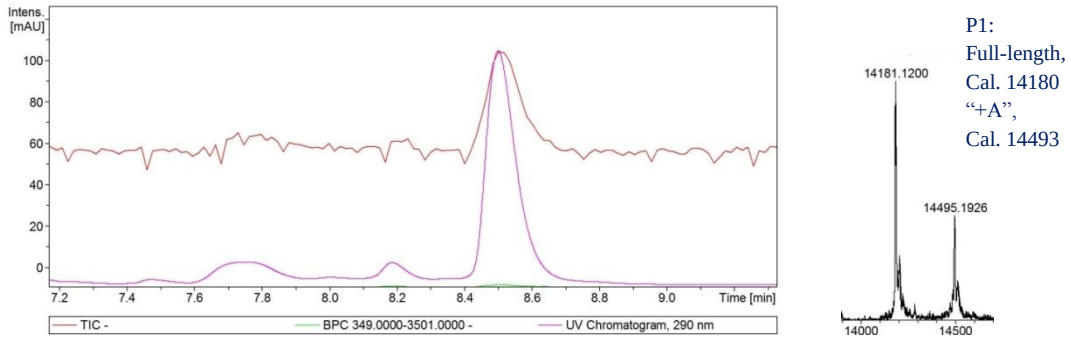


Figure 7.27: M-MuLV reverse transcriptase (wild type) RT reaction mixture (3 mM Mg^{2+}, primer ODN19-"+4") of the triazole template Click F (Backbone C, UzT) after incubation at 37 ℃ for 2 h then desalting by gel-filtration (NAP-10) (see also Figure 9.59).

Although the "-A+A" product can be misassigned as the "full-length" product, which has the same mass, it is unlikely to be the case. The extension using the primer 4 nucleotides after the triazole (+4) gave "full-length" product and its "+A" derivative (Figure 7.27). So the triazole did not completely block the "+A" effect. If the "full-length" product was generated using primers with the 3′-end before the triazole, its "+A" derivative should also be presented. But this "+A" derivative was not observed in Figure 7.25, so for the extensions with the starting points before the triazole, the generation of "full-length" product, without deletion, is unlikely.

Surprisingly, the primer extension from the point, one nucleotide after the triazole (P3), gave one "A" deletion as well (Line 5, Figure 7.24, MS: Figure 7.26). As the potential product sequence after the triazole was "AACC" and the first "A" was already presented in the primer "P3", the deletion was only possible at the second "A" after the triazole or at the end of the reverse transcription product. However, because the extension of the primer (P4) with the 3′-end, 4 nucleotides after the triazole, gave "full-length" and "plus A" product (Figure 7.27), the deletion was unlikely at the end. Therefore, the deletion using "+1" primer was likely at the bordered "U" in the template at the "+2" site (Figure 7.24). As this nucleotide is the second nucleotide on the 5′ side of the triazole, it does not come from 2′-azido-nucleoside triphosphates but is intrinsic to the natural RNA. This makes this backbone unfavourable for the sequencing protocol, because it causes loss of one nucleotide in the sequence information.

Triazole Backbone C, containing the sequence "CzT" at the triazole site, was subsequently evaluated. Due to the "C:G" base pair being stronger than the "T:A" base pair, this backbone was expected to cause less deletion at the triazole site. The PAGE and MS results confirmed this expectation. The longest products (upper bands, Figure 7.28) were identified by MS as

the “full-length” product and its “plus A” derivative, which were obtained regardless of primer length. Therefore, this triazole linkage did not cause deletion when the reverse transcriptase read through. However, it still significantly hindered the reverse transcription as the truncated products were observed using primers P1-P3. They were identified by MS to be correlated with stopping after adding the first nucleotide “G” after triazole (“+1”) and stopping after adding the 7th and 8th nucleotides (“+7” and “+8”) after triazole. MS analysis of both crude reaction products (Figure 7.29) and gel-recovered “+7” and “+8” bands (Appendix, Figure 9.69) were analysed and the results were confirmed.

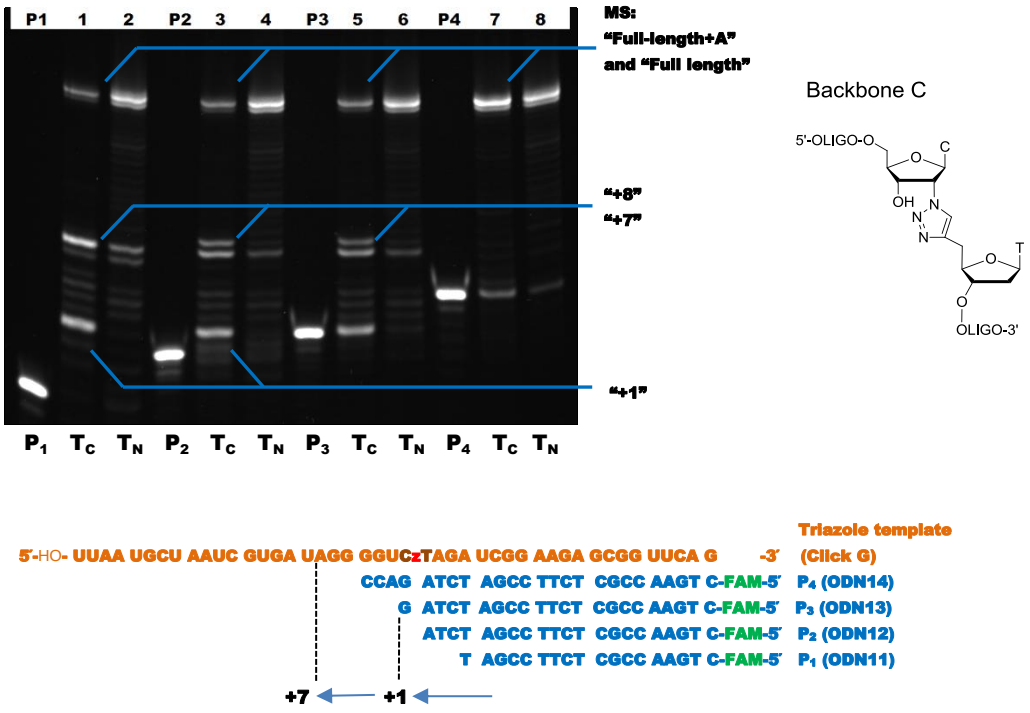


Figure 7.28: M-MuLV Reverse Transcriptase (wild type) test in Mg^{2+}-buffer for read-through of triazole backbone C (CzT) with 4 primers.

RT template: Click G (ORN18 click ORN19), 60 pmol for Lane 1, 3, 5, 7, T_C.
Control: canonical RNA template ORN13 (CU), 60 pmol for Lane 2, 4, 6, 8, T_N.

Lane P1-P4: primers (with 5′-FAM) in water
Lane P1, 1, 2: 60 pmol primer ODN11; Lane P2, 3, 4: 60 pmol primer ODN12;
Lane P3, 5, 6: 60 pmol primer ODN13; Lane P4, 7, 8: 60 pmol primer ODN14.
T_C: reverse transcription of Click template, T_N: reverse transcription of nature RNA template.

Lane 1-8: same reverse transcription condition as that in Figure 7.12.

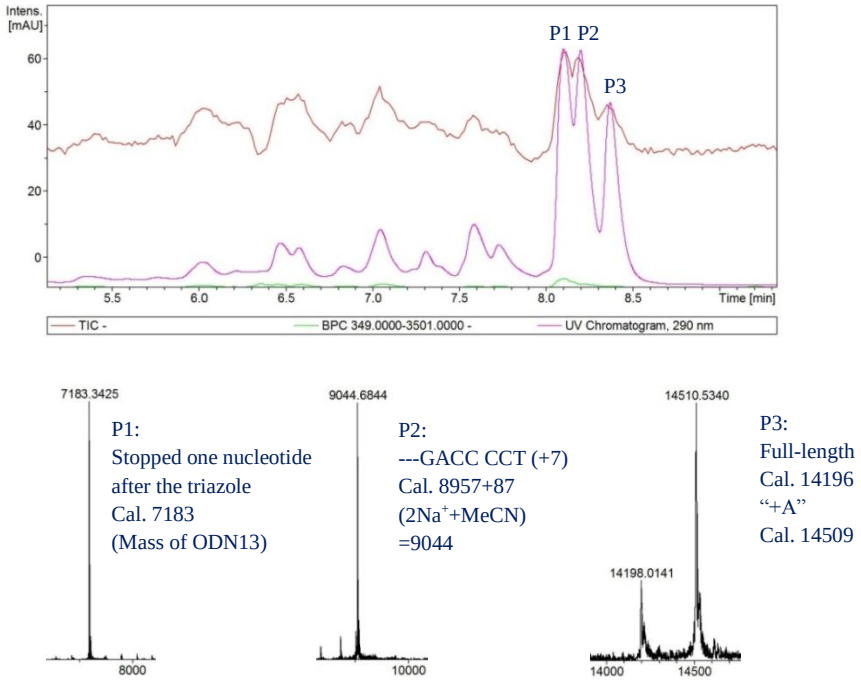


Figure 7.29: M-MuLV reverse transcriptase (wild type) RT reaction mixture (3 mM Mg^{2+}, primer ODN11) of the triazole template Click G (Backbone C, CzT) after incubation at 37 ℃ for 2 h then desalting by gel-filtration (NAP-10) (see also Figure 9.61).

The time-course (Figure 7.30) of the reverse transcription of Backbone C (CzT) clearly showed that the enzyme stopped at the "+1" site when using the shorter primer (P1, "-7") and stopped at the "+7" and "+8" sites when using the longer primer with the 3′-end one nucleotide after the triazole (P2, "+1").

Taking in to account previous experience of working with Backbone B and B′, Backbone C with the sequence "CzC" was potentially a good template for reverse transcription. However, the 5′-alkyne C phosphoramidite is not currently available. The synthesis of this monomer is time-consuming and was not undertaken in this project.

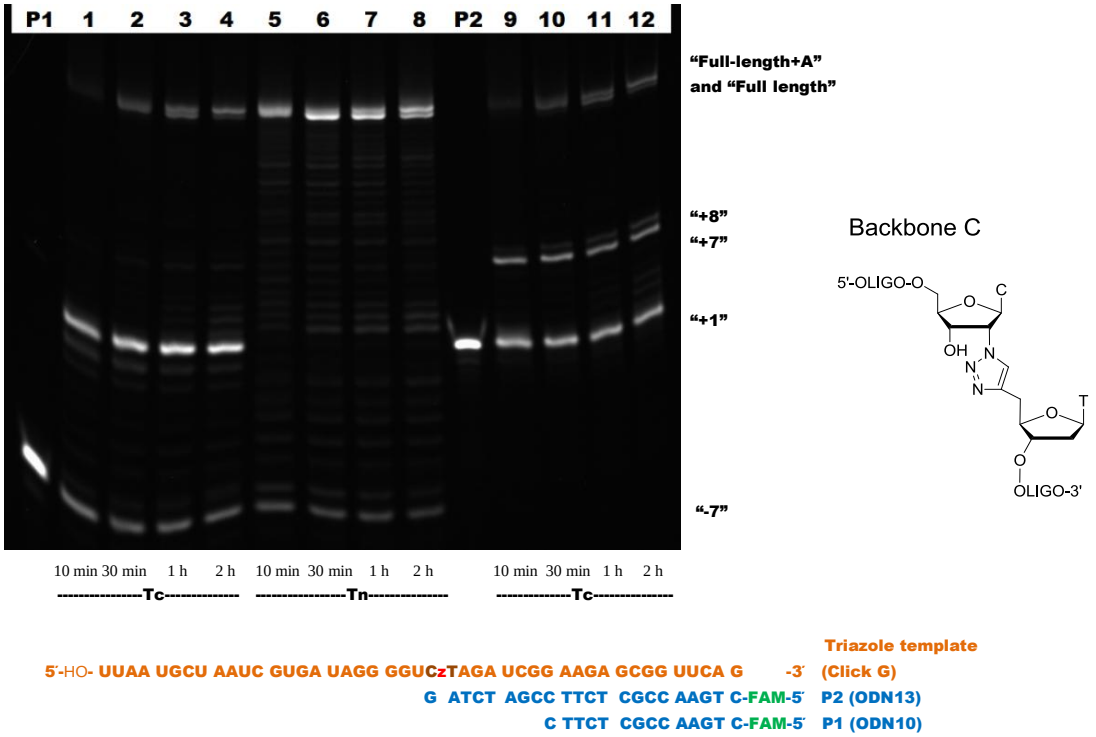


Figure 7.30: Time-course of the reverse transcription of triazole Backbone C (CzT) by M-MuLV Reverse Transcriptase (wild type).

RT template (Tc): Click G (ORN18 click ORN19, CzT), 60 pmol for lane 1-4, 9-12.
Control (Tn): canonical RNA template ORN13 (CU), 60 pmol for lane 5- 8.

Lane P1,P2: primers (with 5´-FAM) in water
Lane P1, 1-8: 60 pmol primer ODN10;
Lane P2, 9-12: 60 pmol primer ODN13.

Lane 1-8: same reverse transcription condition as that in Figure 7.12.
Reverse transcription: 37 ℃, 10 min / 30 min / 1 h / 2 h.

To promote the extension to reach the end of the template, one potential solution is extending the reverse transcription overnight. However, as described above, the wild type M-MuLV Reverse Transcriptase degraded the triazole template in 2 h. When the template was degraded, increasing the incubation time further resulted in no improvement (Figure 7.31, Lanes 1, 4).

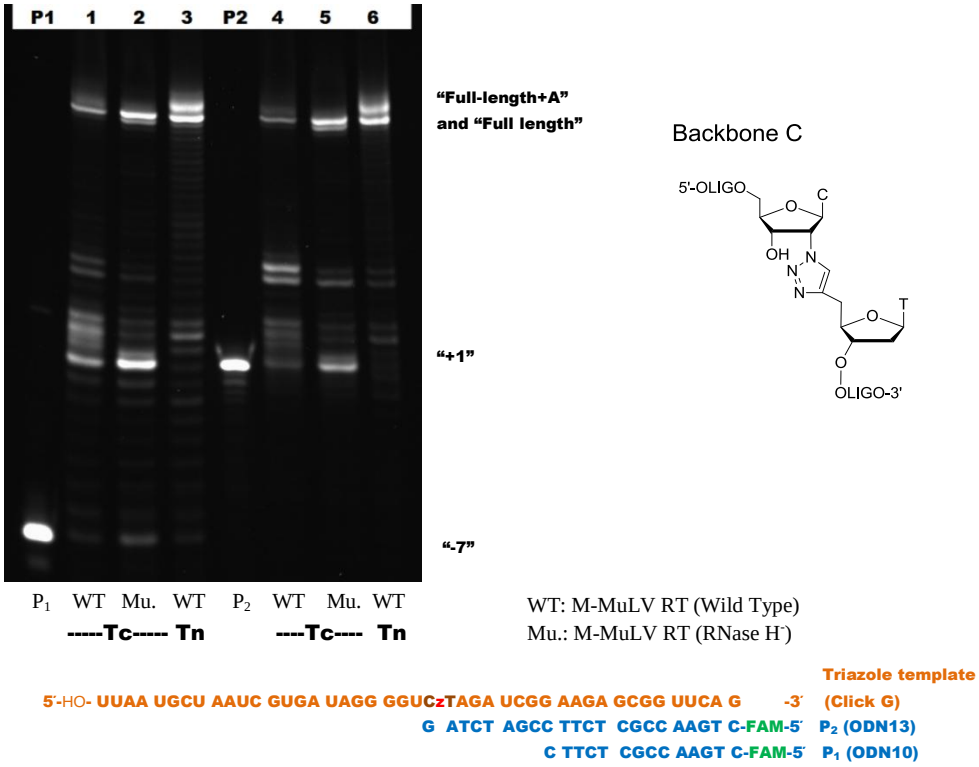


Figure 7.31: Overnight reverse transcription of triazole Backbone C (CzT) by M-MuLV Reverse Transcriptase (wild type) and M-MuLV Reverse Transcriptase (RNase H$^-$)

RT template (Tc): Click G (ORN18 click ORN19, CzT), 60 pmol for Lane 1, 2, 4, 5.
Control (Tn): canonical RNA template ORN13 (CU), 60 pmol for Lane 3, 6.

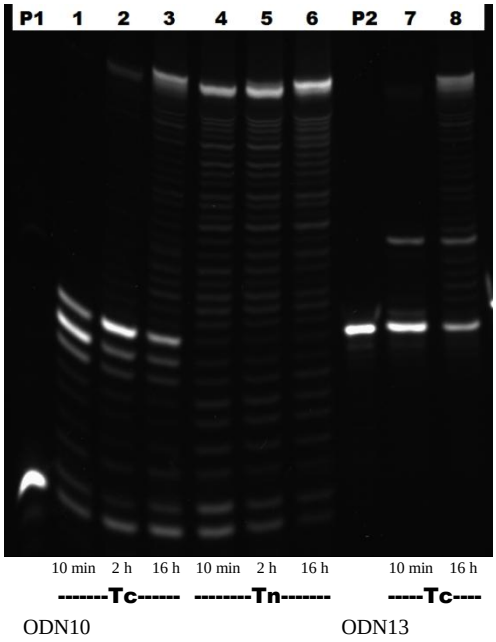
Lane P1, P2: primers (with 5´-FAM) in water
Lane P1, 1-3: 60 pmol primer ODN10;
Lane P2, 4-6: 60 pmol primer ODN13.

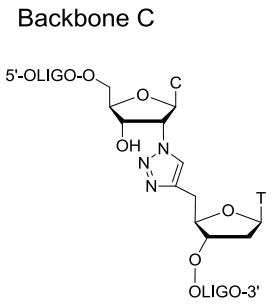
Lane 1, 3, 4, 6: M-MuLV Reverse Transcriptase (Wild Type), same reaction composition as that in Figure 7.12;
Lane 2, 5: M-MuLV Reverse Transcriptase (RNase H$^-$), other concentrations are the same as above.
Reverse transcription: 37 ℃ overnight.

Lane 1, 2, 4, 5: further reaction with M-MuLV Reverse Transcriptase (wild type) at 37 ℃ for 2 h to digest RNA template (adding 0.2 μmol DTT, 200 u (1 μL) M-MuLV Reverse Transcriptase (wild type, NEB®) in 5 μL 1x supplied M-MuLV Reverse Transcriptase (RNase H$^-$) Reaction Buffer).

The mutant M-MuLV Reverse Transcriptase (RNase H$^-$) mentioned above is more suitable for overnight reaction as it has greatly reduced RNase H activity [29]. The MS analysis of the crude reaction mixture, after 2 h incubation with this enzyme, clearly showed the presence of the triazole template (e.g. Figure 7.34). However, to get a clean result using M-MuLV Reverse Transcriptase (RNase H$^-$), additional procedures were developed to solve the band-smearing problem mentioned above (see Experimental chapter for procedures). For PAGE, excessive unlabelled reverse transcription product mimic was added to the reaction mixture

and annealed to the RNA template before gel electrophoresis. For MS, the reaction mixture was treated with RNase H to digest the RNA template after reverse transcription (The RNase H treatment was not performed for all Backbones' RT MS samples mentioned in this section, but was used for all backbones' RT MS samples in the next section). Because of the additional treatment, cleaner PAGE (Figure 7.32) and MS results were obtained for the experiments described below.





Triazole template
5'-HO- UUAA UGCU AAUC GUGA UAGG GGUCzTAGA UCGG AAGA GCGG UUCA G -3' (Click G)
G ATCT AGCC TTCT CGCC AAGT C-FAM-5' P2 (ODN13)
C TTCT CGCC AAGT C-FAM-5' P1 (ODN10)

Figure 7.32: M-MuLV Reverse Transcriptase (RNase H$^-$) reading triazole backbone C (Click G, CzT) time-course

RT template: Click G (ORN18 click ORN19), 60 pmol for Lane 1, 2, 3, 7, 8.
Control: canonical RNA template ORN13 (CU), 60 pmol for Lane 4, 5, 6.

Lane P1, P2: primers (with 5´-FAM) in water
Lane P1, 1, 2: 60 pmol primer ODN10; Lane P2, 3, 4: 60 pmol primer ODN13.
Tc: reverse transcription of Click template, Tn: reverse transcription of nature RNA template.

Lane 1-8: 3 μM primer (60 pmol), 3 μM template (60 pmol), 1x supplied M-MuLV Reverse Transcriptase (RNase H$^-$) Reaction Buffer (NEB®, 50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$, pH 8.3 at r.t.), 10 mM DTT, 0.5 mM dNTP (each triphosphate), 200 u (1 μL) M-MuLV Reverse Transcriptase (RNase H$^-$)(NEB®) in total 20 μL.

Reverse transcription: 37 ℃, 10 min/ 2 h/overnight.

The PAGE of the M-MuLV Reverse Transcriptase (RNase H$^-$) overnight reaction time-course showed an increased percentage of products at the full-length range when leaving the reaction

overnight (Figure 7.32). However, about half of the primer still stopped at the "+1" site. The situation was not improved using the primer "P2" that has the starting point at the "+1" site. It suggests the triazole linkage made the nucleotides at the "+1" site and "+2" site difficult to be reverse transcribed.

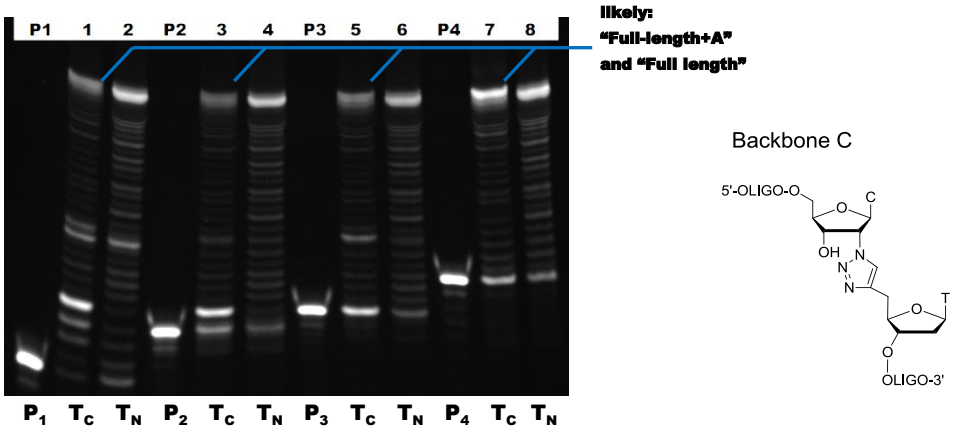


Figure 7.33: M-MuLV Reverse Transcriptase (RNase H$^-$) overnight test in Mg^{2+}-buffer for read-through of triazole backbone C (CzT) with 4 primers.

RT template: Click G (ORN18 click ORN19), 60 pmol for Lane 1, 3, 5, 7.
Control: canonical RNA template ORN13 (CU), 60 pmol for Lane 2, 4, 6, 8.
Lane P1-P4: primers (with 5´-FAM) in water
Lane P1, 1, 2: 60 pmol primer ODN11; Lane P2, 3, 4: 60 pmol primer ODN12;
Lane P3, 5, 6: 60 pmol primer ODN13; Lane P4, 7, 8: 60 pmol primer ODN14.
T_C: reverse transcription of Click template, T_N: reverse transcription of nature RNA template.

Lane 1-8: same reverse transcription condition as that in Figure 7.32.
Reverse transcription: 37 ℃, overnight for Click template, 37 ℃ 2 h for canonical RNA template.

The overnight M-MuLV Reverse Transcriptase (RNase H$^-$) test was repeated using the same 4 primers as the test for the wild type enzyme. Comparing to the similar result obtained using the wild type reverse transcriptase mentioned above (Figure 7.28), the mutant enzyme produced a lower percentage of "+7/+8" bands and higher percentage of full-length product (Figure 7.33). However, about half of primers still stopped at the "+1" site. MS analysis of the overnight reaction samples was not performed. However, MS analysis of the 2 h incubation reactions showed the presence of the "full-length" product for all samples (Figure 7.34-7.37),

these samples for MS were not treated by RNase H to remove the RNA template, so only part of the RT product, not affected by the template, was shown on the HPLC trace as a small peak. However, the results showed little deletion product for all 2 h reactions. This indicates the overnight reactions are unlikely to generate a large percentage of deletion products. According to PAGE (Figure 7.33) the overnight reaction likely produced more "full-length +A" derivative.

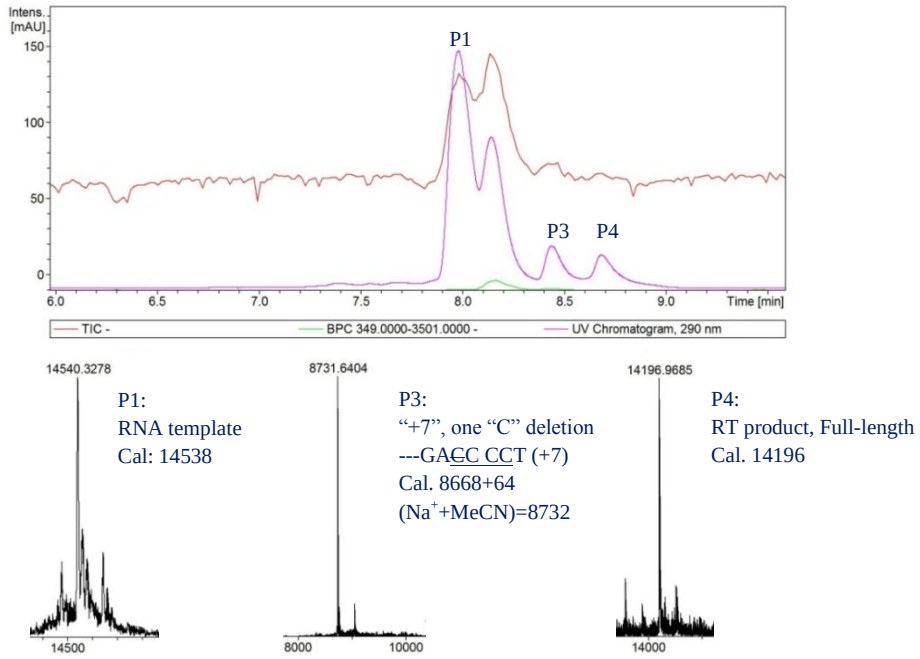


Figure 7.34: M-MuLV reverse transcriptase (RNase H^-) RT reaction mixture (3 mM Mg^{2+}, primer ODN11) of the triazole template Click G (Backbone C, CzT) after incubation at 37 ℃ for 2 h then desalting by gel-filtration (NAP-10). RNase H digestion was not performed. For P3 it is uncertain which of the underlined "C" was deleted (see also Figure 9.65).

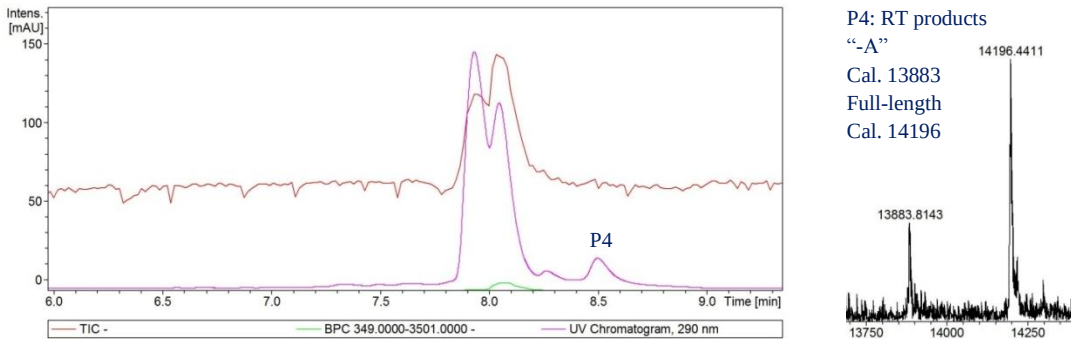


Figure 7.35: M-MuLV reverse transcriptase (RNase H^-) RT reaction mixture (3 mM Mg^{2+}, primer ODN12-"+0") of the triazole template Click G (Backbone C, CzT) after incubation at 37 ℃ for 2 h then desalting by gel-filtration (NAP-10) (see also Figure 9.66). RNase H digestion was not performed.

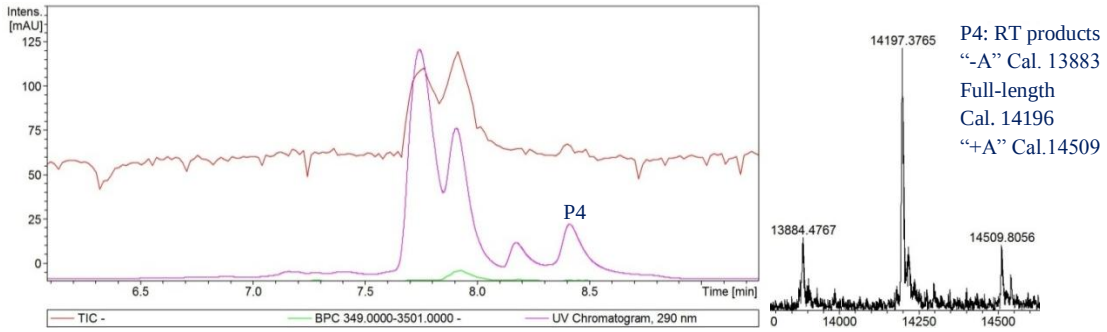


Figure 7.36: M-MuLV reverse transcriptase (RNase H$^-$) RT reaction mixture (3 mM Mg^{2+}, primer ODN13-"+1") of the triazole template Click G (Backbone C, CzT) after incubation at 37 ℃ for 2 h then desalting by gel-filtration (NAP-10) (see also Figure 9.67). RNase H digestion was not performed.

