Assembly of Nucleic Acids with Site-Specific Modifications

A thesis submitted to the Board of Mathematical, Physical and Life Sciences Division, Department of Chemistry in partial fulfilment for the degree of

Master of Science by Research

Ashley Tsai

Linacre College, University of Oxford

Michaelmas Term 2018



Assembly of Nucleic Acids with Site-Specific Modifications

Thesis for the degree of Master of Science by Research

Ashley Tsai, Linacre College, Michaelmas Term 2018

The ability to synthesise large nucleic acids with site-specific chemical modifications would be invaluable for applications including therapeutics, epigenetics, and nanotechnology. Current approaches mainly rely on polymerase enzymes or solid-phase chemical synthesis, both of which have their limitations. The former is restricted to enzyme-tolerated modifications and cannot achieve site-specificity beyond the level of Watson-Crick base pairing; the latter is constrained to oligonucleotide lengths of around 150 nucleotides for DNA (75 nucleotides for RNA) due to imperfect monomer coupling. To address these issues, ligation of chemically synthesised oligonucleotides is investigated for the assembly of long nucleic acids (beyond 100 nucleotides) with site-specific modifications.

The limits of enzymatic ligation are determined by optimising DNA and RNA assembly using ligases, before chemical ligation is explored as a potentially more scalable and cost-effective alternative. In particular, copper(I)-catalysed azide-alkyne cycloaddition (CuAAC) is investigated for the chemical ligation of RNA oligonucleotides. The reaction produces an unnatural triazole backbone at the ligation site. While the triazole backbone has been shown to be tolerated by DNA and RNA polymerases in replication and transcription, it has been unexplored for translation. An assay that involves rolling circle translation of a circular mRNA template, assembled by CuAAC and enzymatic ligation, is developed using a prokaryotic cell-free protein synthesis system. The effects of the local sequence environment and triazole position relative to the codon reading frame are examined, with special

attention placed on the codon wobble position. Prokaryotic ribosomes are able to read triazole-containing mRNA to produce polypeptide translation products. The yields are lower than that of the natural backbone control, possibly due to destabilised hybridisation between the tRNA and mRNA, a feature that requires further optimisation.

I, Ashley Tsai declare that this thesis and the work presented in it are my own and have been generated by me as the result of my own original research.

I confirm that:

- This work was done wholly while in candidature for a research degree at this University;

- 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

Acknowledgements

Behind this thesis is a cornucopia of silly mistakes and dead ends. Fortunately, the support, guidance, and encouragement I have received from my colleagues, friends, and family run even deeper. I am sure I cannot adequately express my gratitude to those who have made this thesis and degree possible, but an attempt shall be made.

First, I would like to thank Professor Tom Brown for inviting me to conduct this project in his laboratory. Thank you for your patience and guidance in transforming me from a clueless student to one hopefully less so over these past three years. I especially appreciate how you have encouraged me to come up with my own ideas, and seriously listened to them all. Thank you, Dr. Afaf El-Sagheer, for graciously sharing your well-earned wisdom, knowledge, and experience. Of course, thank you also for all the oligonucleotides you have tirelessly made for me. Your dedication and hard work are a constant inspiration. Dr. Arun Shivalingam, Dr. Ysobel Baker, and Pui Taemaitree – your help with immunoassays, SDS-PAGE, and RNA CuAAC ligation were invaluable. I can only hope to pass on the generosity and guidance you have so kindly shown me. Additionally, my gratitude to Prof. Brown, Dr. El-Sagheer, and Dr. Shivalingam for their patience in proofreading this thesis and helping to make it much more elegant and comprehensible. To all the Brownies, thank you for all the conversations, food, and memories we've shared throughout this journey. You are the ones who made each day in the laboratory different and enjoyable.

Thank you also to Linacre College for providing such a warm and welcoming home away from home. Outside of the laboratory, I have also learned unforgettable lessons from the wonderfully diverse and passionate members of the college. I would like to give a shout-out to my teammates in the Linacre College Boat Club and the Oxford University Water Polo Club for maintaining my physical and mental health. I will always cherish the challenges and camaraderie we shared, in and out of the water.

To the '25 Friends,' 'Market Life' gang, and other friends, I cannot imagine Oxford life without you. You all kept me going, when it felt like my experiments were not. It is such an immense privilege to count you as friends, and all the moments we've shared are held dearly, from hanging out in the CR to travelling abroad. In particular: Lynda, you've been my #1 dynamic duo partner in crime and everything else since the very beginning, and that is an understatement; Salam, your boundless generosity, encouragement, and knowledge are positively buoying, both before and during this thesis submission process; and Cam, I leave this project in your more than capable mind and hands – good luck, mate!

To my family, I would like to express my eternal love and gratitude. Dad, I strive to emulate your humorous, curious, and dedicated approach to work and life. Mom, you motivate me to keep exploring externally, in the greater world, and internally, in my

own interests. Alex, you challenge me to continue thinking creatively and to seek outside the box. A-Gong, A-Ma, Wai-Gong, Wai-Puo – so much has changed in three generations, and that is in large part to your leadership, sacrifices, and love. And Pepper, you are a constant reminder to enjoy each day to the fullest. You are all a constant source of inspiration, and this undertaking is only possible because of your support, encouragement, and guidance since Day 1!

Thank you God, for all the blessings, privileges, and opportunities that have brought me to this point.

Finally, I stand on the shoulders of giants – including but not limited to the names referenced in this thesis.

“Here, on the edge of what we know, in contact with the ocean of the unknown, shines the mystery and the beauty of the world. And it’s breathtaking.”

-Carlo Rovelli

Abbreviations

ε	Molar extinction coefficient
$A_{230}/A_{260}/A_{280}$	Absorbance at 230, 260, or 280 nm
Å	Ångström
A	Adenine
A site	Aminoacyl site
alk	Alkyne
AMP/ATP	Adenosine monophosphate/triphosphate
APS	Ammonium persulfate
az	Azide
BSA	Bovine serum albumin
c	Concentration
C	Cytosine
°C	Degree(s) Celsius
CFPS	Cell-free protein synthesis
CTP	Cytidine triphosphate
CuAAC	Copper(I)-catalysed azide-alkyne cycloaddition
Da	Dalton
DHFR	Dihydrofolate reductase
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DMT	4,4′-dimethoxytrityl [protecting group]
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
ds	Double-stranded
dT	Deoxythymidine
E. coli	*Escherichia coli*
E site	Exit site
ECL	Enhanced chemiluminescence
EDC	1-ethyl-3-[3-(dimethylamino)propyl]carbodiimide
EDTA	Ethylenediaminetetraacetic acid
EF	Elongation factor

eGFP	Enhanced green fluorescent protein
EK	Enterokinase
ELISA	Enzyme-linked immunosorbent assay
eq.	Equivalent(s)
ermCL	ermC leader [peptide]
ESI	Electrospray ionisation
Exo T	Exonuclease T
FAM	Fluorescein
Fig.	Figure(s)
fMet	N-formylmethionine
g	Gram(s)
G	Guanine
GMP/GDP/GTP	Guanosine monophosphate/diphosphate/triphosphate
h	Hypochromicity
HAA	Hexylammonium acetate [buffer]
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HFIP	Hexafluoroisopropanol
HP	High performance
hr	Hour(s)
HRP	Horseradish peroxidase
I	Inosine
IF	Initiation factor
IgG	Immunoglobulin G
K_d	Dissociation constant
l	length
L	Litre
LC	Liquid chromatography
LCR	Ligase cycling reaction
LNA	Locked nucleic acid
m	Meter(s)
M	Molar concentration
min	Minute(s)
mRNA	Messenger RNA

MS	Mass spectrometry
MW	Molecular weight
NAD^+	Nicotinamide adenine dinucleotide
nt	Nucleotide(s)
NTP	Nucleotide triphosphate
OD	Optical density
oligo(s)	Oligonucleotide(s)
ON	Overnight
p	Phosphate
P site	Peptidyl site
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCA	Polymerase cycling asssembly
PCR	Polymerase cycling reaction
PEG	Polyethylene glycol
pH	Potential of hydrogen ($-\log[H^+]$)
PNK	Polynucleotide kinase
PURE	Protein synthesis using recombinant elements
PVDF	Polyvinylidene difluoride [membrane]
RBS	Ribosome binding site
RCA	Rolling circle amplification
RCT	Rolling circle translation
RF	Release factor
RNA	Ribonucleic acid
RNase	Ribonuclease
RP	Reversed phase
RppH	RNA 5′ pyrophosphohydrolase
rpm	Revolutions per minute
rRNA	Ribosomal RNA
RT	Room temperature
rU	Ribouridine
SDS	Sodium dodecyl sulfate
ss	Single-stranded

SSB	Single-stranded binding [proteins]
T	Thymine
T7RNAP	T7 RNA polymerase
Taq	*Thermus aquaticus*
TBE	Tris/borate/EDTA [buffer]
TBDMS	*Tert*-butyldimethylsilyl [protecting group]
TC	Thiomorpholine-4-carbothioate [protecting group]
TdT	Terminal deoxynucleotidyl transferase
TEA	Triethylamine
TEAA	Triethylammonium acetate [buffer]
TEAB	Triethylammonium bicarbonate [buffer]
TEMED	Tetramethylethylenediamine
THPTA	Tris(3-hydroxypropyltriazolylmethyl)amine
TIC	Total ion chromatogram
Tm	Melting temperature
TOM	Triisopropylsilyloxymethyl [protecting group]
tRNA	Transfer RNA
U	Uracil
UP	Ultra performance
UTP	Uridine triphosphate
UTR	Untranslated region
UV	Ultraviolet
V	Volt(s)
W	Watt(s)

Amino acid codes

Amino Acid	1-letter code	3-letter code
Alanine	A	Ala
Cysteine	C	Cys
Aspartic acid	D	Asp
Glutamic acid	E	Glu
Phenylalaline	F	Phe
Glycine	G	Gly
Histidine	H	His
Isoleucine	I	Ile
Lysine	K	Lys
Leucine	L	Leu
Methionine	M	Met
Asparagine	N	Asn
Proline	P	Pro
Glutamine	Q	Gln
Arginine	R	Arg
Serine	S	Ser
Threonine	T	Thr
Valine	V	Val
Tryptophan	W	Trp
Tyrosine	Y	Tyr

Table of Contents

Abstract……...i
Declaration of Authorship……....iii
Acknowledgements……..……iv
Abbreviations……...……...…vii
Amino acid codes…….……....xi
Table of Contents……...…….…….xii
1. General introduction…….……..1
1.1 Nucleic acids……..……..2
1.1.1 Deoxyribonucleic acid (DNA)...…….……..3
1.1.2 Ribonucleic acid (RNA)…….……...…5
1.2 Polymerase chain reaction (PCR)……...……..6
1.3 *De novo* oligonucleotide synthesis……..8
1.3.1 Chemical oligonucleotide synthesis……...……...8
1.3.2 Enzymatic oligonucleotide synthesis…….…….13
1.4 Artificial gene synthesis……15
1.4.1 Chemical gene synthesis…….……15
1.4.2 Enzymatic gene synthesis…….……..16
1.5 Central dogma of molecular biology…….18
1.5.1 Replication……..19
1.5.2 Transcription……..…….21
1.5.3 Translation……..22
1.6 Gel electrophoresis……24
1.6.1 Polyacrylamide gel electrophoresis (PAGE) for nucleic acids……...25
1.6.2 Agarose gel electrophoresis for nucleic acids…….26
1.6.3 Sodium dodecyl sulfate-PAGE (SDS-PAGE) for proteins…….……27
1.7 Reversed-phase high-performance liquid chromatography (RP-HPLC)……...28
1.8 Ultraviolet (UV) absorbance for nucleic acid quantitation……...…….29
1.9 Objectives…….……32
2. Enzymatic ligation of nucleic acids……..34
2.1 Introduction and aim of the study…….…….35
2.2 Background……..……...36
2.2.1 T4 ligases……...…….36

2.2.2 *Taq* DNA ligase……38
2.2.3 T4 polynucleotide kinase (PNK)……38
2.2.4 Biotin-streptavidin magnetic beads……39
2.3 Enzymatic ligation of DNA……40
2.3.1 Templated enzymatic ligation of DNA……40
2.3.2 Magnetic bead-based ligation and purification……48
2.4 Enzymatic ligation of RNA……51
2.5 Summary and conclusions of the study……59
3. Chemical ligation of nucleic acids……62
3.1 Introduction and aim of the study……63
3.2 Background……64
3.2.l Copper(I)-catalysed azide-alkyne cycloaddition (CuAAC) reaction……64
3.2.2 Phosphoramidate ligation reaction……68
3.3 CuAAC ligation……69
3.4 Phosphoramidate ligation……76
3.5 Summary and conclusions of the study……80
4. Translation of triazole-containing mRNA……82
4.1 Introduction and aim of the study……83
4.2 Background……84
4.2.1 *E. coli* RNA 5′ pyrophosphohydrolase (RppH)……84
4.2.2 T7 RNA polymerase transcription……84
4.2.3 Cell-free protein synthesis (CFPS)……86
4.2.4 Wobble phenomenon……87
4.2.5 Immunoassays: ELISA, dot blot, western blot……89
4.3 CPFS of ermC leader peptide mRNA……93
4.4 CPFS of FLAG-HA peptide mRNA……95
4.5 CFPS of CreiLOV DNA……98
4.6 CFPS of luciferase mRNA……103
4.7 CFPS of His-tagged luciferase mRNA……107
4.8 CFPS of circular FLAGx3 mRNA……113
4.9 Summary and conclusions of the study……122
5. Conclusions and future work……127
6. Experimental……132
6.1 Reagents and buffers……133

6.2 Methods……...……134
7. References……...……..144
8. Appendix……...……155
8.1 Sequences……156
8.2 Supplementary data……..……169

Chapter 1: General introduction

1.1 Nucleic acids

1.1.1 Deoxyribonucleic acid (DNA)

Deoxyribonucleic acid (DNA) is a linear, polymeric macromolecule by which almost all known organisms store and transmit genetic information. Each monomer is called a nucleotide, and consists of a phosphate group, pentose sugar, and heterocyclic base. Without the phosphate group, the sugar and base unit is referred to as a nucleoside. The nucleoside has hydroxyl groups at the C5′ and C3′ ends of the pentose, and a phosphate group can be attached at either site to form the oligonucleotide. In DNA, the sugar is a 2′-deoxy-D-ribofuranose. There are four different monomers that make up DNA, which are differentiated by their four different bases: adenine (A), guanine (G), cytosine (C), and thymine (T). Adenine and guanine are purines, and cytosine and thymine are pyrimidines. Purines are attached to the C1′ of the pentose by their N9 atom, whereas pyrimidines are attached to the C1′ of the pentose by their N1 atom; this bond is called a glycosidic bond. The glycosidic bond usually takes the *anti*-conformation, partially due to steric effects. In natural DNA, the nucleosides are in the β-configuration, where the substituents attached to the C1′ and C5′ atoms extend from the same side of the pentose (**Figure 1-1**). Individual nucleotides are joined together linearly by phosphodiester bonds between the C3′ of one monomer and the C5′ of the next, resulting in directionality with a free 5′ end and a free 3′ end. By convention, DNA sequences are read in the 5′ → 3′ direction, unless otherwise specified. At neutral pH, each phosphate group in the DNA backbone has a charge of -1, meaning DNA molecules as a whole are negatively charged (**Figure 1-2a**).[1-3]

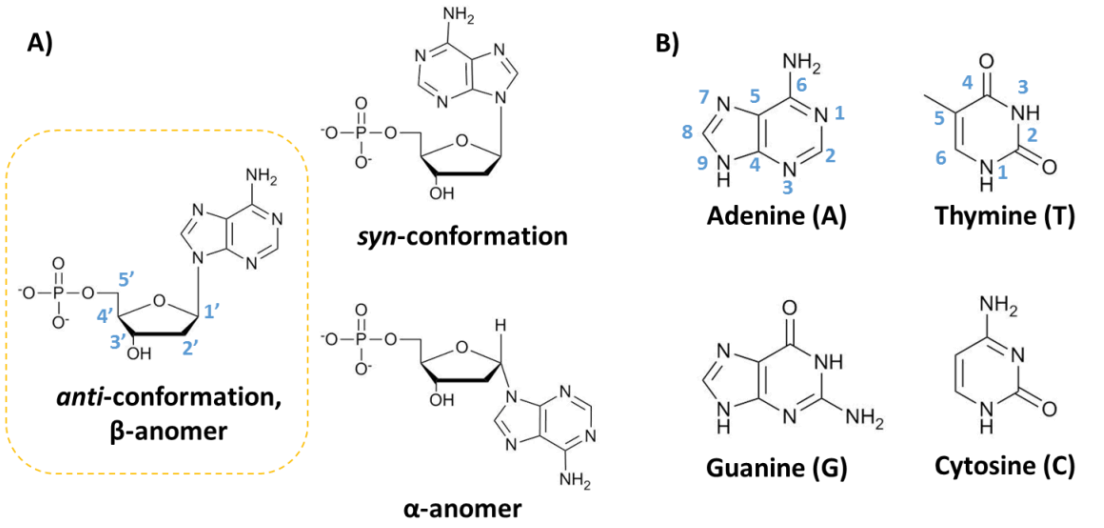


Figure 1-1. A) DNA monophosphate nucleotides usually exist in the *anti*-conformation as a β-anomer. **B)** The four DNA bases are adenine (A), guanine (G), thymine (T), and cytosine (C). A and G are purines, attached to the C1′ of the sugar by their N9 atom; T and C are pyrimidines, attached to the C1′ of the sugar by their N1 atom.

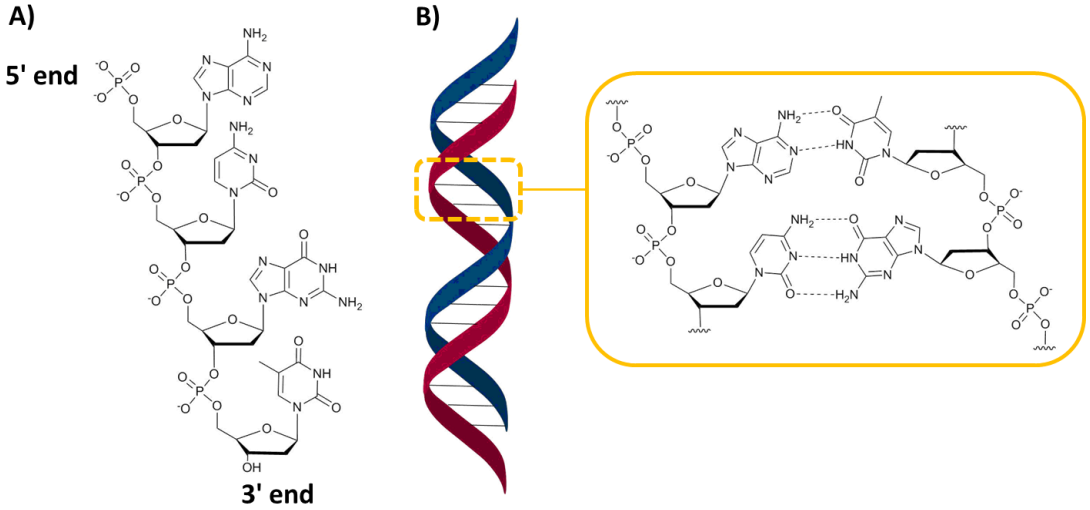


Figure 1-2. A) DNA is a linear polymeric molecule with a 5′ end and a 3′ end. Here, the sequence shown reads 5′-ACGT-3′. The monomers are linked by phosphodiester bonds. **B)** DNA forms a double helix with two antiparallel, complementary strands held together by hydrogen bonds between base pairs. A and T form base pairs with two hydrogen bonds, while C and G form base pairs with three hydrogen bonds.

In 1953, James Watson and Francis Crick proposed that DNA exists as a double helix in nature.[4] This landmark achievement provided an explanation for how

genetic information is copied and passed on. Several key pieces of information provided insights into the double helical structure of DNA, including: 1) Erwin Chargaff's observation that the molar amount of A always equalled that of T, and similarly for G and C,[5] suggesting the base pairing rules; 2) Jerry Donohue's data regarding the dominant tautomeric forms of the bases, indicating the correct hydrogen bonding between bases; and 3) Maurice Wilkins and Rosalind Franklin's X-ray diffraction photographs of DNA fibres,[6] implying its helical form with the planar bases stacked parallel to the helix axis. The Watson-Crick model of B-DNA, the conformation that DNA mainly adopts in biological conditions, has the following characteristics (**Figure 1-2b**):[7]

- The two strands in the double helix are antiparallel, with one running in the 5′ → 3′ direction and the other running in the 3′ → 5′ direction.
- The two strands are coiled around the same axis in a right-handed double helix that is approximately 20 Å in diameter, with the hydrophobic base pairs in the centre and the sugars and phosphates coiled around the perimeter. The helix has a pitch, or rise per turn, of approximately 34 Å.
- The faces of the planar bases are almost perpendicular to the helical axis, arranged in a stack. Each base forms hydrogen bonds with a base on the opposite strand according to the base pairing rules: A with T (two hydrogen bonds), and G with C (three hydrogen bonds). The distance between successive base pairs is approximately 3.4 Å, such that the ideal B-DNA has 10.5 base pairs per turn.

The resulting double helix structure has two grooves winding between the sugar-phosphate backbones. In B-DNA, asymmetries between the top and bottom edges of the base pairs and in the pentose produce a narrower minor groove and a

wider major groove. The double helix structure is stabilised by hydrogen bonding between opposite bases, and by π-π stacking between successive bases.[8] Each Watson-Crick base pair (A-T, T-A, G-C, and C-G) can be exchanged for another without perturbing the sugar-phosphate backbone.[1-3] However, any other pairs results in significant deformation of the double helix structure. By the Watson-Crick base-pairing rules, the exact sequence of one strand can be determined completely from the sequence of its complementary strand. Therefore, the separated strands of a double helix can be used as templates to produce two identical copies of the original double helix, providing a mechanism for genetic information to be passed on. As Watson and Crick wrote in 1953, "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."[4]

1.1.2 Ribonucleic acid (RNA)

Ribonucleic acid (RNA) has a very similar chemical structure to DNA, but with two major differences. First, the pentose in RNA is D-ribofuranose, with a hydroxyl group attached to C2′. This 2′-OH means the sugar conformation in RNA is slightly different than that in DNA, which in turn affects overall structure. Additionally, RNA is more chemically unstable than DNA, because the 2′-OH makes RNA susceptible to cleavage upon intramolecular cyclisation of the 2′-hydroxyl and 3′-phosphate groups. The second major difference is that thymine in DNA is replaced by uracil (U) in RNA, which also forms base pairs with adenine (**Figure 1-3**). While DNA usually exists as a double-stranded helix in cells, RNA is often single stranded, forming instead secondary structure through intramolecular base pairing. RNA plays numerous roles in biological systems, from coding for proteins to catalysing chemical

reactions. Three major types of RNA that are directly involved in protein production are messenger RNA (mRNA) which relays coding information from DNA to ribosome, ribosomal RNA (rRNA) which constitutes part of the ribosome, and transfer RNA (tRNA) which matches the appropriate amino acid to the code specified by the mRNA.[1-3]

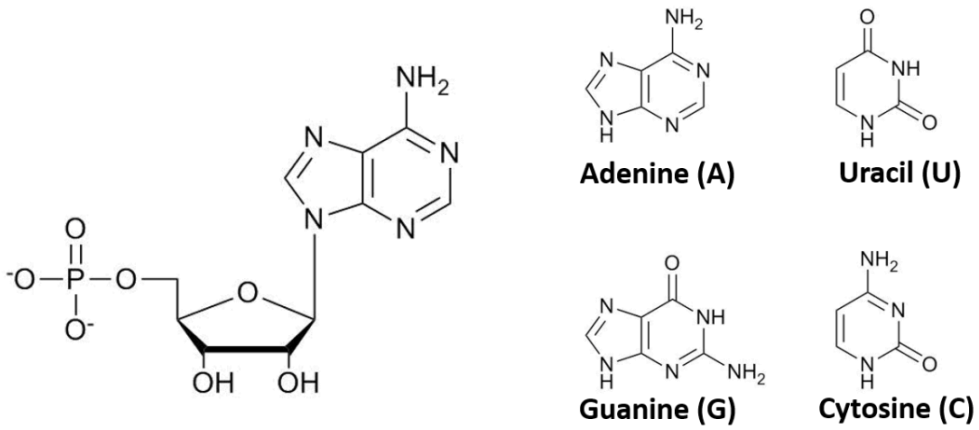


Figure 1-3. Molecular components of RNA: RNA sugars have a hydroxyl group at the C2′ position. Instead of T, uracil (U) forms base pairs with A in RNA molecules.

1.2 Polymerase chain reaction (PCR)

The abilities of DNA polymerases to make copies of a DNA template using the Watson-Crick base pairing rules are exploited in a wide array of laboratory techniques, including polymerase chain reaction (PCR). Generally, PCR relies on a heat-stable DNA polymerase - commonly derived from the bacterium *Thermus aquaticus* (*Taq*) found in hot springs - and repeated thermal cycles of denaturation, annealing, and elongation to amplify a given section of DNA. PCR reaction mixtures include deoxynucleotide triphosphates (dNTPs) and a pair of short primers that are

complementary to the 3′ ends of the target sequence to be amplified, from which the polymerase extends the copy using the free dNTPs. The denaturation step melts the DNA template into single strands at a high temperature (e.g. 95 °C). The temperature is then lowered in the annealing step to allow the primers to anneal to the template (e.g. 55 °C). Next, the temperature is changed to the optimum for the polymerase in the elongation step (e.g. 68 °C for *Taq* polymerase). DNA copies made in previous rounds act as templates in subsequent rounds, and the thermal cycles are repeated to yield exponential dsDNA copies of the template (**Figure 1-4**).[9-10] PCR is such a vital, commonly used laboratory technique that the 1993 Nobel Prize in Chemistry was jointly awarded to its inventor, Kary B. Mullis.[11] However, polymerases (and thus polymerase-based techniques like PCR) are limited in their ability to preserve site-specificity of modifications in nucleic acids, and cannot recognise certain modifications.

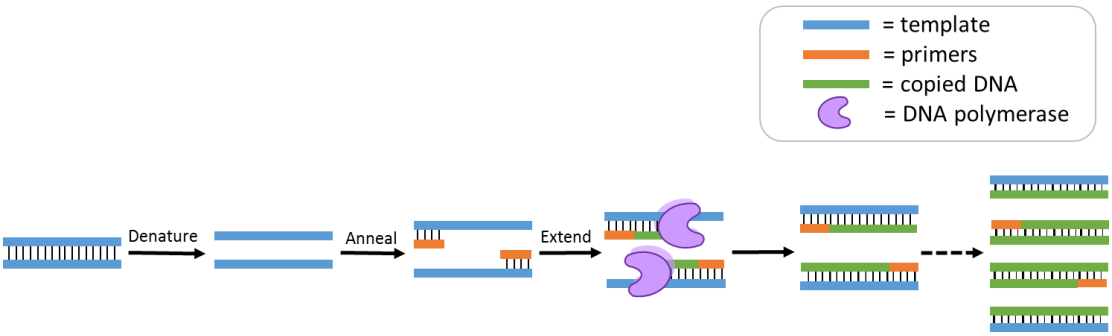


Figure 1-4. A typical PCR mixture contains the target DNA template, short complementary primers, dNTPs, and DNA polymerase in buffer. Multiple thermal cycles of denaturing, annealing, and extension result in exponential amplification of the target DNA sequence.

1.3 *De novo* oligonucleotide synthesis

1.3.1 Chemical oligonucleotide synthesis

Chemical oligonucleotide synthesis has enabled scientists to design and make *de novo* sequences with modifications at specific positions. The first report of chemical oligonucleotide synthesis was published in 1955 by A. M. Michelson and Alexander R. Todd, who described the synthesis of a TT dinucleotide.[12] Since then, three major advances have led to the development of modern oligonucleotide synthesis: Har Gobind Khorana's protecting groups,[13] which allowed for the sequential addition of nucleosides; Robert Letsinger's solid-phase approach,[14] which made the synthesis amenable to automation and greatly simplified the purification process; and Marvin Caruthers' phosphoramidite method,[15] which made possible the commercial distribution of nucleic acid monomers for oligonucleotide synthesis.

Today, synthetic DNA oligonucleotides of up to around 150 nucleotides (nt) can be made, and the vast majority of oligonucleotides are made by solid-phase synthesis with phosphoramidite monomers. Each monomer has a 2-cyanoethyl protecting group and a diisopropylamino group on the 3′-phosphite moiety, and a 4,4′-dimethoxytrityl (DMT) protecting group on the 5′-hydroxyl moiety. Certain bases also have protecting groups on their exocyclic amino groups (**Figure 1-5**).

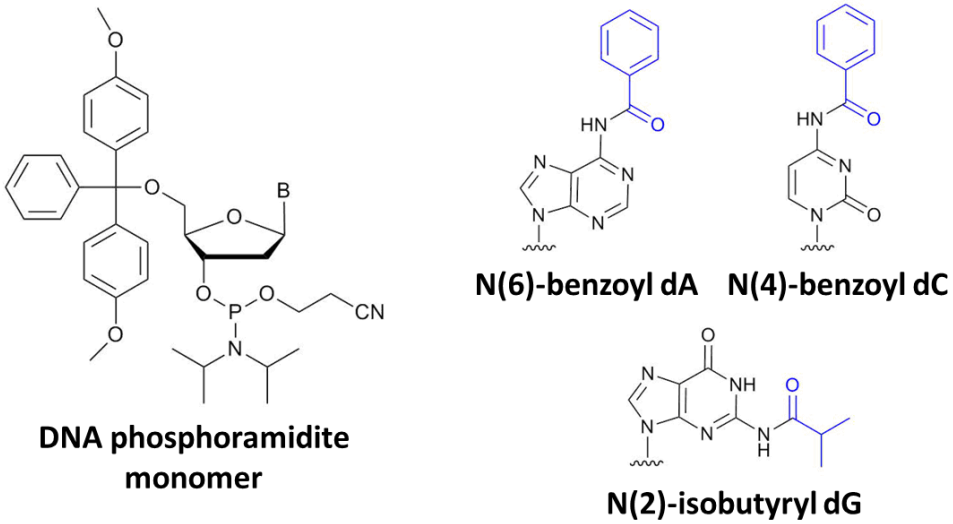


Figure 1-5. Protected DNA phosphoramidite monomer and bases for solid-phase synthesis. B signifies any of the four DNA bases. dT monomers do not require protecting groups on the base.

An advantage of solid-phase synthesis is the ability to drive reactions to completion with large excesses of reagents, and then wash away impurities and excess reagents after each step. Phosphoramidite synthesis occurs in the 3′ → 5′ direction. The 3′ terminal nucleoside is first attached to the solid support, commonly controlled pore glass or polystyrene beads. Synthesis then proceeds through cycles that extend the oligonucleotide one residue at a time. In the first step, detritylation, the DMT group protecting the 5′-hydroxyl of the growing oligonucleotide is removed with an acid such as trichloroacetic acid and washed away. The second step is activation and coupling, in which the next residue is added in excess as a nucleoside phosphoramidite. An activator such as tetrazole, which is a weak acid, protonates the diisopropylamino group of the incoming phosphoramidite monomer, making it a good leaving group. The newly deprotected 5′-hydroxyl group of the bound oligonucleotide then attacks the 3′-phosphorus atom of the monomer, forming a phosphite triester linkage. The diisopropylamino group of the monomer leaves and is washed away, along with excess monomers and reagents. While the coupling reaction

is generally very high-yielding, there are inevitably some 5′-hydroxyl groups that fail to couple. These need to be blocked to prevent further elongation, which would result in oligonucleotides that contain base deletions. Capping happens as the third step, in which a mixture of acetic anhydride and *N*-methylimidazole acetylates any free 5′-hydroxyl groups and blocks them from further reactions. Oxidation, the fourth step, converts the phosphorus atom of the newly formed phosphite triester linkage from an unstable P(III) to a stable P(V) form with iodine, pyridine, and water. The cycle repeats for each addition of a monomer until the whole oligonucleotide sequence is synthesised in the 3′ → 5′ direction. Finally, the oligonucleotide can be cleaved from the solid support with concentrated aqueous ammonia. Heating the oligonucleotide in ammonia at 55 °C also removes the cyanoethyl protecting groups from the phosphates, and the protecting groups (e.g. benzoyl and isobutyryl groups) from the exocyclic primary amino groups on A, C, and G (**Figure 1-6**).[3, 16-18]

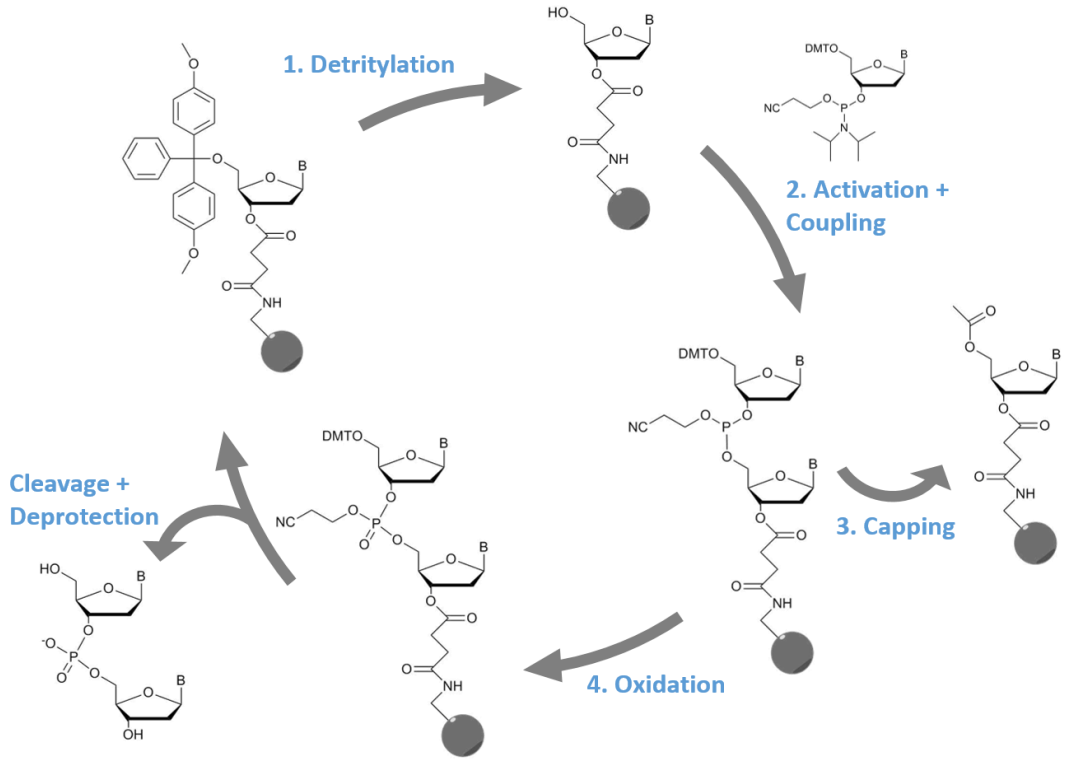


Figure 1-6. Schematic of solid-phase oligonucleotide synthesis. Each addition of a monomer occurs in four steps, and the cycle is repeated until the whole sequence has been synthesised. **1. Detritylation:** first, the DMT group protecting the 5′-hydroxyl group of the growing nucleic acid is removed. **2. Activation and coupling:** the deprotected 5′-hydroxyl group then attacks the 3′-phosphorus atom of the incoming monomer, which has been activated by the protonation of its diisopropylamino group. **3. Capping:** next, any unreacted and deprotected 5′-hydroxyl groups are capped to prevent deletions in the sequence. **4. Oxidation:** finally, the phosphorus atom of the newly added nucleotide is converted from an unstable P(III) to a stable P(V) form. **Steps 1-4** are repeated for the addition of each monomer. **Cleavage and deprotection:** when the sequence is complete, the oligonucleotide is cleaved from the solid support and the protecting groups are removed.

RNA chemical synthesis is similar to that of DNA, although the practical length limit is shorter at approximately 75 nt. The main difference is that the 2′-hydroxyl group of RNA must be protected to prevent 5′ to 2′ coupling of monomers, and to prevent 3′ to 2′ backbone migration or chain cleavage. The lower efficiency of RNA synthesis relative to that of DNA synthesis is in part due to the steric bulk of the 2′-hydroxyl protecting group, which decreases coupling efficiencies. Additionally, the 2′-hydroxyl protecting group has a degree of instability in the basic conditions used to

remove the protecting groups on the bases, leading to potential chain cleavage or 3′ to 2′ phosphate migration. Other differences may include the protecting groups used, such as *tert*-butylphenoxyacetyl protecting groups on the bases, and the capping reagent, such as *tert*-butylphenoxyacetic anhydride, depending on the exact 2′ protecting group on the monomers. One of the most commonly used 2′ protecting groups is *tert*-butyldimethylsilyl (TBDMS) (**Figure 1-7a**), which is removed with a fluoride source such as triethylamine trihydrofluoride.[19] Alternative 2′ protecting groups include triisopropylsilyloxymethyl (TOM), removed with tetra-*n*-butylammonium fluoride in tetrahydrofuran, and thiomorpholine-4-carbothioate (TC), removed with ethylenediamine in toluene (**Figure 1-7b**).[20] Advantages to TC-protected RNA monomers compared to TBDMS monomers include the ability to remove all protecting groups simultaneously, and enhanced coupling due to smaller protecting groups on the bases. The choice of which 2′ protecting group to use also depends on whether their removal methods are compatible with chemical modifications in the sequence.

After synthesis and deprotection, oligonucleotides can be purified by high-performance liquid chromatography or gel electrophoresis to remove truncated strands and incompletely deprotected molecules.[3, 21]

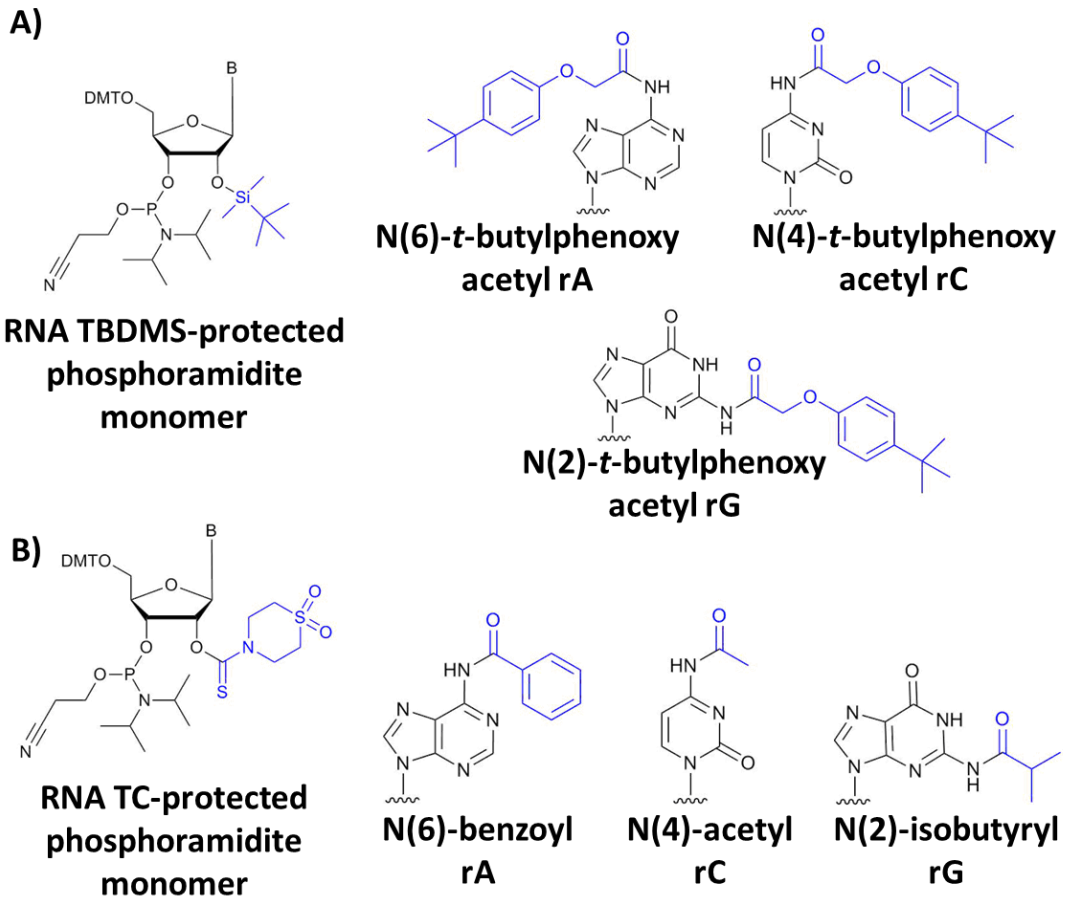


Figure 1-7. Protected RNA phosphoramidite monomers and bases for solid-phase synthesis. B signifies any of the four RNA bases. **A)** TBDMS-protected RNA monomers with corresponding protected bases. **B)** TC-protected RNA monomers with corresponding protected bases. rU monomers do not require protecting groups on the base.

1.3.2 Enzymatic oligonucleotide synthesis

Template-independent enzymatic synthesis of oligonucleotides has been explored in parallel to chemical synthesis, but none of the methods have been widely adopted yet due to issues with coupling yields, coupling times, and oligonucleotide purity. The general mechanism involving sequential cycles of additions, deprotections, and washes is similar to that of chemical oligonucleotide synthesis. For example, Carole Schmitz and Manfred Reetz published in 1999 a method for solid-phase synthesis of oligonucleotides using T4 RNA ligase, with coupling times of 48 hours (for coupling to a ribocytidine) up to 144 hours (for coupling to a

deoxyriboadenine).[22] Despite the current dominance of chemical solid-phase synthesis, interest still remains in enzymatic oligonucleotide synthesis for its potential advantages, such as the following: enzymatic specificity, mild aqueous reaction conditions, longer potential length limits, fewer protecting groups, and high-throughput optimisation via enzyme engineering.

Very recently in July 2018, Sebastian Palluk *et al.* at the University of California, Berkeley, published a method for enzymatic oligonucleotide synthesis using terminal deoxynucleotidyl transferase (TdT), a template-independent polymerase that adds dNTPs to the 3′ end of ssDNA. TdT has been used in previous enzymatic synthesis methods,[23] but their creative approach was to conjugate a single dNTP to the TdT using a cleavable linker. This provided the double advantages of having an extremely high effective concentration of dNTP at the catalytic site, and blocking the 3′ end of the growing oligonucleotide with the TdT itself to prevent further additions. Starting with a DNA primer, each monomer addition occurred in a cycle of two steps. First, a TdT-dNTP conjugate with the desired base was added, extending the primer by one base. Next, the linker between the TdT and dNTP was cleaved and the TdT was inactivated or removed, deprotecting the 3′ end of the oligonucleotide and leaving it ready for the next addition. Cleavage of the linker they used left a propargylamino scar on each base, but *Taq* polymerase was still able to read and replicate the synthesised oligonucleotide. Palluk *et al.* reported coupling times of 1.5-3 minutes, deprotection times of 1 minute, and average stepwise yields of 97.7% for a 10-mer oligonucleotide.[24-25]

1.4 Artificial gene synthesis

1.4.1 Chemical gene synthesis

Chemically synthesised oligonucleotides can be ligated via chemical reactions to produce artificial genes without using enzymes. Advantages of this approach include the abilities to incorporate a wide variety of modifications at specific positions, perform the assembly at a large scale, and use orthogonal ligation reactions simultaneously. In 1991, Zoe Shabarova *et al*. reported the "first non-enzymatic assembly of a biologically active gene" by condensation reactions using cyanogen bromide. The gene was assembled from four oligonucleotides that were 35-53 nt in length. The final 186 nt DNA product contained a promoter, Shine-Dalgarno sequence (also known as the ribosome binding site), an 18 codon-long coding region beginning with an ATG start codon and ending with a TAG stop codon, and a transcription terminator. After enzymatic insertion into a plasmid, the chemically assembled gene was shown to be successfully transcribed. Assembly of the 186 nt gene with the reported chemical ligations (few minutes) was found to be much faster than with T4 DNA ligase (12-14 hours).[26]

More recently, one-pot chemical assemblies of genes several hundred nt in length have been published. Mikiembo Kukwikila *et al.* used copper(I)-catalysed azide-alkyne cycloaddition (CuAAC) to assemble the 335 nt gene for the iLOV fluorescent protein in 2017. Ten self-templating DNA oligonucleotides were used to assemble the sense and antisense strands simultaneously, each of them 50-80 nt long. The oligonucleotides were combined, annealed, and ligated in the same tube, and the fully ligated product was purified by denaturing polyacrylamide gel to give approximately 4.5% final yield. CuAAC produces an unnatural triazole linker at the ligation site, but DNA polymerases were able to replicate the iLOV gene with its four

triazole backbones per strand in PCR and primer extension assays. Furthermore, the CuAAC-ligated gene produced functional iLOV protein when inserted into a plasmid and transformed into *Escherichia coli* (*E. coli*), verifying that it was biocompatible. Versions of the gene with epigenetic modifications and fluorophores at specified locations were also assembled to demonstrate the versatility of the approach.[27] Using the same CuAAC ligation chemistry, Ilenia Manuguerra *et al.* assembled the 762 nt enhanced green fluorescent protein (eGFP) gene from 14 oligonucleotides. In this case, a DNA nanostructure was employed to bring the 3′ and 5′ ends of the oligonucleotides together to improve ligation efficiencies relative to the equivalent unstructured reactions.[28] Afaf El-Sagheer and Tom Brown also showed in 2017 that the 762 nt eGFP gene could be assembled using one-pot phosphoramidate ligation, another chemical reaction that produces an unnatural backbone that was tolerated by DNA and RNA polymerases in *in vitro* replication and transcription, respectively. Furthermore, they demonstrated that CuAAC and phosphoramidate ligation are orthogonal reactions that can occur simultaneously by assembling in one tube a 303-mer DNA with both a triazole and phosphoramidate ligation site.[29]

1.4.2 Enzymatic gene synthesis

Most enzymatic gene synthesis methods utilise ligases, polymerases, or a combination of the two. After the gene is assembled, a final PCR step is often used to isolate and amplify the complete gene, sometimes followed by a cloning step to select for the correct sequence. Presently, enzymatic gene synthesis far surpasses chemical gene synthesis in terms of the length that can be achieved, but one disadvantage is that polymerases generally cannot preserve site-specificity of modifications.[30-33] A comprehensive overview of current enzymatic gene synthesis methods is outside the

scope of this text, but recent reviews can be found in literature. Representative methods that are commonly used are briefly described here.[34-36]

Polymerase cycling assembly (PCA) has similar reaction conditions and components to PCR, but uses multiple overlapping oligonucleotides instead of a pair of primers. The oligonucleotides are fragments of both strands of a dsDNA construct, each with overlapping complementary regions of 20-30 nt. When the overlapping regions are annealed, the polymerase can 'fill in the gaps' by extending in the 5′ → 3′ direction. This extended piece of DNA can then act as a template in further rounds of denaturation, annealing, and extension. Eventually, the whole gene is assembled and can be amplified by PCR with primers complementary to both ends of the target sequence.[37] Ligase chain reaction is similar to PCA, but features a heat-stable ligase that joins oligonucleotides instead of a polymerase that adds dNTPs. The oligonucleotides are fragments of the target sequence without any gaps. Bridging, complementary splints (or oligonucleotides comprising the complementary strand) ensure the correct assembly order. Templated assembly of DNA occurs in repeated rounds of denaturation, annealing, and ligation. Successfully ligated DNA serve as templates in subsequent rounds.[38, 39] Combining features of PCA and LCR, Gibson assembly is an isothermal DNA assembly method that takes place at 50 °C, involving exonucleases, polymerases, and ligases. The exonuclease, which is not heat stable, first digests the 5′ ends of dsDNA to create ssDNA overhangs, and denatures soon after incubation at 50 °C. When complementary overhangs anneal, thermostable DNA polymerase fills in the gaps, and thermostable ligase then seals the nicks to produce the final assembled dsDNA product.[40, 41]

1.5 Central dogma of molecular biology

The central dogma of molecular biology elucidates the flow of genetic information through DNA, RNA, and proteins that underpins all known life. Francis Crick first described the phrase in 1958,[42] and expanded further on it in a paper published in *Nature* in 1970. He wrote, "The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred from protein to either protein or nucleic acid."[43] The central dogma is often conflated with the sequence hypothesis, also put forth by Crick in 1958 - "In its simplest form [the sequence hypothesis] assumes that the specificity of a piece of nucleic acid is expressed solely by the sequence of its bases, and that this sequence is a (simple) code for the amino acid sequence of a particular protein." In Crick's words, the difference was that "the sequence hypothesis was a positive statement, saying that the (overall) transfer nucleic acid → protein did exist, whereas the central dogma was a negative statement, saying that transfers from protein did not exist."[42] Building on these two frameworks, Crick classified all nine possible information transfers between DNA, RNA, and protein into three categories: general transfers, which occur in almost all cells; special transfers, which *may* occur in exceptional cases; and unknown transfers, which are postulated never to occur.[43] Of the three general transfers, replication, transcription, and translation are the processes by which DNA → DNA, DNA → RNA, and RNA → protein information transfers occur, respectively (**Figure 1-8**).

General Transfers	Special Transfers	Unknown Transfers
DNA → DNA	RNA → RNA	Protein → Protein
DNA → RNA	RNA → DNA	Protein → DNA
RNA → Protein	DNA → Protein	Protein → RNA



Figure 1-8. Top, Crick's categories of information transfer; bottom, currently known information transfers. The general transfers are shown in solid black arrows.

1.5.1 Replication

Before a cell divides, its DNA is copied in a process called replication. A group of proteins called the replisome separates the DNA double helix into two single strands, and a complementary copy of each strand is synthesised to yield two exact copies of the original dsDNA. Prokaryotic replisomes include helicase, gyrase, single-stranded binding (SSB) proteins, primase, DNA polymerase III, DNA polymerase I, and DNA ligase. Helicase is an enzyme that moves along a strand of DNA and separates the double helix by breaking the hydrogen bonds between the base pairs. Unwinding the duplex creates supercoils in the DNA, which is relaxed by the enzyme gyrase moving ahead of the helicase. To prevent the separated single strands from re- or self-annealing, SSB proteins bind to the ssDNA strands and stabilise them. DNA polymerase III can then read the ssDNA as templates, and catalyse the synthesis of a complementary DNA strand from free dNTPs. However, DNA polymerases generally cannot initiate synthesis, and require an existing 3′-hydroxyl group in order to add a dNTP. Primase thus initiates DNA synthesis by making a complementary RNA (or DNA) primer from which DNA polymerases can

elongate their DNA chain. After elongation, DNA polymerase I removes the RNA primers and fills in the gaps with DNA.

All known DNA polymerases can only extend in the 5′ → 3′ direction. This means that while one strand of the double helix can be copied continuously, the other strand must be copied in short bursts. The leading strand is synthesised continuously in the same direction in which the helicase is unwinding the double helix, but the lagging strand waits for sections of ssDNA template to become available. It is then synthesised discontinuously, also in the 5′ → 3′ direction, in pieces called Okazaki fragments. These Okazaki fragments are eventually covalently joined by DNA ligase (**Figure 1-9**).[1-3, 44-46]

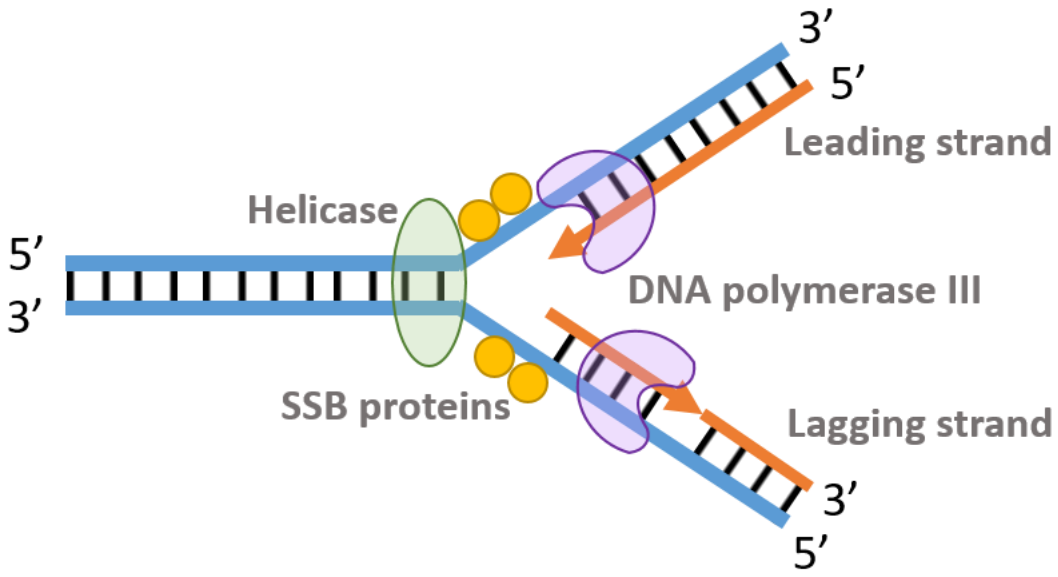


Figure 1-9. A schematic showing a prokaryotic replication fork, with the leading strand synthesised continuously and the lagging strand synthesised in Okazaki fragments. Helicase unwinds the DNA duplex, SSB proteins stabilise the ssDNA strands, and DNA polymerase III synthesises a complementary DNA strand from free dNTPs.

1.5.2 Transcription

Transcription is the process in which DNA is copied to RNA by RNA polymerases. In cells, DNA exists in a double helix form, and only one of the strands is used as a template during transcription. This DNA strand is called the antisense strand. RNA polymerase reads the antisense strand in the 3′ → 5′ direction to synthesise a complementary RNA molecule that is extended in the 5′ → 3′ direction. The resulting RNA has the same sequence as the DNA sense strand, which is also known as the coding strand and is complementary to the antisense strand, except that thymine (T) is replaced with uracil (U).