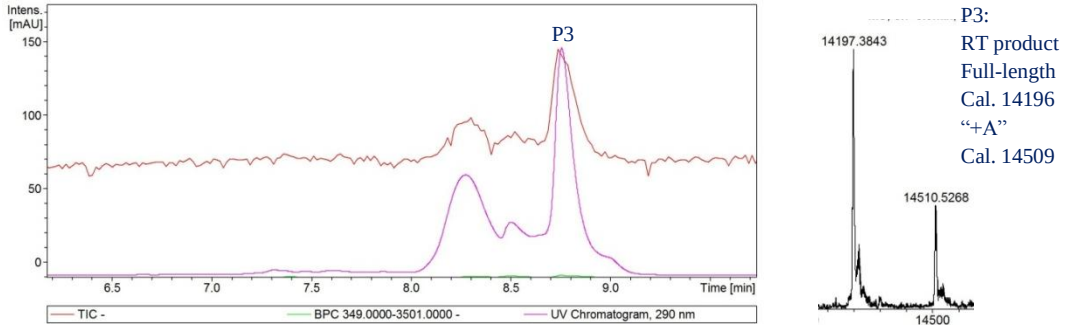


Figure 7.37: M-MuLV reverse transcriptase (RNase H$^-$) RT reaction mixture (3 mM Mg^{2+}, primer ODN14-"+4") of the triazole template Click G (Backbone C, CzT) after incubation at 37 ℃ for 2 h then desalting by gel-filtration (NAP-10) (see also Figure 9.68). RNase H digestion was not performed.

Due to the reverse transcription results described above, the order of 5′-adapter ligation and reverse transcription in the sequencing protocol using copper-catalysed Click ligation needs to be considered. As the reverse transcription gives a mixture of "+A" 3′-terminal addition product and non-addition product, it is more appropriate to ligate the 5′-adapter before reverse transcription to avoid sequence confusion at the 5′ end of the small RNA.

7.8 Reverse Transcription of Copper-free Click ligated template

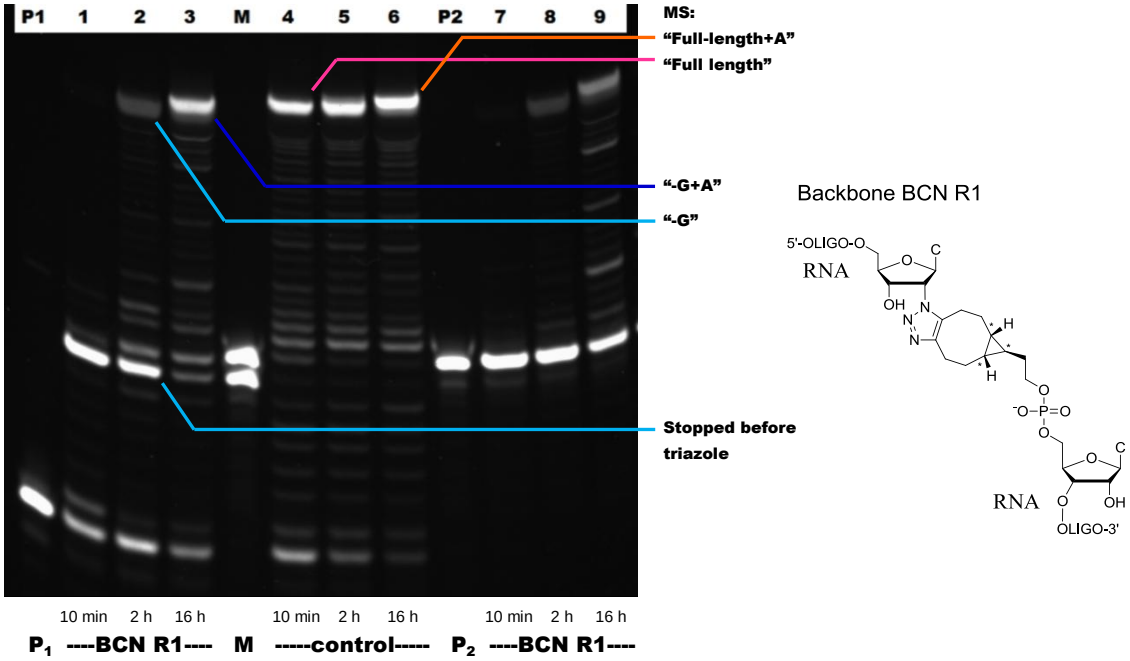
7.8.1 Reverse transcription of backbone BCN R1 (C-BCN-C, RNA adapter)

The enzymatically incorporated 2′-azide group can also be "clicked" to a 5′-BCN labelled adapter. This copper-free Click ligation could significantly simplify the RNA sequencing procedure. As this SPAAC Click reaction does not need copper catalyst, it helps avoid the copper-induced degradation of the RNA sample (discussed in section 7.11). 2′-Azido-2′-dCTP and 2′-azido-2′-dUTP can be incorporated by Yeast PAP, therefore, the corresponding Click BCN backbones with the sequence "C-BCN-C" and "U-BCN-C" needed to be tested for reverse transcription. Cytidine was chosen as the nucleotide at the 3′ side of the BCN to increase the duplex stability. As the BCN linker generated in the SPAAC reaction is much bulkier than the triazole linkage generated by CuAAC, it was expected that it would hinder the reverse transcriptase. Consequently, Click BCN backbones were reverse-transcribed overnight. Due to the extended reaction time, M-MuLV Reverse Transcriptase (RNase H^-), instead of the wild type reverse transcriptase, was used to avoid the degradation of the RNA template by the active RNase H domain during overnight incubation (discussed above, Section 7.7.1 and 7.7.4).

The Backbone BCN R1 was constructed by "clicking" the 2′-azide modified RNA with a 5′-BCN modified RNA adapter. The 2′-azide modified RNA was synthesised from a 2′-azido-2′-deoxynucleoside functionalised resin as mentioned above. The 5′-BCN was added as a phosphoramidite during standard RNA synthesis (see oligonucleotide synthesis section 8.2 in Experimental).

Both PAGE and MS analysis showed that the M-MuLV Reverse Transcriptase (RNase H^-) paused at the triazole site (C-BCN-C) but then read-through this site successfully during overnight incubation (Figure 7.38). The tests were repeated and the result was confirmed. HPLC-MS analysis of the 2 h incubation sample showed two HPLC peaks (about 1:1) representing "stopped before triazole" and "Full-length minus G" ("-G") products (Figure 7.39). Their "+A" derivatives were also observed, accounting for 10 % of their original products comparing to the intensities of each of their main MS peaks. After overnight incubation, the "Stopped before triazole" product disappeared with its "+A" derivative remaining as a minor peak. Most of the primer extended to the end of the template but produced the "Full-length -G+A" product only (Figure 7.40). This "+A" was likely added at the 3′-end of the DNA product during overnight reaction because it was observed only as a minor MS peak in the 2 h incubation sample. The extension of "+4" primer overnight

produced the "Full-length +A" product (Figure 7.41). This further indicated that the "+A" at the end of the strand always occurred using the triazole template, which is the same as the reaction using the control RNA template (Figure 9.77, 9.78, Appendix). So having a "+A" in the middle but no "+A" at the end of the RT product was unlikely.



Triazole template

5'-HO- UUAA UGCU AAUC GUGA UAGG GGUCzCAGA UCGG AAGA GCGG UUCA G -3' (Click BCN R1)

G GTCT AGCC TTCT CGCC AAGT C-FAM-5' P2/M (ODN16)

GTCT AGCC TTCT CGCC AAGT C-FAM-5' M (ODN15)

C TTCT CGCC AAGT C-FAM-5' P1 (ODN10)

Figure 7.38: M-MuLV Reverse Transcriptase (RNase H$^-$) overnight test in Mg^{2+} buffer for read-through of triazole backbone BCN R1 (RNA adapter) with two primers.

RT template: Click BCN R1 (ORN18 click ORN20, RNA adapter), 60 pmol for Lane 1-3, 7-9.
Control: canonical RNA template ORN15 (CC), 60 pmol for Lane 4-6.

Lane P1, P2: primers (with 5´-FAM) in water
Lane P1, 1-3, 4-6: 60 pmol primer ODN10; Lane P2, 7-9: 60 pmol primer ODN16.

Lane M: ODN15 ("+0") and ODN16 ("+1") in water as markers.

Lane 1-9: 3 µM primer (60 pmol), 3 µM template (60 pmol), 1x supplied M-MuLV Reverse Transcriptase (RNase H$^-$) Reaction Buffer (NEB®, 50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$, pH 8.3 at r.t.), 10 mM DTT, 0.5 mM dNTP (each triphosphate), 200 u (1 µL) M-MuLV Reverse Transcriptase (RNase H$^-$)(NEB®) in total 20 µL.
Reverse transcription: 37 ℃, 16 h (overnight).

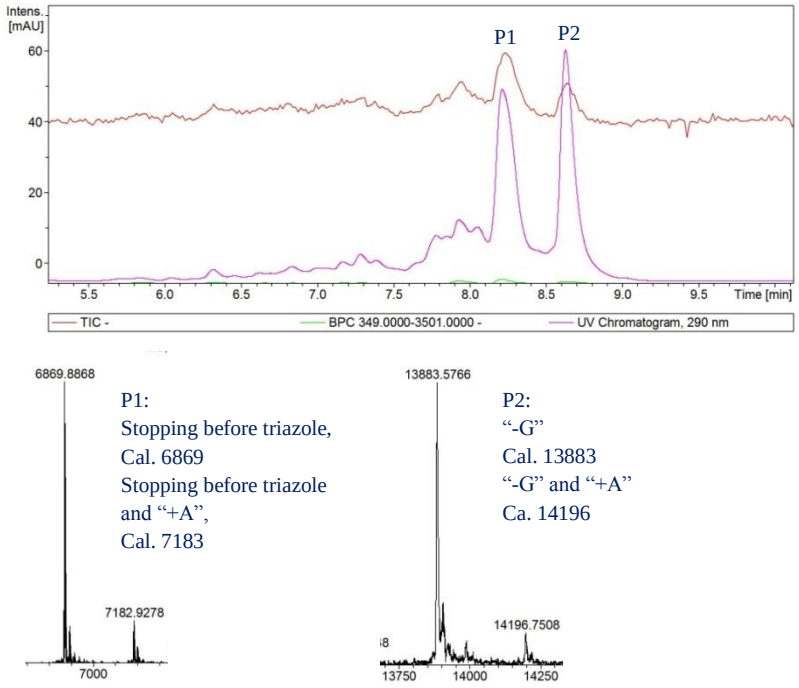


Figure 7.39: M-MuLV reverse transcriptase (RNase H$^-$) RT reaction mixture (3 mM Mg^{2+}, primer ODN10) of the triazole template Click BCN R1 (C-BCN-C, RNA adapter-part) after incubation at 37 ℃ for 2 h, digestion by RNase H and desalting by gel-filtration (NAP-10) (see also Figure 9.75).

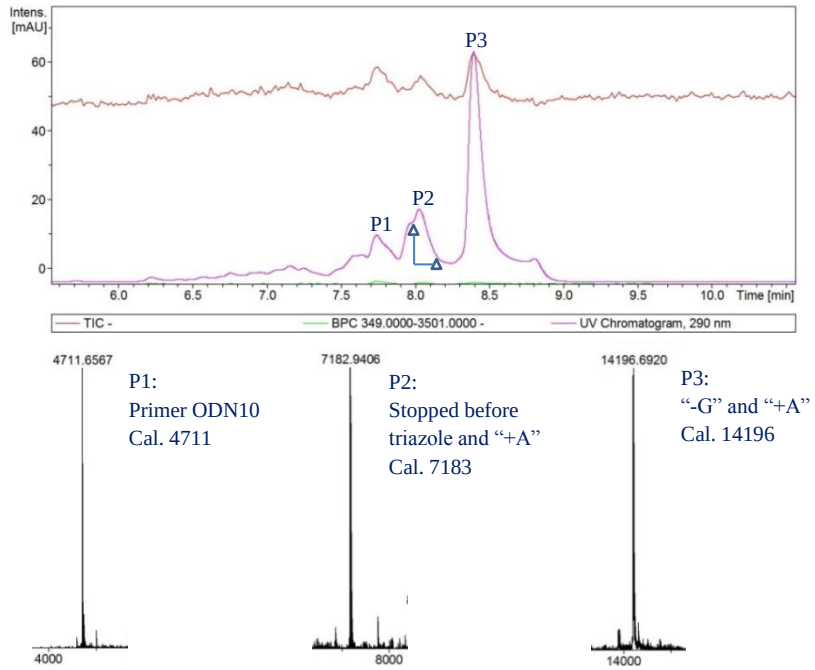


Figure 7.40: M-MuLV reverse transcriptase (RNase H$^-$) RT reaction mixture (3 mM Mg^{2+}, primer ODN10) of the triazole template Click BCN R1 (C-BCN-C, RNA adapter-part) after incubation at 37 ℃ overnight, digestion by RNase H and desalting by gel-filtration (NAP-10) (see also Figure 9.76).

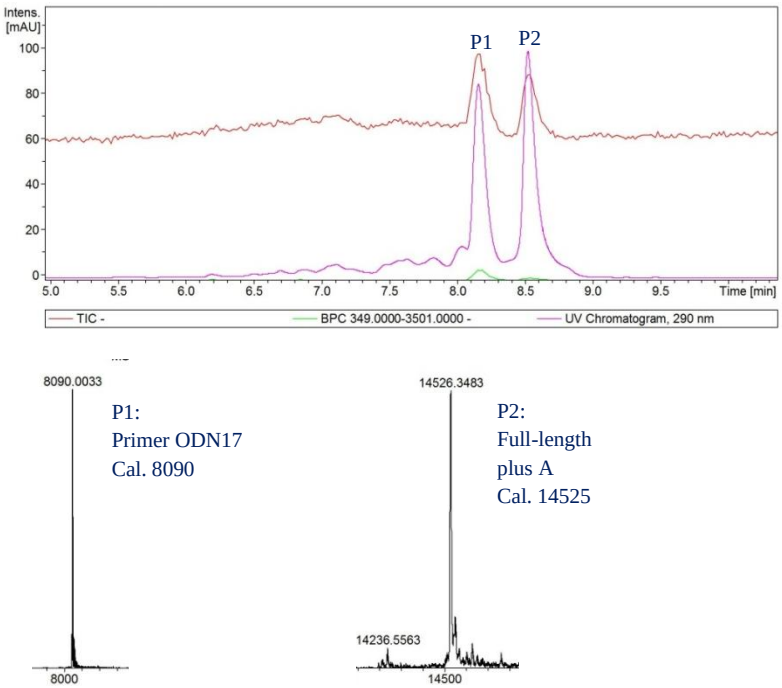


Figure 7.41: M-MuLV reverse transcriptase (RNase H$^-$) RT reaction mixture (3 mM Mg^{2+}, primer ODN17-"+4") of the triazole template Click BCN R1 (C-BCN-C, RNA adapter-part) after incubation at 37 ℃ overnight, digestion by RNase H and desalting by gel-filtration (NAP-10) (see also Figure 9.79).

Sequencing of the RT products may identify which "G" was deleted during reverse transcription. Because the RT product could be a mixture with different sequences, it was further amplified by PCR and cloned (Figure 7.42). Each individual subclone would have a single copy of the insert so the sequencing result can be unambiguous. PCR also transferred the RT product to longer dsDNA to facilitate cloning. Nineteen clones were sequenced and the "-G" deletion at the triazole-BCN site was confirmed for most of the clones. Of the 19 clones that were sequenced, 16 clones have the same sequence representing one "G" deletion as that in Figure 7.43. However, three clones (Figure 7.44, underlined in green) possessed additional sequence alterations near the triazole-BCN site.

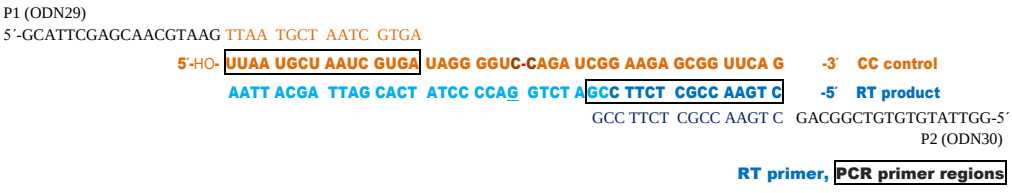


Figure 7.42: PCR primer (P1, P2) alignment for the amplification of reverse transcription product of triazole-BCN template and control RNA template by GoTaq® DNA polymerase.

The consistent one “G” deletion should not affect the accuracy of small RNA sequencing. This is because the corresponding “C” on the triazole-BCN template comes from triphosphate incorporation, so this sequence is not part of the small RNA. However, the possibility of additional sequence alterations such as the “U” to “C” transition (Figure 7.44, clone No. 4 and 11) is not ideal for our small RNA sequencing application. However, the low frequency (3/19) of these sequence alterations means they can be corrected through multiple parallel sequencing. The 3 nucleotide deletion of clone No. 16 was located at the primer region, so this deletion is not caused by the triazole-BCN linkage.

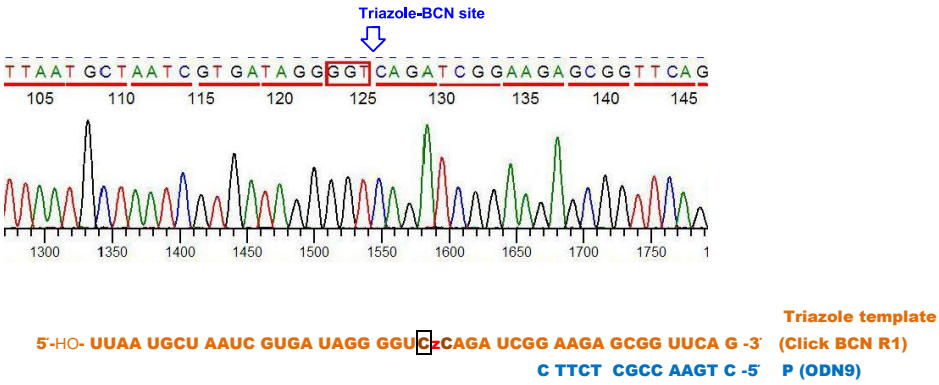


Figure 7.43: Typical sequencing trace of the reverse transcription product of Click BCN R1 (C-BCN-C). The deleted “G” corresponds to the circularized “C” in the triazole template. (The whole datasheet is shown in Figure 9.96.)

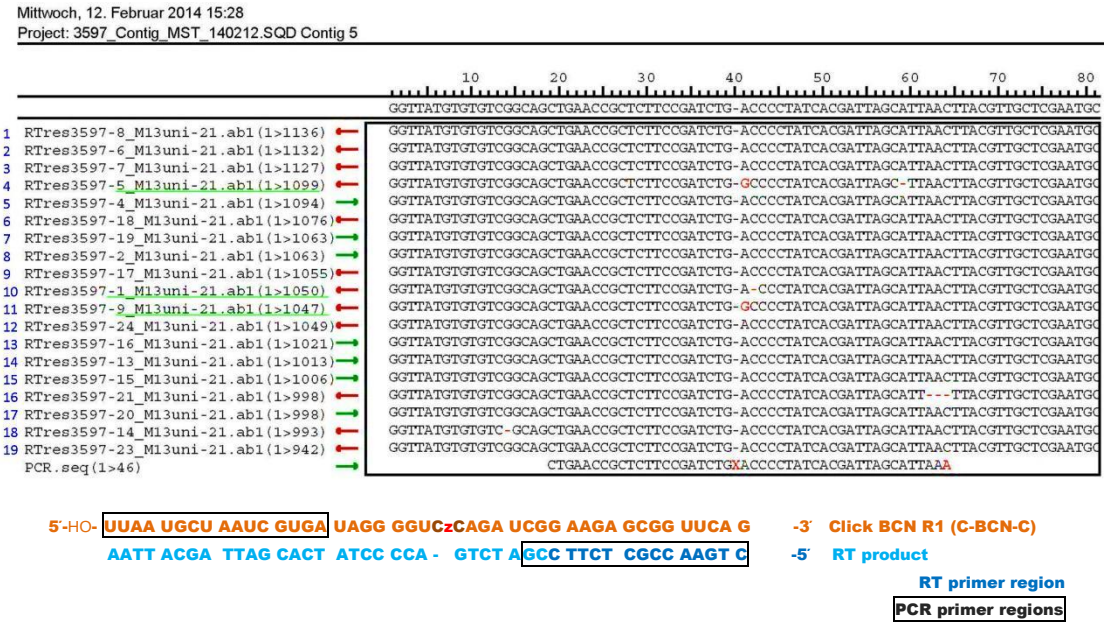


Figure 7.44: Sequences of the 19 clones of the PCR amplified RT products of triazole-BCN template Click BCN R1 (C-BCN-C), aligned to the expected sequence at the bottom.

The sequencing result of the control sample (RT of the canonical RNA template) also presented low-frequency sequence alterations (Figure 7.45). The result of the control sample suggested that some of the additional sequence alterations in Figure 7.44 may not be caused by reverse transcription of the triazole-BCN linkage but was introduced in the subsequent PCR and cloning steps.

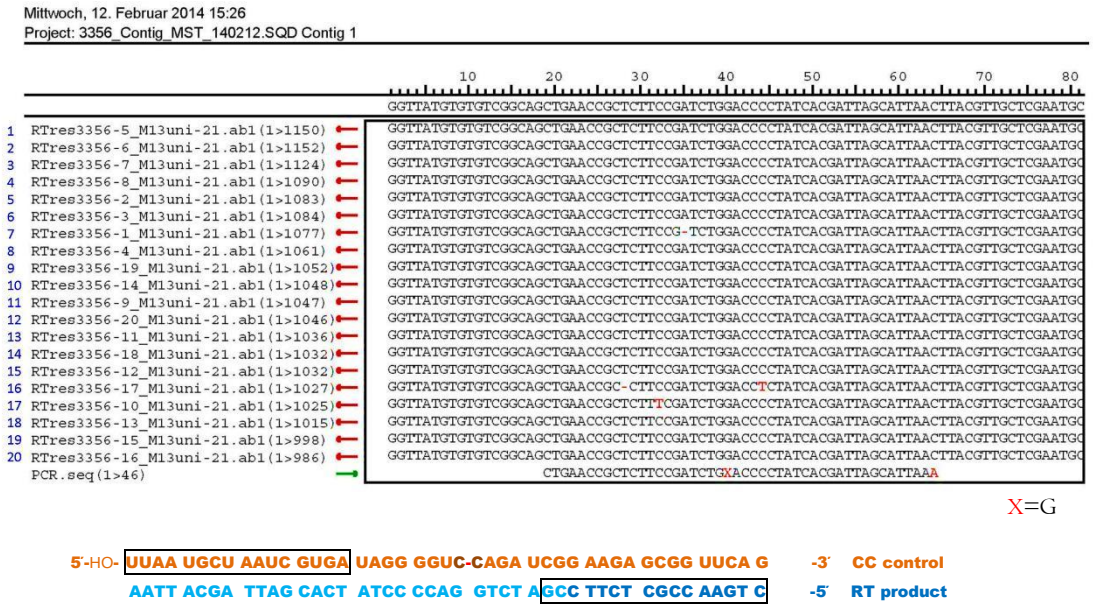


Figure7.45: Sequences of the 20 clones of the PCR amplified RT product of control RNA template (CC), aligned to the expected sequence at the bottom.

In addition, the extension of the "+1" primer was tested under the same RT conditions. The result (Lanes 7-9, Figure 7.38) showed that the extension from the "+1" position was less efficient than the one starting far before the triazole site. It indicates that it is easier for the reverse transcriptase to pass this triazole-BCN linkage during polymerization than to start near the triazole-BCN linkage. The MS of this "+1" primer extension has not been performed.

It was also observed that by changing the divalent cation in the buffer from Mg^{2+} to Mn^{2+}, the reverse transcription was largely completed in 2 h instead of overnight (Figure 7.46). MS analysis confirmed that under these conditions, the same "-G+A" product was produced (Figure 9.80, 9.81, Appendix).

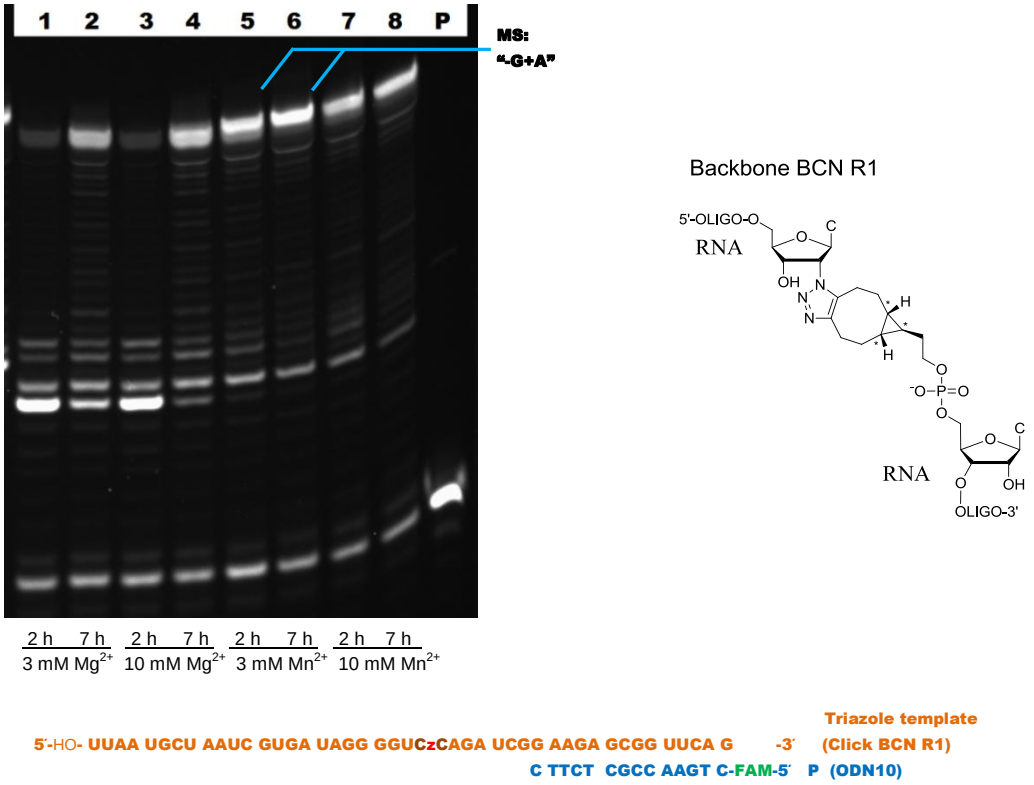


Figure 7.46: Optimization of M-MuLV Reverse Transcriptase (RNase H$^-$) reading of triazole backbone Click BCN R1 (RNA adapter) in different buffers.

RT template: Click BCN R1 (ORN18 click ORN20, RNA adapter), 60 pmol for Lane 1-8;
Lane P1, 1-8: 60 pmol primer ODN10; Lane P: primers (with 5′-FAM) in water.

Lane 1,2: 1x supplied M-MuLV Reverse Transcriptase (RNase H$^-$) Reaction Buffer (50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$, pH 8.3 at r.t.);
Lane 3,4: 1x Mg^{2+}-free reaction buffer, 10 mM $MgCl_2$;
Lane 5,6: 1x Mg^{2+}-free reaction buffer, 3 mM $MnCl_2$;
Lane 7,8: 1x Mg^{2+}-free reaction buffer, 10 mM $MnCl_2$.

Lane 1-12: 3 µM primer (60 pmol), 3 µM template (60 pmol), 10 mM DTT, 0.5 mM dNTP, 200 U (1 µL) M-MuLV Reverse Transcriptase (RNase H$^-$) (NEB®) in total 20 µL.
Reverse transcription: 37 ℃, 2 h or 7 h.

As the triazole-BCN is a much longer linkage than the standard phosphodiester, it was suspected that the nucleotide at the 5′ side of the linkage can template the addition of the second or third nucleotide after the triazole using "+T" or "+T+1" primers that have one nucleotide (T) opposing the triazole-BCN linkage. Thymidine (T) was chosen randomly. However, the result showed that the "+T" primer cannot extend further and the "+T+1" primer gave no significant improvement over the short "-7"-primer with the 3′-end before the triazole (Figure 7.47).