Transcription begins with initiation, in which RNA polymerase, along with transcription factors, bind to a specific sequence in the DNA double helix that is called a promoter. The enzyme then unwinds a short region of the DNA to create a single-stranded transcription bubble. Using the antisense strand as a template, RNA polymerase catalyses bond formation between nucleoside triphosphates (NTPs) that are complementary to the first two bases of the transcription start site. The RNA polymerase continues to read the antisense template strand, elongating the RNA molecule in the 5′ → 3′ direction by adding one complementary NTP at a time. Elongation continues until the termination step (**Figure 1-10**). Bacteria have two main methods of transcription termination, rho-independent and rho-dependent. In rho-independent termination, transcription is stopped by the formation of a hairpin structure in the new RNA transcript. Rho-dependent termination occurs when the protein factor Rho releases the RNA by destabilising interactions between the transcript and the DNA template.[1-3, 47-49]

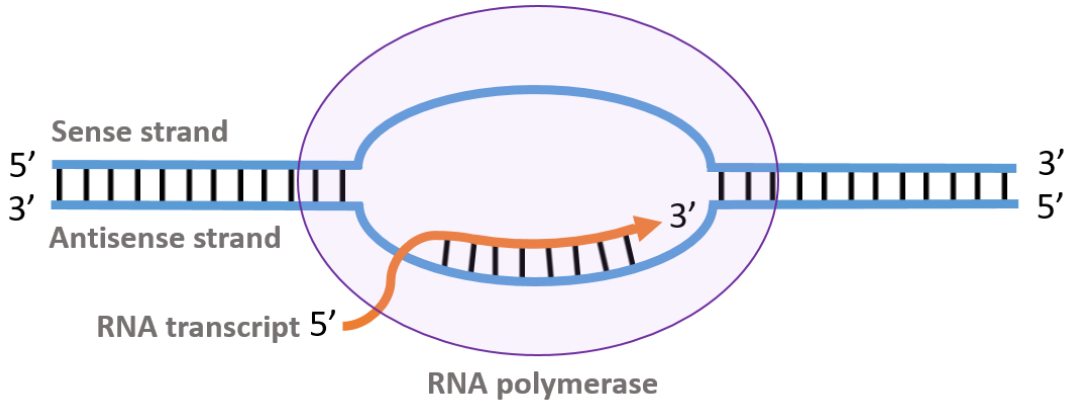


Figure 1-10. A schematic showing the transcription bubble. The RNA transcript (orange) is synthesised by RNA polymerase in the 5′ → 3′ direction. The resulting RNA molecule is complementary to the DNA antisense strand, and has the same sequence as the sense strand (with U instead of T).

1.5.3 Translation

Translation consists of three main steps: initiation (the slowest step), elongation, and termination. In prokaryotes such as *E. coli*, translation initiation requires the assembly of a complex involving the 30S and 50S ribosomal subunits, mRNA template, tRNA charged with N-formylmethionine (fMet), initiation factors IF-1, IF-2, and IF-3, and GTP for energy. First, the three IFs bind to the 30S small ribosomal subunit. The 30S subunit is then positioned on the mRNA by binding to the Shine-Dalgarno sequence, AGGAGG (or similar). The Shine-Dalgarno sequence, also known as the ribosome binding site (RBS), is complementary to a region of the rRNA in the 30S subunit and is usually located 5-8 nt upstream of the AUG start codon. Next, IF-2 recognises the fMet tRNA, which contains the anticodon complementary to the start codon. Binding of the fMet tRNA triggers the release of IF-3, which in turn leads to binding of the 50S large ribosomal subunit to form the complete ribosomal complex. The GTP bound to IF-2 is then hydrolysed to GDP, and both IF-1 and IF-2 are released. The resulting complex is composed of the ribosomal subunits, fMet tRNA, and mRNA, and can proceed to the next step of elongation.

The ribosomal complex has three tRNA sites called the aminoacyl (A) site, peptidyl (P) site, and exit (E) site. Upon initiation, the fMet tRNA is bound at the P site. Elongation begins when a charged tRNA complementary to the second codon of the mRNA enters the A site, arriving with elongation factor (EF) Tu. The GTP that is bound to EF-Tu has to be hydrolysed to GDP before EF-Tu can be released from the ribosome, which is the rate limiting step in elongation. This step is important for proofreading, since it provides time for an incorrect tRNA, which would have weaker binding to the mRNA, to dissociate before its amino acid is incorporated into the growing polypeptide chain. Once EF-Tu is released, the ribosome catalyses formation of a peptide bond between the fMet amino acid of the tRNA in the P site and the new amino acid of the tRNA in the A site. In the process, the growing polypeptide chain is transferred to the tRNA in the A site, leaving the tRNA in the P site uncharged. Elongation proceeds as the ribosome moves along the mRNA in the 5′ → 3′ direction in single-codon steps (each codon consisting of 3 RNA nucleotides). Each step, or translocation, moves the uncharged tRNA from the P to E site, and the tRNA with the polypeptide chain from the A to P site. A new, charged tRNA complementary to the next codon can enter the now empty A site, which causes the release of the uncharged tRNA from the E site. This elongation cycle continues until the ribosome reaches a stop codon on the mRNA (**Figure 1-11**).

Cells do not contain tRNAs complementary to the termination codons UAA, UAG, or UGA. Instead, when the ribosome reaches one of these stop codons, a release factor (RF) enters the A site and binds to the codon. RF-1 recognises UAA and UAG, while RF-2 recognises UAA and UGA. Binding of the RF triggers hydrolysis of the bond between the polypeptide chain and the tRNA in the P site, releasing the translation product and dissociating the ribosome complex.[1-3, 50-54]

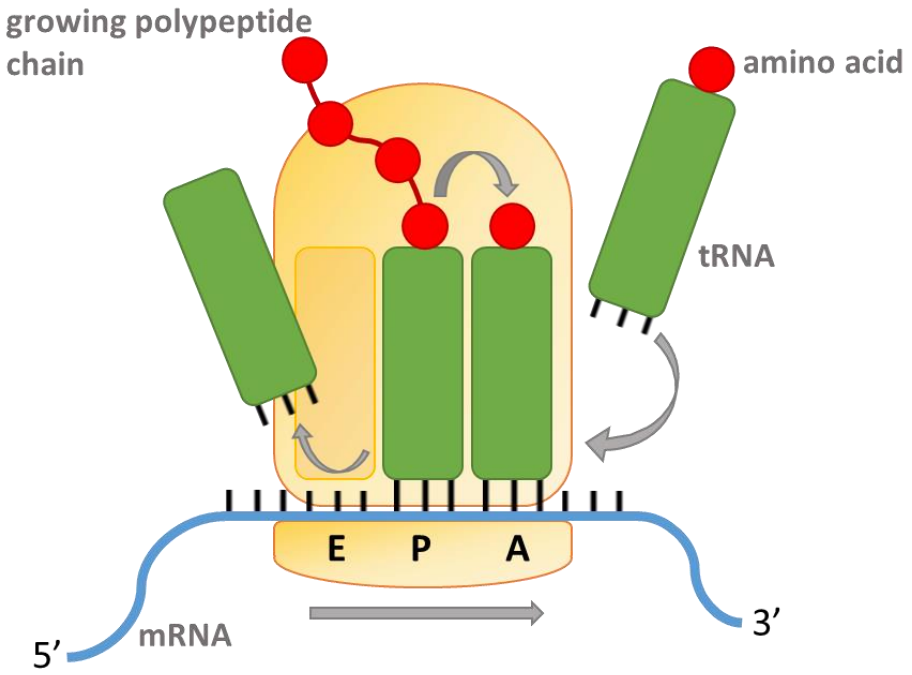


Figure 1-11. A schematic showing the translation complex with the ribosomal subunits, mRNA template, and complementary tRNAs. As the ribosome reads the mRNA in the 5′ → 3′ direction, the polypeptide chain is transferred from the tRNA in the P site to the tRNA in the A site. The now uncharged tRNA in the P site is moved to the E site, and the tRNA with the growing polypeptide chain is moved to the P site. A new, charged tRNA can enter in the now empty A site, triggering the release of the uncharged tRNA from the E site.

1.6 Gel electrophoresis

Gel electrophoresis is a method that separates nucleic acids and proteins based on their size, structure, and charge. It is a useful tool for analysing reactions, as well as for purifying samples by extracting bands of interest from the gel. The samples are loaded in lanes on one end of the gel, and an electrical field is applied so that the charged molecules migrate to the other end. Generally, smaller molecules move through the porous gel material faster than larger ones, so the size of the analytes can be estimated by comparing their positions relative to that of markers of known size (**Figure 1-12**). Protein and nucleic acid gels can be visualised by using stains or labels bound to the analytes, and nucleic acid gels can also be visualised by

ultraviolet (UV) absorbance. Two of the most commonly used gel materials are polyacrylamide and agarose.[55-57]

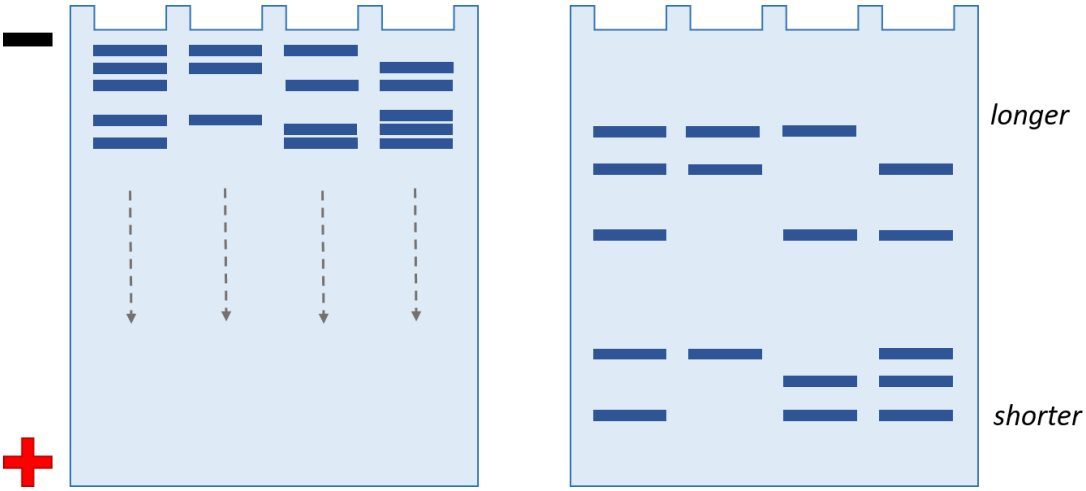


Figure 1-12. In gel electrophoresis, an electrical field is applied to a gel composed of porous material. Sample components are thus separated by size, structure, and charge.

1.7.1 Polyacrylamide gel electrophoresis (PAGE) for nucleic acids

Polyacrylamide gels are made by polymerising a mixture of acrylamide and *N,N′*-methylenebisacrylamide (**Figure 1-13b**). The latter component is a cross-linking agent that facilitates the formation of a gel matrix, without which the acrylamide would polymerise as linear chains. The reaction is a free radical polymerisation, usually initiated with ammonium persulfate (APS) and tetramethylethylenediamine (TEMED) (**Figure 1-13a**). Gel pore size can be tuned by varying the concentration of acrylamide, with a higher concentration resulting in smaller average pores. Higher acrylamide percentage gels are thus better suited for separating shorter oligonucleotides. Typical acrylamide percentages range from 4-20%, and can be used to separate nucleic acids on the order of 1-1,000 nt in length.

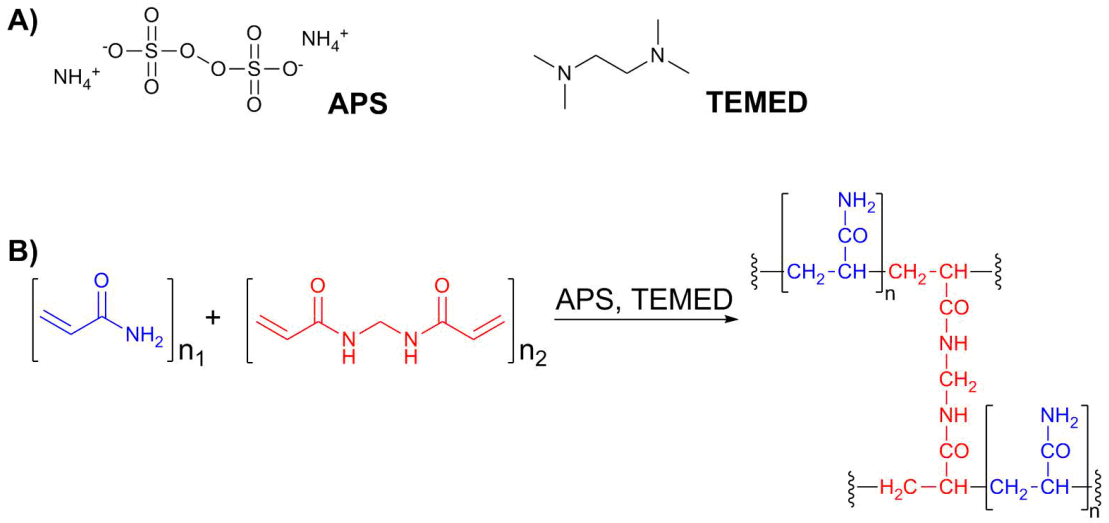


Figure 1-13. **A)** Chemical structures of ammonium persulfate (APS) and tetramethylethylenediamine (TEMED). APS and TEMED are used to initiate free radical polymerisation of acrylamide. **B)** Acrylamide (blue) and *N,N′*-methylenebisacrylamide (red) are polymerised to make PAGE gels. The latter is a cross-linking agent.

Nucleic acids have one negative charge per nucleotide backbone at neutral pH, so they migrate from the cathode towards the anode of the gel. Electrophoretic mobility is affected by not just nucleic acid length and charge, which have a constant linear relationship in nucleic acids, but also structure. Native polyacrylamide gel electrophoresis (PAGE) therefore provides information on both size and structure. In denaturing PAGE, structure is disrupted by polymerising the gel with 7 M urea and loading the samples with formamide. These components eliminate nucleic acid structure by blocking hydrogen bonding between bases, so migration in denaturing PAGE gels depends only on linear length. PAGE can produce sharp bands that provide single-nucleotide resolution.[55-59]

1.6.2 Agarose gel electrophoresis for nucleic acids

Agarose gels are composed of polysaccharide polymers derived from seaweed (**Figure 1-14**). Upon heating the polymers in buffer to dissolve them and then cooling

the solution down, a gel matrix is formed by hydrogen bonding. As with polyacrylamide gels, agarose gel pore size can be tuned by varying the agarose percentage. Typical agarose percentages range from 0.7-2%, and are suitable for nucleic acids that are 100-10,000 nt long. Advantages to agarose gels include nontoxic gel material and the ability to melt the gel for fast nucleic acid extraction.[56, 57, 60, 61]

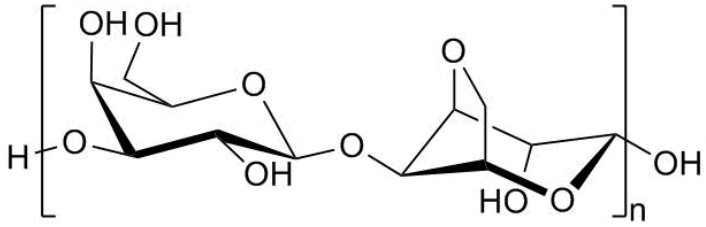


Figure 1-14. Chemical structure of agarose, a linear polymer. Agarose can be dissolved by heating in aqueous solution; upon cooling, it forms a solid three-dimensional gel via hydrogen bonding.

1.6.3 Sodium dodecyl sulfate-PAGE (SDS-PAGE) for proteins

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) is a variant of PAGE for protein separation. Unlike nucleic acids that have a constant negative charge per nucleotide at neutral pH, proteins are composed of amino acids that have a positive charge, a negative charge, or none at all. To address this variance, SDS is used as a negatively charged surfactant to cover the proteins' intrinsic charge and give them a similar charge-to-mass ratio (**Figure 1-15**). SDS is used in both the gel and the loading buffer for the samples. Protein samples are first denatured by boiling for 5-10 minutes in the loading buffer. The buffer contains SDS, which binds to the proteins' backbones and prevents secondary and tertiary structure formation, and a reducing agent like β-mercaptoethanol, which cleaves disulfide bridges. SDS-PAGE gels are usually cast in a discontinuous format consisting of a stacking gel

with wells on the top and a resolving gel on the bottom. The stacking gel usually has 4-6% acrylamide and is near neutral pH, whereas the resolving gel usually has 10-20% acrylamide and is at a slightly basic pH. The proteins in the sample concentrate at the interface between the stacking and resolving gels before entering the latter, resulting in sharper bands. SDS-PAGE gels can also be cast in a gradient format. A gradient of acrylamide percentage enables analysis over a wide range of molecular mass on the same gel. Proteins separated by SDS-PAGE can be visualised with a stain or label bound to the sample, or transferred to a membrane for western blot analysis (see Chapter 4.2.5).[58, 62, 63]

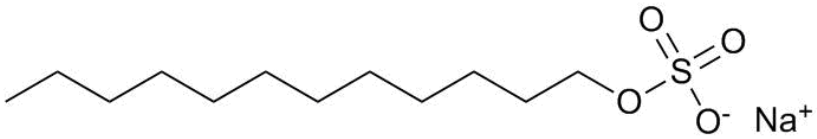


Figure 1-15. Chemical structure of sodium dodecyl sulfate (SDS), an anionic surfactant used to denature proteins and impart a similar charge-to-mass ratio in SDS-PAGE.

1.7 Reversed-phase high-performance liquid chromatography (RP-HPLC)

High-performance liquid chromatography (HPLC) is a purification method that consists of a pressurized liquid solvent (the mobile phase) passing through a column packed with a solid material (the stationary phase). The mobile phase is usually pumped through the column with a gradient of two buffers (buffers A and B). The sample is injected into the column and the various components of the sample interact differently with the stationary phase, leading to different elution times over the buffer gradient. In reversed phase HPLC (RP-HPLC), the stationary phase is made of a nonpolar material while buffer A in the mobile phase is a polar solvent

(e.g. buffered water). Buffer B, on the other hand, contains less polar solvents (e.g. acetonitrile or methanol). As the buffer gradient from A to B is increased, the solvent polarity of the mobile phase is decreased. This results in greater competition of the solvent for binding to the stationary phase compared to the sample. Therefore, more hydrophobic components in the sample have a longer retention time. RP-HPLC is often used for oligonucleotide purification, because oligonucleotides with unremoved protecting groups or chemical modifications differ in their hydrophobicity. A typical oligonucleotide RP-HPLC column contains hydrocarbon chains bound to a silica support as the stationary phase. Nucleic acid elution can be monitored with a UV detector, and the area under absorbance peaks can be compared to determine relative quantities. Commonly used buffers for oligonucleotide purification include ammonium acetate, hexylammonium acetate (HAA), triethylammonium acetate (TEAA), and triethylammonium bicarbonate (TEAB), with a gradient of acetonitrile.[3, 55, 64, 65]

1.8 Ultraviolet (UV) absorbance for nucleic acid quantitation

Oligonucleotide concentrations can be estimated in solution by measuring ultraviolet (UV) absorbance with a spectrophotometer. Absorption is usually measured at a wavelength of 260 nm (A_{260}), since the heterocyclic rings of DNA and RNA nucleotides have an absorption maximum around this wavelength. In a typical A_{260} measurement, the oligonucleotide solution is placed in a cuvette and exposed to UV light of 260 nm on one side. The nucleic acids absorb some of the light, and a detector on the opposite side of the cuvette measures how much light is transmitted. A_{260} is also commonly referred to as the optical density (OD) of the solution, which is

the logarithm of the ratio of the intensity of incident light (I_0) to the intensity of transmitted light (I):

$$OD = log\frac{I_0}{I}$$

The amount of absorbed light is related to the oligonucleotide concentration by the Beer-Lambert Law, which states that the absorbance of a sample is equal to the product of the extinction coefficient (ε), the concentration (c), and the length of the cuvette (l):

$$A = \varepsilon cl$$

Although accurate values for the extinction coefficient must be determined experimentally for each oligonucleotide sequence, there are two main methods for providing good estimates. In the base composition method, ε is given as the sum of the molar extinction coefficients of each individual nucleoside. The sum is then multiplied by a factor of 0.9 to account for the effect of base stacking, which decreases ε_{260} of the polymerised bases relative to free ones. The following equations give the base composition extinction coefficients for DNA and RNA in units of M^{-1} cm^{-1}. An advantage of the base composition method is that it is fast and easy to calculate, and modifications can be easily accommodated by simply adding their individual extinction coefficients.

$$DNA: \varepsilon_{260} = 0.9[nA(15{,}400) + nC(7{,}400) + nG(11{,}500) + nT(8{,}700)]$$

$$= 0.9\sum_{1}^{N} \varepsilon_{individual\ bases}$$

$$RNA: \varepsilon_{260} = 0.9[nA(15{,}400) + nC(7{,}200) + nG(11{,}500) + nU(9{,}900)]$$

$$= 0.9\sum_{1}^{N} \varepsilon_{individual\ bases}$$

In reality, factors like sequence and length affect the extinction coefficient of the oligonucleotide. The nearest-neighbour method takes into account the interactions between bases: it is the sum of the extinction coefficients of dinucleotide pairs (composed of the first and second bases, second and third bases, and so on reading from 5′ → 3′), minus the sum of the extinction coefficients of the individual bases not at the 5′ or 3′ termini. While the nearest-neighbour method is typically very accurate for unmodified oligonucleotides, it does not extend easily to include modifications.

$$Nearest\ neighbor: \varepsilon_{260} = \sum_{1}^{N-1} \varepsilon_{dinucleotide\ pairs} - \sum_{2}^{N-1} \varepsilon_{individual\ bases}$$

Empirically derived values measured at pH 7 and a wavelength of 260 for $\varepsilon_{individual\ bases}$ (**Table 1-1**) and $\varepsilon_{nearest\ neighbour}$ (**Tables 1-2** and **1-3**) are given below in units of $M^{-1}\ cm^{-1}$, with an average error of approximately 4%.

Table 1-1. Individual extinction coefficients for DNA and RNA bases (in $M^{-1}\ cm^{-1}$).[66]

	A	**C**	**G**	**T**
$DNA_{individual\ bases}$	15,400	7,400	11,500	8,700
	A	**C**	**G**	**U**
$RNA_{individual\ bases}$	15,400	7,200	11,500	9,900

Table 1-2. Nearest neighbour extinction coefficients for DNA dinucleotides (in $M^{-1}\ cm^{-1}$).[66]

$DNA_{nearest\ neighbour}$		**3′**			
		A	**C**	**G**	**T**
5′	**A**	24,400	21,200	25,000	22,800
	C	21,200	14,600	18,000	15,200
	G	25,200	17,600	21,600	20,000
	T	23,400	16,200	19,000	16,800

Table 1-3. Nearest neighbour extinction coefficients for RNA dinucleotides (in M^{-1} cm^{-1}).[66]

$RNA_{nearest\ neighbour}$		3′			
		A	**C**	**G**	**U**
5′	**A**	27,400	21,000	25,000	24,000
	C	21,000	14,200	17,800	16,200
	G	25,200	17,400	21,600	21,200
	U	24,600	17,200	20,000	19,600

For double-stranded oligonucleotides, π-stacking of the bases in the duplex lowers the absorbance, an effect called hypochromicity (*h*). The extinction coefficients of dsDNA can be estimated with the following equations:

$$E_{ds} = \left(E_{ss,forward\ strand} + E_{ss,reverse\ strand}\right) * (1 - h)$$

$$h = (0.059 * fraction\ of\ GC\ bases) + (0.287 * fraction\ of\ AT\ bases)$$

UV measurements can also be used to assess contamination of the nucleic acid sample by comparing the ratios of absorbance at different wavelengths. A_{260}/A_{280} and A_{260}/A_{230} are commonly used ratios since proteins absorb near 280 nm, and organic compounds and chaotropic salts such as urea, EDTA, and phenolate ions absorb near 230 nm. An A_{260}/A_{280} close to 1.8 and 2.0 are ideal for DNA and RNA respectively, and an A_{260}/A_{230} ratio close to 1.5-1.8 is ideal for both.[3, 66-69]

1.9 Objectives

The broad objective of this project is to develop methods for the assembly of nucleic acids with site-specific modifications beyond the current length limits of solid-phase synthesis, and bridge the gap to complex genome assemblies. The ability to do so would be a useful tool in the fields of epigenetics, bionanotechnology, nucleic acid therapeutics, and more. The assembly method would have to be high-

yielding enough not to require PCR amplification, because site-specificity would not be preserved by polymerases. To be biologically relevant, the assembled genes would also have to be compatible with the processes of replication, transcription, and translation. The approach described in this project is to ligate chemically synthesised oligonucleotides.

In Chapter 2, ligase assembly of DNA and RNA was optimised to determine the limits of enzymatic ligation. Chapter 3 explored and optimised chemical ligation of RNA as a potentially less expensive, more scalable alternative to enzymatic ligation. In particular, attention was focused on copper(I)-catalysed azide-alkyne cycloaddition, which produces an unnatural triazole backbone at the ligation site. Replication and transcription by DNA and RNA polymerases, respectively, of DNA containing triazole backbones have been described in the literature. Successful translation of RNA with triazole backbones would demonstrate compatibility of the triazole linkage across all core processes of the central dogma of molecular biology, and further open up chemical ligation as a pathway to nucleic acid assembly. Chapter 4 investigated whether chemically ligated RNA with an unnatural triazole backbone could be translated into protein by ribosomes.

Chapter 2: Enzymatic ligation of nucleic acids

2.1 Introduction and aim of the study

Site-specific incorporation of chemical modifications in nucleic acids would provide an invaluable tool for a wide range of applications, from therapeutics to nanotechnology, but current synthesis methods are limited in their ability to achieve that goal. Polymerase-based assembly methods only recognise Watson-Crick base pairs and are restricted to enzyme-tolerated modifications. While chemical synthesis methods enable position-specific incorporation of various modifications, they are limited to oligonucleotide lengths of around 150 nt for DNA (around 75 nt for RNA) due to imperfect monomer coupling. To address these issues, ligation of chemically synthesised oligonucleotides would facilitate the construction of long nucleic acids with site-specific modifications (in this context, on the order of 10^2-10^3 bases).

There are two main ligation routes, namely enzymatic and chemical. Nucleic acid assembly via enzymatic ligation was first reported in the 1960s, and led to the development of methods such as ligase cycling reaction (LCR). In LCR, thermostable ligases are used to covalently join the 3′ and 5′ termini of oligonucleotides in repeated cycles of denaturing, annealing, and ligating. Methods like LCR often include PCR or cloning steps at the end to purify and amplify the desired product.[70-73] However, site-specific modifications introduced during chemical oligonucleotide synthesis would be lost during polymerase copying of the ligation product. One aim of this study was therefore to investigate the limits of ligation without polymerase amplification, by optimising reaction conditions for DNA gene assembly by ligases. DNA enzymatic ligation was chosen as the starting point before moving on to more challenging, less well-studied ligation reactions with RNA. The second aim of this study was to assemble enzymatically ligated RNA templates as controls to compare with chemically ligated RNA (Chapter 3) in translation experiments (Chapter 4).

2.2 Background

2.2.1 T4 ligases

T4 DNA ligase, derived from the enterobacteria phage T4, is one of the most commonly used enzymes for joining together nucleic acid molecules. The ligase is most efficient at ligating nicks in dsDNA, but is also capable of ligating sticky and blunt-ended dsDNA as well as DNA-RNA hybrids. It uses ATP as a cofactor to catalyse the formation of a phosphodiester bond between adjacent 3′-hydroxyl and 5′-phosphate groups. First, an adenylyl group from ATP is transferred to the amino group of a lysine residue in the active site of the ligase. Next, the AMP group is transferred to the 5′-phosphate of a nucleic acid, forming a pyrophosphate group. Finally, the 3′-hydroxyl group of an adjacent nucleic acid attacks the 5′-phosphate, releasing the AMP and resulting in a phosphodiester linkage between the 3′ and 5′ nucleic acid ends (**Figure 2-1**).[74-78]

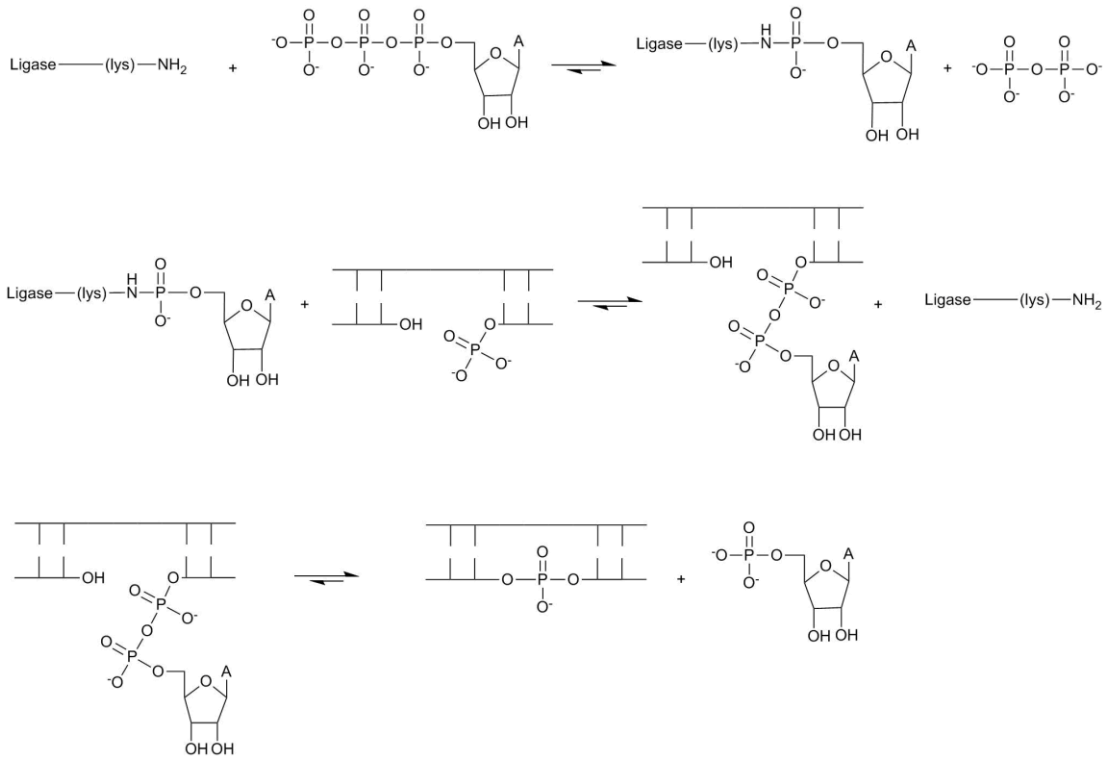


Figure 2-1. Reaction mechanism for T4 DNA ligase, which links adjacent DNA or RNA 3′-hydroxyl and 5′-phosphate termini. ‘A’ denotes an adenine base.

T4 DNA ligase functions at a range of temperatures from 4-37 °C. The most efficient ligation temperature for a given set of oligonucleotides is a balance between the optimal temperature for the enzyme and the hybridisation temperature of complementary oligonucleotides. In general, lower incubation temperatures require longer incubation times. Although ligation efficiency is greatly reduced if there are single mismatches at either side of the ligation site (especially on the 3′ side), higher incubation temperatures further help to minimise the formation of products with mismatches.[79, 80]

The inclusion of polymers such as polyethylene glycol (PEG) 6000 in the reaction mixture has been shown to increase the ligation yield of T4 DNA ligase. Polymers like PEG act as macromolecular crowding agents, effectively increasing the nucleic acid and ligase concentrations in the solution. The effect was especially notable in the ligation of blunt-ended dsDNA. Large PEG polymers with molecular weights ranging from 3,000 to 20,000 all gave similar results, but smaller PEG polymers with molecular weights of 200 to 400 did not increase yield. Published studies have generally used PEG 4000, PEG 6000, or PEG 8000 (with average molecular weights of 4,000, 6,000, and 8,000 respectively) at concentrations of 5-20% (w/v). As a non-charged molecule, PEG can be separated from nucleic acids by PAGE.[81, 82]

T4 RNA ligase 2, also derived from the phage T4, joins the 3′-hydroxyl group of RNA to the 5′-phosphate of either RNA or DNA in a duplex. Its reaction mechanism is similar to that of T4 DNA ligase, also requiring ATP as a cofactor.[83-86]

2.2.2 Taq *DNA ligase*

Taq DNA ligase is derived from *Thermus aquaticus*, a thermophilic bacterium, and is one of the most commonly used thermostable ligases. The enzyme is active at temperatures of 45-65 °C, and can ligate sticky-ended DNA and nicks in dsDNA. The reaction mechanism for *Taq* ligase is similar to that of T4 ligase, but uses nicotinamide adenine dinucleotide (NAD^+) instead of ATP as the AMP donor. Because of its thermal stability, *Taq* DNA ligase has been used in protocols that require elevated temperatures, such as Gibson Assembly and LCR. Advantages to using a thermostable ligase like *Taq* ligase include the ability to run the reaction at or near the oligonucleotide melting temperature, minimising obstructive secondary structure and off-target ligation.[71, 78, 87-89]

2.2.3 T4 polynucleotide kinase (PNK)

T4 polynucleotide kinase (PNK) phosphorylates DNA and RNA by catalysing the transfer of the ɣ-phosphate from ATP to the 5′-hydroxyl group of the nucleic acid (**Figure 2-2**). It is also capable of transferring the ɣ-phosphate to the 5′-OH of nucleoside 3′-monophosphates, and acting as a 5′ and 3′ phosphatase in the presence of ADP. *In vivo*, the T4 phage employs PNK partnered with ligase to repair host-cell tRNAs that have been cleaved by the bacterium's defense system. *In vitro*, PNK is often used in laboratories to label the 5′ termini of nucleic acids with [^{32}P] and to phosphorylate the 5′ ends of nucleic acids for enzymatic ligation. T4 PNK is more efficient at phosphorylating single-stranded substrates than blunt-ended duplexes, perhaps due to the narrow entrance to its kinase active site.[90-92]

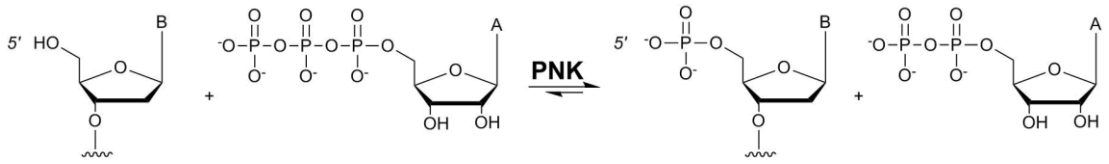


Figure 2-2. Phosphorylation of the 5′ terminus of a nucleic acid by PNK. 'A' denotes an adenine base, and 'B' denotes any of the four nucleobases.

2.2.4 Biotin-streptavidin magnetic beads

The binding between streptavidin, a 52.8 kDa protein, and biotin, a small molecule, is one of the strongest known non-covalent interactions in nature ($K_d = 10^{-14 \text{ to } -15}$ M). It is stable over a wide range of pH and temperatures. In addition to this strong binding, several factors make the interaction attractive for use in purification and labelling assays. Biotin is a small and stable molecule, so it can be attached to molecules like DNA and protein with relatively little impact on structure and function. The binding between streptavidin and biotin is specific, and each streptavidin molecule can also bind up to four biotin molecules.[93, 94] Several commercially available platforms exploit these advantageous properties, such as streptavidin-coated magnetic beads. Magnetic beads can be easily washed, separated, and concentrated without having to perform centrifugation, filtering, or precipitation steps. Streptavidin magnetic beads can be used to capture biotin-labelled DNA, and act as a solid support for subsequent reactions and wash steps (**Figure 2-3**). Such a system has been utilised to synthesise DNA,[95, 96] hybridise DNA,[97] generate ssDNA from PCR products,[98] detect point mutations,[99] and sequence captured DNA.[100]

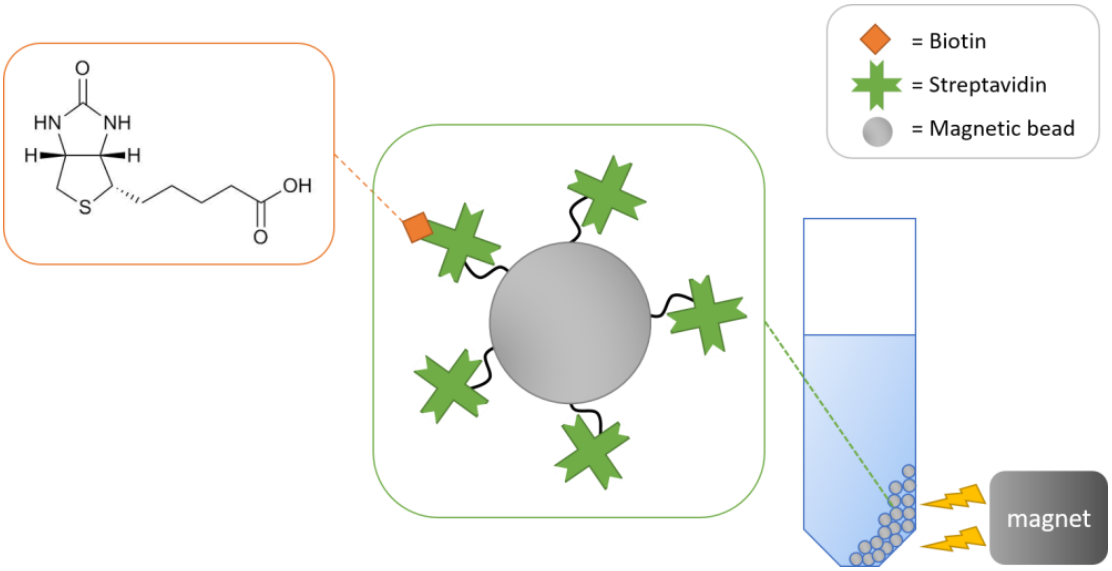


Figure 2-3. Streptavidin-coated magnetic beads can be used to capture and purify biotin-labelled molecules.

Results and Discussion

2.3 Enzymatic ligation of DNA

2.3.1 Templated enzymatic ligation of DNA

The limitations of enzymatic DNA ligation were investigated to understand the issues that contribute to low yield, and whether they are enzyme-specific. Enzymatic ligation of an unmodified ssDNA system with complementary bridging splints was chosen as a starting point for optimisation, due to its relative simplicity and ease of synthesis. The 762 nt eGFP gene was divided into ten oligonucleotide fragments (**O1-10**; 59-95 nt each), with nine bridging splints (**S1-9**; 30 nt each) such that **S1** was complementary to the last 15 nt of **O1** and the first 15 nt of **O2** (**Figure 2-4**; see Appendix for sequences). Preliminary reactions with T4 ligase revealed that yield decreased (and incomplete ligation by-products increased) with increasing numbers of ligation sites. Each successive oligonucleotide pair was ligated and

analysed separately to ensure that the ligation efficiency was not related to individual oligonucleotide design (**Figure 2-5**).

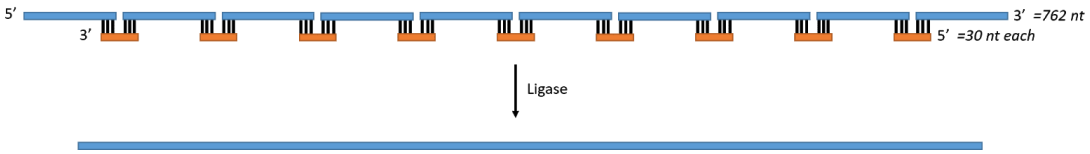


Figure 2-4. Schematic of one-pot assembly of the 762 nt eGFP gene by enzymatically ligating ten ssDNA oligonucleotides templated with nine complementary splints (shown in orange).

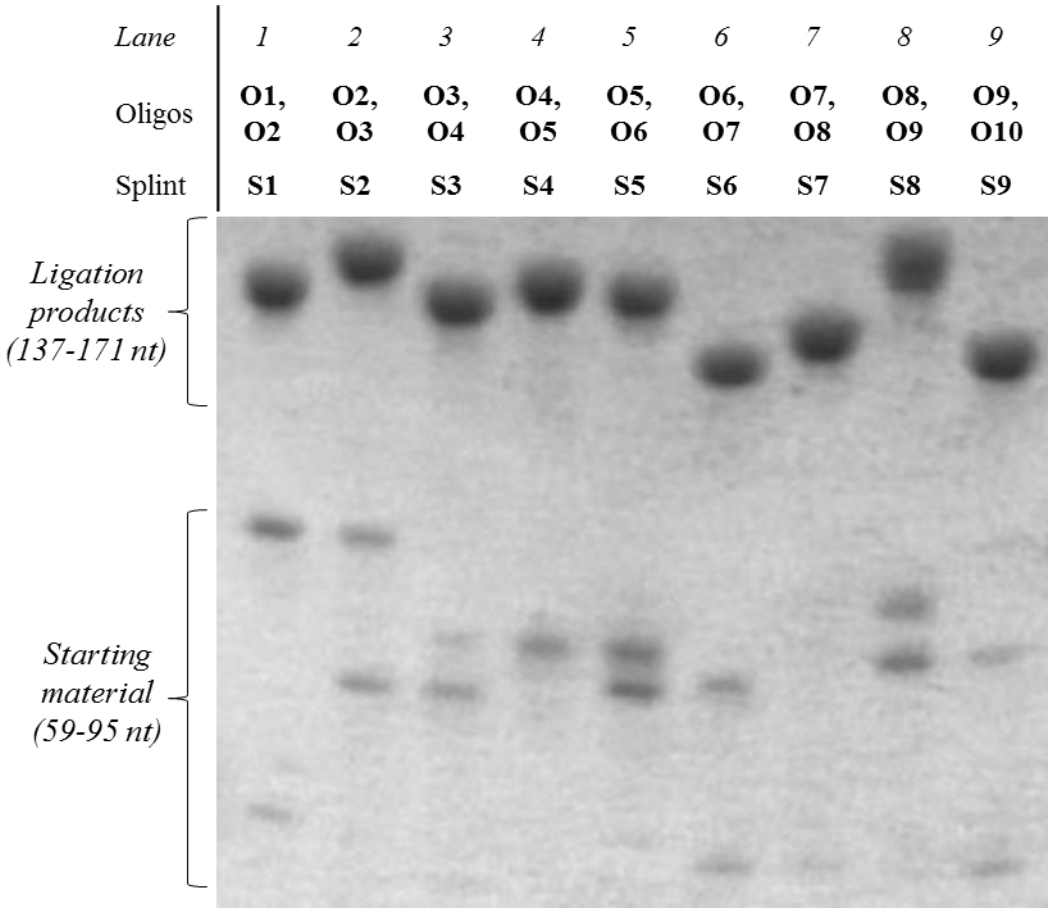


Figure 2-5. 10% denaturing PAGE analysing ligation of each oligonucleotide pair with T4 DNA ligase, in a 10-oligonucleotide system containing the eGFP gene.

All five steps in the general enzymatic ligation workflow were then systematically investigated, as summarised in **Table 2-1** (additional gels shown in

Appendix **Figures A1**-**A4**): 1) combining oligonucleotides and splints, 2) adding ligase buffer, 3) annealing, 4) adding ligase, and 5) incubating the reaction. For the majority of cases, changing the oligonucleotide to splint ratio, reaction buffer composition, annealing rates, and incubation conditions did not lead to significant changes in yield. Therefore, the T4 ligation reaction itself appeared to be relatively robust and mainly limited by the number of ligation sites. While one-pot ligation of ten oligonucleotides to assemble the complete 762 nt eGFP gene was possible, the yield and purity were poor due to incompletely ligated by-products. Ligating the two halves of the gene separately, extracting them from a PAGE gel, and then ligating the extracted products to form the whole gene greatly improved purity (**Figures 2-6** and **2-7**).

Table 2-1. Conditions tested for the enzymatic ligation of DNA oligonucleotides with T4 DNA ligase, templated by complementary splits. Additional gels are shown in the appendix (**Figures A1**-**A4**).

Step in the protocol: Control conditions	**Conditions tested**
1. Combine oligonucleotides and splints in a tube: equimolar amounts of oligonucleotide and splint (1:1)	- Phosphorylate with PNK before ligation - Oligonucleotide to splint ratio: 2:1 or 1:2 - Ligate sets of oligonucleotides separately, extract from PAGE gel, then ligate together
2. Add ligase buffer: no NaCl, 10 mM $MgCl_2$, 1 mM ATP (from manufacturer)	- Monovalent cations: add 20 mM NaCl - Divalent cations: increase to 13 or 16 mM $MgCl_2$ - Increase to 2 mM ATP
3. Anneal: 5 min at 95 °C, then slowly cooled to RT	- Fast annealing: place sample in -20 °C after heating - No annealing
4. Add ligase: T4 DNA ligase	- Ligase amount: 2x
5. Incubate reaction: 4 hr at RT	- Time: 1 hr, or ON - Temperature: 4 °C, RT, or 37 °C - Thermal cycling (1 cycle = 1 hr at 37 °C → 5 min at 95 °C): 1, 3 or 5 cycles, with fresh ligase added before each cycle

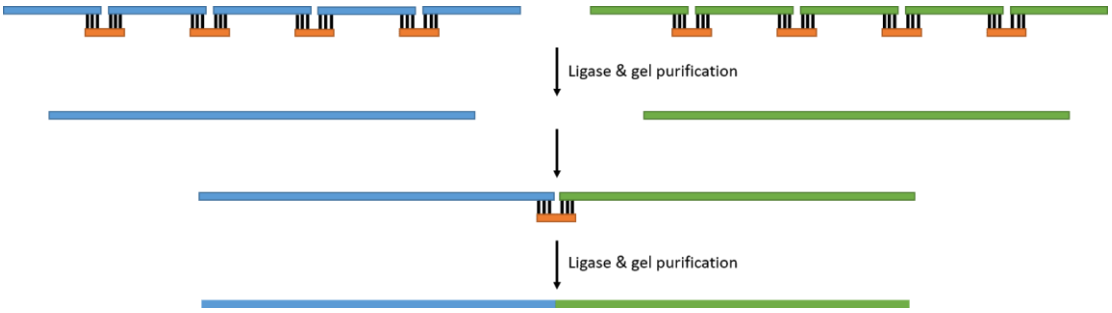


Figure 2-6. Schematic of the assembly of the eGFP gene by ligating the two halves of the gene separately, purifying the products by PAGE, and then ligating them together to form the complete gene. Templating splints are shown in orange.

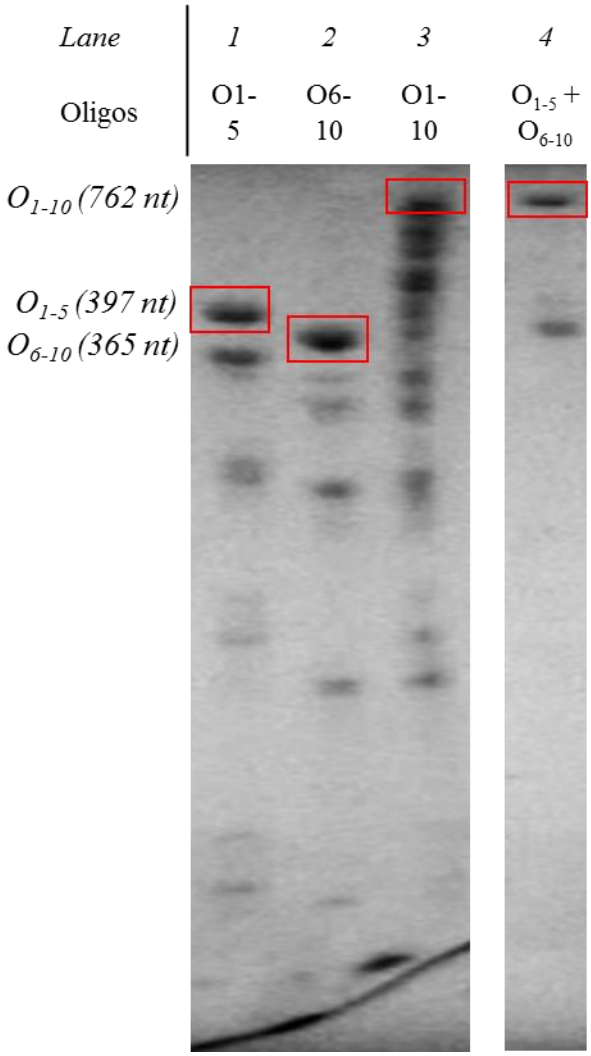


Figure 2-7. 5% denaturing PAGE of DNA oligonucleotides bridged by complementary splints and ligated with T4 DNA ligase. **Lane 1:** One-pot ligation of five oligonucleotides that make up the first half of the eGFP gene (O1-5); **Lane 2:** One-pot ligation of five oligonucleotides that make up the second half of the eGFP gene (O6-10); **Lane 3:** One-pot ligation of ten oligonucleotides that make up the 762 nt eGFP gene (O1-10); **Lane 4:** Ligation of the gel-extracted products from Lanes 1 and 2 to assemble the 762 nt eGFP gene (O_{1-5}+O_{6-10}).

Thermal cycling, in which oligonucleotides are melted and re-hybridised in each cycle, was investigated as a way to account for imperfect or incomplete annealing. Similar to the LCR process, each cycle consisted of a denaturing step followed by an annealing and ligation step (**Figure 2-8**). DNA ligase from the thermophilic bacterium *Thermus aquaticus* (*Taq*) was also tested, as its thermal stability allows for facile thermal cycling without having to add fresh ligase before each cycle. Optimisation experiments were conducted for *Taq* DNA ligation, and are summarised in **Table 2-2** (additional gels shown in Appendix **Figures A5-A7**). *Taq* ligase consistently produced lower yield than T4 ligase when ligating 5 or more oligonucleotides, so *Taq* ligations were not investigated in more depth (Appendix **Figure A5**).

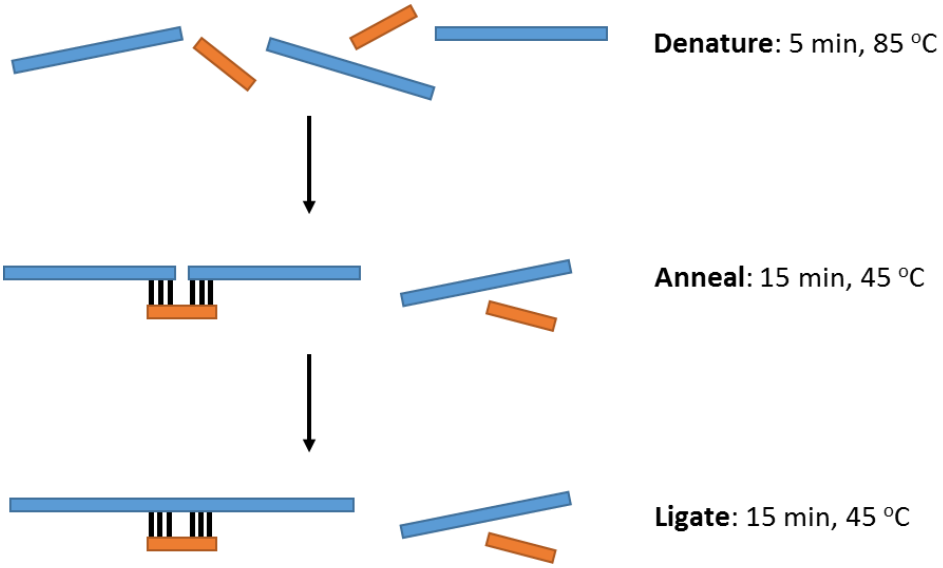


Figure 2-8. Schematic of one thermal cycle for ligation with *Taq* ligase, which is repeated multiple times. Annealing and ligation both occur during the same 15 min step at 45 °C. Templating splints are shown in orange.

Table 2-2. Conditions tested for the enzymatic ligation of DNA oligonucleotides with *Taq* DNA ligase, templated by complementary splits. Gels are shown in the appendix (**Figures A5-A7**).

Step in the protocol: Control conditions	**Conditions tested**
1. Combine oligonucleotides and splints in a tube: equimolar amounts of oligonucleotide and splint (1:1)	- Excess oligo: 4:1, 4:2, or 4:3 oligonucleotide to splint ratio - Excess splint: 1:2 or 1:4 oligonucleotide to splint ratio
2. Add ligase buffer: (from manufacturer)	
3. Anneal: 5 min at 95 °C, then slowly cooled to RT	
4. Add ligase: *Taq* DNA ligase	- Ligase amount: 2x - Add fresh ligase after 10th thermal cycle (out of 20)
5. Incubate reaction: 20 cycles (1 cycle = 5 min at 85 °C → 15 min at 45 °C)	- Thermal cycles (1 cycle = 5 min at 85 °C → 15 min at 45 °C): 5, 10, 15, or 20 cycles - No cycles: ON at 55 °C

Simultaneous enzymatic assembly of both the sense and antisense strands was also explored, to determine if the more rigid, fully double helical structure would enhance ligation yields. The system consisted of overlapping oligonucleotide fragments of approximately 40 nt each that formed a double-stranded construct upon hybridisation (**Figure 2-9**). The first 232 nt of the eGFP gene and its complementary sequence were used to test the dsDNA approach (see Appendix for sequences). However, ligation yield did not improve relative to the splinted system, and the products were increasingly streaky and difficult to isolate by PAGE with increasing length (**Figure 2-10**). Ligation of a fully double-stranded construct was thus not investigated further.

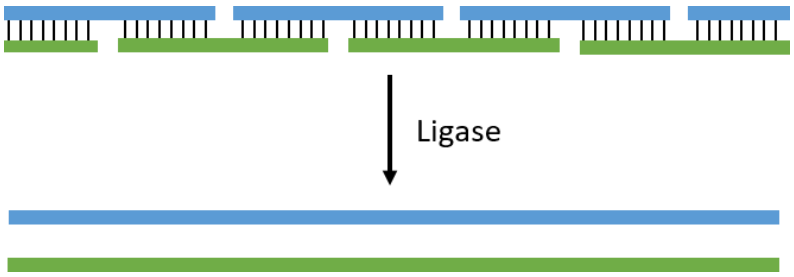


Figure 2-9. Schematic of DNA assembly by ligating overlapping, complementary dsDNA oligonucleotides.

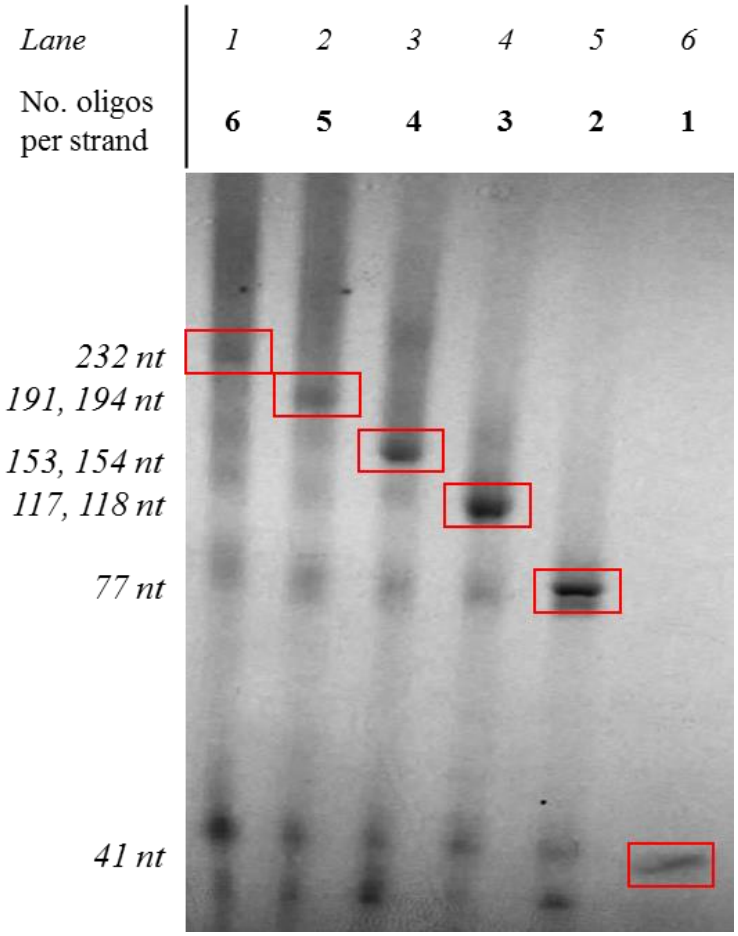


Figure 2-10. 8% denaturing PAGE of overlapping, complementary DNA oligonucleotides (oligos) ligated with T4 DNA ligase. The oligonucleotides form a dsDNA construct upon hybridisation. Some lanes have two ligation products of slightly different lengths, corresponding to the two strands of the dsDNA construct. **Lane 1:** One-pot ligation of twelve oligonucleotides total, six per strand (F1-6 and R1-6); **Lane 2:** One-pot ligation of ten oligonucleotides total, five per strand (F2-6 and R2-6); **Lane 3:** One-pot ligation of eight oligonucleotides total, four per strand (F3-6 and R3-6); **Lane 4:** One-pot ligation of six oligonucleotides total, three per strand (F4-6 and R4-6); **Lane 5:** One-pot ligation of four oligonucleotides total, two per strand (F5-6 and R5-6); **Lane 6:** Single oligonucleotide (F6).

2.3.2 Magnetic bead-based ligation and purification

Ligating the desired product in PAGE-purified sections to assemble the whole construct greatly improved product purity, but PAGE purification is time consuming and labour intensive. To streamline this issue, magnetic beads were investigated as both a solid support for ligation reactions and as a capture mechanism for isolating the desired product. In the former method, the idea was similar to solid phase oligonucleotide synthesis, in that splints and oligonucleotides were added sequentially with wash steps in between. The first oligonucleotide was attached via a 5′-biotin to streptavidin-coated magnetic beads (**Figure 2-11a**), each successive splint or oligonucleotide fragment was hybridised followed by a wash step (**Figure 2-11b**), and ligase was added at the end (**Figure 2-11c**). Such a procedure would allow each successive oligonucleotide piece to be added in excess, and theoretically decrease the amount of different by-products.

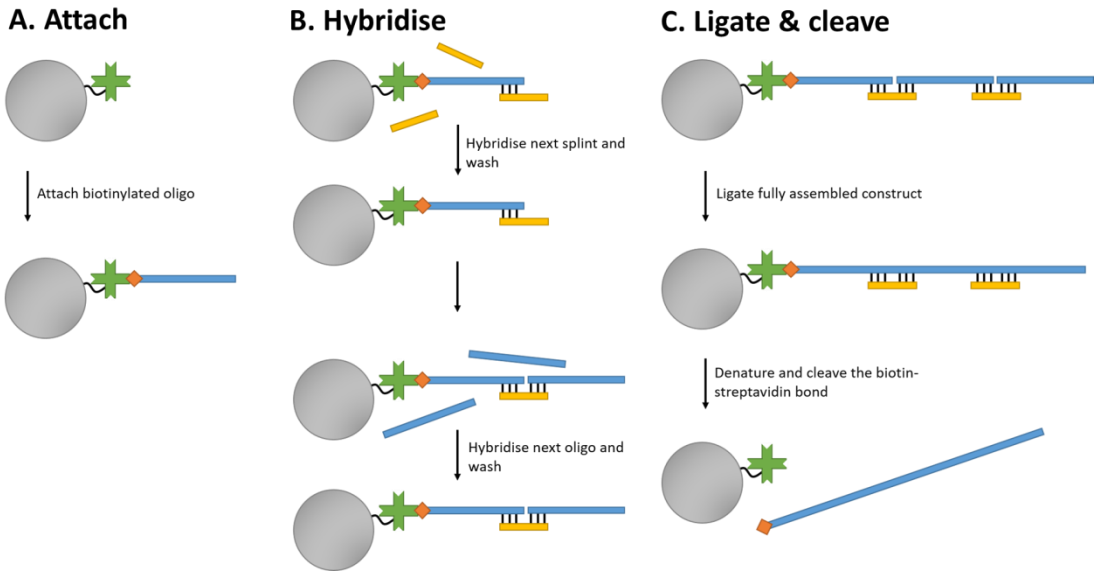


Figure 2-11. Schematic for ligation reactions using streptavidin-coated magnetic beads as a solid support. **A)** The first oligonucleotide with a 5′-biotin is attached to the beads by binding with the streptavidin. **B)** Each successive splint or oligonucleotide is hybridised, with any excess unbound ones removed by washing the beads. **C)** Once the whole construct is assembled, ligase is added, and the ligated product is cleaved from the beads.

In the latter method, the streptavidin-coated magnetic beads would be used to purify and isolate the complete ligation product. First, the oligonucleotides would be ligated in solution. An oligonucleotide complementary to the first 30 bases from the 5′ end of the desired product would be attached to one batch of beads, and an oligonucleotide complementary to the last 30 bases from the 3′ end would be attached to another batch of beads. The ligated oligonucleotides would be hybridised to one batch of beads, washed, and removed (**Figure 2-12a**), and then the collected oligonucleotides would be hybridised to the other batch of beads, washed, and removed (**Figure 2-12b**). Theoretically, only DNA containing both the first 30 bases from the 5′ end and last 30 bases from the 3′ end should remain, i.e. only the completely ligated product.

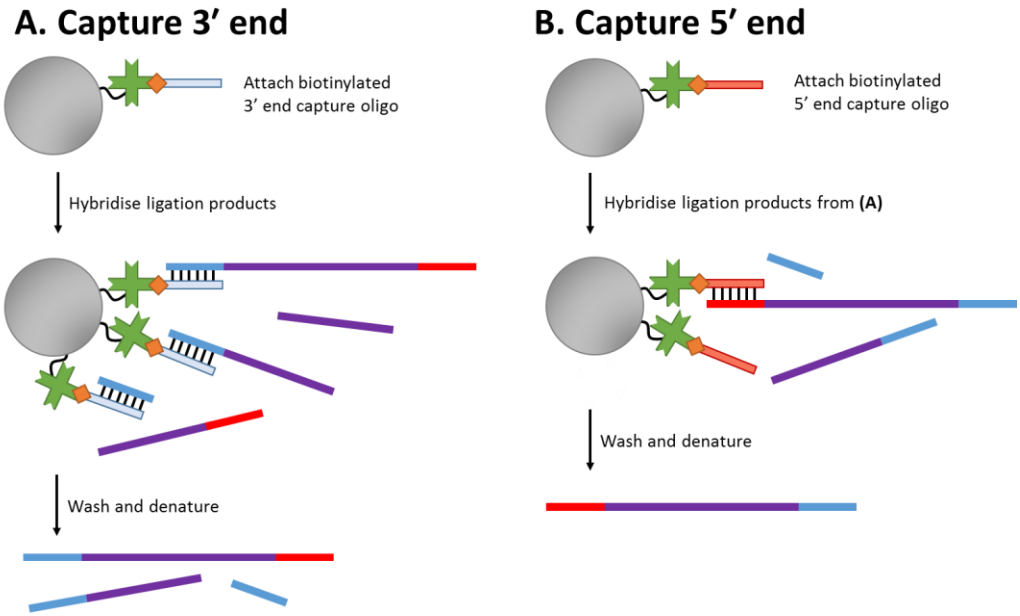


Figure 2-12. Schematic for capture and purification of ligation products with streptavidin-coated magnetic beads. One batch of beads would have biotinylated oligonucleotides complementary to the 3′ end of the complete ligation product, while the other batch would have biotinylated oligonucleotides complementary to the 5′ end. **A)** Ligation products with the 3′ sequence hybridise to the first batch of beads, and any unbound DNA is washed away. **B)** Next, the collected ligation products from **(A)** are incubated with the second batch of beads that are complementary to the 5′ end, and any unbound DNA is washed away. Theoretically, only complete ligation products should remain (i.e. with the correct sequences at both the 3′ and 5′ ends).

Experiments summarised in **Table 2-3** (additional gels shown in Appendix **Figures A8**-**A11**) were conducted to optimise DNA attachment and ligation or capture on the beads, but yields were consistently lower than for the same reaction performed in solution (**Figure 2-13a**). Switching from hydrophobic beads with a 2.8 µm diameter (M-280 Dynabeads^{TM}, Thermo Fisher Scientific) to hydrophilic beads with a 1 µm diameter (MyOne C1 Dynabeads^{TM}, Thermo Fisher Scientific) improved loading capacity. While the initial loading was higher for the smaller hydrophilic beads, the ligation product yield was still low (**Figure 2-13b**). Likewise, capture of the ligation product was also low. These results, combined with information from the manufacturer's manuals, indicated it was possible that steric hindrance and charge repulsion were obstructing annealing or capture of the oligonucleotides. Combined with the high cost of the beads, the poor yield rendered the streptavidin magnetic beads unsuitable for the aims of this project.

Table 2-3. Conditions tested for DNA ligation on and capture with streptavidin-coated magnetic beads. Gels are shown in the appendix (**Figures A8**-**A11**).

Conditions Tested
Add complementary splint and oligonucleotide pairs at the same time, or wash in between
Add each splint or oligonucleotide twice with a wash in between
Wash beads with 0.1% BSA solution
Increase ligation time, ligase amount, incubation temperature, or incubation volume
Compare T4 and *Taq* ligase
Compare hydrophobic and hydrophilic beads of different sizes
Attach capture oligonucleotide to beads first or hybridise to ligation product in solution first

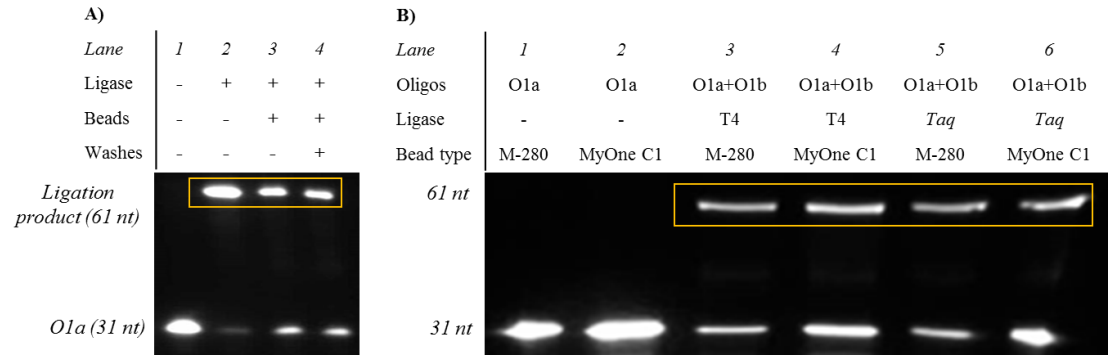


Figure 2-13. O1a, a 31 nt DNA oligonucleotide with a 5′ biotin and FAM fluorophore label, was used to investigate binding to and ligation on streptavidin-coated magnetic beads. **A)** 10% denaturing PAGE gel showing ligation of two DNA oligonucleotides (O1a and O1b) on Dynabeads^{TM} M-280 streptavidin-coated beads (**lanes 3** and **4**), compared to ligation in solution (**lane 2**). In **lane 3**, the splint and the donor oligonucleotide were added at the same time; in **lane 4**, the splint was added first and followed by a gentle wash before addition of the donor oligonucleotide. **B)** 12.5% denaturing PAGE gel showing attachment and ligation on two different kinds of streptavidin-coated beads, with two different ligases. Dynabeads^{TM} M-280 have a hydrophobic surface and 2.8 µm diameter, whereas Dynabeads^{TM} MyOne C1 have a hydrophilic surface and 1.0 µm diameter. Ligation reactions were incubated overnight (ON) at room temperature (RT) or at 55 °C for T4 and *Taq* DNA ligases, respectively.

2.4 Enzymatic ligation of RNA

Ligases were used to assemble linear and circular mRNA templates with natural phosphodiester backbone linkages for *in vitro* translation studies that are described further in Chapter 4 (**Figure 2-14**; see Chapter 4 and Appendix for sequences). Both DNA ligase and RNA ligase 2 can ligate an RNA 3′-hydroxyl to an RNA 5′-phosphate in either an RNA:DNA or RNA:RNA duplex.[76, 83] Liquid chromatography-mass spectrometry (LC-MS) measurements of a 99 nt RNA ligation product purified by PAGE gel extraction showed that the measured mass (32,097) closely matched the expected mass (32,099 nt), demonstrating the fidelity of the enzymatic reaction (**Figure 2-15**; Appendix **Figure A13**). To optimise ligation yield, several conditions were tested: type of ligase, type of splint, incubation time,

incubation temperature, addition of PEG 6000, treatment with PNK, and annealing of the oligonucleotides prior to ligation (**Table 2-4**; additional gel shown in Appendix **Figure A12**). Initially, RNA ligase 2 was used because it was found to produce higher ligation yields. However, a switch was eventually made to DNA ligase because it performed more consistently and was more cost-effective (**Figure 2-16**). DNA splints were also compared to RNA splints for templating the ligation reactions, and DNA splints were found to be more suitable. Use of RNA splints resulted in extra bands on PAGE gels that ran slower than DNA-templated ligation products. This unexpected result with RNA splints happened in both DNA ligase and RNA ligase 2 reactions, and even in templated chemical ligation reactions that are described further in Chapter 3. As such, the extra band produced when using RNA splints was likely not caused by the ligases (**Figure 2-17**).

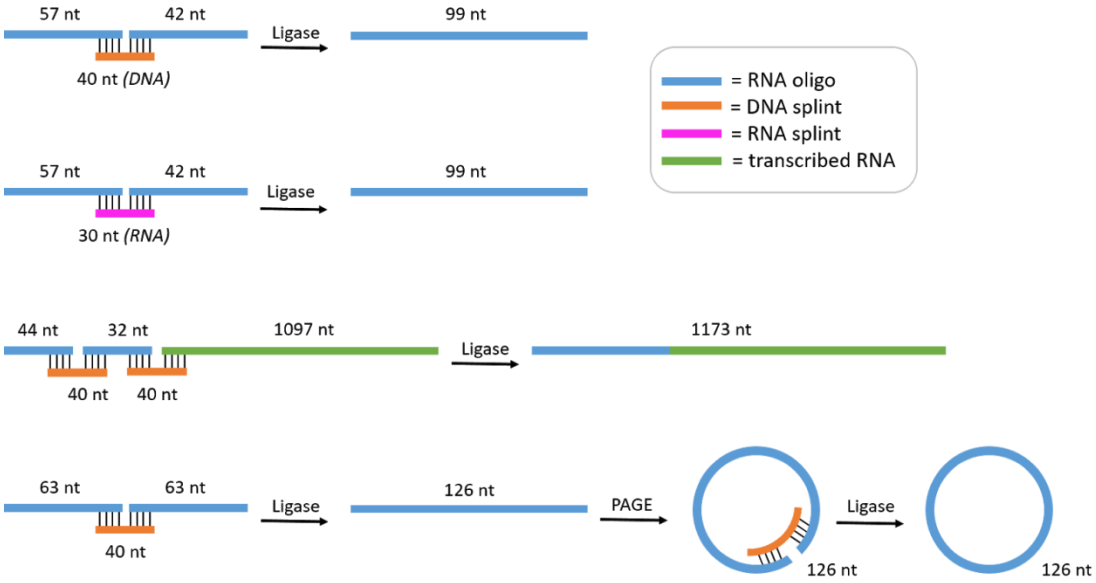


Figure 2-14. Linear and circular RNA constructs that were assembled by enzymatic ligation. The constructs were used in translation reactions described in Chapter 4.

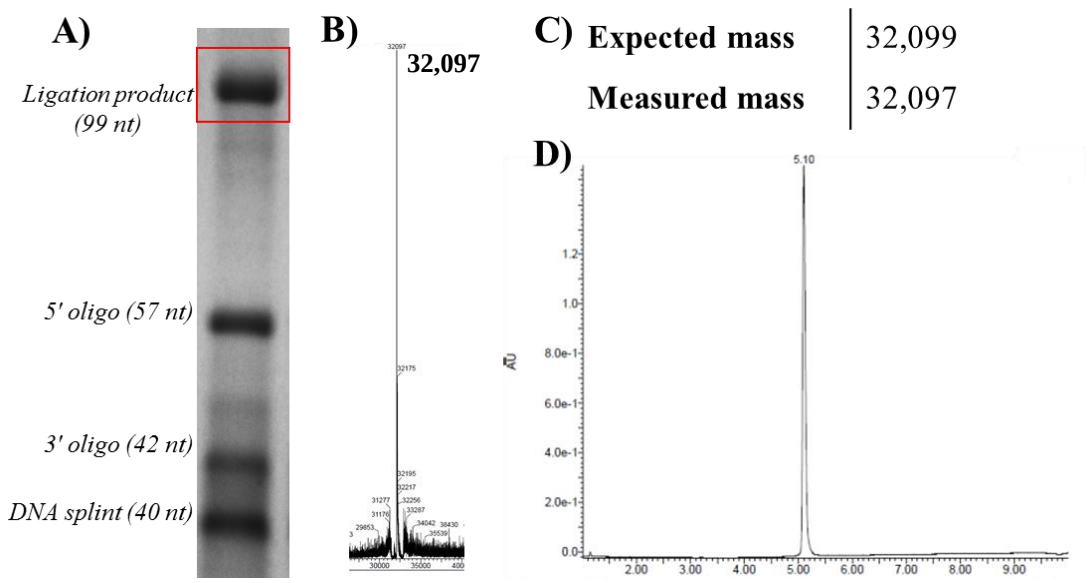


Figure 2-15. A) 12% denaturing PAGE gel showing enzymatic ligation of a 57 nt RNA and a 42 nt RNA to form a 99 nt product. The ligation product was extracted from the gel and analysed by LC-MS. **B)** Deconvoluted mass spectrum of the ligation product, showing a molecular mass of 32,097. **C)** The measured molecular mass of 32,097 was very close to the expected mass of 32,099. **D)** UV chromatogram, showing a sharp single peak. (Oligonucleotides: FLAGHA-R1-e, FLAGHA-R2-e, and FLAGHA-S)

Table 2-4. Conditions tested for the enzymatic ligation of RNA oligonucleotides templated with complementary splints, using T4 DNA ligase or T4 RNA ligase 2. An additional gel is shown in the appendix (**Figure A12**).

Step in the protocol: Control conditions	**Conditions tested**
1. Combine oligonucleotides and splints in a tube: equimolar amounts of RNA oligonucleotide and DNA splint (1:1)	- Phosphorylate with PNK before ligation - Splint length: 24 nt or 40 nt - Splint type: DNA or RNA
2. Add ligase buffer: (from manufacturer)	- Add PEG: 6.7% PEG 6000 or 25% PEG 4000
3. Add ligase: T4 DNA ligase	- Ligase type: T4 DNA ligase or T4 RNA ligase 2
4. Incubate reaction: 6 hr at 37 °C	- Incubation temperature: RT or 37 °C - Incubation time: 1, 2, 4, 6, or 16 hr

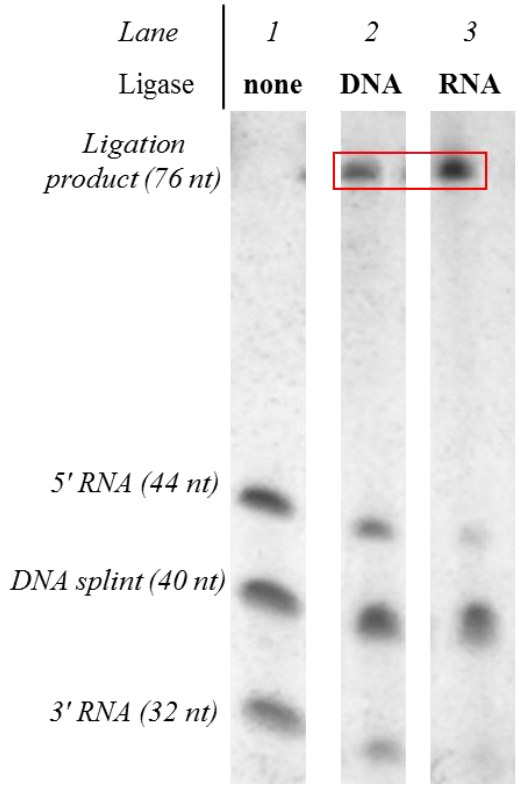


Figure 2-16. 8% denaturing PAGE gel comparing the ligation efficiencies of T4 DNA ligase and T4 RNA ligase 2. (Oligonucleotides: HisLuc-R1-e, HisLuc-R2-e, and HisLuc-S12)

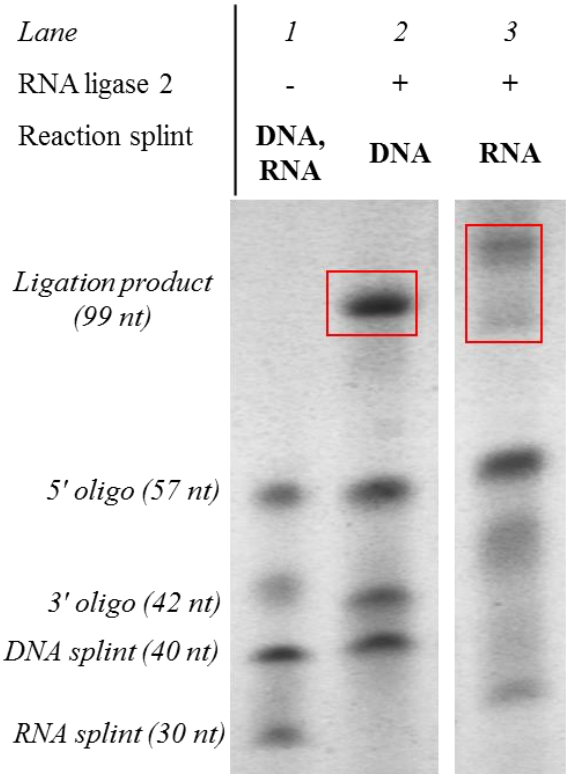


Figure 2-17. 12% denaturing PAGE gel investigating the effects of using a 40 nt DNA splint compared to a 30 nt RNA one. (Oligonucleotides: FLAGHA-R1-e, FLAGHA-R2-e, FLAGHA-s, and FLAGHA-S_{RNA})

Incubation at 37 °C resulted in higher yield than incubation at RT, so enzymatic RNA ligation reactions were incubated at 37 °C for up to 16 hours (**Figure 2-18**). Adding PEG 6000 as a crowding agent to the ligation reaction appeared to slightly improve yield, and since PEG is an inexpensive component, it was included in subsequent reaction mixtures. Treating the oligonucleotides with PNK to ensure that the 5′ termini were efficiently phosphorylated prior to ligation did not noticeably impact yield, so consequently it was not used in future reactions (**Figure 2-19**). Annealing synthetic RNA oligonucleotides with DNA splints before ligation also did not significantly affect yield (**Figure 2-20**). On the other hand, it was found that annealing long RNA (transcribed RNA of approximately 1,000 nt) in four different buffers all led to extensive degradation of the RNA, compared to incubating for an equal amount of time in the same buffer at room temperature (RT) or performing the same annealing protocol in water (**Figure 2-21**). RNAs were not annealed thereafter to avoid potential heat-related degradation.

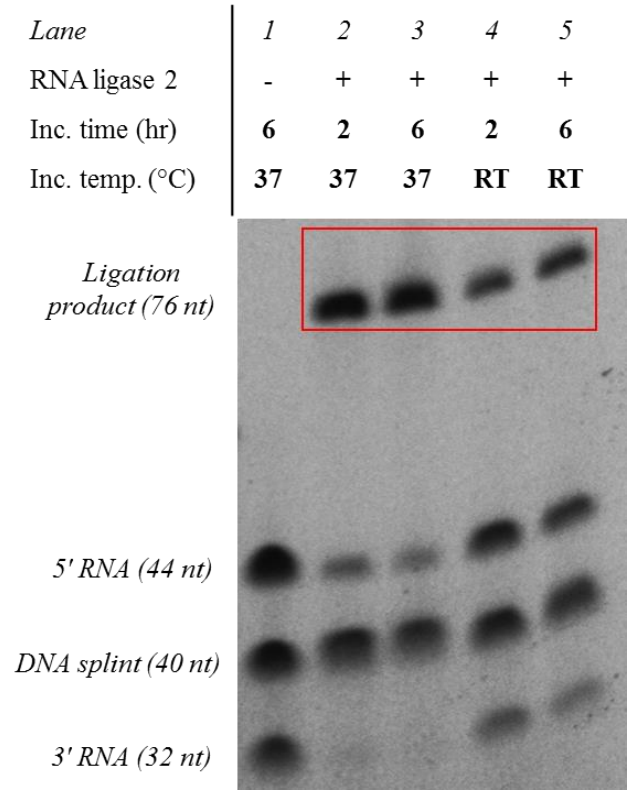


Figure 2-18. 6% PAGE gel comparing incubation (inc.) temperatures of RT and 37 °C, and incubation times of 2 and 6 hr. (Oligonucleotides: HisLuc-R1-e, HisLuc-R2-e, and HisLuc-S1)

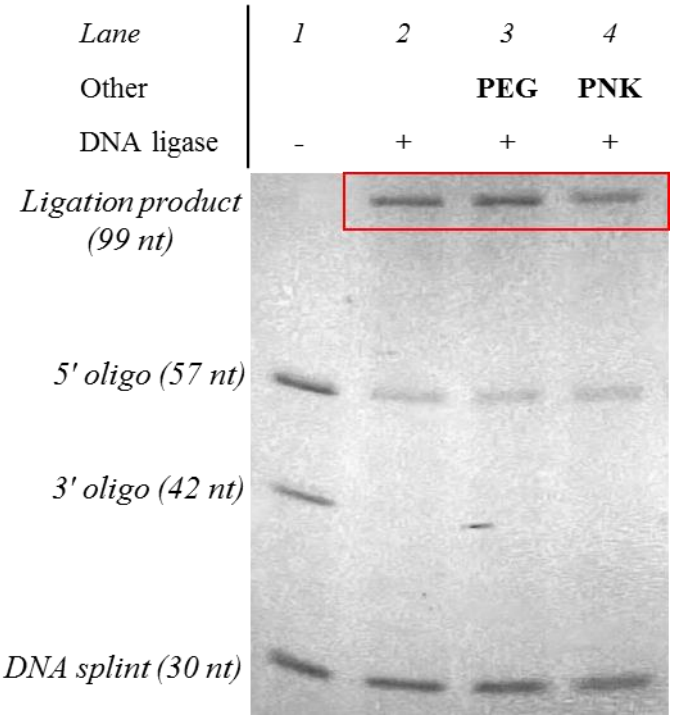


Figure 2-19. 12.5% denaturing PAGE gel investigating the effects of adding 6.7% PEG 6000 to the reaction mixture (**Lane 3**) and treating the oligonucleotides with PNK before ligation (**Lane 4**). (Oligonucleotides: ermCL-R1-e, ermCL-R2-e, and ermCL-S)

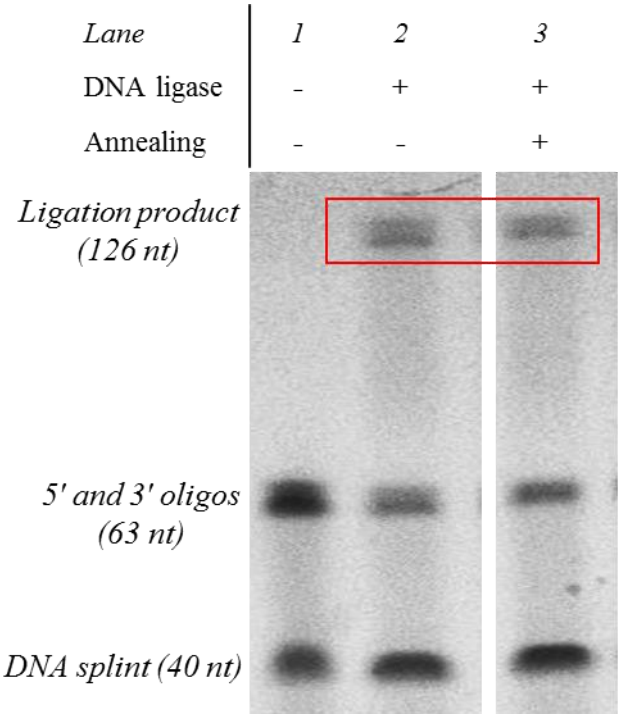
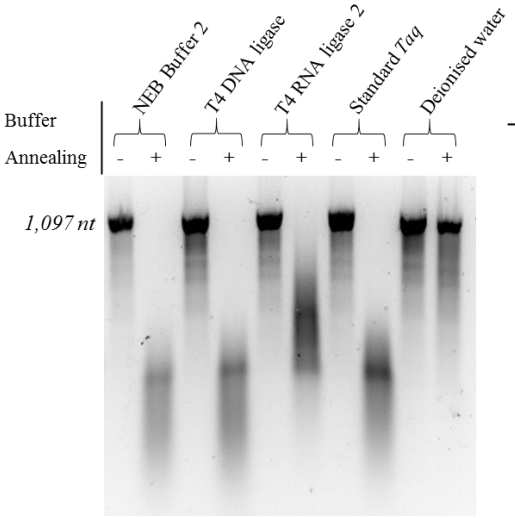


Figure 2-20. 6% denaturing PAGE gel investigating the effects of annealing the RNA and complementary splint prior to ligation. The annealing protocol was 2 min at 80 °C, followed by cooling at -1 °C until at 25 °C. (Oligonucleotides: RCT-enz-1, RCT-enz-2, and RCT-SA21-long)



Buffer	Components (1x)	pH at 25 °C
NEB buffer 2	50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT	7.9
NEB T4 DNA ligase buffer	50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT	7.5
NEB T4 RNA ligase 2 buffer	50 mM Tris-HCl, 2 mM $MgCl_2$2, 400 µM ATP, 1 mM DTT	7.5
NEB Standard *Taq* buffer	50 mM KCl, 10 mM Tris-HCl, 1.5 mM $MgCl_2$	8.3

Figure 2-21. 2% agarose gel showing annealing of a 1,097 nt transcribed RNA (HisLuc-R3) in different buffers. The annealing protocol was 5 min at 95 °C, followed by cooling at a rate of -1 °C until at 25 °C. Annealed samples were compared to samples kept at RT for the same amount of time and in the same buffers.

Enzymatic ligation yield of RNA was found to vary significantly by RNA length and sequence. Ligation yield was observed to decrease with increasing RNA

length. For example, yield for ligation of a 44 nt and 32 nt RNA oligonucleotide (76 nt final) was higher than yield for ligation of two 63 nt RNA oligonucleotides (126 nt final); yield for ligation of a 1,097 nt transcript to either a 32 nt or 76 nt oligonucleotide was extremely low (see Chapter 4), which is consistent with reports in literature on the ligation of kilobase-long RNA.[101] Possible reasons for this trend include more RNA impurity and degradation with increased length, formation of secondary structures in longer RNAs that obstruct the ligase, and increased complexity that might reduce the specificity of splint hybridisation.

Enzymatic ligation was also used to construct a circular mRNA template for rolling circle translation (RCT) from two sequential ligations of two 63 nt RNAs. The two 63 nt RNAs were first ligated to form a 126 nt linear RNA, which was purified by PAGE gel extraction. The linear RNA was then cyclised via intramolecular ligation to form a 126 nt circular RNA. A shorter 24 nt DNA splint was found to produce higher yield and fewer by-products than a longer 40 nt DNA splint during the cyclisation step (**Figure 2-22**).

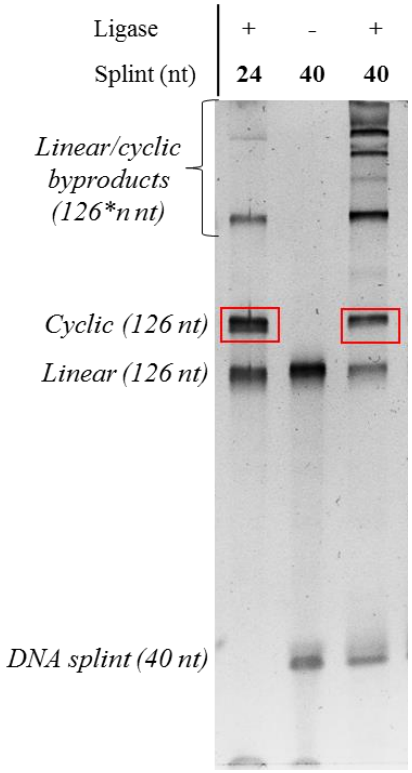


Figure 2-22. 6% denaturing PAGE showing intramolecular ligation of a 126 nt linear RNA (purified from the linear ligation of RCT-enz-1 and RCT-enz-2), resulting in cyclisation. The reaction also produced intermolecular ligation by-products that were 126*n nt in length, where n is a positive integer ≥ 2. Two different splint lengths of 24 nt (RCT-SA-12) and 40 nt (RCT-SA-12-long) were compared. Cyclisation was confirmed with an exonuclease T digestion assay (further described in Chapter 4).

2.5 Summary and conclusions of the study

Enzymatic ligation of both DNA and RNA were investigated, and several reaction conditions were tested to optimise ligation yield. First, DNA ligase was used to assemble the 762 nt eGFP gene from ten ssDNA oligonucleotide fragments templated with nine complementary splints. Ligation yield decreased with increasing numbers of oligonucleotides (or in other words, increasing numbers of ligation sites), especially above five oligonucleotides (four ligation sites). In general, changing the oligonucleotide to splint ratio, reaction buffer composition, annealing conditions, and incubation time and temperature did not significantly impact ligation yield. While T4

DNA ligase was used for the majority of studies, *Taq* ligase was also examined for use in thermocycling ligation reactions. T4 DNA ligase consistently produced higher yield when ligating five or more oligonucleotides, despite the ability to repeatedly melt and re-anneal the oligonucleotides in thermal cycles when using the thermostable *Taq* DNA ligase. To enzymatically assemble the 762 nt eGFP gene with high purity from ten DNA oligonucleotides, each half of the gene was first ligated separately from five oligonucleotides, PAGE-purified, and then ligated together to make the whole product. Ligation of DNA in a fully double-stranded duplex was also attempted to investigate the effects of double helix structure. The results for enzymatic ligation of dsDNA were similar to that of ssDNA, although the ligation products were more difficult to analyse and isolate by PAGE.

Next, magnetic beads were explored as a solid support for ligation reactions, and as a capture mechanism for ligated products. Switching from hydrophobic beads to hydrophilic beads with a smaller diameter increased loading capacity of the initial oligonucleotide, but yield of the complete ligation product remained low in both solid-support and capture cases. Steric hindrance was hypothesised to be a major obstruction preventing the hybridisation of DNA to oligonucleotides bound to the beads.

Finally, templated enzymatic ligation of RNA oligonucleotides was analysed. Yield decreased with longer RNA length, potentially due to increased impurities, degradation, and secondary structure. Incubation at 37 °C generally led to higher yields than incubation at RT, and longer incubation times improved yield in some cases. DNA and RNA ligases were compared, as well as DNA and RNA splints. Although RNA ligase 2 occasionally produced higher yield, DNA ligase was found to be more consistent and cost-effective. Using RNA splints to template the ligation

reaction resulted in extra bands on PAGE gels, so DNA splints were used to template both RNA ligase 2 and DNA ligase reactions. Both intermolecular linear ligation and intramolecular cyclic ligation of RNA were used to assemble mRNA templates for rolling circle translation (see Chapter 4). For intramolecular cyclic ligations, shorter splints of 24 nt produced higher yield of the desired product than longer splints of 40 nt.

In general, the conditions tested to optimise the enzymatic ligation reactions only resulted in modest improvements at best. The main limitation appeared to be the number of ligation sites, and also the length of the starting material in the case of RNA. This meant that to obtain large amounts of ligation product – e.g. enough to not require amplification by PCR, cloning, or other enzymatic processes – the reactions had to be scaled up. Scaling up the reaction would necessitate using correspondingly larger amounts of ligase enzyme, which could become prohibitively costly. Additionally, enzymes are proteins susceptible to denaturing and degradation. During the course of this study, stock tubes of ligase would occasionally lose their activity. It was not possible to distinguish inactive ligase prior to performing the reaction and analysing the results, leading to lost time and materials. In the next chapter, chemical ligations are explored as a cheaper, more easily scalable, and potentially more consistent alternative method for nucleic acid assembly.

Chapter 3: Chemical ligation of nucleic acids

3.1 Introduction and aim of the study

Chemical ligation of oligonucleotides is an attractive alternative to enzymatic ligation due to potential advantages including scalability, cost, compatibility with modifications, tolerance for a wide range of reaction conditions, and potential for one-pot assembly with orthogonal reactions. Two chemical reactions, copper(I)-catalysed azide-alkyne cycloaddition (CuAAC) and phosphoramidate ligation, have previously been used to successfully assemble DNA templates for replication and transcription. An unnatural backbone linkage is produced at the resulting ligation site, namely a triazole or phosphoramidate linkage, respectively (**Figure 3-1**). These linkages were shown to be biocompatible with DNA and RNA polymerases.[27, 28] However, much less is known about translation of chemically ligated RNA containing these unnatural backbones. The objectives of this study were to characterise and optimise chemical ligation of RNA, and to assemble mRNA templates with unnatural backbones for translation studies described in Chapter 4.

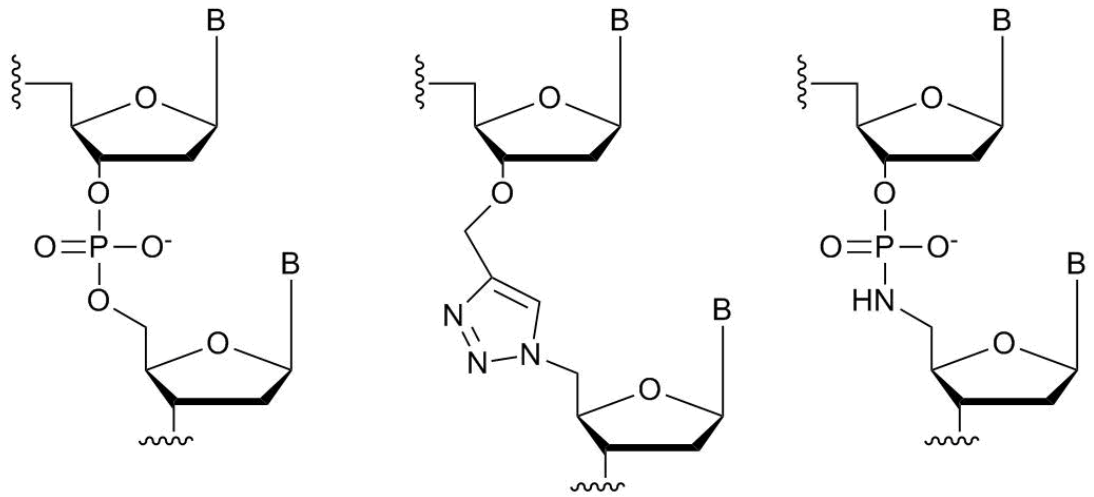


Figure 3-1. From left to right: natural phosphate backbone, triazole backbone (resulting from the CuAAC reaction), and phosphoramidate backbone in DNA. 'B' denotes any of the DNA bases.

3.2 Background

3.2.1 Copper(I)-catalysed azide-alkyne cycloaddition (CuAAC) reaction

Copper(I)-catalysed azide-alkyne cycloaddition (CuAAC) is perhaps the best known and most commonly used example of a "click chemistry" reaction, a term coined by K. Barry Sharpless.[103] Features of CuAAC that fulfil the requirements of click chemistry and make it attractive for nucleic acid ligation include: high yield, fast reaction rate, regiospecificity, physiological stability, no byproducts, ease of handling, and mild reaction conditions (e.g. aqueous solution at room temperature). In fact, templated CuAAC ligation of DNA has been reported to be faster and more selective than the same ligation with T4 DNA ligase.[103] Furthermore, azides and alkynes, the reactive groups in CuAAC, are rarely found in nature and thus make the reaction highly bioorthogonal.[105-109]

Ligation of two oligonucleotides via the CuAAC reaction requires one to have a 3′-propargyl group and the other to have a 5′-azido group. The former is obtained by solid-phase chemical synthesis using a solid-support that has 3′-propargyl monomer. The latter is introduced after completion of the synthesis, but before deprotection and cleavage from the resin. The 5′-hydroxyl group is first converted to a 5′-iodo intermediate with methyltriphenoxyphosphonium iodide, followed by treatment with a sodium azide solution to give a 5′-azide (**Figure 3-2**).

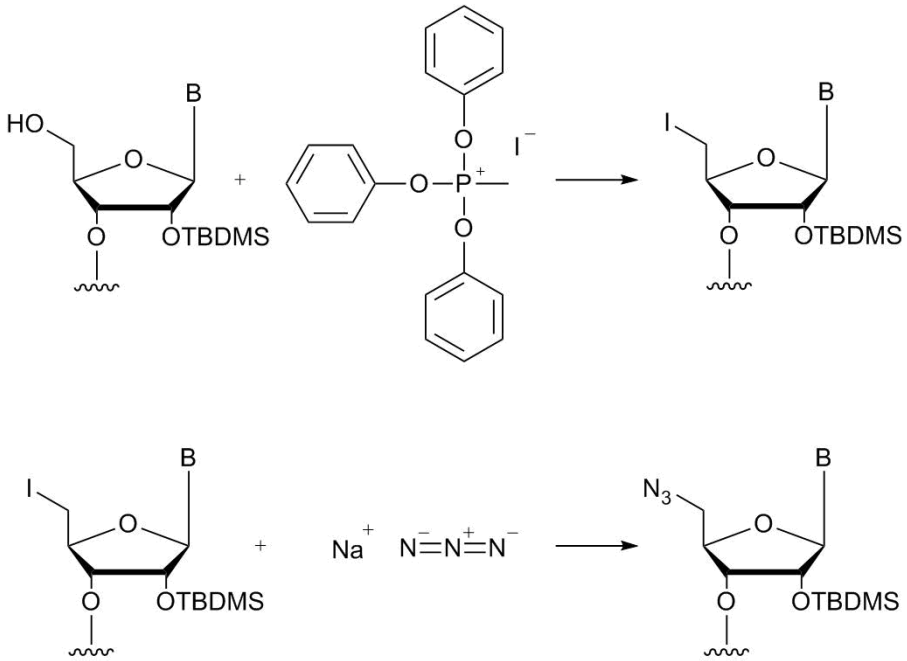


Figure 3-2. Conversion of an RNA 5′-hydroxyl group to a 5′-azido group. The hydroxyl group is first converted to an iodo intermediate with methyltriphenoxyphosphonium iodide, and then converted to an azido group with sodium azide. The bases (B) are still protected, as are the 2-hydroxyl groups with TBDMS.

During the CuAAC reaction, the Cu(I) catalyst is generated from Cu(II) (added in the form of $CuSO_4$) using sodium ascorbate as an *in situ* reducing agent. Water-soluble tris(3-hydroxypropyltriazolylmethyl)amine (THPTA) ligand (**Figure 3-3a**) is also added to stabilise the Cu(I) in aqueous solution and minimise its interaction with the biomolecules. The reaction is performed in anaerobic conditions at room temperature: the solution is degassed using argon to remove oxygen, which oxidises Cu(I) to Cu(II). The product of the CuAAC ligation reaction is a triazole backbone that links the 3′ and 5′ ends of the oligonucleotides that formerly contained the 3′-propargyl and 5′-azido groups (**Figure 3-3b**).[104-111]

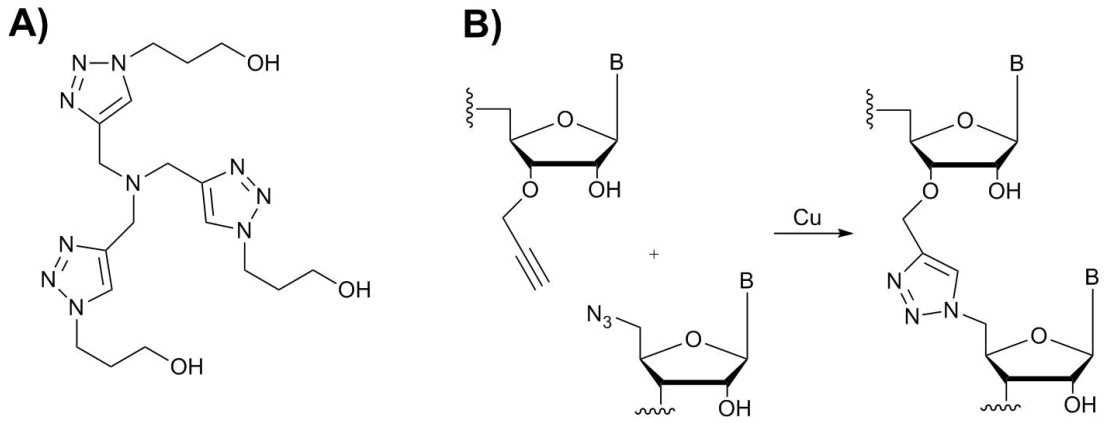


Figure 3-3. A) Chemical structure of the THPTA ligand. **B)** CuAAC ligation of oligonucleotides containing a 3′-propargyl group and a 5′-azido group results in an unnatural, biocompatible triazole backbone.

In 2011, Afaf El-Sagheer *et al*. published the design of a biocompatible triazole backbone that was accurately copied by DNA polymerases in PCR, linear copy, and rolling circle amplification (RCA) assays.[112] Shortly afterwards, successful *in vitro* transcription of a DNA template containing the biocompatible triazole in the coding region was demonstrated with T7 RNA polymerase, at approximately 77% yield compared to the native control. Interestingly, there was no RNA product when the triazole was located in the promoter region at the +4/5 position, perhaps due to inhibition of the DNA-protein binding necessary for transcription initiation.[113] The artificial triazole backbone was also found to be tolerated in both prokaryotic and eukaryotic living cells. Plasmids containing the triazole were functional in *E. coli* for both essential (ampicillin resistance) and non-essential (fluorescent protein) genes,[114] and triazole-containing DNA was successfully transcribed in human cells.[115] The versatility and biocompatibility of the linkage was further demonstrated by the one-pot assembly of a 335 nt iLOV fluorescent protein dsDNA gene containing eight triazole linkages (resulting from CuAAC ligation of ten DNA oligonucleotides, four linkages per strand) that was functional *in vitro* and in *E. coli*.[27]

Literature on RNA containing triazole backbones is more limited. Functional ribozymes have been assembled by CuAAC ligation of chemically synthesised RNA oligonucleotides, which enabled site-specific labelling with fluorophores and incorporation of DNA/RNA hybrid segments.[116] Tomoko Fujino *et al*. reported in 2016 cell-free translation of a 13 nt RNA template containing two triazole linkages, different from the one designed by the Brown group (**Figure 3-4**). The triazoles were located between the first/second and second/third bases of the second codon in the sequence, and the RNA coded for the dipeptide methionine-phenylalaline (Met-Phe): 5′- Gp ApUpGp **UtUt**Cp UpApAp GpCpU -3′, where 'p' indicates a natural phosphate backbone and 't' indicates a triazole. They found that although the 'correct target to misread' ratio was lower for the triazole-containing RNA than for the natural control (57:43 and 77:23, respectively), the total amount of Met-Phe obtained was two-fold higher for the triazole-containing template.[117] Taken together, these studies suggest that RNA with unnatural triazole backbones retain some functionality as ribozymes and mRNA templates for translation.

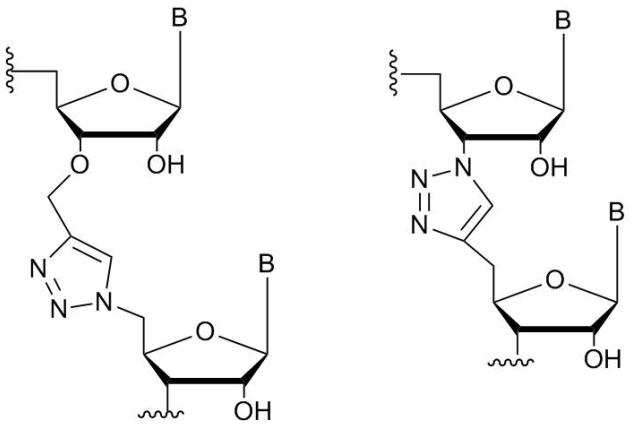


Figure 3-4. Left, triazole linkage from the Brown group. Right, triazole linkage from the Fujino group.

3.2.2 Phosphoramidate ligation reaction

Phosphoramidate ligation of oligonucleotides occurs between a 3′-phosphate and 5′-amino group in the presence of 1-ethyl-3-[3-(dimethylamino)propyl]carbodiimide (EDC) and 1-(2-hydroxyethyl) imidazole (**Figures 3-5a** and **3-5b**). It has been used to assemble DNA origami,[118] as well as DNA templates for amplification and transcription.[29] The reaction shares several similarities with CuAAC: it can occur in aqueous environments at room temperature, and results in an unnatural backbone at the ligation site (**Figure 3-5c**). In 2017, El-Sagheer and Brown reported one-pot assembly of the 762 nt eGFP gene from ten overlapping DNA oligonucleotides using solely phosphoramidate ligation. Like the biocompatible triazole backbone, the phosphoramidate linkage can be accurately read through by DNA and RNA polymerases. Conveniently, the two reactions are orthogonal, as demonstrated by simultaneous, one-pot ligation of three DNA oligonucleotides to create a 303 nt molecule containing one triazole and one phosphoramidate linkage.[29] However, phosphoramidate ligation of RNA has not been widely reported in literature. Furthermore, only the DNA dT monomer is currently commercially available with a 5′-amino group (**Figure 3-5d**).

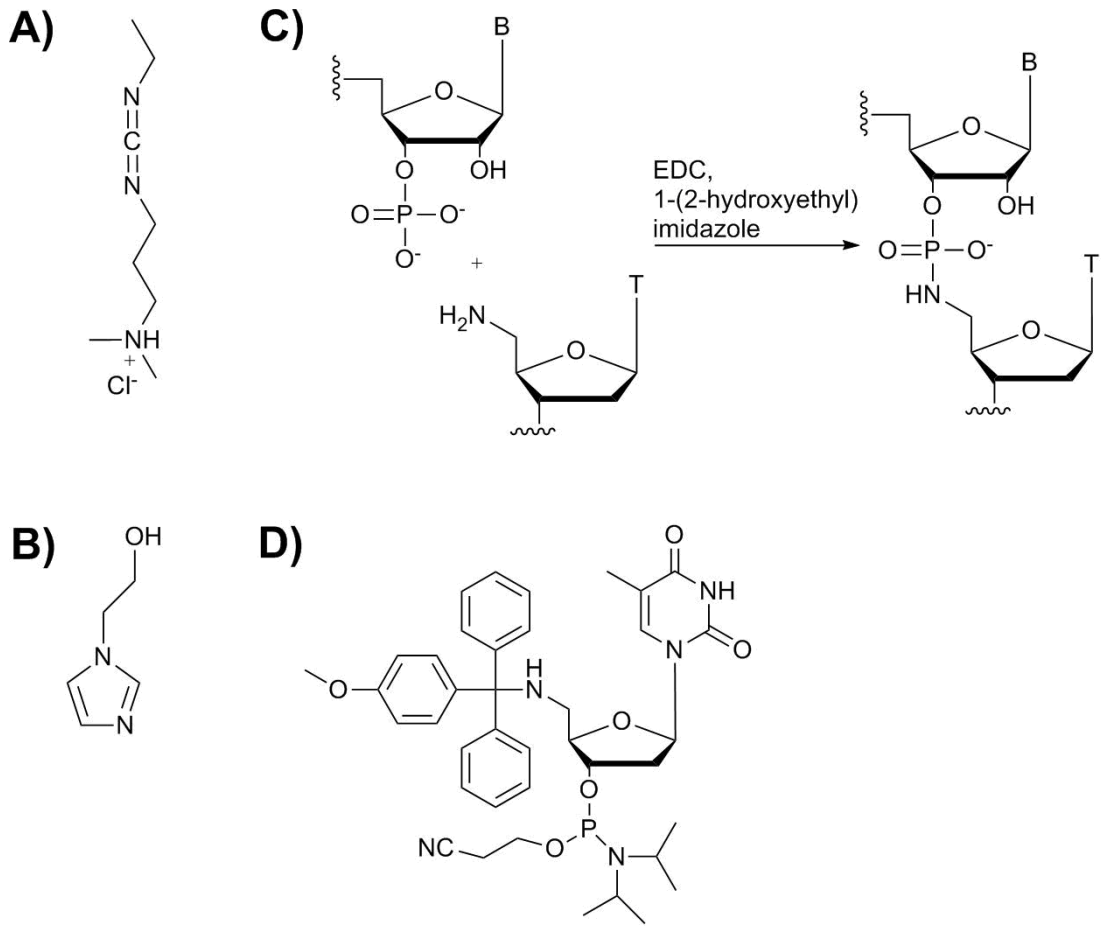


Figure 3-5. A) Structure of 1-ethyl-3-[3-(dimethylamino)propyl]carbodiimide hydrochloride (EDC-HCl). **B)** Structure of 1-(2-hydroxyethyl) imidazole. **C)** Phosphoramidate ligation of a 3′-phosphate group and a 5′-amino group results in an unnatural, phosphoramidate backbone. **D)** Protected 5′-amino dT phosphoramidite monomer. Currently, only dT is commercially available as a 5′-amino monomer.

Results and Discussion

3.3 CuAAC ligation

CuAAC ligation of chemically synthesised RNA oligonucleotides was optimised to make linear and circular mRNA templates for *in vitro* translation studies described further in Chapter 4 (**Figure 3-6**). Upon reaction of a 3′-propargyl and a 5′-azide, a triazole linkage was formed between the two oligonucleotide termini in place of the natural phosphodiester backbone. To stop the CuAAC reaction and remove the

reagents, buffer exchange was performed with either a NAP™ size exclusion column or spin filter column. Next, the samples were loaded onto a denaturing PAGE gel for analysis and/or purification. For downstream applications, bands corresponding to the ligation products were excised and extracted from the gel. The purified RNA products were then used in further ligation reactions (e.g. with ligase) or translation reactions. Mass spectrometry measurements of a 99 nt RNA ligation product purified by PAGE gel extraction showed that the measured mass (32,098) closely matched the expected mass (32,099), demonstrating the reliability of the CuAAC reaction (**Figure 3-7**; Appendix **Figure A13**).