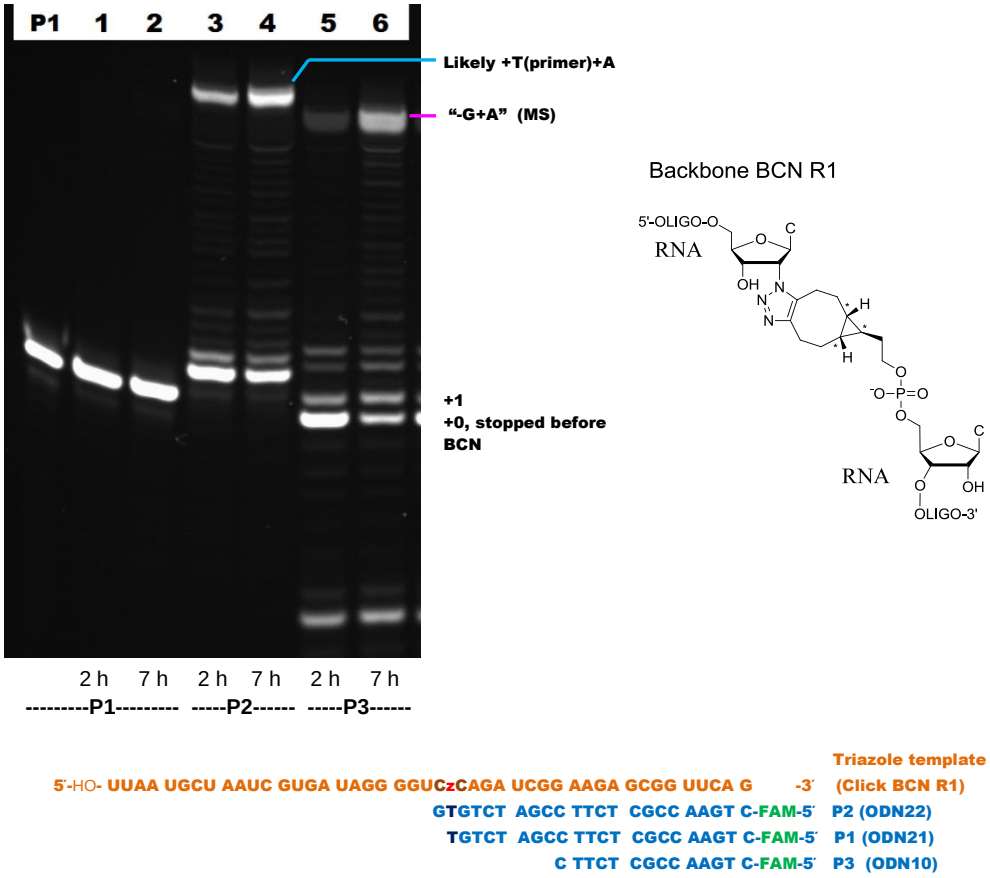


Figure 7.47: M-MuLV Reverse Transcriptase (RNase H$^-$) 2 h and 7 h test in Mg^{2+} buffer for reading of triazole backbone Click BCN R1 (RNA adapter) using different primers.

RT template: Click BCN R1 (ORN18 click ORN20, RNA adapter), 60 pmol for Lane 1-6.

Lane P1: primer P1 (with 5′-FAM) in water
Lane P1, 1-2: 60 pmol primer P1; Lane 3-4: 60 pmol primer P2; Lane 5-6: 60 pmol primer P3.

Lane 1-6: same reverse transcription condition as that in Figure 7.38.
Reverse transcription: 37 ℃, 2 h or 7 h.

7.8.2 Reverse transcription of triazole Backbone BCN1 (C-BCN-C, DNA adapter) and Backbone BCN2 (C-BCN-G-clamp, RNA adapter)

As mentioned in the next generation sequencing section above, it is uncertain whether it is necessary to use an RNA adapter instead of a DNA adapter for reverse transcription. No side-by-side comparative study has been published. The Backbone BCN1 with a DNA adapter, instead of an RNA adapter, was synthesised by us for reverse transcription. Comparing to Backbone BCN R1 with a RNA adapter, the reverse transcription of Backbone BCN1 was significantly less efficient (Figure 7.45). The reverse transcription of the DNA part before the triazole-BCN linkage is slower and a lower percentage of primer successfully transcribed past the triazole-BCN linkage.

To increase the duplex stability at the triazole-BCN site, an RNA adapter with a "G-clamp" nucleotide (Figure 7.48) [36] was also synthesised and the corresponding Backbone BCN2 was made and tested (Figure 7.49). Compared to the result of Backbone BCN R1, the reverse transcription of the Backbone BCN2 quickly produced a "+1" product, which was slow to extend further. The "+1" band of the overnight reaction sample (Figure 7.49, Lane 9) was identified by MS as a "plus A" derivative of "stopped before triazole" product ("0+A") (Figure 7.50). Although the identity of the "+1" band of the 2 h incubation sample (Figure 7.49, Lane 8) was not confirmed by MS, it is likely to be the "0+A" product as well, with no "0+G" product, because the overnight reaction produced only "0+A" product and "Full-length –G+A" product. It was assumed that the G-clamp promotes the terminal addition of A to the 'product stopped before the triazole'. However, the Backbone BCN3 with the sequence "U-BCN-C" also generated "0+A" quickly after 2 h (see next section), so this quick "+A" is not necessarily caused by G-clamp.

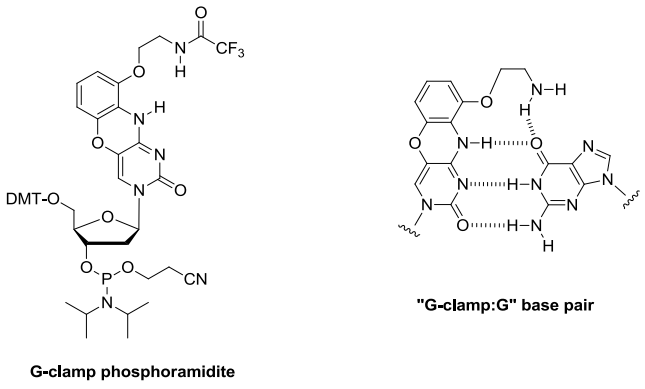


Figure 7.48: G-clamp Phosphoramidite from Glen Research® and the base pair of G-clamp and "G"

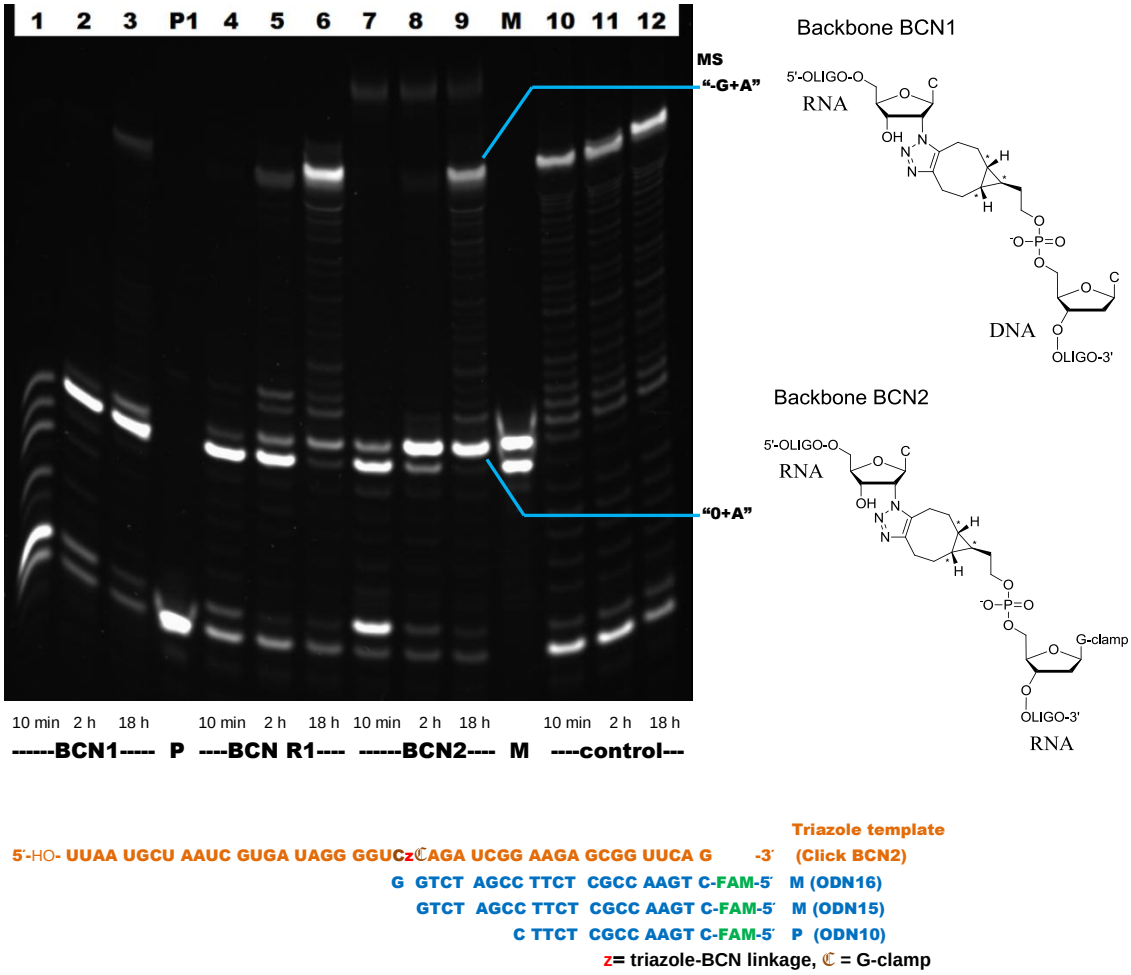


Figure 7.49: Reverse transcription comparison of triazole backbone BCN1 (DNA adapter), BCN R1 (RNA adapter), BCN2 ("CzG-clamp", RNA adapter) and canonical RNA template in Mg^{2+}-buffer.

RT template:
Click BCN1 (ORN18 click ODN28, DNA adapter), 60 pmol for Lane 1-3;
Click BCN R1 (ORN18 click ORN20, RNA adapter), 60 pmol for Lane 4-6;
Click BCN2 (ORN18 click ORN21, RNA adapter with G-clamp), 60 pmol for Lane 7-9;
Control: canonical RNA template ORN15 (CC), 60 pmol for Lane 10-12.

Lane P1: primers (with 5´-FAM) in water
Lane P1, 1-12: 60 pmol primer ODN10.
Lane M: ODN15 (+0) and ODN16 (+1) in water as markers.

Lane 1-12: same reverse transcription condition as that in Figure 7.38.
Reverse transcription: 37 ℃, 18 h (overnight).

After the terminal "A" addition to the primer, extended before the triazole-BCN linkage after

2 h, this “0+A” primer was difficult to extend further (Figure 7.49, Lane 8, 9). The backbone with the sequence U-BCN-(G-clamp) was suspected to have better reverse transcription efficiency because the added “A” matches the “U” next to the triazole-BCN linkage. This backbone has not been made and investigated. However, the test of Backbone BCN3 (see next section) with the sequence “U-BCN-C”, without G-clamp, showed that the “0+A” product was also quickly formed but was difficult to extend further. This indicates that the U-BCN-(G-clamp) sequence may generate the same difficulty. Alternatively, placing the G-clamp at the 5′-side of the triazole-BCN linkage instead of the 3′-side may provide more significant help for reverse transcription. However, placing G-clamp at the 5′ side of the triazole-BCN can only be achieved through incorporation of 2′-azido-G-clamp nucleoside triphosphate. This idea is challenging and has not yet been investigated.

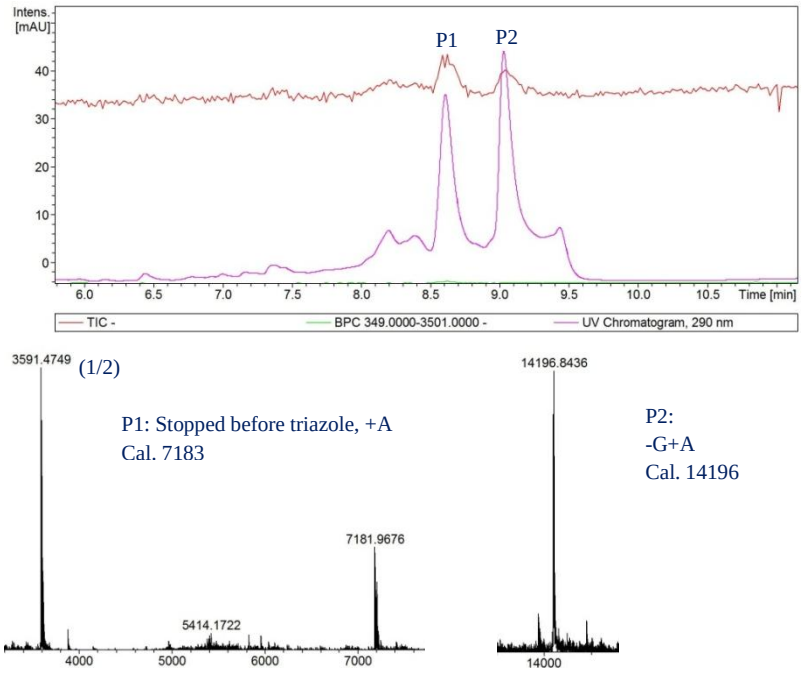


Figure 7.50 (Figure 9.82): M-MuLV reverse transcriptase (RNase H$^-$) RT reaction mixture (3 mM $MgCl_2$ in buffer, primer ODN10) of the triazole template Click BCN2 (C-BCN-ℂ, RNA adapter-part, ℂ=G-clamp) after incubation at 37 ℃ overnight, digestion by RNase H and desalting by gel-filtration (NAP-10).

It was found that by adding Mn^{2+} to the buffer, the reverse transcription of Backbone BCN1 was significantly improved (Figure 7.51). However, the reverse transcription of Backbone BCN R1 finished in 2 h in Mn^{2+}-buffer, which was still faster than this optimised result.

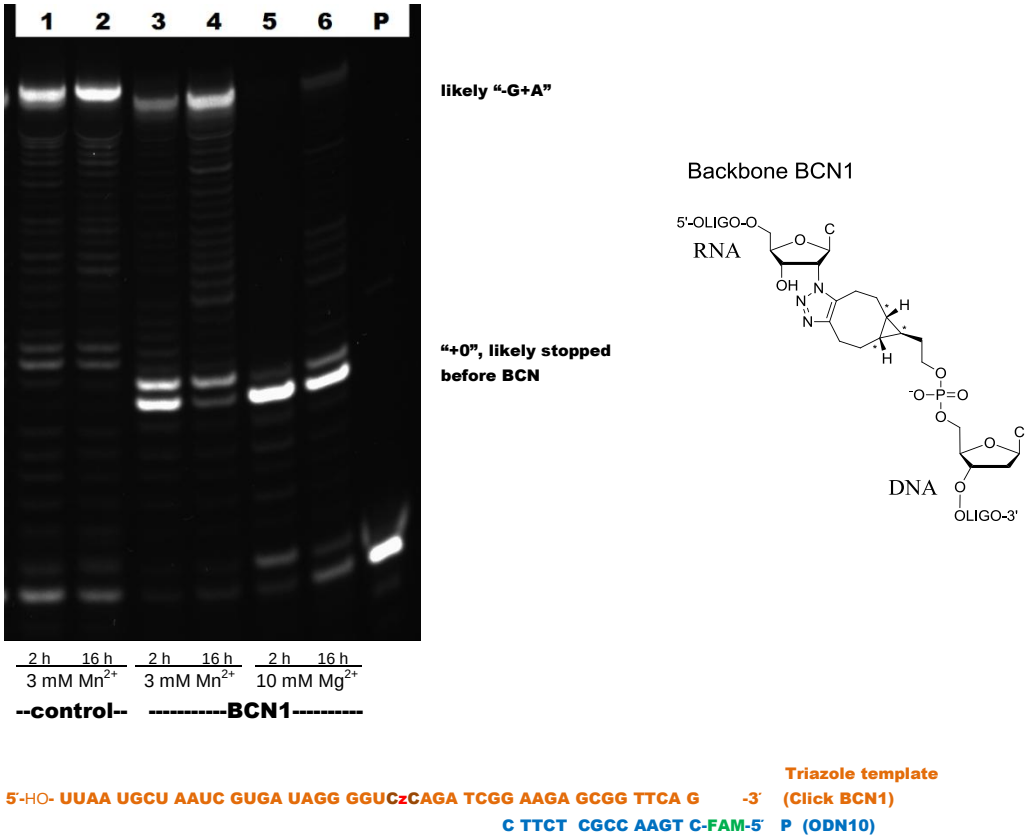


Figure 7.51: Optimization of M-MuLV Reverse Transcriptase (RNase H$^-$) Reading-through of triazole backbone BCN1 (DNA adapter) in different buffers.

RT template:
Control: canonical RNA template ORN15 (CC), 60 pmol for Lane 1, 2;
Click BCN1 (ORN18 click ODN28, CzC, DNA adapter), 60 pmol for Lane 3-6.

Lane P: primers (with 5´-FAM) in water
Lane P, 1-6: 60 pmol primer ODN10.

Lane 1-4: 3 µM primer (60 pmol), 3 µM template (60 pmol), 1x supplied M-MuLV Reverse Transcriptase (RNase H$^-$) Reaction Buffer (NEB®, 50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$, pH 8.3 at r.t.), 3 mM $MnCl_2$, 10 mM DTT, 0.5 mM dNTP, 200 u (1 µL) M-MuLV Reverse Transcriptase (NEB®) in total 20 µL.
Lane 5, 6: 1x M-MuLV Reverse Transcriptase (RNase H$^-$) Reaction Buffer (Mg^{2+}-free, 50 mM Tris-HCl, 75 mM KCl, pH 8.3 at r.t.), 10 mM $MgCl_2$. The rest is the same as above.
Reverse transcription: 37 ℃, 2 h/16 h.

7.8.3 Reverse transcription of backbone BCN3 (U-BCN-C, RNA adapter)

As illustrated above the 2´-azido-2´-dUTP was incorporated by Yeast PAP and multiple-addition was less noticeable compared to that of 2´-azido-2´-dCTP. This can generate Backbone BCN3 with the sequence "U-BCN-C" instead of "C-BCN-C" (Backbone BCN R1). However, the reverse transcription of this backbone by M-MuLV Reverse Transcriptase (RNase H^-) in standard Mg^{2+}-buffer had difficulty passing the triazole-BCN linkage (Figure 7.52, MS: Figure 7.53, 7.54). The problem was not solved by using "+1" primer as it was also not extended (Lane 10).

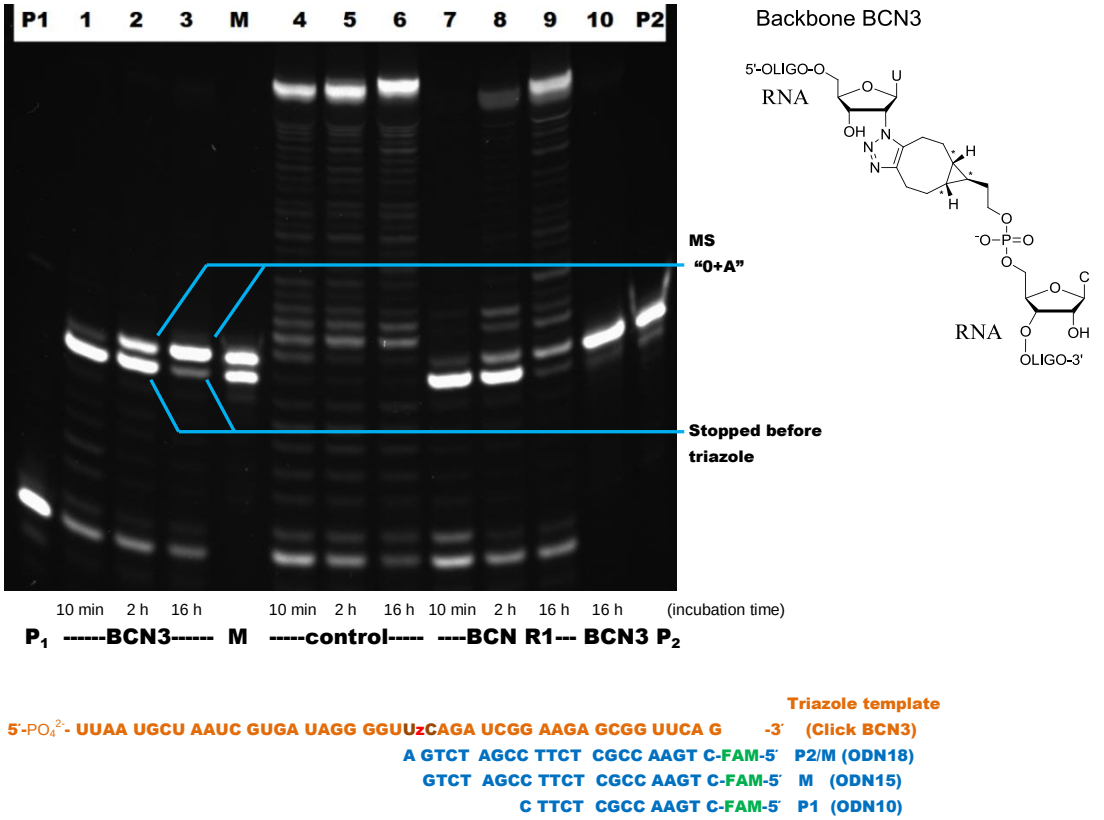


Figure 7.52: M-MuLV Reverse Transcriptase (RNase H^-) in Mg^{2+} buffer overnight reverse transcription comparison of triazole backbone BCN3 (RNA adapter, U-BCN-C) and BCN R1 (RNA adapter, C-BCN-C).

RT primers (with 5´-FAM, Lane P1, P2: primers in water):

Lane P1, 1-9: 60 pmol primer ODN10; Lane 10, P2: 60 pmol primer ODN18.

Lane M: ODN15 ("+0") and ODN18 ("+1") as markers

RT template:

Click BCN3 (ORN16 click ORN20), 60 pmol for Lane 1-3, 10;

Control: canonical RNA template ORN15 (CC), 60 pmol for Lane 4-6;

Click BCN R1 (ORN18 click ORN20), 60 pmol for Lane 7-9.

Lane 1-10: same reverse transcription condition as that in Figure 7.38. Reverse transcription: 37 ℃, 10 min/2 h/16 h.

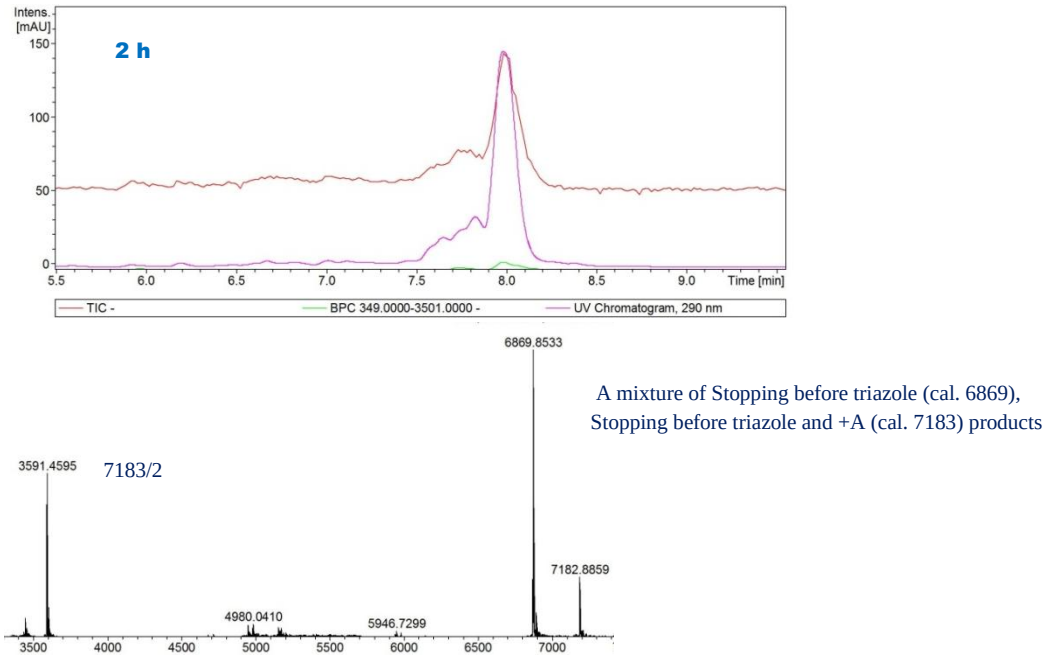


Figure 7.53: M-MuLV reverse transcriptase (RNase H$^-$) RT reaction mixture (3 mM $MgCl_2$ in buffer, primer ODN10) of the triazole template Click BCN3 (U-BCN-C, RNA adapter-part) after incubation at 37 ℃ for 2 h, digestion by RNase H and desalting by gel-filtration (NAP-10) (see Figure 9.83).

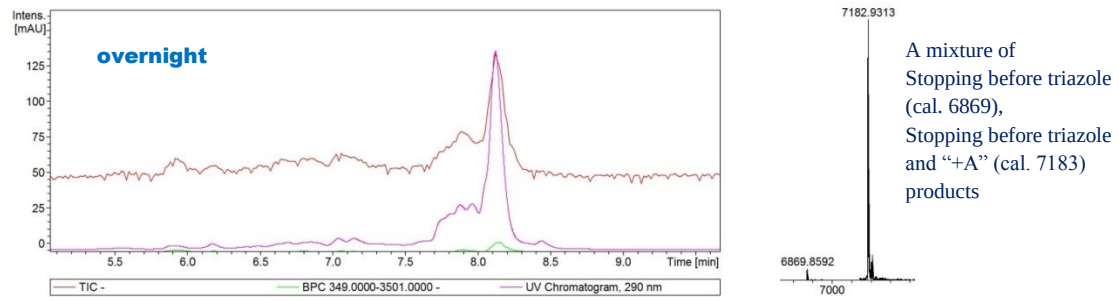
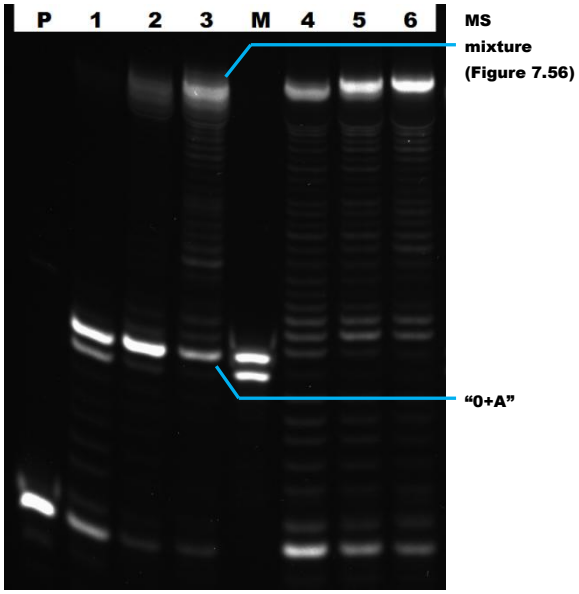
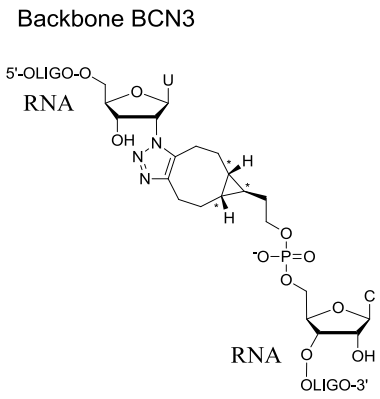


Figure 7.54 (Figure 9.84): M-MuLV reverse transcriptase (RNase H$^-$) RT reaction mixture (3 mM $MgCl_2$ in buffer, primer ODN10) of the triazole template Click BCN3 (U-BCN-C, RNA adapter-part) after incubation at 37 ℃ overnight, digestion by RNase H and desalting by gel-filtration (NAP-10).

The reaction buffers with various divalent cation and KCl concentrations were screened (data not shown) and the buffer containing 3 mM Mg^{2+} and 3 mM Mn^{2+} was found to be the optimal buffer for the M-MuLV Reverse Transcriptase to read-through this "U-BCN-C" linkage. However, both PAGE (Figure 7.55) and MS analysis (Figure 7.56) showed that the reverse transcription produced a mixture of products including two-nucleotide-deletion (likely nucleotides encircled), one-nucleotide-deletion and "deletion+mismatch" products as their "+A" derivatives.





Triazole template

5'-HO- UUAA UGCU AAUC GUGA UAGG GGUUzCAGA UCGG AAGA GCGG UUCA G -3' (Click BCN3)

A GTCT AGCC TTCT CGCC AAGT C-FAM-5' M (ODN18)

GTCT AGCC TTCT CGCC AAGT C-FAM-5' M (ODN15)

C TTCT CGCC AAGT C-FAM-5' P (ODN10)

Figure 7.55: M-MuLV Reverse Transcriptase (RNase H$^-$) overnight test in Mg^{2+} +Mn^{2+} buffer for read-through of triazole backbone BCN3 (RNA adapter)

RT primers (with 5´-FAM, Lane P: primers in water):
Lane P1, 1-6: 60 pmol primer ODN10.
Lane M: ODN15 ("+0") and ODN18 ("+1") as markers
RT template:
Click BCN3 (ORN16 click ORN20, UzC), 60 pmol for Lane 1-3;
Control: canonical RNA template ORN15 (CC), 60 pmol for Lane 4-6.