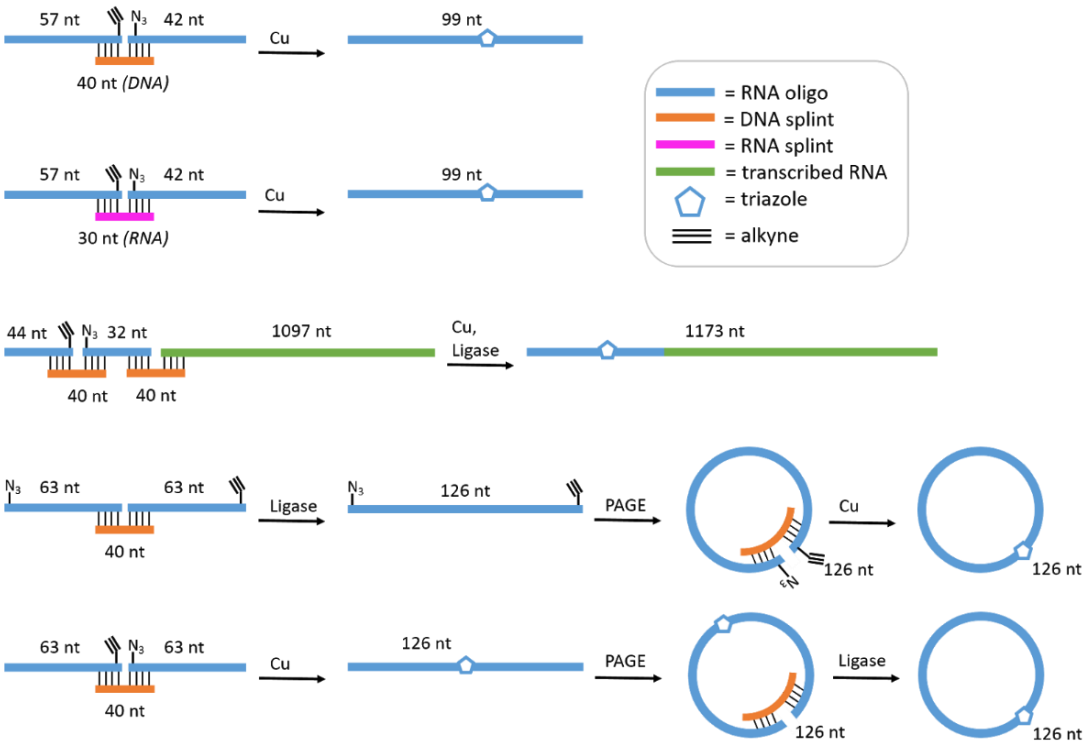


Figure 3-6. Various linear and circular RNA templates containing a triazole backbone were assembled with CuAAC, to be used in translation studies described in Chapter 4.

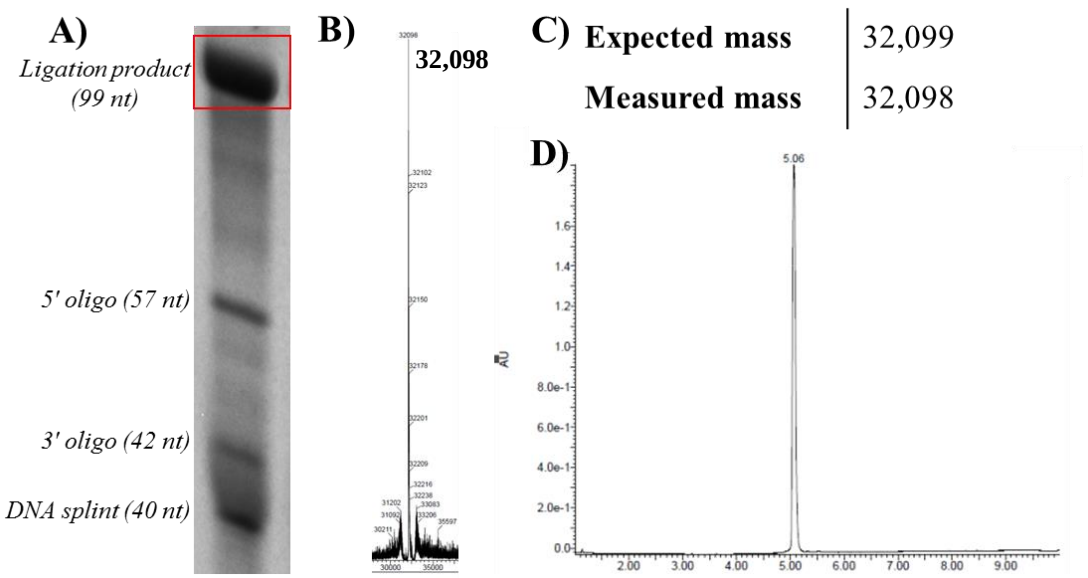


Figure 3-7. A) 12% denaturing PAGE gel showing CuAAC ligation of a 57 nt RNA (FLAGHA-R1-c) and a 42 nt RNA (FLAGHA-R2-c) to form a 99 nt product with a triazole backbone. The ligation product was extracted from the gel and analysed by LC-MS. **B)** Deconvoluted mass spectrum of the ligation product, showing a molecular mass of 32,098. **C)** The measured molecular mass of 32,098 was very close to the expected mass of 32,099. **D)** UV absorbance of the LC-MS chromatogram, showing a sharp single peak.

Both templated and untemplated CuAAC ligations were performed. In templated CuAAC, a splint complementary to both the 3′-propargyl and 5′-azide ends of the RNAs was used to hold them together in aqueous solution with 0.2 M NaCl to facilitate hybridisation. Previous work from the group had optimised the molar ratio of copper(II) pentahydrate, sodium ascorbate, and THPTA ligand to be 1:10:7.[112] Keeping that ratio constant, different amounts of copper were tested to optimise ligation yield. It was found that 100 equivalents (eq.) of copper to RNA performed better than 50 eq., but no significant differences were observed between 100, 200, and 500 eq. (**Figures 3-8** and **3-9**). For the typical CuAAC RNA ligation in this study, 200-250 eq. of copper were used. Additionally, doubling the ratio of splint to RNA oligonucleotide was compared to using equimolar amounts, and produced negligible differences in yield (**Figure 3-9**).

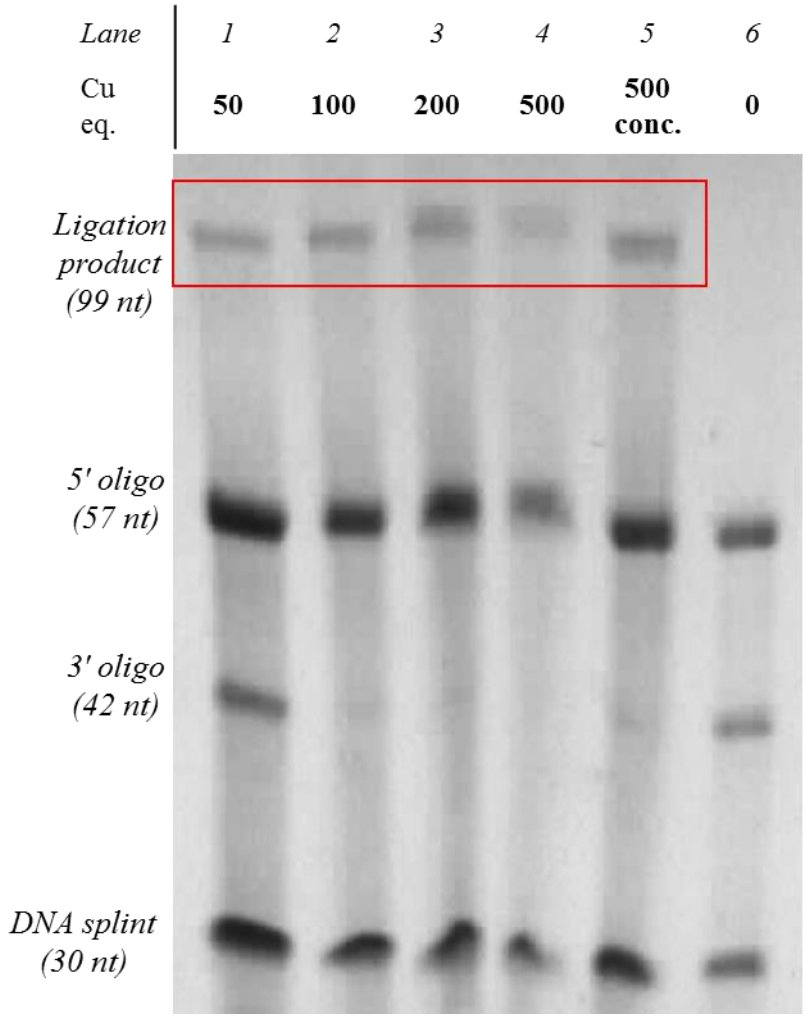


Figure 3-8. 12.5% denaturing PAGE gel analysing CuAAC yield using different molar equivalents of copper relative to RNA oligonucleotide. In **Lane 5**, 500 molar equivalents of copper were used, but the reaction volume was decreased so that the concentration of copper was increased relative to **Lane 4**. In **Lanes 2-5**, although the product bands are faint, the 3′ oligonucleotide is almost entirely consumed by the ligation reaction. (Oligonucleotides: ermCL-R1-c, ermCL-R2-c, and ermCL-S)

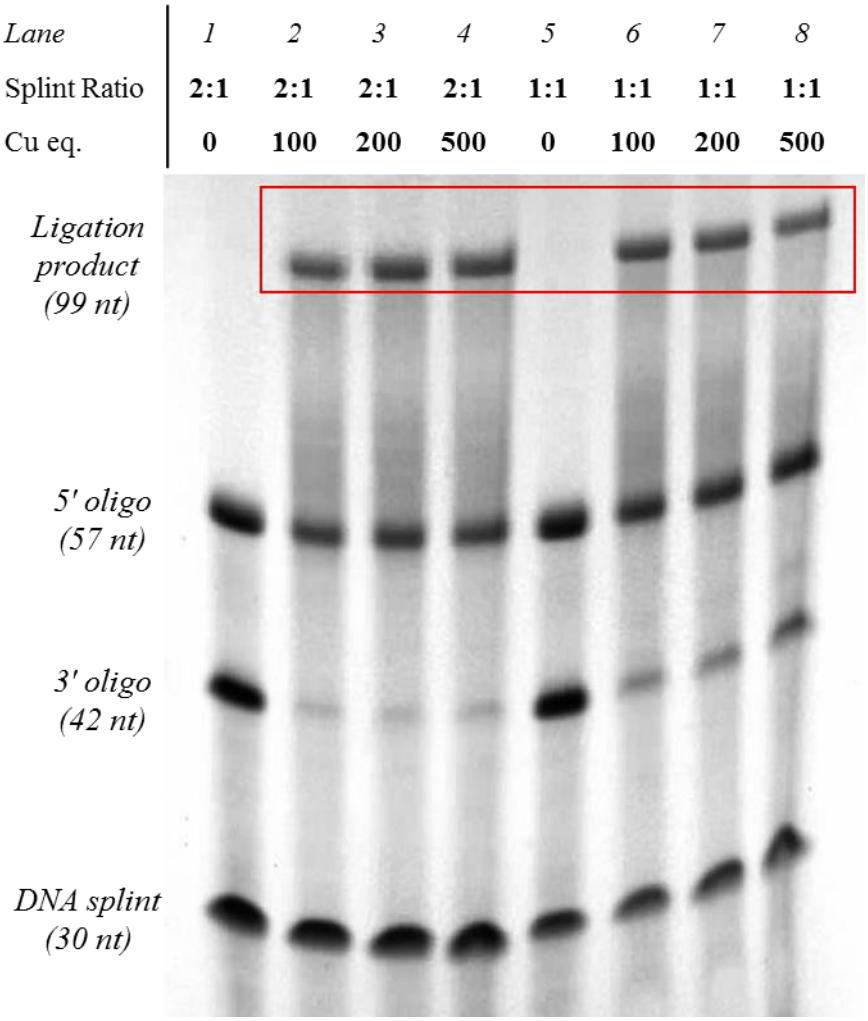


Figure 3-9. 12.5% denaturing PAGE analysing CuAAC yield using different molar equivalents of copper and splint relative to RNA oligonucleotide. (Oligonucleotides: ermCL-R1-c, ermCL-R2-c, and ermCL-S)

Templated reactions with DNA splints were also compared to ones with RNA splints. Reactions templated with an RNA splint resulted in two distinct bands by denaturing PAGE compared to one band for reactions templated with a DNA splint, with the extra band running slightly slower on the gel (**Figure 3-10a**). The same phenomenon was also observed in enzymatic ligations of RNA templated with RNA splints (Chapter 2.4), so it was not specific to the CuAAC reaction. In the negative control sample with no copper added, the RNA splint seemed to interact with the 5′-azide oligonucleotide to produce two bands. It was unclear whether the two bands represented incomplete denaturing, a side reaction, or some other phenomenon. LC-

MS analysis of gel-extracted ligation products templated by RNA splints showed two peaks on the UV chromatogram, compared to one peak for products templated by DNA splints (**Figure 3-10b** and **Figure 3-10c**). While the first peak contained the expected mass for the ligation product, the deconvoluted mass spectrum from the second peak was extremely noisy and did not produce a distinct mass measurement (Appendix **Figure A14**). Since yields were higher with DNA splints and DNA is easier to synthesise and handle, DNA splints were used for subsequent ligation reactions and the extra band was not investigated further. Untemplated reactions (i.e. without splints) were also performed in 50% DMSO to minimise RNA secondary structure formation, using a 1:5:2 molar ratio of copper, ascorbic acid, and THPTA ligand. Yields for untemplated reactions were slightly lower than for templated reactions with the same ratios of copper, ascorbic acid, and THPTA ligand (**Figure 3-11**).

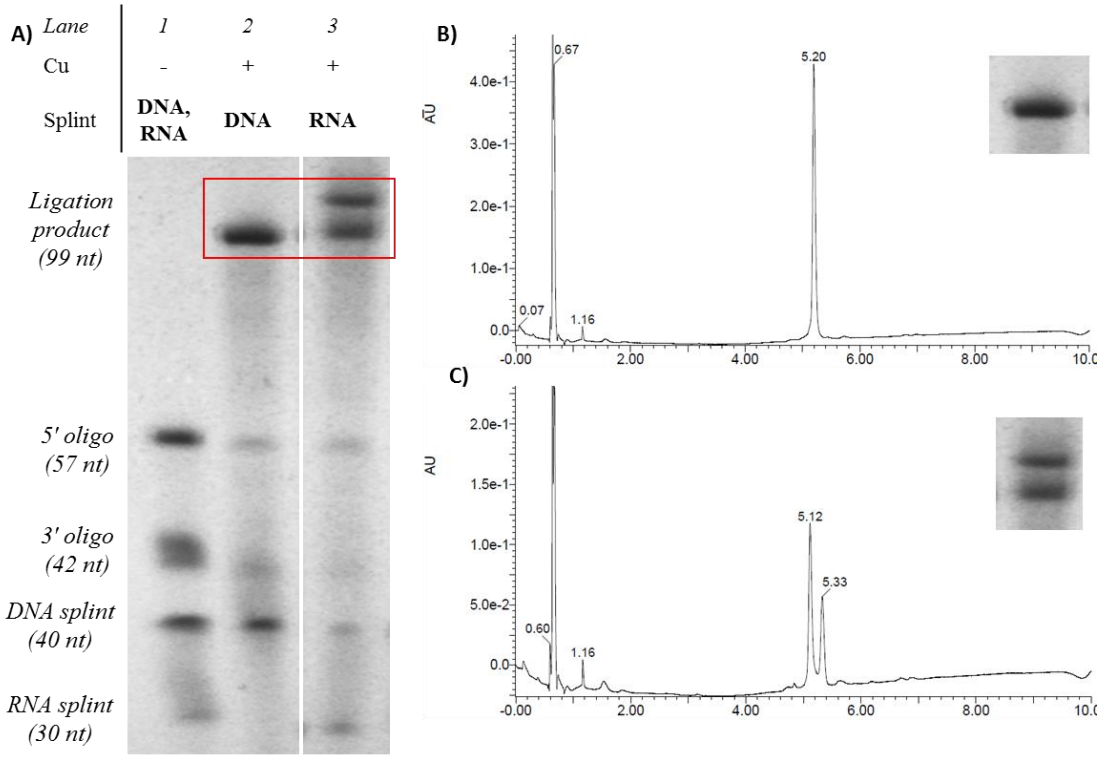


Figure 3-10. A) 12.5% denaturing PAGE gel comparing CuAAC ligation reactions templated with a 40 nt DNA splint (**Lane 2**) to reactions templated with a 30 nt RNA splint (**Lane 3**). In **Lane 1**, both DNA and RNA splints were added at half the concentration. **B)** UV chromatogram trace for CuAAC ligation templated with a 40 nt DNA splint, extracted from **Lane 2** of the gel. **C)** UV chromatogram trace for CuAAC ligation templated with a 30 nt RNA splint, extracted from **Lane 3** of the gel. (Oligonucleotides: FLAGHA-R1-c, FLAGHA-R2-c, FLAGHA-S and FLAGHA-S_{RNA})

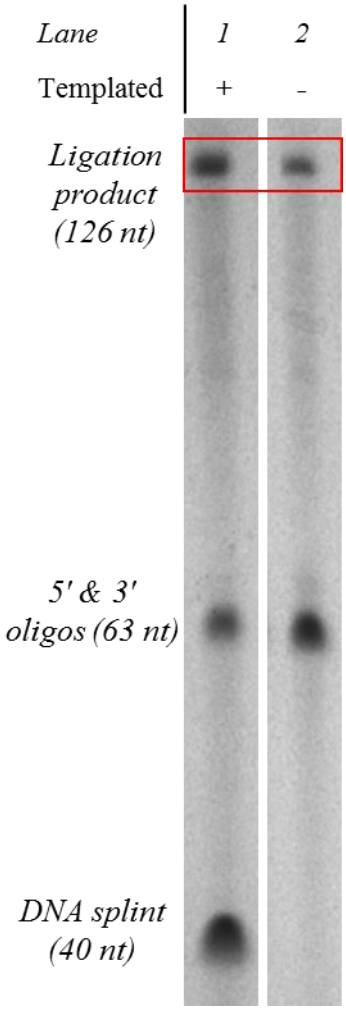


Figure 3-11. 8% denaturing PAGE gel comparing a CuAAC ligation reaction templated with a DNA splint (**Lane 1**) and an untemplated ligation reaction (i.e. without a splint) in 50% DMSO (**Lane 2**). The same ratios and amounts of copper, sodium ascorbate, and THPTA ligand were used for both reactions. (Oligonucleotides: RCT-CU31-1, RCT-CU31-2, and RCT-SA12-long)

3.4 Phosphoramidate ligation

Phosphoramidate ligation was explored as an orthogonal chemical reaction to CuAAC for ligating RNA. Reaction of a 3′-phosphate and 5′-amino group with EDC-HCl (hereafter referred to as EDC) and 1-(2-hydroxyethyl) imidazole results in a phosphoramidate linkage. This linkage is a minimal modification of the natural phosphodiester backbone, compared to the completely artificial triazole linkage from CuAAC. Previous work in the Brown group had shown that like the triazole linkage,

the phosphoramidate linkage can be read by both DNA and RNA polymerases to replicate and transcribe DNA, respectively.[29]

Initially, phosphoramidate ligation was attempted between an RNA oligonucleotide with a 3′-phosphate and another RNA oligonucleotide with a 5′-amino group, templated by a complementary DNA splint (note: the 5′-amino terminal nucleotide was dT, because it is the only commercially available monomer with a 5′-amino group). 1-(2-hydroxyethyl) imidazole and EDC were added to final concentrations of 0.1 M and 0.5 M, respectively, and the reaction was kept at room temperature for up to 2 hours. These conditions were optimised conditions published in literature for the phosphoramidate ligation of DNA.[29] However, the phosphoramidate reaction with RNA produced no detectable product, so reaction conditions were first investigated with DNA oligonucleotides. Keeping the ratio of 1-(2-hydroxyethyl) imidazole to EDC the same, final concentrations of 0.1 M/0.5 M, 1 M/5 M, and 5 M/25 M of 1-(2-hydroxyethyl) imidazole/EDC were tested for templated ligation of two DNA oligonucleotides (**Figure 3-12**). Final concentrations of 0.1 M 1-(2-hydroxyethyl) imidazole and 0.5 M EDC resulted in good ligation yield of the DNA, but higher concentrations produced faint and smeared bands on denaturing PAGE gels. Therefore, concentrations of 0.1 M 1-(2-hydroxyethyl) imidazole and 0.5 M EDC or lower were used for subsequent experiments.

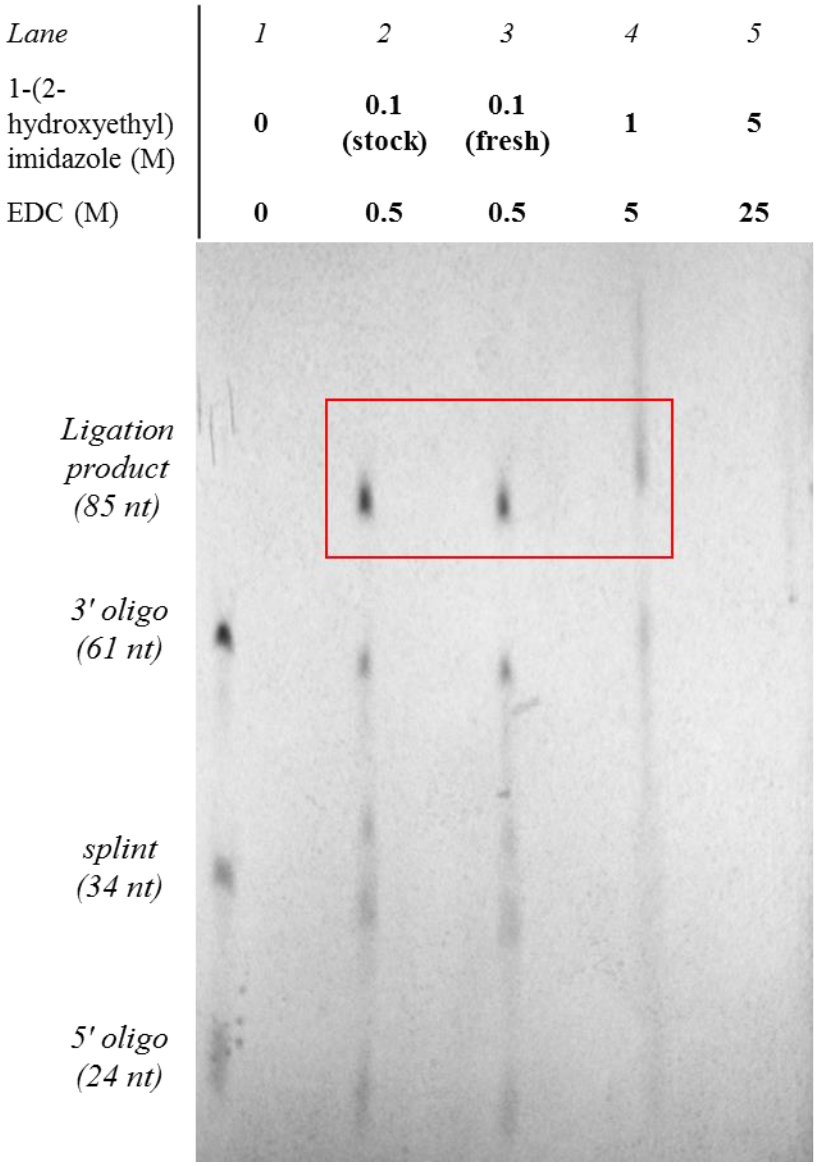


Figure 3-12. 8% denaturing PAGE gel showing phosphoramidate ligation of two DNA oligonucleotides with different concentrations of 1-(2-hydroxyethyl) imidazole and EDC. (Oligonucleotides: PA-DNA-1, PA-DNA-2, and PA-DNA-S)

Based on the good ligation yields for DNA but none for RNA using identical conditions, it was hypothesised that the 2′-hydroxyl group reacts with the terminal 3′-phosphate intramolecularly in the presence of the reagents, forming a ring that prevents ligation with a second RNA molecule (**Figure 3-13**). To investigate this issue, an RNA oligonucleotide was synthesised with a 3′-phosphate on a 3′-terminal DNA nucleotide (i.e. it lacked a terminal 2′-hydroxyl group). Phosphoramidate ligations with different ratios of 1-(2-hydroxyethyl) imidazole and EDC were

investigated between this RNA-DNA hybrid and a 5′-amino RNA oligonucleotide, and compared to phosphoramidate ligation between two DNA molecules. Based on the denaturing PAGE results, it appeared that the EDC concentration affected ligation yield more than the relative ratios of EDC to 1-(2-hydroxyethyl) imidazole, where lowering EDC concentration decreased yield. Ligation product was observed from the RNA oligonucleotide with a 3′ DNA nucleotide, but the yield was still poor for all conditions tested (**Figure 3-14**).

Hence, the phosphoramidate reaction was not pursued further. Additional optimisation of reaction conditions and protecting groups may improve yield enough to be practical for assembling RNA templates; that is left for future studies.

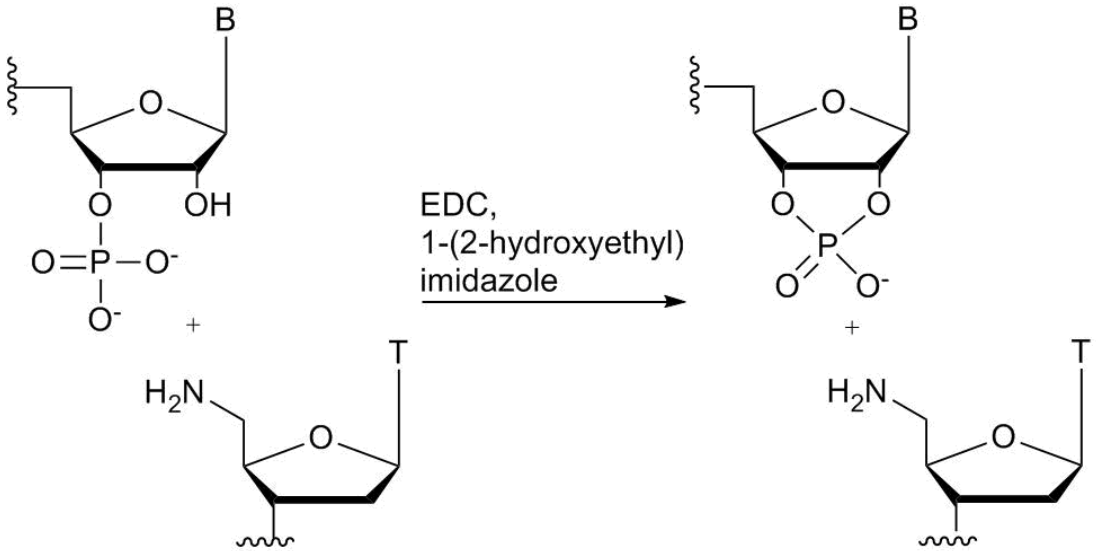


Figure 3-13. Hypothesised self-cyclisation of the terminal 2′-hydroxyl and 3′-phosphate groups in the presence of EDC and 1-(2-hydroxyethyl) imidazole, forming a ring that prevents ligation with a second RNA molecule.

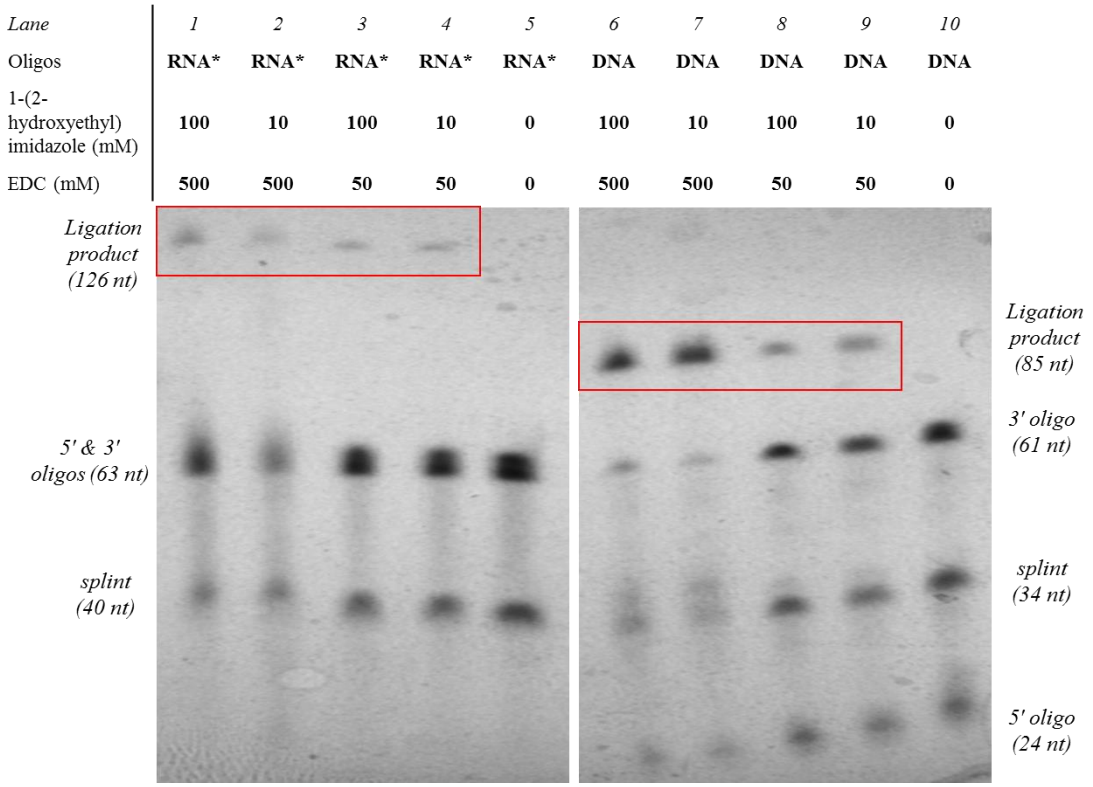


Figure 3-14. 8% denaturing PAGE gel showing phosphoramidate ligation of nucleic acids with different ratios and concentrations of 1-(2-hydroxyethyl) imidazole and EDC. In **Lanes 1-5**, 'RNA*'indicates that the 5′ RNA oligonucleotide has a DNA nucleotide at the 3′ terminus (i.e. no terminal 2′-hydroxyl group). **Lanes 6-10** show ligation of DNA with equivalent reaction conditions as a comparison. (Oligonucleotides: RNA*: PA-CU31-1*, PA-CU31-2, and RCT-SA12-long; DNA: PA-DNA-1, PA-DNA-2, and PA-DNA-S)

3.5 Summary and conclusions of the study

Two chemical reactions were investigated for assembling RNA oligonucleotides, CuAAC and phosphoramidate ligation. The attractive features of these methods include mild reaction conditions (e.g. aqueous solution and room temperature incubation) and relatively fast reaction times. Both ligation reactions produce an unnatural backbone linkage. Previous publications have demonstrated enzymatic (polymerase) replication and transcription of DNA with these unnatural

linkages, but very little work has been reported on translation of RNA containing these backbones.

The templated CuAAC ligation of RNA with DNA splints was optimised in aqueous 0.2 M NaCl solution with 200 equivalents of copper. Copper(II), sodium ascorbate, and THPTA ligand were added in a 1:10:7 molar ratio. Untemplated ligation was also evaluated in 50% DMSO, with a 1:5:2 molar ratio of copper(II), ascorbic acid, and THPTA. CuAAC ligation reactions were successfully used to assemble RNA templates containing an unnatural triazole linkage, to be used in translation reactions described in the next chapter. Yields were generally comparable to enzymatic ligations with T4 DNA ligase (see Chapter 4 for side by side comparisons).

Phosphoramidate ligation reactions were attempted as an orthogonal reaction to CuAAC, and as a chemical reaction to produce a close analogue to the natural phosphodiester backbone. However, phosphoramidate ligation of RNA did not yield enough product to be practical for downstream applications. Varying EDC and 1-(2-hydroxyethyl) imidazole ratios and concentrations did not significantly improve yield, nor did replacing the terminal 3′ RNA nucleotide with a DNA equivalent to avoid cross-reactions with the 2′-hydroxyl group. Due to the low yield for RNA assembly, phosphoramidate ligation was not investigated further in this study.

Chapter 4: Translation of triazole-containing mRNA

4.1 Introduction and aim of the study

In Chapter 3, CuAAC was shown to be a promising reaction for the chemical ligation of oligonucleotides. Previously, the Brown group had shown that DNA and RNA polymerases could read through the artificial triazole linkage that results from the CuAAC reaction in replication and transcription processes, respectively.[27, 112-115] That left translation as the remaining uninvestigated step in the core central dogma of molecular biology. The objective of this study was therefore to determine whether ribosomes can tolerate the unnatural triazole backbone in an mRNA template during translation. Although Fujino *et al*. had published a report on the translation of RNA containing a triazole backbone (see Section 3.2.1), the structure of their linkage was different from the triazole backbone developed by the Brown group.[117] Furthermore, their report used a very short 13mer RNA template containing two adjacent triazole backbones within the same codon. Because the motivation of this study was to evaluate chemical ligation as a method for oligonucleotide assembly in a biologically relevant context, RNA templates of around 100 nt or longer were used (i.e., beyond the length limit of solid phase oligonucleotide synthesis). The effects of the local sequence environment and triazole position relative to the codon reading frame were also investigated, with special interest in the wobble position of the codon.

Prokaryotic cell-free protein synthesis (CFPS) was chosen as the starting point, as it is a relatively simple translation system. RNA templates were constructed by enzymatic ligation for natural backbone controls, by CuAAC ligation for triazole-containing templates, or by a combination of the two. Six different template iterations were designed, and the general workflow of each experiment was to assemble the mRNA template by ligation, add it to CFPS reaction mixture, and analyse the translation reactions with a variety of assays.

4.2 Background

4.2.1 E. coli *RNA 5′ pyrophosphohydrolase (RppH)*

RNA 5′ pyrophosphohydrolase (RppH) converts the 5′-triphosphate of RNA to 5′-monophosphate by removing a pyrophosphate (**Figure 4-1**). In *E. coli*, RppH plays a role in mRNA degradation, since RNase E efficiency is greatly increased when the RNA 5′ terminus is a monophosphate. RppH activity can be inhibited by secondary structure at the 5′ terminus, such as a stem-loop. Furthermore, *E. coli* RppH requires at least two, preferably three or more, unpaired bases at the 5′ end. The enzyme can accept any unpaired 5′ sequence, although a modest preference for G at position +2 has been reported. In this study, RppH was used to convert the terminal 5′-triphosphate to monophosphate in *in vitro* transcribed RNA, so that the RNA could be accepted as a substrate by ligase enzymes.[119-123]

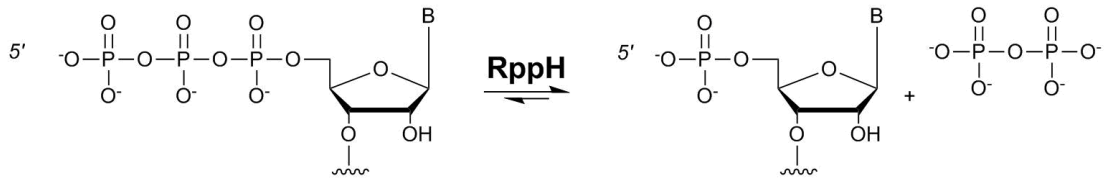


Figure 4-1. Conversion of an RNA 5′-triphosphate to monophosphate by RppH.

4.2.2 T7 RNA polymerase transcription

Bacteriophage T7 RNA polymerase (T7RNAP) has several advantages that have made it a commonly used tool for *in vitro* transcription of RNA. A relatively simple single unit of 99 kDa, the enzyme is both accurate and efficient. Importantly, the polymerase very specifically and strictly initiates transcription downstream of a double-stranded promoter sequence, TAATACGACTCACTATAGGG, with the underlined G as the first base of the RNA transcript. T7RNAP then uses the opposite strand (antisense strand) as the template, transcribing in the 5′ to 3′ direction (**Figure**

4-2). While the promoter sequence has to be double-stranded, the rest of the DNA template can be either single- or double-stranded. This feature of T7RNAP enables well-defined transcription of an ssDNA template by hybridising a DNA oligonucleotide containing the promoter sequence. Double-stranded linear PCR products or circular plasmids containing the promoter sequence can also be transcribed by T7RNAP. As mentioned in Chapter 1, rho-independent transcription termination occurs in bacteria by the formation of a hairpin structure in the new RNA template. T7RNAP transcription can therefore be halted at a predetermined site by the inclusion of a hairpin-forming terminator sequence in the DNA template. A U-rich sequence following the hairpin has also been reported to enhance termination. Together with the hairpin structure, the U residues destabilise interactions between the enzyme, DNA template, and RNA transcript, leading to transcription termination. At the end of the transcript, T7RNAP can add a few extra untemplated bases, resulting in heterogeneous 3′ ends. In this study, *in vitro* transcription with T7RNAP was used to obtain long RNA molecules that would have otherwise required ligation of many chemically synthesised RNA oligonucleotides.[124-129]

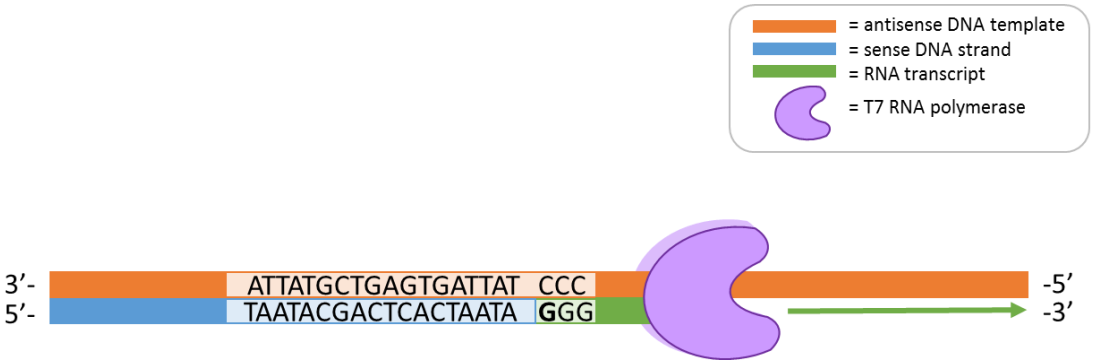


Figure 4-2. Transcription of a DNA template by T7 RNA polymerase. The promoter sequence is shown in blue, and the first nucleotide of the RNA transcript is the bolded **G** (in green). The RNA polymerase uses the antisense DNA strand (orange) as a complementary template to synthesise the RNA transcript.

4.2.3 Cell-free protein synthesis (CFPS)

Cell-free protein synthesis (CFPS) is an *in vitro* method for polypeptide translation using extracted or recombinant cellular components outside of a living cell. While not currently suitable for large-scale protein production, CFPS has several advantages over *in vivo* protein synthesis. First, it is generally much faster and simpler than *in vivo* methods, especially with the availability of commercial CFPS kits. The lack of cell walls and membranes allows for direct access to and control over the translation environment. Additionally, cell viability and homeostasis do not have to be taken into account. Taken together, these features of CFPS enable high-throughput testing of a wide range of conditions, addition of external factors such as unnatural amino acids, and synthesis of toxic proteins.[130-132]

The most common prokaryotic CFPS system is based on *E. coli* cells, and the most common eukaryotic ones are based on rabbit reticulocyte, wheat germ, and insect cells. Each has their own advantages and disadvantages. In general, prokaryotic systems are simpler and tend to be higher-yielding, while eukaryotic systems are capable of post-translational modifications.[133] CFPS systems often contain components for both transcription and translation, also referred to as coupled systems. T7RNAP is regularly used for the transcription reaction in coupled CFPS. DNA, especially in the form of circular plasmids, is typically much more degradation-resistant than RNA. Thus, a major advantage of using DNA templates in coupled CFPS is that fresh mRNA is continually being synthesised. A greater amount of exogenous RNA template than of DNA template is usually needed to achieve the same CFPS yield as a result. Furthermore, transcription and translation naturally occur simultaneously in prokaryotes, which lack membrane-bound nuclei. Multiple

ribosomes can begin translating the 5′ end of the mRNA while the 3′ end is still being transcribed, thereby reducing RNA degradation and increasing protein yield.[129, 134]

Protein synthesis using recombinant elements, or PURE, is a CFPS system developed by Takuya Ueda's group that has since been commercialised. The PURE system consists of only the components necessary for T7 transcription and *E. coli* translation, which have been highly purified to remove endogenous nucleases and proteases. This increases translation yield by minimising RNA and protein degradation. CFPS kits based on the PURE platform were used to investigate translation of triazole-containing mRNA in this study.[135]

4.2.4 Wobble phenomenon

The genetic code is degenerate but not ambiguous. There are $4^3 = 64$ possible codons for 20 standard amino acids and a stop signal, and each codon only maps to a single amino acid or stop signal. If standard base pairing rules were strictly applied to all three positions in the codon, organisms would require at least 64 distinct tRNA species. However, the wobble phenomenon takes codon degeneracy into account such that the same tRNA molecule can bind to multiple mRNA codons.

In 1966, Francis Crick published his 'Wobble Hypothesis,' based on logical deductions from structural knowledge and experimental data showing that some tRNAs could recognise multiple codons that differed in their third base. He proposed that while the first two base pairs of an mRNA codon follow strict Watson-Crick base pairing rules with its complementary tRNA, there is a certain amount of 'wobble' in the third one that allows for non-standard base pairing. The structure formed by the first two Watson-Crick base pairs imposes constraints that ensure the third base pair is also structurally similar to that of a Watson-Crick base pair, but allows for minor

deviations within those limits. These constraints lead to wobble base pairing patterns that are repeatedly observed in experimental data.[1, 2, 136] **Table 4-1** below lists the wobble base pairing rules that Crick proposed, as well as updated ones from Frank V. Murphy IV and Venki Ramakrishnan (2004).[137, 138] Practically, the wobble hypothesis implies that there is a degree of conformational flexibility tolerated in the third position of the codon (**Figure 4-3**).

Table 4-1. Original wobble base pairing rules proposed by Crick (left), and updated rules proposed by Murphy and Ramakrishnan (right). Non-Watson-Crick base pairs are underlined. I, inosine; k^2C, lysidine; xm^5s^2U, 5-methyl-2-thiouridine derivatives; xm^5Um, 5-methyl-2′-O-methyluridine derivatives; Um, 2′-O-methyluridine; xm^5U, 5-methyluridine derivatives; xo^5U, 5-hydroxyuridine derivatives. Taken from Cochella & Green (2004).[141]

Crick (1966)		**Murphy & Ramakrishnan (2004)**	
Anticodon (tRNA)	**Codon (mRNA)**	**Anticodon**	**Codon**
G	C, U	G	C, U
C	G	C	G
		k^2C	A
A	U	A	U, C, G, (A)
U	A, G	U	A, U, G, (C)
		xm^5s^2U, xm^5Um, Um, xm^5U	A, (G)
		xo^5U	A, U, G
		I	A, C, U

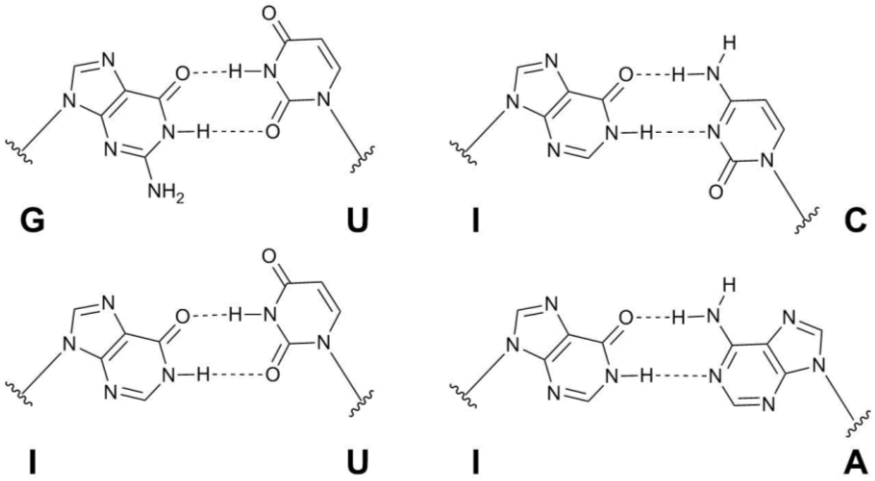


Figure 4-3. Examples of non-canonical wobble base pairs, clockwise from top left: G-U, inosine (I)-C, I-A, and I-U.

Based on this phenomenon, it was hypothesised that an unnatural backbone linkage would be best tolerated either immediately in front of or behind the wobble base pair of a codon. Therefore, mRNA templates were tested in which the triazole was located between the second and third bases of a codon (the '23' position) or between the third base of a codon and the first base of the next codon (the '31' position). At first, attention was focused primarily on the '31' position, as it was thought that positioning the triazole in between codons would be more favourable than positioning it within a codon. The '12' position was not tested, in which the triazole would be located between the first and second bases of a codon, because of the strict Watson-Crick base-pairing requirements as explicated by the wobble hypothesis.

4.2.5 Immunoassays: ELISA, dot blot, western blot

Immunoassays utilise the highly specific binding abilities of antibodies to detect and measure target molecules. The analyte can be detected directly by primary antibodies that recognise the target molecule itself; it can also be detected indirectly

by secondary antibodies that recognise a primary antibody bound to the analyte. The second mechanism often results in signal amplification, since multiple secondary antibodies can bind to a single primary antibody. Reporter molecules are usually conjugated to the antibodies to provide an output signal, such as fluorophores or enzymes that catalyse chemiluminescent reactions. In this chapter, three types of immunoassays are used: enzyme-linked immunosorbent assays (ELISAs), dot blots, and western blots. All three of these immunoassays provide information about the presence and relative quantities of the target molecule, and western blots also provide additional data about size.

ELISAs enable specific detection and semi-quantification of target molecules by adsorption of either the analyte or antibody to polystyrene multi-well plates, followed by several washing steps (**Figure 4-4**). In the first method, the analyte is incubated in the wells to allow passive adsorption to the plate, and then the wells are washed to remove anything not adsorbed. Next, a blocking buffer is applied to block remaining surfaces and prevent non-specific binding. The primary antibody is then added, which binds to the analyte adsorbed to the plate. Another wash step removes excess and unbound antibody. Signal can be read from the primary antibody (direct ELISA), or a secondary antibody that recognises the primary antibody can be applied, washed, and detected (indirect ELISA). In a sandwich ELISA, a capture antibody specific for the analyte is first adsorbed to the wells. The wells are washed, the samples are applied, and any target molecules in the samples bind to the adsorbed capture antibodies. After blocking any remaining sites on the capture antibodies, another primary antibody that is not reactive with the capture antibody is added. This antibody also recognises the target molecule and binds to analyte already bound to the capture antibodies, resulting in an antibody-analyte-antibody ‘sandwich.’ If this

antibody is labelled, it operates as the detection antibody. Otherwise, a labelled secondary antibody specific for the primary antibody on the free end of the 'sandwich' can be used to provide the signal output. Sandwich ELISAs are generally more robust than ELISAs in which the analyte is adsorbed to the plate, since two different primary antibodies are used. The sandwich ELISAs performed in this study utilised fluorophore-labelled antibodies with different excitation and emission wavelengths, and the assays were analysed using a plate reader that detected fluorophore emission (note: even though fluorophore-labelled antibodies do not involve enzymes, the term ELISA is often used to encompass assays of this type regardless).[139-142]

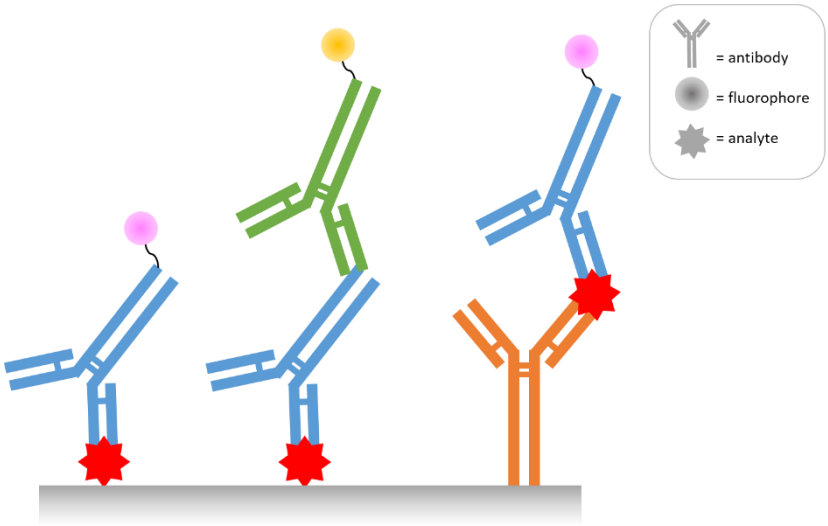


Figure 4-4. Types of ELISAs, from left to right: direct, in which analyte adsorbed to the plate is detected by a primary antibody; indirect, in which analyte adsorbed to the plate is detected by a secondary antibody bound to the primary antibody; and sandwich, in which the analyte is 'sandwiched' between capture and detection antibodies.

Dot blots are similar in idea to ELISAs, in that labelled antibodies are used to detect immobilised analyte. First, a drop of the analyte sample is applied to an activated polyvinylidene difluoride (PVDF) membrane. The sample components bind

to the membrane as it dries. The membrane is then blocked with blocking buffer to prevent non-specific binding, and incubated with a primary antibody that recognises and binds to the target molecule. After washing off unbound primary antibody, a labelled secondary antibody that binds to the primary antibody is added. Secondary antibodies labelled with horseradish peroxidase (HRP) were used for the dot blots in this study. HRP catalyses a chemiluminescent reaction upon addition of luminol and peroxide. The intensity of the chemiluminescence can be captured by camera in an imager and compared to determine relative quantities of the target molecule.[143-145]

Western blots follow the same procedure as dot blots, except the analyte sample is first separated by size on an SDS-PAGE gel (**Figure 4-5**). The samples are then transferred to a PVDF membrane electrically, retaining their size separation. Once the analytes are transferred to the membrane, western blots essentially follow the same blocking, primary antibody, secondary antibody, and detection steps as dot blots.[146-148] The same HRP-labelled secondary antibodies used for the dot blots were also used for the western blots in this study.

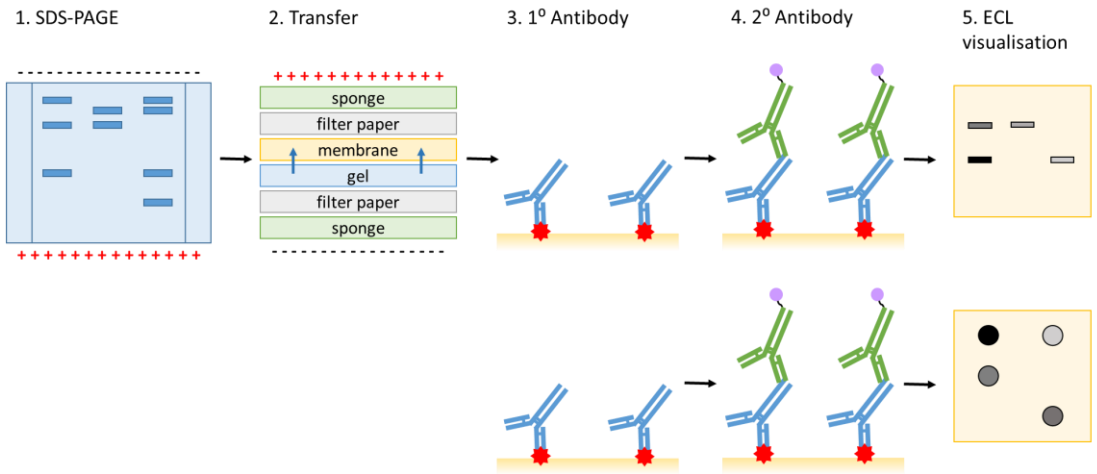


Figure 4-5. Top: a schematic for western blots. First, the protein sample is separated by size by SDS-PAGE. The proteins are then transferred to a membrane. The membrane is incubated with primary antibody, followed by HRP-conjugated secondary antibody. The secondary antibody is visualised by enhanced chemiluminescence (ECL). **Bottom:** a schematic for dot blots. The antibody incubation and visualisation processes are the same as for western blots, but the protein sample is spotted directly onto the membrane without any size separation.

Results and Discussion

4.3 CPFS of ermC leader peptide mRNA

CFPS with a triazole-containing RNA was first attempted with a 99 nt mRNA template for the bacterial ermC leader (ermCL) peptide (**Tables 4-2** and **4-3**). The template was based off of work published by Thomas Hoernes *et al.* (2016), in which they used the PURExpress® bacterial CFPS system to translate synthetic mRNA.[149] This sequence was chosen because its length allowed the whole template to be assembled by ligating two chemically synthesised RNA oligonucleotides. The complete sequence included a 5′ untranslated region (UTR) with a ribosome binding site (RBS), in addition to start and stop codons within the translated region. The ligation site was designed to be in between codons (‘31’ position), located after the sixth codon in the translated region. The mRNA templates were constructed by ligating the RNA either enzymatically or by CuAAC. Ligation reactions were

analysed by denaturing PAGE, and the correct ligation product bands were excised and extracted from the gel. After desalting and resuspending in water, RNA concentration was measured by UV absorbance (**Figure 4-6**).

Table 4-1. Oligonucleotide sequences for the assembly of the ermC leader peptide template. The RBS is underlined, and the start and stop codons are in bold. ermCL-R1 was ligated to ermCL-R2 to form the complete RNA template, and ermCL-S was a complementary DNA splint to template the ligation reaction. For CuAAC, ermCL-R1 had a 3′-alkyne (alk) group and ermCL-R2 had a 5′-azide (az) group; for the enzymatic ligation, ermCL-R1 had a 3′-OH group and ermCL-R2 had a 5′-PO_4 group.

Name	Length (nt)	Sequence (5′ → 3′)
ermCL-R1	57	UAUUUUAAUAAUUAAUUAACUUUAGUAAGGAGGAACAAU **AUG**GGCAUUUUUAGUAUU *–alk/OH*
ermCL-R2	42	*az/PO_4–* UUUGUAAUCAGCACAGUUCAUUAUCAACCAA ACAAAAAA**UAA**
ermCL-S	30	TGTGCTGATTACAAAAATACTAAAAATGCC

Table 4-2. The translation product of the ermC leader peptide mRNA template.