Lane 1-6: 3 µM primer (60 pmol), 3 µM template (60 pmol), 1x supplied M-MuLV Reverse Transcriptase (RNase H$^-$) Reaction Buffer (NEB®, 50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$, pH 8.3 at r.t.), 3 mM $MnCl_2$, 10 mM DTT, 0.5 mM dNTP, 200 u (1 µL) M-MuLV Reverse Transcriptase (NEB®) in total 20 µL.
Reverse transcription: 37 ℃, 10 min/2 h/16 h.

As the reverse transcription of Click BCN3 produced a mixture, the RT product of Click BCN3 was also PCR amplified and sequenced using the same strategy for Click BCN R1 and control template mentioned above. Most of the sequences of 20 clones (Figure 7.57) were consistent with the MS result (Figure 7.56). Eight out of 20 clones had the sequence of "-AC" (such as clone No.2), which corresponds to the mass of 13907 (about 20 % of RT products in the MS result). Nine out of 20 clones had the sequence of "-A" (such as clone No.4), which corresponds to the highest MS peak with the mass of 14196. Only 1 clone had the sequence of "-A and U:G mismatch" (clone No.11), which corresponds to the mass of 14212. Based on the

MS and sequencing results, it is not clear which of the 4 consecutive "G's" after the triazole was deleted or mismatched.

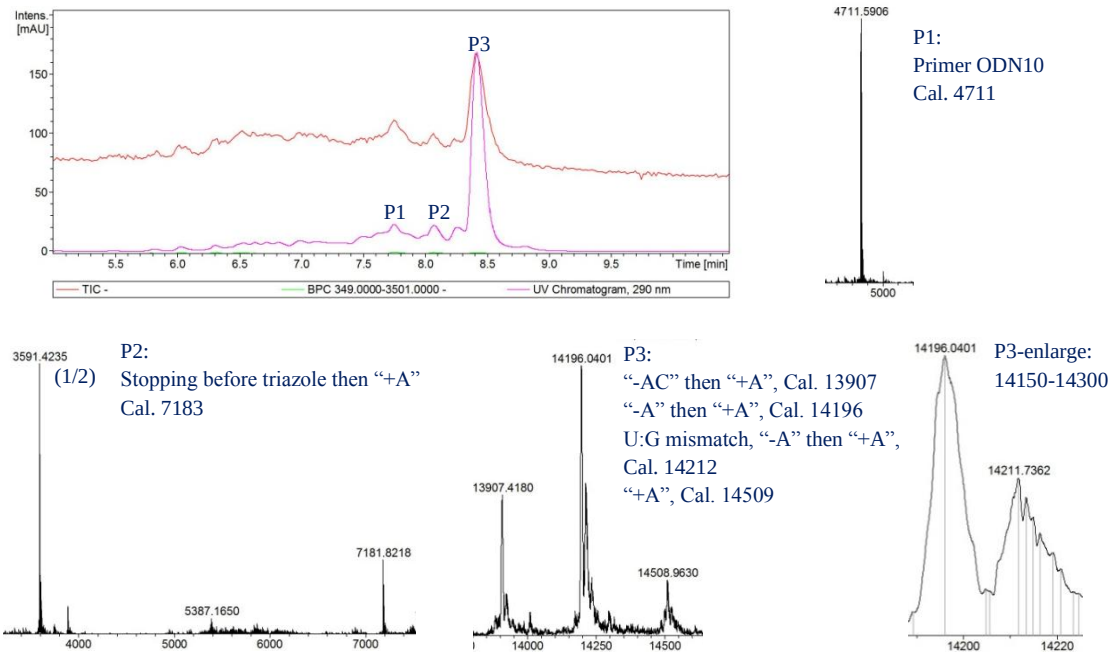


Figure 7.56: M-MuLV reverse transcriptase (RNase H$^-$) RT reaction mixture (3 mM $MgCl_2$ and 3 mM $MnCl_2$ in buffer, primer ORN10) of the triazole template Click BCN3 (U-BCN-C, RNA adapter-part) after incubation at 37 ℃ overnight, digestion by RNase H and desalting by gel-filtration (NAP-10) (see also Figure 9.85).

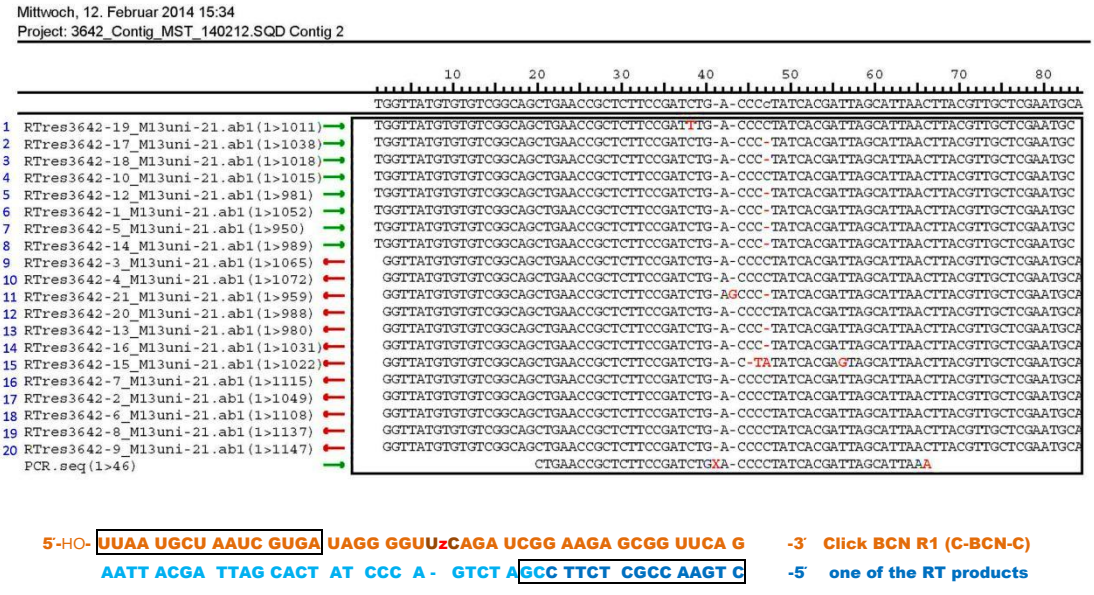


Figure7.57: Sequences of the 20 clones of the PCR amplified RT products of triazole-BCN template Click BCN3 (U-BCN-C), aligned to the expected sequence at the bottom.

Overnight reverse transcription requires minimised RNase H contamination, which could become a problem for natural RNA samples. Another potential problem of overnight reverse transcription is that the long incubation time may allow some second-strand synthesis to take place [6]. Although a sequencing library consists of dsDNAs, the sequences of the 5′-adapter and 3′-adapter indicate which single strand of the dsDNA comes from a small RNA. But if the reverse transcription primer annealed to a small RNA region with semi-complimentary sequence instead of 3′-adapter region, it may initiate second-strand synthesis during overnight incubation (Figure 7.58). This will introduce fake sequences complimentary to the actual small RNAs to the sequencing library, because these second-strands will have the sequence complementary to the 3′-adapter sequence from the RT primer. This is a reported problem for mRNA sequencing [37], however, for small RNA sequencing the second-strand synthesis may not be a problem. This is because small RNAs are much shorter than mRNA, so the reverse transcription is unlikely to stop before reaching the end, therefore the generated second-strand will have both the sequence complementary to the 3′-adapter at the 5′-end and 3′-adapter sequence at the 3′-end (Figure 7.58). This makes these fake sequences distinguishable so they can be easily removed.

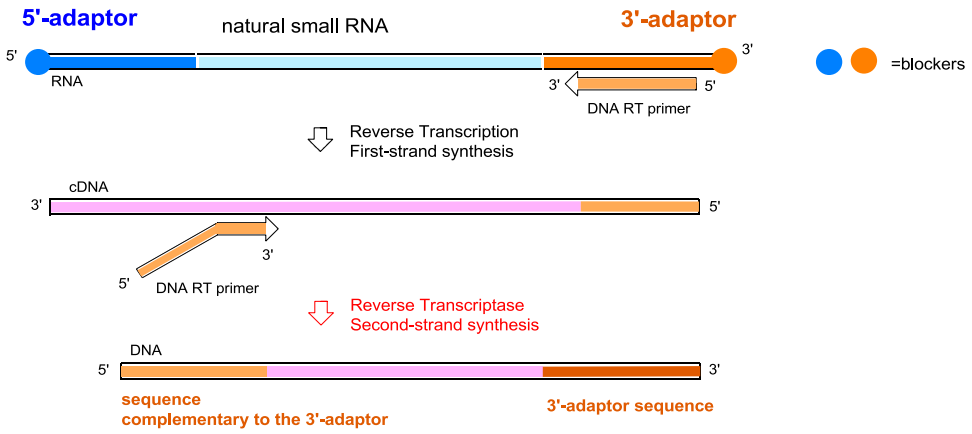


Figure 7.58: Potential problem of second-strand synthesis during reverse transcription

7.9 Conclusions

RNA Click ligation is proposed for 3′-adapter ligation during small RNA sequencing library preparation. Read-through of the triazole Backbones B and B′ (formed from 3′-*O*-propargyl RNA and 5′-azido adapter) by M-MuLV Reverse Transcriptase (wild type) was achieved with the sequence "CzC" but not well with the sequence "CzU" at the triazole site. The reverse transcription caused a one "G" deletion. However, introducing 3′-*O*-propargyl group by nucleoside triphosphate was difficult. This reduced the potential of triazole Backbone B and B′.

As the 2′-azide can be introduced by a nucleoside triphosphate, the reverse transcription of triazole Backbone C (formed from 2′-azido RNA and 5′-alkynyl adapter) became important. The triazole linkage on Backbone C with the sequence "UzT" was successfully read-through by a M-MuLV Reverse Transcriptase (wild type), but the reverse transcription generated a deletion at the small RNA section of the template. Triazole Backbone C with the sequence "CzT" was also examined but the RT efficiency was lower than triazole Backbone B and B′.

More importantly, the triazole-BCN linkage in template Click BCN R1 (C-BCN-C, formed from 2′-azido RNA and 5′-BCN RNA adapter), generated by SPAAC Click ligation, was read-through overnight by M-MuLV Reverse Transcriptase and the reaction produced a clean, one "G" deletion product. The deletion was confirmed by sequencing to be the first nucleotide after triazole-BCN linkage. This suggests that copper-free Click ligation has the potential to be utilised to further modify the small RNA sequencing library preparation protocol. Based on the triphosphate incorporation results, template Click BCN3 (U-BCN-C) was also examined for reverse transcription. However, this template generated a mixture of RT products under optimal conditions, so it is not ideal for our sequencing application.

7.10 Further discussion I– Phosphoramidate backbone and 5′-adapter ligation

In our modified small RNA sequencing protocol, the 3′-adapter was ligated through copper-catalysed Click reaction or copper-free Click reaction. The enzymatic 5′-adapter ligation can also be substituted by an enzyme-free chemical ligation strategy. As mentioned above, miRNA, siRNA and piRNA all naturally possess a 5′-monophosphate group. This can be utilised to form a P3′→N5′ phosphoramidate link with a 3′-amine group on a 5′-adapter.

Oligonucleotide phosphoramidates includes P3′→N5′ phosphoramidate, N3′→P5′ phosphoramidate and phosphoramidate with a free amine on the phosphorus (Figure 7.59). The P3′→N5′ phosphoramidate link was first made in 1976 [38]. The 5′-P-N bond was introduced on to a 5′-amino-5′-deoxythymidine-5′-triphosphate, accepted by DNA polymerase I. As this triphosphate with a P-N link was mixed with other natural dNTPs during reaction, the P3′→N5′ phosphoramidate links were mixed with natural phosphodiester link in the "DNA chain". N3′→P5′ Phosphoramidate, however, generated much more commercial interest because the oligo-2′-deoxynucleotide N3′→P5′ phosphoramidate (containing only this linkage) formed stable duplexes with complementary DNA and RNA and also triplexes with dsDNAs [39]. It is expected to be a good antisense agent [40]. The phosphoramidate oligomer with a free amine on the phosphorus has also been synthesised using the *H*-phosphonate approach [41,42]. The I_2 oxidation of *H*-phosphonate in presence of ammonia generated this phosphoramidate. The free amine on this backbone can be further functionalized with other residues to increase its duplex stability with another normal ssDNA [43].

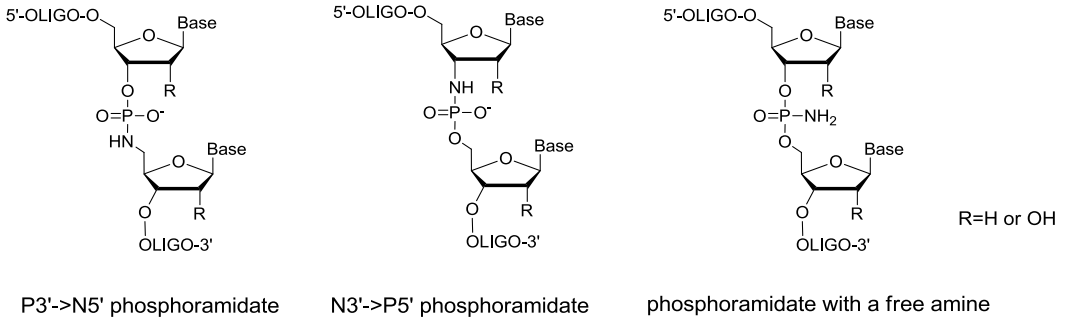


Figure 7.59: Three classes of oligonucleotide phosphoramidates

Both P3′→N5′ phosphoramidate and N3′→P5′ phosphoramidate are unstable at pH lower than 5 but stable in alkaline and neutral media. This is because the protonated nitrogen makes

the phosphorus susceptible to nucleophilic attack by water. N3′→P5′ Phosphoramidates are highly resistant to cleavage by 3′-exonucleases and endonucleases (which catalyse cleavage of "3′-O-P" bond). P3′→N5′ Phosphoramidates are labile to 3′-exonucleases (similar to phosphodiester link) but resistant to 5′-exonucleases which cleaves "P-O-5′" bond [39].

N3′→P5′ Phosphoramidates were recently utilised for enzyme-free RNA 5′-labelling [44]. The reaction went to completion in aqueous buffer at room temperature (Figure 7.60). However, the ligation requires the help of a template, which generates additional complexity to our sequencing protocol. In the published study, the 5′-fragment of the ligation pair with a 3′-NH_2 group was a DNA. It is possible to use the same strategy to ligate an RNA 5′-adapter with a 3′-NH_2. This might be necessary for efficient reverse transcription.

Due to its structural similarity with the phosphodiester linkage, N3′→P5′ phosphoramidate is likely to be read-through by a reverse transcriptase. The ability of N3′→P5′ phosphoramidate single linkage in the template to be read-through by DNA polymerase during primer extension, including PCR, was mentioned in a patent document [45]. However, the read-through of this link during reverse transcription has not been reported. In addition to this, because Click ligation and phosphoramidate ligation are orthogonal to each other, it is also possible to use a single-adapter approach similar to the one discussed above.

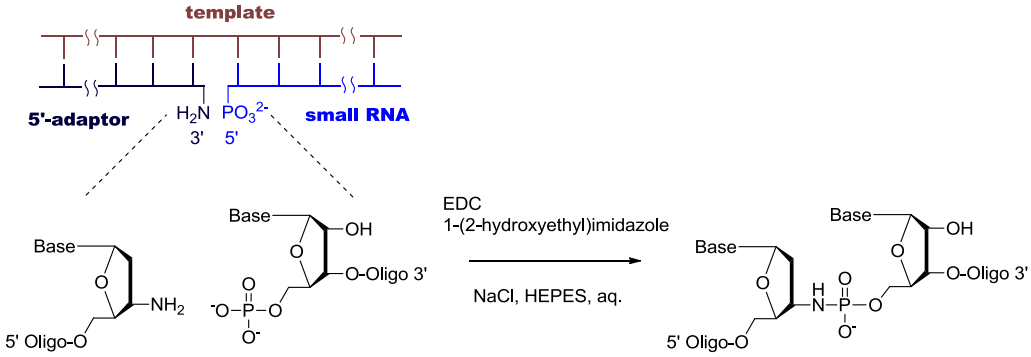


Figure 7.60: N3′→P5′ Phosphoramidate has the potential for small RNA sequencing.

7.11 Copper-catalysed cleavage of RNA

During preparation of RNA triazole backbones, it was repeatedly found that the gel-purified products were partially degraded (e.g. Appendix, Figure 9.30). Monitored by a time-course experiment using one RNA strand, without the adapter to "click", the degradation was seen to follow a distinctive pattern. It progressively removed two uridine nucleotides at the 5′-end of the RNA strand and the cleavage was complete in 2 h (Figure 7.61).

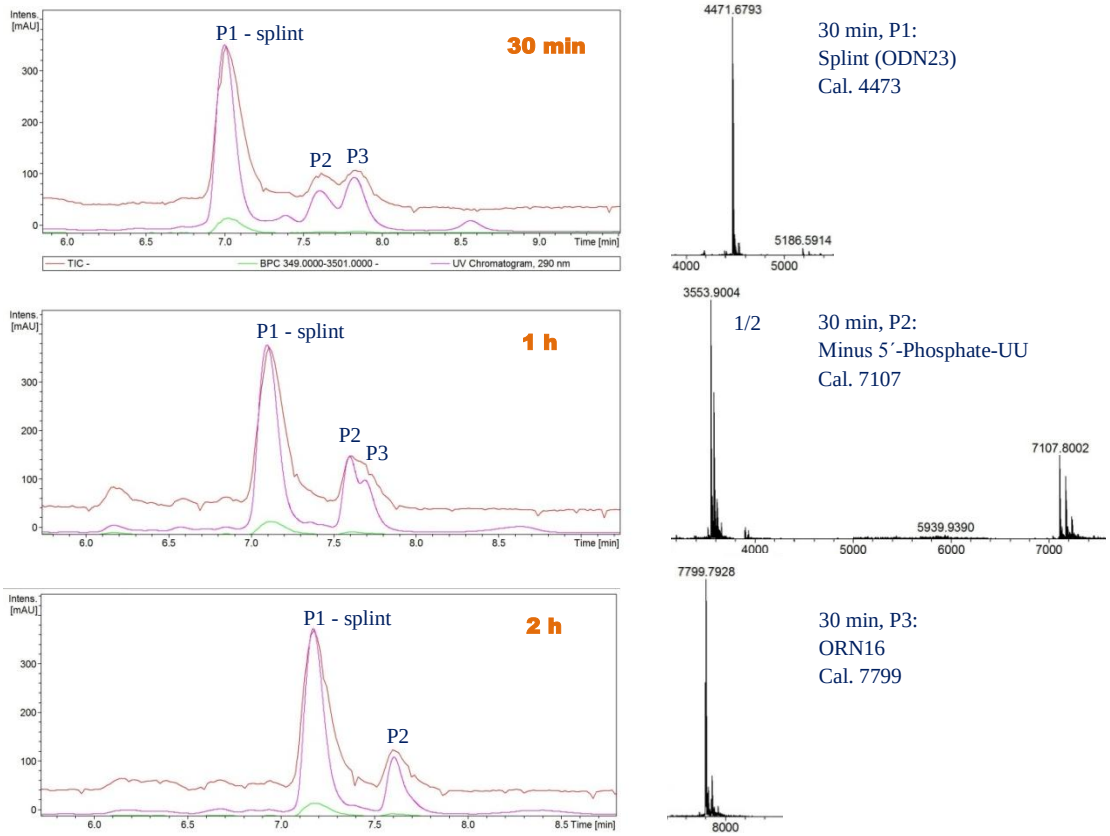


Figure 7.61: Progressive copper catalysed degradation of ssRNA ORN16 in 2 h (From Appendix, Figure 9.86, 9.87, 9.88).

By screening different reaction conditions, it was found that the cleavage happened at the 5′-end of RNA with either a 5′-OH or a 5′-phosphate group. The "5′-UU" cleavage of RNA with a 5′-phosphate was faster and finished in 2 h (Appendix, Figure 9.95), while the cleavage of the RNA with a 5′-OH was incomplete in 2 h (Appendix, Figure 9.92). The DNA splint is not

essential for the cleavage. Both Cu^{I} (with sodium ascorbate) and Cu^{II} can catalyse cleavage and the reaction also requires salt (NaCl) and copper-ligand (THPTA) (Table 7.2).

		5´-OH				5´-Phosphate				RNA primer
		No splint		with splint		No splint		with splint		
		-	+	-	+	-	+	-	+	THPTA
Cu^{I}	-	N/A	No	N/A	N/A	N/A	No	N/A	N/A	
	+	N/A	Yes	N/A	Yes	N/A	Yes	N/A	Yes	
Cu^{II}	-	N/A	No	N/A	N/A	N/A	N/A	N/A	N/A	
	+	No	Yes[1]	N/A	Yes	N/A	Yes[2]	N/A	Yes	
	0.2 M NaCl									

Table 7.2: Copper-cleavage condition screening. Yes: cleavage observed, No: no cleavage, N/A: not tested. 1: Figure 9.92, 2: Figure 9.95.

The copper-kanamycin complex (Figure 7.62) was reported in 1999 to catalyse fast and specific cleavage of RNA under mild conditions [46]. In our study, THPTA was used as a copper ligand instead of an aminoglycoside. Our cleavage reaction at room temperature has a similar rate to this report. It suggests that the cleavage found in our study is a similar copper-catalysed reaction. As the cleavage also requires a special secondary structure of RNA, NaCl (salt) is an essential requirement for the reaction.

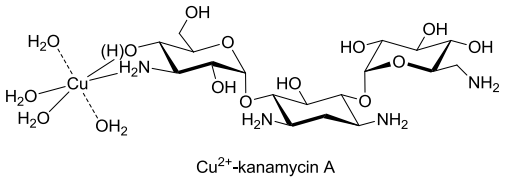


Figure 7.62: Cu^{2+}-kanamycin A complex [46].

7.12 Further discussion II– *in situ* RT-PCR of specific small RNA

In situ RT-PCR is an available technique to compare expression level of one mRNA among adjacent cells in a section of fixed tissue. It has been utilised to detect the spatial distribution of a neuropeptide-related receptor mRNA in various mouse tissue sections such as sections from spinal cord [47]. This method has not been reported for small RNA distribution detection. By selecting different parts of the embryos for Illumina® small RNA sequencing, it has been reported that miRNA expression is spatially localised in early frog embryos [48]. It is uncertain if the spatial pattern of small RNA expression is linked to embryo development and other Pathogenic or cancer developments.

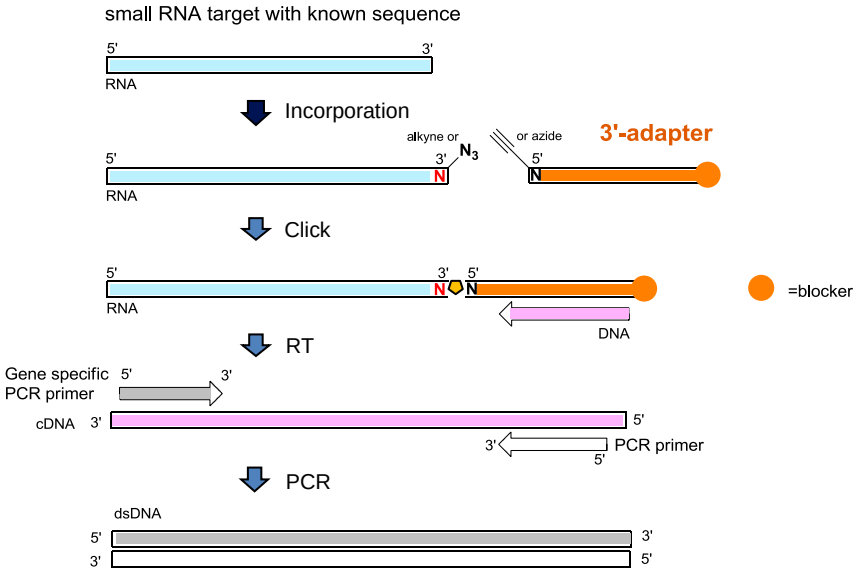


Figure 7.63: Proposal of small RNA RT-PCR using Click ligation

In situ RT-PCR may be difficult for small RNAs. Although ligation of small RNAs from a total RNA extract has been reported for labelling and detection of small RNAs [49], it has not been reported for small RNAs in fixed tissues. For *in situ* RT-PCR of mRNA the reverse transcription is initiated using a poly(T) primer, which anneals to the poly(A) tail of mRNA [47]. This tail is not available for small RNAs, so for small RNA *in situ* RT-PCR, one 3´-adapter is needed to be ligated to the 3´-end of small RNAs (Figure 7.63). The protocol is a simplified version of the sequencing protocol described above, with only one ligation step.

Similar to our small RNA sequencing protocol, the crucial requirement of Click ligation for *in situ* RT-PCR is that the combination of labelling and Click ligation needs to be more efficient than the T4 RNA ligase catalysed adapter ligation. As the ligation of small RNAs needs to be performed upon fixed tissue on a microscope slide, Click ligation may be favorable. However, the efficiency of the ligase-catalysed adapter ligation and enzymatic 3´-labelling in fixed tissues needs to be compared.

1. Memczak, S. et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333-338 (2013).
2. Zhuang, F., Fuchs, R.T. & Robb, G.B. Small RNA expression profiling by high-throughput sequencing: Implications of enzymatic manipulation. *J. Nucleic Acids.* **2012**, 360358-360358 (2012).
3. Carthew, R.W. & Sontheimer, E.J. Origins and mechanisms of miRNAs and siRNAs. *Cell* **136**, 642-655 (2009).
4. Rana, T.M. Illuminating the silence: understanding the structure and function of small RNAs. *Nat. Rev. Mol. Cell Biol.* **8**, 23-36 (2007).
5. Wang, Z., Gerstein, M. & Snyder, M. RNA-seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57-63 (2009).
6. Mamanova, L. & Turner, D.J. Low-bias, strand-specific transcriptome Illumina sequencing by on-flowcell reverse transcription (FRT-seq). *Nat. Protoc.* **6**, 1736-1747 (2011).
7. Hafner, M. et al. Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods* **44**, 3-12 (2008).
8. Soares, A.R., Pereira, P.M. & Santos, M.A.S. Next-generation sequencing of miRNAs with Roche 454 GS-FLX technology: steps for a successful application. *Methods in molecular biology (Clifton, N.J.)* **822**, 189-204 (2012).
9. Morin, R.D. et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.* **18**, 610-621 (2008).
10. Zhang, H. et al. Genome-wide analysis of small RNA and novel microRNA discovery in human acute lymphoblastic leukemia based on extensive sequencing approach. *PLoS One* **4**, 10 (2009).
11. Bullard, D.R. & Bowater, R.P. Direct comparison of nick-joining activity of the nucleic acid ligases from bacteriophage T4. *Biochem. J* **398**, 135-144 (2006).

12. Gerard, G.F. & D'Alessio, J.M. Reverse transcriptase (EC 2.7.7.49): The use of cloned Moloney murine leukemia virus reverse transcriptase to synthesize DNA from RNA. *Methods in Molecular Biology; Enzymes of molecular biology* **16**, 73-93 (1993).
13. Speckmann, W.A., Terns, R.M. & Terns, M.P. The Box C/D motif directs snoRNA 5'-cap hypermethylation. *Nucleic Acids Res.* **28**, 4467-4473 (2000).
14. Munafo, D.B. & Robb, G.B. Optimization of enzymatic reaction conditions for generating representative pools of cDNA from small RNA. *RNA* **16**, 2537-2552 (2010).
15. Zhelkovsky, A.M. & McReynolds, L.A. Simple and efficient synthesis of 5' pre-adenylated DNA using thermostable RNA ligase. *Nucleic Acids Res.* **39**, E117-U71 (2011).
16. Alefelder, S., Patel, B.K. & Eckstein, F. Incorporation of terminal phosphorothioates into oligonucleotides. *Nucleic Acids Res.* **26**, 4983-4988 (1998).
17. Linsen, S.E.V. et al. Limitations and possibilities of small RNA digital gene expression profiling. *Nat. Methods* **6**, 474-476 (2009).
18. Hafner, M. et al. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA* **17**, 1697-1712 (2011).
19. Jayaprakash, A.D., Jabado, O., Brown, B.D. & Sachidanandam, R. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res.* **39**(2011).
20. Zhuang, F.L., Fuchs, R.T., Sun, Z.Y., Zheng, Y. & Robb, G.B. Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Res.* **40**, 14 (2012).
21. Sorefan, K. et al. Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence* **3**, 4 (2012).
22. Zhelkovsky, A.M. & McReynolds, L.A. Structure-function analysis of methanobacterium thermoautotrophicum RNA ligase - engineering a thermostable ATP independent enzyme. *BMC Mol. Biol.* **13**(2012).
23. Torchia, C., Takagi, Y. & Ho, C.K. Archaeal RNA ligase is a homodimeric protein that catalyzes intramolecular ligation of single-stranded RNA and DNA. *Nucleic Acids Res.* **36**, 6218-6227 (2008).
24. Winz, M.L., Samanta, A., Benzinger, D. & Jaschke, A. Site-specific terminal and internal labeling of RNA by poly(A) polymerase tailing and copper-catalyzed or copper-free strain-promoted click chemistry. *Nucleic Acids Res.* **40**, 13 (2012).
25. England, T.E. & Uhlenbeck, O.C. Enzymatic oligoribonucleotide synthesis with T4 RNA ligase. *Biochemistry* **17**, 2069-2076 (1978).

26. Lingner, J. & Keller, W. 3'-End labeling of RNA with recombinant yeast poly(A) polymerase. *Nucleic Acids Res.* **21**, 2917-2920 (1993).
27. Fujino, T., Yasumoto, K., Yamazaki, N., Hasome, A., Sogawa, K. & Isobe, H. Triazole-linked DNA as a primer surrogate in the synthesis of first-strand cDNA. *Chem. Asian J.* **6**, 2956-2960 (2011).
28. Elbashir, S.M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21-and 22-nucleotide RNAs. *Genes Dev.* **15**, 188-200 (2001).
29. Kotewicz, M.L., Sampson, C.M., Dalessio, J.M. & Gerard, G.F. Isolation of cloned Moloney murine leukemia virus reverse transcriptase lacking ribonuclease H activity. *Nucleic Acids Res.* **16**, 265-277 (1988).
30. Scolnick, E., Rands, E., Aaronson, S.A. & Todaro, G.J. RNA-dependent DNA polymerase activity in 5 RNA viruses: divalent cation requirements. *Proc. Natl. Acad. Sci. USA* **67**, 1789-& (1970).
31. Roth, M.J., Tanese, N. & Goff, S.P. Purification and characterization of murine retroviral reverse-transcriptase expressed in *Escherichia coli*. *J. Biol. Chem.* **260**, 9326-9335 (1985).
32. El-Sagheer, A.H., Sanzone, A.P., Gao, R., Tavassoli, A. & Brown, T. Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **108**, 11338-11343 (2011).
33. El-Sagheer, A.H. & Brown, T. Efficient RNA synthesis by *in vitro* transcription of a triazole-modified DNA template. *Chem. Commun.* **47**, 12057-12058 (2011).
34. Fauster, K. et al. 2'-Azido RNA, a versatile tool for chemical biology: Synthesis, X-ray structure, siRNA applications, click labeling. *ACS Chem. Biol.* **7**, 581-589 (2012).
35. Chapuis, H., Bui, L., Bestel, I. & Barthelemy, P. 2'-lipid-modified oligonucleotides via a 'Staudinger-Vilarrasa' reaction. *Tetrahedron Lett.* **49**, 6838-6840 (2008).
36. Lin, K.Y. & Matteucci, M.D. A cytosine analogue capable of clamp-like binding to a guanine in helical nucleic acids. *J. Am. Chem. Soc.* **120**, 8531-8532 (1998).
37. Wu, J.Q. et al. Systematic analysis of transcribed *loci* in encode regions using RACE sequencing reveals extensive transcription in the human genome. *Genome Biol.* **9**, R3 (2008).
38. Letsinger, R.L., Hapke, B., Petersen, G.R. & Dumas, L.B. Enzymatic-synthesis of duplex circular phiX174 DNA containing phosphoramidate bonds in (-) strand. *Nucleic Acids Res.* **3**, 1053-1063 (1976).
39. Gryaznov, S.M. & Banait, N.S. DNA and RNA analogues - oligonucleotide phosphoramidates with bridging nitrogen. *Expert Opin. Ther. Pat.* **12**, 543-559 (2002).
40. Gryaznov, S. et al. Oligonucleotide n3'->p5' phosphoramidates as antisense agents. *Nucleic Acids Res.* **24**, 1508-1514 (1996).

41. Froehler, B.C. Deoxynucleoside H-phosphonate diester intermediates in the synthesis of internucleotide phosphate analogs. *Tetrahedron Lett.* **27**, 5575-5578 (1986).
42. Froehler, B., Ng, P. & Matteucci, M. Phosphoramidate analogs of DNA - synthesis and thermal-stability of heteroduplexes. *Nucleic Acids Res.* **16**, 4831-4839 (1988).
43. Letsinger, R.L., Bach, S.A. & Eadie, J.S. Effects of pendant groups at phosphorus on binding-properties of d-ApA analogs. *Nucleic Acids Res.* **14**, 3487-3499 (1986).
44. Vogel, H. & Richert, C. Labeling small RNAs through chemical ligation at the 5' terminus: Enzyme-free or combined with enzymatic 3'-labeling. *ChemBioChem* **13**, 1474-1482 (2012).
45. Ellis, N.M., Kuimelis, R.G., Heiner, C.R., Lazaruk, K.D. & Walsh, P.S. Phosphoramidate-phosphodiester oligonucleotide chimera as primers. *Official Gazette of the United States Patent and Trademark Office Patents* **1232**(2000).
46. Sreedhara, A., Patwardhan, A. & Cowan, J.A. Novel reagents for targeted cleavage of RNA sequences: towards a new family of inorganic pharmaceuticals. *Chem. Commun.*, 1147-1148 (1999).
47. Lossi, L., Gambino, G., Salio, C. & Merighi, A. Direct *in situ* RT-PCR. *Methods Mol. Biol.*, **789**, 111-126 (2011).
48. Harding, J.L. et al. Small RNA profiling of *Xenopus* embryos reveals novel miRNAs and a new class of small RNAs derived from intronic transposable elements. *Genome Res.* **24**, 96-106 (2014).
49. Maroney, P.A., Chamnongpol, S., Souret, F. & Nilsen, T.W. Direct detection of small RNAs using splinted ligation. *Nat. Protoc.* **3**, 279-287 (2008).

Chapter 8 Experimental Details

8.1 Chemical synthesis:

All reagents were purchased from Aldrich®, Avocado®, Fluka®, Proligo®, Applied Biosystems®, Glen Research® or Link Technologies® and were used without purification with the exception that the following solvents were purified by distillation: THF (over sodium wire and benzophenone), DCM, diisopropylethylamine and pyridine (over calcium hydride).

Most chemical reactions were carried out under an argon atmosphere using oven-dried glassware with purified and distilled solvents. The organic phase thin layer chromatography (TLC) was performed using Merck Kieselgel 60 F24 (0.22 mm thickness, aluminium backed) and the compounds were visualised by 254 nm UV irradiation and by staining as detailed below. Reverse phase TLC was performed using Merck RP-18 F_{254s} TLC plates. Column chromatography was carried out under pressure using Fisher scientific DAVISIL 60A (35-70 micron) silica. Reverse-phase chromatography was performed using ISOLUTE® Flash C18 columns. The Reversed-phase HPLC was performed on a Gilson® system using a Luna 10u C8 (2) 100Å 10.00×250 mm column (10 micron).

NAP gel-filtration columns were purchased from GE Healthcare® and used according to the manufacturer's instructions.

500 Å 5′-*O*-(4,4′-dimethoxytrityl)(5′-DMT)-2′-*O*-propargyluridine-3′-lcaa CPG resin, 500 Å 5′-DMT-3′-*O*-propargyluridine-2′-lcaa CPG resin, 500 Å 5′-DMT-2′-*O*-propargyladenosine (N-Bz)-3′-lcaa CPG resin and 500 Å 5′-DMT-3′-*O*-propargyladenosine (N-Bz)-2′-lcaa CPG resin were purchased from ChemGenes®. The 2′/3′-*O*-propargyl modified resins were utilised for triphosphorylation, synthesis of "1-4mer" for T4 RNA ligase 1 labelling and synthesis of small RNA section to make the triazole template for reverse-transcription studies (See oligonucleotide synthesis section below).

TLC Stains used: *p*-anisaldehyde

p-anisaldehyde	9.2 mL
glacial acetic acid	3.8 mL
conc. sulfuric acid	12.5 mL
ethanol (95%)	338 mL

^{1}H NMR spectra were measured at 300 MHz on a Bruker AC300 spectrometer or 400 MHz on a Bruker DPX400 spectrometer. ^{13}C NMR spectra were measured at 75 MHz or 100 MHz on the same spectrometers respectively. Multiplicities of ^{13}C signals were determined using the DEPT spectral editing technique. ^{31}P NMR spectra were measured at 121 MHz. Experiments were carried out using $CDCl_3$, DMSO-d_6, MeOD or D_2O. Spectra were referenced to the appropriate solvent peak. Chemical shifts are given in ppm relative to tetramethylsilane and J values are given in Hz and are corrected to within 0.5 Hz. All spectra were internally referenced to the appropriate residual undeuterated solvent signal. Assignment was also aided by ^{1}H-^{1}H COSY and ^{1}H-^{13}C HMQC.

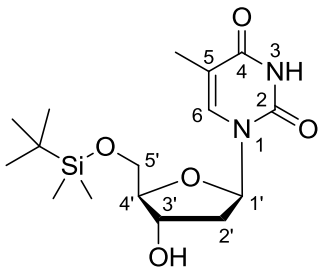
Low-resolution mass spectra were recorded using chemical impact or electron ionization technique on a Thermo Quest Trace GC-MS instrument in DCM (HPLC grade), or using ESI^{+}/ESI^{-} technique on a Micromass Platform II. HPLC-Electrospray Mass spectrometry of the nucleoside triphosphates and oligonucleotides was recorded in water using a Bruker micrOTOF™ II focus ESI-TOF MS instrument in ES^{-} mode and data was processed using Brucker Daltonics® DataAnalysis.

In most cases the best yields have been stated.

dA	15.4	rA	15.4	ATDBio®
dC	7.4	rC	7.2	
dG	11.5	rG	11.5	
T	8.7	U	9.9	
5-Me-dC	5.7			Glen Research®

Table 8.1: ε for triphosphate and oligonucleotide concentration calculation ($x10^3$ $M^{-1}\cdot cm^{-1}$):

5′-*O*-TBDMS-thymidine (18) [1]



Thymidine (2.5 g, 10.3 mmol) was dissolved in anhydrous pyridine with 0.05 eq. of 4-dimethylaminopyridine (DMAP, 64.7 mg, 0.53 mmol). This solution was placed under argon, and 1.1 eq. of *tert*-butyldimethylsilyl chloride (1.720 g, 11.41 mmol) was then added. The mixture was stirred overnight at room temperature. The solvent was removed under reduced pressure and the crude residue purified using flash chromatography with an elution gradient of petroleum ether/DCM/MeOH (start from 20% petroleum in DCM, gradually change to DCM then to 10% MeOH in DCM). Pure product was recovered as a white solid in 78.8% yield. R_f = 0.45 (DCM/MeOH; 9/1; V/V).

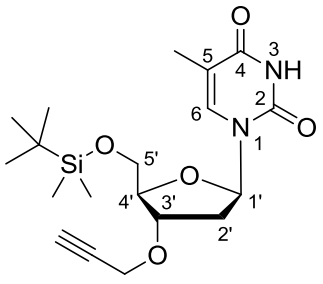
[1]^{1}H NMR (300 MHz, DMSO, d^6): *Thymine*: 11.31 (br s, 1H, N-H), 7.47 (d, 1H, $J_{H6,\ CH3}$ = 0.7 Hz, H_6), 1.77 (d, 3H, $J_{CH3,\ H6}$ = 0.6 Hz, CH_3); *ose*: 6.17 (dd, 1H, J = 6.3 Hz, J = 7.6 Hz, $H_{1'}$), 5.26 (s, 1H, OH), 4.20 (m, 1H, $H_{3'}$), 3.89 (dd, 1H, J = 3.2 Hz, J = 11 Hz, $H_{5'a}$), 3.81 (m, 1H, $H_{4'}$), 3.79 (dd, 1H, J = 3.2 Hz, J = 11.0 Hz, $H_{5'b}$), 2.08 (ddd, 1H, J = 3.0 Hz, J = 6.3 Hz, J = 13.2 Hz, $H_{2'a}$), 2.05 (ddd, 1H, J = 5.8 Hz, J = 7.6 Hz, J = 13.2 Hz, $H_{2'b}$); *TBDMS*: 0.88 (s, 9H, C-CH_3), 0.08 (s, 6H, Si-CH_3).

^{13}C NMR (300 MHz, DMSO-d^6): 163.7 (C4), 150.4 (C2), 135.5 (C6), 109.4 (C5), 86.8 (C4′), 83.8 (C1′), 70.5 (C3′), 63.3 (C5′), 39.4 (C2′), 12.2 (CH_3); *TBDMS*: 25.8 (C-$\underline{C}H_3$), 18,0 ($\underline{C}$-CH_3), -5.4 (Si-CH_3).

LRMS (ESI$^+$, MeCN) m/z = 379.3 ([M + H]$^+$, 18.9%), 420.3 ([M+Na+MeCN]$^+$, 42.4%), 735.5 ([2M+Na]$^+$, 100%), 1091.9 ([3M+Na]$^+$, 15.4%).

3´-*O*-propargyl-5´-*O*-TBDMS-thymidine (19) [1]

*Following the modified propargylation method Used for 5´-O-*DMT-thymidine *[2].*



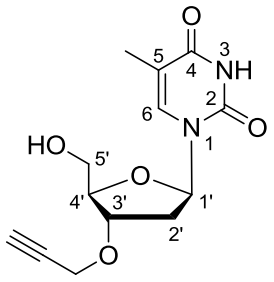
5´-TBDMS-thymidine (**18**, 0.576 g, 1.62 mmol) was co-evaporated with THF 3 times before drying over KOH under vacuum for 3 h. The compound was dissolved in THF(21 mL) and stirred at 0 ℃ on an ice-bath. Sodium hydride (60% dispersion in mineral oil, 0.162 g, 4.04 mmol) was added in portions. The reaction mixture became turbid and was left to stir at r.t. for 1 h then was put back on the ice-bath. Propargyl bromide (80% in toluene, 0.396 mL, 3.56 mmol) was added at 0 ℃ and the reaction was left to stir under reflux for 17 h. The reaction was quenched by adding 10 mL of water and the solvents were removed in vacuo. The residue was dissolved in DCM, washed with water and saturated brine. The organic layer was separated, dried over sodium sulphate and filtered. The solvent was removed in vacuo. The crude product was purified by column chromatography with an elution gradient EtOAc in DCM. Pure product was recovered as a viscous oil in 74.7% yield. R_f = 0.40 (DCM/EtOAc; 2/1; V/V).

[1]^{1}H NMR (300 MHz, DMSO-d^6): *thymine*: 11.34 (br s, 1H, N-H), 7.47 (d, 1H, $J_{H6,\ CH3}$ = 0.8 Hz, H_6), 1.78 (d, 3H, $J_{CH3,\ H6}$ = 0.8 Hz, CH_3); *ose*: 6.09 (dd, 1H, J = 5.7 Hz, J = 8.7 Hz, $H_{1'}$), 4.24 (m, 1H, $H_{3'}$), 3.99 (m, 1H, $H_{4'}$), 3.79 (dd, 1H, J = 4.3 Hz, J = 11.2 Hz, $H_{5'a}$), 3.74 (dd, 1H, J = 3.7 Hz, J = 11.2 Hz, $H_{5'b}$), 2.28 (ddd, 1H, J = 1.6 Hz, J = 5.7 Hz, J = 13.7 Hz, $H_{2'a}$), 2.10 (ddd, 1H, J = 6 Hz, J = 8.7 Hz, J = 13.7 Hz, $H_{2'b}$); *TBDMS*: 0.89 (s, 9H, C-CH_3), 0.09 (s, 6H, Si-CH_3); *propargyl*: 4.22 (d, 2H, J = 2.3 Hz, CH_2), 3.47 (t, 1H, J = 2.3 Hz, C-H).

^{13}C NMR (300 MHz, DMSO-d^6): 163.5 (C4), 150.3 (C2), 135.2 (C6), 109.5 (C5), 83.8 (C4'), 83.8 (C1'), 78.2 (C3'), 63.1 (C5'), 35.9 (C2'), 12.2 (CH_3); *TBDMS*: 25.7 (C-$\underline{C}H_3$), 17,9 ($\underline{C}$-CH_3), -5.6 (CH_3); *Propargyl*: 79.9 (C), 77.3 (CH), 55.7 (CH_2).

LRMS (ESI$^+$, MeCN) m/z = 417.2 ([M + Na]$^+$, 8.4%), 458.3 ([M+Na+MeCN]$^+$, 14.6%), 811.6 ([2M+Na]$^+$, 100%), 869.6 ([2M+2MeCN+H]$^+$, 59.2%).

3′-*O*-propargyl-thymidine (20) [3]



To a solution of 3′-*O*-propargyl-5′-*O*-TBDMS-thymidine (**19**, 3.4054 g, 8.6313 mmol, 1.0 eq.) in dry THF (90 mL) was added 1.0 M TBAF in THF (10.36 mL, 10.36 mmol, 1.2 eq.). The mixture became cloudy after 10 min. The reaction mixture was stirred at r.t. for 40 min. To prevent TBAF from sticking on the product, both water and EtOAc were added to dilute the reaction mixture to prevent it from drying-up during evaporation. The EtOAc and THF were evaporated under vacuum and the rest was mixed with EtOAc (300 mL) and then washed with water (2×100 mL). The product in the aqueous layer of the second round wash was back extracted with EtOAc (100 mL) and the EtOAc layer was separated and used to extract the product in the aqueous solution of the first round water-wash. The procedure was repeated 3 times to insure that no significant product remained in the water. The combined organic layer was dried over Na_2SO_4. The solvent was evaporated to dryness under vacuum affording the crude product as an off-white solid. It was then purified by column chromatography with an elution of 0% - 5% methanol in DCM to offer **20** as a white powder (2.24 g, 92.8%). R_f = 0.44 (EtOAc/MeOH/ammonia; 5.6/0.3/0.3; V/V).

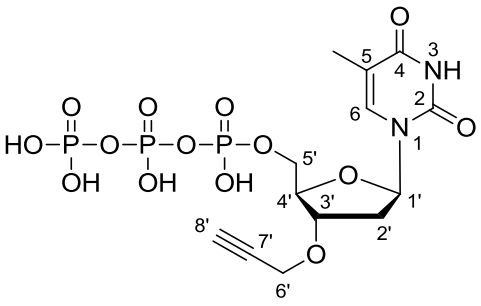
^{1}H-NMR (400 MHz, DMSO-d^6): *thymine*: 11.29 (br s, 1H, N-H), 7.69 (m, 1H, H_6), 1.78 (d, 3H, $J_{CH3,\ H6}$ = 0.8 Hz, CH_3); *ose*: 6.10 (dd, 1H, J = 6.1 Hz, J = 8.1 Hz, $H_{1'}$), 4.24 (m, 1H, $H_{3'}$), 3.94 (m, 1H, $H_{4'}$), 3.60 (dd, 1H, J = 4.3 Hz, J = 11.2 Hz, $H_{5'a}$), 3.56 (dd, 1H, J = 3.7 Hz, J = 11.8 Hz, $H_{5'b}$), 2.26 (ddd, 1H, J = 1.6 Hz, J = 5.7 Hz, J = 13.7 Hz, $H_{2'a}$), 2.13 (ddd, 1H, J = 6.1 Hz, J = 8 Hz, J = 13.8 Hz, $H_{2'b}$); *propargyl*: 4.21 (d, 2H, J = 1.0 Hz, CH_2), 3.45 (m, 1H, C-H).

^{13}C NMR (400 MHz, DMSO-d^6): 163.7 (C4), 150.4 (C2), 135.9 (C6), 109.5 (C5), 84.4 (C4′), 83.8 (C1′), 78.7 (C3′), 61.4 (C5′), 36.0 (C2′), 12.2 (CH_3); *Propargyl*: 80.2 (C), 77.2 (CH), 55.9 (CH_2).

LRMS (ESI$^+$, MeCN) m/z = 242.4 (TBAF, 100%), 344.2 ([M+Na+MeCN]$^+$, 37.0%), 583.3 ([2M+Na]$^+$, 14.2%), 605.3 ([2M+Na+MeCN]$^+$, 9.0%), 863.5 ([3M+Na]$^+$, 9.9%), 1143.8 ([4M+Na]$^+$, 3.0%).

3′-*O*-propargyl thymidine triphosphate(16)

Following the modified triphosphorylation method.[4]



Trimethylphosphate was dried by molecular sieves overnight. 3′-*O*-propargyl thymidine (100 mg, 0.357 mmol, 1.0 equiv., **20**) was dried by dry pyridine evaporation 3 times in a 10 mL flask then dried over KOH under vacuum overnight together with 114 mg proton sponge in a vial. Aliquoted *tris*(tetrabutylammonium) hydrogen pyrophosphate was also further dried over KOH under high-vacuum. 0.4 mL dry DMF was used to dissolve each of the three 500 mg aliquots of *tris*(tetrabutylammonium) hydrogen pyrophosphate in vials and left for 2 h in the presence of molecular sieves to dry the solution. Step 1: dried proton sponge (114 mg, 1.5 equiv), activated molecular sieves and stirrer bar were put inside the 10 mL 3′-*O*-propargyl deoxythymidine flask under argon. The flask was then sealed. The dry trimethylphosphate (1.16 mL) was then injected to the flask. The solution in the flask was cooled by an ice-bath and stirred. $POCl_3$ (0.10 mL, 3.0 equiv.) was dropped in and a white precipitation was observed immediately. The mixture was further stirred for 5 h but the reaction was incomplete as was shown by TLC. An additional 1 equiv. of $POCl_3$ (0.03 mL) was added and the mixture stirred for 50 min, however no significant change was observed. Step 2: A suspension of *tris*(tetrabutylammonium) hydrogen pyrophosphate dried over molecular sieves (1580 mg, 4.9 equiv) in DMF (1.6 mL) and tri-*n*-butylamine (0.36 mL, does not dissolve in DMF) was injected to the flask and the white precipitate disappeared . After 3 h the reaction was quenched by triethylammonium bicarbonate buffer (1 M, pH 7.5, 10 mL) to prevent formation of the lower impurity bands of TLC. The reaction mixture was dried under vacuum then high-vacuum to obtain a slightly yellow oil. Before column chromatography the crude product was dissolved in 30 mL water and washed by 3×20 mL DCM to remove the impurities. The DCM only removed a limited amount of the oil. The product was found to be absent from the DCM and EtOAc layers. (TLC, UV and anisaldehyde stain). The crude product was divided to two portions (to prevent blocking the flash column) and purified by reverse-phase chromatography using an ISOLUTE Flash C18 column with a gradient of MeCN in water from 0% to 40% as the eluent. In both two attempts the majority of the monophosphate and the triphosphate came out very late at about 20%-30% MeCN and the more polar triphosphate band migrated slower than the monophosphate band. This reversed order was also observed in the HPLC chromatogram.

HPLC was carried out using a C8 reverse-phase column with the method "RPSAVE" using a shallow gradient of Buffer B up to 15-20%. The best result was achieved using 5-15% gradient of Buffer B over 11.5 min then pausing at 15% for 3 min for peak 2 to elute (Figure 9.1, Appendix). Buffer A: 0.1 M TEA-bicarbonate buffer (pH 7.8), Buffer B: 0.1 M TEA-bicarbonate buffer (pH 7.8) and 40% MeCN.

HPLC method: RPSAVE

Time (min)	% buffer B	
0.00	0.0	
3.00	0.0	
3.50	3.0	
15.00	15.0*	
16.00	100.0	
17.00	100.0	
17.50	0.0	
20.00	0.0	*: adjustable

The total O.D./mL obtained after HPLC was 55.8 + 52.0 $O.D._{260\ nm}$/mL in 1 mL. As c=A/εL and regarding ε=8500 $L{\cdot}mol^{-1}{\cdot}cm^{-1}$ (Thermo Scientific®), the amount of total triphosphate sample obtained was 16 µmol (4.5%). R_f = 0.12 (*n*-propanol/H_2O/ammonia; 3/1/1; V/V).

^{1}H-NMR (300 MHz, D_2O): *thymine*: 7.70 (d, 1H, $J_{H6,\ CH3}$ = 0.8 Hz, H_6), 1.88 (d, 3H, $J_{CH3,\ H6}$ = 0.8 Hz, CH_3); *ose*: 6.27 (dd, 1H, J = 5.9 Hz, J = 8.8 Hz, $H_{1'}$), 4.52 (m, 1H, $H_{3'}$), 4.32 (m, 1H, $H_{4'}$), 4.18 (m, 1H, $H_{5'a}$), 4.14 (m, 1H, $H_{5'b}$), 2.47 (m, 1H, $H_{2'a}$), 2.30 (m, 1H, $H_{2'b}$); *propargyl*: 4.27 (d, 2H, J = 2.2 Hz, CH_2), 2.87 (t, 1H, J = 2.2 Hz, C-H) ; *TEA*: 3.15 (q, J = 7.3 Hz), 1.23 (t, J = 7.1 Hz).

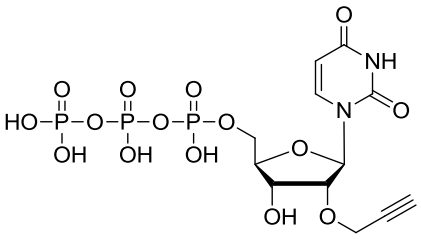
^{31}P-NMR (300 MHz, D_2O): -10.41, -11.15, -23.00.

LRMS (ESI^-, H_2O, micrOTOF™ II) m/z = 519.0 (calculated: 519, $[M-H]^-$)

PrdTTP-Na^+ was obtained by passing the PrdTTP-TEA solution through a small glass column filled with DOWEX 50WX8-200 (Na^+) ion-exchange resin. The DOWEX 50WX8-200 (Na^+) resin was prepared by treating the DOWEX® 50WX8-200 hydrogen form resin (Sigma Aldrich®) with 1 M NaOH solution and water repeatedly. The PrdTTP-Na^+ solution was freeze-dried and the product confirmed by MS-ESI^+ showing no TEA peak. Comparing to PrdTTP-TEA, PrdTTP-Na^+ is less stable to the repeated freeze-thaw cycles according to TLC.

2′-*O*-propargyluridine-5′-triphosphate-TEA salt (30)

Following a modified solid-phase triphosphorylation method.[5]



Step 1: 60 mg of 500 Å 2′-*O*-propargyluridine-3′-lcaa CPG resin (**11**, 37.1 μmol/g, 2.23 μmol) was placed in a glass vial and dried over KOH under vacuum for 2 h then put inside a dried glass column. The column is then sealed and filled with argon and a balloon was attached.

Step 2: 1.5 mL TCA (3% in DCM) solution was added onto the column for 10 min then removed from the column by argon flow. The resin was washed with 3x 1 mL dry DCM then 3x 1 mL dry MeCN.

Step 3: The resin was washed by 0.4 mL dry dioxane + 0.2 mL dry pyridine (mixed inside the column) which were then removed by argon pressure. The tap was quickly closed.

Step 4: The resin was suspended in 0.08 mL dioxane.

Step 5 (Step I): 50 mg aliquoted salicyl chlorophosphite (0.23 mmol, needs to be pure) in a small vial was dissolved in 0.16 mL dry dixoane and 0.08 mL dry pyridine (smoke generated when mixing the pyridine to the salicyl solution, pyridine was refluxed over CaH_2 for at least 2 h then distilled) then dropped onto the resin and left for 15 min. Note: the colour of the resin and the solution changed from light yellow to light olive green over time. It is decided to stop this step at 15 min instead of 1 h to prevent the resin becoming dark green. The initial attempt with 1 h reaction time for this step failed to generate any triphosphate band on TLC plate.

Step 6: The reagents were removed by argon pressure and the resin was washed with 3x 1 mL dry dioxane followed by 3x 1 mL MeCN.

Step 7: The resin was suspended in 0.1 mL DMF.

Step 8: 0.4 mL dry DMF was injected to dissolve the aliquoted 500 mg *tris*(tetrabutylammonium) hydrogen pyrophosphate (The total volume was measured by a syringe to be 0.6 mL). It was used for two triphosphorylation reactions of 2′ and 3′-*O*-propargyluridine.

Step 9 (Step II): 0.2 mL dry tributylamine (did not dissolve in DMF) was prepared in the 1 mL syringe. Half of the pyrophosphate solution (0.3 mL, 0.27 mmol) was taken inside the syringe then both of them were injected inside the column for 2 h (r.t.). The resin turned immediately from off-white to scarlet red as the pyrophosphate solution was injected. The colour did not change over time.

Step 10: After 2 h the reagents were removed from the column by argon pressure and the resin was washed with 3x 1 mL dry DMF followed by 3x 1 mL dry MeCN.

Step 11: 1 mL 0.02 M I_2 in water /pyridine/THF (Applied Biosystems® DNA synthesis reagent) was added to the resin for 30 min. I_2 solution was removed by argon flow and the resin was washed thoroughly with HPLC grade MeCN.

Step 12: After drying by argon flow the resin was transferred from the column to a glass vial and treated with concentrated aqueous NH_3 solution at r.t. for 1 h.

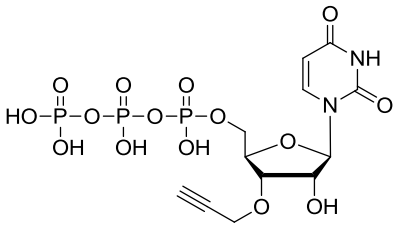
Step 13: The resin was removed by filtration and the solvent was removed under low vacuum on a Buchi then under high vacuum and analysed by TLC.