Name	Length (aa)	MW (kDa)	Sequence
ermCL	19	2.2	MGIFSIFVISTVHYQPNKK

Translation reactions were performed by adding PAGE-purified RNA to PURExpress® CFPS mixture (New England Biolabs) for a final concentration of approximately 10 μM. The reactions were incubated for 4 hr at 37 °C, as suggested by the manufacturer. The original plan was to analyse the translation reactions by mass spectrometry and Coomassie-stained Tris-Tricine SDS-PAGE. However, no translation product was detected from either the enzymatically-ligated or CuAAC-ligated templates by mass spectrometry (Appendix **Figure A15**). Difficulties were also encountered when trying to visualise test peptides of approximately the same molecular weight as the ermCL peptide on Tris-Tricine SDS-PAGE gels (16% resolving gel) with Coomassie staining, thought to be due to the small size of the

peptide. In the Hoernes paper on which the ermCL RNA sequence was based, translation products were visualised by incorporating radioactive amino acids during CPFS and analysing the reactions on 16% Tris-Tricine SDS-PAGE gels.[149] Radioactive detection is much more sensitive than Coomassie staining, but it was avoided in this study due to safety concerns. Therefore, a new mRNA template was designed that could take advantage of other sensitive assays suitable for detecting small peptides.

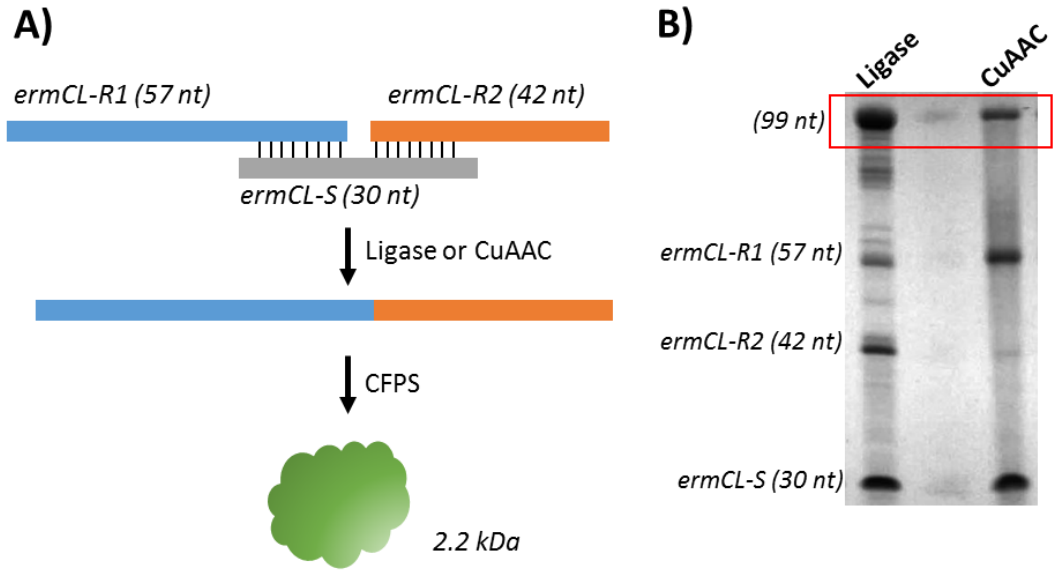


Figure 4-6. A) Schematic for the assembly and translation of the ermCL peptide mRNA template. **B)** 12.5% denaturing PAGE showing enzymatic and CuAAC ligation of the 99 nt ermCL mRNA template.

4.4 CFPS of FLAG-HA peptide mRNA

An mRNA template coding for a peptide containing a FLAG and an HA tag with a linker in between was designed for CFPS. Both FLAG and HA tags are well-established epitope tags with widely available antibodies. This template was designed to take advantage of the sensitivity and specificity of antibodies in order to detect

translation products in immunoassays. Namely, an anti-FLAG antibody could be used to capture the peptide, and an anti-HA antibody could be used to detect the captured peptide in a sandwich ELISA (and vice versa). The RNA template was designed so that the ligation site was in the region coding for the linker, in the '31' position after the 11th translated codon. This meant that theoretically the ribosome would read and translate the FLAG tag before reaching the triazole backbone, and would only translate the HA tag after going through the triazole. Again, the RNA template contained an RBS, a start codon, and a stop codon (**Tables 4-4** and **4-5**).

Table 4-4. Oligonucleotide sequences for the assembly of the FLAG-HA peptide template. The RBS is underlined, and the start and stop codons are in bold. FLAGHA-S_{RNA} is an RNA splint, whereas FLAGHA-S is a DNA splint.

Name	Length (nt)	Sequence (5′ → 3′)
FLAGHA-R1	57	UUAACUUUAUAAGGAGGUAACAAU**AUG**GAUUAUAA AGAUGACGAUGACAAAGCGGGC *–alk/OH*
FLAGHA-R2	42	*az/PO$_4$–* CUGGGUGCGGGUUAUCCGUACGAUGUG CCGGAUUAUGCC**UAA**
FLAGHA-S_{RNA}	30	AUAACCCGCACCCAGGCCCGCUUUGUCAUC
FLAGHA-S	40	TACGGATAACCCGCACCCAGGCCCGCTTTGTCATC GTCAT

Table 4-5. The translation product of the FLAG-HA peptide mRNA template. The FLAG and HA tags, connected by a linker region, are underlined.

Name	Length (aa)	MW (kDa)	Sequence
FLAGHA	24	2.6	MDYKDDDDKAGLGAGYPYDVPDYA

The 99 nt RNA template was assembled either by enzymatic or CuAAC ligation of two RNA oligonucleotides. CFPS reactions were also performed as described in section 4.3, and the translation reactions were then purified with Vivaspin 2® spin-filter columns (Sartorius) with a 5000 molecular weight cut-off to remove ribosomes, T7RNAP, and other CFPS components. Sandwich ELISA was performed by coating black 96-well solid bottom plates (Porvair Sciences) with anti-

FLAG antibodies conjugated with Alexa 488 fluorophores, or anti-HA antibodies conjugated with Alexa 647 fluorophores (Life Technologies). CFPS reaction samples were then added, followed by a blocking step and incubation with the other antibody. The plate was imaged at the excitation and emission wavelengths of the two fluorophores to see if any translated FLAG-HA peptide could be detected. However, no appreciable difference was detected between negative controls with no RNA template and CFPS reactions with enzymatically ligated RNA (**Figure 4-7**). Based on the lack of detection signal from CPFS reactions with enzymatically ligated RNA, CFPS reactions with CuAAC-ligated RNA were not tested.

It was thought that perhaps peptides were more difficult to synthesise and detect than larger proteins. Initiation is the slowest step in translation, and compared to proteins, peptides are much smaller. The ratio of initiation steps per incorporated amino acid is thus much higher for peptides. Additionally, the size of proteins makes them more amenable to SDS-PAGE analysis than peptides, and the functional capabilities of some proteins open up the possibility of even more assays. As such, the next experimental iteration used a template for the production of a fluorescent protein.

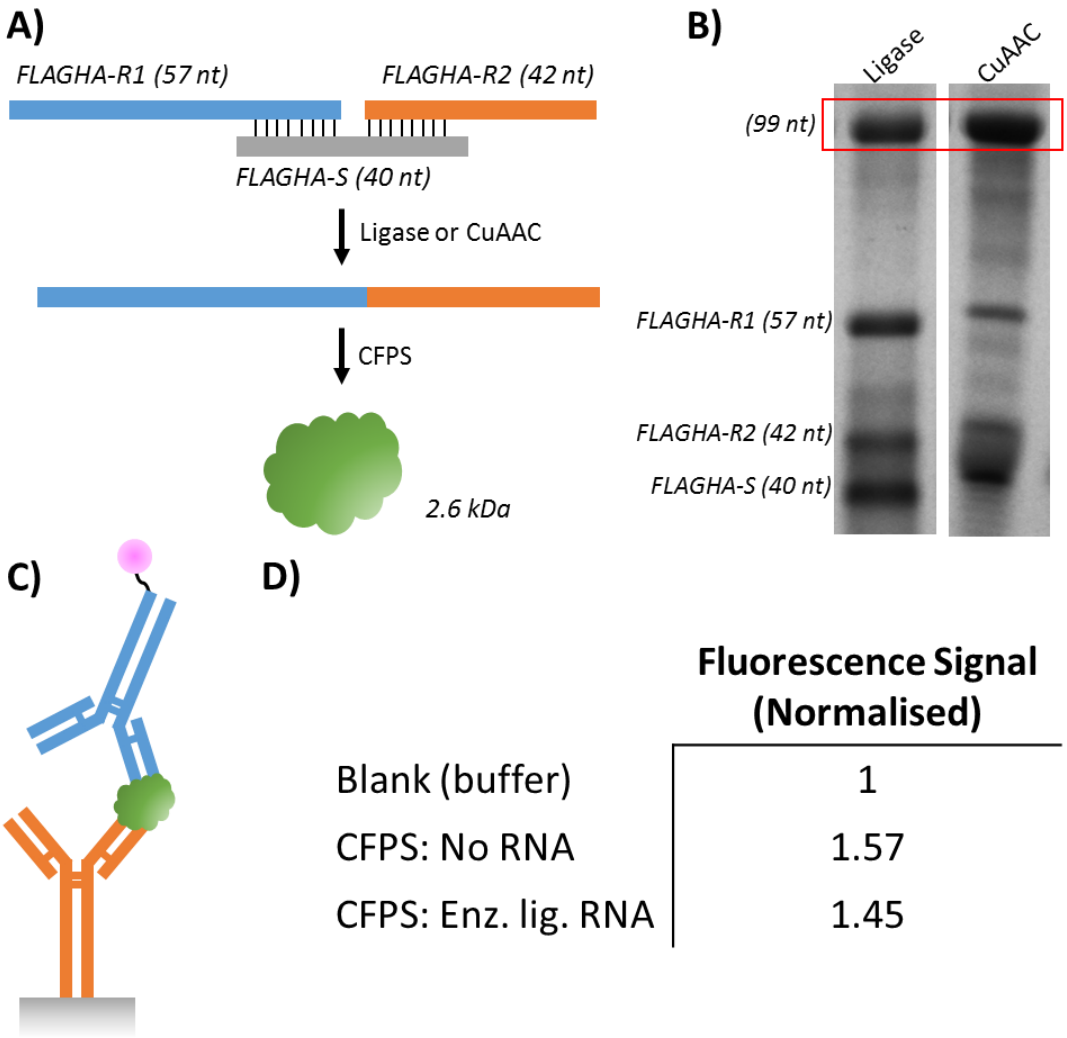


Figure 4-7. A) Schematic for the assembly and translation of the FLAG-HA peptide mRNA template. **B)** 12% denaturing PAGE showing enzymatic and CuAAC ligation of the 99 nt FLAG-HA mRNA template. **C)** Schematic for sandwich ELISA detection of the FLAG-HA peptide with a fluorophore-conjugated antibody. **D)** Fluorescence readings from sandwich ELISA, normalised by the blank (in which buffer was added in lieu of the sample). There was no significant difference between CFPS with no RNA, and CFPS with enzymatically ligated RNA template (enz. lig. RNA). Based on the negative results from CFPS with enzymatically-ligated RNA, CuAAC-ligated RNA was not tested.

4.5 CFPS of CreiLOV DNA

The next experimental iteration was to use chemically synthesised, enzymatically ligated DNA coding for the fluorescent CreiLOV protein. The idea was to develop a working CFPS system with DNA first, before applying it to RNA. There

were four main advantages to the DNA CreiLOV template: 1) the increased stability of DNA over RNA, to ensure that the template was not being degraded; 2) the coupling of transcription with translation, mimicking the natural phenomenon in prokaryotes; 3) the increased size of the protein product, making it easier to analyse by SDS-PAGE; and 4) the fluorescent properties of the protein upon addition of flavin mononucleotide, opening up the possibility of functional assays.

The DNA sequence for CreiLOV was partially based on work published by Arnab Mukherjee *et al.* (2015),[150] and also contained a T7RNAP promoter region and RBS. Initially, the DNA template used was 410 nt total, assembled from three oligonucleotides (CreiLOV-D1, -D2, and -D3). The oligonucleotides were annealed with templating splints, and then ligated with T4 DNA ligase (**Tables 4-6** and **4-7**). The ligation reaction was analysed by denaturing PAGE, and the correct ligation product was extracted from the gel and purified by phenol:chloroform extraction and ethanol precipitation. The CreiLOV DNA template was then added to PURExpress® CFPS mixture, along with excess T7RNAP promoter strand for coupled transcription-translation. However, analysis by SDS-PAGE and Coomassie staining did not detect any CreiLOV product. On the other hand, CFPS with a positive control plasmid coding for dihydrofolate reductase (DHFR) supplied by the manufacturer produced a band of the expected size on the gel (**Figure 4-8**).

Separately transcribing the CreiLOV DNA and adding the transcribed RNA to the CFPS reaction mixture was also tested. CreiLOV DNA template was incubated with T7RNAP promoter strand, ATP, CTP, GTP, UTP, and T7RNAP at 37 °C, either for four hr or overnight (ON). The transcription reactions were analysed by denaturing PAGE, and the correct RNA product was extracted from the gel and used

in CFPS reactions. As before, SDS-PAGE results did not show any detectable CreiLOV protein product (**Figure 4-9**).

Upon the suggestion of the PURExpress® manufacturer, a truncated T7RNAP terminator sequence (CreiLOV-T) was added to the 5′ end of the CreiLOV DNA template (meaning the terminator sequence would be at the 3′ end of the transcribed RNA; **Table 4-6**). The T7 terminator was initially not included, because it was thought that the main role of the terminator was to regulate mRNA transcription from a plasmid and was therefore unnecessary for linear templates. However, the T7 terminator sequence also plays a stabilising role by forming a stem loop structure at the 3′ end of the RNA.[151] Therefore, a new CreiLOV DNA template was assembled using T4 DNA ligase as described above, with the 48 nt truncated terminator sequence added to the 3′ UTR region. Despite the addition of the truncated terminator sequence to the template, still no CreiLOV translation product was detected (**Figure 4-10**).

Table 4-6. Oligonucleotide sequences for the assembly of the CreiLOV template. The sequences corresponding to the T7RNAP promoter and terminator are italicized and underlined, the RBS is underlined, and the start and stop codons are in bold. They are the reverse complements (with the exception of CreiLOV-P, the promoter sequence) since the DNA template is the antisense strand. CreiLOV-T contains the suggested truncated T7RNAP terminator sequence. CreiLOV-P contains the T7RNAP promoter sequence, and is complementary to CreiLOV-D3.

Name	Length (nt)	Sequence (5′ → 3′)
CreiLOV-D1	138	*PO_4-* **TTA**CGCCAGCGCCTTGCCCTCCGTCTTGCTGGTCAC ATCCACCTGCACGCCCACGAACTTGGACACCCGGCCGTCG GGCGTCTTGATGGGCGTGACGGTGAGCAGGTTCCAAAAGG GCGTGCCGTCCTTGCGGTAGTT
CreiLOV-D2	136	*PO_4-* GAGCAGGCGCACGGAGCAGGCCTCGCCCTTCTTGA TGGCGTCCCTGATCTTCTGGACCTCCTTGGGGTCCGTGCC CTCGCCTTGGAGGAAGCGTGCGTTGTGACCAAGCACCTCA TCGGGTCCATATCCAGTCATG
CreiLOV-D3	136	*PO_4-* GCATAGAACCCCTCGCTGGCGTAGACCAGCGGACA GTCCGGAAGCGTGGCATCTGCGACGACGAAAGTATGCCGC AAGCCAGC**CAT**ATTGTTACCTCCTTATAAAGTTAACCC*TA TAGTGAGTCGTATTA*GGTTGC
CreiLOV-T	48	*CAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGT T*ATGCTAG
CreiLOV-S12	40	CCTGCTCCGTGCGCCTGCTCAACTACCGCAAGGACGGCAC
CreiLOV-S23	40	GCCAGCGAGGGGTTCTATGCCATGACTGGATATGGACCCG
CreiLOV-ST1	40	AGGGCAAGGCGCTGGCGTAACTAGCATAACCCCTTGGGGC
CreiLOV-P	60	GCAACC*TAATACGACTCACTATA*GGGTTAACTTTATAAGG AGGTAACAAT**ATG**GCTGGCT

Table 4-7. The translation product of the CreiLOV template.

Name	Length (aa)	MW (kDa)	Sequence
CreiLOV	119	13	MAGLRHTFVVADATLPDCPLVYASEGF YAMTGYGPDEVLGHNARFLQGEGTDPK EVQKIRDAIKKGEACSVRLLNYRKDGT PFWNLLTVTPIKTPDGRVSKFVGVQVD VTSKTEGKALA

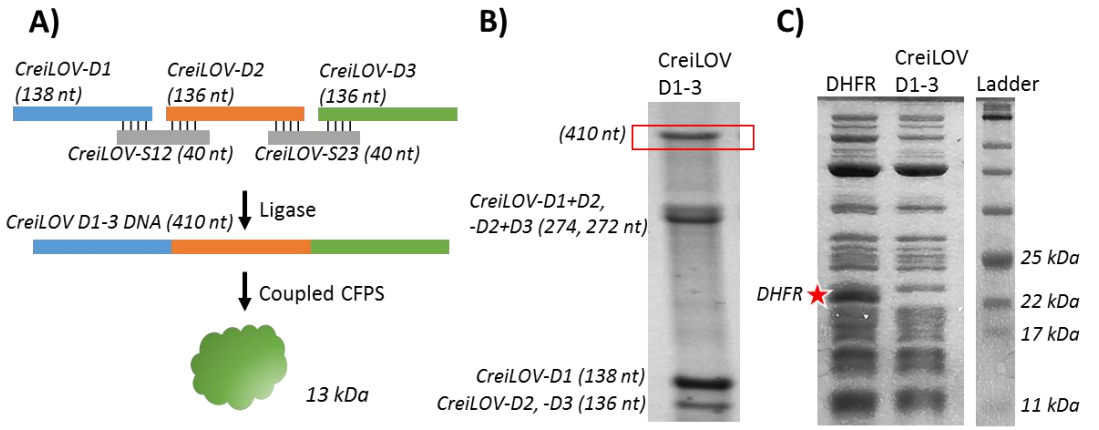


Figure 4-8. A) Schematic for the assembly and coupled transcription-translation of the original CreiLOV DNA template. **B)** 5% denaturing PAGE showing enzymatic ligation of the 410 nt CreiLOV DNA template. **C)** 16% Tris-Tricine SDS-PAGE with Coomassie staining showing coupled transcription-translation of the positive control DHFR plasmid (approximately 20 kDa protein product) and of the CreiLOV (D1-3) DNA template (13 kDa protein product).

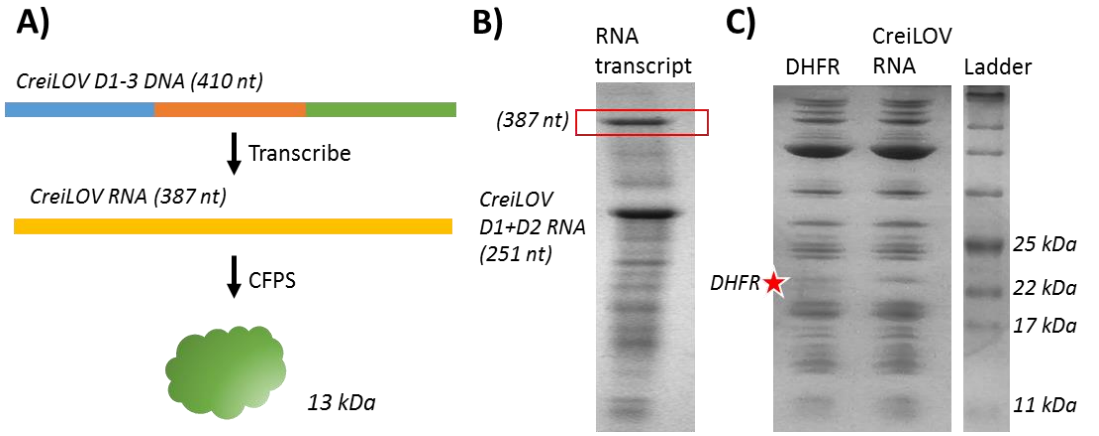


Figure 4-9. A) Schematic for the separate transcription and translation of the original CreiLOV DNA template. **B)** 5% denaturing PAGE showing the 387 nt RNA transcribed from the ligated CreiLOV DNA template. **C)** 16% Tris-Tricine SDS-PAGE with Coomassie staining showing translation of the positive control DHFR plasmid (approximately 20 kDa protein product) and transcribed CreiLOV (D1-3) RNA template (13 kDa protein product).

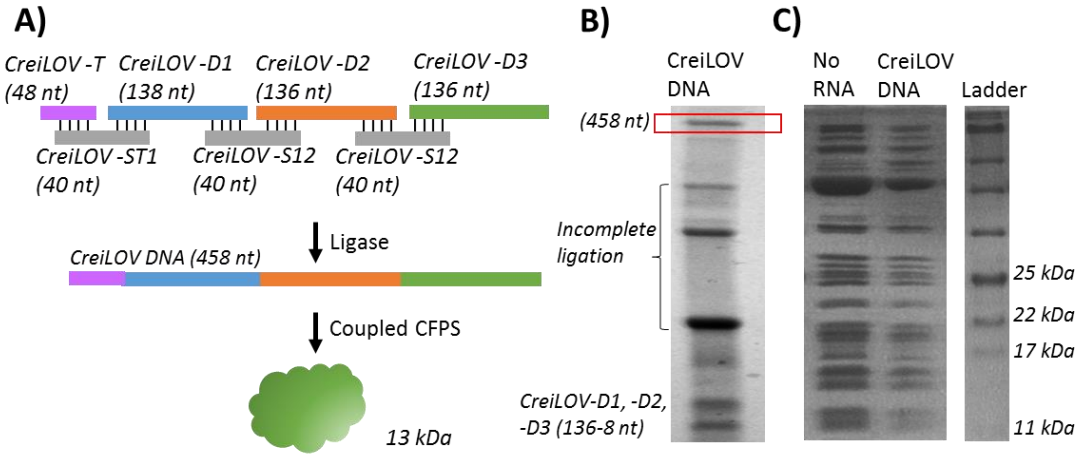


Figure 4-10. A) Schematic for the assembly and coupled transcription-translation of the CreiLOV DNA template, with the added truncated T7RNAP terminator sequence (purple). The terminator sequence was added at the 5′ end of the DNA template, so that it would be at the 3′ end of the transcribed RNA. **B)** 5% denaturing PAGE showing enzymatic ligation of the 458 nt CreiLOV DNA template with the truncated terminator sequence. **C)** 16% Tris-Tricine SDS-PAGE with Coomassie staining showing coupled transcription-translation of a negative control with no template and of CreiLOV DNA template containing the terminator sequence.

4.6 CFPS of luciferase mRNA

The goal of the next experimental design was to further elucidate the roles of the RNA 5′ and 3′ UTRs in translation, by comparing CFPS reactions with RNA containing the same coding region but different UTRs. RNA templates with the truncated UTRs suggested by the PURExpress® manufacturer (which were used in the previous CreiLOV study) were compared to RNA templates with longer UTRs from a control luciferase plasmid. Luciferase is a commonly used 36 kDa protein that allows for straightforward analysis by SDS-PAGE and functional luminescence assays. The luciferase plasmid was transcribed with T7RNAP to produce an RNA template containing a 32 nt 5′-UTR (Luc-5′UTR) and 172 nt 3′ UTR (Luc-3′UTR). For the template with truncated UTRs, linear DNA with a 26 nt 5′-UTR (Luc-5′UTR-trunc) and 35 nt 3′-UTR (Luc-3′UTR-trunc) was first produced by PCR of the

plasmid with tailed primers. The linear DNA was then transcribed with T7RNAP to produce the RNA template with truncated UTRs (**Tables 4-8** and **4-9**). Both the full-length 172 nt and truncated 35 nt 3′ UTRs could form stable hairpin secondary structures, as calculated by The Vienna RNAfold web server (**Figure 4-11**).[155] The 35 nt 3′ UTR was similar in structure and sequence to the hairpin formed at the 3′ end of the 172 nt 3′ UTR.

Table 4-8. Sequences for the full-length and truncated untranslated regions (UTRs) of the luciferase mRNA templates. The RBS are underlined, and the stop codons are in bold.

Name	Length (nt)	Sequence (5′ → 3′)
Luc-5′UTR	32	GGCAATTTTAACTTTACTAAGGAGAATTCACC
Luc-3′UTR	172	**TAA**ACGAATTCGGGCTCGGTACCCGGGGATCCTCT AGAGTCGACCTGCAGGCATGCAAGCTGATCCGGCT GCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGC TGCCACCGCTGAGCAATAACTAGCATAACCCCTTG GGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG
Luc-5′UTR-trunc	26	GGGCTTAAGTATAAGGAGGAAAAAAT
Luc-3′UTR-trunc	35	**TAA**CTAGCATAACCCCTCTCTAAACGGAGGGGTTT

Table 4-9. The translation product of the luciferase template.

Name	Length (aa)	MW (kDa)	Sequence
Luc	312	36	MASKVYDPEQ RKRMITGPQW WARCKQMNVL DSFINYYDSE KHAENAVIFL HGNAASSYLW RHVVPHIEPV ARCIIPDLIG MGKSGKSGNG SYRLLDHYKY LTAWFELLNL PKKIIFVGHD WGACLAFHYS YEHQDKIKAI VHAESVVDVI ESWDEWPDIE EDIALIKSEE GEKMVLENNF FVETMLPSKI MRKLEPEEFA AYLEPFKEKG EVRRPTLSWP REIPLVKGGK PDVVQIVRNY NAYLRASDDL PKMFIESDPG FFSNAIVEGA KKFPNTEFVK VKGLHFSQED APDEMGKYIK SFVERVLKNE QV

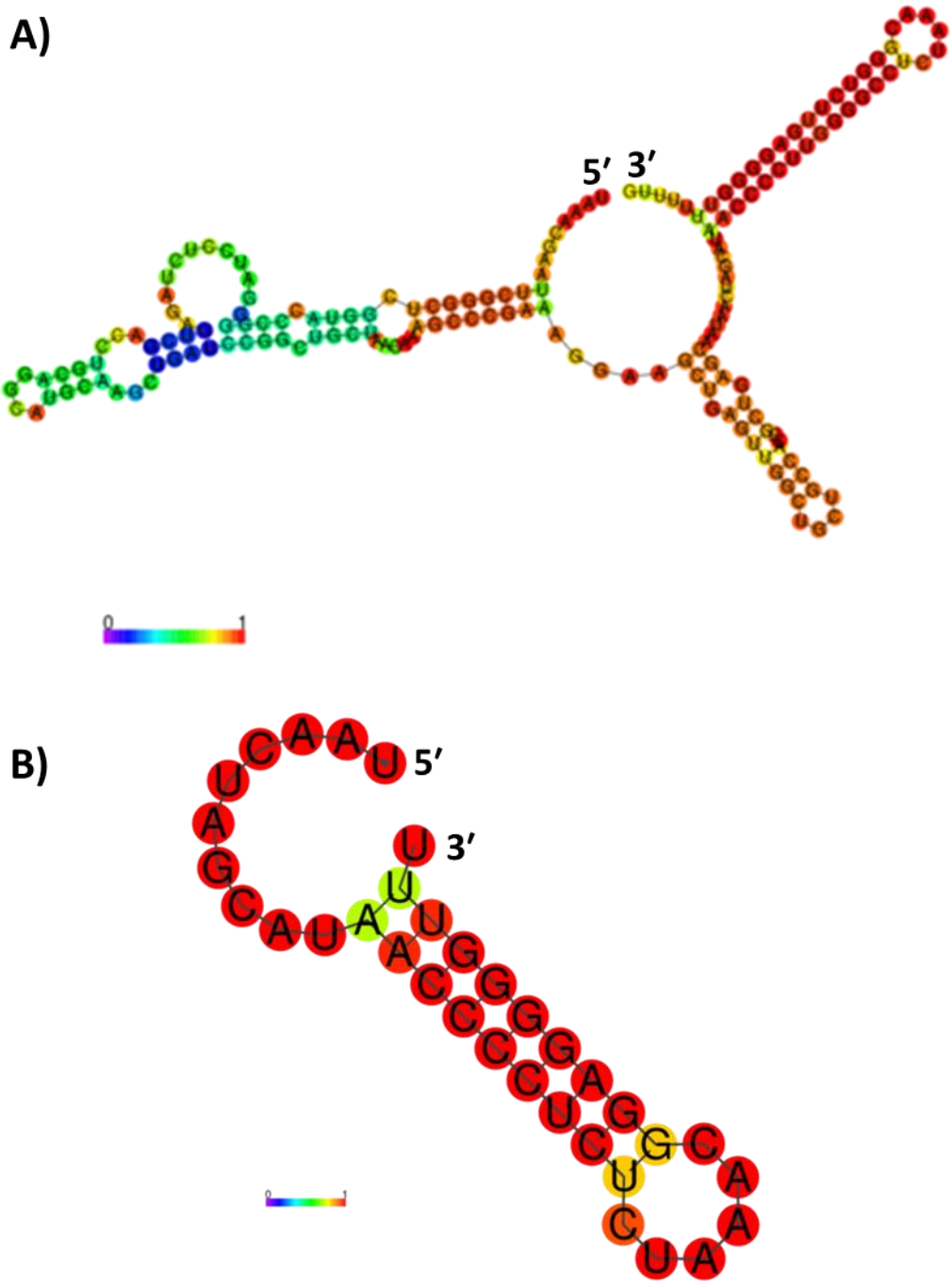


Figure 4-11. Secondary structure of the RNA 3′ UTRs, as predicted by The Vienna RNAfold web server.[152] The structures are colour-coded by probability (red being most probable and purple being least probable). **A)** Secondary structure formed by the full-length 3′ UTR. **B)** Secondary structure formed by the truncated 3′ UTR suggested by the PURExpress® manufacturer.

Both RNA templates were purified by phenol:chloroform extraction and ethanol precipitation, and equal amounts were added to CFPS reaction mixture. The

translation reactions were then analysed by SDS-PAGE and Coomassie staining. A clear protein band was detected for the reaction with mRNA containing the full-length UTRs. On the other hand, no appreciable difference was detected between the reaction with RNA containing truncated UTRs and the negative control with no RNA added. These results suggested that UTRs play a crucial role in CFPS with RNA templates, and that the truncated UTRs were insufficient for producing significant translation yield. Luminescence assays further demonstrated that CFPS reactions with RNA templates containing the full-length UTRs produced functional luciferase (**Figure 4-12**).

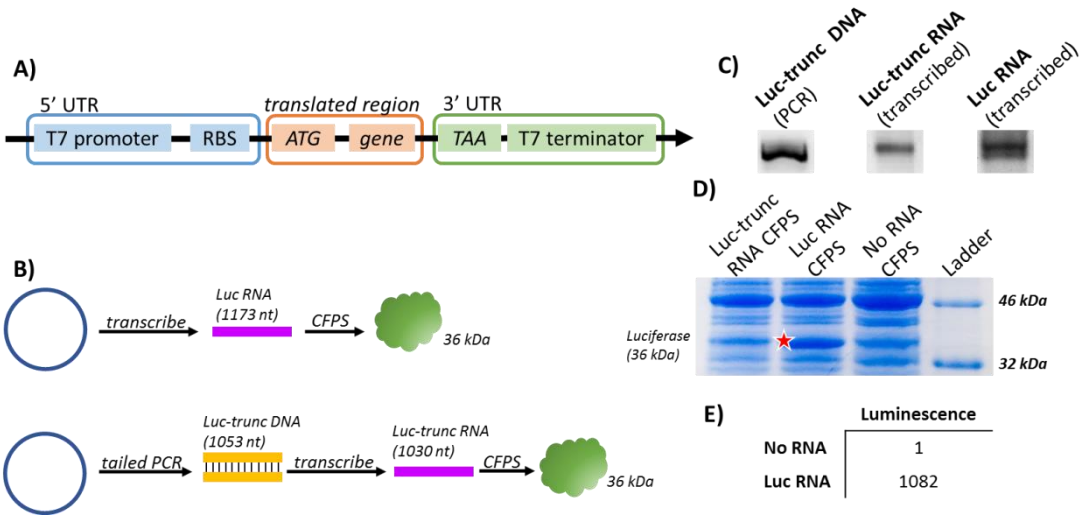


Figure 4-12. A) Diagram of the luciferase RNA template. The 5′ untranslated region (UTR; in blue) contained the T7RNAP promoter sequence and the RBS. The translated region (orange) contained the start codon and code for the luciferase protein. The 3′ UTR (green) contained the stop codon and T7RNAP terminator sequence. **B)** Schematic for the assembly and translation of the luciferase mRNA templates. The mRNA templates with the full-length UTRs were directly transcribed from the luciferase plasmid (top). The mRNA templates with the truncated UTRs were transcribed from linear DNA templates obtained from tailed PCR of the plasmid (bottom). **C)** 2% agarose gels stained with SYBR™ gold showing the tailed PCR and transcription products. **D)** 16% SDS-PAGE gels stained with Coomassie blue showing the CFPS reactions with luciferase templates (36 kDa protein product). **E)** Luminescence measurements of CFPS reactions, normalised by the negative control with no RNA added.

Given the proper UTRs, this study showed that exogenous RNA added to CFPS reactions could generate enough functional protein to be detected by SDS-PAGE and Coomassie staining. Since the full-length UTRs alone were already 204 nt long combined, any template containing a coding region and these UTRs would be too long to chemically synthesise without many ligation sites. Therefore, the next template iteration was designed so that chemically synthesised RNA containing a triazole backbone and coding for a small peptide tag would be enzymatically ligated to transcribed RNA coding for luciferase and containing the full-length 3′ UTR.

4.7 CFPS of His-tagged luciferase mRNA

Based on the results from the previous experiment in which luciferase was translated from RNA containing long UTRs, a new template was designed that coded for a 6xHis peptide tag followed by luciferase. The first part of the template, which contained the 5′ UTR and 6xHis tag coding region, was assembled by ligating two RNA oligonucleotides (76 nt total): either enzymatically for the control, or by CuAAC in the ‘31’ position after the fourth translated codon. The second part, containing the luciferase coding region and 3′ UTR, was formed by enzymatically transcribing a DNA template obtained by tailed PCR of the luciferase plasmid used in the previous experiment (1,097 nt). The two parts would then be enzymatically ligated to assemble the complete mRNA template (1,173 nt). The rationale for this template design was that the 6xHis tag would allow for purification with a nickel-nitrilotriacetic acid (Ni-NTA) matrix, as well as for detection and semi-quantification with anti-His tag immunoassays. Additionally, the luciferase would enable analysis by SDS-PAGE and luminescence assays (**Tables 4-10** and **4-11**).

Table 4-10. Oligonucleotide sequences for the assembly of the His-tagged luciferase template. The T7RNAP terminator sequence is italicized and underlined, the RBS is underlined, and the start and stop codons are in bold.

Name	Length (nt)	Sequence (5′ → 3′)
HisLuc-R1	44	GGCAAUUUUAACUUUACUAAGGAGAAUUCACC**AUG**AA ACAUCAC
HisLuc-R2	32	CAUCACCAUCACGCGAUCGCCAUGGCUUCCAA
HisLuc-R3	1,097	GGUGUACGACCCCGAGCAACGCAAACGCAUGAUCACU GGGCCUCAGUGGUGGGCUCGCUGCAAGCAAAUGAACG UGCUGGACUCCUUCAUCAACUACUAUGAUUCCGAGAA GCACGCCGAGAACGCCGUGAUUUUUCUGCAUGGUAAC GCUGCCUCCAGCUACCUGUGGAGGCACGUCGUGCCUC ACAUCGAGCCCGUGGCUAGAUGCAUCAUCCCUGAUCU GAUCGGAAUGGGUAAGUCCGGCAAGAGCGGGAAUGGC UCAUAUCGCCUCCUGGAUCACUACAAGUACCUCACCG CUUGGUUCGAGCUGCUGAACCUUCCAAAGAAAAUCAU CUUUGUGGGCCACGACUGGGGGGCUUGUCUGGCCUUU CACUACUCCUACGAGCACCAAGACAAGAUCAAGGCCA UCGUCCAUGCUGAGAGUGUCGUGGACGUGAUCGAGUC CUGGGACGAGUGGCCUGACAUCGAGGAGGAUAUCGCC CUGAUCAAGAGCGAAGAGGGCGAGAAAAUGGUGCUUG AGAAUAACUUCUUCGUCGAGACCAUGCUCCCAAGCAA GAUCAUGCGGAAACUGGAGCCUGAGGAGUUCGCUGCC UACCUGGAGCCAUUCAAGGAGAAGGGCGAGGUUAGAC GGCCUACCCUCUCCUGGCCUCGCGAGAUCCCUCUCGU UAAGGGAGGCAAGCCCGACGUCGUCCAGAUUGUCCGC AACUACAACGCCUACCUUCGGGCCAGCGACGAUCUGC CUAAGAUGUUCAUCGAGUCCGACCCUGGGUUCUUUUC CAACGCUAUUGUCGAGGGAGCUAAGAAGUUCCCUAAC ACCGAGUUCGUGAAGGUGAAGGGCCUCCACUUCAGCC AGGAGGACGCUCCAGAUGAAAUGGGUAAGUACAUCAA GAGCUUCGUGGAGCGCGUGCUGAAGAACGAGCAGGUU ***UAA****ACGAAUUCGGGCUCGGUACCCGGGGAUCCUCUAG AGUCGACCUGCAGGCAUGCAAGCUGAUCCGGCUGCUA ACAAAGCCCGAAAGGAAGCUGAGUUGGCUGCUGCCAC CGCUGAGCAAUAACUAGCAUAACCCCUUGGGGCCUCU AAACGGGUCUUGAGGGGUUUUUUG*

HisLuc-S12	40	GCGATCGCGTGATGGTGATGGTGATGTTTCATGGTGA ATT
HisLuc-S23	40	GTTGCTCGGGGTCGTACACCTTGGAAGCCATGGCGAT CGC

Table 4-11. The translation product of the His-tagged luciferase template. The 6xHis tag is underlined, and the luciferase protein begins with the M in position 12 (bolded).

Name	Length (aa)	MW (kDa)	Sequence
HisLuc	323	39.7	MK<u>HHHHHH</u>AI A**M**ASKVYDPE QRKRMITGPQ WWARCKQMNV LDSFINYYDS EKHAENAVIF LHGNAASSYL WRHVVPHIEP VARCIIPDLI GMGKSGKSGN GSYRLLDHYK YLTAWFELLN LPKKIIFVGH DWGACLAFHY SYEHQDKIKA IVHAESVVDV IESWDEWPDI EEDIALIKSE EGEKMVLENN FFVETMLPSK IMRKLEPEEF AAYLEPFKEK GEVRRPTLSW PREIPLVKGG KPDVVQIVRN YNAYLRASDD LPKMFIESDP GFFSNAIVEG AKKFPNTEFV KVKGLHFSQE DAPDEMGKYI KSFVERVLKN EQV

Both T4 RNA ligase 2 and T4 DNA ligase require a 5′-monophosphate at the ligation site, but transcription with T7RNAP produces transcripts with a 5′-triphosphate.[74, 83, 153] Therefore, several methods to obtain a 5′-monophosphate on transcribed RNA were investigated using a test template that was short enough to evaluate by mass spectrometry (**Figure 4-13**). The first method was to treat the purified RNA transcript with RNA 5′ pyrophosphohydrolase (RppH), an enzyme that removes pyrophosphate from the 5′-triphosphate of RNA to leave a 5′-monophosphate. The second method was to treat the transcript with RppH, followed by T4 polynucleotide kinase (PNK) to ensure the RNA had a 5′-monophosphate and was not completely dephosphorylated. Finally, the third method was to prime the transcription reaction itself with guanosine monophosphate (GMP). T7RNAP always begins its transcript with a G, and is able to initiate transcription with GMP but can only elongate the transcript with GTP. Taking advantage of this fact, reports in literature have obtained transcripts with a 5′-monophosphate by adding GMP in excess of GTP in the transcription reaction.[153, 154] Priming transcription reactions with two-fold and four-fold excesses of GMP to GTP were tested in this study.

Both transcripts treated with RppH and transcription reactions primed with two-fold excess GMP matched the calculated mass of the transcript with a 5′-monophosphate. On the other hand, for the untreated control sample (with a 5′-triphosphate), there was a +64 mass difference between the measured and calculated mass; this was surmised to be due to ions binding to the triphosphate. The mass difference between the RppH/GMP samples and untreated control was -220, not -158 as expected, the calculated molecular weight of $(PO_3)_2$; neither was it -237, the calculated molecular weight of $(PO_3)_3$, which would have suggested complete dephosphorylation (**Table 4-12**). Reacting RppH-treated samples with PNK, which adds 5′-monophosphates to 5′-OH groups, did not change the mass, further implying that the RppH-treated and GMP-primed samples had the correct 5′-monophosphate. Since treatment with RppH produced the same mass results as priming the transcription reaction with two-fold GMP but produced higher RNA yield, it was decided to proceed with RppH treatment.

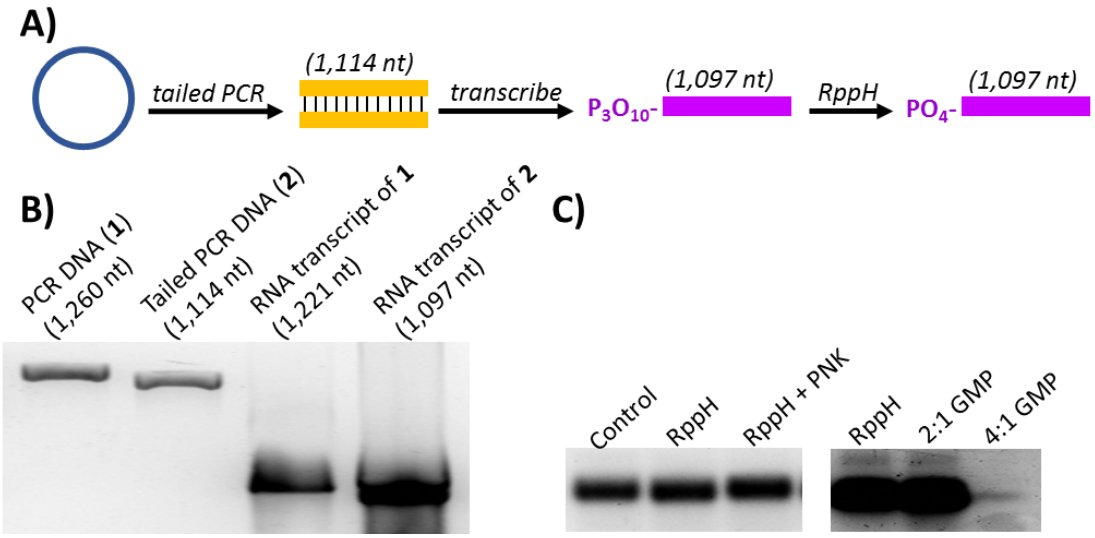


Figure 4-13. A) Schematic for the transcription and monophosphorylation of HisLuc-R3 (1,097 nt RNA). **B)** 2% agarose gels showing PCR and transcription products. HisLuc-R3 was obtained from transcription of tailed PCR products; the non-tailed PCR product of the original luciferase gene (**1**) and its transcript are shown for comparison. **C)** 2% agarose gels showing the transcription products after testing various methods of obtaining a 5′-monophosphate. The only sample that showed significant degradation or lack of RNA was the transcription reaction primed with a 4:1 ratio of GMP:GTP.

Table 4-12. Mass spectrometry results using a short (48 nt) RNA transcript to test various methods of obtaining a 5′-monophosphate. The untreated control transcript should have a 5′-triphosphate. Both treatment with RppH in T4 RNA ligase 2 buffer and priming the transcription reaction with two-fold GMP produced the expected mass for the transcript with a 5′-monophosphate (highlighted).

Method	Expected 5′ group	Expected mass	Measured mass	Mass difference
Untreated control	$(PO_3)_3O$-	15,723	15,787	64
RppH (NEB Buffer 2)	$(PO_3)O$-	15,566	15,788	222
RppH (T4 RNA ligase 2 buffer)	$(PO_3)O$-	15,566	15,567	-1
RppH → PNK	$(PO_3)O$-	15,566	15,567	-1
2:1 GMP:GTP	$(PO_3)O$-	15,566	15,567	-1
4:1 GMP:GTP	$(PO_3)O$-	15,566	*No sample obtained*	*No sample obtained*

Two ligation pathways to obtain the fully assembled 6xHis-luciferase RNA template were attempted. In the first pathway, the middle oligonucleotide (HisLuc-R2) was enzymatically ligated to the transcribed RNA (HisLuc-R3), followed by

ligation of the first oligonucleotide (HisLuc-R1) by either T4 DNA ligase or CuAAC. In the second pathway, the two RNA oligonucleotides (HisLuc-R1 and HisLuc-R2) were ligated first by either ligase or CuAAC, followed by enzymatic ligation of the transcribed RNA (HisLuc-R3). Both enzymatic and CuAAC ligation of the two RNA oligonucleotides were successfully performed, and the products were purified by denaturing PAGE for ligation to the transcribed RNA. However, it was difficult to determine whether ligation of HisLuc-R3 was successful by PAGE. Because HisLuc-R3 was so much longer, ligation of the other oligonucleotides would only produce a very slight bandshift by PAGE. To address this problem, fluorophores were added to HisLuc-R1 and HisLuc-R2 to make it easier to determine if they were indeed being ligated to HisLuc-R3 (**Figure 4-14**).

Even with the fluorophore tags, it was still difficult to separate complete ligation assemblies from unligated HisLuc-R3. Moreover, the whole assembly process was cumbersome given the low and inconsistent yield of the HisLuc-R3 ligations, consistent with literature reports on the ligation of kilobase-long RNA.[101] It was therefore desirable to design a template that would combine the ease of assembly of relatively short RNA oligonucleotides with the ease of analysis of relatively large translation products.

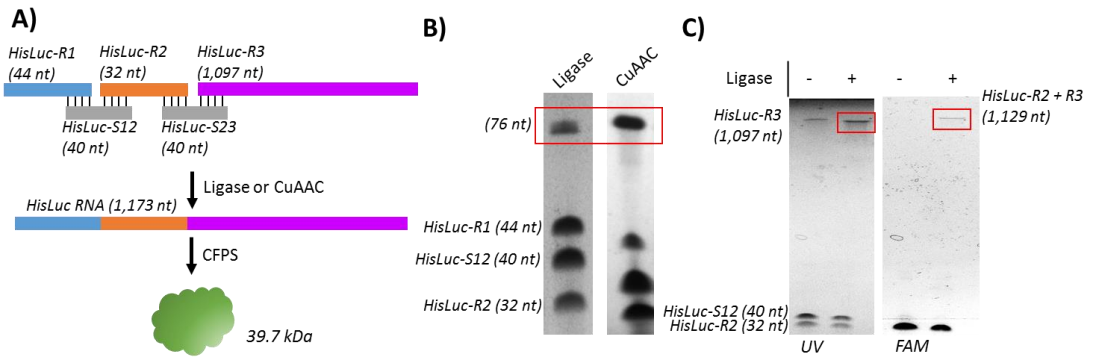


Figure 4-14. A) Schematic for the assembly and translation of the His-tagged luciferase mRNA templates. HisLuc-R1 and HisLuc-R2 (blue and orange) were chemically synthesised oligonucleotides, and were ligated enzymatically or by CuAAC. HisLuc-R3 (purple) was an RNA transcript, and was enzymatically ligated to HisLuc-R2. **B)** 8% denaturing PAGE showing enzymatic and CuAAC ligation of HisLuc-R1 and HisLuc-R2. **C)** 5% denaturing PAGE showing enzymatic ligation of FAM-labelled HisLuc-R2 and unlabelled HisLuc-R3. On the left, the gel is visualised by UV absorbance; on the right, the same gel is visualised by the FAM fluorophore. Only free HisLuc-R2 and HisLuc-R2 ligated to HisLuc-R3 should be visible on the right.

4.8 CFPS of circular FLAGx3 mRNA

In rolling circle translation (RCT), a ribosome reads a circular RNA template without a stop codon to produce a translation product consisting of multiple repeating peptide sequences. In this way, relatively large translation products can be generated from short RNA templates without the need for long UTRs to stabilise free termini. Naoko Abe *et al* (2013) reported successful RCT of repeating FLAG tags from short circular RNA using prokaryotic CFPS. Their highest yielding template was a 126 nt cyclic RNA coding for three FLAG tags per round of the circle.[155, 156]

The RNA templates used in the RCT studies described here were based on the 126 nt template from Abe *et al*. The 126 nt circular templates were assembled from two chemically synthesised RNA oligonucleotides of approximately 63 nt each. First, the two RNA strands were ligated with either T4 DNA ligase for the natural backbone

control or CuAAC for the triazole backbone to form a 126 nt linear strand. The ligation reaction was analysed by denaturing PAGE, and the correct 126 nt band was extracted from the gel. Next, the linear strand was circularised by ligating its 3′ and 5′ ends together with T4 DNA ligase. This reaction was also analysed and purified by denaturing PAGE (**Figure 4-15** and **Table 4-13**). Cyclisation of the RNA template was checked by incubating the extracted RNA with exonuclease T (Exo T), a nuclease that requires a free 3′ terminus to digest ssRNA and ssDNA in the 3′ → 5′ direction. Linear templates were digested by Exo T, but not circular ones (**Figure 4-16**). The extracted circular RNA templates were then used in prokaryotic CFPS reactions, which were further analysed by dot blot and western blot. The immunoblots were performed with mouse anti-FLAG primary antibody (optimised at 1:2,000 dilution) and anti-mouse IgG, peroxidase-conjugated secondary antibody (optimised at 1:250,000 dilution), and visualised with enhanced chemiluminescence (ECL) substrate.

Table 4-13. The translation product of one cycle around the circular FLAG RNA template. The FLAG tags (DYKDDDDK) are underlined.

Name	Length (aa)	MW (kDa)	Sequence
RCT	42	5.1	MII<u>DYKDDDD K</u>II<u>DYKDDDD</u> <u>K</u>LLII<u>DYKDD DDK</u>IIIIKEI YP

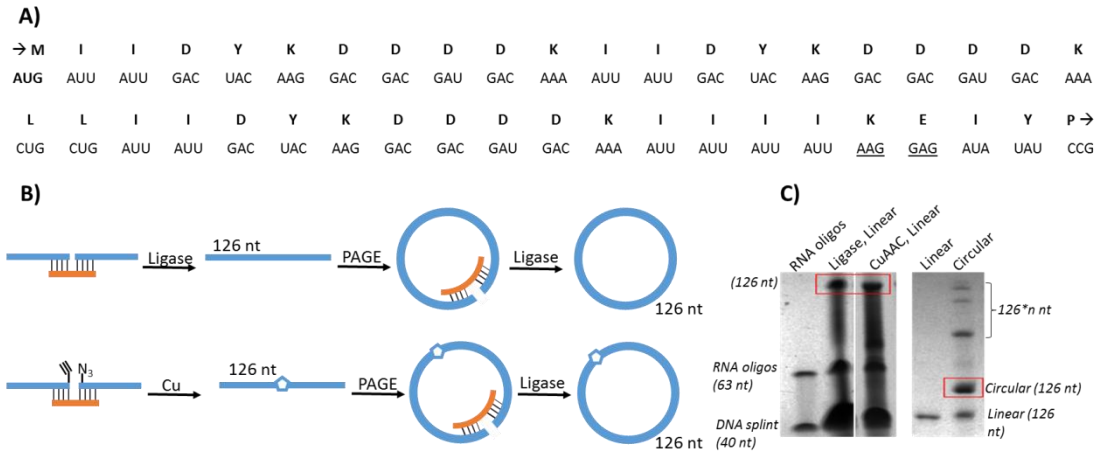


Figure 4-15. A) The nucleic acid sequence of the circular FLAG template, with the corresponding amino acid code. The start codon is in bold, and the RBS is underlined. The RBS, as part of the circular RNA template, is also translated. There is no terminator sequence or stop codon. The arrows signify that the template is cyclic. **B)** Schematic for the assembly and translation of the circular FLAG template, without a triazole linkage (top) and with a triazole (bottom). **C)** 8% denaturing PAGE of the intermolecular linear ligation (enzymatic or CuAAC) and intramolecular circular ligation (enzymatic) steps to assemble the circular FLAG template. The streakiness in the ligase and CuAAC linear ligation gels is due to the large amount of oligonucleotides loaded into the gel wells.

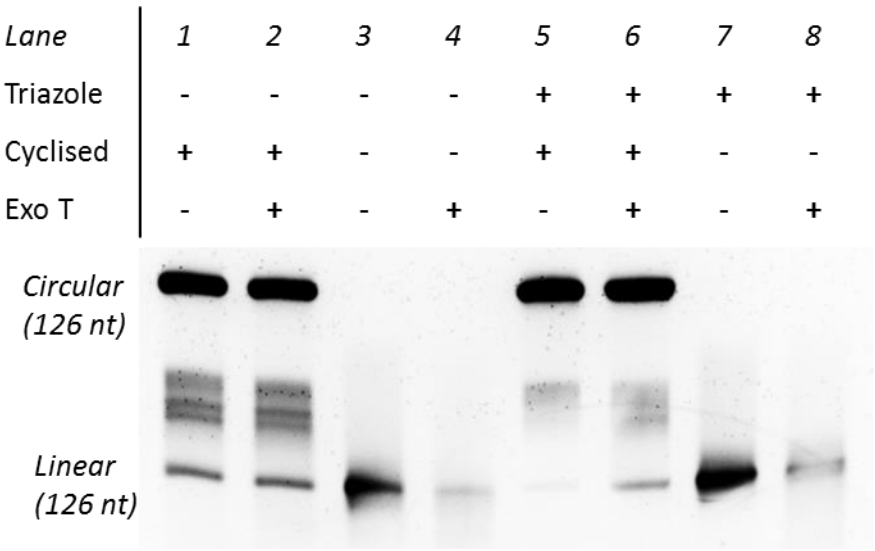


Figure 4-16. 12.5% denaturing PAGE confirming cyclisation of the RNA templates with exonuclease T (Exo T) digestion. The RNA in **Lanes 1-4** contain only natural phosphodiester backbones, and the RNA in **Lanes 5-8** contain an internal triazole linkage. **Lanes 1-2** and **Lanes 5-6** contain circular RNA, while **Lanes 3-4** and **Lanes 7-8** contain linear RNA. Exo T requires a free 3′ terminus to digest ssRNA and ssDNA in the 3′ → 5′ direction

Six different triazole-containing circular templates were investigated, in addition to the natural backbone control with no triazoles. All templates consisted of the same 126 nt sequence, but with the triazole in different positions. Factors that were tested included the position of the triazole relative to the codon reading frame, its position relative to the AUG start site, and the local sequence around the triazole linkage. First, the templates were designed so that the triazole was either in between codons, i.e. between the third base of a codon and the first base of the next codon ('31' templates), or between the second and third bases of the same codon ('23' templates). These two positions were chosen as the most likely to tolerate the unnatural triazole linkage, being immediately before or after the wobble base pair of the codon. The '31' position would also be between tRNA binding sites. Secondly, previous work from the Brown group with triazole-containing DNA templates had found that yield from DNA and RNA polymerases increased when the triazole was located farther away from the initiation site.[113] Therefore, in two of the templates the triazole linkage was located farther away from the AUG start codon. Finally, the G/C content around the triazole was varied, since G/C base pairs have three hydrogen bonds compared to two for A/U base pairs. As a result, increased G/C content generally leads to greater duplex stability and higher melting temperatures (**Table 4-14**; see Appendix for diagrams of the circular templates with different triazole positions).