Step 14: After measuring the UV absorption, the crude product was directly purified by reverse-phase HPLC using a shallow gradient (4-15%) of Buffer B (Buffer A: 0.1 M TEA-bicarbonate buffer, pH 7.8; Buffer B: 40% MeCN in 0.1 M TEAB buffer, pH 7.8). The first round HPLC did not completely separate the triphosphate compound (peak 2, Figure 9.6) from higher TLC bands (peak 1 and 3, Figure 9.6). A second round of reverse-phase HPLC using 4-15% of Buffer B (25% MeCN in 0.1 M TEAB buffer, pH 7.8) was performed and the triphosphate was isolated but with reduced yield. The triphosphate was lyophilised to give a white powder (66 nmol, 10.7%, yield calculated according to $O.D._{260\ nm}$ of the total cleavage products and the final purified triphosphate, ε=8500 $L \cdot mol^{-1} \cdot cm^{-1}$. R_f = 0.12 (*n*-propanol/H_2O/ammonia; 3/1/1; V/V).

LRMS (ESI$^-$, H_2O, micrOTOF^{TM} II) m/z = 521.0 (calculated: 521, $[M-H]^-$)

3′-*O*-propargyluridine-5′-triphosphate-TEA salt (31)

Following a modified solid-phase triphosphorylation method.[5]

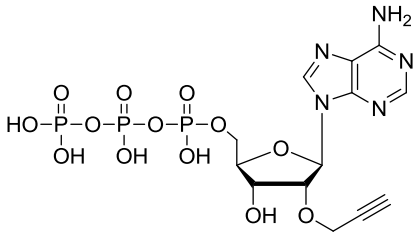


The triphosphorylation starts with 60 mg of 500 Å 3′-*O*-propargyluridine-2′-lcaa CPG resin (**12**, 34.6 µmol/g, 2.08 µmol). The procedure is similar to that for 2′-*O*-propargyluridine triphosphate, except that the salicyl chlorophosphite (Step I) was injected onto the resin for 1 h. The colour of the resin remained off-white. When the pyrophosphate solution (step II) was injected the resin turned to lemon yellow. Only one round of HPLC was performed to successfully isolate the triphosphate (TLC). After measuring the UV absorption the purified product was lyophilised to get an off-white powder (84 nmol, 9.6%, yield calculated according to $O.D._{260\ nm}$ of the total cleavage products and the final purified triphosphate, ε=8500 $L \cdot mol^{-1} \cdot cm^{-1}$). R_f = 0.13 (*n*-propanol/H_2O/ammonia; 3/1/1; V/V).

LRMS (ESI$^-$, H_2O, micrOTOF^{TM} II) m/z = 520.9582 (calculated: 521, $[M-H]^-$)

2′-*O*-propargyladenosine-5′-triphosphate-TEA salt (32)

Following a modified solid-phase triphosphorylation method.[5]



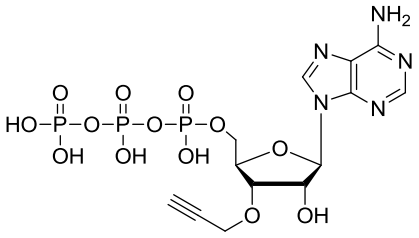
The triphosphorylation was performed on 101 mg of 500 Å 5′-DMT-2′-*O*-propargyladenosine (*N*-Bz)-3′-lcaa CPG resin (**34**, 34 μmol/g, 3.4 μmol). The procedure is similar to that for 2′-*O*-propargyluridine triphosphate except that the amounts of salicyl chlorophosphite and *tris*(tetrabutylammonium) hydrogen pyrophosphate and the solvents to dissolve these reagents have been doubled. The salicyl chlorophosphite (Step I) was injected onto the resin for 15 min. The colour of the resin turned from off-white to yellow then olive-green gradually. When the pyrophosphate solution was injected (step II), the resin turned immediately to scarlet red. After washing off the iodine and drying the resin. It was transferred to a glass vial and concentrated ammonia was allowed to treat the resin at 55 ℃ for 2 h for the cleavage of the oligonucleotide and removal of benzoyl group. One round of HPLC was performed to successfully isolate the triphosphate (monitored by TLC). After measuring the UV absorption, the purified product was lyophilised to get a white powder (439 nmol, 15.8%, yield calculated according to $O.D._{260\ nm}$ of the total cleavage products and the final purified triphosphate, ε=15400 $L \cdot mol^{-1} \cdot cm^{-1}$).

R_f = 0.21 (*n*-propanol/H_2O/ammonia; 3/1/1; V/V).

LRMS (ESI^-, H_2O, micrOTOF^{TM} II) m/z = 544.0 (calculated: 544, $[M-H]^-$)

3′-*O*-propargyladenosine-5′-triphosphate-TEA salt (33)

Following a modified solid-phase triphosphorylation method.[5]



The triphosphorylation was performed on 102.0 mg of 500 Å 5′-DMT-3′-*O*-propargyladenosine (*N*-Bz)-2′-lcaa CPG resin (**35**, 21 µmol/g, 2.1 µmol). The procedure is similar to that for 2′-*O*-propargyladenosine triphosphate synthesis. The salicyl chlorophosphite (Step I) was injected onto the resin for 15 min. The colour of the resin turned from off-white to yellow-olive green gradually. When the pyrophosphate solution was injected (step II), the resin turned immediately to scarlet red. After washing off the iodine and drying the resin. It was transferred to a glass vial and concentrated ammonia was allowed to treat the resin at 55 ℃ for 2 h for the cleavage of the oligonucleotide and benzoyl group deprotection. One round of HPLC was performed to successfully isolate the triphosphate (TLC). After measuring the UV absorption the purified product was lyophilised to get a white powder (44.4 nmol, 8.37%, yield calculated according to $O.D._{260\ nm}$ of the total cleavage products and the final purified triphosphate, ε=15400 $L \cdot mol^{-1} \cdot cm^{-1}$).

R_f = 0.20 (*n*-propanol/H_2O/ammonia; 3/1/1; V/V).

LRMS (ESI$^-$, H_2O, micrOTOF^{TM} II) m/z = 543.9 (calculated: 544, $[M-H]^-$)

8.2 Oligonucleotide synthesis and purification.

8.2.1 Solid-phase chain assembly

Standard DNA and RNA phosphoramidites, solid supports (including 2′-*O*-Me-Ac-C RNA CPG), and additional reagents were purchased from Link Technologies® and Applied Biosystems® Ltd. For the DNA oligonucleotides, *N*(2)-dmf dG, *N*(6)-benzoyl dA, *N*(4)-acetyl dC and unprotected dT phosphoramidites were employed (Applied Biosystems®, Link Technologies®). RNA oligonucleotides were prepared using 2′-TBDMS protected RNA phosphoramidite monomers with *t*-butylphenoxyacetyl protected A, G and C nucleobases and unprotected uracil (Proligo®). 5′-Fluorescein (6-FAM, Link Technologies®) and 5′-BCN (bicyclo[6.1.0]nonyne, 5′-Click-easy™ BCN CEP I, Berry & Associates®) were added as phosphoramidite monomers. 0.3 M Benzylthiotetrazole in acetonitrile (Link Technologies®) was used as the coupling agent, *t*-butylphenoxyacetic anhydride as the capping agent and 0.1 M iodine as the oxidizing agent (Proligo®). Oligonucleotide synthesis was carried out on an Applied Biosystems 394 automated DNA/RNA synthesiser using the standard 1.0 μmole RNA phosphoramidite cycle.

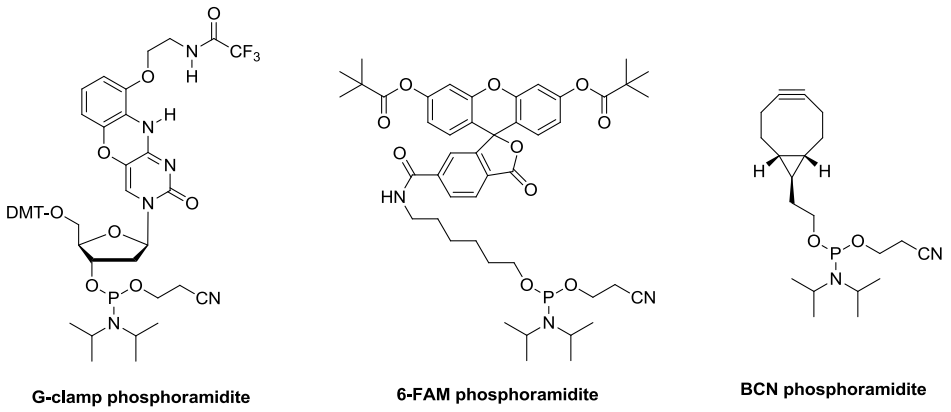


Figure 8.1: AP-dC-CE Phosphoramidite from Glen Research® (G-clamp), 5'-Fluorescein-CE Phosphoramidite (6-FAM) from Link Technologies® and Bicyclo[6.1.0]nonyne (BCN) phosphoramidite monomer from Berry & Associates® (5′-Click-easy™ BCN CEP I).

500 Å 5′-DMT-3′-*O*-propargylcytidine (N-Bz)-2′-lcaa CPG resin (**42**) was purchased from ChemGenes® together with other 2′ or 3′-*O*-propargyl resins described above in the chemical synthesis section. For G-clamp on Backbone BCN2, AP-dC-CE Phosphoramidite was purchased from Glen Research® (Cat. # 10-1097, Figure 8.1) and used with standard methods of assembling and deprotection.

5′-DMT-3′-*O*-propargyl-5-methyldeoxycytidine on solid support (**10**) were synthesised and provided by Dr Afaf H. El-Sagheer [2]. 5′-DMT-2′-azido-2′-deoxycytidine (N-Ac)/uridine on solid support (Figure 8.4) were synthesised and provided by Dr Afaf H. El-Sagheer. These two 2′-azido nucleoside resins were utilised for synthesis of small RNA section to make the triazole template for reverse-transcription study (See oligonucleotide synthesis section below for details and citations). 5′-Alkyne T phosphoramidite (unpublished, Figure 8.2) was synthesised and provided by Dr Afaf H. El-Sagheer.

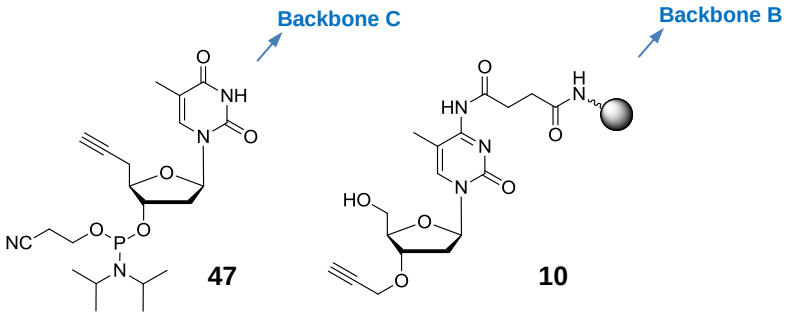


Figure 8.2: Left, 5′-Alkyne T phosphoramidites (**47**) synthesised by Dr Afaf H. El-Sagheer (unpublished). It was used for the synthesis of triazole Backbone C for reverse transcription studies. Right, 5′-DMT-3′-*O*-propargyl-5-methyldeoxycytidine on solid support (**10**) synthesised and provided by Dr Afaf H. El-Sagheer [2]. It was used for the synthesis of triazole Backbone B for reverse transcription studies.

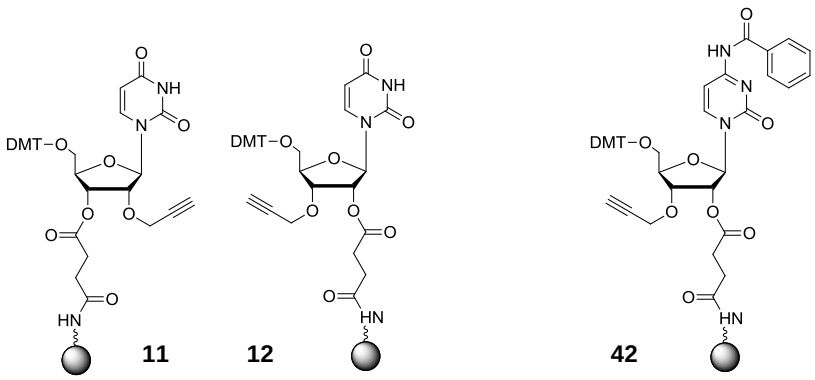


Figure 8.3: Commercially available 2′-*O*-propargyl uridine 3′-LCAA CPG resin (**11**), 3′-*O*-propargyl uridine 2′-LCAA CPG resin (**12**) and 5′-DMT-3′-*O*-propargyl cytidine (*N*-Bz)-2′-lcaa CPG resin (**42**) from ChemGenes® used for 2′/3′-*O*-propargyl oligoribonucleotide synthesis.

For the synthesis of the 3′-*O*-propargyl RNA or 2′-azido RNA, the corresponding 5′-*O*-(4,4′-dimethoxytrityl)-3′-*O*-propargyluridine/propargyladenosine (*N*-Bz)/propargylcytidine (*N*-Bz)-2′-lcaa CPG resin or 5′-DMT-2′-azido-2′-deoxycytidine (*N*-Ac)/uridine solid support (27.6 mg to

35 mg maximum according to the loading of resins, 1.0 μmol scale solid-phase synthesis using phosphoramidites) were used which were packed into twist columns (Glen Research®).

The 5′-*O*-(4,4′-dimethoxytrityl)-2′-azido-2′-deoxycytidine (*N*-Ac)/uridine resins for the synthesis of 2′-azido RNAs were unpublished and were made and granted by Dr Afaf H. El-Sagheer. They were made by functionalising resins with published 5′-*O*-(4,4′-dimethoxytrityl)-2′-azido-2′-deoxycytidine (*N*-Ac) [6] and 5′-*O*-(4,4′-dimethoxytrityl)-2′-azido-2′-deoxyuridine [7]. The oligonucleotide chain was assembled from 3′ to 5′ via standard phosphoramidite chemistry.

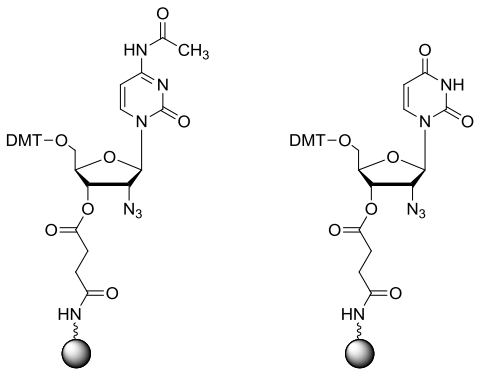
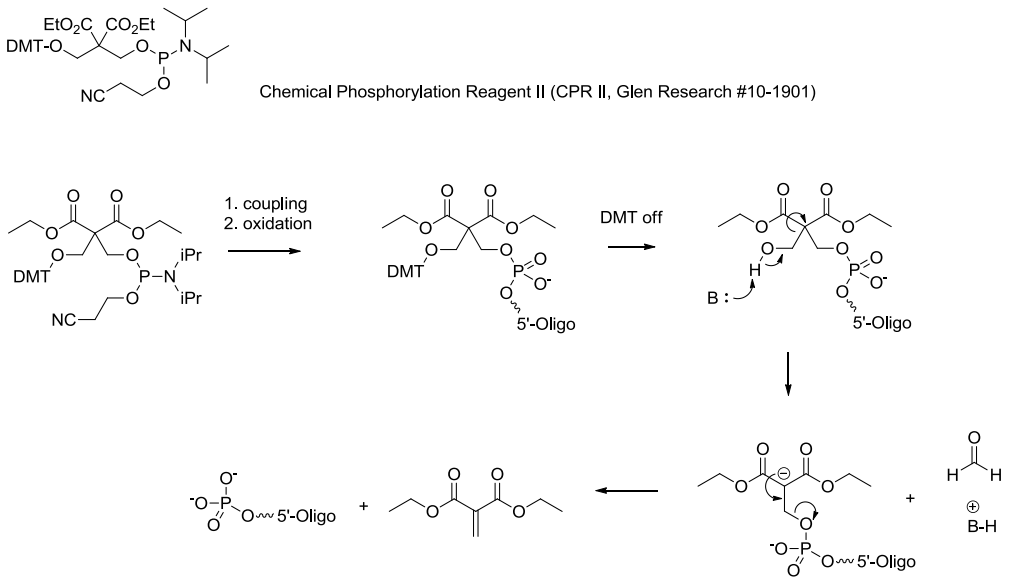


Figure 8.4: 5′-*O*-(4,4′-dimethoxytrityl)-2′-azido-2′-deoxycytidine (*N*-Ac)/uridine resins for the synthesis of 2′-azido RNAs , provided by Dr Afaf H. El-Sagheer

All β-cyanoethyl phosphoramidite monomers were dissolved in anhydrous acetonitrile to a concentration of 0.1 M immediately prior to use. The coupling time for dA, dG, dC, and dT monomers was 35 s and this was extended to 10 min for all other monomers. Stepwise coupling efficiencies were determined through automated trityl cation conductivity monitoring and in all cases were >96.5%. Cleavage of oligonucleotides from the solid support and deprotection were achieved by exposure to concentrated aqueous ammonia/ethanol (3/1, v/v) for 2 h at room temperature followed by heating in a sealed tube for 45 min at 55 ℃.

The 5′-phosphate was assembled using Chemical Phosphorylation Reagent II (CPR II, Glen Research #10-1901) (Scheme 8.1). It is compatible with 5′-DMT-ON purification because DMT protection prevents elimination in ammonia. It also enables cyanoethyl deprotection of oligonucleotide before cleagave and nucleobase-deprotection (ammonia, 55 ℃) to avoid alkylation of the N3 position of thymidine by acrylonitrile (+53 MS peak) under nucleobase-deprotection conditions (Glen Report). The procedure of cyanoethyl deprotection is as below: 10-15 mL 10% DEA/MeCN in a syringe was pushed through the column drop by drop over 10 min and disposed to prevent reverse-alkylation. (The treatment time was limited to 10 min otherwise the oligonucleotide might be cleaved from the resin. The heating of this step was also improper because the 3′-phosphate may migrate to 2′-OH during heating.) The column was then washed

with dry MeCN, dried by argon flow and then stored for ammonia cleavage as mentioned above (3:1 ammonia:ethanol, 55 ℃, 45 min).



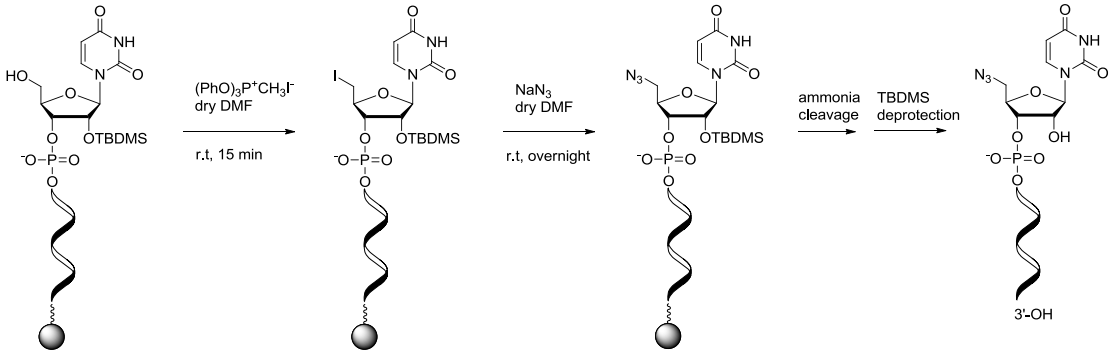
Scheme 8.1: Chemical Phosphorylation Reagent II (CPR II, Glen Research #10-1901) and its mechanism of 5′-phosphorylation

8.2.2 2′-TBDMS deprotection of oligoribonucleotides

After cleavage from the solid support and deprotection of the nucleobases, oligonucleotides containing solution were concentrated to a small volume in vacuo, transferred to 15 mL plastic tubes and freeze dried. The residue was dissolved in DMSO (300 µL) and triethylamine trihydrofluoride (300 µL) was added, after which the reaction mixtures were kept at 65 ℃ for 1.5 h. Sodium acetate (3 M, 50 µL) and butan-1-ol (3 mL) were added with vortexing and the samples were kept at −80 ℃ for 30 min then centrifuged at 4 ℃ at 13,000 rpm for 10 min. The supernatant was decanted and the precipitate was washed twice with ethanol (0.75 mL) then dried under vacuum. The fully deprotected oligonucleotides were purified by reversed-phase HPLC.

8.2.3 Synthesis of 5′-azido RNAs (ORN12 and ORN14)

After assembly on the 1.0 µmol scale (trityl-off), the 5′-OH of the protected RNA oligonucleotide was converted to 5′-azide by treating with a 0.5 M solution of methyl-triphenoxyphosphonium iodide following the established procedure (Scheme 8.2) [2] which is described below.



Scheme 8.2: Synthesis of RNA ORN12 with 5´-azide (Z).

Methyltriphenoxyphosphonium iodide (0.26 g) was weighted in a 25 mL flask. The flask is then sealed and filled with argon. 1 mL dry DMF was injected to dissolve the compound and the solution was sucked out in a 1 mL syringe and periodically passed through the column to another 1 mL syringe then back and forth over 15 min at room temperature. The column was then washed several times with dry DMF. Sodium azide (50 mg) was suspended in 1 mL dry DMF, heated at 70 ℃ for 10 min in a glass vial on the heating block then cooled down (only a small portion of the NaN_3 had dissolved). The supernatant was taken into a 1 mL syringe and passed back and forth through the column then left to react at room temperature overnight. The column was washed with DMF and MeCN and dried by argon flow. The 5´-azido-RNA was cleaved from the solid support. The procedure of TBDMS deprotection is the same as that for other RNA oligonucleotide.

8.2.4 Purification of oligonucleotides

The Reversed-phase HPLC was performed on a Gilson system using a Luna 10u C8 (2) 100Å pore Phenomenex 10.00×250 mm column (10 micron) with a gradient of acetonitrile in 0.1 M TEA-acetate solution (0% to 35% buffer B over 15 min, flow rate 4 mL/min, buffer A: 0.1 M TEA-acetate, pH 7.0, buffer B: 0.1 M TEA-acetate, pH 7.0, with 50% MeCN). Elution was monitored by UV absorption. After HPLC purification, oligonucleotides were desalted using NAP-10 columns.

The same HPLC system was used for triphosphate purification but with a shallower gradient of acetonitrile in 0.1 M TEA-bicarbonate solution (0% to 15% buffer B over 15 min, flow rate 4 mL/min, buffer A: 0.1 M TEA-acetate, pH 7.0, buffer B: 0.1 M TEA-acetate, pH 7.8, with 40% MeCN). Elution was monitored by UV absorption. After HPLC purification, the solution was concentrated under high-vacuum then transferred and lyophilised.

8.3 Tables of ODNs and ORNs in this study

ODN	name	5′-3′	Cal. Mass	Found	Note
ODN1	Res2360	5′-N_3-CGCGTCTAACC	3302	3301.5989	primer for Figure 3.2
ODN2	B1149	5′-FAM-ATTT TACC CTCT GAAG GCTC C	6868	6869.2324	primer for Figure 3.3
ODN3	Res2465	5′-FAM-GGTTATGTGTGTCGGCAG	6138	6138.0531	Primer for others
ODN4	Res2466	CAGTCACAAAACTGCCGACACACATAACC	8778	8777.54	template
ODN5	Res2607	KXACTAGCTCGAATTCATGCCAG	6814	6814.3118	duplex with ODN6
ODN6	Res2608	KXACTGGCATGAATTCGAGCTAG	6894	6894.3682	duplex with ODN5
ODN7	Res2609	KXAACTAGCTCGAATTCATGCCAG	7127	7127.3470	duplex with ODN8
ODN8	Res2610	KXAACTGGCATGAATTCGAGCTAG	7207	7207.3433	duplex with ODN7

K = 5′-hexynol, X = C12 spacer, HO-(-CH_2-)$_{12}$-OH, 202.33 (M.W.)+61.96=264.29

Table 8.2: DNA oligonucleotides synthesised for DNA polymerase incorporation of triphosphates

name	5′-3′	Cal. Mass	Found	Note
ODN3+PrT	5′-FAM-GGTTATGTGTGTCGGCAGT_k	6480	6481.2171	9 °N_m DNA pol.
ODN3+AZT	5′-FAM-GGTTATGTGTGTCGGCAGT_z	6467	6468.0931 6467.9773	TdT (NEB) 9 °N_m DNA pol.
ODN4+T	CAGTCACAAAACTGCCGACACACATAACCT	9082	9081.6898	MS: Mg^{2+}-buffer
ODN5+AZT	KXACTAGCTCGAATTCATGCCAGT_z	7143	7143.2426	duplex with ODN6
ODN6+AZT	KXACTGGCATGAATTCGAGCTAGT_z	7223	7223.2460	duplex with ODN5
ODN7+AZT	KXAACTAGCTCGAATTCATGCCAGT_z	7456	7456.2764	duplex with ODN8
ODN8+AZT	KXAACTGGCATGAATTCGAGCTAGT_z	7536	7536.2694	duplex with ODN7
ODN7+ODN8 +AZT	One AZT incorporation, one end sealed 7456 + 7536 =14992, 14992-304(T)-25(N_3)=14663	14663	14663.7976	
ODN7+AZT+ ODN8+AZT	Two AZT incorporation, expected two ends sealed Two AZT incorporation, expected one end sealed	14992	14993.6664 14993.5537	
EcoRI cleavage product	EcoRI cleavage of the two ends sealed dsDNA	7514	7514.3692	

T_k: 3′-*O*-propargyl T, T_z: AZT

Table 8.3: DNA Elongation and subsequent CuAAC ligation products

ORN	name	5′-3′	Cal. Mass	Found	Note
ORN1	Res2882	5′-P-X	361	360.0415	1-mer for T4 RNA ligase
ORN2	Res2883	5′-P-CX_1	666	665.0752	2-mer
ORN3	Res2884	5′-P-CUX_1	972	971.0885	3-mer
ORN4	Res2885	5′-P-CUAX_1	1302	1300.1297	4-mer
ORN5	Res1864	UUAAUGCUAAUCGUGAUAGGGGU	7388	7389.0801	Primer, sequence of miR-155
ORN6	Res2887	5′-FAM-UUAAUGCUAAUCGUGAUAGGGGUC	8230	8230.8387	Labelling Target miR-155 + C (for ligase) double HPLC purification
ORN7	Res3099	5′-FAM-UUAAUGCUAAUCGUGAUAGGGGUX_8	8244	8244.7803	Labelling Target, 2′-*O*-Me-rC
ORN8	Res3357	5′-P-CUAC UCGG AAGA GCGG UUCA X_4	6868	6868.6924	For T4 RNA ligase 1, not used

P = phosphate, X_1 = 3′-propargyl C (2′-OH, from RNA resin), X_8=2′-*O*Me-rC.

X_4=3′-*O*-propargyl G, M.W. 321.29, ε = 11.5 x10^3 M^{-1}· cm^{-1}

Table 8.4: RNA oligonucleotides synthesised for 3′-labelling studies

RNA enzymatic labelling products:

name	5′-3′	Cal. Mass	Found	Note
ORN5+2′-N_3-2′-dU	UUAAUGCUAAUCGUGAUAGGGGUU_z	7719	7720.0813 7720.0200	Yeast PAP *E.coli* PAP
ORN5+ 2x (2′-N_3-2′-dU)	UUAAUGCUAAUCGUGAUAGGGGUU_zU_z	8050	8051.0459	*E.coli* PAP
ORN2+ORN6	5′-FAM-UUAAUGCUAAUCGUGAUAGGGGUC CX_1	8878	8881.0821	T4 RNA ligase 1
ORN3+ ORN6	5′-FAM-UUAAUGCUAAUCGUGAUAGGGGUC CUX_1	9184	9187.0601	T4 RNA ligase 1
ORN4+ ORN6	5′-FAM-UUAAUGCUAAUCGUGAUAGGGGUC CUAX_1	9513	9515.1262	T4 RNA ligase 1

X_1 = 3′-propargyl C (2′-OH, from RNA resin)

Table 8.5: RNA enzymatic labelling products

For reverse transcription studies:

ORN	database	5′-3′	modifications	MS cal.	MS found	Note
ORN9	Res2877	UUAA UGCU AAUC GUGA UAGG GGUX_1	X_1=3′-*O*-propargyl C M.W. 281.26, ε = 7.2	7731	7731.6215	K for Click Old
ORN10	Res3352	5′-P-UUAA UGCU AAUC GUGA UAGG GGUX_1	X_1=3′-*O*-propargyl C M.W. 281.26, ε = 7.2	7811	7811.9655	Same seq. as Res2877
ORN11	Res3360	5′-P-UUAA UGCU AAUC GUGA UAGG GGUX_6	X_6=3′-*O*-propargyl-2′-$d^{Me}C$ M.W. 279.29, ε = 5.7 (Glen Research)	7809	7809.8294	Click with Res2879, Res3354
ORN12	Res2879 Res3288	UAGA UCGG AAGA GCGG UUCA G	5′-N_3	6814+25 = 6839	6839.6515	Z for Click Old, sequence from Illumina® 3′-adaptor [8]
ORN13	Res3059	UUAA UGCU AAUC GUGA UAGG GGUC UAGA UCGG AAGA GCGG UUCA G		14570	14572.2511	RT control-CU

ORN14	Res3354	CAGA UCGG AAGA GCGG UUCA G	5′-N_3	6814+25 = 6839	6839.0134	Click with Res3352
ORN15	Res3356	UUAA UGCU AAUC GUGA UAGG GGUC CAGA UCGG AAGA GCGG UUCA G	CC control	14569	14570.1726	RT control-CC
ORN16	Res3353	5′-P-UUAA UGCU AAUC GUGA UAGG GGUX_2	X_2=2′-azido-2′-**dU** M.W. 269.21, ε = 9.9	7799	7799.9485	Click with Res3355
ORN17	Res3359	5′-P-UUAA UGCU AAUC GUGA UAGG GGUX_5	X_5=2′-azido-2′-**dC** M.W. 268.23, ε = 7.2	7798	7798.7584	Click with Res3355
ORN18	Res3401 Res3535	UUAA UGCU AAUC GUGA UAGG GGUX_5	X_5=2′-azido-2′-**dC** M.W. 268.23, ε = 7.2	7718	7719.1195	
ORN19	Res3355	X_3AGA UCGG AAGA GCGG UUCA G	X_3=5′-alkyne T M.W. 250.25 ε = 8.7	6821	6820.7702	Click with Res3353/3359
ORN20	Res3358	UUAA UGCU AAUC GUGA UAGG GGUU UAGA UCGG AAGA GCGG UUCA G		14571	14571.4996	RT control-UU
ORN20	Res3551	X_7CAGA UCGG AAGA GCGG UUCA G	X_7=BCN-linker M.W. 226	7040	7039.8983	
ORN21	Res3553	$X_7$5AGA UCGG AAGA GCGG UUCA G	X_7=BCN-linker, M.W. 226, 5=G clamp (2′-H), M.W 376, ε=8.9	7173	7172.7924	

Table 8.6: RNAs used for reverse transcription triazole backbone assembly

			5′	3′		MS cal.	MS found
1	Click Old	CzU	ORN9 3′-OPr-C, 5′-OH	ORN12 5′-N_3-U	Triazole Backbone B′	14570	14570.7741
2	Click A*	CzU	ORN10 3′-OPr-C, 5′-P	ORN12 5′-N_3-U	Triazole Backbone B′	14570+80 =14650	13960.2183* 14652.3054
3	Click D	CzC	ORN10 3′-OPr-C, 5′-P	ORN14 5′-N_3-C	Triazole Backbone B′	14649	14651.1693
4	Click B*	CzU	ORN11 3′-OPr-dC^{Me}, 5′-P	ORN12 5′-N_3-U	Triazole Backbone B	14664-16 =14648	13957.2601* 14650.3048
5	Click E	CzC	ORN11 3′-OPr-dC^{Me}, 5′-P	ORN14 5′-N_3-C	Triazole Backbone B	14647	14648.2202
6	Click F	UzT	ORN16 2′-N_3-2′-dU, 5′-P	ORN19 5′-alkyne-T	Triazole Backbone C	14619	14621.0202
7	Click C*	CzT	ORN17 2′-N_3-2′-dC, 5′-P	ORN19 5′-alkyne-T	Triazole Backbone C	14664-29-17 =14618	13928.1506* 14619.7717
8	Click G Re "C"	CzT	ORN18 2′-N_3-2′-dC, 5′-OH	ORN19 5′-alkyne-T	Triazole Backbone C	14584-29-17 =14538	14540.6360
9	Click BCN1	C-BCN-C	ORN18 2′-N_3-2′-dC, 5′-OH	ODN28 (DNA) 5′-BCN-dC	N/A	7718+6746 =14464	14465.4474
10	Click BCN R1	C-BCN-C	ORN18 2′-N_3-2′-dC, 5′-OH	ORN20 5′-BCN-rC	N/A	7718+7040 =14758	14759.2700
11	Click BCN2	C-BCN-C	ORN18 2′-N_3-2′-dC, 5′-OH	ORN21 5′-BCN-dG-clamp (as dC)	N/A	7718+7173 =14891	14892.6179
12	Click BCN3	U-BCN-C	ORN16 2′-N_3-2′-dU, 5′-P	ORN20 5′-BCN-rC	N/A	7799+7040 =14839	14839.0112

5′-P = 5′-phosphate

*: Significant degradation (minus 5′-P-UU) was observed (Figure 9.30). Click A and Click C was not used for reverse transcription. Click B was used for one experiment (Figure 7.12) demonstrating that the reveres transcription stopped at the triazole site and the result was not affected by this template 5′-degradation.

Table 8.7: Summary of triazole backbones for reverse transcription studies

			MS cal.	found	
ODN9	Res2922	C TGAA CCGC TCTT C	4175	4174.4790	For sequencing
ODN10	Res3116	5´-FAM-C TGAA CCGC TCTT C	4711	4711.6831	RT primer "-7"
ODN11	Res3117	5´-FAM-C TGAA CCGC TCTT CCGA T	5947	5946.8457	RT primer P_1, "-3"
ODN12	Res3344	5´-FAM-C TGAA CCGC TCTT CCGA TCTA	6853	6854.0229	For CU, primer P_2,"+0"
ODN13	Res3324	5´-FAM-C TGAA CCGC TCTT CCGA TCTA G	7183	7183.1024	For CU, primer P_3,"+1"
ODN14	Res3118	5´-FAM-C TGAA CCGC TCTT CCGA TCTA GACC	8074	8075.0938	For CU, primer P_4,"+4"
ODN15	Res3347	5´-FAM-C TGAA CCGC TCTT CCGA TCTG	6869	6870.0019	For CC, primer P_2,"+0"
ODN16	Res3346	5´-FAM-C TGAA CCGC TCTT CCGA TCTG G	7199	7199.1145	For CC, primer P_3,"+1"
ODN17	Res3345	5´-FAM-C TGAA CCGC TCTT CCGA TCTG GACC	8090	8090.2862	For CC, primer P_4,"+4"
ODN18	Res3623	5´-FAM-C TGAA CCGC TCTT CCGA TCTG A	7183	7182.8157	For UC, primer P_3,"+1"
ODN19	Res3394	5´-FAM-C TGAA CCGC TCTT CCGA TCTA AACC	8058	8058.1514	For UT, primer "+4"
ODN20	Res3395	5´-FAM-C TGAA CCGC TCTT CCGA TCTA A	7167	7167.0125	For UT, primer "+1"
ODN21	Res3565	5´-FAM-C TGAA CCGC TCTT CCGA TCTG T	7174	7173.8307	Additional T
ODN22	Res3566	5´-FAM-C TGAA CCGC TCTT CCGA TCTG TG	7503	7502.8141	Additional TG
ODN23	Res2921	CCGA TCTA - GACC CCT	4473	4472.5292	CU splint
ODN24	Res3392	CCGA TCTG - GACC CCT	4489	4488.6523	CC splint
ODN25	Res3393	CCGA TCTA - AACC CCT	4457	4456.6662	UT splint
ODN26	Res3552	C TGAA CCGC TCTT CCGA TCTA GACC CCTA TCAC GATT AGCA TTAA	13660	13657.9715	CU product MIMIC
ODN27	Res3601	C TGAA CCGC TCTT CCGA TCTG GACC CCTA TCAC GATT AGCA TTAA	13676	13674.6408	CC product MIMIC
ODN28	Res3494	X_7CAGA TCGG AAGA GCGG TTCA G — X_7=BCN-linker M.W. 226	6746	6746.3406	For Click BCN1
ODN29	Res3645 B6287	GC ATTC GAGC AACG TAAG TTAA TGCT AATC GTGA	10490	10489.8929	PCR primer P1 Annealing region
ODN30	Res3646 B6288	GGT TATG TGTG TCGG CAGC TGAA CCGC TCTT CCG	10457	10456.8508	PCR primer P2 Annealing region

Table 8.8: DNA oligonucleotides synthesised for reverse-transcription studies

RT product sequence:

			MS Cal.	
	RT of CzU/CzT Full-length	5′-FAM-C TGAA CCGC TCTT CCGA TCTA GACC CCTA TCAC GATT AGCA TTAA	14196	
	RT of CzU/CzT Stopped before triazole	5′-FAM-C TGAA CCGC TCTT CCGA TCTA	6853 as ODN12	
	RT of CzU/CzT Stopped before triazole +A	5′-FAM-C TGAA CCGC TCTT CCGA TCTA A	7167 as ODN20	
	RT of CzU/CzT “+7” (no deletion)	5′-FAM-C TGAA CCGC TCTT CCGA TCTA GACC CCT	8957	
	RT of CzU/CzT “+8” (no deletion)	5′-FAM-C TGAA CCGC TCTT CCGA TCTA GACC CCTA	9270	
	RT of CzC Full-length	5′-FAM-C TGAA CCGC TCTT CCGA TCTG GACC CCTA TCAC GATT AGCA TTAA	14212	
	RT of CzC Deletion G (~~G~~)	5′-FAM-C TGAA CCGC TCTT CCGA TCTG ~~G~~ACC CCTA TCAC GATT AGCA TTAA	13883	
	RT of CzC Full-length +A	5′-FAM-C TGAA CCGC TCTT CCGA TCTG GACC CCTA TCAC GATT AGCA TTAA A	14525	
	RT of CzC Deletion G (~~G~~) +A	5′-FAM-C TGAA CCGC TCTT CCGA TCTG ~~G~~ACC CCTA TCAC GATT AGCA TTAA A	14196	
	RT of UzC Full-length	5′-FAM-C TGAA CCGC TCTT CCGA TCTG AACC CCTA TCAC GATT AGCA TTAA	14196	

Table 8.9: Calculated mass of common RT products

The mass of the “-G”, “+A” products were calculated by minus the corresponding numbers listed below:
-dA: 313
-dT: 304
-dC: 289
-dG: 329

-rA: 329
-rU: 306
-rC: 305
-rG: 345

(No -61.96)

8.4 General materials and procedures for enzymatic reaction and analysis

3′-Azido-2′,3′-dideoxythymidine-5′-triphosphate (AzdTTP) sodium salt was purchased from Jena Bioscience®. 2′-N_3-2′-dUTP sodium salt was purchased from IBA-lifesciences®. 2′-N_3-2′-dCTP lithium salt and 2′-N_3-2′-dATP lithium salt were purchased from TriLink BioTechnologie®.

Incubation was performed on a BIO-RAD T100™ Thermal Cycler. The enzymatic reaction was stopped by freezing the sample in liquid nitrogen then mixing with equal volume of formamide then directly loaded onto the denaturing gel for PAGE analysis (20% acrylamide, 7.0 M urea in 1xTBE buffer, 600 V const. 5.5 h) or diluting to 1 mL and desalted by NAP-10 then freeze-dried overnight for HPLC-MS analysis. 10xTBE buffer was made by dissolving 108 g Tris-base, 55 g boric acid, 9.3 g EDTA acid in water to a total volumn of 1 L.

For non-labelled (without 5′-FAM) oligonucleotides, the acrylamide gel was either visualised by UV-shadowing against a white background on a gel imager (Syngene Gene Genius) or stained by SYBR Gold (Invitrogen®, 20 µL SYBR Gold solution in 100 mL sterilized water) for 30 min in the dark, and then visualised on the gel imager using the transilluminator function. For 5′-FAM labelled oligonucleotide the gel was directly visualised on the gel imager using the transilluminator function.

8.5 DNA 3′-end labelling

8.5.1 Enzymes used for DNA 3′-end labelling experiments

TdT enzyme was purchased from Promega® (#M1871) and New England Biolab® (#M0315). EcoRI was purchased from Promega® (#R6011). Therminator® II (9°N$_m$ DNA pol., D141A/E143A/A485L/Y409V, NEB® #M0266), DNA Polymerase I, Large (Klenow) Fragment (NEB® #M0210) and T7 DNA Polymerase (unmodified, NEB® #M0274) were purchased from New England Biolab. GoTaq™ DNA polymerase was purchased from Promega® (#M3171). EcoRI restriction enzyme was purchased from Promega® (#R6011).

8.5.2 Typical composition for TdT (Promega®)

For unlabeled primer (Figure 3.2): 42 µM primer (ODN1, 1.05 nmol), 0.232 mM triphosphate (5.81 nmol), 1x Terminal Transferase Buffer (Promega®, 1 mM $CoCl_2$), 24 u TdT (Promega®, 30 u/µL) in total 25 µL, 37 °C for 4h. 5x Terminal Transferase Buffer (Promega® #M189A)**:** 500 mM cacodylate buffer (pH 6.8), 5 mM $CoCl_2$ and 0.5 mM DTT.

For 5′-FAM labelled primer (Figure 3.3): 1.88 µM primer (B1149 with 5′-FAM, 47.0 pmol), 0.288~16 mM triphosphate, 1x Terminal Transferase Buffer (Promega®, 1 mM $CoCl_2$), 15 u TdT (Promega®, 30 u/µL) in total 25 µL, incubated at 37 ℃ for 1 h.

8.5.3 Typical composition for TdT (NEB®)

5.75 µM primer (ODN3 with 5′-FAM, 115 pmol), 0.4 mM triphosphate (8 nmol), 1x Terminal Transferase Reaction Buffer (NEB®, 10 mM $Mg(OAc)_2$), 0.25 mM $CoCl_2$, 15 u TdT (NEB®, 20 u/µL) in total 20 µL, incubated at 37 ℃ for 16 h.

1x Terminal Transferase Reaction Buffer (NEB®): 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM Magnesium Acetate, pH 7.9 @ 25 ℃.

8.5.4 Primer extension without thermocycling

For the available DNA polymerases experiment (Figure 3.10), 40 nmol PrdTTP-Na^+ (1.0 mM in 40 µL, regarded as 40 nmol triphosphate and 40 nmol monophosphate/diphosphate according to TLC, caused by degradation of the PrdTTP-Na^+ sample) or 8 nmol dTTP (0.2 mM in 40 µL) was used for each reaction.

DNA Polymerase I, Large (Klenow) Fragment: 1.0 µM ODN3 (5′-FAM primer, 0.04 nmol), 1.0 µM ODN4 (template, 0.04 nmol), 1.0 mM PrdTTP or 0.2 mM dTTP, 1x NEBuffer2, 2.5 u Klenow pol.(5 u/µL, NEB®) in total 40 µL. The reactions were incubated at 25 ℃ (r.t.) for 30 min, then stopped by addition of 4 µL 100 mM EDTA followed by incubation at 72 ℃ 15 min.

T7 polymerase (unmodified): 1.0 µM ODN3 (5′-FAM primer, 0.04 nmol), 1.0 µM ODN4 (template, 0.04 nmol), 1.0 mM PrdTTP or 0.2 mM dTTP, 1x T7 DNA pol. buffer, 0.8 µL(10 mg/mL) BSA, 5 u T7 DNA pol.(10 u/µL, NEB®) in total 40 µL. The reactions were incubated at 37 ℃ for 30 min, then stopped by incubating at 72 ℃ for 15 min.

GoTaq™ polymerase (Promega®): 1.0 µM ODN3 (5′-FAM primer, 0.04 nmol), 1.0 µM ODN4 (template, 0.04 nmol), 1.0 mM PrdTTP or 0.2 mM dTTP, 1x GoTaq colourless buffer, 0.5 µL GoTaq pol. (5 u/µL, Promega®) in total 40 µL. The reactions were incubated at 72 ℃ for 30 min. The polymerase was inactivated by mixing with 40 µL formamide and then freezed or put on an ice-bath.

For the 9 °N_m DNA polymerase reaction without thermocycling (Figure 3.15), the reaction mixture composition is the same as below.

8.5.5 9 °N_m DNA pol. (Therminator® II) incorporation of PrdTTP with thermocycling

For PAGE, 1.65 μM ODN3 (33 pmol, fraction HPLC2-2), 3.30 μM ODN4 (66 pmol, fraction HPLC2-1), 80 μM PrdTTP or dTTP (1.6 nmol), 1x ThermoPol® buffer (NEB®, 20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton® X-100, pH 8.8 at 25 °C) or 1x ThermolPol II®-Mn^{2+} buffer (2 mM $MnCl_2$), 6 u Therminator II (NEB®, 2 u/μL) in total of 20.4 μL. 10x ThermolPol II®-Mn^{2+} buffer was made by adding 1 M $MnCl_2$ solution (Sigma®) to 10x ThermolPol II® buffer (Mg^{2+}-free, NEB®) to adjust the Mn^{2+} concentration to 20 mM.

For HPLC-MS, the reaction was scaled up to 40 μL with the same concentrations as above.

Thermocycling (20/40 μL) program [9]: 20 cycles of 94 °C for 20 sec, 46 °C for 40 sec, 60 °C for 90 sec (method name: PNAS2006). Upon finishing, the samples were topped-up to 1 mL and gel-filtered (NAP-10) then freeze-dried before PAGE instead of direct loading on gel to eliminate the smearing of the bands.

For incorporation of AzdTTP, both the primer and template concentrations were doubled. The rest was the same as above.

8.6 End-sealed dsDNA construction

For Figure 4.2: 10 μM ODN5-ODN6 or ODN7-ODN8 (0.4 nmol each), 200 μM AzdTTP (8 nmol), 1x ThermoPol® (Mg^{2+}) buffer, 6 u Therminator II (NEB®) in total 40 μL. Thermocycling Program: PNAS2006, same as above.

For Figure 4.5, one incorporation reaction was scaled up to 160 μL total volume with the same concentrations as that above. The products were filtered by NAP-10 and freeze-dried. 1/4 of the incorporation product was loaded for Lane P and the rest was used for CuAAC. Also for Figure 4.5-Lane P* an additional incorporation reaction (scaled up to 80 μL total volume) was performed and the products were filtered by NAP-10 and freeze-dried. Half of this additional incorporation reaction product was digested by EcoRI without Click and loaded for Lane P*.

EcoRI cleavage: original duplex or PCR products (0.4 nmol each strand) or PAGE purified CuAAC product (Figure 4.7, assuming 0.4 nmol was recovered) were firstly annealed in 13.3 μL 0.2 M NaCl by heating at 80 °C then slowly cool down. The oligonucleotide solutions were then mixed with Buffer H and EcoRI. Final concentration: 10 μM each strand of the PCR products,

66.5 mM NaCl from oligonucleotide solution, 1x Buffer H (Promega®, 90 mM Tris-HCl, 50 mM NaCl, 10 mM $MgCl_2$, pH 7.5 at 37 ℃), 120 u EcoRI (Promega® #R6011, 12 u/μL) in total 40 μL, incubation at 37 ℃ for 3 h.

Cu^I solution preparation (sequentially added, degassed by argon after each step):

1. 500 μL 0.2 M NaCl
2. 0.3 mg THPTA - Cu^I ligand
3. 2 μL sodium ascorbate solution (100mg sodium ascorbate/mL, 0.5 M)
4. 1.0 μL $CuSO_4$ solution (25 mg $CuSO_4$ $5H_2O$/mL, 100 nmol/μL, 0.1 M)

Copper concentration in Cu^I solution: 100 nmol/500 μL, 0.20 nmol/μL.

The CuAAC reaction (sequentially added, degassed by argon after adding NaCl solution and Cu^I solution):

1. 0.4 nmol (each strand) PCR product
2. 200 μL 0.2 M NaCl
3. 50 μL Cu^I solution (10 nmol, 25 equiv. of Cu^I to the oligonucleotide)

The reaction was left at r.t. for 1 h then stopped by gel-filtering using NAP-10.

The reaction was scaled up (x3) for the Click reaction shown by PAGE (Figure 4.5 and 4.7). 1/3 of this CuAAC reaction product was loaded for Figure 4.5 Lane C. The rest 2/3 was loaded for PAGE purification (Figure 4.7 Lane Click).

For the sealed dsDNA sample analysed by HPLC-MS (Figure 4.6), the enzymatic incorporation reaction and CuAAC reaction were both scaled up (x4). The Click product was diluted to 1 mL and desalted by NAP-10, then freeze-dried and concentrated in 10 μL for HPLC-MS (ESI⁻). The MS acquisition method was changed to "ST300 DG8 N3pint0 Cap-100 skim-30TUNE" to minimize the TEA salt addition.

8.7 RNA 3′-end labelling

8.7.1 Enzymes used for RNA 3′-labelling

E.coli Poly (A) Polymerase (*E.coli* PAP) was purchased from New England Biolab® (NEB #M0276). Cid1 Poly(U) Polymerase (Cid1 PUP, fission yeast) was purchased from New England Biolabs® (NEB #M0337). Yeast Poly (A) Polymerase (Yeast PAP) was purchased from

Affymetrix-USB® (#74225) and Molecular Cloning Laboratories® (MCLAB, #PAPY). RNase H was purchased from New England Biolabs® (NEB #M0297).

8.7.2 Yeast PAP (USB®) and Cid1 PUP 3´-labelling of RNA without 5´-FAM

Yeast PAP (USB®): 3 µM (0.0767 nmol) primer, 5 µL 5x Poly(A) Polymerase Reaction Buffer (USB®, 100 mM Tris-HCl, pH 7.0, 3.0 mM $MnCl_2$, 0.1 mM EDTA, 1 mM DTT, 500 µg/ml acetylated BSA, 50% glycerol), 160 µM (4 nmol) triphosphate, 600 u (1 µL) yeast PAP (USB®) in total 25 µL.

Cid1 PUP: 3 µM (0.0767 nmol) primer, 2.5 µL 10x NEBuffer 2 (NEB®, 500 mM NaCl, 100 mM Tris-HCl, 100 mM $MgCl_2$, 10 mM DTT, pH 7.9 at 25 °C), 160 µM (4 nmol) triphosphate, 2 u (1 µL) Cid1 PUP (NEB®) in total 25 µL.

The reaction mixture was incubated at 37 °C for 1 h. When the reaction was finished, the solution was mixed with 25 µL formamide for 20% PAGE (650 v const., 5 h 30 min). The bands on the gel were visualised by SYBR Gold (Invitrogen®) stain.

8.7.3 Yeast PAP (MCLAB®) 3´-end labelling of RNA with 5´-FAM

Typical labelling reaction using Yeast PAP and 2´-*O*-Pr-ATP, 2´-N_3-2´-dCTP or 2´-N_3-2´-dUTP is as below: 1.5 µM primer (ORN6 or ORN7, with 5´-FAM), 80 µM triphosphate (calculated according to $O.D._{260\ nm}$. CTP, ε = 9000, UTP, ε = 10000, ATP, ε = 15400, Thermo Scientific®), 1x Yeast PAP reaction buffer (MCLAB®, 20 mM Tris-HCl, pH 7.0, 0.6 mM $MnCl_2$, 20 µM EDTA, 0.2 mM DTT, 100 µg/mL Acetylated BSA, 10% Glycerol), 10 mM additional DTT (NEB), 5 u (1 µL) Yeast PAP (MCLAB®, #PAPY-30) in total of 20 µL. For MS the reaction was scaled up to 50 µL in total. Samples were incubated at 37 °C for 1 h.

8.7.4 T4 RNA ligase 1 labelling studies

The reaction mixture was as below: 1.5 µM RNA primer (ORN6 or ORN7, with 5´-FAM, 30 pmol), 10% DMSO, 50 µM monomer-tetramer (1 nmol "2-4 mer"), 1x T4 RNA ligase buffer, 1 mM ATP, 10 u T4 RNA ligase 1 (NEB®) in total 20 µL. The reactions were incubated at 16 °C for 16 h (overnight).

8.8 Preparation of the triazole backbones for reverse transcription studies

8.8.1 Click Old (ORN9 click ORN12) with triazole Backbone B´

The protocol is similar to the hammerhead ribozyme assembly [2]. The reactant concentrations were kept the same except that 20 equiv. of Cu^{I} to alkyne/azide instead of 10 equiv. was applied.

Step1: The $CuSO_4$ and sodium ascorbate solutions in 0.2 M NaCl were prepared as below:
25 mg $CuSO_4 \cdot 5H_2O$ / 1 mL 0.2 M NaCl (0.1 M)
100 mg sodium ascorbate / 1 mL 0.2 M NaCl (0.5 M)
Step 2: 33.1 nmol alkyne oligonucleotide (ORN9, K), azide oligonucleotide (ORN12, Z) and splint (ODN23, T) were annealed in 662 µL 0.2 M NaCl by heating the solution at 80 ℃ for 5 min then cooling slowly to room temperature.
Step 3: 2.1 mg THPTA was dissolved in 350 µL 0.2 M NaCl. The solution was degassed by argon flow. 14 µL 0.5 M sodium ascorbate solution and 7 µL 0.1 M $CuSO_4$ solution were sequentially added each followed degassing.
Step 4: 350 µL of this Cu^{I} solution was quickly added to the degassed oligonucleotide solution and the reaction was left at room temperature for 1 h.
Step 5: The reaction was terminated by gel-filtration (NAP-10). The 1.5 mL final solution was freeze-dried for PAGE purification (20% gel in 1x TBE buffer: 10.8 g Tris(hydroxymethyl) aminomethane, 5.5 g boric acid, 0.93 g EDTA acid in 1 L water).

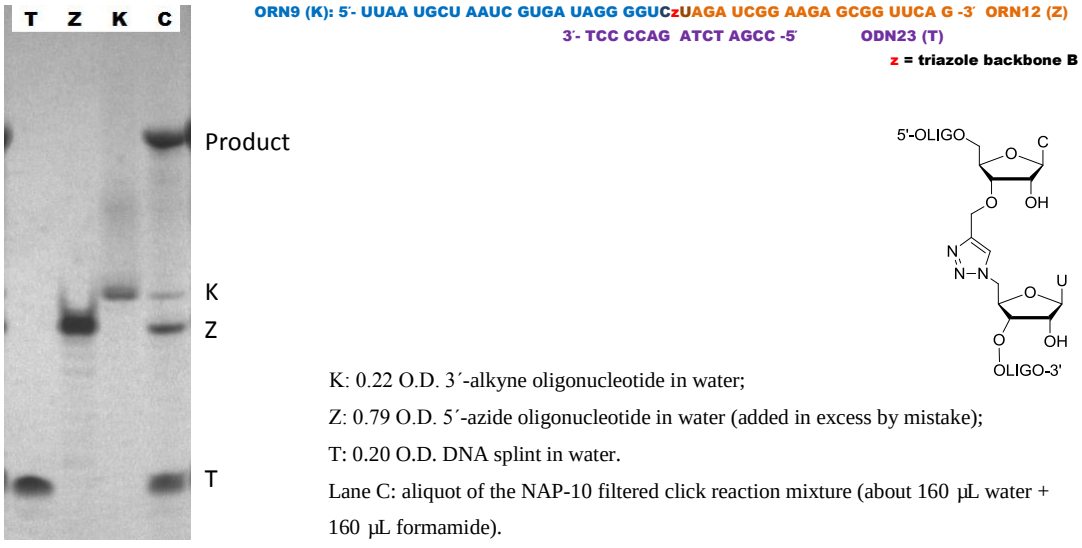
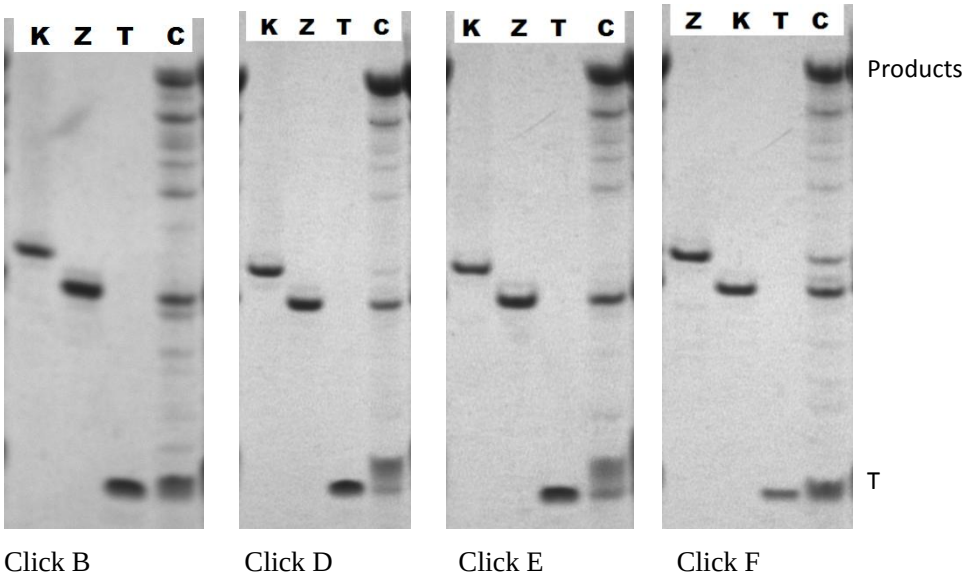


Figure 8.5: Preparation and PAGE-purification of triazole backbone B (Click Old) for the reverse transcription study.

The ligation product (P) was retrieved from the gel by incubating the excised gel slices with 6 mL Tris-HCl buffer (50 mM Tris-base, 25 mM NaCl, pH 7.5) at 37 ℃ overnight and desalted by NAP-25 and NAP-10 gel-filtration then lyophilised.