Table 4-14. The different triazole positions tested (denoted by '**t**'), relative to the codon reading frame and to the AUG start site. The templates all share the same sequence, but with the triazole located in different places within the cylic template. The templates are named so that the letters denote the RNA bases immediately before and after the triazole linkage, and the numbers indicate where the triazole is positioned relative to the codon reading frame. In '31' templates, the triazole is between the third base of a codon and the first base of the next codon; in '23' templates, the triazole is between the second and third bases of the same codon. The codon reading frame is indicated by brackets under the 'Local Sequence' column.

Name	Local sequence	No. bases after *AUG* [nt]
Enz	*no triazole*	*no triazole*
CU31	[GAC]**t**[UAC]	39
(G)AC23	[GA**t**C]	38
GC31	[CUG]**t**[CUG]	63
UU23	[AU**t**U]	68
AC31	[AAA]**t**[CUG]	60
(U)AC23	[UA**t**C]	41

At first, RCT reactions were incubated at 37 °C as recommended by the manufacturer and in literature. Translation product was detected for the natural backbone control, but no product was detected for any of the triazole-containing templates. Lower incubation temperatures were then tested based on the rationale that triazole backbones are known to destabilise duplex structures and lower melting temperatures.[102, 157, 158] At 37 °C, the incubation temperature may have been above the melting temperature (Tm) of the tRNA to the triazole-containing mRNA. Indeed, while still not as strong as the signal from the natural backbone control template, incubation at 25 °C produced detectable translation product for all but one of the triazole-containing templates (UU23). Translation product was also detected from reactions incubated at 16 °C, but yield was lower than for reactions incubated at 25 °C (**Figure 4-17**). Therefore, it appeared that the optimum temperature for translation of triazole-containing mRNA was below the Tm of the tRNA-mRNA duplex, but balanced with the optimal reaction temperature for the translation machinery.

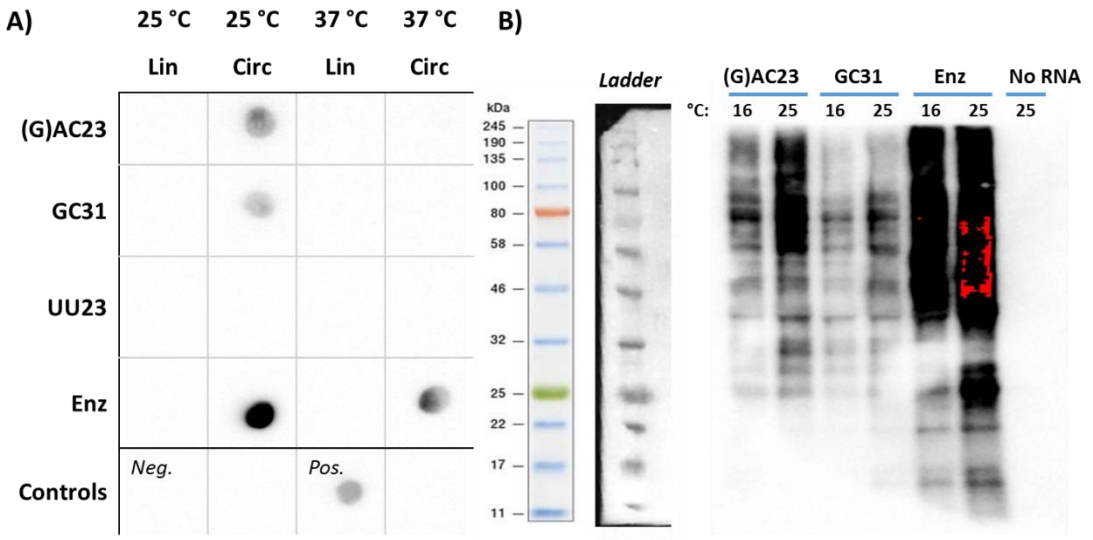


Figure 4-17. A) Dot blot comparing CFPS reactions incubated at 37 °C and 25°C. Templates with the triazole at different positions were tested, along with a template containing only natural phosphate backbones ('Enz'). Circular (circ) and linear (lin) versions of the templates were also compared. A positive control (a commercial FLAG-tagged protein; '+FLAG') and negative control (CFPS with no RNA) are shown on the bottom. **B)** A western blot transferred from a 10-20% gradient SDS-PAGE gel comparing CFPS reactions incubated at 25 °C and 16 °C with different mRNA templates. The red areas in the 'Enz – 25 °C' lane indicate pixel saturation. A ladder visualised with dyes is shown on the left, along with corresponding molecular weights (in units of kDa; from New England Biolabs).

The critical role of the Tm was further underscored by the varied yields from the different triazole-containing templates. Based on the immunoassay results, G/C content around the triazole was the most important factor, presumably because G/C base pairs tend to increase Tm by forming one more hydrogen bond than A/U base pairs. Position relative to the codon also influenced translation yield, whereas the position of the triazole relative to the AUG start site had little impact. The triazole-containing mRNA template that consistently produced the highest translation yield was (G)AC23. In the best cases, (G)AC23 produced 70-75% of translation yield relative to the natural backbone-only template, based on pixel density analysis of dot blots (**Figure 4-18** and **Figure 4-19**, '100 ng' spots).

First, the results suggested that at least one G/C base pair was essential immediately before or after the triazole. UU23 was the only template that did not generate any detectable translation product, and it was also the only template without a G/C base pair next to the triazole. Furthermore, having two G/C base pairs in the triazole-containing codon was more favourable than having just one, as (G)AC23 produced over 5 times more translation yield than (U)AC23 (dot blot from **Figure 4-19**, based on the relative intensities of the '24 hr' spots). Secondly, the results also indicated that the '23' triazole position was more favourable than the '31' position. (G)AC23 consistently generated higher yields than the following (**Figure 4-19**): CU31, in which the triazole was shifted by only one base (1.5 times higher); GC31, despite the triazole being immediately surrounded by two G/C base pairs in GC31 (4.7 times higher); and AC31, in which the immediate environment (e.g. the bases) next to the triazole was identical (2.3 times higher).

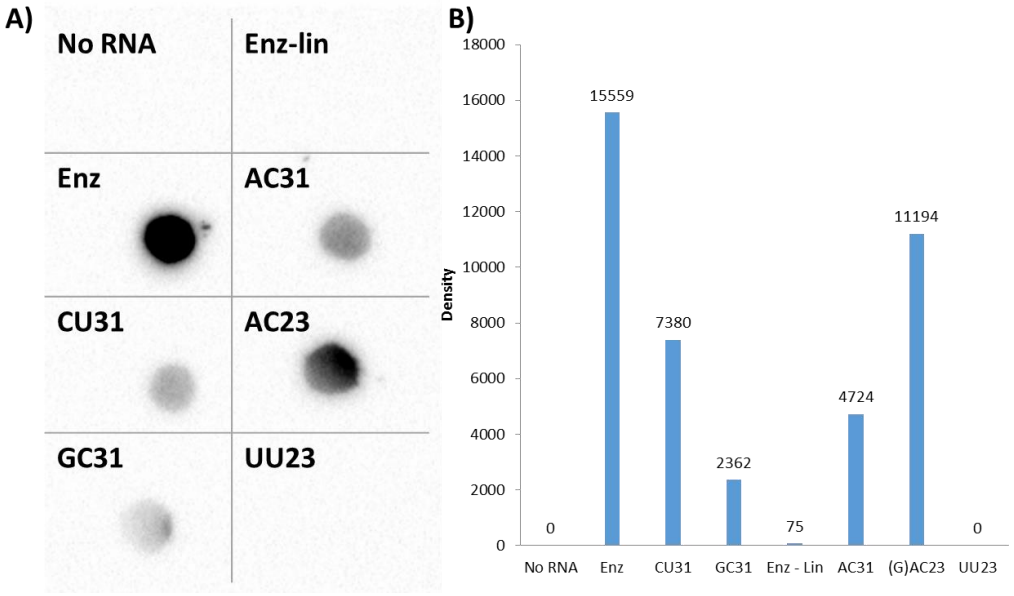


Figure 4-19. A) Dot blot comparing CFPS reactions using mRNA templates containing a triazole linkage at different positions, or with no triazole ('Enz'). For 'Enz-lin,' a linear RNA template with only natural backbones was used. **B)** Pixel densities of the dots were analysed by ImageJ to compare relative amounts of translated FLAG tag.

Translation reactions with 126 nt linear templates were compared to those with circularised templates. Negligible signal was detected for any of the reactions with linear templates, including the natural backbone linear template. On western blots, RCT with the circular templates resulted in a wide size distribution, demonstrating that the ribosomes were going around the cyclic mRNAs multiple times to create high molecular weight proteins (**Figure 4-17**). Experiments varying incubation time and RNA template amount showed that in general, longer incubation times increased yield, but RNA amounts higher than a certain level actually decreased yield (**Figure 4-19**). Increasing incubation time from 2 to 8 to 24 hr steadily increased translation yield for the two triazole-containing templates, but the differences in yield were much smaller for the Enz template with only natural backbones. In a similar vein, increasing the RNA amount from 25 ng to 100 ng increased translation yields for the triazole-containing templates, but not the Enz template. These results suggest that increasing incubation time and optimising mRNA amount can help to increase yield from triazole-containing templates, relative to natural backbone-only templates.

RCT samples were also digested with enterokinase (EK), a protease that specifically cleaves after lysine in an Asp-Asp-Asp-Asp-Lys peptide sequence. Since the FLAG tag is composed of Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys, EK should cleave correctly translated products after each FLAG tag. Analysis by western blot of the EK assays showed that the high molecular weight RCT products were indeed digested (**Figure 4-20**).

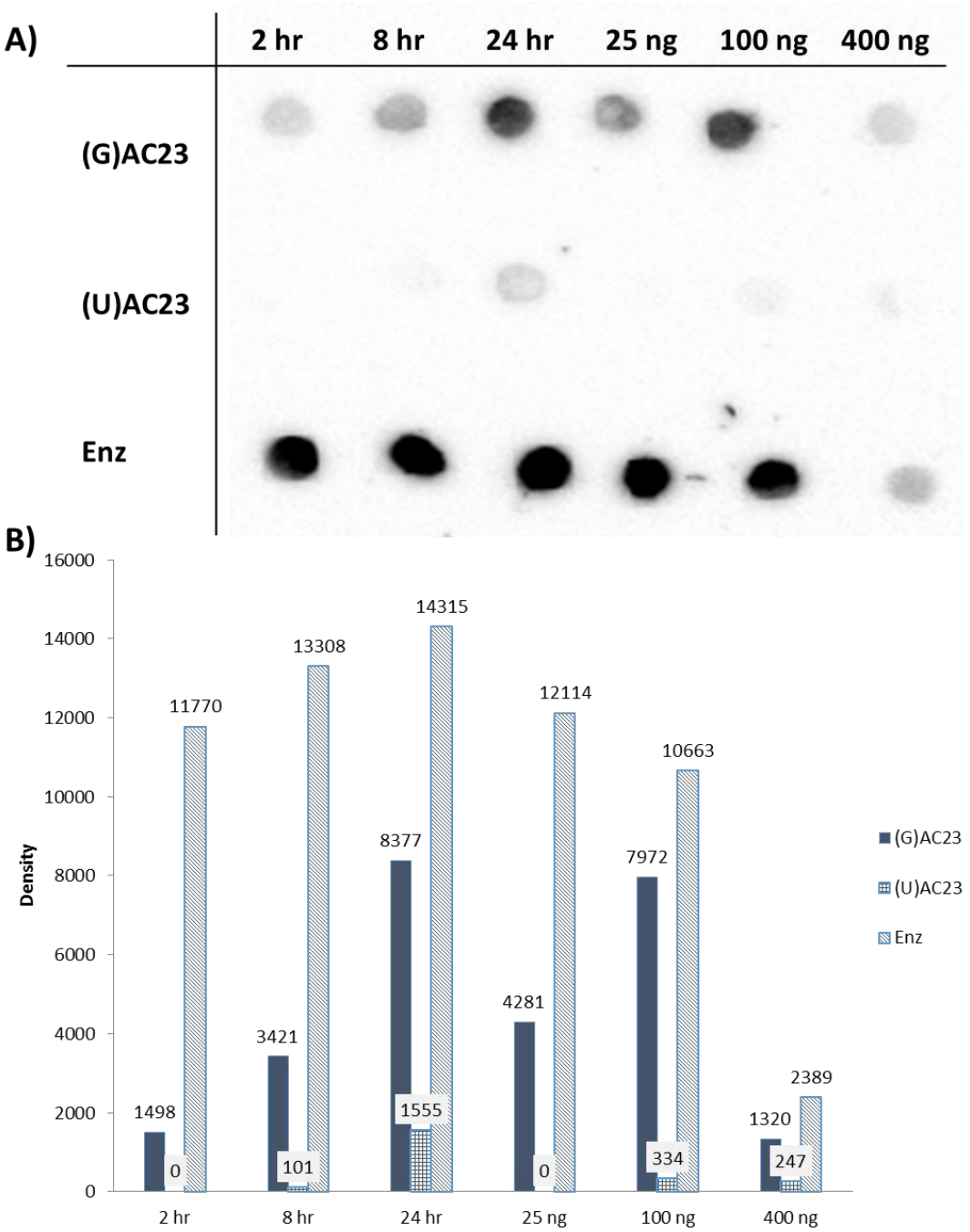


Figure 4-19. A) Dot blot comparing CFPS reactions with different incubation times and RNA template amounts. The reactions were incubated for 24 hr at 25 °C and used 100 ng RNA per 2.5 µL reaction, unless otherwise stated. **B)** Pixel density of the dots were analysed by ImageJ.

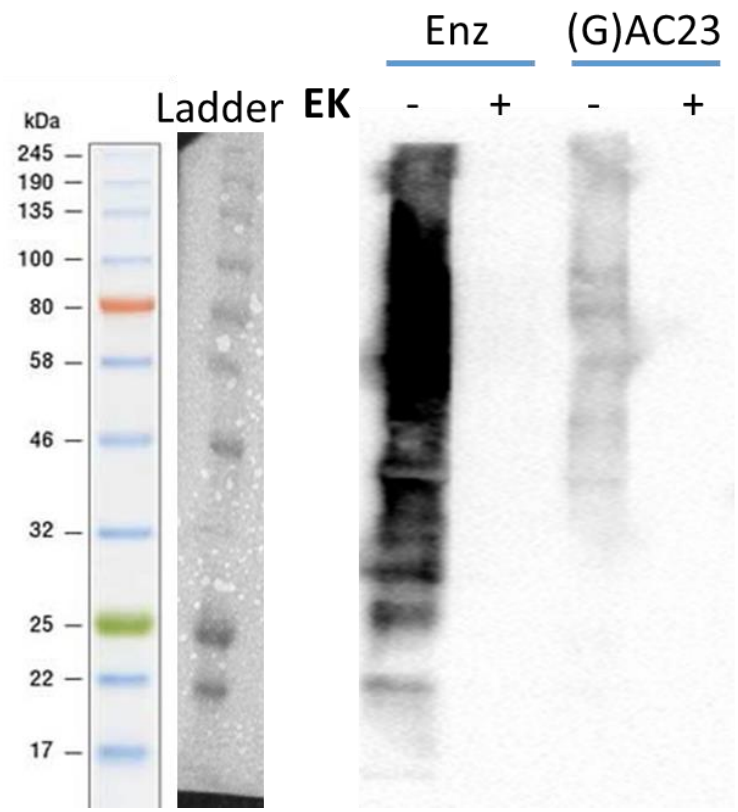


Figure 4-20. Western blot of a 10-20% gradient SDS-PAGE gel showing CFPS reactions digested with enterokinase (EK). EK is a specific protease that recognises an amino acid sequence (DDDK) that is in the FLAG tag, and cleaves the peptide bond after the sequence. A ladder visualised with dyes is shown on the left, along with corresponding molecular weights (in units of kDa; from New England Biolabs).

4.9 Summary and conclusions of the study

Several iterations of templates were tested to study prokaryotic cell-free protein synthesis (CFPS) from mRNA containing an unnatural triazole backbone. The triazole backbone was introduced through a CuAAC reaction between the 3′-alkyne of one RNA oligonucleotide and the 5′-azide of another RNA oligonucleotide, resulting in chemical ligation of the two RNA molecules. Control templates with only natural phosphodiester backbones were also assembled by enzymatic ligation.

First, a template coding for the 19mer ermC leader peptide was tested.[149] The template was assembled by ligating two RNA oligonucleotides, and the sequence was

based off of a study published in literature. No translation product was detected by SDS-PAGE or mass spectrometry, so a new template was designed that could take advantage of the high sensitivity and specificity of immunoassays. The second template was also assembled by ligation of two synthetic RNAs, and coded for a 24mer peptide containing the FLAG and HA tags. However, still no translation product was detected by ELISA using anti-FLAG and anti-HA antibodies. To better understand prokaryotic CFPS and to eliminate RNA degradation as a factor, the third iteration design was a DNA template assembled from four synthetic oligonucleotides coding for the luminescent CreiLOV protein.[150] DNA is not only easier to synthesise and more resistant to degradation, but it is also more commonly used than RNA in prokaryotic CFPS (in which transcription and translation are often coupled) with higher reported yields. Even so, still no translation product was detected by SDS-PAGE or functional luminescence assays.

Attention was next directed to the untranslated regions (UTRs) of the mRNA template. Previously, they had not been investigated because the 5′ and 3′ UTRs suggested by the CFPS kit manufacturer consisted primarily of the T7 RNA polymerase (T7RNAP) promoter and terminator sequences, respectively. The role of T7RNAP is to transcribe RNA to DNA in coupled transcription-translation reactions. Since the original aim of these studies was to use RNA templates directly, the T7RNAP promoter and terminator regions were not thought necessary. Correspondence with the CFPS kit manufacturer revealed that even in cases where RNA templates are used directly, the 3′ T7RNAP terminator region plays a key stabilising role by forming a hairpin structure. Thus, in the fourth experimental design, tailed PCR was used to add different UTRs to a luciferase gene, which was then enzymatically transcribed to RNA. The transcribed RNA was purified and added

as direct templates in CFPS reactions. Interestingly, RNA containing the full-length UTRs produced functional luciferase that was detected by SDS-PAGE and luminescence assays, whereas RNA with the shortened UTRs suggested by the CFPS kit manufacturer did not. These results proved that directly adding RNA templates with full-length UTRs to CFPS reactions could produce detectable amounts of protein.

Since the full-length 3′ UTR itself was already 172 nt long, a fifth template was designed in which synthetic RNA containing a triazole linkage would be enzymatically ligated to transcribed RNA containing the 3′ UTR. The synthetic RNA was assembled from two oligonucleotides by either enzymatic or CuAAC ligation, and coded for a 6xHis tag to allow for Ni-NTA purification and immunodetection; the transcribed RNA coded for luciferase, to allow for analysis by SDS-PAGE and luminescence assays. However, obstacles arose in attempting to enzymatically ligate the 6xHis tag RNA to the luciferase RNA, as the luciferase RNA (1,097 nt) was significantly longer than the 6xHis tag RNA (76 nt).[101]

Finally, learning from the previous difficulties both in analysing small peptide translation products and in assembling long RNA templates, a sixth template was designed for rolling circle translation (RCT). The advantage of RCT is that a short RNA template can produce relatively large translation products, since the ribosome theoretically reads through the circular template for multiple rounds without termination. Furthermore, RCT eliminates the need for lengthy UTRs and an initiation step before each read-through of the template, which is the slowest step in translation. The 126 nt circular RNA template used was based on work published by Abe *et al.* (2013),[155] and coded for three FLAG tags with linkers in between. Lack of detectable translation product from the linear versions of the 126 nt template

supported the results from the earlier iterations of the CFPS studies, in which mRNA UTRs were found to play an important role in translation yield.

Melting temperature of the tRNA to the mRNA codon appeared to play a key role in RCT of triazole-containing mRNA. Initially, translation product was detected by immunoblot with anti-FLAG antibodies from reactions with the natural backbone control template, but not the templates containing a triazole backbone. It is known that binding strength between the mRNA and tRNA plays a crucial role in translation elongation,[1-2] and that triazole backbones destabilise duplexes and decrease their melting temperatures.[102, 157, 158] Working with this knowledge, lower RCT incubation temperatures of 25 °C and 16 °C were tested and compared to the conventional incubation temperature of 37 °C. Lowering the incubation temperature resulted in detectable translation products for the triazole-containing templates, with 25 °C generating higher yields than 16 °C. As such, the optimal RCT incubation temperature seemed to be a balance between tRNA Tm and ribosomal activity. The lack of translation product at 37 °C was hypothesised to be because triazole backbones have the effect of lowering duplex melting temperatures, and 37 °C may have been above the melting temperature of tRNA hybridising to triazole-containing mRNA.

Additionally, UU23 was the only template that did not generate any detectable product, even at lowered incubation temperatures. It was also the only template tested that did not contain at least one G/C base pair immediately before or after the triazole. Because G/C base pairs involve 3 hydrogen bonds instead of 2 for A/U base pairs, they have a higher Tm. Based on the dot blot and western blot results, it was hypothesised that at least one G/C base pair immediately before or after the triazole is necessary for successful translation, and that positioning the triazole immediately

before the wobble position in a codon is more favourable than positioning it in between codons. Triazole position and location were eliminated as reasons for the lack of translation product from UU23, since other templates with the triazole in the '23' position or with the triazole in a similar location relative to the AUG start site generated translation product.

In summary, ribosomes were capable of reading through a triazole linkage in a circular mRNA template for multiple rounds without frameshifting, as indicated by the high molecular weight products detected by anti-FLAG western blot. This claim is also premised on the fact that any frameshift of the template used in this study would have resulted in a stop codon before the ribosome completed a whole round of the circular template. However, translation yields from the triazole-containing templates were lower than from the control template with only natural phosphodiester backbones - in the best cases, 70-75% of the yield from the control template. Further studies are needed to determine if the triazole linkage was slowing the ribosome, occasionally stopping the ribosome, or increasing the likelihood of a frameshift (resulting in termination). Additionally, future studies could determine whether the triazole backbone caused substitution mutations, and if so, at what rate. Translation through triazole backbones could be further investigated and optimised by introducing modifications to increase the Tm of the mRNA-tRNA duplex, such as locked nucleic acids (LNA).[157-159]

Chapter 5: Conclusions and future work

The use of enzymatic and chemical ligation methods for the assembly of nucleic acids was investigated in this study. First, conditions were tested for the optimisation of enzymatic DNA ligation to determine maximum yields of the most commonly used ligation techniques. It was found that varying annealing protocols, incubation time and temperature, buffer composition, and nucleic acid structure did not significantly affect ligation yield. Instead, the number of ligation sites had the most significant impact on yield. The conclusions from this first part of the study were 1) for increased purity, sections of up to four ligation sites could be assembled separately, purified, and then ligated together for the final product; and 2) because none of the factors investigated significantly improved yield, the ligation reaction must be scaled up to obtain greater amounts of the product. Scaling up would require the use of correspondingly more ligase, which could become cost-prohibitive. Furthermore, as enzymes are proteins, they are susceptible to inactivation due to degradation and denaturing. Chemical ligation was thus investigated as a less expensive and potentially more consistent alternative method that would be more easily scalable.

The next part of this project investigated enzymatic and chemical ligation of RNA oligonucleotides. Literature on the assembly of RNA (e.g. mRNA) is much less developed than for DNA (e.g. genes). The comparative dearth of publications regarding RNA oligonucleotides is in part due to the relative instability of RNA and difficulty in synthesising it. RNA is more susceptible to chemical and nuclease cleavage, and solid-phase synthesis of RNA takes longer, is lower yielding, and has a shorter length limit than that of DNA. At the same time, RNA has wonderfully diverse catalytic abilities and plays essential roles in biological protein production. Assembly of artificial RNA would facilitate studies in RNA epigenetics, therapeutics,

nanotechnology, and more. Out of the two chemical ligation methods tested, CuAAC was found to produce comparable yields to enzymatic ligation for the assembly of two RNA oligonucleotides. Phosphoramidite ligation, while successfully demonstrated for DNA, was not effective for RNA. The reason was hypothesised to be due to undesirable reactions involving the 2′-hydroxyl group of ribose and the adjacent 3′ activated phosphate. New protecting group strategies may circumvent this issue.

As an efficient, bioorthogonal reaction, CuAAC is a useful tool for RNA assembly. However, the reaction leaves an artificial triazole linkage at the ligation site which may interfere with the natural activity of the RNA. The final part of this project studied the compatibility of CuAAC-ligated mRNA with ribosomal translation. After several experimental design iterations, a rolling circle translation (RCT) assay was developed, involving circular ssRNA templates of 126 nt in total with one triazole linkage. Advantages of RCT included the ability to generate large protein products from short RNA templates, greatly improved RNA stability without the need for long untranslated regions, and repeated read-through of the template without the need for multiple translation initiation steps. The circular RNA template used in this study coded for repeating units of the FLAG peptide tag, which could be detected with anti-FLAG antibodies in dot blots and western blots. Several RNA templates were assembled with the triazole linkage in different positions and surrounded by different base pairs.

Immunodetection of the translation products revealed that ribosomes could successfully translate triazole-containing mRNA, but with decreased efficiency (up to approximately 74%). One major factor could be that the triazole lowers the duplex stability (i.e. Tm) of the hybrid between the tRNA and mRNA codon.[102, 160, 161] As

evidence, RCT with triazole-containing mRNA produced protein products at 25 °C but not 37 °C, while mRNA with only natural phosphate backbones produced protein at both incubation temperatures. Furthermore, a template without any G/C base pairs flanking the triazole yielded no product at any of the temperatures tested. It was found that the highest-yielding template contained the triazole between the second and third bases of the same codon, compared to templates with the triazole in between two codons. The former triazole position may be better tolerated by the ribosome due to the wobble phenomenon.

Future studies could attempt to elucidate why triazole-containing mRNA is translated less efficiently - whether the triazole causes the ribosome to slow down, terminate more frequently, or frameshift more often. Additionally, it would be useful to know substitution mutation rates and patterns. An immediate next step would be to construct an mRNA template entirely from chemically synthesised and ligated oligonucleotides, without using any enzymes at all. For example, the 126 nt circular template used in this project could be assembled from two CuAAC ligations, resulting in an entirely chemically synthesised mRNA molecule containing two triazole linkages. Additional site-specific modifications could be introduced, such as a locked nucleic acids (LNA) next to the triazole to increase the Tm of hybridisation with the tRNA,[157-159] and thus potentially increase translation efficiency. Other interesting studies would be to test translation of triazole-containing mRNA *in vivo*, and whether having an exonuclease-resistant triazole linkage at the 3′ terminus of linear RNA would obviate the need for a stabilising 3′ UTR. Finally, it is envisaged that the RCT protocol described here could be used to study the effects of epigenetic modifications on translation efficiency, and in applications involving proteins with

multiple repeating units such as structural biomaterials (e.g. spider silk) and protein hydrogel fabrication.[160]

Chapter 6: Experimental

6.1 Reagents and buffers

Reagents

Reagents were used as purchased, without further purification.

Oligonucleotides

Oligonucleotides were synthesised by ATDBio or in the Brown group by Dr. Afaf El-Sagheer. Oligonucleotides from ATDBio were used as received. Oligonucleotides synthesised in the Brown group were purified by HPLC using 0.1 M TEAB buffer, and the product was confirmed using mass spectrometry as described below.

Buffers

HPLC:

0.1 M triethylammonium acetate (TEAA) buffer (1 L): 2.0 M aqueous triethylammonium acetate, pH 7.0 (50 mL; Link Technologies), deionised water (950 mL)

0.1 M triethylamine bicarbonate (TEAB) buffer (1 L): 1.0 M triethylamine (100 mL), deionised water (900 mL); bubble carbon dioxide through solution until pH 7-7.4

0.1 M hexylammonium acetate (HAA) buffer (1 L): acetic acid (5.99 g), hexylamine (10.12 g), deionised water (981 g); adjust to approximately pH 8.0.

2x binding & washing (B&W) buffer: 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 2 M NaCl

10x Tris/borate/EDTA (TBE) buffer (1 L): 0.9 M Tris base (108 g), 0.9 M boric acid (55 g), 31 mM ethylenediaminetetraacetic acid (9.2 g)

10x RNA gel extraction buffer: 500 mM Tris-HCl, 250 mM NaCl; adjust to pH 7.5

Tris-tricine SDS-PAGE:

3x gel buffer: 3 M Tris base, 1 M HCl, 0.3% SDS (w/v); adjust to pH 8.45

10x cathode buffer: 1 M Tris base, 1 M tricine, 1% SDS (w/v); pH approximately 8.25 (no adjustment)

10x anode buffer: 1 M Tris base, 225 mM HCl; adjust to pH 8.9

Tris-glycine SDS-PAGE:

10x gel running buffer: 1.92 M glycine, 0.25 M Tris base, 1% SDS (w/v); adjust pH to 8-8.3

5x lower buffer (resolving gel): 1.25 M Tris-HCl, 0.5% SDS (w/v); adjust pH to 8.8

5x upper buffer (stacking gel): 0.5 M Tris-HCl, 0.5% SDS (w/v); adjust pH to 6.8

2x SDS-PAGE loading buffer: 125 mM Tris-HCl pH 6.8, 30% glycerol, 2% SDS (w/v), 5% β-mercaptoethanol (v/v), 0.03% bromophenol blue (w/v)

Coomassie staining:
Coomassie stain: 10% acetic acid, 25% methanol, 0.1% Coomassie Brilliant Blue G-250 (w/v)
Coomassie destain I: 10% acetic acid, 40% methanol, 50% water
Coomassie destain II: 7% acetic acid, 10% methanol, 83% water

Western blot transfer:
10x transfer buffer: 1.92 M glycine, 0.25 M Tris base; adjust to pH 8.3
1x Western blot transfer buffer: 10% 10x transfer buffer, 20% methanol

PBS: 1 PBS Dulbecco A tablet (typical formula in g/L: NaCl 8.0, KCl 0.2, Na_2HPO_4 1.15, KH_2PO_4 0.2, pH 7.3 ± 0.2 at 25 °C; Oxoid) per 100 mL deionised water

PBS-T: PBS with 0.1% Tween-20 (v/v)

Dry milk blocking buffer: 5% dry skim milk (w/v; Tesco) in PBS-T

Carbonate/bicarbonate buffer: 15 mM carbonate, 35 mM bicarbonate; adjust to pH 9.6

6.2 Methods

DNA oligonucleotide synthesis
DNA oligonucleotides were synthesised on an Applied Biosystems 394 automated DNA/RNA synthesiser using a standard 1.0 μmole phosphoramidite cycle of acid-catalysed detritylation, coupling, capping, and iodine oxidation, with DNA phosphoramidites, solid supports, and additional reagents from Link Technologies and Applied Biosystems Ltd. Stepwise coupling efficiencies and overall yields were determined by an automated trityl cation conductivity monitoring facility, and in all cases were >98.0%. β-cyanoethyl phosphoramidite monomers were dissolved in anhydrous acetonitrile to a concentration of 0.1 M immediately prior to use. The coupling time for normal A, G, C and T monomers was 1 min. The 5′ FAM, biotin, and phosphate modifications were introduced by using the corresponding phosphoramidite, with extended coupling times of 10 min. Cleavage of the oligonucleotides from the solid support and deprotection were achieved by exposure to concentrated aqueous ammonia solution (1 hr at RT) followed by heating in a sealed glass vial (5 hr at 55 °C). The ammonia was evaporated, and the oligonucleotides were resuspended in deionised water for HPLC purification.

RNA oligonucleotide synthesis
RNA was synthesized on an Applied Biosystems 394 automated DNA/RNA synthesizer using a standard 1.0 μmole phosphoramidite cycle of acid-catalysed detritylation, coupling, capping, and iodine oxidation. The building blocks for the RNA analogues were prepared using 2′-TBDMS or 2′-TC protected RNA

phosphoramidite monomers with *t*-butylphenoxyacetyl protection of the A, G and C nucleobases and unprotected U (Sigma-Aldrich). A solution of 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 (Sigma-Aldrich). All phosphoramidite monomers were dissolved in anhydrous acetonitrile to a concentration of 0.1 M immediately prior to use, and the coupling time for all monomers was 6 min. Stepwise coupling efficiencies were determined by automated trityl cation conductivity monitoring and in all cases were >97%.

Conversion of RNA 5′-OH to 5′-Azide

A 0.5 M solution of methyltriphenoxyphosphonium iodide in dry DMF was prepared. Two 1 mL syringes were used to periodically pass 2 mL of the solution through the RNA synthesis column over 30 min at RT. Using new 1 mL syringes, the column was then washed 5 times with 1 mL dry DMF. Next, 100 mg of sodium azide was added to 2 mL of dry DMF in a glass vial, and heated at 70 °C for 10 min. When the solution had cooled to RT, 2 1 mL syringes were used to pass the sodium azide solution through the column, avoiding picking up any non-dissolved solids. The column was then incubated for 5 hr at 55 °C with the sodium azide solution. Lastly, the column was washed 5 times with 1 mL DMF, followed by 5 washes with 1 mL acetonitrile. The resin was dried by blowing argon through the column.

RNA cleavage

The RNA synthesis column was disassembled, and the resin transferred to a glass vial. A 3 mL solution of 3:1 (v/v) concentrated aqueous ammonia:ethanol solution was added to the vial, and the vial was heated for 2-4 hr at 55 °C. The vial was then cooled to RT, and most of the liquid was evaporated. The remaining liquid was transferred to a 15 mL Falcon™ tube. Acetonitrile was used to collect any residual RNA, and added to the same 15 mL Falcon™ tube. The RNA was then freeze-dried.

RNA 2′-TBDMS deprotection

Dry RNA was dissolved in 300 µL DMSO and 300 µL triethylamine trihydrofluoride, and incubated for 2.5 hr at 65 °C. Butanol precipitation was then performed by adding 3 mL butanol and 50 µL of 3 M sodium acetate. After mixing by vortexing, the solution was kept at -80 °C for at least 1 hr. The solution was then centrifuged for 30 min at 9,500 rpm and 4 °C. The supernatant was discarded, and the pellet was washed twice with cold ethanol. Any remaining ethanol was removed by freeze drying.

RNA 2′-TC deprotection

600 µL of a 1:1 dry toluene:dry ethylenediamine (v/v) solution was slowly added to the RNA synthesis column with no agitation, and incubated for 6 hr at RT. The column was then washed twice with 1 mL dry toluene, and dried with argon. The

RNA was eluted from the resin in the column by slowly pushing 1 mL water through the column, and repeating with another 1 mL water.

Oligonucleotide quantification
Oligonucleotide concentration was determined by measuring absorbance at 260 nm with a Varian Cary 50 Bio UV-Vis spectrophotometer (Agilent) in deionised water at RT, and using the Beer-Lambert law, $c = A\varepsilon l$ (c = concentration, A = absorbance, ε = molar extinction coefficient, l = cuvette length). The molar extinction coefficient was determined using the nearest neighbour method.

Oligonucleotide mass spectrometry (MS)
Mass spectrometry samples were prepared as 20 μM solutions in deionised water, and analysed with a UPLC-MS system consisting of a Waters XEVO G2-QTOF mass spectrometer (ESI negative mode) and equipped with an Acquity UPLC BEH C18 column (130 Å pore, 1.7 μm, 2.1 mm x 50 mm; Waters). A gradient of methanol in TEA and hexafluoroisopropanol (HFIP) was used (buffer A: 8.6 mM TEA, 200 mM HFIP in 5% methanol/water (v/v); buffer B: buffer A in 80% methanol): 0 to 70% buffer B over 8 minutes at a flow rate of 0.2 mL/min. Raw data were deconvoluted using the MassLynx v4.1 software package.

Peptide mass spectrometry (MS)
Mass spectrometry samples were prepared in deionised water, and analysed with a LC-MS system consisting of a Waters XEVO G2-QTOF mass spectrometer (ESI positive mode) and equipped with an Aeris Peptide XB-C18 column (100 Å pore, 2.6 μm, 2.1 mm x 100 mm; Phenomenex). A gradient of acetonitrile in formic acid was used (buffer A: 0.1% formic acid/water (v/v); buffer B: acetonitrile): 5 to 99% buffer B over 6.5 minutes at a flow rate of 0.3 mL/min. Raw data were deconvoluted using the MassLynx v4.1 software package.

Reversed-phase high performance liquid chromatography (RP-HPLC) purification
The HPLC system consisted of a Gilson 118 UV detector, 306 pumps, 805 manometric module, and 811c dynamic mixer, equipped with a Luna 10 μm C8 column (100 Å pore; 10 mm x 250 mm; Phenomenex). A gradient of acetonitrile in 0.1 M TEAB buffer was used to elute the oligonucleotides at RT (buffer A: 0.1 M TEAB, pH 7.0-7.4; buffer B: buffer A with 50% acetonitrile). A typical gradient used was 10 to 30% buffer B over 21.5 min at a flow rate of 4 mL/min. Nucleic acid elution was monitored by UV absorbance. Collected fractions were freeze-dried overnight, and resuspended in deionised water.

Denaturing polyacrylamide gel electrophoresis (PAGE)
Denaturing polyacrylamide gels with 7 M urea were cast using an SE400 vertical slab gel electrophoresis unit from Hoefer Inc. (1.5 mm depth x 18 cm width x 24 cm height), and pre-run for at least 30 min at constant power (20 W) using a PowerPac power supply (BioRad) with 1x TBE buffer. Before loading, DNA samples were

denatured by adding formamide in a 1:1 ratio (v/v) and heating at 95 °C for 5 min. RNA samples were denatured by adding formamide, but were not heated. Xylene cyanol and bromophenol blue were loaded as dye markers. The gels were run at 20 W for approximately 2 hr. The gels were visualised by UV shadowing, fluorescent dye labelling, or SYBR™ Gold (Thermo Fisher Scientific) staining on a GeneSnap system (SynGene).

Native PAGE

Native polyacrylamide gels were cast using an SE400 vertical slab gel electrophoresis unit from Hoefer Inc. (1.5 mm depth x 18 cm width x 24 cm height), and pre-run for at least 1 hr at constant voltage (120 V) using a PowerPac power supply (BioRad) with 1x TBE buffer. A 10% sucrose solution was used as loading buffer, and added to samples before loading. Xylene cyanol and bromophenol blue were loaded as dye markers. The gels were run at 120 V for approximately 12 hr, and visualised by UV shadowing on a GeneSnap system (SynGene).

Polyacrylamide gel extraction

The desired gel bands were excised, crushed, and submerged in deionised water (for DNA) or extraction buffer (for RNA) in microcentrifuge tubes. The tubes were incubated on an Eppendorf ThermoMixer overnight at 37 °C and 900 rpm. Two size exclusion NAP™ columns (GE Healthcare) were used per sample to remove salts and other impurities according to the manufacturer's protocol, and the extracted oligonucleotides were stored in deionised water.

Agarose gel electrophoresis

For a 2% agarose gel, 2.00 g of agarose powder were added to 100 mL of 1x TBE buffer. The solution was microwaved 30 seconds at a time until the agarose was fully dissolved. When the solution had cooled to approximately 40 °C, 5 µL of 10000x SYBR™ Gold (Thermo Fisher Scientific) was added (to a final concentration of 0.5x) and cast on the gel unit. Samples were prepared by mixing with 5x Green GoTaq® Reaction Buffer (Promega) before loading onto the gel. The gel was cast and run on a SUB10 Mini-Plus Submarine Gel Electrophoresis Unit (Hoefer) with 1x TBE buffer at 126 V.

Enzymatic ligation

In a typical enzymatic ligation reaction, equimolar amounts of each oligonucleotide were combined in a 200 µL microcentrifuge tube. 10x ligase buffer (from the manufacturer) was added, and nuclease-free water was used to adjust the buffer concentration as necessary. To anneal the oligonucleotides, the sample tube was heated for 5 min at 95°C, and then cooled at a rate of -1 °C per minute until at 22 °C. RNA samples were typically not annealed. Ligase was then added, and the reaction was incubated at the appropriate temperature. Typical DNA ligation reactions with T4 DNA ligase (Promega or New England Biolabs) were incubated for 4-24 hr at RT. Typical reactions with *Taq* ligase (New England Biolabs) were incubated for 20

temperature cycles (1 cycle = 15 min at 45 °C → 5 min at 85 °C) on a Bio-Rad T100TM Thermal Cycler. Typical RNA ligation reactions with T4 DNA ligase or T4 RNA ligase 2 (New England Biolabs) were incubated for 16 hr at 37 °C. In a typical reaction, 1 μL of ligase was used per 10 μL of reaction.

Ligation on streptavidin magnetic beads

Dynabeads™ M-280 or MyOne C1 streptavidin magnetic beads (Thermo Fisher Scientific) were vortexed to resuspend the beads, and 50 μL were transferred to a 1.5 mL microcentrifuge tube. The beads were washed 3 times with 100 μL 2x binding and washing (B&W) buffer, and resuspended in 50 μL 2x B&W buffer. To attach the oligonucleotides, 50 μL of 20 μM biotinylated oligonucleotide were added, and the sample tube was incubated for 1 hr at RT with gentle shaking. The tube was placed on a magnetic stand to remove the supernatant, followed by 3 washes with 100 μL 2X B&W buffer. The beads were then resuspended in 50 μL 2x B&W buffer and 50 μL of the next oligonucleotide in the sequence (20 μM concentration), and incubated for 30 min at RT with gentle shaking to hybridise the oligonucleotides. This was repeated until the entire desired sequence was hybridised. Next, the beads were washed with 50 μL 1x ligase buffer and resuspended in 20 μL 1x ligase buffer. T4 or *Taq* ligase was added (2 μL), and the reaction was incubated at the appropriate temperature. To cleave the oligonucleotides, 1:1 formamide (v/v) was added to the beads, and the mixture was heated at 95°C for 10 min. The supernatant was collected for analysis by denaturing PAGE.

Capture on streptavidin magnetic beads

First, oligonucleotides were ligated in solution and magnetic beads were washed, as described above. For the "attach-then-hybridise" method, the beads were resuspended in equal volume 20 μM biotinylated capture oligonucleotide and 2x B&W buffer, and incubated on an Eppendorf ThermoMixer for 2 hr at 22 °C and 300 rpm (the "attach" part). The beads were then washed twice with 2x B&W buffer, resuspended with the ligation products, and incubated for another 2 hr at 22 °C and 300 rpm (the "hybridise" part). For the "hybridise-then-attach" method, an equimolar amount of biotinylated capture oligonucleotide was added to the ligation products. The mixture was incubated for 2 hr at RT (the "hybridise" part), added to the magnetic beads, and incubated for 2 hr at 22 °C and 300 rpm (the "attach" part). Finally, for both methods, the products were cleaved by adding 1:1 formamide (v/v) and heating at 95°C for 10 min. The supernatant was collected for analysis by denaturing PAGE.

Templated RNA copper(I)-catalysed azide alkyne cycloaddition (CuAAC) ligation

Freshly prepared copper(II) sulfate pentahydrate, sodium ascorbate, and THPTA ligand in aqueous solution were added in a molar ratio of 1:10:7. A typical RNA CuAAC reaction used 200-250 molar equivalents of copper. The reaction was flushed with argon, and then incubated for 1 hr at RT. CuAAC reagents were removed with an illustra™ NAP™ size exclusion column (GE Healthcare) before loading the oligonucleotides on a PAGE gel.

Untemplated RNA CuAAC ligation
Equimolar amounts of the oligonucleotides were combined and freeze-dried in a microcentrifuge tube. The oligonucleotides were resuspended in reaction solution, such that the final concentrations were 5 mM $MgCl_2$, 0.2 M TEAA buffer, and 50% DMSO. Freshly dissolved ascorbic acid and a stock solution of 0.1M copper(II) sulfate pentahydrate-0.2 M THPTA in 55% DMSO were added in a molar ratio of 1:5:2 of copper:ascorbic acid:THPTA. A typical RNA CuAAC reaction used 200-250 molar equivalents of copper. The reaction was flushed with argon, and then incubated for approximately 2 hours at RT. CuAAC reagents were removed with an illustra™ NAP™ size exclusion column (GE Healthcare) before loading the oligonucleotides on a PAGE gel.

Templated RNA phosphoramidite ligation
Equimolar amounts of the oligonucleotides were combined and freeze-dried in a microcentrifuge tube. The oligonucleotides were resuspended in reaction buffer, such that the final concentrations were 0.2 M HEPES (pH 7.5) and 0.4 M NaCl. In a typical reaction, 1-(2-hydroxyethyl) imidazole and freshly dissolved EDC-HCl were then added to final concentrations of 0.1 M and 0.5 M, respectively. The reactions were incubated for 2 hr at RT. Before PAGE analysis, salts and reagents were removed with an illustra™ NAP™ size exclusion column (GE Healthcare).

MEGAscript® transcription
Transcription was performed with a MEGAscript® transcription kit (Invitrogen) according to the manufacturer's protocol. For each reaction, 2 µL each of ATP, CTP, GTP, UTP, and 10x reaction buffer were combined in a tube. 0.5 µL of RiboLock RNase Inhibitor, 1 µg of DNA template, and 2 µL of enzyme mix were then added, and the total volume was made up to 20 µL with nuclease-free water. The reaction was incubated overnight at 37 °C.

For transcription reactions primed with GMP, GTP was added at a ratio of 3:5 relative to each of the other triphosphates. GMP was then added at a ratio of 2:1 or 4:1 relative to GTP. All other factors were kept the same as unprimed transcription reactions.

Phenol:chloroform extraction
An equal volume of 25:24:1 phenol:chloroform:isoamyl alcohol (Invitrogen) was added to the sample vial, and mixed well by vortexing. The sample was then centrifuged for 5 minutes at 17,000 x g. The aqueous layer on top was removed and kept.

Ethanol precipitation
A double volume of cold absolute ethanol was added to the sample, and 3 M sodium acetate was added to a final concentration of 0.3 M. The sample was mixed by

vortexing, and kept at -80 °C for at least 1 hr. The sample was then centrifuged for 5 min at 17,000 x g. After removing the supernatant and washing the pellet twice with ethanol, the pellet was freeze dried to remove any residual ethanol.

MEGAclear™ transcription clean-up

MEGAclear™ kit (Invitrogen) was used according to the manufacturer's protocol. The volume of the RNA transcription reaction was brought up to 100 µL with Elution Solution. 350 µL of Binding Solution and 250 µL of ethanol were then added to the sample for a total volume of 700 µL. Next, the sample was added to a filter cartridge and centrifuged for 1 min at 12,500 x g, with the flow-through discarded. Two washes were performed by adding 500 µL of Wash Solution to the cartridge, centrifuging for 1 min at 12,500 x g, and discarding the flow-through. The transcribed RNA was eluted by adding 50 µL of elution solution to the cartridge, incubating for 10 min at 65 °C, and centrifuging again. The elution step was repeated to maximise RNA recovery. Elution Solution, Binding Solution, and Wash Solution were supplied with the kit.

Polymerase chain reaction with Taq *DNA polymerase*

For a typical 25 µL reaction, the following were combined in a tube: 2.5 µL 10x reaction buffer, 0.5 µL 10 mM dNTPs, 0.5 µL 10 µM forward primer, 0.5 µL 10 µM reverse primer, and 0.125 µL *Taq* DNA polymerase (New England Biolabs). 1 µg of template DNA was added, and the final volume made up to 25 µL with nuclease-free water. The reaction was incubated on a thermocycler with the following protocol: 2 min at 95 °C for initial denaturation; 30 cycles of 20 seconds at 95 °C, 30 seconds at 60 °C, and 60 seconds at 68 °C; and 5 min at 68 °C for the final extension.

QIAquick® PCR purification

PCR reactions were cleaned up with QIAquick® PCR Purification Kit (Qiagen) according to the manufacturer's protocol. Briefly, 5 volumes of Buffer PB was added to the PCR reaction, mixed, and applied to a QIAquick column. The column was centrifuged until all the liquid had passed through the column, for approximately 1 min. 0.75 mL of Buffer PE was added to the column and the column was centrifuged. To remove residual wash buffer, the column was centrifuged again for 1 min. Finally, 50 µL of water was added to the column, the column was centrifuged for 1 min, and the eluent was collected. Buffers PB and PE were supplied with the kit.

T4 polynucleotide kinase (PNK)

RNA or DNA was combined with buffer and T4 polynucleotide kinase (PNK; New England Biolabs). In the case of RNA, RiboLock RNase inhibitor (Thermo Fisher Scientific) was also included. T4 DNA ligase or T4 RNA ligase 2 buffer was used, since PNK is active in these buffers and they contain the required ATP. The reaction was incubated for 30 min at 37 °C, and PNK was inactivated by heating at 65 °C for 20 min. 10 units of RppH (1 µL) were used for for 300 pmol of 5′ termini.

RNA 5′ pyrophosphohydrolase (RppH)
RNA was combined with RiboLock RNase inhibitor (Thermo Fisher Scientific), RNA ligase 2 buffer, and RNA 5′ pyrophosphohydrolase (RppH; New England Biolabs). The reaction was incubated for 1 hr at 37 °C, and RppH was inactivated by heating at 65 °C for 5 min. 5 units of RppH (1 µL) were used for each 30 µL reaction.

Exonuclease T (Exo T)
RNA was combined with exonuclease T (Exo T; New England Biolabs) and Exo T buffer. The reaction was incubated for 20 min at 25 °C. 5 units of Exo T (1 µL) were used for each 20 µL reaction.

Enterokinase
10 µg of enterokinase (EK; Raybiotech) was suspended in 100 µL of the following buffer (for a final concentration of 0.1 µg/µL): 20 mM Tris-HCl (pH 8), 200 mM NaCl, 2 mM $CaCl_2$, and 50% glycerol. 0.5 µL of EK was added to 150 ng of RNA, for a total reaction volume of 5 µL in buffer (20 mM Tris-HCl, pH 8, 50 mM NaCl, 2 mM $CaCl_2$). The reaction was incubated for 16 hr at 25 °C.

PURExpress® in vitro *protein synthesis*
Following the manufacturer's protocol, 10 µL of Solution A were combined with 7.5 µL of Solution B. 1 µL of RiboLock RNase Inhibitor (Thermo Fisher Scientific, 40 U/µL) and DNA or RNA template were then added. RiboLock was omitted for translation reactions with cyclic RNA templates. The total volume was made up to 25 µL with nuclease-free water, and incubated for 16 hrs at 25 °C. The reaction was scaled up or down accordingly.

Tris-Tricine SDS-PAGE
Tris-tricine SDS-PAGE gels were run using an omniPAGE Mini vertical gel electrophoresis unit from Cleaver Scientific (1 mm depth x 10 cm width x 10 cm height). First, a 16% resolving gel was poured and allowed to set with a layer of water on top (per gel: 2.4 mL 40% acrylamide, 2 mL 3x gel buffer, 1.6 mL water, 50 µL 10% APS, 5 µL TEMED). The layer of water was tipped out, and a 4% stacking gel with wells was cast on top of the 16% resolving gel (per gel: 0.3 mL 40% acrylamide, 1 mL 3x gel buffer, 1.7 mL water, 45 µL 10% APS, 4.5 µL TEMED). Samples were prepared by mixing with 2x SDS-PAGE loading buffer, and boiling at 100 °C for 10 min. Color Prestained Protein Standard, Broad Range (New England Biolabs) was used as a marker and loaded directly onto the gel. Cathode buffer was added so that it covered the top of the gels, and anode buffer was added so that it covered the bottom of the gels. Once the samples were loaded, the gel was run at 60 V until the samples reached the resolving gel, and then the voltage was increased to 120 V. After the run was complete, the gel was washed with water and then left in Coomassie stain overnight. The gel was destained with Coomassie destain I for approximately 15 min, followed by Coomassie destain II until the background was clear.

Tris-Glycine gradient SDS-PAGE

Novex™ 10-20% Tris-Glycine mini gels were purchased from Thermo Fisher Scientific, and used with an omniPAGE Mini vertical gel electrophoresis unit from Cleaver Scientific (1 mm depth x 10 cm width x 10 cm height). Samples were prepared by mixing with 2x SDS-PAGE loading buffer, and boiling at 100 °C for 10 min. Color Prestained Protein Standard, Broad Range (New England Biolabs) was used as a marker and loaded directly onto the gel. The gel was submerged in 1x Tris-Glycine gel running buffer and run at 80 V until the sample reached the resolving gel, after which the voltage was increased to 180 V.

Dot blot

PVDF membranes (Amersham™ Hybond®, 0.2 µm pore size) were prepared by soaking in methanol for 15 seconds, water for 2 min, and PBS-T for 10 min. While still damp, 1.5-2 µL of translation reaction was pipetted onto the membrane. The membrane was then allowed to dry at RT. After rinsing first with PBS, the membrane was blocked in 5% milk/PBS-T for 30 min at RT and 300 rpm. Primary anti-FLAG antibody (mouse anti-FLAG M2 monoclonal antibody; Sigma) was added to the blocking solution at a 1:2,000 dilution, and incubated ON at 4 °C and 300 rpm. The membrane was washed by rinsing twice with PBS, and then washing three times in PBS-T for 5 min at RT and 400 rpm. Secondary anti-mouse antibody (rabbit anti-mouse IgG peroxidase conjugate; Sigma) was added in 5% milk/PBS-T at a 1:250,000 dilution, and incubated for 1 hr at RT and 300 rpm. The wash step was repeated, followed by two more rinses with PBS. The blot was visualised by applying SuperSignal™ West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific), blotting the excess with a paper towel, and imaging with a Bio-Rad Universal Hood III and Image Lab™ software.

Western blot

PVDF membranes (Amersham™ Hybond®, 0.2 µm pore size) were prepared by soaking in methanol for 15 seconds, water for 2 min, and 1x transfer buffer for 10 min. The transfer sandwich was assembled as follows: anode support grid, sponge, filter paper, membrane, SDS-PAGE gel, filter paper, sponge, cathode support grid. The sandwich was then placed in an omniPAGE Mini tank with 1x transfer buffer, and run at 400 mA for 1.5 hr. Ice was placed around the tank to cool the temperature. After rinsing the membrane with PBS, the blocking, primary antibody, secondary antibody, and visualisation steps were the same as described above for dot blots.

Fluorophore-conjugated antibody assay

Black 96-well MaxiSorp™ plates (Thermo Fisher Scientific) were coated with 100 µL of either FLAG-Alexa 488 or HA-Alexa 647 antibody (Life Technologies) in carbonate/bicarbonate buffer at 2.5 ng/µL. The plate was covered and incubated at 4 °C ON. The coating solution was removed and the plate was washed 3 times, for 5 min with 300 µL PBS-T per wash. The wells were blocked with 200 µL blocking

buffer and incubated for 2 hr at RT, and the wash step was repeated. 100 µL of the sample in blocking buffer was added to each well and incubated for 1.5 hr at 37 °C, followed by the same wash step. Either FLAG-Alexa 488 or HA-Alexa 647 antibody was added in 100 µL blocking buffer at 2.5 ng/µL as the detection antibody. The plate was incubated for 2 hr at RT and washed 5 times. The wells were resuspended with 200 µL PBS-T and the plate was read with a CLARIOstar plate reader (Alexa 488: excitation at 488 nm, emission at 535 nm; Alexa 647: excitation at 625 nm, emission at 680 nm).

Luciferase glow assay

Translation reactions with luciferase templates were assayed for activity with the Pierce® Renilla Luciferase Glow Assay Kit from Thermo Fisher Scientific. 2 µL of translation reaction were added to 48 µL of lysis buffer in a 96-well solid white opaque plate (Porvair Science). Working solution was prepared from coelenterazine and Renilla glow assay buffer according to the manufacturer's protocol, and 50 µL was added to each plate. The samples were mixed by pipetting, and imaged after waiting approximately 10 min. Luminescence was measured using a PHERAstar microplate reader (LUM plus module) and Bio-Rad Universal Hood III imager.

Chapter 7: References

References

1. Voet, D., & Voet, J. G. (2004). *Biochemistry* (3rd ed.). John Wiley & Sons, Inc.
2. Cooper, G. (2000). *The Cell: A molecular approach. 2nd edition*. ASM Press.
3. Brown, T., & Brown, T. J. *Nucleic Acids Book*. https://www.atdbio.com/nucleic-acids-book
4. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids. *Nature* **171,** 737–738 (1953).
5. Chargaff, E. Chemical Specificity of Nucleic Acids and Mechanisms of Their Enzymatic Degradation. *Experientia* **6,** 201-209 (1950).
6. Wilkins, M. H. F. & Randall, J. T. Crystallinity in sperm heads: Molecular structure of nucleoprotein in vivo. *Biochim. Biophys. Acta* **10,** 192–193 (1953).
7. Wing, R. *et al.* Crystal structure analysis of a complete turn of B-DNA. *Nature* **287,** 755–758 (1980).
8. Yakovchuk, P., Protozanova, E. & Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **34,** 564–574 (2006).
9. Saiki, R. K. *et al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239,** 487–91 (1988).
10. Chien, A., Edgar, D. B. & Trela, J. M. Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus. *J. Bacteriol.* **127,** 1550–7 (1976).
11. Kary B. Mullis. Nobel Lecture: The Polymerase Chain Reaction. Available at: https://www.nobelprize.org/prizes/chemistry/1993/mullis/lecture/
12. Michelson, A. M. & Todd, A. R. Nucleotides part XXXII. Synthesis of a dithymidine dinucleotide containing a 3′: 5′-internucleotidic linkage. *J. Chem. Soc.* **0,** 2632–2638 (1955).
13. Gilham, P. T. & Khorana, H. G. Studies on Polynucleotides. I. A New and General Method for the Chemical Synthesis of the C_5''-C_3'' Internucleotidic Linkage. Syntheses of Deoxyribo-dinucleotides. *J. Am. Chem. Soc.* **80,** 6212–6222 (1958).
14. Letsinger, R. L. & Mahadevan, V. Stepwise Synthesis of Oligodeoxyribonucleotides on an Insoluble Polymer Support. *J. Am. Chem. Soc.* **88,** 5319–5324 (1966).
15. Beaucage, S. L. & Caruthers, M. H. Deoxynucleoside phosphoramidites—A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* **22,** 1859–1862 (1981).
16. Caruthers, M. H. Gene synthesis machines: DNA chemistry and its uses. *Science* **230,** 281–5 (1985).
17. Caruthers, M. H. A brief review of DNA and RNA chemical synthesis. *Biochem. Soc. Trans.* **39,** 575–580 (2011).
18. Caruthers, M. H. The Chemical Synthesis of DNA/RNA: Our Gift to Science. *J. Biol. Chem.* **288,** 1420–1427 (2013).

19. Sproat, B. S. RNA Synthesis Using 2'-O-(Tert-Butyldimethylsilyl) Protection. *Methods Mol. Bio.* **288,** 17-32 (2005).
20. Dellinger, D. J. *et al.* Streamlined Process for the Chemical Synthesis of RNA Using 2′- *O* -Thionocarbamate-Protected Nucleoside Phosphoramidites in the Solid Phase. *J. Am. Chem. Soc.* **133,** 11540–11556 (2011).
21. Baronti, L., Karlsson, H., Marušič, M. & Petzold, K. A guide to large-scale RNA sample preparation. *Anal. Bioanal. Chem.* **410,** 3239–3252 (2018).
22. Schmitz, C. & Reetz, M. T. Solid-Phase Enzymatic Synthesis of Oligonucleotides. *Org. Lett.***1,** 1729-1731 (1999).
23. Jensen, M. A. & Davis, R. W. Template-Independent Enzymatic Oligonucleotide Synthesis (TiEOS): Its History, Prospects, and Challenges. *Biochemistry* **57,** 1821–1832 (2018).
24. Palluk, S. *et al.* De novo DNA synthesis using polymerase-nucleotide conjugates. *Nat. Biotechnol.* **36,** 645–650 (2018).
25. Clore, A. A new route to synthetic DNA. *Nat. Biotechnol.* **36,** 593–595 (2018).
26. Shabarova, Z. A. *et al.* Chemical ligation of DNA: the first non-enzymatic assembly of a biologically active gene. *Nucleic Acids Res.* **19,** 4247–4251 (1991).
27. Kukwikila, M., Gale, N., El-Sagheer, A. H., Brown, T. & Tavassoli, A. Assembly of a biocompatible triazole-linked gene by one-pot click-DNA ligation. *Nat. Chem.* **9,** 1089–1098 (2017).
28. Manuguerra, I. *et al.* Gene assembly via one-pot chemical ligation of DNA promoted by DNA nanostructures. *Chem. Commun. (Camb).* **54,** 4529–4532 (2018).
29. El-Sagheer, A. H. & Brown, T. Single tube gene synthesis by phosphoramidate chemical ligation. *Chem. Commun.* **53,** 10700–10702 (2017).
30. Gibson, D. G. *et al.* Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **319,** 1215–20 (2008).
31. Smith, H. O., Hutchison, C. A., Pfannkoch, C. & Venter, J. C. Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 15440–5 (2003).
32. Kodumal, S. J. *et al.* Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 15573–8 (2004).
33. Stemmer, W. P. C., Crameri, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164,** 49–53 (1995).
34. Schindler, D., Dai, J. & Cai, Y. Synthetic genomics: a new venture to dissect genome fundamentals and engineer new functions. *Curr. Opin. Chem. Biol.* **46,** 56–62 (2018).
35. Wang, L. *et al.* Synthetic Genomics: From DNA Synthesis to Genome Design. *Angew. Chemie Int. Ed.* **57,** 1748–1756 (2018).
36. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **11,** 499–507 (2014).

37. Xiong, A.-S. *et al.* Non-polymerase-cycling-assembly-based chemical gene synthesis: Strategies, methods, and progress. *Biotechnol. Adv.* **26,** 121–134 (2008).
38. Au, L.-C., Yang, F.-Y., Yang, W.-J., Lo, S.-H. & Kao, C.-F. Gene Synthesis by a LCR-Based Approach: High-Level Production of Leptin-L54 Using Synthetic Gene in Escherichia coli. *Biochem. Biophys. Res. Commun.* **248,** 200–203 (1998).
39. Kneidinger, B., Graninger, M. & Messner, P. Scaling up the ligase chain reaction-based approach to gene synthesis. *Biotechniques* **3,** 249–252 (2001).
40. Gibson, D. G. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res.* **37,** 6984–6990 (2009).
41. Gibson, D. G. Enzymatic Assembly of Overlapping DNA Fragments. *Methods Enzymol.* **498,** 349–361 (2011).
42. Crick, F. H. C. On protein synthesis. *Symp Soc Exp Biol* **12,** 138–163 (1958).
43. Crick, F. Central dogma of Molecular Biology. *Nature* **227,** 561–563 (1970).
44. Mott, M. L. & Berger, J. M. DNA replication initiation: mechanisms and regulation in bacteria. *Nat. Rev. Microbiol.* **5,** 343–354 (2007).
45. Alberts, B. M. Prokaryotic DNA replication mechanisms. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **317,** 395–420 (1987).
46. Griffiths AJF, Miller JH, Suzuki DT, et al. An Introduction to Genetic Analysis. 7th edition. New York: W. H. Freeman; 2000. Mechanism of DNA replication. Available from: https://www.ncbi.nlm.nih.gov/books/NBK21862/
47. Mcclure, W. R. Mechanism and control of transcription initiation in prokaryotes. *Ann. Rev. Biochem* **54,** 171-204 (1985).
48. Platt, T. Transcription termination and the regulation of gene expression. *Ann. Rev. Biochem.* **55,** 339–372 (1986).
49. Henkin, T. M. Control of transcription termination in prokaryotes. *Annu. Rev. Genet.* **30,** 35–57 (1996).
50. Kozak, M. Initiation of translation in prokaryotes and eukaryotes. *Gene* **234,** 187–208 (1999).
51. McCarthy, J. E. G. & Brimacombie, R. Prokaryotic translation: the interactive pathway leading to initiation. *Trends Genet.* **10,** 402–407 (1994).
52. Gualerzi, C. O. & Pon, C. L. Initiation of mRNA Translation in Prokaryotes. *Biochemistry* **29,** 5881–5889 (1990).
53. Gingold, H. & Pilpel, Y. Determinants of translation efficiency and accuracy. *Mol. Syst. Biol.* **7,** 481 (2011).
54. Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361,** 13–37 (2005).
55. Zhang, Q. *et al.* Recent Methods for Purification and Structure Determination of Oligonucleotides. *Int. J. Mol. Sci.* **17,** (2016).
56. Stellwagen, N. C. Electrophoresis of DNA in agarose gels, polyacrylamide gels and in free solution. *Electrophoresis* **30,** S188–S195 (2009).
57. Barril, P. & Nates, S. Introduction to Agarose and Polyacrylamide Gel Electrophoresis Matrices with Respect to Their Detection Sensitivities. *Gel Electrophoresis – Principles and Basics* (2012).

58. Chrambach, A. & Rodbard, D. Polyacrylamide gel electrophoresis. *Science* **172,** 440–451 (1971).
59. Hegyi, G. *et al.* 7.3. Polyacrylamide gel electrophoresis (PAGE). *Introduction to Practical Biochemistry* (2013). Available at: http://elte.prompt.hu/sites/default/files/tananyagok/IntroductionToPracticalBiochemistry/ch07s03.html. (Accessed: 30th September 2018)
60. Serwer, P. Agarose gels: Properties and use for electrophoresis. *Electrophoresis* **4,** 375–382 (1983).
61. Lee, P. Y., Costumbrado, J., Hsu, C.-Y. & Kim, Y. H. Agarose gel electrophoresis for the separation of DNA fragments. *J. Vis. Exp.* **62**, 3923 (2012).
62. Kurien, B. T. & Scofield, R. H. (2012). *Protein Electrophoresis*. Humana Press.
63. Hames, B. D. (1998) *Gel electrophoresis of proteins : a practical approach*. Oxford University Press.
64. Sinha, N. D. & Jung, K. E. Analysis and Purification of Synthetic Nucleic Acids Using HPLC. in *Current Protocols in Nucleic Acid Chemistry* **61,** 10.5.1-10.5.39 (John Wiley & Sons, Inc., 2015).
65. Andrus, A. & Kuimelis, R. G. Analysis and Purification of Synthetic Nucleic Acids Using HPLC. in *Current Protocols in Nucleic Acid Chemistry* **Chapter 10,** 10.5.1-10.5.13 (John Wiley & Sons, Inc., 2000).
66. Calculate extinction coefficient of DNA and RNA on-line. Available at: http://www.owczarzy.net/extinctionDNA.htm. (Accessed: 30th September 2018)
67. Li, X. *et al.* Comparison of three common DNA concentration measurement methods. *Anal. Biochem.* **451,** 18–24 (2014).
68. Understanding and measuring variations in DNA sample quality | Articles | Oxford Gene Technology. Available at: https://www.ogt.com/resources/literature/483_understanding_and_measuring_variations_in_dna_sample_quality. (Accessed: 30th September 2018)
69. Gallagher, S. Quantitation of Nucleic Acids with Absorption Spectroscopy. in *Current Protocols in Protein Science* **13,** A.4K.1-A.4K.3 (John Wiley & Sons, Inc., 1998).
70. Guo, C., Kong, D., Lei, Y. & Hili, R. Expanding the Chemical Diversity of DNA. *Synlett* **29**, 1405-1414 (2018).
71. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).
72. Kok, S. de *et al.* Rapid and Reliable DNA Assembly *via* Ligase Cycling Reaction. *ACS Synth. Biol.* **3,** 97–106 (2014).
73. Chandran, S. Rapid Assembly of DNA via Ligase Cycling Reaction (LCR). *Methods Mol. Biol.* **1472,** 105–110 (2017).
74. Lehman, I. R. DNA ligase: structure, mechanism, and function. *Science* **186,** 790–7 (1974).
75. Gaastra, W. & Hansen, K. Ligation of DNA with T4 DNA Ligase. *Methods Mol. Biol.* **2**, 225–230 (1985).
76. Kershaw, C. J. & O'Keefe, R. T. Splint ligation of RNA with T4 DNA ligase. *Methods Mol. Biol.* **941,** 257–69 (2012).

77. Bullard, D. R. & Bowater, R. P. Direct comparison of nick-joining activity of the nucleic acid ligases from bacteriophage T4. *Biochem. J.* **398,** 135–44 (2006).
78. Lohman, G. J. S., Tabor, S. & Nichols, N. M. DNA Ligases. in *Current Protocols in Molecular Biology* **94,** 3.14.1-3.14.7 (John Wiley & Sons, Inc., 2011).
79. Cherepanov, A. V. & De Vries, S. Kinetics and thermodynamics of nick sealing by T4 DNA ligase. *Eur. J. Biochem.* **270,** 4315–4325 (2003).
80. Ferretti, L. & Sgaramella, V. Temperature dependence of the joining by T4 DNA ligase of termini produced by type II restriction endonucleases. *Nucleic Acids Res.* **9,** 85–93 (1981).
81. Pheiffer, B. H. & Zimmerman, S. B. Polymer-stimulated ligation: enhanced blunt-Or cohesive-end ligation of DNA or deoxyribooligonucleotides by T4 DNA ligase in polymer solutions. *Nucleic Acids Res.* **11,** 7853–7871 (1983).
82. Hayashi, K., Nakazawa, M., Ishizaki, Y., Hiraoka, N. & Obayashi, A. Regulation of inter- and intramolecular ligation with T4 DNA ligase in the presence of polyethylene glycol. *Nucleic Acids Res.* **14,** 7617–7631 (1986).
83. Ho, C. K. & Shuman, S. Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 12709–14 (2002).
84. Stark, M. R., Pleiss, J. A., Deras, M., Scaringe, S. A. & Rader, S. D. An RNA ligase-mediated method for the efficient creation of large, synthetic RNAs. *RNA* **12,** 2014–9 (2006).
85. Chauleau, M. & Shuman, S. Kinetic mechanism of nick sealing by T4 RNA ligase 2 and effects of 3′ -OH base mispairs and damaged base lesions. *RNA* **19,** 1–8 (2013).
86. Nandakumar, J., Ho, C. K., Lima, C. D. & Shuman, S. RNA substrate specificity and structure-guided mutational analysis of bacteriophage T4 RNA ligase 2. *J. Biol. Chem.* **279,** 31337–47 (2004).
87. Smith, H. O., Hutchison, C. A., Pfannkoch, C. & Venter, J. C. Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 15440–5 (2003).
88. Gibson, D. G., Smith, H. O., Hutchison, C. A., Venter, J. C. & Merryman, C. Chemical synthesis of the mouse mitochondrial genome. *Nat. Methods* **7,** 901–903 (2010).
89. Roth, T. L., Milenkovic, L. & Scott, M. P. A Rapid and Simple Method for DNA Engineering Using Cycled Ligation Assembly. *PLoS One* **9,** e107329 (2014).
90. Lillehaug, J. R. & Kleppe, K. Kinetics and Specificity of T4 Polynucleotide Kinase. *Biochemistry* **14,** 1221–1225 (1975).
91. Wang, L. K., Lima, C. D. & Shuman, S. Structure and mechanism of T4 polynucleotide kinase: an RNA repair enzyme. *EMBO J.* **21,** 3873–80 (2002).
92. Caldecott, K. W. Polynucleotide Kinase: A Versatile Molecule Makes a Clean Break. *Structure* **10,** 1151–1152 (2002).
93. Chivers, C. E., Koner, A. L., Lowe, E. D. & Howarth, M. How the biotin-streptavidin interaction was made even stronger: investigation via crystallography and a chimaeric tetramer. *Biochem. J.* **435,** 55–63 (2011).

94. Holmberg, A. *et al.* The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis* **26,** 501–510 (2005).
95. Jensen, M. A., Akhras, M. S., Fukushima, M., Pourmand, N. & Davis, R. W. Direct oligonucleotide synthesis onto super-paramagnetic beads. *J. Biotechnol.* **167,** 448–53 (2013).
96. Pengpumkiat, S., Koesdjojo, M., Rowley, E. R., Mockler, T. C. & Remcho, V. T. Rapid Synthesis of a Long Double-Stranded Oligonucleotide from a Single-Stranded Nucleotide Using Magnetic Beads and an Oligo Library. *PLoS One* **11,** e0149774 (2016).
97. Wang, J., Kawde, A.-N., Erdem, A. & Salazar, M. Magnetic bead-based label-free electrochemical detection of DNA hybridization. *Analyst* **126,** 2020–2024 (2001).
98. Wilson, R. Preparation of Single-Stranded DNA from PCR Products with Streptavidin Magnetic Beads. *Nucleic Acid Ther.* **21,** 437–440 (2011).
99. Su, Q., Xing, D. & Zhou, X. Magnetic beads based rolling circle amplification–electrochemiluminescence assay for highly sensitive detection of point mutation. *Biosens. Bioelectron.* **25,** 1615–1621 (2010).
100. Hultman, T., Stahl, S., Homes, E. & Uhlén, M. Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support. *Nucleic Acids Res.* **17,** 4937–4946 (1989).
101. Zhovmer, A. & Qu, X. Proximal disruptor aided ligation (ProDAL) of kilobase-long RNAs. *RNA Biol.* **13,** 613–621 (2016).
102. Dallmann, A. *et al.* Structure and Dynamics of Triazole-Linked DNA: Biocompatibility Explained. *Chem. - A Eur. J.* **17,** 14714–14717 (2011).
103. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chemie Int. Ed.* **40,** 2004–2021 (2001).
104. Osman, E. A., Gadzikwa, T. & Gibbs, J. M. Quick Click: The DNA-Templated Ligation of 3′- *O* -Propargyl- and 5′-Azide-Modified Strands Is as Rapid as and More Selective than Ligase. *ChemBioChem* (2018).
105. 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. Chemie Int. Ed.* **41,** 2596–2599 (2002).
106. Liang, L. & Astruc, D. The copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) "click" reaction and its applications. An overview. *Coord. Chem. Rev.* **255,** 2933–2945 (2011).
107. Zhu, L., Brassard, C. J., Zhang, X., Guha, P. M. & Clark, R. J. On the Mechanism of Copper(I)-Catalyzed Azide-Alkyne Cycloaddition. *Chem. Rec.* **16,** 1501–1517 (2016).
108. Hong, V., Presolski, S. I., Ma, C. & Finn, M. G. Analysis and Optimization of Copper-Catalyzed Azide-Alkyne Cycloaddition for Bioconjugation. *Angew. Chemie* **121,** 10063–10067 (2009).
109. Finn, M. G. & Fokin, V. V. Click chemistry: function follows form. *Chem. Soc. Rev.* **39,** 1231 (2010).

110. El-Sagheer, A. H. & Brown, T. Click chemistry with DNA. *Chem. Soc. Rev.* **39,** 1388 (2010).
111. El-Sagheer, A. H. & Brown, T. A triazole linkage that mimics the DNA phosphodiester group in living systems. *Q. Rev. Biophys.* **48,** 429–436 (2015).
112. El-Sagheer, A. H., Sanzone, A. P., Gao, R., Tavassoli, A. & Brown, T. Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 11338–43 (2011).
113. El-Sagheer, A. H. & Brown, T. Efficient RNA synthesis by in vitro transcription of a triazole-modified DNA template. *Chem. Commun.* **47,** 12057 (2011).
114. Sanzone, A. P., El-Sagheer, A. H., Brown, T. & Tavassoli, A. Assessing the biocompatibility of click-linked DNA in Escherichia coli. *Nucleic Acids Res.* **40,** 10567–75 (2012).
115. Birts, C. N. *et al.* Transcription of Click-Linked DNA in Human Cells. *Angew. Chemie Int. Ed.* **53,** 2362–2365 (2014).
116. 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. U. S. A.* **107,** 15329–34 (2010).
117. Fujino, T. *et al.* Chimeric RNA Oligonucleotides Incorporating Triazole-Linked Trinucleotides: Synthesis and Function as mRNA in Cell-Free Translation Reactions. *J. Org. Chem.* **81,** 8967–8976 (2016).
118. Kalinowski, M. *et al.* Phosphoramidate Ligation of Oligonucleotides in Nanoscale Structures. *ChemBioChem* **17,** 1150–1155 (2016).
119. Vasilyev, N. & Serganov, A. Structures of RNA Complexes with the Escherichia coli RNA Pyrophosphohydrolase RppH Unveil the Basis for Specific 5′-End-Dependent mRNA Decay. *J. Biol. Chem.* **290,** 9487–9499 (2015).
120. Foley, P. L., Hsieh, P.-K., Luciano, D. J. & Belasco, J. G. Specificity and Evolutionary Conservation of the Escherichia coli RNA Pyrophosphohydrolase RppH. *J. Biol. Chem.* **290,** 9478–9486 (2015).
121. Deana, A., Celesnik, H. & Belasco, J. G. The bacterial enzyme RppH triggers messenger RNA degradation by 5′ pyrophosphate removal. *Nature* **451,** 355–358 (2008).
122. Celesnik, H., Deana, A. & Belasco, J. G. Initiation of RNA Decay in Escherichia coli by 5′ Pyrophosphate Removal. *Mol. Cell* **27,** 79–90 (2007).
123. Richards, J. *et al.* An RNA Pyrophosphohydrolase Triggers 5′-Exonucleolytic Degradation of mRNA in Bacillus subtilis. *Mol. Cell* **43,** 940–949 (2011).
124. Shis, D. L. & Bennett, M. R. Synthetic biology: the many facets of T7 RNA polymerase. *Mol. Syst. Biol.* **10,** 745 (2014).
125. Arnaud-Barbe, N., Cheynet-Sauvion, V., Oriol, G., Mandrand, B. & Mallet, F. Transcription of RNA templates by T7 RNA polymerase. *Nucleic Acids Res.* **26,** 3550–3554 (1998).
126. Tabor, S. Expression Using the T7 RNA Polymerase/Promoter System. in *Current Protocols in Molecular Biology* **Chapter 16,** Unit16.2 (John Wiley & Sons, Inc., 2001).

127. Tabor, S. & Richardson, C. C. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl. Acad. Sci. U. S. A.* **82,** 1074–8 (1985).
128. Mairhofer, J., Wittwer, A., Cserjan-Puschmann, M. & Striedner, G. Preventing T7 RNA Polymerase Read-through Transcription—A Synthetic Termination Signal Capable of Improving Bioprocess Stability. *ACS Synth. Biol.* **4,** 265–273 (2015).
129. Köhrer, C., Mayer, C., Gröbner, P. & Piendl, W. Use of T7 RNA polymerase in an optimized Escherichia coli coupled in vitro transcription-translation system. Application in regulatory studies and expression of long transcription units. *Eur. J. Biochem.* **236,** 234–9 (1996).
130. Chong, S. Overview of cell-free protein synthesis: historic landmarks, commercial systems, and expanding applications. *Curr. Protoc. Mol. Biol.* **108,** 16.30.1-11 (2014).
131. Cai, Q. *et al.* A simplified and robust protocol for immunoglobulin expression in Escherichia coli cell-free protein synthesis systems. *Biotechnol. Prog.* **31,** 823–31 (2015).
132. Martin, R. W. *et al.* Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids. *Nat. Commun. 2018 91* **9,** 1203 (2018).
133. Zemella, A., Thoring, L., Hoffmeister, C. & Kubick, S. Cell-Free Protein Synthesis: Pros and Cons of Prokaryotic and Eukaryotic Systems. *Chembiochem* **16,** 2420–31 (2015).
134. Nevin, D. E. & Pratt, J. M. A coupled in vitro transcription-translation system for the exclusive synthesis of polypeptides expressed from the T7 promoter. *FEBS Lett.* **291,** 259–263 (1991).
135. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19,** 751–755 (2001).
136. Crick, F. H. Codon--anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* **19,** 548–55 (1966).
137. Murphy, F. V & Ramakrishnan, V. Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nat. Struct. Mol. Biol.* **11,** 1251–1252 (2004).
138. Cochella, L. & Green, R. Wobble during decoding: more than third-position promiscuity. *Nat. Struct. Mol. Biol.* **11,** 1160–1162 (2004).
139. Overview of ELISA. Available at: https://www.thermofisher.com/uk/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-elisa.html.
140. Van Weemen, B. K. & Schuurs, A. H. W. M. Immunoassay using antigen-enzyme conjugates. *FEBS Lett.* **15,** 232–236 (1971).
141. Engvall, E. & Perlmann, P. Enzyme-linked immunosorbent assay (ELISA) quantitative assay of immunoglobulin G. *Immunochemistry* **8,** 871–874 (1971).

142. Lequin, R. M. Enzyme immunoassay (EIA)/enzyme-linked immunosorbent assay (ELISA). *Clin. Chem.* **51,** 2415–8 (2005).
143. Bergmans, H. E. N. & Gaastra, W. Dot-Blot Hybridization Method. in *New Nucleic Acid Techniques* **4,** 385–390 (Humana Press, 1988).
144. Rupprecht, K. R. *et al.* Development of a dot-blot assay for screening monoclonal antibodies to low-molecular-mass drugs. *Anal. Biochem.* **407,** 160–164 (2010).
145. Tian, G. *et al.* Quantitative dot blot analysis (QDB), a versatile high throughput immunoblot method. *Oncotarget* **8,** 58553–58562 (2017).
146. Towbin, H., Staehelin, T. & Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. U. S. A.* **76,** 4350–4 (1979).
147. Burnette, W. N. "Western Blotting": Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal. Biochem.* **112,** 195–203 (1981).
148. Mahmood, T. & Yang, P.-C. Western blot: technique, theory, and trouble shooting. *N. Am. J. Med. Sci.* **4,** 429–34 (2012).
149. Hoernes, T. P. *et al.* Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. *Nucleic Acids Res.* **44,** 852–862 (2016).
150. Mukherjee, A. *et al.* Engineering and Characterization of New LOV-Based Fluorescent Proteins from *Chlamydomonas reinhardtii* and *Vaucheria frigida*. *ACS Synth. Biol.* **4,** 371–377 (2015).
151. 1. Ren, G.-X., Guo, X.-P. & Sun, Y.-C. Regulatory 3' Untranslated Regions of Bacterial mRNAs. *Front. Microbiol.* **8,** 1276 (2017).
152. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna RNA Websuite. *Nucleic Acids Res.* (2008).
153. Sampson, J. R. & Uhlenbeck, O. C. Biochemical and physical characterization of an unmodified yeast phenylalanine transfer RNA transcribed *in vitro*. *Proc. Natl. Acad. Sci. USA* **85,** 1033-1037 (1988).
154. Luo, Y., Eldho, N. V., Sintim, H. O. & Dayie, T. K. RNAs synthesized using photocleavable biotinylated nucleotides have dramatically improved catalytic efficiency. *Nucleic Acids Res.* **39,** 8559–8571 (2011).
155. Abe, N. *et al.* Rolling Circle Amplification in a Prokaryotic Translation System Using Small Circular RNA. *Angew. Chemie Int. Ed.* **52,** 7004–7008 (2013).
156. Abe, N. *et al.* Rolling Circle Translation of Circular RNA in Living Human Cells. *Sci. Rep.* **5,** 16435 (2015).
157. Kumar, P., Truong, L., Baker, Y. R., El-Sagheer, A. H. & Brown, T. Synthesis, Affinity for Complementary RNA and DNA, and Enzymatic Stability of Triazole-Linked Locked Nucleic Acids (t-LNAs). *ACS Omega* **3,** 6976–6987 (2018).

158. Kumar, P., El-Sagheer, A. H., Truong, L. & Brown, T. Locked nucleic acid (LNA) enhances binding affinity of triazole-linked DNA towards RNA. *Chem. Commun. (Camb).* **53,** 8910–8913 (2017).
159. El-Sagheer, A. H. & Brown, T. Combined nucleobase and backbone modifications enhance DNA duplex stability and preserve biocompatibility. *Chem. Sci.* **5,** 253–259 (2014).
160. El-Sagheer, A. H. & Brown, T. Click Nucleic Acid Ligation: Applications in Biology and Nanotechnology. *Acc. Chem. Res.* **45,** 1258–1267 (2012).

Appendix

Appendix: Sequences

Key: p = phosphate, FAM = fluorescein, Cy3 = cyanine 3, alk = alkyne, az = azide, am = amine, EM = expected mass, MM = measured mass, MD = mass difference

eGFP ssDNA Sequences

Note: The 5′-biotin and 3′-propanol groups on S1-9 (splints) were for another application, and not necessary for the ligation reactions discussed in this study.

Name	nt	5′	Sequence (5′ → 3′)	3′	EM	MM	MD
O1	61	p	TCGACGGTACCGCGGGCCCGGGATCCACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGC		18973	18973	0
O2	95	p	TGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACC		29541	29542	+1
O3	76	p	TACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCC		23193	23193	0
O4	83	p	TGACCTACGGTGTACAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGC		25463	25464	+1
O5	82	p	TACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCC		25321	25323	+2
O6	77	p	TGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC		23980	23981	+1
O7	60	p	TACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAAC		18562	18561	-1
O8	88	p	TTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGC		26986	26988	+2
O9	81	p	TGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGC		24817	24819	+2
O10	59	p	TGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGC		18318	18317	-1

S1	30	biotin	CCACCCCGGTGAACAGCTCCTCGCCCTTGC	propanol	9566	9568	-2
S2	30	biotin	GGTCAGCTTGCCGTAGGTGGCATCGCCCTC	propanol	9732	9734	-2
S3	30	biotin	GTACACCGTAGGTCAGGGTGGTCACGAGGG	propanol	9894	9896	-2
S4	30	biotin	GCGCTCCTGGACGTAGCCTTCGGGCATGGC	propanol	9757	9759	-2
S5	30	biotin	CGATGCGGTTCACCAGGGTGTCGCCCTCGA	propanol	9741	9743	-2
S6	30	biotin	GTTGTGGCTGTTGTAGTTGTACTCCAGCTT	propanol	9782	9784	-2
S7	30	biotin	GTGGCGGATCTTGAAGTTCACCTTGATGCC	propanol	9770	9772	-2
S8	30	biotin	GGTTGTCGGGCAGCAGCACGGGGCCGTCGC	propanol	9847	9849	-2
S9	30	biotin	CGGTCACGAACTCCAGCAGGACCATGTGAT	propanol	9742	9744	-2
O1a	30	biotin, FAM	TCGACGGTACCGCGGGCCCGGGATCCACCG		10113	10112	-1
O1b	31	p	GTCGCCACCATGGTGAGCAAGGGCGAGGAGC		9716	9717	+1
S0	20		TGGTGGCGACCGGTGGATCC		3190	6189	-1
O10 capture	40	biotin	TTTTTTTTTTGCTTTACTTGTACAGCTCGTCCATGCCGAG		12596	12596	0

eGFP dsDNA Sequences

Name	nt	5′	Sequence (5′ → 3′)	3′	EM	MM	MD
F1	38	p, biotin, dU	TCGACGGTACCGCGGGCCCGGGATCCACCGGTCGCCAC		12313	12313	0
F2	41	p	CATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGC		12908	12909	+1
F3	35	p	CCATCCTGGTCGAGCTGGACGGCGACGTAAACGGC		10863	10863	0
F4	41	p	CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC		12782	12783	+1
F5	36	p	CTACGGCAAGCTGACCCTGAAGTTCATCTGCACCAC		11030	11030	0
F6	41	p	CGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCC		12470	12470	0
R1	39	p	TTGCTCACCATGGTGGCGACCGGTGGATCCCGGGCCCGC		12058	12058	0
R2	37	p	CAGGATGGGCACCACCCCGGTGAACAGCTCCTCGCCC		11346	11347	+1
R3	37	p	TGAACTTGTGGCCGTTTACGTCGCCGTCCAGCTCGAC		11388	11388	0
R4	41	p	CTTGCCGTAGGTGGCATCGCCCTCGCCCTCGCCGGACACGC		12556	12556	0
R5	36	p	CAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAG		11237	11236	-1
R6	41	p	CCGTAGGTCAGGGTGGTCACGAGGGTGGGCCAGGGCACGGG		12934	12934	0

ermC Leader Peptide RNA Template Sequences

Name	nt	5′	Sequence (5′ → 3′)	3′	EM	MM	MD
ermCL-R1-c	57	p	UAUUUUAAUAAUUAAUUAACUUUAGUAAGGAGGAACAAUAUGGGCAUUUUUAGUAUU	alk	18340	18338	-2
ermCL-R2-c	42	az	UUUGUAAUCAGCACAGUUCAUUAUCAACCAAACAAAAAAUAA		13392	13392	0
ermCL-S	30		TGTGCTGATTACAAAAATACTAAAAATGCC		9206	9207	+1
ermCL-R1-e	57	p	UAUUUUAAUAAUUAAUUAACUUUAGUAAGGAGGAACAAUAUGGGCAUUUUUAGUAUU		18302	18300	-2
ermCL-R2-e	42	p	UUUGUAAUCAGCACAGUUCAUUAUCAACCAAACAAAAAAUAA		13447	13446	-1

FLAG-HA Peptide RNA Template Sequences

Name	nt	5′	Sequence (5′ → 3′)	3′	EM	MM	MD
FLAGHA-R1-c	57	p	UUAACUUUAUAAGGAGGUAACAAUAUGGAUUAUAAAGAUGACGAUGACAAAGCGGGC	alk	18578	18578	0
FLAGHA-R2-c	42	az	CUGGGUGCGGGUUAUCCGUACGAUGUGCCGGAUUAUGCCUAA		13521	13521	0
FLAGHA-S	40		TACGGATAACCCGCACCCAGGCCCGCTTTGTCATCGTCAT		12153	12153	0
FLAGHA-S_{RNA}	30		AUAACCCGCACCCAGGCCCGCUUUGUCAUC		9444	9444	0
FLAGHA-R1-e	57	p	UUAACUUUAUAAGGAGGUAACAAUAUGGAUUAUAAAGAUGACGAUGACAAAGCGGGC		18540	18538	-2
FLAGHA-R2-e	42	p	CUGGGUGCGGGUUAUCCGUACGAUGUGCCGGAUUAUGCCUAA		13576	13576	0

CreiLOV DNA Template Sequences

Name	nt	5′	Sequence (5′ → 3′)	3′	EM	MM	MD
CreiLOV-D1	138	p	TTACGCCAGCGCCTTGCCCTCCGTCTTGCTGGTCACATCCACCTGCACGCCCACGAACTT GGACACCCGGCCGTCGGGCGTCTTGATGGGCGTGACGGTGAGCAGGTTCCAAAAGGGCGT GCCGTCCTTGCGGTAGTT		42537	42538	+1
CreiLOV-D2	136	p	GAGCAGGCGCACGGAGCAGGCCTCGCCCTTCTTGATGGCGTCCCTGATCTTCTGGACCTC CTTGGGGTCCGTGCCCTCGCCTTGGAGGAAGCGTGCGTTGTGACCAAGCACCTCATCGGG TCCATATCCAGTCATG		41934	41935	+1
CreiLOV-D3	136	p	GCATAGAACCCCTCGCTGGCGTAGACCAGCGGACAGTCCGGAAGCGTGGCATCTGCGACG ACGAAAGTATGCCGCAAGCCAGCCATATTGTTACCTCCTTATAAAGTTAACCCTATAGTG AGTCGTATTAGGTTGC		42039	42038	-1
CreiLOV-T	48		CAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAG		14740	14741	+1
CreiLOV-S12	40		CCTGCTCCGTGCGCCTGCTCAACTACCGCAAGGACGGCAC		12164	12163	-1
CreiLOV-S23	40		GCCAGCGAGGGGTTCTATGCCATGACTGGATATGGACCCG		12378	12378	0
CreiLOV-ST1	40		AGGGCAAGGCGCTGGCGTAACTAGCATAACCCCTTGGGGC		12372	12372	0
CreiLOV-P	60		GCAACCTAATACGACTCACTATAGGGTTAACTTTATAAGGAGGTAACAATATGGCTGGCT		18530	18529	-1

Luciferase UTR Sequences

Note: The luciferase transcript was too long to perform mass spectrometry measurements on.

Name	nt	5′	Sequence (5′ → 3′)	3′	EM	MM	MD
Luc-5'UTR-trunc primer	74		GCGAATTAATACGACTCACTATAGGGCTTAAGTATAAGGAGGAAAAAATATGAAACATCAACATCAGCACCAAG		22901	22902	+1
Luc-3'UTR-trunc primer	54		AAACCCCTCCGTTTAGAGAGGGGTTATGCTAGTTAAACCTGCTCGTTCTTCAGC		16563	16564	+1
luciferase transcript with long UTRs (transcribed)	1173	ppp	GGCAAUUUUAACUUUACUAAGGAGAAUUCACCAUGAAACAUCAACAUCAGCACCAAGCGAUCGCCAUGGCUUCCAAGGUGUACGACCCCGAGCAACGCAAACGCAUGAUCACUGGGCCUCAGUGGUGGGCUCGCUGCAAGCAAAUGAACGUGCUGGACUCCUUCAUCAACUACUAUGAUUCCGAGAAGCACGCCGAGAACGCCGUGAUUUUUCUGCAUGGUAACGCUGCCUCCAGCUACCUGUGGAGGCACGUCGUGCCUCACAUCGAGCCCGUGGCUAGAUGCAUCAUCCCUGAUCUGAUCGGAAUGGGUAAGUCCGGCAAGAGCGGGAAUGGCUCAUAUCGCCUCCUGGAUCACUACAAGUACCUCACCGCUUGGUUCGAGCUGCUGAACCUUCCAAAGAAAAUCAUCUUUGUGGGCCACGACUGGGGGGCUUGUCUGGCCUUUCACUACUCCUACGAGCACCAAGACAAGAUCAAGGCCAUCGUCCAUGCUGAGAGUGUCGUGGACGUGAUCGAGUCCUGGGACGAGUGGCCUGACAUCGAGGAGGAUAUCGCCCUGAUCAAGAGCGAAGAGGGCGAGAAAAUGGUGCUUGAGAAUAACUUCUUCGUCGAGACCAUGCUCCCAAGCAAGAUCAUGCGGAAACUGGAGCCUGAGGAGUUCGCUGCCUACCUGGAGCCAUUCAAGGAGAAGGGCGAGGUUAGACGGCCUACCCUCUCCUGGCCUCGCGAGAUCCCUCUCGUUAAGGGAGGCAAGCCCGACGUCGUCCAGAUUGUCCGCAACUACAACGCCUACCUUCGGGCCAGCGACGAUCUGCCUAAGAUGUUCAUCGAGUCCGACCCUGGGUUCUUUUCCAACGCUAUUGUCGAGGGAGCUAAGAAGUUCCCUAACACCGAGUUCGUGAAGGUGAAGGGCCUCCACUUCAGCCAGGAGGACGCUCCAGAUGAAAUGGGUAAGUACAUCAAGAGCUUCGUGGAGCGCGUGCUGAAGAACGAGCAGGUUUAAACGAAUUCGGGCUCGGUACCCGGGGAUCCUCUAGAGUCGACCUGCAGGCAUGCAAGCUGAUCCGGCUGCUAACAAAGCCCGAAAGGAAGCUGAGUUGGCUGCUGCCACCGCUGAGCAAUAACUAGCAUAACCCCUUGGGGCCUCUAAACGGGUCUUGAGGGGUUUUUUG	--	--	--	--

His-tagged Luciferase RNA Template Sequences

Note: The HisLuc-R3 transcript was too long to perform mass spectrometry measurements on.

Name	nt	5′	Sequence (5′ → 3′)	3′	EM	MM	MD
HisLuc-R1-c	44		GGCAAUUUUAACUUUACUAAGGAGAAUUCACCAUGAAACAUCAC	alk	14065	14064	-1
HisLuc-R2-c	32	az	CAUCACCAUCACGCGAUCGCCAUGGCUUCCAA		10127	10127	0
HisLuc-S12	40		GCGATCGCGTGATGGTGATGGTGATGTTTCATGGTGAATT		12484	12484	0
HisLuc-S23	40		GTTGCTCGGGGTCGTACACCTTGGAAGCCATGGCGATCGC		12345	12347	+2
Luc-P	34		TAATACGACTCACTATAGGTGTACGACCCCGAGC		10396	10397	+1
HisLuc-R1-e	44		GGCAAUUUUAACUUUACUAAGGAGAAUUCACCAUGAAACAUCAC		14027	14026	-1
HisLuc-R2-e	32	p	CAUCACCAUCACGCGAUCGCCAUGGCUUCCAA		10182	10182	0
HisLuc-R3 (transcribed)	1097	P	GGUGUACGACCCCGAGCAACGCAAACGCAUGAUCACUGGGCCUCAGUGGUGGGCUCGCUGCAA GCAAAUGAACGUGCUGGACUCCUUCAUCAACUACUAUGAUUCCGAGAAGCACGCCGAGAACGC CGUGAUUUUUCUGCAUGGUAACGCUGCCUCCAGCUACCUGUGGAGGCACGUCGUGCCUCACAU CGAGCCCGUGGCUAGAUGCAUCAUCCCUGAUCUGAUCGGAAUGGGUAAGUCCGGCAAGAGCGG GAAUGGCUCAUAUCGCCUCCUGGAUCACUACAAGUACCUCACCGCUUGGUUCGAGCUGCUGAA CCUUCCAAAGAAAAUCAUCUUUGUGGGCCACGACUGGGGGGGCUUGUCUGGCCUUUCACUACUC CUACGAGCACCAAGACAAGAUCAAGGCCAUCGUCCAUGCUGAGAGUGUCGUGGACGUGAUCGA GUCCUGGGACGAGUGGCCUGACAUCGAGGAGGAUAUCGCCCUGAUCAAGAGCGAAGAGGGCGA GAAAAUGGUGCUUGAGAAUAACUUCUUCGUCGAGACCAUGCUCCCAAGCAAGAUCAUGCGGAA ACUGGAGCCUGAGGAGUUCGCUGCCUACCUGGAGCCAUUCAAGGAGAAGGGCGAGGUUAGACG GCCUACCCUCUCCUGGCCUCGCGAGAUCCCUCUCGUUAAGGGAGGCAAGCCCGACGUCGUCCA GAUUGUCCGCAACUACAACGCCUACCUUCGGGCCAGCGACGAUCUGCCUAAGAUGUUCAUCGA GUCCGACCCUGGGUUCUUUUCCAACGCUAUUGUCGAGGGAGCUAAGAAGUUCCCUAACACCGA GUUCGUGAAGGUGAAGGGCCUCCACUUCAGCCAGGAGGACGCUCCAGAUGAAAUGGGUAAGUA CAUCAAGAGCUUCGUGGAGCGCGUGCUGAAGAACGAGCAGGUUUAAACGAAUUCGGGCUCGGU ACCCGGGGAUCCUCUAGAGUCGACCUGCAGGCAUGCAAGCUGAUCCGGCUGCUAACAAAGCCC GAAAGGAAGCUGAGUUGGCUGCUGCCACCGCUGAGCAAUAACUAGCAUAACCCCUUGGGGCCU CUAAACGGGUCUUGAGGGGUUUUUUG		--	--	--
Luc-5′tailed primer	34		TAATACGACTCACTATAGGTGTACGACCCCGAGC		10396	10396	0
Luc-3′ primer	21		CAAAAAACCCCTCAAGACCCG		6322	6323	+1

Phosphoramidate Ligation FLAGx3 RNA Template Sequences

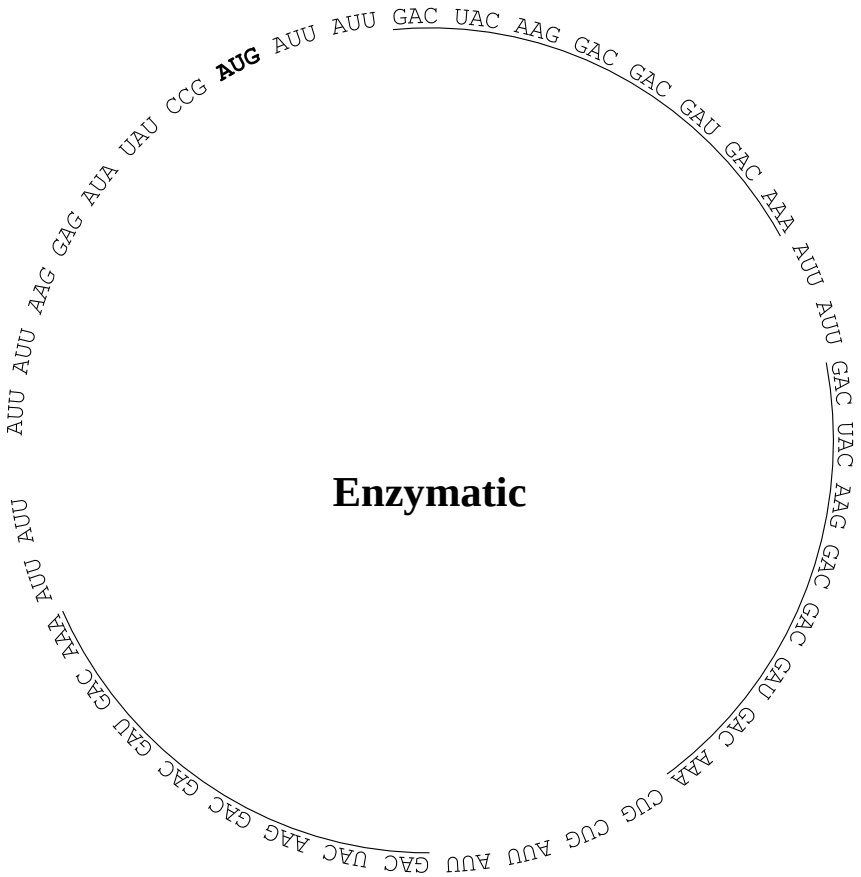
Note: PA-CU31-1* has a DNA monomer on the 3′ terminus, denoted by lowercase ‘c’ (i.e. no terminal 2′-OH).

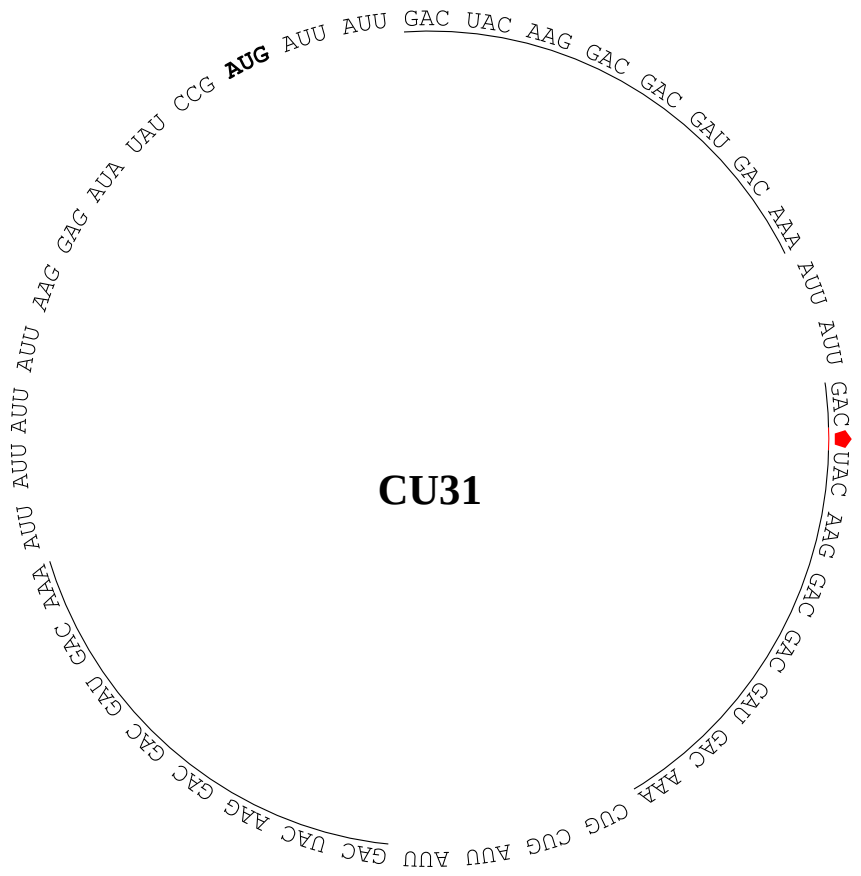
Name	nt	5′	Sequence (5′ → 3′)	3′	EM	MM	MD
PA-CU31-1	63	p	AUUAUUAAGGAGAUAUAUCCGAUGAUUAUUGACUACAAGGACGACGAUGACAAAAUUAUUGAC	p	20423	20422	-1
PA-CU31-2	63	am	TACAAGGACGACGAUGACAAACUGCUGAUUAUUGACUACAAGGACGACGAUGACAAAAUUAUU		20297	20297	0
PA-CU31-1*	63	p	AUUAUUAAGGAGAUAUAUCCGAUGAUUAUUGACUACAAGGACGACGAUGACAAAAUUAUUGAc	p	20407	20404	-3
PA-DNA-1	24		TGCAATCACAGGTTCGTGACGGCT	p	7448	7449	+1
PA-DNA-2	61	am	TCATAAGTCTCAACCTCCTAGTCCTATAGTGAGTCGTATTAGTGCCTGCCAAATCTGTTAG		18678	18678	0
PA-DNA-S	34	Cy3	GAGACTTATGGAGCCGTCACGAACCTGTGATTGC		10991	10989	-2

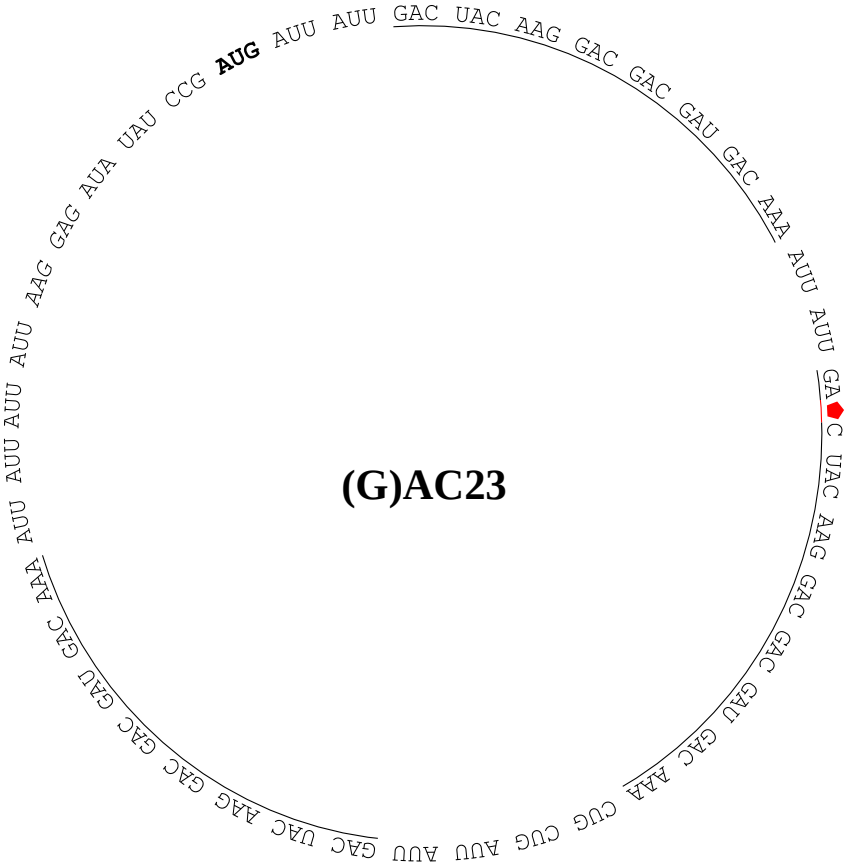
RCT FLAGx3 RNA Template Sequences (diagrams of circularised RNA shown below)

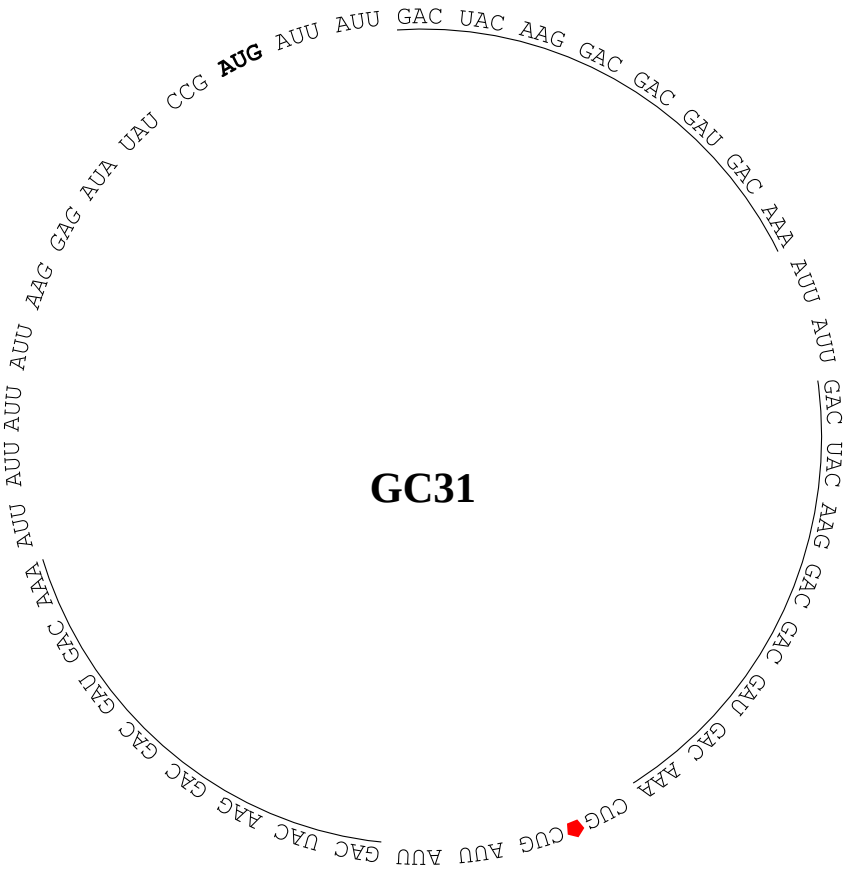
Start codons (AUG) are in bold.