8.8.2 Preparation of the reverse-transcription template Click B, Click D, Click F

The method is the same as above except that different alkyne oligonucleotide, azide oligonucleotide and splint were used for the assemblies.



Click B K (ORN11): 5′-Ⓟ- UUAA UGCU AAUC GUGA UAGG GGUC^{Me}zUAGA UCGG AAGA GCGG UUCA G -3′ Z (ORN12)
3′- TCC CCAG ATCT AGCC -5′ T (ODN23)

Click D K (ORN10): 5′-Ⓟ- UUAA UGCU AAUC GUGA UAGG GGUCzCAGA UCGG AAGA GCGG UUCA G -3′ Z (ORN14)
3′- TCC CCAG GTCT AGCC -5′ T (ODN24)

Click E K (ORN11): 5′-Ⓟ- UUAA UGCU AAUC GUGA UAGG GGUC^{Me}zCAGA UCGG AAGA GCGG UUCA G -3′ Z (ORN14)
3′- TCC CCAG GTCT AGCC -5′ T (ODN24)

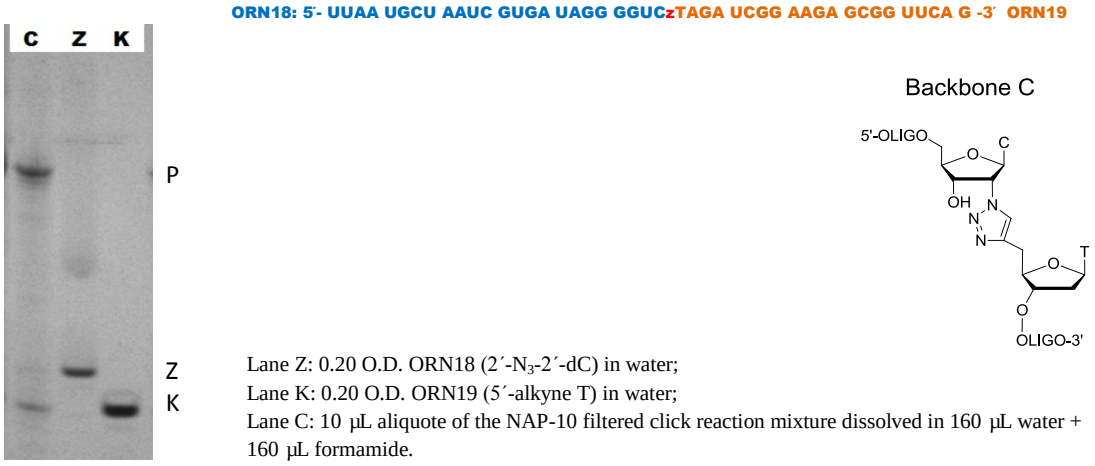
Click F Z (ORN16): 5′-Ⓟ- UUAA UGCU AAUC GUGA UAGG GGUUzTAGA UCGG AAGA GCGG UUCA G -3′ K (ORN19)
3′- TCC CCAA ATCT AGCC -5′ T (ODN25)

Figure 8.6: Preparation and PAGE-purification of triazole backbones via templated CuAAC

8.8.3 Preparation of Click G using non-templated CuAAC in water to avoid degradation

The procedure is similar to the CuAAC in 0.2 M NaCl described above. However water was used to dissolve all reagents and oligonucleotides. The two RNAs were mixed with 8 time higher concentration comparing to CuAAC-in-buffer protocol and without annealing. In detail, 33.1 nmol

azide oligonucleotide (ORN18, Z, 2´-N_3-2´-dC) and alkyne oligonucleotide (ORN19, K, 5´-alkyne T) were dissovled in 82.8 µL water. 2.1 mg THPTA was dissolved in 25 µL water. The solution was degassed by argon flow. 14 µL 0.5 M sodium ascorbate solution (in water) and 7 µL 0.1 M $CuSO_4$ solution (in water) were sequentially added each followed by degassing by argon flow. 43.8 µL of this Cu^I solution was quickly added to the degassed oligonucleotide solution and the reaction was left at room temperature for 2 h. The reaction was terminated by gel-filtration (NAP-10). The 1.5 mL final solution was freeze-dried and purified by PAGE as above.

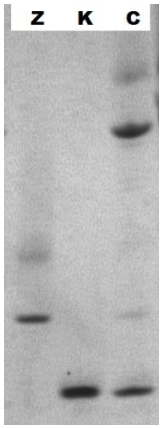


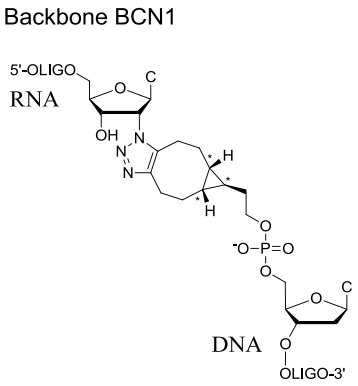
Lane Z: 0.20 O.D. ORN18 (2´-N_3-2´-dC) in water;
Lane K: 0.20 O.D. ORN19 (5´-alkyne T) in water;
Lane C: 10 µL aliquote of the NAP-10 filtered click reaction mixture dissolved in 160 µL water + 160 µL formamide.

Figure 8.7: PAGE (20%) purification of Click G (ORN18 click ORN19).

8.8.4 Preparation of the reverse-transcription template Click BCN1 (C-BCN-C)

30 nmol azide oligonucleotide ORN18 (Z, 2′-N_3) and 30 nmol alkyne oligonucleotide ODN28 (K, 5′-BCN) were mixed in 60 µL water. The reaction was kept at r.t. for 2 h then freezed for polyacrylamide gel electrophoresis (PAGE) next day. The ligation product solution was diluted to 160 µL by water and mixed with equal volume of formamide and purified by PAGE (20% denaturing gel). The gel was visualised by UV shadowing in a gel imager (Syngene® Gene Genius). The ligation product (P) was retrieved from the gel by incubating the cutted gel with 6 mL Tris-HCl buffer (50 mM Tris-base, 25 mM NaCl, pH 7.5) at 37 ℃ overnight and desalted by NAP-25 and NAP-10 gel-filtration then lyophilised. The Same procedure was also used for retrieving enzymatic reaction products from the gel.





ORN18: 5′- UUAA UGCU AAUC GUGA UAGG GGUC$_{BCN}$CAGA TCGG AAGA GCGG TTCA G -3′ ODN28

BCN: triazole-BCN link

Figure 8.8: PAGE purification of Click BCN1 made by non-templated RSAAC (copper-free Click) in water.

ORN18 (Z, 2′-N_3-2′-dC): 30.0 nmol, aliquote 3-1, 3-2.
ODN28 (K, 5′-BCN, DNA): 30.0 nmol, aliquote 3-1.

RSAAC: 60 µL in total, r.t., 2 h then freezed.

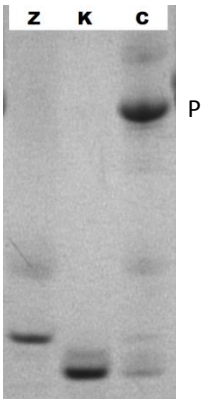
Z: 0.20 O.D. ORN18 (2′-N_3-2′-dC) in water;
K: 0.20 O.D. ODN28 (5′-BCN) in water.
Lane C: 10 µL aliquote of the 320 µL click reaction mixture (60 µL solution + 100 µL water + 160 µL formamide)

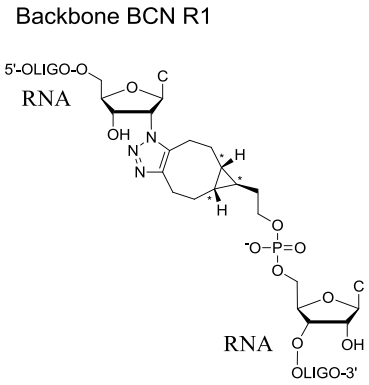
8.8.5 Preparation of the reverse-transcription template Click BCN R1 (C-BCN-C), Click BCN3 (U-BCN-C) and Click BCN2 (C-BCN-G-clamp)

Click BCN R1: 30 nmol azide oligonucleotide ORN18 (Z, 2′-N_3) and 30 nmol alkyne oligonucleotide ORN20 (K, 5′-BCN) were mixed in 60 µL water. The reaction was kept at r.t. for 2 h then freezed for polyacrylamide gel electrophoresis (PAGE) next day. For Click BCN3 and

Click BCN2, different oligonucleotides shown in the figures were “clicked” with the same method above.

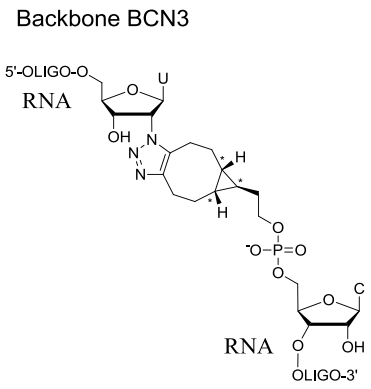
ORN18: 5′- UUAA UGCU AAUC GUGA UAGG GGUC$_{BCN}$CAGA UCGG AAGA GCGG UUCA G -3′ ORN20





Z: 0.20 O.D. ORN18 (2′-N_3-2′-dC) in water;
K: 0.20 O.D. ORN20 (5′-BCN, RNA) in water.

Figure 8.9: PAGE purification of Click BCN R1 made by non-templated RSAAC (copper-free Click) in water.



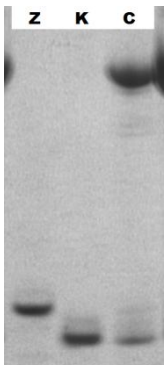
ORN16: 5′- UUAA UGCU AAUC GUGA UAGG GGUU$_{BCN}$CAGA UCGG AAGA GCGG UUCA G -3′ ORN20

Figure 8.10: PAGE purification of Click BCN3 made by non-templated RSAAC (copper-free Click) in water.

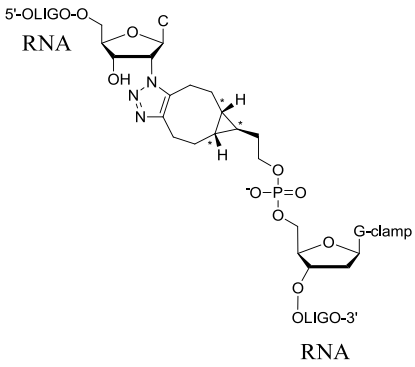
ORN16 (Z, 2′-N_3-2′-dU): 4.6 nmol.
ORN20 (K, 5′-BCN, RNA): 4.6 nmol.
RSAAC: 25.2 µL in total, r.t., 2 h then freeze.

Z: 0.4 nmol (about 0.10 O.D) ORN16 (2′-N_3-2′-dU) in water;
K: 0.4 nmol (about 0.10 O.D) ORN20 (5′-BCN, RNA) in water.

Lane C: 4 µL aliquote of the 92 µL click reaction mixture (dissolved the freeze-dried product in 46 µL water + 46 µL formamide)



Backbone BCN2



ORN18: 5′- UUAA UGCU AAUC GUGA UAGG GGUCBCN**CAGA UCGG AAGA GCGG UUCA G -3′ ORN21**

C: G-clamp

Figure 8.11: PAGE purification of Click BCN2 made by non-templated RSAAC (copper-free click) in water.

ORN18 (Z, 2′-N_3-2′-dC, Res3535): 30.0 nmol, aliquote 3-1.
ORN21 (K, 5′-BCN-G-clamp, RNA): 30.0 nmol, aliquote 3-1.

RSAAC: 60 µL in total, r.t., 2 h then freeze.

Z: 0.20 O.D. ORN18 (2′-N_3-2′-dC) in water;
K: 0.20 O.D. ORN21 (5′-BCN-G-clamp) in water.

Lane C: 10 µL aliquote of the 320 µL click reaction mixture (60 µL solution + 100 µL water + 160 µL formamide)

8.9 Reverse transcription studies

AMV Reverse Transcriptase was purchased from New England Biolab® (NEB # M0277). Superscript® III Reverse Transcriptase was purchased from Life Technologies (# 18080-044). M-MuLV Reverse Transcriptase (wild type) and M-MuLV Reverse Transcriptase (RNase H^-) (also named ProtoScript® II Reverse Transcriptase) were purchased from New England Biolab® (NEB® #M0253, M0368).

The standard condition for MuLV Reverse Transcriptase (wild type) catalysed reverse-transcription is as below: 3 µM primer, 3 µM template, 1x M-MuLV Reverse Transcriptase (RNase H^-) Reaction Buffer (NEB, 50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$, pH 8.3 at r.t.), 10 mM DTT (NEB), 0.5 mM dNTP (Promega®, 0.5 mM each dNTP), 200 U M-MuLV Reverse Transcriptase (wild type, 1 µL, NEB®). For PAGE samples the reactions were performed in 20 µL

total volume. For MS and sequencing samples the reaction was scaled-up to 50 µL and 100 µL total volumes respectively. The samples were incubated at 37 ℃ for 2 h.

The reaction composition for MuLV Reverse Transcriptase (RNase H^-, NEB®) is the same as above. The samples were incubated at 37 ℃ for 2 h or 16 h.

The optimised reaction conditions were summarized in Table 8.10. $MgCl_2$, $MnCl_2$ solution were added to the reaction buffer to adjust the divalent cation concentrations. For Click BCN3 template (U-BCN-C), additional 3 mM $MnCl_2$ beside 3 mM $MgCl_2$ was added to ensure efficient read-through of the triazole-BCN site.

For AMV Reverse Transcriptase (NEB®). Superscript® III Reverse Transcriptase (Life Technologies®), the reaction compositions were listed in the figure caption (Figure 7. 11)

		Backbone assembly		M-MuLV RT (wild type)		Incubation time at 37 ℃	M-MuLV RT (RNase H^-)		Incubation time at 37 ℃
		5′	3′	Mg^{2+}	Mn^{2+}		Mg^{2+}	Mn^{2+}	
Click Old	CzU	3′-OPr-C,	5′-N_3-U	0	3 mM	2 h	0	3 mM	2 h
Click D	CzC	3′-OPr-C,	5′-N_3-C	3 mM	0	2 h			
Click B*	CzU	3′-OPr-dC^{Me},	5′-N_3-U	3 mM (Did not optimise)	0	2 h			
Click E	CzC	3′-OPr-dC^{Me},	5′-N_3-C	3 mM	0	2 h	3 mM	0	2 h
Click F	UzT	2′-N_3-2′-dU,	5′-alkyne-T	3 mM	0	2 h			
Click G Re "C"	CzT	2′-N_3-2′-dC,	5′-alkyne-T	3 mM	0	2 h	3 mM	0	Over-night
Click BCN1	C-BCN-C	2′-N_3-2′-dC,	(DNA) 5′-BCN-dC				0	3 mM	Over-night
Click BCN R1	C-BCN-C	2′-N_3-2′-dC,	5′-BCN-rC				3 mM	0	Over-night
Click BCN2	C-BCN-C	2′-N_3-2′-dC,	5′-BCN-dG-clamp (as dC)				3 mM (Did not optimise)	0	Over-night
Click BCN3	U-BCN-C	2′-N_3-2′-dU,	5′-BCN-rC				3 mM	3 mM	Over-night

Table 8.10: Optimised M-MuLV Reverse Transcriptase (wild type) and M-MuLV Reverse Transcriptase (RNase H^-) reaction conditions for read-through of triazole backbones. The greyed out options were not examined.

The MuLV Reverse Transcriptase (RNase H^-, NEB®) RT samples were treated under the following procedures to minimize the co-migration of the RNA template and the DNA products during PAGE or HPLC by duplex formation. For PAGE, 10 time excess of the unlabelled RT product mimics (ODN26-CU for Click G product; ODN27-CC for Click BCN1 RT products, Click BCN R1 RT product, Click BCN2 RT products, Click BCN3 RT products and CC control

RT products) were added together with the reverse-transcription product and heated at 90 ° for 10 min then cooled down to r.t. The samples were then mixed with equal volume of formamide and loaded for PAGE. For HPLC-MS, the RNA template in the sample was digested by RNase H after reverse-transcription. The reaction was carried out by adding 25 μL solution of 1x MuLV RT (RNase H^-) buffer, 7.5 μL 0.1 M DTT (additional 10 mM in total 75 μL reaction) and 12.5 u RNase H (NEB® #M0297, 5 u/μL) to the 50 μL finished reverse-transcription reaction solution, then incubated at 37 ℃ for 6 h. The products were then gel-filtered (NAP-10) and lyophilised, then dissolved in 10 μL water for MS.

8.10 Reverse transcription and PAGE purification of RT product for sequencing

The reverse transcription products were obtained using unlabelled RT primer (ODN9, no 5´-FAM, same sequence to ODN10)

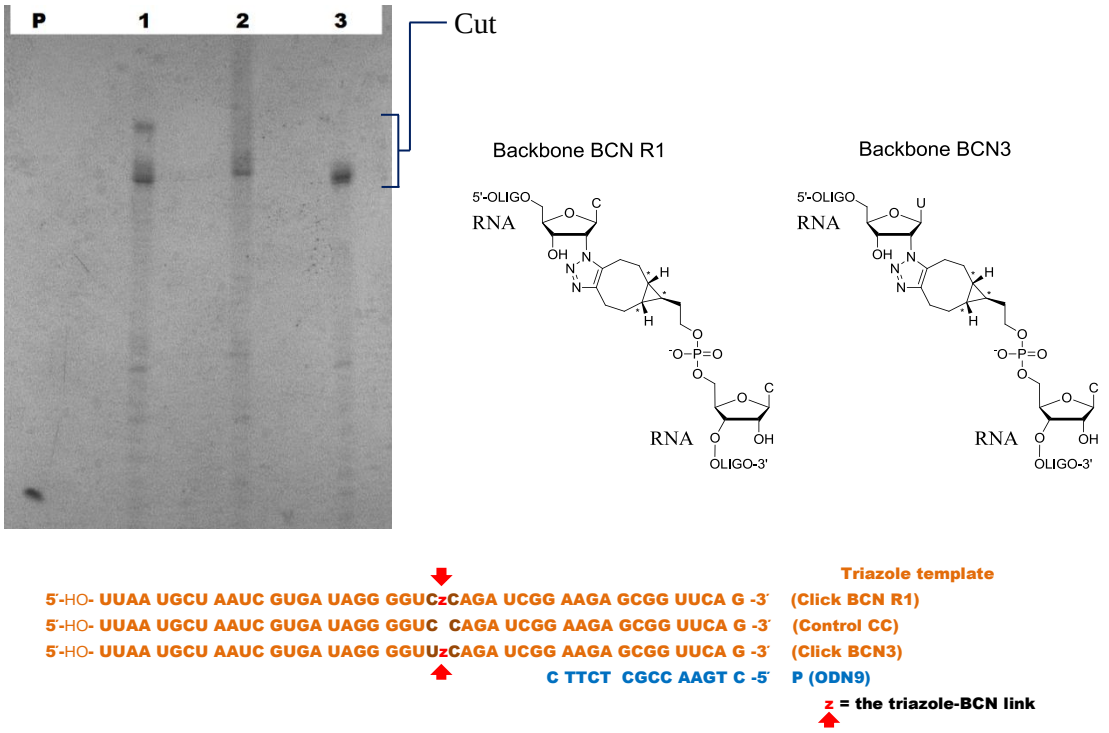


Figure 8.12: M-MuLV Reverse Transcriptase (RNase H$^-$) overnight RT of triazole-BCN backbones in Mg^{2+}-buffer (Click BCN R1, control-CC) or Mg^{2+} +Mn^{2+}-buffer (Click BCN3) with unlabelled primer. Bands in the bracketed region were cut separately for sequencing.

RT template:
Lane 1: 300 pmol Click BCN R1 (CzC, ORN18 click ORN20, RNA adapter),
Lane 2: 300 pmol canonical RNA (ORN15, CC),
Lane 3: 300 pmol Click BCN3 (UzC, ORN16 click ORN20, RNA adapter).

Lane P: primer ODN9 (no 5´-FAM) in water

Lane 1-2: 3 mM $MgCl_2$ in buffer, Lane 3: 3 mM $MgCl_2$ and 3 mM $MnCl_2$ in buffer

RT condition: 3 µM primer (300 pmol), 3 µM template (300 pmol), 1x supplied M-MuLV Reverse Transcriptase (RNase H$^-$) Reaction Buffer (50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$, 0 or 3 mM $MnCl_2$, pH 8.3 at r.t.), 10 mM DTT, 0.5 mM dNTP (each triphosphate), 1000 u (5 µL) M-MuLV Reverse Transcriptase (RNase H$^-$) (NEB®) in total 100 µL. The reactions were incubated at 37 ℃ for 18 h (overnight) then stopped by heating samples at 90 ℃ for 10 min.

RNase H digestion: 50 µL RNase H reaction solution containing 1x M-MuLV Reverse Transcriptase (RNase H$^-$) buffer (NEB®), 15 µL 0.1 M DTT and RNase H (NEB®) and 25 u RNase H (NEB®, 5 u/µL). The reactions were incubated at 37 ℃ for 6 h.

The smeared bands in Figure 8.12 were likely caused by insufficient RNase H digestion of the RNA RT template. So the bracketed regions of the smeared bands were cut for product extraction (incubated with 6 mL water at 37 ℃ overnight, then filtered by NAP-25, NAP-10) in order to increase the amount of oligonucleotide retrieved from gel. The remaining RNA templates were expected to be degraded in water overnight.

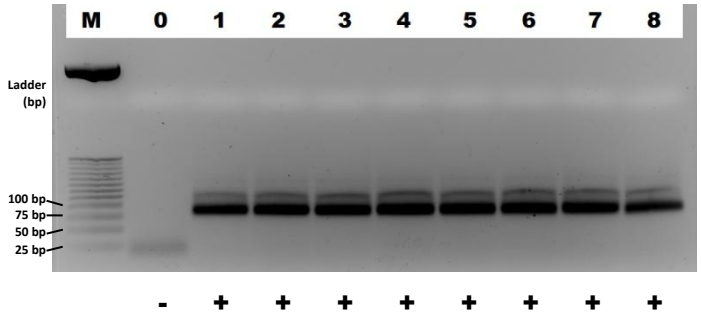
The ssDNAs extracted from gel were freeze-dried, dissolved in 1 mL water for O.D. measurement:

1. Click BCN R1 RT product: 0.1165 O.D./mL, about 0.2 μM
2. ORN15 (CC control) RT product: 0.1214 O.D./mL, about 0.2 μM
3. Click BCN3 RT product: 0.0947 O.D./mL, about 0.2 μM

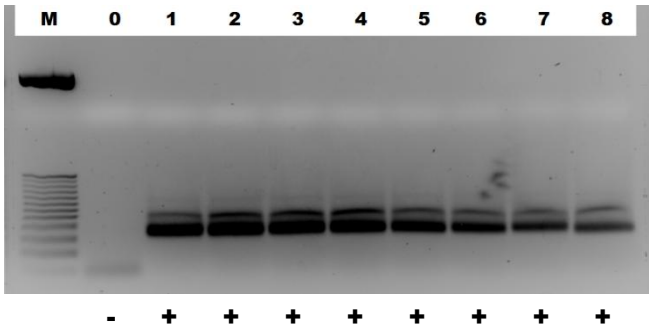
The samples were then freeze-dried and dissolved in 50 μL water, 1 μL of these solutions were diluted 1000 times and about 4 fmol (1 μL) of these purified RT products were amplified by PCR.

8.11 PCR of the reverse transcription product for sequencing

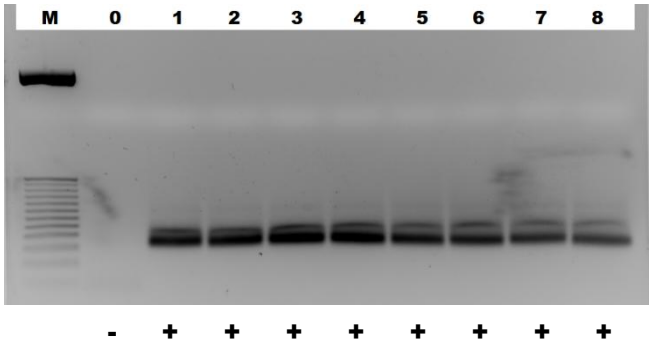
Gel 1: M-MuLV RT (RNase H$^-$) read Click BCN R1



Gel 2: M-MuLV RT (RNase H$^-$) read ORN15 (CC control)



Gel 3: M-MuLV RT (RNase H$^-$) read Click BCN3



P1 (ODN29)
5´-GCATTCGAGCAACGTAAG TTAA TGCT AATC GTGA
5´-HO- UUAA UGCU AAUC GUGA UAGG GGUC-CAGA UCGG AAGA GCGG UUCA G -3´ CC control
AATT ACGA TTAG CACT ATCC CCAG GTCT AGCC TTCT CGCC AAGT C -5´ RT product
GCC TTCT CGCC AAGT C GACGGCTGTGTGTATTGG-5´
P2 (ODN30)

Figure 8.13: 2% Agarose gel electrophoresis purification of 3 PCR products from 3 RT products. The main band (about 90 bp in length) was cut for sequencing.

PCR template:
Gel 1: M-MuLV RT (RNase H$^-$) read Click BCN R1
Gel 2: M-MuLV RT (RNase H$^-$) read ORN15 (CC control)
Gel 3: M-MuLV RT (RNase H$^-$) read Click BCN3

Lane 0: no template, PCR negative control;
Lane 1-8: about 4 fmol (10^9 copies) RT product for each 50 µL reaction.

Lane 0-8: 0.5 µM each primer (25 pmol), 0 or 4 fmol template, 1x Green GoTaq® Reaction Buffer, 0.5 mM dNTP (Promega®, 2.5 µL of 10 mM solution), 6.25 u GoTaq® DNA polymerase (5u/µL) in total 50 µL.

PCR thermocycling: 95 ℃ for 2 min, then 30 cycles of 95 ℃ for 30 sec, 45 ℃ for 30 sec, 72 ℃ for 30 sec.
Agarose gel electrophoresis: 2%, 100 V const., 30 min.

The main PCR band was cut and the dsDNA PCR product was extracted from gel using QIAGEN® QIAquick gel extraction kit following manufacturer's protocol.

The purified PCR products were cloned and sequenced by Eurofins® MWG GmbH.

8.12 Copper catalysed degradation of RNA

Cu^{I} solution in water: 0 or 0.3 mg THPTA, 50 µL water, 2 µL sodium ascorbate solution (100 mg/1 mL water), 1 µL $CuSO_4$ solution (25 mg $CuSO_4 \cdot 5H_2O$/1 mL water) were sequentially added followed by degassing.

Cu^{II} solution was made with the same concentrations as above but without sodium ascorbate.

The cleavage reaction procedure was the same as the CuAAC Click reaction mentioned above for the preparation of triazole backbones. The scale of the reaction was changed to 1 nmol oligoribonucleotide for cleavage. The composition was as below: 1 nmol RNA, 0 or 1 nmol DNA splint, 0 or 0.2 M NaCl, 10.6 µL Cu^{I} or Cu^{II} solution in total 30.6 µL. After 30 min/1 h/2 h, the reaction was stopped by topping-up to 1 mL and gel-filtration by NAP-10. The 1.5 mL filtered solutions were then freeze-dried for HPLC-MS analysis.

1. Lucas, R. et al. Microwave-assisted synthesis of a triazole-linked 3'-5' dithymidine using click chemistry. *Tetrahedron Lett.* **49**, 1004-1007 (2008).
2. El-Sagheer, A.H. & Brown, T. New strategy for the synthesis of chemically modified RNA constructs exemplified by hairpin and hammerhead ribozymes. *Proc. Natl. Acad. Sci. USA* **107**, 15329-15334 (2010).
3. James, D. et al. A 'click chemistry' approach to the efficient synthesis of modified nucleosides and oligonucleotides for PET imaging. *Tetrahedron Lett.* **51**, 1230-1232 (2010).
4. Litosh, V.A.L.V.A. et al. Improved nucleotide selectivity and termination of 3'-OH unblocked reversible terminators by molecular tuning of 2-nitrobenzyl alkylated HOMedU triphosphates. *Nucleic Acids Res.* **39**, 13 (2011).
5. Gaur, R.K., Sproat, B.S. & Krupp, G. Novel solid-phase synthesis of 2'-*O*-methylribonucleoside 5'-triphosphates and their alpha-thio analogs. *Tetrahedron Lett.* **33**, 3301-3304 (1992).
6. Fauster, K. et al. 2'-Azido RNA, a versatile tool for chemical biology: Synthesis, X-ray structure, siRNA applications, click labeling. *ACS Chem. Biol.* **7**, 581-589 (2012).
7. Chapuis, H., Bui, L., Bestel, I. & Barthelemy, P. 2'-Lipid-modified oligonucleotides via a 'Staudinger-Vilarrasa' reaction. *Tetrahedron Lett.* **49**, 6838-6840 (2008).

8. Mamanova, L. & Turner, D.J. Low-bias, strand-specific transcriptome Illumina sequencing by on-flowcell reverse transcription (FRT-seq). *Nat. Protoc.* **6**, 1736-1747 (2011).
9. Ju, J.Y. et al. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc. Natl. Acad. Sci. USA* **103**, 19635-19640 (2006).