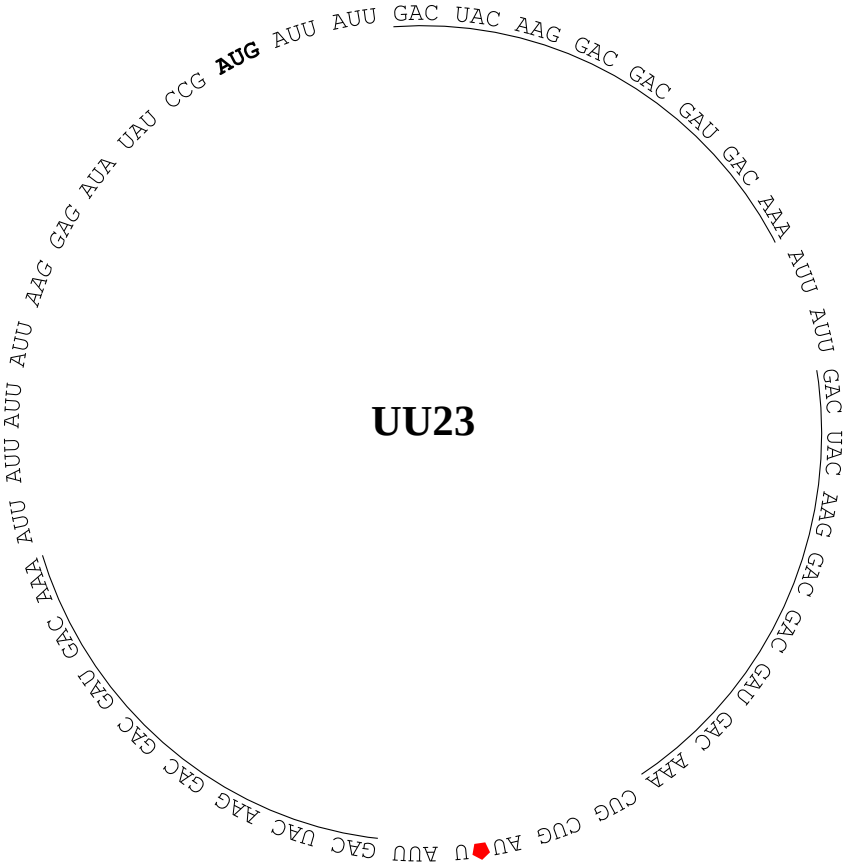
Name	nt	5′	Sequence (5′ → 3′)	3′	EM	MM	MD
RCT-enz-1	63	p	AUUAUU*AAGGAG*AUAUAUCCG**AUG**AUUAUUGACUACAAGGACGACGAUGACAAAAUUAUUGAC		20343	20342	-1
RCT-enz-2	63	p	UACAAGGACGACGAUGACAAACUGCUGAUUAUUGACUACAAGGACGACGAUGACAAAAUUAUU		20379	20379	0
RCT-CU31-1	63	p	AUUAUU*AAGGAG*AUAUAUCCG**AUG**AUUAUUGACUACAAGGACGACGAUGACAAAAUUAUUGAC	alk	20381	20381	0
RCT-CU31-2	63	az	UACAAGGACGACGAUGACAAACUGCUGAUUAUUGACUACAAGGACGACGAUGACAAAAUUAUU		20324	20323	-1
RCT-GAC23-1	62	p	AUUAUU*AAGGAG*AUAUAUCCG**AUG**AUUAUUGACUACAAGGACGACGAUGACAAAAUUAUUGA	alk	20076	20077	+1
RCT-GAC23-2	64	az	CUACAAGGACGACGAUGACAAACUGCUGAUUAUUGACUACAAGGACGACGAUGACAAAAUUAUU		20630	20629	-1
RCT-GC31-1	63	p	AUUAUUGACUACAAGGACGACGAUGACAAAAUUAUUGACUACAAGGACGACGAUGACAAACUG	alk	20440	20440	0
RCT-GC31-2	63	az	CUGAUUAUUGACUACAAGGACGACGAUGACAAAAUUAUUAUUAUU*AAGGAG*AUAUAUCCG**AUG**		20265	20264	-1
RCT-UU23-1	68	p	AUUAUUGACUACAAGGACGACGAUGACAAAAUUAUUGACUACAAGGACGACGAUGACAAACUGC UGAU	alk	22032	22030	-2
RCT-UU23-2	58	az	UAUUGACUACAAGGACGACGAUGACAAAAUUAUUAUUAUU*AAGGAG*AUAUAUCCG**AUG**		18673	18674	+1
RCT-AC31-1	63	p	**AUG**AUUAUUGACUACAAGGACGACGAUGACAAAAUUAUUGACUACAAGGACGACGAUGACAAA	alk	20464	20463	-1
RCT-AC31-2	63	az	CUGCUGAUUAUUGACUACAAGGACGACGAUGACAAAAUUAUUAUUAUU*AAGGAG*AUAUAUCCG		20241	20239	-2
RCT-UAC23-1	64	p	UUAUU*AAGGAG*AUAUAUCCG**AUG**AUUAUUGACUACAAGGACGACGAUGACAAAAUUAUUGACUA	alk	20687	20686	-1
RCT-UAC23-2	62	az	CAAGGACGACGAUGACAAACUGCUGAUUAUUGACUACAAGGACGACGAUGACAAAAUUAUUA		20018	20018	0
RCT-SA12	21		GTCCTTGTAGTCAATAATTTT		6401	6401	0
RCT-SA21	24		CTCCTTAATAATAATAATTTTGTC		7276	7275	-1
RCT-SA12-long	40		TTGTCATCGTCGTCCTTGTAGTCAATAATTTTGTCATCGT		12224	12223	-1
RCT-SA21-long	40		GGATATATCTCCTTAATAATAATAATTTTGTCATCGTCGT		12249	12248	-1
RCT-SB21	25		GTAGTCAATAATCATCGGATATATC		7664	7664	0
RCT-SB12	38		CCTTGTAGTCAATAATCAGCAGTTTGTCATCGTCGTCC		11595	11595	0

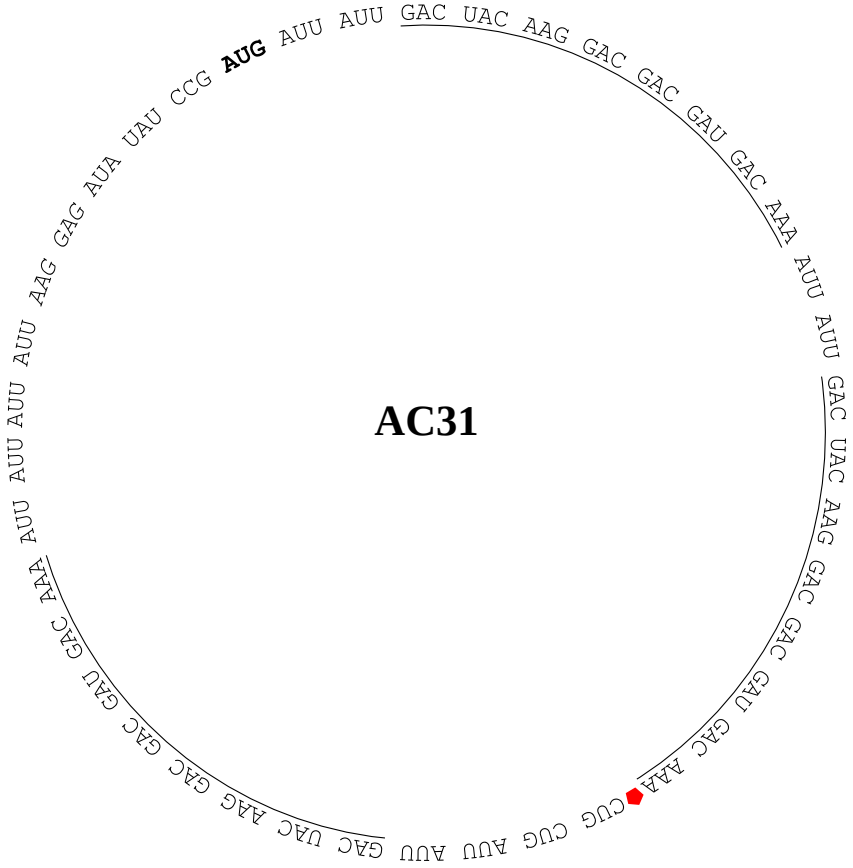

AUG AUU AUU GAC UAC AAG GAC GAC GAU GAC AAA AUU AUU GAC UAC AAG GAC GAC GAU GAC AAA CUG CUG AUU AUU GAC UAC AAG GAC GAC GAU GAC AAA AUU AUU AUU AAG GAG AUA UAU CCG
Enzymatic

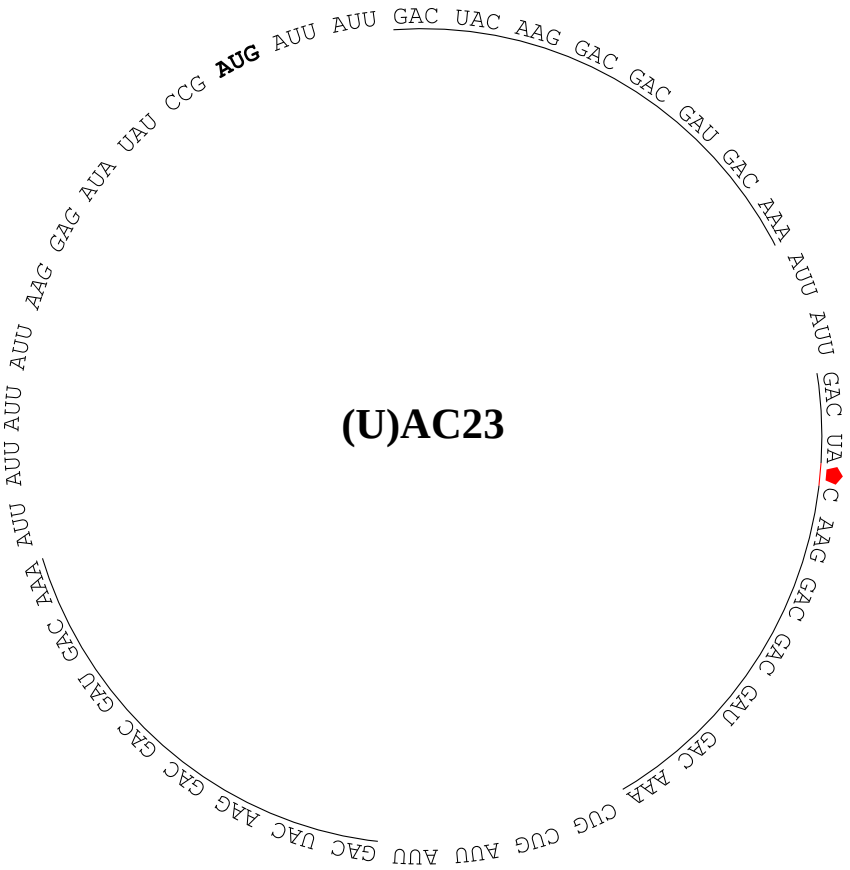

AUG AUU AUU GAC UAC AAG GAC GAC GAU GAC AAA AUU AUU GAC UAC AAG GAC GAC GAU GAC AAA CUG CUG AUU AUU GAC UAC AAG GAC GAC GAU GAC AAA AUU AUU AUU AAG GAG AUA UAU CCG
CU31


GAC UAC AAG GAC GAC GAU GAC AAA CUG CUG AUU AUU GAC UAC AAG GAC GAC GAU GAC AAA AUU AUU GAC UAC AAG GAC GAC GAU GAC AAA AUU AUU AUG CCG UAU AUA GAG AAG AUU AUU AUU AUU GAC GAC GAU GAC GAC UAC AAG GAC GAC AAA AUU AUU
(G)AC23


GAC UAC AAG GAC GAC GAU GAC AAA CUG CUG AUU AUU GAC UAC AAG GAC GAC GAU GAC AAA AUU AUU GAC UAC AAG GAC GAC GAU GAC AAA AUU AUU AUG CCG UAU AUA GAG AAG AUU AUU AUU AUU GAC GAC GAU GAC GAC UAC AAG GAC GAC AAA AUU AUU
GC31


AUG AUU AUU GAC UAC AAG GAC GAC GAU GAC AAA AUU AUU GAC UAC AAG GAC GAC GAU GAC AAA CUG CUG AUU AUU GAC UAC AAG GAC GAC GAU GAC AAA AUU AUU AUU AUU AAG GAG AUA UAU CCG
UU23


AUG AUU AUU GAC UAC AAG GAC GAC GAU GAC AAA AUU AUU GAC UAC AAG GAC GAC GAU GAC AAA CUG CUG AUU AUU GAC UAC AAG GAC GAC GAU GAC AAA AUU AUU AUU AUU AAG GAG AUA UAU CCG
AC31



Diagrams of the circular mRNA templates showing the codon reading frame: the RBS is italicized, the AUG start codon is in bold, the FLAG tag codons (coding for DYKDDDDK) are underlined, and the triazole linker is signified by '⬟.'

Appendix: Supplementary data

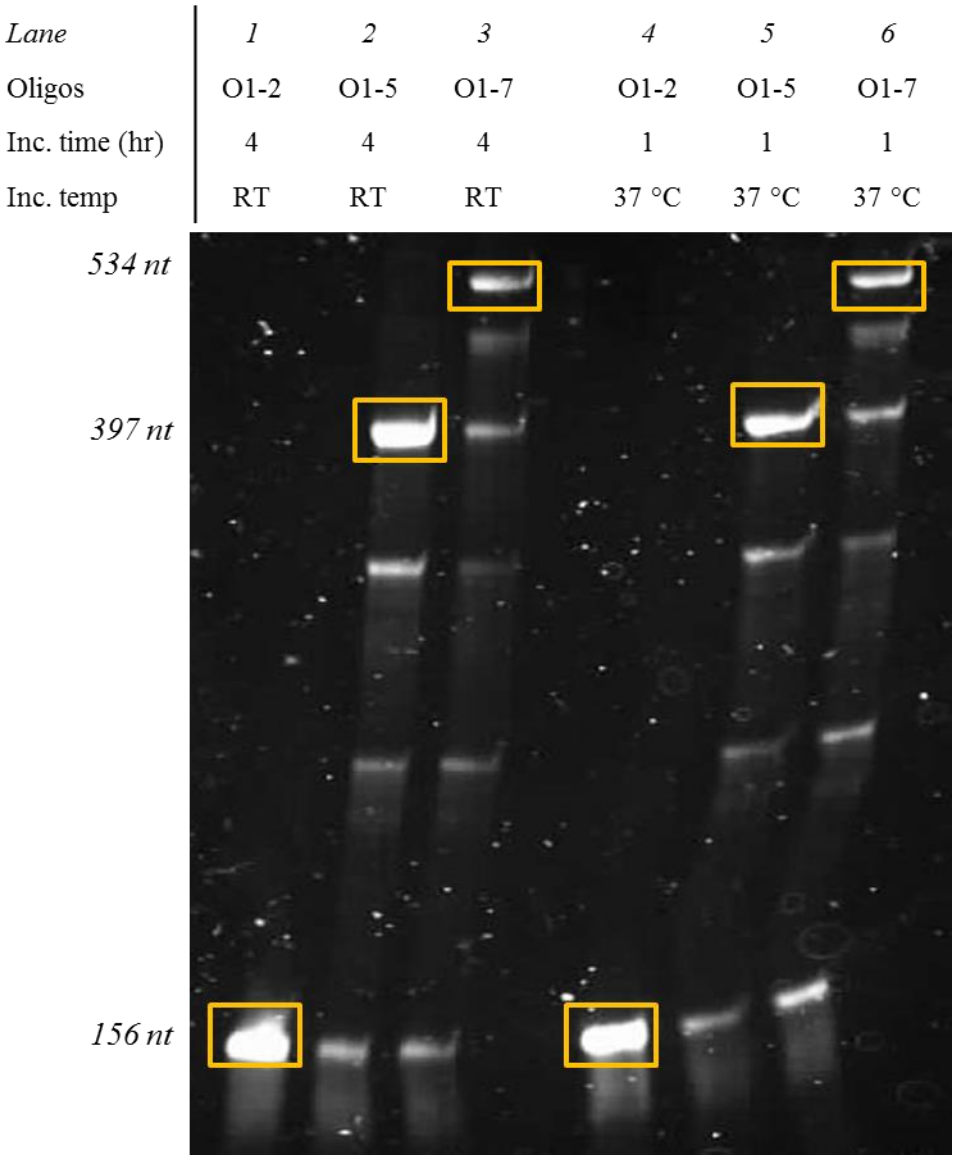


Figure A1. 3.5% denaturing PAGE gel analysing two incubation (inc.) conditions for the enzymatic ligation of templated ssDNA: 4 hr at RT (**Lanes 1-3**), or 1 hr at 37 °C (**Lanes 4-6**). The gel was visualised by SYBR™ Gold stain.

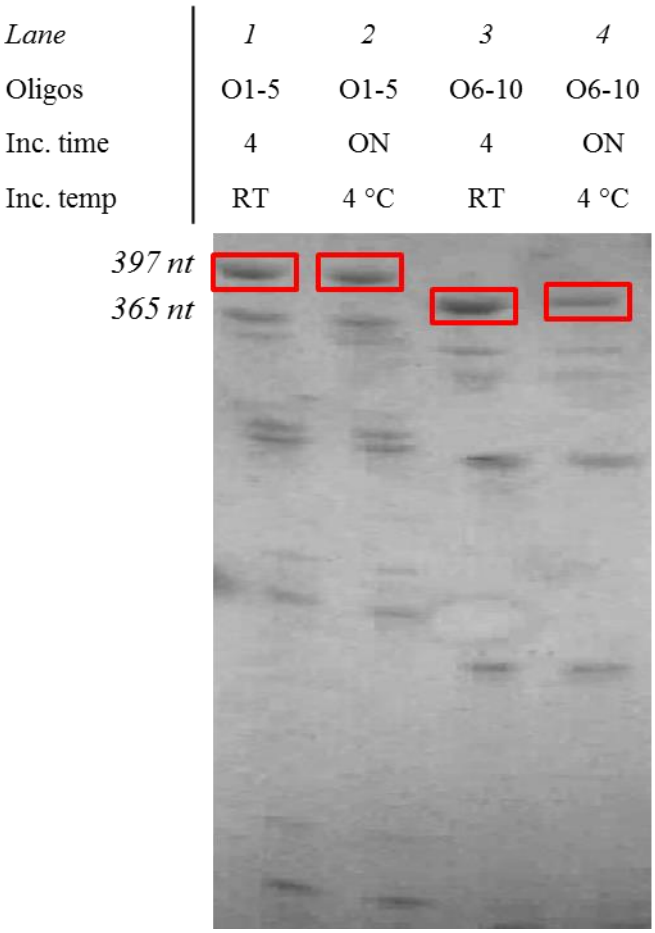


Figure A2. 5% denaturing PAGE gel analysing two incubation (inc.) conditions for the enzymatic ligation of templated ssDNA: 4 hr at RT (**Lanes 1** and **3**), or ON at 4 °C (**Lanes 2** and **4**).

Lane	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*
Oligos	O1-5	O1-5	O1-5	O1-5	O1-5	O1-5	O1-5	O6-10	O6-10	O6-10	O6-10	O6-10	O6-10	O6-10
Variable		2x ligase	2:1 (O:S)	1:2 (O:S)	1 cycle	3 cycles	5 cycles		2x ligase	2:1 (O:S)	1:2 (O:S)	1 cycle	3 cycles	5 cycles

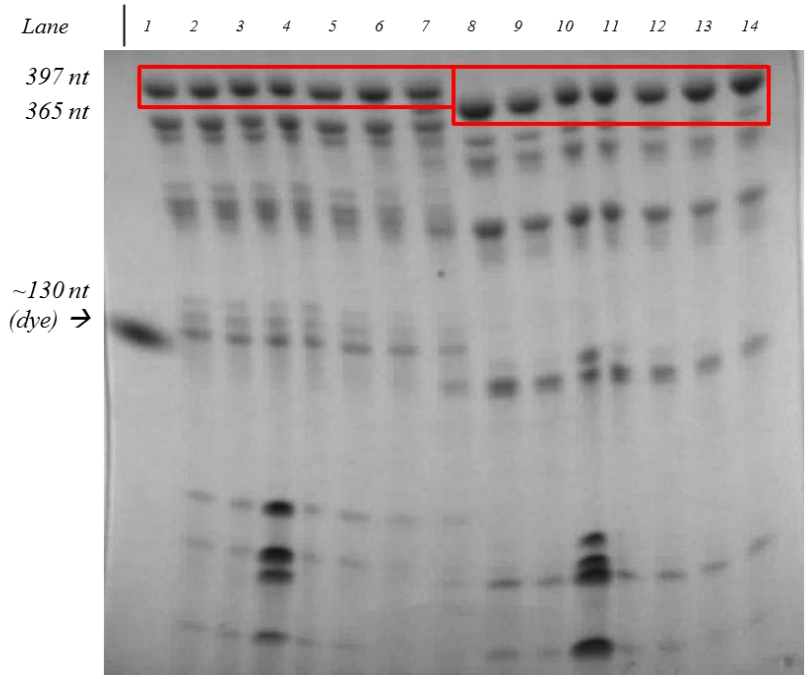


Figure A3. 5% denaturing PAGE gel analysing different conditions for the enzymatic ligation of templated ssDNA. In the control reactions (**Lanes 1** and **8**), a 1:1 oligonucleotide to splint ratio (O:S) was used, and the samples were incubated for 4 hr at RT. These conditions were used for the other reactions unless otherwise specified. **Lanes 2** and **9**: double the amount of ligase; **Lanes 3** and **10**: double the amount of oligo; **Lanes 4** and **11**: double the amount of splint. **Lanes 5** and **12**: 1 incubation cycle of 1 hr at 37 °C followed by 5 min at 95 °C; **Lanes 6** and **13**: 3 incubation cycles, with fresh ligase added before each cyle; **Lanes 7** and **14**: 5 incubation cycles, with fresh ligase added before each cycle.

Lane	1	2	3	4	5	6	7	8	9
Variable		30 min, 37 °C	PNK	Fast annealing	No annealing	20 mM NaCl	13 mM $MgCl_2$	16 mM $MgCl_2$	2 mM ATP

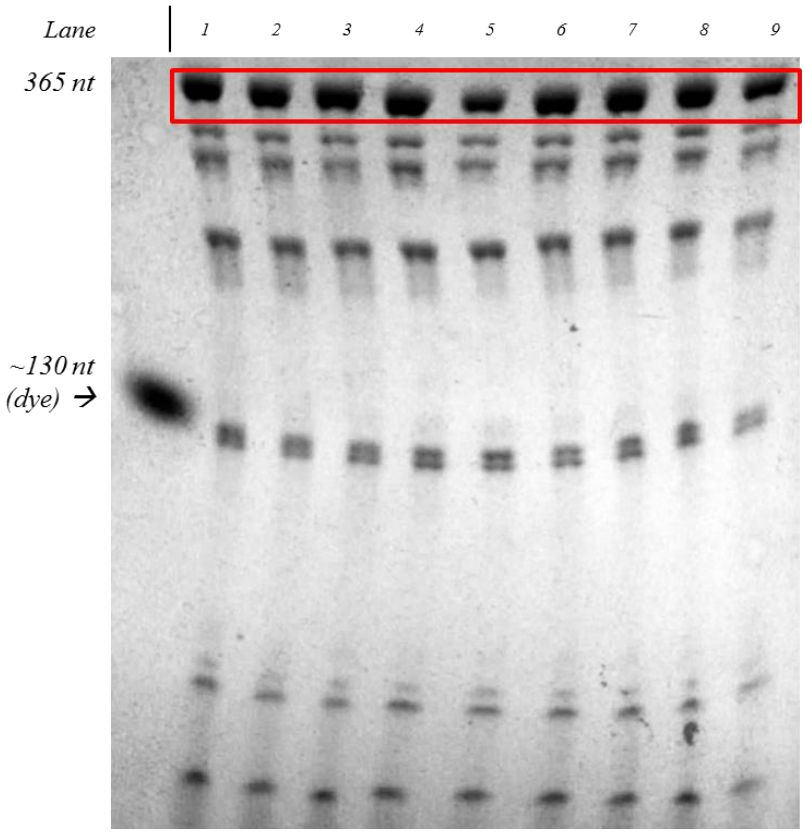


Figure A4. 5% denaturing PAGE gel analysing different conditions for the enzymatic ligation of templated ssDNA (O1-6). In the control reaction (**lane 1**), samples were annealed by heating for 5 min at 95 °C and left to cool to RT; the reaction buffer contained 10 mM $MgCl_2$, 1 mM ATP, and no NaCl. These conditions were used unless specified otherwise. **Lane 2**: the ligation reaction was incubated for 30 min at 37 °C. **Lane 3**: the oligonucleotides were treated with polynucleotide kinase (PNK) prior to adding ligase; the PNK reaction was incubated for 30 min at 37 °C and then heat inactivated for 20 min at 65 °C. **Lane 4**: fast annealing was performed by placing the oligonucleotides in a -20 °C freezer after heating for 5 min at RT. **Lane 5**: the annealing step was omitted. **Lanes 6-9**: the buffer composition was changed to have final concentrations of 20 mM NaCl, 13 mM $MgCl_2$, 16 mM $MgCl_2$, or 2 mM ATP, respectively. The DNA oligonucleotides used in the ligation reaction were O6-10.

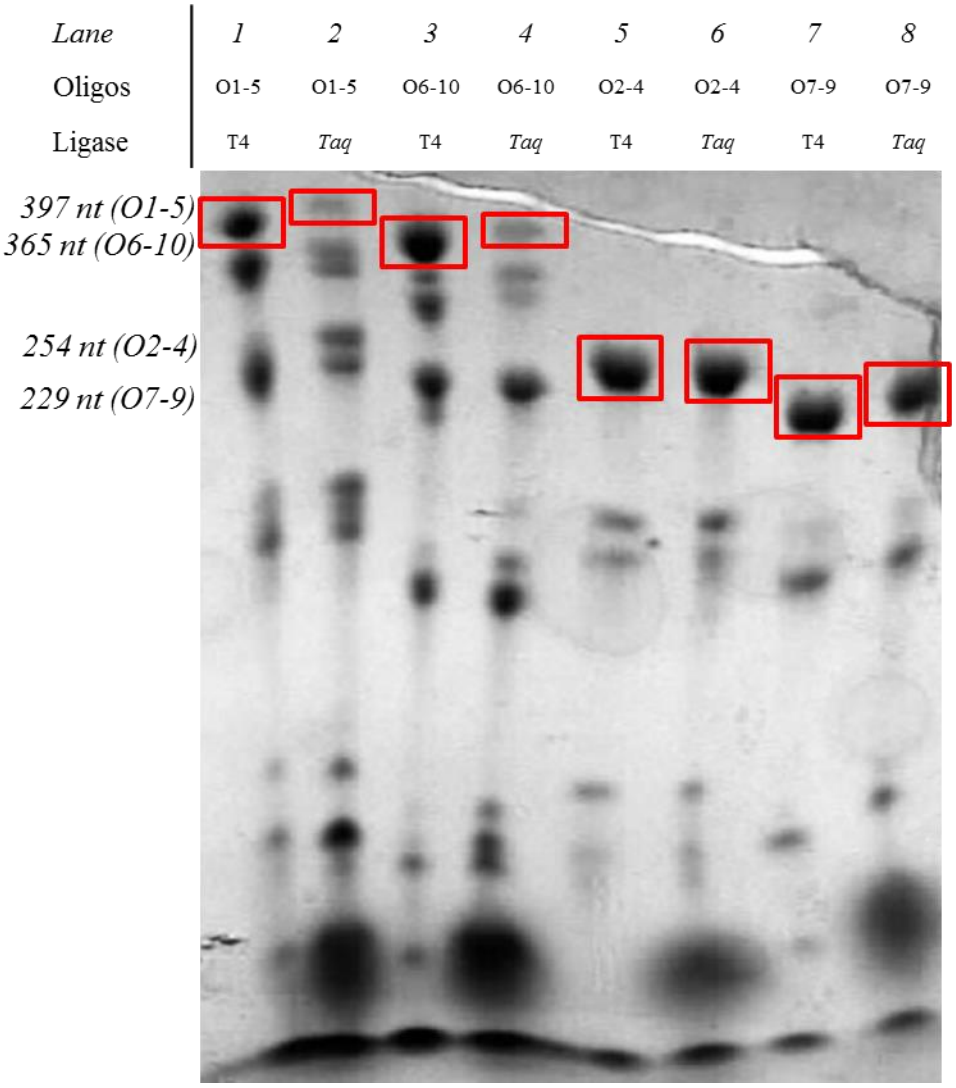


Figure A5. 5% denaturing PAGE gel comparing T4 and *Taq* DNA ligases for the enzymatic ligation of templated ssDNA. T4 (**Lanes 1**, **3**, **5**, and **7**) and *Taq* (**Lanes 2**, **4**, **6**, and **8**) DNA ligations were incubated overnight at RT or 55 °C, respectively.

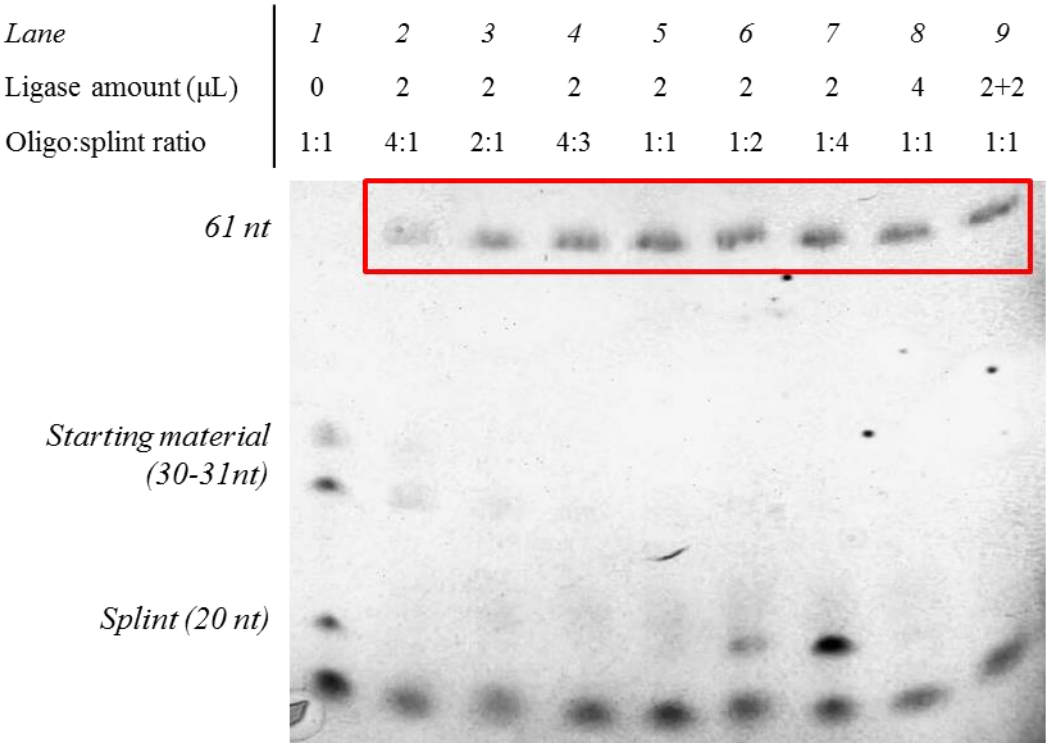


Figure A6. 12.5% denaturing PAGE gel investigating oligo:splint ratios and ligase amounts for the templated ligation of ssDNA with *Taq* DNA ligase. Reactions were incubated for 20 thermal cycles, with one cycle consisting of 15 min at 45 °C followed by 5 min at 85 °C. In **Lane 9**, fresh ligase was added after the 10th cycle. The DNA oligonucleotides used in the ligation reaction were O1a (31 nt) and O1b (30 nt).

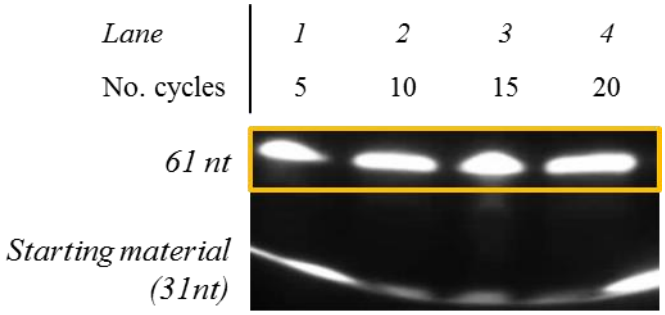


Figure A7. 5% denaturing PAGE gel testing different incubation conditions for the templated ligation of ssDNA with *Taq* DNA ligase. One thermal cycle consisted of 15 min at 45 °C followed by 5 min at 85 °C. The DNA oligonucleotides used in the ligation reaction were O1a (31 nt) and O1b (30 nt).

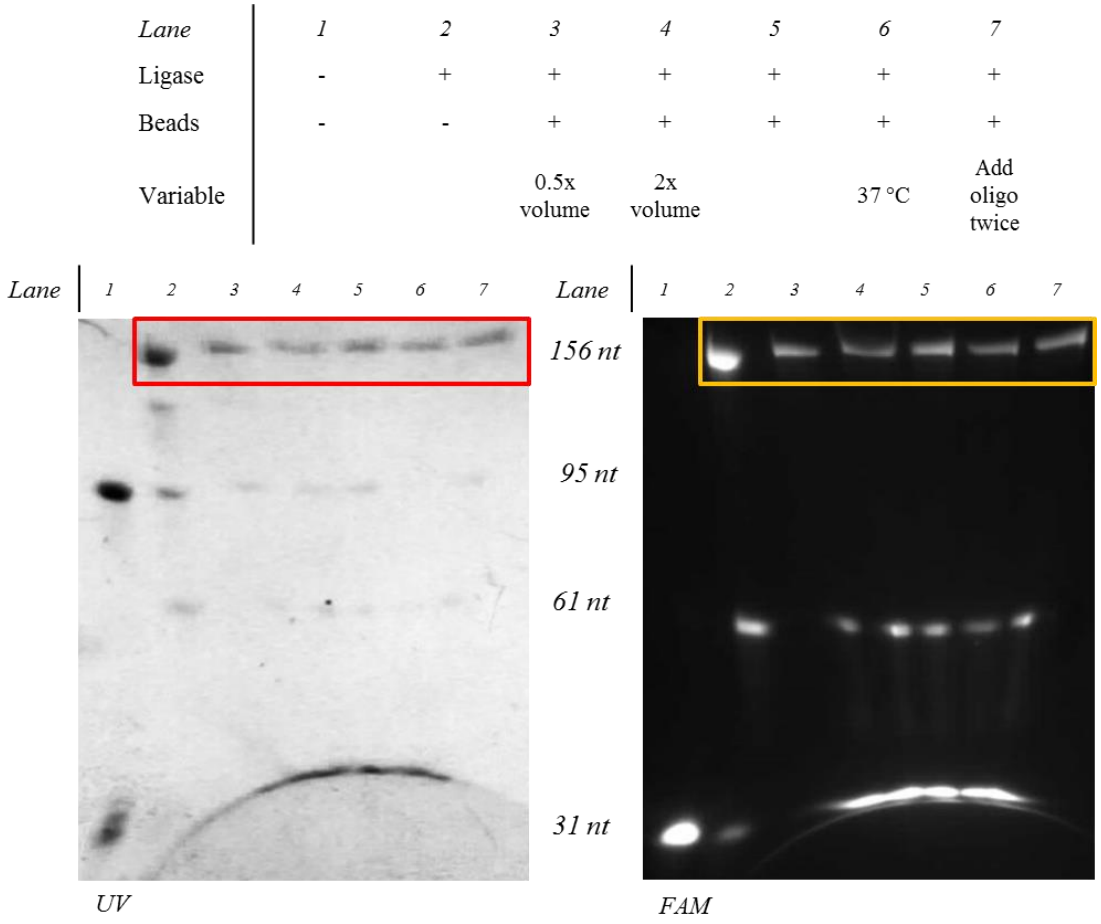


Figure A8. 10% denaturing PAGE gel investigating T4 DNA ligase reactions using Dynabeads™ M-280 streptavidin beads (Thermo Fisher) as a solid support. **Lanes 1-2** were performed in solution without beads. The reactions were incubated at RT, except for **Lane 6**. The DNA oligonucleotides used in the ligation reaction were O1a (31 nt), O1b (30 nt), and O2 (95 nt). The gels were visualised by UV absorbance (left), and by a FAM fluorophore attached to O1a (right).

Lane	*1*	*2*	*3*	*4*	*5*	*6*	*7*
Ligase	-	+	+	+	+	+	+
Beads	-	-	+	+	+	+	+
Oligos	O1a-4	O1a-4	O1a-4	O1a-4	O1a-4	O1a-4	O1a-2
Variable			ON inc.	2x ligase	BSA wash	Add oligo twice	

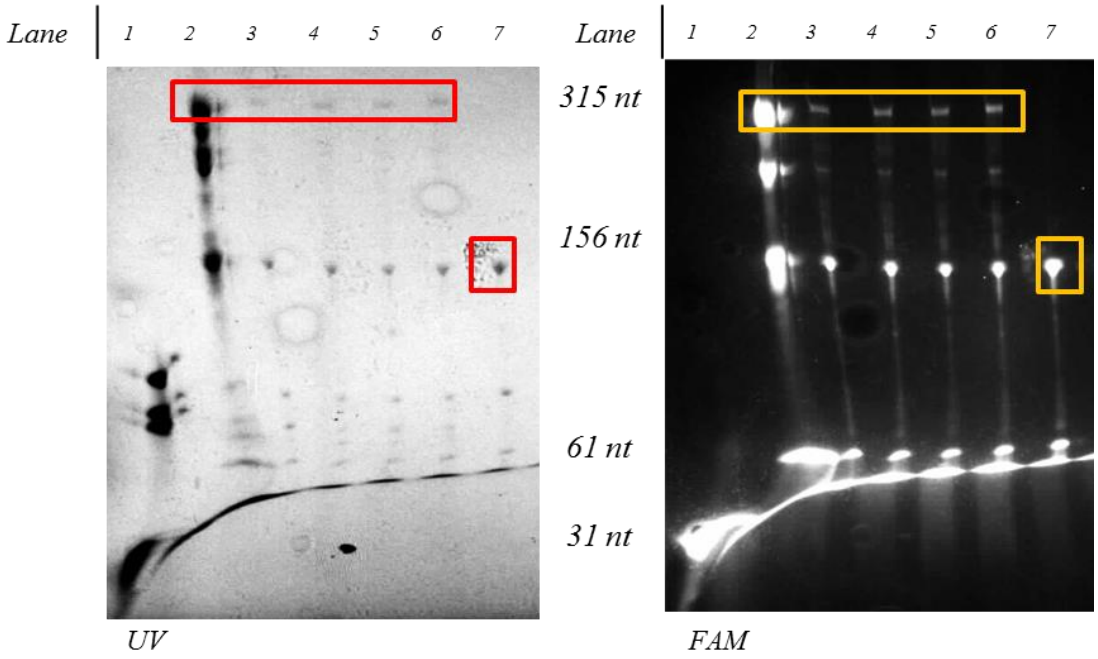


Figure A9. 5% denaturing PAGE gel investigating T4 DNA ligase reactions using Dynabeads™ M-280 streptavidin beads (Thermo Fisher) as a solid support. **Lanes 1-2** were performed in solution without beads. In **Lane 3**, the reaction was incubated overnight; otherwise the reactions were incubated 4 hr at RT. In **Lane 4**, double the amount of ligase was used. In **Lane 5**, 0.1% BSA was added to the wash solution used to wash the beads. In **Lane 6**, each oligonucleotide and splint were added twice, with wash steps in between. The gels were visualised by UV absorbance (left), and by a FAM fluorophore attached to O1a (right).

Lane	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Amount of beads (µL)	0	25	25	25	25	50	50	50	50
Bead type: MyOne C1 (C) or M-280 (M)	--	C	C	M	M	C	C	M	M
Attach (A) or hybridise (H) first	--	A	H	A	H	A	H	A	H

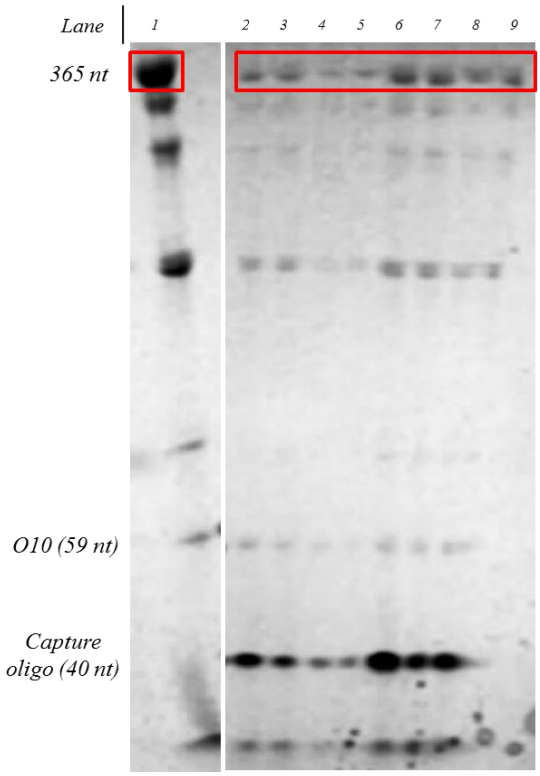


Figure A10. 8% denaturing PAGE gel investigating the use of different magnetic streptavidin beads as a capture mechanism for oligonucleotides. Two bead amounts were used to see how the oligonucleotide capture scaled with the amount of beads used. Two bead types, Dynabeads^{TM} MyOne C1 (C; hydrophilic surface with 1.0 µm diameter) and M-280 (M; hydrophobic surface with 2.8 µm diameter) were compared. The capture oligonucleotide was a 40mer consisting of a 5´-biotin, T_{10} spacer, and 30 nt complementary to the last 30 nt of O10. O6-10 were first ligated in solution. Two methods of capture were tested: the capture oligonucleotide was first attached to the beads and then hybridised with the ligation reaction (attach first – A), or the capture oligonucleotide was hybridised with the ligation reaction in solution first and then bound to the beads (hybridise first – H).

Lane	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Amount of beads (µL)	25	25	25	25	50	50	50	50
Bead type: MyOne C1 (C) or M-280 (M)	C	C	M	M	C	C	M	M
Attach (A) or hybridise (H) first	A	H	A	H	A	H	A	H

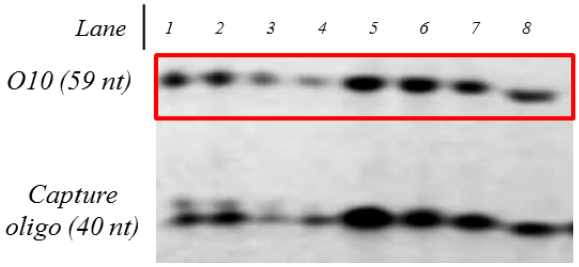


Figure A11. 8% denaturing PAGE gel investigating the use of different magnetic streptavidin beads as a capture mechanism for oligonucleotides. Two bead amounts were used to see how the oligonucleotide capture scaled with the amount of beads used. Two bead types, Dynabeads™ MyOne C1 (C; hydrophilic surface with 1.0 µm diameter) and M-280 (M; hydrophobic surface with 2.8 µm diameter) were compared. The capture oligonucleotide was a 40mer consisting of a 5´-biotin, T_{10} spacer, and 30 nt complementary to the last 30 nt of O10. Two methods of capture were tested: the capture oligonucleotide was first attached to the beads and then hybridised with O10 (attach first – A), or the capture oligonucleotide was hybridised with O10 in solution first and then bound to the beads (hybridise first – H).

Lane	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Ligase	-	DNA	DNA	DNA	DNA	RNA	RNA	RNA	RNA
Inc. time (hr)	4	1	4	16	4	1	4	16	4
25% PEG 4000	-	-	-	-	+	-	-	-	+

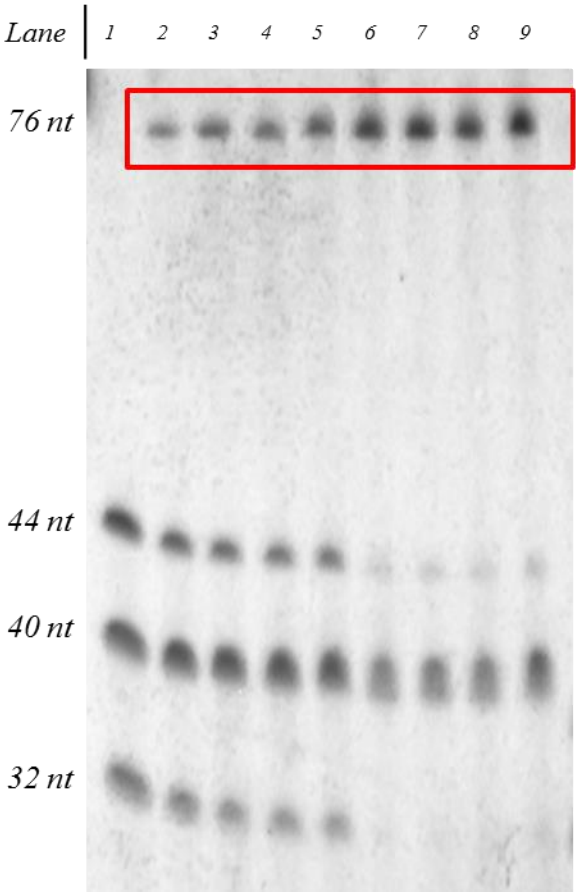
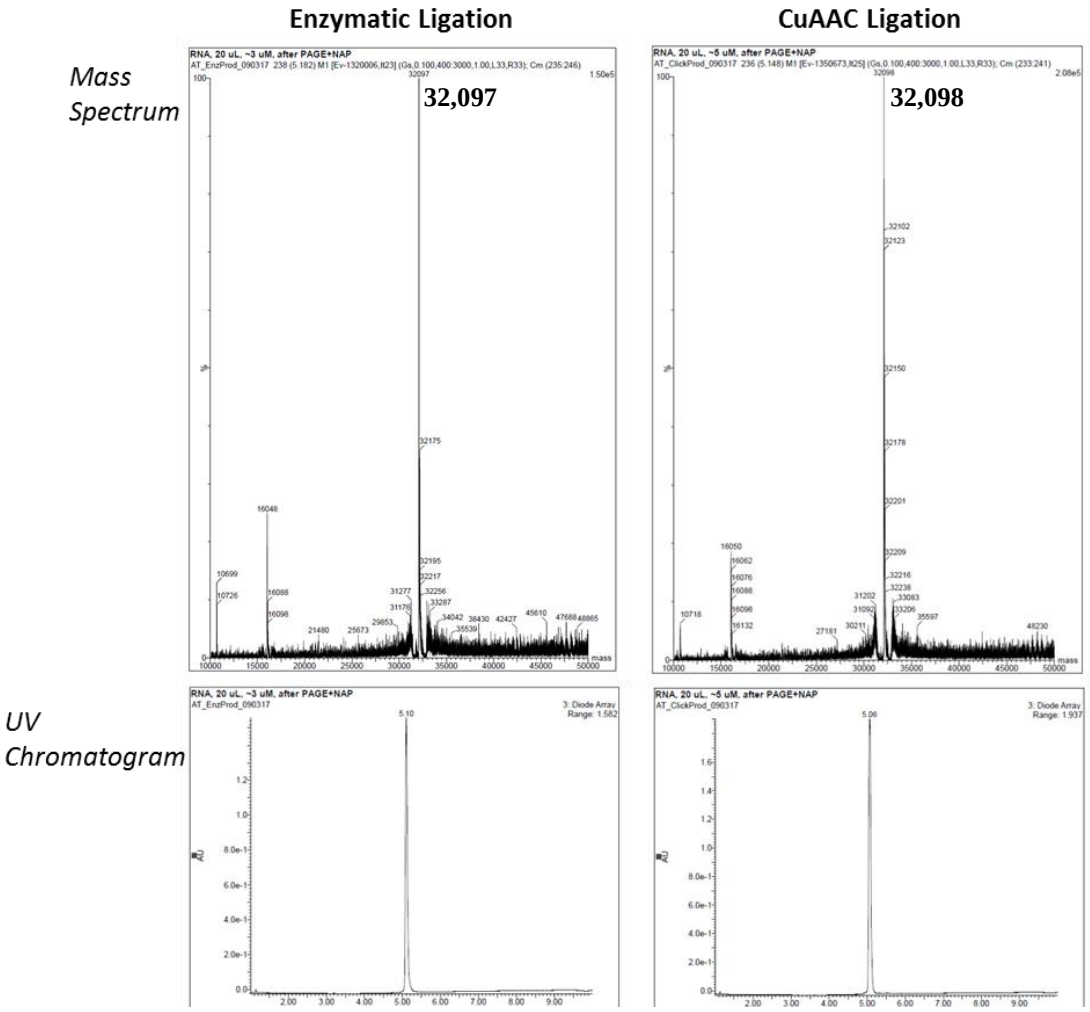
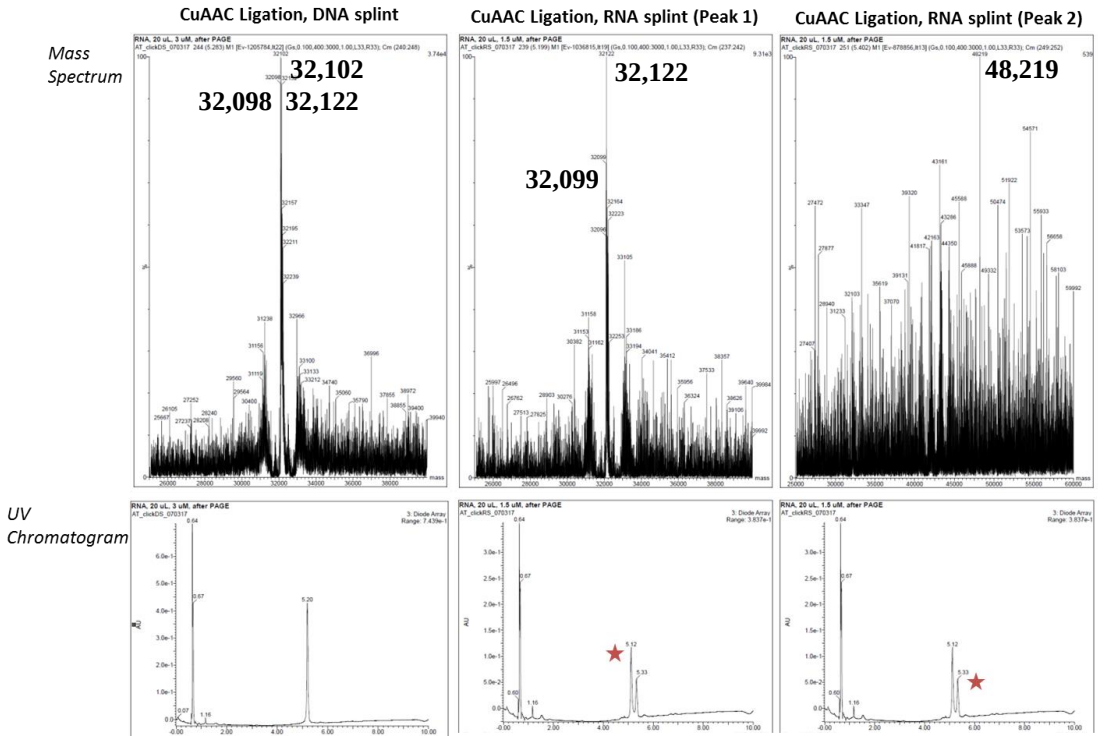


Figure A12. 8% denaturing PAGE gel investigating different incubation times, ligase types, and the addition of 25% PEG 4000 for the templated enzymatic ligation of ssRNA. T4 DNA ligase and T4 RNA ligase 2 were compared. The RNA oligonucleotides used in the ligation reaction were HisLuc-R1-e (44 nt) and HisLuc-R2-e (32 nt), templated with a 40 nt DNA splint (HisLuc-S12).



	Expected Mass	Measured Mass
Enzymatic FLAG-HA RNA ligation product	32,099	32,097
CuAAC FLAG-HA RNA ligation product	32,099	32,098

Figure A13. LC-MS data for RNA ligation products extracted from denaturing PAGE gels. Both the enzymatically and CuAAC-ligated RNA (FLAG-HA RNA template, 99 nt) showed a single sharp peak on the UV chromatogram, and had mass measurements close to the expected mass values.



	Expected Mass	Measured Mass
CuAAC FLAG-HA RNA ligation product – DNA splint	32,099	32,098; 32,102; 32,122
CuAAC FLAG-HA RNA ligation product – RNA splint (peak 1)	32,099	32,099; 32,122
CuAAC FLAG-HA RNA ligation product – RNA splint (peak 1)	32,099	(48,219)

Figure A14. LC-MS data for RNA CuAAC ligation products (FLAG-HA RNA template, 99 nt) extracted from denaturing PAGE gels, templated with a DNA splint (left) or RNA splint (middle and right). While the reaction templated with a DNA splint showed a single sharp peak on the UV chromatogram, the reaction templated with an RNA splinted showed two sharp peaks. The second peak from the reaction templated with an RNA splint did not yield a clear mass measurement.

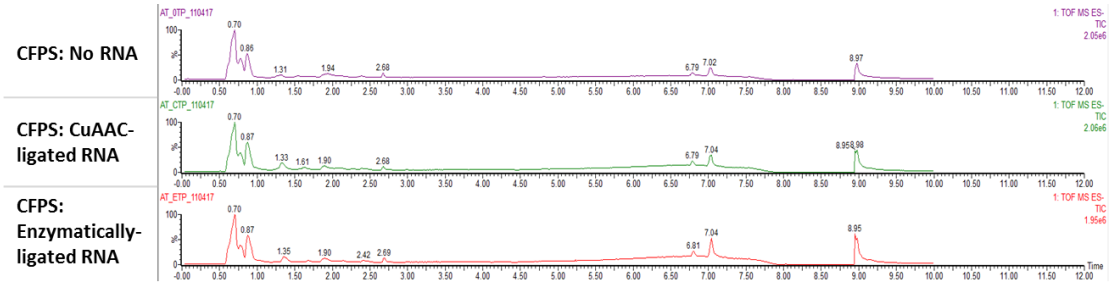


Figure A15. The total ion chromatogram (TIC) from LC-MS analysis of CFPS reactions with no RNA (top, purple), CuAAC-ligated RNA (middle, green), and enzymatically-ligated RNA (bottom, red). No translation product was detected from from the CFPS reactions with the ermCL RNA templates.